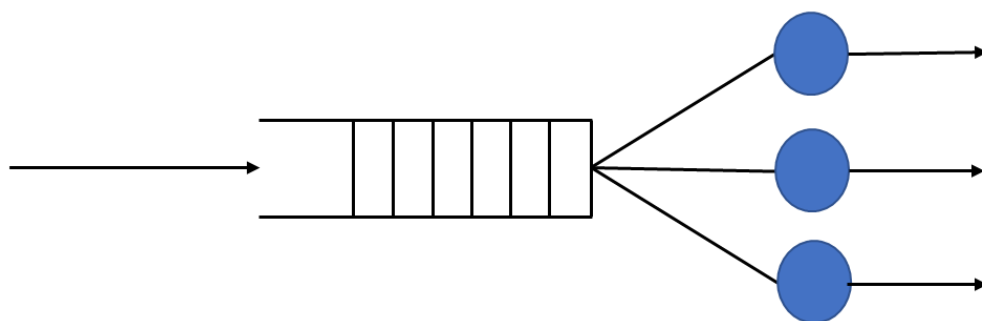


ΘΕΩΡΙΑ ΟΥΡΩΝ ΑΝΑΜΟΝΗΣ

Αντώνιος Θ. Οικονόμου

Καθηγητής

Εθνικό & Καποδιστριακό Πανεπιστήμιο Αθηνών



Αθήνα 2022

ΚΑΛΛΙΠΟΣ

531ΚΙΚΤΕΣ
εκδόσεις
ακαδημαϊκές



ΘΕΩΡΙΑ ΟΥΡΩΝ ΑΝΑΜΟΝΗΣ

Θεωρία Ουρών Αναμονής

Μια σύντομη εισαγωγή

Αντώνιος Θ. Οικονόμου

Καθηγητής

Εθνικό & Καποδιστριακό Πανεπιστήμιο Αθηνών

Τίτλος πρωτοτύπου: «*Θεωρία Ουρών Αναμονής*»

Copyright © 2022, ΚΑΛΛΙΠΟΣ, ΑΝΟΙΚΤΕΣ ΑΚΑΔΗΜΑΪΚΕΣ ΕΚΔΟΣΕΙΣ



Το παρόν έργο διατίθεται με τους όρους της άδειας Creative Commons Αναφορά Δημιουργού – Μη Εμπορική Χρήση – Παρόμοια Διανομή 4.0. Για να δείτε τους όρους της άδειας αυτής επισκεφτείτε τον ιστότοπο <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.el>

Αν τυχόν κάποιο τμήμα του έργου διατίθεται με διαφορετικό καθεστώς αδειοδότησης, αυτό αναφέρεται ρητά και ειδικώς στην οικεία θέση.

Συντελεστές έκδοσης

Γλωσσική επιμέλεια: ???

Γραφιστική επιμέλεια: ???

Τεχνική επεξεργασία: ???

ΚΑΛΛΙΠΟΣ

Εθνικό Μετσόβιο Πολυτεχνείο

Ηρώων Πολυτεχνείου 9

15780 Ζωγράφου

www.kallipos.gr

Βιβλιογραφική αναφορά: Αντώνιος Θ. Οικονόμου, (2022). *Θεωρία Ουρών Αναμονής*. Αθήνα: Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις

Διαθέσιμο στο: <https://www.kallipos.gr/>

ISBN: ???

Αφιερώνεται
στη σύζυγό μου Πόλυ Ρετσίλα
και στα παιδιά μας Θοδωρή και Κυριακή
Βασιλική.

ΠΕΡΙΕΧΟΜΕΝΑ

Πίνακας συντομεύσεων - ακρωνυμίων	ix
Εισαγωγή	xi
I Τα θεμέλια της Θεωρίας Ουρών Αναμονής	1
1 Περιγραφή, ονοματολογία και μέτρα απόδοσης	3
1.1 Βασική περιγραφή και ονοματολογία Kendall	3
1.2 Αναγεννητικότητα και μέτρα απόδοσης συστήματος	6
1.3 Εμφυτευμένες διαδικασίες σε στιγμές αφίξεων και αναχωρήσεων	12
1.4 Ασκήσεις	14
1.5 Σχόλια	16
2 Βασικά αποτελέσματα	19
2.1 Ρυθμός συνωστισμού - Ευστάθεια	19
2.2 Ιδιότητα μεμονωμένων μεταβάσεων και ιδιότητα PASTA	20
2.3 Νόμος του Little	22
2.4 Ασκήσεις	24
2.5 Σχόλια	25
3 Βασικά στοιχεία πιθανοτήτων	27
3.1 Δεσμευμένη μέση τιμή	27
3.2 Πιθανογεννήτριες - Μετασχηματισμοί Laplace-Stieltjes	30
3.3 Η εκθετική κατανομή	40
3.4 Ασκήσεις	42
3.5 Σχόλια	44

II	Αποτίμηση απόδοσης: Μαρκοβιανά Μοντέλα	47
4	Στοχαστικές διαδικασίες για Μαρκοβιανά μοντέλα	49
4.1	Μαρκοβιανές αλυσίδες συνεχούς χρόνου: Ορισμός και μεταβατική συμπεριφορά	49
4.2	Μαρκοβιανές αλυσίδες συνεχούς χρόνου: Οριακή συμπεριφορά	58
4.3	Μαρκοβιανές αλυσίδες συνεχούς χρόνου: Αμοιβές και κόστη	62
4.4	Διαδικασία Poisson	63
4.5	Ασκήσεις	68
4.6	Σχόλια	70
5	Ανάλυση μέσης τιμής για Μαρκοβιανά μοντέλα	71
5.1	Ανάλυση μέσης τιμής στην $M/M/1/1$ ουρά	71
5.2	Ανάλυση μέσης τιμής στην $M/M/1$ ουρά	72
5.3	Ανάλυση μέσης τιμής στην $M/M/1$ ουρά με χρόνους επανεκκίνησης	73
5.4	Ανάλυση μέσης τιμής στην $M/M/1$ ουρά με την K -πολιτική ενεργοποίησης	74
5.5	Ασκήσεις	76
5.6	Σχόλια	77
6	Απλές Μαρκοβιανές ουρές	79
6.1	Ορισμός και βασικοί υπολογισμοί	79
6.2	Ο ρυθμός διαπέρασης και οι εμφυτευμένες κατανομές	81
6.3	Η $M/M/1/1$ ουρά	83
6.4	Η $M/M/1$ ουρά	85
6.5	Τροποποιήσεις της $M/M/1$ ουράς	89
6.5.1	Η $M/M/1/k$ ουρά	89
6.5.2	Η $M/M/1$ ουρά με αποθαρρυνόμενους πελάτες	91
6.5.3	Η $M/M/1$ ουρά με μεταβλητή ταχύτητα εξυπηρέτησης	92
6.6	Η $M/M/c$ ουρά	93
6.7	Η $M/M/c/c$ ουρά και το μοντέλο Engset	95
6.8	Ασκήσεις	98
6.9	Σχόλια	99
7	Γενικές Μαρκοβιανές ουρές	101
7.1	Η $M/M/1$ ουρά με ομαδικές αφίξεις	101
7.1.1	Η περίπτωση των μεμονωμένων αφίξεων	103
7.1.2	Η περίπτωση των ομάδων γεωμετρικού μεγέθους	104
7.1.3	Κατανομές ισορροπίας πλήθους πελατών σε στιγμές αφίξεων	104
7.2	Το Μαρκοβιανό σύστημα εκκαθαρίσεων με ομαδικές αφίξεις	105
7.2.1	Η περίπτωση των μεμονωμένων αφίξεων	107
7.2.2	Η περίπτωση των ομάδων γεωμετρικού μεγέθους	108
7.3	Η $M/M/1$ ουρά με ομαδικές εξυπηρετήσεις	108
7.4	Η $M/M/$ άπειρο ουρά με ομαδικές αφίξεις	113
7.5	Ασκήσεις	115
7.6	Σχόλια	116
8	Η μέθοδος των φάσεων	119
8.1	Η μέθοδος των φάσεων για κατανομές Erlang	119
8.2	Η γενική μέθοδος των φάσεων	121
8.3	Προσαρμογή κατανομών σε δεδομένα	123
8.4	Ασκήσεις	124

8.5	Σχόλια	125
9	Διδιάστατες Μαρκοβιανές ουρές	127
9.1	Η $M/M/1$ ουρά με ενεργές και ανενεργές περιόδους εξυπηρέτησεων	128
9.2	Η $M/M/1$ ουρά με χρόνους επανεκκίνησης	131
9.3	Η $M/M/1/1$ ουρά με επαναπροσπάθειες	134
9.4	Ασκήσεις	137
9.5	Σχόλια	138
10	Αντιστρέψιμες Μαρκοβιανές ουρές	141
10.1	Αντίστροφη στοχαστικής διαδικασίας	141
10.2	Αντιστρέψιμες στοχαστικές διαδικασίες	143
10.3	Η $M/M/c$ ουρά με ετερογενείς υπηρέτες	145
10.3.1	Η $M/M/2$ ουρά με ετερογενείς υπηρέτες	146
10.4	Η $M/M/1$ ουρά με ανενεργές περιόδους αφίξεων-εξυπηρέτησεων	147
10.5	Δυο παράλληλες $M/M/1$ ουρές με κοινό χώρο αναμονής	149
10.6	Η διαδικασία αναχωρήσεων σε απλές Μαρκοβιανές ουρές με Poisson διαδικασία αφίξεων	149
10.7	Ασκήσεις	151
10.8	Σχόλια	152
11	Απλά Μαρκοβιανά δίκτυα ουρών	153
11.1	Μαρκοβιανά δίκτυα - Απλά Μαρκοβιανά δίκτυα	153
11.2	Δίκτυα Jackson	155
11.3	Ρυθμοί διαπέρασης ανοικτών δικτύων Jackson	156
11.4	Κατανομή ισορροπίας ανοικτών δικτύων Jackson	157
11.5	Σχετικοί ρυθμοί διαπέρασης κλειστών δικτύων Jackson	163
11.6	Κατανομή ισορροπίας κλειστών δικτύων Jackson	164
11.7	Ασκήσεις	167
11.8	Σχόλια	169
III	Αποτίμηση απόδοσης: Μη-Μαρκοβιανά Μοντέλα	171
12	Στοχαστικές διαδικασίες για μη-Μαρκοβιανά Μοντέλα	173
12.1	Μαρκοβιανές αλυσίδες διακριτού χρόνου	173
12.2	Ανανεωτικές διαδικασίες	180
12.3	Ασκήσεις	192
12.4	Σχόλια	196
13	Ανάλυση μέσης τιμής για μη-Μαρκοβιανά μοντέλα	197
13.1	Ανάλυση μέσης τιμής στην $M/G/1/1$ ουρά	198
13.2	Ανάλυση μέσης τιμής στην $M/G/1$ ουρά	198
13.3	Ανάλυση μέσης τιμής στην $M/G/1$ ουρά με πολλαπλές διακοπές υπηρέτη	200
13.4	Ανάλυση μέσης τιμής στην $M/G/1$ ουρά με την K -πολιτική ενεργοποίησης	201
13.5	Ασκήσεις	202
13.6	Σχόλια	203
14	Κλασικές μη-Μαρκοβιανές ουρές	205
14.1	Η $M/G/1$ ουρά	205
14.1.1	Ο αριθμός των πελατών στο σύστημα	206

14.1.2	Ο χρόνος παραμονής πελάτη	208
14.1.3	Η διάρκεια και ο αριθμός των πελατών ενός κύκλου απασχόλησης	210
14.2	H G/M/1 ουρά	213
14.2.1	Ο αριθμός των πελατών στο σύστημα	213
14.2.2	Ο χρόνος παραμονής πελάτη	215
14.3	H M/G/άπειρο ουρά	215
14.4	Το σύστημα με Poisson αφίξεις και ανανεωτική διαδικασία εκκαθαρίσεων	217
14.5	Ασκήσεις	218
14.6	Σχόλια	220
IV	Βελτιστοποίηση, έλεγχος και στρατηγική συμπεριφορά	223
15	Βέλτιστος σχεδιασμός και στρατηγική συμπεριφορά	225
15.1	Βασικές έννοιες από τη Θεωρία Παιγνίων	226
15.2	Στρατηγική αλληλεπίδραση μεταξύ των πελατών σε συστήματα ουρών αναμονής	227
15.3	Συμπεριφορές απόφυγε-το-πλήθος και ακολουθήσε-το-πλήθος	228
15.4	Στρατηγικές κατωφλίου	229
15.5	Πλαίσιο κοινωνικής βελτιστοποίησης και βελτιστοποίησης μονοπωλίου	229
15.6	Στρατηγικές εισόδου στη μη-παρατηρήσιμη M/M/1 ουρά	230
15.7	Στρατηγικές εισόδου στην παρατηρήσιμη M/M/1 ουρά	233
15.8	Πληροφορία και στρατηγική συμπεριφορά	237
15.9	Ασκήσεις	238
15.10	Σχόλια	240
16	Κλασικά προβλήματα βέλτιστου σχεδιασμού και βέλτιστου ελέγχου	243
16.1	Ένας γρήγορος ή πολλοί αργοί υπηρέτες;	243
16.2	Κοινή ουρά ή ουρά ανά υπηρέτη;	244
16.3	Γρήγορος ακριβός υπηρέτης ή αργός οικονομικός υπηρέτης;	245
16.4	Προς ποιον υπηρέτη να κατευθυνθούν οι πελάτες;	247
16.5	Ασκήσεις	248
16.6	Σχόλια	249
Επίλογος		251
Ιστορία, εφαρμογές και κατευθύνσεις		253
Βιβλιογραφία		261

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

1.1	Εξέλιξη του πλήθους πελατών σε σύστημα D/D/1 με ενδιάμεσους χρόνους αφίξεων $a = 10$ και χρόνους εξυπηρέτησης $b = 9$	12
1.2	Εξέλιξη του πλήθους πελατών σε σύστημα D/D/1 με ενδιάμεσους χρόνους αφίξεων $a = 10$ και χρόνους εξυπηρέτησης $b = 1$	12
4.1	Διάγραμμα ρυθμών μετάβασης μιας Μαρκοβιανής αλυσίδας με χώρο καταστάσεων $\{0, 1, 2, 3\}$	51
4.2	Διάγραμμα ρυθμών μετάβασης διαδικασίας Poisson.	55
4.3	Διάγραμμα ρυθμών μετάβασης M/M/1/1 ουράς.	56
4.4	Διάγραμμα ρυθμών μετάβασης M/M/1 ουράς.	56
4.5	Διάγραμμα ρυθμών μετάβασης διαδικασίας γέννησης-θανάτου.	62
6.1	Διάγραμμα ρυθμών μετάβασης M/M/1/k ουράς.	89
6.2	Διάγραμμα ρυθμών μετάβασης M/M/1 ουράς με αποθαρρυνόμενους πελάτες.	91
6.3	Διάγραμμα ρυθμών μετάβασης M/M/1 ουράς με μεταβλητή ταχύτητα εξυπηρέτησης.	93
6.4	Διάγραμμα ρυθμών μετάβασης M/M/c ουράς.	94
6.5	Διάγραμμα ρυθμών μετάβασης M/M/c/c ουράς.	96
6.6	Διάγραμμα ρυθμών μετάβασης του μοντέλου του Engset.	97
7.1	Διάγραμμα ρυθμών μετάβασης M/M/1 ουράς με ομαδικές αφίξεις.	102
7.2	Διάγραμμα ρυθμών μετάβασης Μαρκοβιανού συστήματος εκκαθαρίσεων με ομαδικές αφίξεις.	106
7.3	Διάγραμμα ρυθμών μετάβασης Μαρκοβιανού συστήματος εκκαθαρίσεων με μεμονωμένες αφίξεις.	107
7.4	Διάγραμμα ρυθμών μετάβασης M/M/1 ουράς με ομαδικές εξυπηρετήσεις.	109
7.5	Διάγραμμα ρυθμών μετάβασης M/M/∞ ουράς με ομαδικές αφίξεις.	113
8.1	Διάγραμμα ρυθμών μετάβασης $E_2/M/1$ ουράς.	120
9.1	Διάγραμμα ρυθμών μετάβασης M/M/1 ουράς με ενεργές και ανενεργές περιόδους εξυπηρέτησεων.	129
9.2	Διάγραμμα ρυθμών μετάβασης M/M/1 ουράς με χρόνους επανεκκίνησης.	132
9.3	Διάγραμμα ρυθμών μετάβασης M/M/1/1 ουράς με επαναπροσπάθειες.	135

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

1	Παραδείγματα συστημάτων εξυπηρέτησης	xi
3.1	Συναρτήσεις και αντίστοιχοι μετασχηματισμοί Laplace-Stieltjes.	39
4.1	Πίνακας μεταβάσεων στοχαστικής διαδικασίας Poisson.	55
4.2	Πίνακας μεταβάσεων M/M/1/1 ουράς.	55
4.3	Πίνακας μεταβάσεων M/M/1 ουράς.	56
4.4	Πίνακας μεταβάσεων D/M/1 ουράς.	57
7.1	Πίνακας μεταβάσεων M/M/1 ουράς με ομαδικές αφίξεις.	101
7.2	Πίνακας μεταβάσεων Μαρκοβιανού συστήματος εκκαθαρίσεων με ομαδικές αφίξεις.	106
7.3	Πίνακας μεταβάσεων M/M/1 ουράς με ομαδικές εξυπηρετήσεις.	108
7.4	Πίνακας πλήθους ριζών της $A(z) = z^N - a(z)$ για πιθανογεννήτρια $a(z)$ και θετικό ακέραιο N	110
7.5	Πίνακας μεταβάσεων M/M/∞ ουράς με ομαδικές αφίξεις.	113
8.1	Πίνακας μεταβάσεων $E_2/M/1$ ουράς.	120
8.2	Πίνακας μεταβάσεων $E_k/E_s/1$ ουράς.	121
8.3	Πίνακας μεταβάσεων Μαρκοβιανής αλυσίδας αναπαράστασης φάσεων για την κατανομή Erlang(k, α).	121
9.1	Πίνακας μεταβάσεων M/M/1 ουράς με ενεργές και ανενεργές περιόδους εξυπηρέτησεων.	128
9.2	Πίνακας μεταβάσεων M/M/1 ουράς με χρόνους επανεκκίνησης.	131
9.3	Πίνακας μεταβάσεων M/M/1/1 ουράς με επαναπροσπάθειες.	135
10.1	Πίνακας μεταβάσεων M/M/c ουράς με ετερογενείς υπηρέτες.	145
10.2	Πίνακας μεταβάσεων M/M/2 ουράς με ετερογενείς υπηρέτες.	146
10.3	Πίνακας μεταβάσεων M/M/1 ουράς με νεκρές περιόδους.	148

ΠΙΝΑΚΑΣ ΣΥΝΤΟΜΕΥΣΕΩΝ - ΑΚΡΩΝΥΜΙΩΝ

ΜΚΔ	Μέγιστος Κοινός Διαιρέτης
ATC	Avoid-the-Crowd
BR	Best Response
D	Deterministic
E_r	Erlang- r
FCFS	First-Come-First-Served
FIFO	First-In-First-Out
FTC	Follow-the-Crowd
G	General
GI	General-Independent
LAS	Least-Attained-Service
LCFS	Last-Come-First-Served
LCFS-PR	Last-Come-First-Served-Preemptive-Resume
LIFO	Last-In-First-Out
LS	Laplace-Stieltjes
M	Memoryless, Markovian
PASTA	Poisson Arrivals See Time Averages
SIRO	Service-In-Random-Order
SSTF	Shortest-Service-Time-First

ΕΙΣΑΓΩΓΗ

Η Θεωρία Ουρών Αναμονής (Queueing Theory) είναι ένα διεπιστημονικό πεδίο που βρίσκεται στην τομή των Εφαρμοσμένων Μαθηματικών (κυρίως της Εφαρμοσμένης Θεωρίας Πιθανοτήτων και της Επιχειρησιακής Έρευνας), των Επιστημών Μηχανικού (κυρίως Ηλεκτρολόγου Μηχανικού, Μηχανολόγου Μηχανικού και Μηχανικού Διοίκησης και Παραγωγής), της Πληροφορικής καθώς και της Διοικητικής Επιστήμης (κυρίως της Διοίκησης Λειτουργιών). Αντικείμενο μελέτης της είναι η μελέτη συστημάτων εξυπηρέτησης (service systems), γνωστών και ως ουρών αναμονής (queueing systems). Τα συστήματα εξυπηρέτησης είναι μαθηματικά πρότυπα που παριστάνουν συστήματα εισόδου-εξόδου πελατών που δέχονται κάποια μορφή εξυπηρέτησης από υπηρέτες, στα οποία υπαισέρχεται τυχαιότητα. Οι έννοιες «πελάτης», «εξυπηρέτηση» και «υπηρέτης» ερμηνεύονται με μεγάλη ευρύτητα και έτσι οι εφαρμογές της Θεωρίας Ουρών Αναμονής εκτείνονται σε πάρα πολλά πεδία της επιστήμης και της ανθρώπινης δραστηριότητας γενικότερα. Πράγματι, τέτοια συστήματα εμφανίζονται σε πολλές καταστάσεις που παρατηρούνται στην καθημερινή ζωή, καθώς και σε περίπλοκα τεχνολογικά συστήματα. Ενδεικτικά, στον πίνακα 1 παρατίθενται μερικά παραδείγματα συστημάτων που μπορούν να προτυποποιηθούν ως συστήματα εξυπηρέτησης και κατόπιν να μελετηθούν στα πλαίσια της Θεωρίας Ουρών Αναμονής.

Ιστορικά, η συγκεκριμένη επιστημονική περιοχή άρχισε να αναπτύσσεται στις αρχές του 20ου αιώνα, όταν ο Agner Krarup Erlang (1878-1929) δημοσίευσε κάποιες εργασίες του για τη μαθηματική μοντελοποίηση του συνωστισμού σε τηλεφωνικά δίκτυα. Η μεγάλη επιτυχία αυτών των μεθόδων στη μελέτη πραγματικών

Σύστημα εξυπηρέτησης	Πελάτης	Εξυπηρέτηση	Υπηρέτης
Ταμείο	Οφειλέτης	Πληρωμή	Ταμίας
Τηλεφωνικό κέντρο	Κλήση	Συνδιάλεξη	Υπάλληλος
Νοσοκομείο	Ασθενής	Εξέταση/θεραπεία	Ιατρός/νοσηλεύτης
Ψυχαγωγικό πάρκο	Επισκέπτης	Ψυχαγωγία	Μέλος προσωπικού
Μέσο μεταφοράς	Επιβάτης	Μεταφορά	Οδηγός
Τουριστικό κατάλυμα	Επισκέπτης	Φιλοξενία	Ξενοδόχος
Αεροδρόμιο	Αποσκευή	Μεταφορά	Σύστημα διαχ. αποσκ.
Αεροδρόμιο	Αεροπλάνο	Απογείωση/Προσγείωση	Πύργος ελέγχου
Χώρα	Πολίτης	Διαβίωση	Κράτος

Πίνακας 1: Παραδείγματα συστημάτων εξυπηρέτησης

συστημάτων έδωσε τεράστια ώθηση στην περαιτέρω ανάπτυξη της Θεωρίας Ουρών Αναμονής καθώς και των εφαρμογών της και σε άλλα πεδία.

Στη Θεωρία Ουρών Αναμονής, η μελέτη ενός συστήματος μπορεί να περιλαμβάνει όλα ή κάποια από τα εξής στάδια:

- μαθηματική προτυποποίηση (μοντελοποίηση),
- αποτίμηση απόδοσης,
- βέλτιστο σχεδιασμό,
- βέλτιστο δυναμικό έλεγχο,
- προσδιορισμό στρατηγικής συμπεριφοράς,
- σύγκριση με άλλα συστήματα εξυπηρέτησης.

Η μαθηματική προτυποποίηση αποτελεί το πρώτο και απαραίτητο στάδιο για την οποιαδήποτε μελέτη στη Θεωρία Ουρών Αναμονής. Κατά τη διάρκεια του σταδίου αυτού ο ερευνητής επικεντρώνεται στον εντοπισμό των κύριων χαρακτηριστικών του συστήματος που θα πρέπει να αναπαρασταθούν με το πρότυπο που θα υιοθετήσει. Στο σημείο αυτό η δυσκολία έγκειται στην εύρεση μιας καλής ισορροπίας μεταξύ πιστότητας και επιλυσιμότητας του προτύπου. Συνήθως, όσο πιο πιστά αναπαριστά ένα πρότυπο το αρχικό σύστημα, τόσο δυσκολότερο είναι να μελετηθεί με μαθηματικές μεθόδους. Επομένως, θα πρέπει να επιλεγεί το απλούστερο δυνατό πρότυπο που όμως απεικονίζει τα σημαντικότερα χαρακτηριστικά του αρχικού συστήματος. Στη διαδικασία της προτυποποίησης, επιλέγεται κάποια οικογένεια στοχαστικών διαδικασιών στο πλαίσιο της οποίας παριστάνεται το σύστημα.

Κατόπιν, το ενδιαφέρον εστιάζεται στην αποτίμηση της απόδοσης του συστήματος, δηλαδή σε ερωτήματα του τύπου «Πόσοι πελάτες βρίσκονται στο σύστημα κατά μέσο όρο;» ή «Πόσο χρόνο παραμένει στο σύστημα ένας πελάτης κατά μέσο όρο;» ή «Ποιο είναι το ποσοστό του χρόνου που ένας υπηρέτης είναι απασχολημένος;». Το βασικό εργαλείο σε αυτό το στάδιο είναι η Θεωρία Πιθανοτήτων και ειδικότερα το δυναμικό της μέρος που αντιστοιχεί στις στοχαστικές διαδικασίες.

Αν ενδιαφέρει ο σχεδιασμός ενός νέου συστήματος, τότε έχουμε να αντιμετωπίσουμε ένα πρόβλημα βελτιστού σχεδιασμού. Στην περίπτωση αυτή, κάποιες παράμετροι λειτουργίας ενός συστήματος θεωρούνται ως δοσμένες, π.χ. ο ρυθμός αφίξεων των πελατών ή ο ρυθμός εξυπηρέτησης ανά υπηρέτη (που μπορεί να έχουν προκύψει από στατιστικά στοιχεία που τηρούνται για παρόμοια συστήματα). Κάποιες άλλες, όμως, παράμετροι λειτουργίας μπορούν να ελεγχθούν από τον σχεδιαστή του συστήματος, π.χ. ο μέγιστος αριθμός των υπηρέτων, δηλ. ο αριθμός των θέσεων εξυπηρέτησης που θα προβλεφθούν. Στην περίπτωση αυτή, υπολογίζουμε το μέτρο απόδοσης του συστήματος που μας ενδιαφέρει να βελτιστοποιήσουμε ως συνάρτηση των παραμέτρων που ελέγχουμε και κατόπιν βελτιστοποιούμε τη συνάρτηση με κλασικές μεθόδους βελτιστοποίησης (π.χ. βελτιστοποίηση για μια μεταβλητή με μηδενισμό της παραγώγου και σύγκριση των τιμών της συνάρτησης στα κρίσιμα σημεία ή πιο προχωρημένες μεθόδους μαθηματικού προγραμματισμού με συνθήκες Lagrange, Karush-Kuhn-Tucker, ακέραιο προγραμματισμό κλπ.).

Κάποιες μεταβλητές ελέγχου μπορούν να μεταβάλλονται όσο το σύστημα βρίσκεται σε λειτουργία, π.χ. ο αριθμός των ενεργών υπηρέτων μπορεί να προσαρμόζεται ανάλογα με τον συνωστισμό που παρατηρείται στο σύστημα. Στην περίπτωση που εστιάζουμε στον έλεγχο τέτοιων μεταβλητών μιλάμε για ένα πρόβλημα βέλτιστου δυναμικού ελέγχου και το κατάλληλο εργαλείο μελέτης είναι ο στοχαστικός δυναμικός προγραμματισμός. Η διαφορά των προβλημάτων αυτών σε σχέση με τα προβλήματα βέλτιστου σχεδιασμού έγκειται στο ότι οι μεταβλητές μπορούν να αλλάζουν τιμές ανάλογα με την κατάσταση του συστήματος, ενώ στα προβλήματα βέλτιστου σχεδιασμού η τιμή τους πρέπει να προσδιοριστεί μια φορά και για πάντα.

Σε όλα τα προαναφερθέντα στάδια μελέτης ενός συστήματος εξυπηρέτησης, οι πελάτες θεωρούνται παθητικές οντότητες που δεν λαμβάνουν αποφάσεις. Βέβαια, από την εμπειρία μας από τα πραγματικά συστήματα

εξυπηρέτησης, γνωρίζουμε ότι οι πελάτες είναι ενεργητικές οντότητες που λαμβάνουν αποφάσεις. Π.χ., σε ένα εστιατόριο, ένας πελάτης επηρεάζεται (είτε θετικά είτε αρνητικά) από τον συνωστισμό που συναντά φθάνοντας σε αυτό. Μεγάλος συνωστισμός μπορεί να είναι μια ένδειξη καλής ποιότητας ή ένδειξη αργής εξυπηρέτησης. Σε κάθε περίπτωση, πάντως, ο συνωστισμός επηρεάζει τις αποφάσεις των δυνητικών πελατών ενός συστήματος. Στην περίπτωση αυτή, έχουμε να αντιμετωπίσουμε ένα πρόβλημα λήψης αποφάσεων με πολλούς αποφασίζοντες: τους πελάτες, τους υπηρέτες και τον διαχειριστή. Ο καθένας από αυτούς λαμβάνει την απόφασή του, έχοντας υπόψη ότι και οι άλλες οντότητες λαμβάνουν αποφάσεις, η καθεμιά με σκοπό τη μεγιστοποίηση του δικού της οφέλους. Οπότε, δημιουργείται ένα «παίγνιο» μεταξύ των διαφόρων οντοτήτων (με την έννοια της σύγκρουσης των συμφερόντων τους) και το κατάλληλο πλαίσιο μελέτης είναι η Θεωρία Παιγνίων. Τα προβλήματα αυτά αναφέρονται και ως προβλήματα στρατηγικής συμπεριφοράς στα πλαίσια της Θεωρίας Ουρών.

Τέλος, συχνά εμφανίζεται η ανάγκη να συγκριθούν δυο συστήματα εξυπηρέτησης. Αυτό προκύπτει, π.χ., όταν ένας οργανισμός σκέφτεται την αντικατάσταση ενός παλαιού συστήματος εξυπηρέτησης από ένα νέο και επομένως επιθυμεί να συγκρίνει τα δυο συστήματα και να ποσοτικοποιήσει τη διαφορά τους, ώστε να καταλήξει στο αν είναι συμφέρουσα η αντικατάσταση. Στην περίπτωση αυτή χρησιμοποιείται ένα ή περισσότερα από τα στάδια μελέτης που αναπτύξαμε, ανάλογα με την προτεινόμενη αλλαγή.

Από τα παραπάνω, γίνεται φανερό ότι η Θεωρία Ουρών είναι ένα πολυεπίπεδο γνωστικό πεδίο που χρησιμοποιεί εργαλεία από πολλές περιοχές των Μαθηματικών και εφαρμόζει τα αποτελέσματα σε πολλές περιοχές της καθημερινότητας και της τεχνολογίας, ανάλογα με τα ιδιαίτερα χαρακτηριστικά των συστημάτων εξυπηρέτησης που εμφανίζονται. Επομένως, μια εισαγωγή στη Θεωρία Ουρών πρέπει αναγκαστικά να είναι επιλεκτική σχετικά με τα στάδια μελέτης που θα πραγματοποιηθεί από τα παραπάνω, τα μαθηματικά εργαλεία που θα χρησιμοποιήσει και τα πεδία εφαρμογών που θα παρουσιάσει.

Το παρόν σύγγραμμα επικεντρώνεται κυρίως στη μοντελοποίηση και στην αποτίμηση απόδοσης συστημάτων εξυπηρέτησης με βάση Μαρκοβιανές αλυσίδες. Οι Μαρκοβιανές αλυσίδες είναι το μαθηματικά απλούστερο (και ταυτόχρονα εξαιρετικά ευρύ από πλευράς εφαρμογών) είδος στοχαστικών διαδικασιών που μπορεί να κατανοηθεί από κάποιον που έρχεται σε επαφή για πρώτη φορά με το αντικείμενο της Θεωρίας Ουρών Αναμονής.

Το βασικό προαπαιτούμενο για την κατανόηση του συγγράμματος είναι ένα πρώτο μάθημα Πιθανοτήτων, όπως αυτά που διδάσκονται στα πρώτα δυο έτη σπουδών σε προγράμματα Θετικών και Οικονομικών Επιστημών, Πολυτεχνικών Σχολών και Σχολών Διοίκησης Επιχειρήσεων. Επιπλέον, χρειάζεται γνώση βασικών τεχνικών για υπολογισμούς αθροισμάτων και ολοκληρωμάτων, στο επίπεδο ενός πρώτου μαθήματος Γενικών Μαθηματικών ή ενός μαθήματος Απειροστικού Λογισμού. Η γνώση στοχαστικών διαδικασιών και ιδιαίτερα Μαρκοβιανών αλυσίδων, προχωρημένων μεθόδων βελτιστοποίησης ή εννοιών από τη Θεωρία Παιγνίων δεν προϋποτίθεται. Το αναγκαίο υλικό από τις περιοχές αυτές των Μαθηματικών αναπτύσσεται συνοπτικά στο σύγγραμμα στο βάθος που χρειάζεται για τη συγκεκριμένη εισαγωγή στη Θεωρία Ουρών Αναμονής. Βεβαίως, η ανάπτυξη αυτού του υλικού εστιάζεται στην κατανόηση των εννοιών και τη χρήση των αποτελεσμάτων και όχι στην αυστηρή απόδειξή τους. Σε κάθε περίπτωση, όμως, οι διάφοροι ορισμοί και τα αποτελέσματα (θεωρήματα) που χρησιμοποιούμε διατυπώνονται με μαθηματική ακρίβεια, ώστε να καθίσταται σαφές στον αναγνώστη ποιες υποθέσεις πρέπει να ελέγχονται κάθε φορά για την εφαρμογή ενός θεωρητικού εργαλείου.

Το βιβλίο διαρθρώνεται σε τέσσερα μέρη. Στο πρώτο μέρος του βιβλίου γίνεται μια εισαγωγή στις βασικές έννοιες της Θεωρίας Ουρών Αναμονής, στην προβληματική που αναπτύσσεται στα πλαίσια της μελέτης συστημάτων εξυπηρέτησης και σε κάποια βασικά αποτελέσματα που έχουν μεγάλο εύρος εφαρμογής. Επιπλέον, γίνεται μια σύντομη επισκόπηση γνώσεων πιθανοτήτων. Το δεύτερο μέρος ξεκινά με μια εισαγωγή στις Μαρκοβιανές αλυσίδες συνεχούς χρόνου που χρησιμοποιούνται ως το βασικό εργαλείο για τη μοντελοποίηση και την αποτίμηση απόδοσης συστημάτων εξυπηρέτησης. Στο μέρος αυτό του βιβλίου επικεντρωνόμαστε σε μια συστηματική εισαγωγή στην αποτίμηση απόδοσης συστημάτων εξυπηρέτησης. Στο τρίτο μέρος παρουσιάζουμε κάποια μαθηματικά εργαλεία για την αποτίμηση απόδοσης συστημάτων που ξεφεύγουν από το πλαίσιο των Μαρκοβιανών αλυσίδων συνεχούς χρόνου και μελετάμε κάποια σημαντικά μη-Μαρκοβιανά συστήματα. Τέλος, στο τέταρτο μέρος προχωράμε σε μια σύντομη εισαγωγή στα υπόλοιπα στάδια μελέτης που αναφέ-

ραμε. Σε αυτό το τέταρτο μέρος δίνουμε μεγαλύτερη έμφαση στη μελέτη της στρατηγικής συμπεριφοράς σε συστήματα εξυπηρέτησης, καθώς η τρέχουσα ερευνητική δραστηριότητα στη Θεωρία Ουρών καθώς και οι καινοτόμες εφαρμογές της φαίνεται να αφορούν κυρίως τέτοια προβλήματα.

Κλείνοντας τον πρόλογο, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα της διδακτορικής διατριβής μου κ. Δημήτρη Φακίνο που με εισήγαγε στην περιοχή της Θεωρίας Ουρών πριν από περίπου 28 χρόνια, τον αγαπημένο φίλο και συνάδελφο κ. Απόστολο Μπουρνέτα με τον οποίο συνεργάζομαι εδώ και 16 περίπου χρόνια σε θέματα Ουρών Αναμονής - και όχι μόνο -, τους διδακτορικούς φοιτητές μου και πλέον καλούς φίλους και συναδέλφους κ.κ. Σπύρο Δήμου, Σπυριδούλα Κάντα, Στέλλα Καποδίστρια, Δημήτρη Λογοθέτη, Αθανασία Μάνου και Όλγα Μπούνταλη, και τους μεταπτυχιακούς και προπτυχιακούς φοιτητές μου που όλα αυτά τα χρόνια παρακολούθησαν μαθήματα Θεωρίας Ουρών Αναμονής στο Τμήμα Μαθηματικών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών. Επίσης, ευχαριστώ τους συσυγγραφείς μου σε διάφορες δημοσιεύσεις κ.κ. Γιώργο Βασιλειάδη, Γιάννη Δημητρακόπουλο, Στέφανο Λεονάρδο, Ivo Adan, Opher Baron, Antonio Gómez-Corral, Maria Jesús López-Herrero, Fikri Karaesmen και Jacques Resing. Θερμές ευχαριστίες οφείλω ιδιαίτερα στους κ.κ. Παναγιώτη Ανδρέου, Γιώργο Βασιλειάδη, Γιάννη Δημητρακόπουλο, Σπύρο Δήμου, Γιάννη Δημητρίου, Παύλο Ζουμπούλογλου, Αθανασία Μάνου, και Απόστολο Μπουρνέτα που διάβασαν παλαιότερες εκδόσεις του παρόντος βιβλίου και/ή το χρησιμοποίησαν σε μαθήματα που δίδαξαν και μου υπέδειξαν λάθη και παραλείψεις. Όλοι αυτοί συνέβαλαν και συμβάλλουν ώστε να διατηρώ τον ενθουσιασμό μου για το αντικείμενο και να εμβαθύνω στην κατανόησή του. Ιδιαίτερη μνεία αξίζει στον Jesús Artalejo με τον οποίο συνεργάστηκα εντατικά για μια δεκαετία. Δυστυχώς δεν είναι πια μαζί μας. Ευχαριστώ πολύ, επίσης, το Τμήμα Μαθηματικών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών που προσφέρει ένα δημιουργικό και γεμάτο ερεθίσματα περιβάλλον που προάγει την επιστημονική δραστηριότητα μου. Ένα μεγάλο «Ευχαριστώ» στη μητέρα μου Όλγα, στους υπόλοιπους συγγενείς, στους φίλους και στους δασκάλους μου. Και τέλος, θα ήθελα ολόψυχα να ευχαριστήσω τη σύζυγό μου Πόλυ και τα παιδιά μας Θοδωρή και Κυριακή Βασιλική που ομορφαίνουν και νοηματοδοτούν τη ζωή μου και διαμορφώνουν ένα ήρεμο και ζεστό οικογενειακό περιβάλλον που μου επιτρέπει να είμαι δημιουργικός.

Καλή Ανάγνωση!

Αντώνιος Θ. Οικονόμου

Αθήνα, 5 Αυγούστου 2022.

Μέρος I

ΤΑ ΘΕΜΕΛΙΑ ΤΗΣ ΘΕΩΡΙΑΣ ΟΥΡΩΝ ΑΝΑΜΟΝΗΣ

ΚΕΦΑΛΑΙΟ 1

ΠΕΡΙΓΡΑΦΗ, ΟΝΟΜΑΤΟΛΟΓΙΑ ΚΑΙ ΜΕΤΡΑ ΑΠΟΔΟΣΗΣ

Στο κεφάλαιο αυτό περιγράφουμε τα βασικά χαρακτηριστικά που καθορίζουν ένα σύστημα εξυπηρέτησης. Επιπλέον, αναφερόμαστε στην ονοματολογία-ταξινόμηση του David George Kendall (1918-2007) για τα συστήματα εξυπηρέτησης. Κατόπιν, εστιάζουμε στα σημαντικότερα μέτρα απόδοσης που χρησιμοποιούνται στη μελέτη των συστημάτων εξυπηρέτησης και βλέπουμε ότι έχουν πολλαπλές ερμηνείες που συνδέονται με τη λεγόμενη αναγεννητικότητα των συστημάτων.

1.1 Βασική περιγραφή και ονοματολογία Kendall

Ένα σύστημα εξυπηρέτησης ή ουρά αναμονής (service system, queueing system, queue) είναι στην ουσία ένα σύστημα εισόδου-εξόδου διακριτών οντοτήτων (μονάδων, πελατών), στο οποίο υπεισέρχεται τυχαιότητα.

Τα βασικά χαρακτηριστικά ενός συστήματος εξυπηρέτησης είναι:

- η διαδικασία αφίξεων (arrival process),
- οι χρόνοι εξυπηρέτησης (service times),
- ο αριθμός των (παράλληλων) παρόχων εξυπηρέτησης - υπηρετών (number of servers),
- η χωρητικότητα του συστήματος (system capacity), και
- η πειθαρχία ουράς (queue discipline).

Για τον λόγο αυτό ο D.G. Kendall εισήγαγε στην εργασία Kendall 1953 ένα σύστημα ονοματολογίας για τις πιο απλές ουρές που περιγράφει συνοπτικά αυτά τα χαρακτηριστικά. Η ονοματολογία του Kendall έχει τη μορφή $A/B/c/k(\quad)$, όπου τα A, B είναι γράμματα, τα c, k αριθμοί και μέσα στην παρένθεση γράφεται μια ακροστοιχίδα γραμμάτων. Καθεμιά από τις 5 παραμέτρους της ονοματολογίας του Kendall αναφέρεται στα 5 χαρακτηριστικά που περιγράψαμε παραπάνω.

Η διαδικασία αφίξεων περιγράφει το πώς έρχονται οι πελάτες στο σύστημα. Οι αφίξεις είναι διαδοχικά γεγονότα που συμβαίνουν στον χρόνο και επομένως μπορούν να περιγραφούν από τις λεγόμενες σημειακές διαδικασίες και τις αντίστοιχες απαριθμήτριές τους που είναι τα γενικά πρότυπα για τη μελέτη γεγονότων που συμβαίνουν με κάποιο βαθμό τυχαιότητας. Πιο συγκεκριμένα, αν έχουμε μια ακολουθία μη-αρνητικών τυχαίων μεταβλητών X_1, X_2, \dots μπορούμε να ορίσουμε τις τυχαίες μεταβλητές $S_0 = 0$ και $S_n = \sum_{i=1}^n X_i$, $n \geq 1$, για τις οποίες ισχύει $S_0 \leq S_1 \leq S_2 \leq \dots$. Η στοχαστική διαδικασία $\{S_n : n \geq 0\}$ αναφέρεται ως σημειακή διαδικασία με χρόνους γεγονότων S_n , $n \geq 0$, και ενδιάμεσους χρόνους γεγονότων X_n , $n \geq 1$. Η απαριθμήτρια (σημειακή) διαδικασία $\{N(t) : t \geq 0\}$ της $\{S_n : n \geq 0\}$ ορίζεται από τη σχέση

$$N(t) = \sup\{n \geq 0 : S_n \leq t\}, \quad t \geq 0,$$

δηλαδή, η $N(t)$ μετράει το πλήθος των γεγονότων που έχουν συμβεί στο $(0, t]$.

Το πιο απλό μοντέλο για τη μελέτη γεγονότων που συμβαίνουν στον χρόνο, με τυχαίο τρόπο και κάποια μορφή πιθανοθεωρητικής περιοδικότητας, είναι μια ανανεωτική διαδικασία, της οποίας ειδική περίπτωση είναι η συχνά χρησιμοποιούμενη διαδικασία Poisson. Πιο συγκεκριμένα, έχουμε τον ακόλουθο ορισμό.

Ορισμός 1.1 (Ανανεωτική διαδικασία - Διαδικασία Poisson) Έστω $\{S_n : n \geq 0\}$ σημειακή διαδικασία με $S_0 = 0$ και $S_n = \sum_{i=1}^n X_i$, $n \geq 1$, όπου X_1, X_2, \dots ανεξάρτητες και ισόνομες μη-αρνητικές τυχαίες μεταβλητές με κατανομή $F_X(t) = \Pr[X_i \leq t]$, που αντιστοιχούν στους ενδιάμεσους χρόνους γεγονότων. Έστω, επίσης, $\{N(t)\}$ η αντίστοιχη απαριθμήτρια διαδικασία με

$$N(t) = \sup\{n \geq 0 : S_n \leq t\}, \quad t \geq 0$$

να είναι το πλήθος των γεγονότων που έχουν συμβεί στο $(0, t]$. Η $\{N(t) : t \geq 0\}$ αναφέρεται ως (απαριθμήτρια) ανανεωτική διαδικασία που γεννάται από την ακολουθία των τυχαίων μεταβλητών $\{X_n : n \geq 1\}$.

Μια ανανεωτική διαδικασία με εκθετικούς ενδιάμεσους χρόνους με παράμετρο λ λέγεται διαδικασία Poisson με ρυθμό λ . ◊

Στη Θεωρία Ουρών Αναμονής, η διαδικασία των αφίξεων είναι συνήθως μια ανανεωτική στοχαστική διαδικασία. Το γράμμα A της ονοματολογίας Kendall προσδιορίζει τον τύπο της ανανεωτικής διαδικασίας. Οι πιο συχνές τιμές που παίρνει στη βιβλιογραφία είναι GI (General independent) (ή G (General)), M (Memoryless, Markovian), D (Deterministic) και E_r (Erlang- r) για τις περιπτώσεις που οι ενδιάμεσοι χρόνοι μεταξύ των αφίξεων είναι γενικοί, εκθετικοί, σταθεροί και Erlang- r αντίστοιχα. Υπάρχουν βέβαια και άλλες τιμές για το γράμμα A που αντιστοιχούν σε κατανομές που εμφανίζονται σπανιότερα στη βιβλιογραφία.

Οι χρόνοι εξυπηρέτησης θεωρούνται στα κλασικά μοντέλα επίσης ανεξάρτητοι και ισόνομοι και αντιστοιχούν στο γράμμα B της ονοματολογίας Kendall, που παίρνει τις ίδιες τιμές με το γράμμα A. Οι τιμές GI και G για το A και το B σηματοδοτούν σε κάποια βιβλία ακριβώς το ίδιο, δηλαδή ανεξάρτητους ισόνομους χρόνους με γενική κατανομή. Σε κάποια, όμως, βιβλία, μόνο το GI σηματοδοτεί ανεξάρτητους και ισόνομους χρόνους, ενώ το G αναφέρεται σε γενικούς χρόνους που μπορεί να είναι και εξαρτημένοι. Στα παλαιά κλασικά συγγράμματα, που αναφέρονται σε συστήματα εξυπηρέτησης με ανεξάρτητους ισόνομους ενδιάμεσους χρόνους αφίξεων και χρόνους εξυπηρέτησεων, συνηθίζεται η τιμή GI για το A και η τιμή G για το B. Στο παρόν βιβλίο, θα χρησιμοποιούμε την τιμή G τόσο για το A, όσο και για το B, εφόσον αναφερόμαστε σε ανεξάρτητους και ισόνομους χρόνους.

Ο αριθμός των υπηρετών αναφέρεται στο πόσοι είναι οι παράλληλοι υπηρέτες που εξυπηρετούν τη ροή των πελατών που εισέρχεται στο σύστημα. Με την έννοια «παράλληλοι» υπηρέτες εννοούμε ότι υπάρχει μια κοινή ουρά για όλους και οι πελάτες πηγαίνουν στον πρώτο υπηρέτη που θα αδειάσει, αν όλοι οι υπηρέτες είναι απασχολημένοι, ή διαλέγουν στην τύχη κάποιον από τους άδειους υπηρέτες, αν υπάρχουν ελεύθεροι υπηρέτες. Ο αριθμός των υπηρετών αντιστοιχεί στον αριθμό c της ονοματολογίας Kendall.

Η χωρητικότητα του συστήματος εκφράζει το μέγιστο πλήθος πελατών που μπορεί να χωρέσει το σύστημα, συμπεριλαμβανομένων τόσο αυτών που περιμένουν να εξυπηρετηθούν όσο και αυτών που βρίσκονται σε διαδικασία εξυπηρέτησης. Αν ένα σύστημα έχει φτάσει στο μέγιστο της χωρητικότητάς του και αφιχθεί ένας πελάτης, τότε στα κλασικά μοντέλα ουρών αναμονής ο πελάτης απορρίπτεται και θεωρείται χαμένος για πάντα.

Φυσικά υπάρχουν και μοντέλα στα οποία οι πελάτες που αποχωρούν λόγω της περιορισμένης χωρητικότητας επανέρχονται αργότερα με την ελπίδα να υπάρχει διαθέσιμη θέση στο σύστημα. Στην περίπτωση αυτή μιλάμε για μοντέλα με επαναπροσπάθειες ή επαναδοκιμές (retrials). Σε τέτοια μοντέλα είναι απαραίτητο να αποσαφηνισθεί η διαδικασία με την οποία οι πελάτες επανέρχονται στο σύστημα και έτσι τα μοντέλα αυτά δεν περιγράφονται στο πλαίσιο της ονοματολογίας Kendall. Προς το παρόν, επομένως, μένουμε στο πλαίσιο των μοντέλων χωρίς επαναπροσπάθειες, όπου η χωρητικότητα του συστήματος αντιστοιχεί στον αριθμό k της ονοματολογίας του Kendall.

Η πειθαρχία ουράς είναι ο τρόπος με τον οποίο το σύστημα διαλέγει ποιον πελάτη θα εξυπηρετήσει, μόλις βρεθεί κάποιος διαθέσιμος υπηρέτης. Η πιο συνηθισμένη πειθαρχία ουράς είναι η FCFS (First-Come-First-Served) ή FIFO (First-In-First-Out), κατά την οποία οι πελάτες επιλέγονται να εξυπηρετηθούν σύμφωνα με τη σειρά της άφιξής τους. Έτσι, μόλις αδειάσει ένας υπηρέτης, επιλέγεται για εξυπηρέτηση ο πελάτης που έχει αφιχθεί πρώτος από όλους που περιμένουν. Η πειθαρχία αυτή μοιάζει η πιο δίκαιη με πρώτη ματιά και χρησιμοποιείται περισσότερο από οποιαδήποτε άλλη στην περίπτωση που οι πελάτες είναι άνθρωποι και μάλιστα έχουν οπτική επαφή με το τι συμβαίνει στο σύστημα (και επομένως μπορούν και βλέπουν πότε φθάνουν οι άλλοι πελάτες). Σε διάφορες εφαρμογές, πάντως, χρησιμοποιούνται και άλλες πειθαρχίες ουράς, όπως η LCFS (Last-Come-First-Served) κατά την οποία οι πελάτες εξυπηρετούνται αντίστροφα από τη σειρά άφιξής τους, η SIRO (Service-In-Random-Order) όπου οι πελάτες εξυπηρετούνται τυχαία, χωρίς να λαμβάνεται υπόψη η σειρά άφιξής τους, η SSTF (Shortest-Service-Time-First) όπου επιλέγεται προς εξυπηρέτηση ο πελάτης που έχει τον μικρότερο χρόνο εξυπηρέτησης κ.α. Υπάρχουν επίσης πειθαρχίες ουράς για την περίπτωση που υπάρχουν περισσότερα του ενός είδη πελατών και το σύστημα τους αντιμετωπίζει διαφορετικά. Στην περίπτωση αυτή, είναι δυνατόν κάποια είδη πελατών να έχουν προτεραιότητα έναντι κάποιων άλλων, οπότε μιλάμε για πειθαρχίες ουράς με προτεραιότητες. Γενικά, αν σκεφθούμε πρακτικές εφαρμογές των ουρών αναμονής θα συνειδητοποιήσουμε ότι υπάρχει μεγάλη ποικιλία στις πειθαρχίες ουράς που χρησιμοποιούνται (π.χ. σκεφτείτε τα ταμεία για πελάτες με λίγα προϊόντα στα supermarkets, τα ταμεία για επιχειρηματικές συναλλαγές στις τράπεζες, τις κρατήσεις θέσεων σε εστιατόρια, τα τηλεφωνικά κέντρα που εξυπηρετούν πελάτες σε περισσότερες από μια γλώσσες κλπ.).

Η χωρητικότητα του συστήματος k και/ή η πειθαρχία ουράς μπορεί να παραλείπονται στην ονοματολογία Kendall. Αυτό συμβαίνει στην περίπτωση που έχουμε απεριόριστη χωρητικότητα ($k = \infty$) ή πειθαρχία FCFS αντίστοιχα.

Παράδειγμα 1.2 Η $M/M/1$ ουρά είναι ένα σύστημα εξυπηρέτησης με Poisson διαδικασία αφίξεων (δηλ. ανεξάρτητους εκθετικούς ενδιάμεσους χρόνους αφίξεων), εκθετικούς χρόνους εξυπηρέτησης και 1 υπηρέτη που έχει άπειρη χωρητικότητα και λειτουργεί υπό την πειθαρχία ουράς FCFS. ◊

Παράδειγμα 1.3 Η $MG/E_2/1/5$ (SIRO) ουρά είναι ένα σύστημα εξυπηρέτησης με ανανεωτική διαδικασία αφίξεων (δηλ. ανεξάρτητους και ισόνομους ενδιάμεσους χρόνους αφίξεων), Erlang-2 χρόνους εξυπηρέτησης και 1 υπηρέτη που έχει χωρητικότητα για 5 πελάτες και λειτουργεί υπό την πειθαρχία ουράς SIRO. ◊

Πολλές φορές η διαδικασία αφίξεων και/ή η κατανομή των χρόνων εξυπηρέτησης δεν είναι ακριβώς γνωστή. Στην περίπτωση αυτή μπορεί να παρέχεται κάποια αδρή πληροφορία για το πώς έρχονται και πώς εξυπηρετούνται οι πελάτες. Π.χ., μπορεί να δίνεται ο μέσος ενδιάμεσος χρόνος a μεταξύ δυο διαδοχικών αφίξεων και ο μέσος χρόνος εξυπηρέτησης b . Ισοδύναμα, μπορεί να δίνεται ο ρυθμός αφίξεων $\lambda = 1/a$ και ο ρυθμός εξυπηρέτησης $\mu = 1/b$. Τα a και b έχουν τη φυσική έννοια της (μέσης) περιόδου των διαδικασιών των αφίξεων και των εξυπηρέτησεων αντίστοιχα, ενώ τα λ και μ αντιστοιχούν στη φυσική έννοια της συχνότητας (ρυθμού) των αφίξεων και των εξυπηρέτησεων. Με τόσο ελλιπή πληροφορία, βεβαίως, τα αποτελέσματα που θα προκύψουν από τη μαθηματική μελέτη μπορεί να μην περιγράφουν τα μέτρα απόδοσης ενός συστήματος με ακρίβεια. Για την εξαγωγή ασφαλέστερων συμπερασμάτων χρειάζεται να είναι γνωστή η κατανομή των αντίστοιχων χρόνων ή τουλάχιστον κάποιες ροπές ανώτερης τάξης. Μετά τη μέση τιμή, η διασπορά των χρόνων μεταξύ των αφίξεων και/ή των χρόνων εξυπηρέτησης επηρεάζει σημαντικά την απόδοση ενός συστήματος.

1.2 Αναγεννητικότητα και μέτρα απόδοσης συστήματος

Αφού περιγραφεί ένα σύστημα εξυπηρέτησης, το πρόβλημα που τίθεται είναι να προβλέψουμε πώς θα συμπεριφέρεται. Στη Θεωρία Ουρών Αναμονής, εστιάζουμε κυρίως σε συστήματα που εμφανίζουν μια πιθανοθεωρητική περιοδικότητα, δηλαδή ο τρόπος λειτουργίας τους παραμένει ομοιόμορφος στον χρόνο και ξαναρχίζει από την αρχή κάθε φορά που ολοκληρώνεται ένας κύκλος λειτουργίας. Το μαθηματικό πρότυπο για την περιγραφή ενός συστήματος που εξελίσσεται στον χρόνο με τυχαίο τρόπο με κάποια μορφή περιοδικότητας είναι μια αναγεννητική διαδικασία. Διαισθητικά μια διαδικασία $\{X(t)\}$ είναι αναγεννητική, αν υπάρχει μια ανανεωτική διαδικασία $\{N(t)\}$ με χρόνους γεγονότων S_n , έτσι ώστε η εξέλιξη της $\{X(t)\}$ σε κάθε ενδιάμεσο χρόνο (ανανεωτικό κύκλο) της $\{N(t)\}$ να είναι πιθανοθεωρητικά ίδια και ανεξάρτητη από ό,τι συμβαίνει στους άλλους κύκλους. Πιο συγκεκριμένα, έχουμε τον ακόλουθο ορισμό:

Ορισμός 1.4 (Αναγεννητική διαδικασία) Μια στοχαστική διαδικασία $\{X(t)\}$ λέγεται αναγεννητική διαδικασία, αν υπάρχει μια μη-αρνητική τυχαία μεταβλητή S_1 με $\Pr[S_1 = 0] < 1$ και $\Pr[S_1 < \infty] = 1$, τέτοια ώστε

- (i) οι $\{X(t) : t \geq 0\}$ και $\{X(t + S_1) : t \geq 0\}$ να είναι στοχαστικά ισοδύναμες (δηλαδή για οποιοδήποτε επιλογές χρονικών στιγμών να έχουν τις ίδιες κατανομές) και
- (ii) οι $\{X(t) : 0 \leq t < S_1\}$ και $\{X(t + S_1) : t \geq 0\}$ να είναι ανεξάρτητες. ◊

Ο ορισμός αυτός συνεπάγεται την ύπαρξη μιας αύξουσας ακολουθίας χρόνων S_1, S_2, \dots , τέτοιων ώστε οι διαδικασίες $\{X(t) : t \geq 0\}$ και $\{X(t + S_n) : t \geq 0\}$ να είναι στοχαστικά ισοδύναμες. Επιπλέον, για κάθε $n \geq 1$, έχουμε ότι οι $\{X(t) : 0 \leq t < S_n\}$ και $\{X(t + S_n) : t \geq 0\}$ είναι ανεξάρτητες. Επομένως, οι χρόνοι $X_n = S_n - S_{n-1}$, $n \geq 1$ (με τη σύμβαση $S_0 = 0$) είναι ανεξάρτητοι και ισόνομοι και επομένως οι S_n μπορούν να θεωρηθούν ως οι χρόνοι των γεγονότων μιας ανανεωτικής διαδικασίας. Σε κάθε τέτοιο γεγονός, η διαδικασία $\{X(t)\}$ «ξεχνά» το παρελθόν της (αφού οι $\{X(t) : 0 \leq t < S_n\}$ και $\{X(t + S_n) : t \geq 0\}$ είναι ανεξάρτητες) και ξαναρχίζει την εξέλιξή της σαν να ξεκινούσαν όλα όπως τη χρονική στιγμή 0 (αφού οι $\{X(t) : t \geq 0\}$ και $\{X(t + S_n) : t \geq 0\}$ είναι στοχαστικά ισοδύναμες). Για το λόγο αυτό λέμε ότι η $\{X(t)\}$ αναγεννάται στοχαστικά σε αυτά τα σημεία και αναφέρεται ως αναγεννητική διαδικασία.

Στα πλαίσια της Θεωρίας Ουρών Αναμονής, οι στιγμές S_1, S_2, \dots που αναγεννάται ένα σύστημα εξυπηρέτησης είναι συνήθως οι στιγμές που φθάνει ένας πελάτης που βρίσκει κενό το σύστημα. Στις στιγμές αυτές αρχίζει ταυτόχρονα να «τρέχει» ένας νέος ενδιάμεσος χρόνος αφίξεων και ένας χρόνος εξυπηρέτησης και όλα αρχίζουν από την αρχή. Οπότε η λειτουργία ενός συστήματος χωρίζεται σε κύκλους λειτουργίας (busy cycles) και ένας τυπικός κύκλος αντιστοιχεί στο διάστημα από τη στιγμή που ένας πελάτης φθάνει σε κενό σύστημα μέχρι την επόμενη άφιξη πελάτη που βρίσκει το σύστημα κενό. Στα πλαίσια του παρόντος συγγράμματος θα δούμε ότι όλα τα συστήματα που θα μελετήσουμε είναι αναγεννητικά και οι στιγμές της αναγέννησής τους είναι ακριβώς οι διαδοχικές στιγμές αφίξεων πελατών που βρίσκουν κενό σύστημα.

Θα δούμε τώρα πώς η αποτίμηση απόδοσης των συστημάτων εξυπηρέτησης διευκολύνεται λόγω της αναγεννητικότητάς τους. Τυπικά ερωτήματα αποτίμησης απόδοσης στη Θεωρία Ουρών Αναμονής είναι τα εξής:

1. Πόσοι πελάτες θα βρίσκονται στο σύστημα κατά μέσο όρο μια τυχούσα χρονική στιγμή;
2. Πόσο χρόνο θα περάσει στο σύστημα κατά μέσο όρο ένας πελάτης;
3. Ποιο ποσοστό του χρόνου του θα βρίσκεται απασχολημένος ένας υπηρέτης που δουλεύει σε ένα συγκεκριμένο σύστημα;

Όπως βλέπουμε, υπάρχουν ερωτήματα που απασχολούν τον διαχειριστή του συστήματος, που βλέπει το σύστημα συνολικά, σαν εξωτερικός παρατηρητής (ερώτημα 1), ερωτήματα που απασχολούν τους πελάτες, που επιδρούν στο σύστημα μόνο παροδικά και κατόπιν φεύγουν (ερώτημα 2) και ερωτήματα που απασχολούν τους υπηρέτες, που απασχολούνται για μεγάλα χρονικά διαστήματα στο σύστημα (ερώτημα 3). Είναι σημαντικό να κατανοηθεί ότι αυτοί οι παράγοντες του συστήματος (διαχειριστής, πελάτες και υπηρέτες) έχουν

διαφορετικές οπτικές και για τον λόγο αυτό υπάρχουν μέτρα απόδοσης του συστήματος που σχετίζονται με την οπτική του καθενός. Για να αποκτήσει ένας ερευνητής μια συνολική εικόνα της απόδοσης ενός συστήματος θα πρέπει να το μελετήσει από όλες αυτές τις οπτικές.

Για τον διαχειριστή του συστήματος η πιο σημαντική πληροφορία είναι ο αριθμός των πελατών που βρίσκονται στο σύστημα μια τυχαία χρονική στιγμή. Έτσι ορίζουμε

- $Q(t)$ τον αριθμό των πελατών στο σύστημα τη στιγμή t ,
- $Q_q(t)$ τον αριθμό των πελατών στον χώρο αναμονής τη στιγμή t (ο δείκτης q μπαίνει για να θυμίζει ότι αναφερόμαστε στο πλήθος των πελατών στην ουρά ($q \rightarrow$ queue)),
- $Q_s(t)$ τον αριθμό των πελατών στον χώρο εξυπηρέτησης τη στιγμή t (ο δείκτης s μπαίνει για να θυμίζει ότι αναφερόμαστε στο πλήθος των πελατών υπό εξυπηρέτηση ($s \rightarrow$ service)).

Φυσικά ισχύει

$$Q(t) = Q_q(t) + Q_s(t).$$

Όπως αναφέραμε, ένα σύστημα εξυπηρέτης συνήθως αναγεννάται στοχαστικά στις στιγμές αφίξεων πελατών που βρίσκουν το σύστημα κενό, οπότε η στοχαστική διαδικασία του αριθμού των πελατών, $\{Q(t)\}$, είναι αναγεννητική, αφού κάθε φορά που ένας πελάτης φθάνει σε κενό σύστημα όλα ξαναρχίζουν από την αρχή. Αυτό μας επιτρέπει να χρησιμοποιήσουμε ένα κεντρικό αποτέλεσμα των αναγεννητικών διαδικασιών που δίνει εναλλακτικούς τρόπους για τη μελέτη της μακροπρόθεσμης συμπεριφοράς μιας αναγεννητικής διαδικασίας. Συγκεκριμένα, έχουμε το ακόλουθο:

Θεώρημα 1.5 (Εργοδικό θεώρημα αναγεννητικών διαδικασιών) Έστω μια αναγεννητική διαδικασία $\{X(t)\}$ με χώρο καταστάσεων \mathbb{R} και δεξιά συνεχείς πραγματοποιήσεις με αριστερά όρια. Έστω, επίσης, S ο χρόνος της αναγέννησης της $\{X(t)\}$ και $1_{\{X(u) \leq x\}}$ η δείκτρια τυχαία μεταβλητή του ενδεχομένου $\{X(u) \leq x\}$ «η $X(u)$ να μην υπερβαίνει το x ». Τότε, η τυχαία μεταβλητή $\int_0^S 1_{\{X(u) \leq x\}} du$ εκφράζει τον συνολικό χρόνο στο $(0, S]$ που η $\{X(t)\}$ είναι το πολύ x . Αν $E[S] < \infty$, ορίζουμε

$$F_X(x) = \frac{E[\int_0^S 1_{\{X(u) \leq x\}} du]}{E[S]}, \quad x \in \mathbb{R}, \tag{1.1}$$

την κατανομή ισορροπίας της $\{X(t)\}$. Τότε έχουμε

$$\lim_{t \rightarrow \infty} \frac{\int_0^t 1_{\{X(u) \leq x\}} du}{t} = F_X(x), \quad \text{με πιθανότητα } 1, \quad x \in \mathbb{R}, \tag{1.2}$$

(δηλαδή το μακροπρόθεσμο ποσοστό του χρόνου που η $\{X(t)\}$ είναι μικρότερη ή ίση με x είναι ίσο με $F_X(x)$),

$$\lim_{t \rightarrow \infty} \frac{E[\int_0^t 1_{\{X(u) \leq x\}} du]}{t} = F_X(x), \quad x \in \mathbb{R}, \tag{1.3}$$

(δηλαδή το μακροπρόθεσμο μέσο ποσοστό του χρόνου που η $\{X(t)\}$ είναι μικρότερη ή ίση με x είναι ίσο με $F_X(x)$), και

$$\lim_{t \rightarrow \infty} \frac{\int_0^t \Pr[X(u) \leq x] du}{t} = F_X(x), \quad x \in \mathbb{R}, \tag{1.4}$$

(δηλαδή η C-οριακή πιθανότητα η $\{X(t)\}$ να είναι μικρότερη ή ίση με x είναι ίση με $F_X(x)$).

Αν, επιπλέον, η κατανομή του S είναι απεριοδική, δηλ. δεν υπάρχει $a > 0$ ώστε $\sum_{n=0}^{\infty} \Pr[S = na] = 1$, τότε

$$\lim_{t \rightarrow \infty} \Pr[X(t) \leq x] = F_X(x), \quad x \in \mathbb{R}, \quad (1.5)$$

(δηλαδή η οριακή πιθανότητα η $\{X(t)\}$ να είναι μικρότερη ή ίση με x δίνεται από την $F_X(x)$).

Επίσης, έστω $c : \mathbb{R} \rightarrow \mathbb{R}$ μια φραγμένη άνω ή κάτω συνάρτηση, όπου $c(u)$ εκφράζει το κόστος ανά χρονική μονάδα παραμονής της $\{X(t)\}$ στην κατάσταση u (δηλαδή τον ρυθμό κόστους παραμονής της $\{X(t)\}$ στην κατάσταση u), και

$$C(t) = \int_0^t c(X(u))du, \quad t \geq 0, \quad (1.6)$$

είναι το συνολικό κόστος που συσσωρεύεται στο $(0, t]$. Έστω ότι X είναι μια τυχαία μεταβλητή με την κατανομή ισορροπίας (1.1) της $\{X(t)\}$.

(i) Για τον μακροπρόθεσμο ρυθμό κόστους έχουμε

$$\lim_{t \rightarrow \infty} \frac{C(t)}{t} = E[c(X)], \quad \text{με πιθανότητα } 1. \quad (1.7)$$

(ii) Για τον μακροπρόθεσμο μέσο ρυθμό κόστους έχουμε

$$\lim_{t \rightarrow \infty} \frac{E[C(t)]}{t} = E[c(X)]. \quad (1.8)$$

Η μέση τιμή $E[c(X)]$ υπολογίζεται ως

$$E[c(X)] = \frac{E[\int_0^S c(X(u))du]}{E[S]} = \frac{E[C(S)]}{E[S]} \quad (1.9)$$

$$= \int_{-\infty}^{\infty} c(x)dF_X(x) \quad (1.10)$$

$$= \begin{cases} \sum_x c(x)f_X(x) & \text{αν } X \text{ διακριτή με συν. πιθ. } f_X(x), \\ \int_{-\infty}^{\infty} c(x)f_X(x)dx & \text{αν } X \text{ συνεχής με συν. πυκν. πιθ. } f_X(x), \end{cases} \quad (1.11)$$

όπου S ο χρόνος 1ης αναγέννησης της $\{X(t)\}$. Αν επιπλέον η κατανομή του S είναι απεριοδική, έχουμε και ότι

$$E[c(X)] = \lim_{t \rightarrow \infty} E[c(X(t))].$$

◊

Το εργοδικό θεώρημα λέει ότι κάτω από πολύ γενικές συνθήκες η οριακή πιθανότητα μια αναγεννητική διαδικασία $\{X(t)\}$ να παίρνει τιμές μικρότερες ή ίσες του x ισούται με το μακροπρόθεσμο ποσοστό του χρόνου που παίρνει τιμές μικρότερες ή ίσες του x . Γενικότερα, η οριακή πιθανότητα μια αναγεννητική διαδικασία $\{X(t)\}$ να παίρνει τιμές σε ένα σύνολο ισούται με το μακροπρόθεσμο ποσοστό του χρόνου που παίρνει τιμές σε αυτό το σύνολο. Και μάλιστα το ποσοστό αυτό μπορεί να υπολογιστεί επικεντρώνοντας την προσοχή μας σε έναν αναγεννητικό κύκλο. Επιπλέον, ο υπολογισμός του μακροπρόθεσμου ρυθμού κόστους μιας αναγεννητικής διαδικασίας, ανάγεται στην εύρεση μιας μέσης τιμής. Αν η $\{X(t)\}$ παίρνει διακριτές τιμές $0, 1, 2, \dots$ είναι χρήσιμο να σκεφτόμαστε και την εκδοχή του θεωρήματος που το ενδεχόμενο $\{X(u) \leq x\}$ αντικαθίσταται από το ενδεχόμενο $\{X(u) = j\}$.

Για να καταλάβουμε τη σημασία του εργοδικού θεωρήματος αναγεννητικών διαδικασιών για τη μελέτη των συστημάτων εξυπηρέτησης, ας θεωρήσουμε την αναγεννητική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών και ας δούμε πώς ερμηνεύονται οι ποσότητες που εμφανίζονται στο θεώρημα. Αυτό που μας ενδιαφέρει

είναι η μακροπρόθεσμη συμπεριφορά της $\{Q(t)\}$, οπότε επικεντρωνόμαστε στην εύρεση του μακροπρόθεσμου ποσοστού του χρόνου που βρίσκονται j πελάτες στο σύστημα, για τις διάφορες τιμές του j . Το ποσοστό αυτό εκφράζεται μαθηματικά ως

$$p_j = \lim_{t \rightarrow \infty} \frac{\int_0^t 1_{\{Q(u)=j\}} du}{t}, \quad (1.12)$$

όπου $1_{\{Q(u)=j\}}$ είναι η δείκτρια τυχαία μεταβλητή του ενδεχομένου $\{Q(u) = j\}$ που παίρνει την τιμή 1 όταν $Q(u) = j$ και την τιμή 0 διαφορετικά. Παρατηρήστε ότι ο αριθμητής του κλάσματος εκφράζει τον συνολικό χρόνο που στο σύστημα βρίσκονται j πελάτες στο χρονικό διάστημα $(0, t]$, ενώ ο παρονομαστής είναι το μήκος του αντίστοιχου διαστήματος, οπότε πράγματι το κλάσμα εκφράζει το ποσοστό του χρόνου που στο σύστημα βρίσκονται j πελάτες στο διάστημα $(0, t]$. Θέτοντας $t \rightarrow \infty$, αναπαριστούμε μαθηματικά την έννοια «μακροπρόθεσμο». Το εργοδικό θεώρημα εκμεταλλεύεται την αναγεννητικότητα της διαδικασίας $\{Q(t)\}$ και μας επιτρέπει να υπολογίσουμε το ποσοστό αυτό περιοριζόμενοι σε έναν κύκλο λειτουργίας του συστήματος. Έχουμε, δηλαδή, ότι

$$p_j = \frac{E[\int_0^Z 1_{\{Q(u)=j\}} du]}{E[Z]}, \quad (1.13)$$

όπου Z είναι η διάρκεια ενός κύκλου λειτουργίας του συστήματος. Η έκφραση αυτή δίνει το p_j ως τον μέσο χρόνο που η $\{Q(t)\}$ μένει στην κατάσταση j σε έναν κύκλο λειτουργίας προς τη μέση διάρκεια του κύκλου, αντιστοιχώντας και πάλι σε ένα είδος ποσοστού χρόνου.

Ένας εναλλακτικός τρόπος να σκεφτόμαστε την πιθανότητα ο αριθμός των πελατών να είναι j σε μια τυχούσα χρονική στιγμή είναι ο εξής: Επιλέγουμε τυχαία μια στιγμή στο $(0, t]$, σύμφωνα με μια ομοιόμορφη τυχαία μεταβλητή U_t , για μεγάλο t , και κατόπιν παρατηρούμε την $\{Q(t)\}$ την επιλεγείσα χρονική στιγμή και μας ενδιαφέρει η πιθανότητα να βρίσκεται στην κατάσταση j . Το εργοδικό θεώρημα βεβαιώνει ότι η πιθανότητα αυτή ισούται με το μακροπρόθεσμο ποσοστό του χρόνου που υπάρχουν j πελάτες στο σύστημα. Έχουμε, δηλαδή, ότι

$$p_j = \lim_{t \rightarrow \infty} \Pr[Q(U_t) = j] = \lim_{t \rightarrow \infty} \int_0^t \Pr[Q(u) = j] \frac{1}{t} du, \quad j \geq 0, \quad (1.14)$$

αφού μια συνάρτηση πυκνότητας πιθανότητας της U_t είναι ίση με $\frac{1}{t}$ στο $(0, t]$ και 0 οπουδήποτε αλλού.

Ένας ακόμη τρόπος να σκεφτόμαστε την πιθανότητα ο αριθμός των πελατών ενός συστήματος να είναι j μια τυχούσα χρονική στιγμή είναι ως την οριακή πιθανότητα να υπάρχουν j πελάτες στο σύστημα, καθώς ο χρόνος μεγαλώνει, δηλαδή ως την $\lim_{t \rightarrow \infty} \Pr[Q(t) = j]$. Το εργοδικό θεώρημα διαβεβαιώνει ότι η οριακή πιθανότητα της κατάστασης j (εφόσον υπάρχει, πράγμα που εξασφαλίζεται όταν ο κύκλος λειτουργίας του συστήματος έχει απεριοδική κατανομή) ισούται με το μακροπρόθεσμο ποσοστό του χρόνου που η $\{Q(t)\}$ περνάει στην j , δηλαδή ότι

$$p_j = \lim_{t \rightarrow \infty} \Pr[Q(t) = j], \quad j \geq 0. \quad (1.15)$$

Επομένως, το εργοδικό θεώρημα εφαρμόζόμενο στην αναγεννητική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών ενός συστήματος δίνει τέσσερις διαφορετικές ερμηνείες και τρόπους υπολογισμού για την ποσότητα p_j που διαισθητικά αντιστοιχεί στην πιθανότητα να υπάρχουν j πελάτες στο σύστημα:

1. Η p_j είναι ο λόγος του μέσου χρόνου που υπάρχουν j πελάτες στο σύστημα σε έναν κύκλο λειτουργίας προς τη μέση διάρκεια του κύκλου λειτουργίας.
2. Η p_j είναι το μακροπρόθεσμο ποσοστό του χρόνου που υπάρχουν j πελάτες στο σύστημα, για σχεδόν όλες τις πραγματοποιήσεις της $\{Q(t)\}$. Το ενδεχόμενο να μην ισχύει αυτό για κάποια πραγματοποίηση έχει πιθανότητα 0.
3. Η p_j είναι η πιθανότητα να βρίσκονται j πελάτες στο σύστημα, αν το κοιτάξουμε σε μια τυχαία, ομοιόμορφα επιλεγμένη χρονική στιγμή σε ένα διάστημα μεγάλου μήκους.

4. Η p_j είναι η πιθανότητα να βρίσκονται j πελάτες στο σύστημα, αν το κοιτάξουμε μια χρονική στιγμή μετά την παρέλευση μεγάλου χρονικού διαστήματος (εφόσον ο κύκλος λειτουργίας του συστήματος έχει απεριοδική κατανομή).

Θα αναφέρουμε την $(p_j : j \geq 0)$ ως κατανομή ισορροπίας, οριακή κατανομή ή στάσιμη κατανομή της $\{Q(t)\}$ και θα γράφουμε εκφράσεις του τύπου $\Pr[Q = j] = p_j$, εννοώντας ότι η Q είναι μια τυχαία μεταβλητή που εκφράζει τον αριθμό των πελατών στο σύστημα σε κατάσταση ισορροπίας. Θα μπορούσε να αναρωτηθεί κανείς γιατί δεν επικεντρωνόμαστε στη μελέτη της $Q(t)$ για κάθε συγκεκριμένο χρόνο t που έχει περάσει από την έναρξη της λειτουργίας του συστήματος, δηλαδή στην εύρεση της λεγόμενης μεταβατικής κατανομής της $\{Q(t)\}$ τη στιγμή t , $\mathbf{p}(t) = (p_j(t) : j \geq 0)$ με $p_j(t) = \Pr[Q(t) = j]$. Καταρχήν, ο υπολογισμός αυτός είναι ένα μαθηματικά απαιτητικό πρόβλημα, που ακόμη και για πολύ απλά συστήματα δεν επιδέχεται λύση σε αναλυτική μορφή. Αλλά το σημαντικότερο είναι ότι μετά την παρέλευση μικρού σχετικά χρόνου η επίδραση της αρχικής κατάστασης του συστήματος (π.χ. αρχικά κενό σύστημα) χάνεται και η κατανομή της $\{Q(t)\}$ συγκλίνει στην κατανομή ισορροπίας.

Για μια εκτενέστερη συζήτηση πάνω στο θέμα των αναγεννητικών διαδικασιών, ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει στα βιβλία Wolff 1989, Φακίνος 2003, Tijms 2008 και Kulkarni 2010. Σχεδόν όλα τα συστήματα που θα μελετήσουμε είναι αναγεννητικά και επιπλέον ο κύκλος λειτουργίας του συστήματος έχει απεριοδική κατανομή. Πράγματι, η απεριοδικότητα του κύκλου λειτουργίας εξασφαλίζεται όταν η κατανομή των ενδιάμεσων χρόνων αφίξεων ή των χρόνων εξυπηρέτησης είναι συνεχής. Κατά συνέπεια, οι οριακές πιθανότητες που αφορούν τον αριθμό των πελατών στο σύστημα ισούνται με τα αντίστοιχα μακροπρόθεσμα ποσοστά του χρόνου που το σύστημα έχει συγκεκριμένο αριθμό πελατών, όπως είδαμε παραπάνω. Για τον λόγο αυτό, σε ό,τι ακολουθεί θα χρησιμοποιούμε ελεύθερα κάποια από τις ισοδύναμες εκφράσεις, χωρίς περαιτέρω μνεία. Αυτό θα συμβαίνει και για άλλα μεγέθη που αφορούν άλλες τυχαίες μεταβλητές που έχουν ενδιαφέρον για τη μελέτη των συστημάτων. Π.χ., συμβολίζοντας με Q την τυχαία μεταβλητή που έχει την κατανομή ισορροπίας $(p_j : j \geq 0)$ της $\{Q(t)\}$ και θεωρώντας μια συνάρτηση κόστους $c(u)$, θα γράφουμε

$$E[c(Q)] = \lim_{t \rightarrow \infty} E[c(Q(t))] = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t c(Q(u)) du, \quad (1.16)$$

χρησιμοποιώντας το δεύτερο μέρος του εργοδικού θεωρήματος που εισάγει την ιδέα της συνάρτησης κόστους και του ρυθμού κόστους. Αν έχουμε προσδιορίσει την $(p_j : j \geq 0)$ (πράγμα που δεν είναι πάντα εύκολο) τότε μπορούμε να υπολογίσουμε την $E[c(Q)]$ από τη σχέση $E[c(Q)] = \sum_{j=0}^{\infty} c(j)p_j$. Ειδικότερα, για $c(u) = u$, έχουμε $E[Q] = \sum_{j=0}^{\infty} jp_j$. Σε κάποιες περιπτώσεις η $E[Q]$ μπορεί να υπολογιστεί και απευθείας με πιθανοθεωρητικά επιχειρήματα, χρησιμοποιώντας κάποια βασικά αποτελέσματα που θα παρουσιαστούν παρακάτω. Ομοίως με την $E[Q]$ μπορούμε να υπολογίσουμε τις $E[Q_q]$ και $E[Q_s]$, δεδομένου ότι οι αντίστοιχες διαδικασίες $\{Q_q(t)\}$ και $\{Q_s(t)\}$ είναι επίσης αναγεννητικές.

Για τους πελάτες το πιο σημαντικό μέτρο απόδοσης είναι ο χρόνος που παραμένουν στο σύστημα. Έτσι ορίζουμε

- S_n τον χρόνο παραμονής στο σύστημα του n -οστού πελάτη,
- W_n τον χρόνο αναμονής στην ουρά (μέχρι να αρχίσει η εξυπηρέτηση) του n -οστού πελάτη,
- B_n τον χρόνο εξυπηρέτησης του n -οστού πελάτη.

Προφανώς έχουμε

$$S_n = W_n + B_n.$$

Ενδιαφερόμαστε και πάλι για το τι συμβαίνει σε κατάσταση ισορροπίας και γι αυτό ενδιαφερόμαστε για την κατανομή ισορροπίας του χρόνου παραμονής. Και πάλι υπάρχουν πολλαπλές ερμηνείες για την κατανομή αυτή που συμπίπτουν όταν το σύστημα έχει αναγεννητικό χαρακτήρα και ισχύει η υπόθεση της απεριοδικότητας των αναγεννητικών κύκλων. Έτσι για την κατανομή ισορροπίας $F_S(x)$ του χρόνου παραμονής πελάτη έχουμε

$$\begin{aligned}
F_S(x) &= \frac{1}{E[A(Z)]} E \left[\sum_{k=1}^{A(Z)} 1_{\{S_k \leq x\}} \right] \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n 1_{\{S_k \leq x\}} \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \Pr[S_k \leq x] \\
&= \lim_{n \rightarrow \infty} \Pr[S_n \leq x], \quad x \geq 0,
\end{aligned} \tag{1.17}$$

όπου Z ο πρώτος κύκλος λειτουργίας και $A(Z)$ το πλήθος των αφίξεων σε αυτόν. Εδώ, η $1_{\{S_k \leq x\}}$ είναι η δείκτρια συνάρτηση του ενδεχομένου $\{S_k \leq x\}$ (ο k -οστός πελάτης να παραμείνει στο σύστημα το πολύ για χρόνο x) που παίρνει την τιμή 1 όταν $S_k \leq x$ και την τιμή 0 διαφορετικά. Οπότε έχουμε ότι το μακροπρόθεσμο ποσοστό των πελατών που παραμένουν στο σύστημα για χρόνο το πολύ x ισούται με την οριακή πιθανότητα ο n -οστός πελάτης να μείνει στο σύστημα για χρόνο το πολύ x , για $n \rightarrow \infty$. Η ισότητα που αφορά το μακροπρόθεσμο ποσοστό των πελατών που παραμένουν στο σύστημα για χρόνο το πολύ x πρέπει να ερμηνευθεί ότι ισχύει με πιθανότητα 1, καθώς το ποσοστό είναι τυχαία μεταβλητή.

Συμβολίζοντας με S την τυχαία μεταβλητή που έχει την κατανομή ισορροπίας ($F_S(x) : x \geq 0$) για τον οριακό μέσο χρόνο παραμονής έχουμε αντίστοιχα ότι

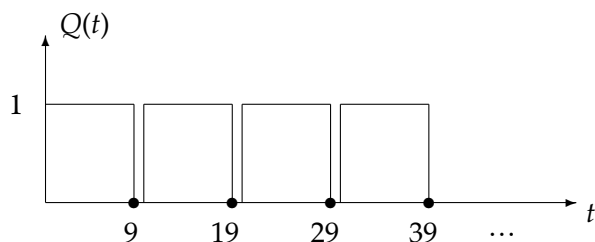
$$E[S] = \frac{1}{E[A(Z)]} E \left[\sum_{k=1}^{A(Z)} S_k \right] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n S_k = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n E[S_k] = \lim_{n \rightarrow \infty} E[S_n], \tag{1.18}$$

όπου και πάλι η ισότητα ισχύει με πιθανότητα 1 για τον μακροπρόθεσμο μέσο

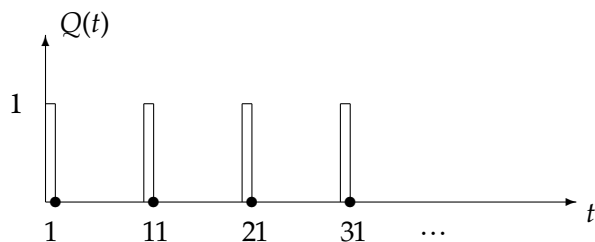
$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n S_k,$$

που είναι τυχαία μεταβλητή. Ο υπολογισμός του $E[S]$ είναι εύκολος αν έχουμε υπολογίσει την κατανομή ($F_S(x) : x \geq 0$) ή την αντίστοιχη συνάρτηση πυκνότητας πιθανότητας ($f_S(x) : x \geq 0$). Πράγματι αρκεί να χρησιμοποιήσουμε τη σχέση $E[S] = \int_0^{\infty} x f_S(x) dx = \int_0^{\infty} (1 - F_S(x)) dx$ και επομένως πρόκειται για έναν υπολογισμό ρουτίνας (ο οποίος μπορεί πάντως να έχει αρκετές πράξεις). Σε κάποιες περιπτώσεις, όμως, ο $E[S]$ μπορεί να υπολογιστεί και απευθείας, ταυτόχρονα με την $E[Q]$, με πιθανοθεωρητικά επιχειρήματα.

Ένα άλλο σημαντικό μέγεθος που αξίζει μελέτης είναι ο κύκλος απασχόλησης ή λειτουργίας (busy cycle) του συστήματος που ορίζεται ως το διάστημα από την άφιξη ενός πελάτη που βρίσκει το σύστημα κενό, μέχρι την επόμενη άφιξη πελάτη που θα βρει το σύστημα κενό. Κάθε τέτοιος κύκλος αρχίζει με ένα χρονικό διάστημα που το σύστημα είναι συνεχώς απασχολημένο. Το διάστημα αυτό αναφέρεται ως μια περίοδος συνεχούς λειτουργίας του συστήματος (busy period). Κατόπιν το σύστημα παραμένει κενό μέχρι να εμφανιστεί ο επόμενος πελάτης. Το διάστημα αυτό αναφέρεται ως μια περίοδος αργίας (idle period) του συστήματος. Οι διάρκειες των περιόδων συνεχούς λειτουργίας Y_n , αργίας I_n , και των κύκλων απασχόλησης $Z_n = Y_n + I_n$ ενδιαφέρουν κυρίως από την οπτική σκοπιά των υπηρετών και του διαχειριστή του συστήματος αφού οι ποσότητες αυτές είναι σημαντικές για να ληφθούν αποφάσεις σχετικά με τη συντήρηση του συστήματος ή με διακοπές-διαλείμματα των υπηρετών. Για τον λόγο αυτό ενδιαφερόμαστε για τη μελέτη των οριακών κατανομών $F_Z(x)$, $F_Y(x)$ και $F_I(x)$, ή τουλάχιστον των αντίστοιχων μέσων τιμών τους $E[Z]$, $E[Y]$ και $E[I]$.



Σχήμα 1.1: Εξέλιξη του πλήθους πελατών σε σύστημα D/D/1 με ενδιάμεσους χρόνους αφίξεων $a = 10$ και χρόνους εξυπηρέτησης $b = 9$.



Σχήμα 1.2: Εξέλιξη του πλήθους πελατών σε σύστημα D/D/1 με ενδιάμεσους χρόνους αφίξεων $a = 10$ και χρόνους εξυπηρέτησης $b = 1$.

1.3 Εμφυτευμένες διαδικασίες σε στιγμές αφίξεων και αναχωρήσεων

Όπως είπαμε παραπάνω, ο διαχειριστής ενός συστήματος, που μπορεί να θεωρηθεί ως ένας εξωτερικός παρατηρητής του συστήματος, έχει μια διαφορετική αντίληψη από τους πελάτες του συστήματος όσον αφορά τον συνωστισμό. Για τον λόγο αυτό ορίσαμε και διαφορετικά μέτρα απόδοσης που αντιστοιχούν στις διαφορετικές οπτικές. Για να γίνει περισσότερο κατανοητή η διαφορά των δυο οπτικών, διαχειριστή-εξωτερικού παρατηρητή και πελατών ας δούμε ένα συγκεκριμένο παράδειγμα συστήματος εξυπηρέτησης υπό τις δύο οπτικές.

Παράδειγμα 1.6 (Διαφορά οπτικής διαχειριστή και πελατών) Ας φανταστούμε ότι έχουμε ένα D/D/1 σύστημα και ας εξετάσουμε τι αντιλαμβάνεται ο διαχειριστής και τι οι πελάτες.

Σε ένα πρώτο σενάριο, ας υποθέσουμε ότι έχουμε αφίξεις σε σταθερά χρονικά διαστήματα, κάθε 10 χρονικές μονάδες και ότι οι χρόνοι εξυπηρέτησης είναι επίσης σταθεροί και ίσοι με 9 χρονικές μονάδες. Τότε ο διαχειριστής βλέπει το σύστημα αρκετά απασχολημένο, για την ακρίβεια βλέπει ότι το 90% του χρόνου στο σύστημα υπάρχει 1 πελάτης ενώ μόνο ένα 10% του χρόνου το σύστημα είναι άδειο. Από την άλλη μεριά κάθε πελάτης βλέπει το σύστημα άδειο τη στιγμή που φθάνει σε αυτό (αφού ο προηγούμενος πελάτης έχει φύγει πριν 1 χρονική μονάδα). Μια τυπική πραγμάτωση της διαδικασίας του πλήθους των πελατών, $\{Q(t)\}$, απεικονίζεται στο σχήμα 1.1.

Σε ένα δεύτερο σενάριο, ας υποθέσουμε ότι οι χρόνοι μεταξύ των αφίξεων είναι και πάλι σταθεροί και ίσοι με 10 χρονικές μονάδες, αλλά τώρα θεωρούμε ότι οι χρόνοι εξυπηρέτησης διαρκούν 1 χρονική μονάδα για κάθε πελάτη. Τώρα ο διαχειριστής βλέπει το σύστημα πολύ λίγο απασχολημένο, αφού το 10% μόλις του χρόνου υπάρχει 1 πελάτης ενώ το 90% του χρόνου το σύστημα είναι άδειο. Από την άλλη μεριά η εντύπωση που αποκομίζει κάθε πελάτης φθάνοντας στο σύστημα δεν διαφέρει από την εντύπωση που έχουν οι πελάτες στο πρώτο σενάριο: Και πάλι κάθε πελάτης βλέπει το σύστημα άδειο τη στιγμή που εισέρχεται σε αυτό. Μια τυπική πραγμάτωση της διαδικασίας του πλήθους των πελατών, $\{Q(t)\}$, απεικονίζεται στο σχήμα 1.2.

Το συμπέρασμα είναι ότι οι οπτικές ενός εξωτερικού παρατηρητή και ενός αφικνούμενου πελάτη μπορεί να δίνουν πολύ διαφορετικές εικόνες για το ίδιο σύστημα. Επίσης δυο συστήματα μπορεί να μοιάζουν παρόμοια υπό τη μία οπτική (όπως τα δυο σενάρια υπό την οπτική των πελατών) και να είναι πολύ διαφορετικά υπό την άλλη οπτική (όπως τα δυο σενάρια υπό την οπτική του διαχειριστή). Αν σκεφτούμε ότι οι πελάτες δεν είναι παθητικές οντότητες αλλά μπορεί να αποφασίζουν τι θα κάνουν σε σχέση με το σύστημα (π.χ. να μπουν σε αυτό ή να φύγουν) έχει μεγάλη σημασία να ποσοτικοποιήσουμε με κάποιο τρόπο το πώς αντιλαμβάνονται τον συνωστισμό του συστήματος οι πελάτες κατά την άφιξή τους ή την αναχώρησή τους.

Προς τον σκοπό αυτό, έστω $A_1 < A_2 < A_3 < \dots$ οι διαδοχικές στιγμές αφίξεων και $D_1 < D_2 < D_3 < \dots$ οι διαδοχικές στιγμές αναχωρήσεων των πελατών. Ξεκινώντας από τη στοχαστική διαδικασία $\{Q(t)\}$ που περιγράφει τον αριθμό των πελατών στο σύστημα σε συνεχή χρόνο, ορίζουμε τις εμφυτευμένες διαδικασίες $\{Q_n^-\}$ και $\{Q_n^+\}$ σε στιγμές αφίξεων και αναχωρήσεων πελατών αντίστοιχα. Έτσι ορίζουμε

- $Q_n^- = Q(A_n^-)$ τον αριθμό των πελατών αμέσως πριν την n -οστή άφιξη πελάτη (δηλαδή τον αριθμό των παρόντων πελατών που βλέπει ο n -οστός πελάτης καθώς εισέρχεται στο σύστημα),
- $Q_n^+ = Q(D_n^+)$ τον αριθμό των πελατών αμέσως μετά την n -οστή αναχώρηση πελάτη (δηλαδή τον αριθμό των πελατών που αφήνει πίσω του κατά την έξοδο του ο n -οστός πελάτης που φεύγει από το σύστημα).

Ενδιαφερόμαστε για τις αντίστοιχες κατανομές ισορροπίας ή οριακές κατανομές που περιγράφουν την εντύπωση που διαμορφώνει ένας πελάτης τη στιγμή που εισέρχεται στο σύστημα και τη στιγμή που αναχωρεί από αυτό, εφόσον το σύστημα βρίσκεται σε κατάσταση ισορροπίας. Μιλάμε, αντίστοιχα για τις κατανομές ισορροπίας ή οριακές κατανομές του αριθμού των πελατών σε στιγμές αφίξεων και αναχωρήσεων. Η οριακή πιθανότητα σε στιγμή άφιξης να βρίσκονται j πελάτες στο σύστημα (εξαιρουμένης της νέας άφιξης), a_j , θα έχει και πάλι εναλλακτικές εκφράσεις ως C -οριακή πιθανότητα και ως μακροπρόθεσμο ποσοστό, εφόσον το σύστημα έχει αναγεννητικό χαρακτήρα με απεριοδικούς κύκλους απασχόλησης. Πιο συγκεκριμένα στην περίπτωση αυτή θα έχουμε:

$$\begin{aligned}
 a_j &= \frac{1}{E[A(Z)]} E \left[\sum_{k=1}^{A(Z)} 1_{\{Q_k^- = j\}} \right] \\
 &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n 1_{\{Q_k^- = j\}} \\
 &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \Pr[Q_k^- = j] \\
 &= \lim_{n \rightarrow \infty} \Pr[Q_n^- = j], \quad j \geq 0,
 \end{aligned} \tag{1.19}$$

όπου Z ο πρώτος κύκλος λειτουργίας και $A(Z)$ το πλήθος των αφίξεων σε αυτόν. Ομοίως, η οριακή πιθανότητα σε στιγμή αναχώρησης να βρίσκονται j πελάτες στο σύστημα (εξαιρουμένης της νέας αναχώρησης), d_j , είναι

$$\begin{aligned}
 d_j &= \frac{1}{E[A(Z)]} E \left[\sum_{k=1}^{A(Z)} 1_{\{Q_k^+ = j\}} \right] \\
 &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n 1_{\{Q_k^+ = j\}} \\
 &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \Pr[Q_k^+ = j] \\
 &= \lim_{n \rightarrow \infty} \Pr[Q_n^+ = j], \quad j \geq 0.
 \end{aligned} \tag{1.20}$$

Οι a_j και d_j είναι τα μακροπρόθεσμα ποσοστά των πελατών που βρίσκουν ή αφήνουν πίσω τους j πελάτες στο σύστημα σε στιγμές αφίξεων και αναχωρήσεων αντίστοιχα, ενώ η p_j είναι το μακροπρόθεσμο ποσοστό του χρόνου που στο σύστημα υπάρχουν j πελάτες (δηλ. είναι η οριακή πιθανότητα σε συνεχή χρόνο). Όπως και η p_j , έτσι και οι a_j και d_j ισούνται με τις αντίστοιχες οριακές πιθανότητες ένας πελάτης να δει j άτομα κατά την άφιξή του στο σύστημα ή κατά την αναχώρησή του, εφόσον βέβαια αυτές οι οριακές πιθανότητες υπάρχουν (πράγμα που συμβαίνει όταν η κατανομή του κύκλου απασχόλησης του συστήματος είναι απεριοδική).

Δεν υπάρχει κάποιος λόγος οι κατανομές ισορροπίας (p_j) , (a_j) και (d_j) να συμπίπτουν και γενικά αυτό δεν ισχύει.

Παράδειγμα 1.7 (Συνέχεια του παραδείγματος 1.6) *Ας συγκρίνουμε τα δυο σενάρια για την D/D/1 ουρά που περιγράψαμε στο παράδειγμα 1.6. Και στα δύο είχαμε $a_0 = 1$ και $a_j = 0$, $j \geq 1$. Όμως, στο πρώτο σενάριο είχαμε $p_0 = 0.1$, $p_1 = 0.9$, $p_j = 0$, $j \geq 2$, ενώ στο δεύτερο είχαμε $p_0 = 0.9$, $p_1 = 0.1$, $p_j = 0$, $j \geq 2$. Δηλαδή, και στα δυο σενάρια είναι $(p_j : j \geq 0) \neq (a_j : j \geq 0)$. Και μάλιστα, στην πραγματικότητα δεν μπορεί να βγει κανένα ουσιώδες συμπέρασμα από το a_j για το αντίστοιχο p_j (π.χ. $a_0 = 1$ και στα δυο σενάρια, ενώ $p_0 = 0.1$ στο πρώτο σενάριο και $p_0 = 0.9$ στο δεύτερο). Βέβαια στα δυο σενάρια ισχύει $(a_j : j \geq 0) = (d_j : j \geq 0)$, αφού όλοι οι πελάτες βρίσκουν το σύστημα κενό και όλοι το αφήνουν κενό.* ◊

Παράδειγμα 1.8 (Κατανομές ισορροπίας σε στάση μεταφορικού μέσου) *Θεωρούμε την περίπτωση που επιβάτες φθάνουν σε μια στάση ενός μεταφορικού μέσου αεροδρομίου σύμφωνα με μια ανανεωτική διαδικασία και εκεί υπάρχει ένα βανάκι που αναχωρεί πάντα γεμάτο, όταν συγκεντρωθούν N επιβάτες. Αν εστιάσουμε στη στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των επιβατών που περιμένουν στη στάση, τότε βλέπουμε ότι η $\{Q(t)\}$ σε κάθε κύκλο λειτουργίας του συστήματος περνάει διαδοχικά από τις καταστάσεις $1, 2, 3, \dots, N-1$ και 0 και σε κάθε κατάσταση παραμένει για έναν ενδιάμεσο χρόνο αφίξεων (παρατηρήστε ότι μόλις συγκεντρωθούν N πελάτες το βανάκι τους απομακρύνει ακαριαία, οπότε δεν παρατηρούμε την κατάσταση N και η στάση αδειάζει). Επομένως, αν ο μέσος ενδιάμεσος χρόνος αφίξεων είναι a , τότε σε έναν κύκλο λειτουργίας η $\{Q(t)\}$ μένει στην κατάσταση $j = 1, 2, \dots, N-1$ και 0 κατά μέσο όρο για a χρονικές μονάδες, ενώ η μέση διάρκεια του κύκλου λειτουργίας είναι Na , όσο απαιτείται για να συγκεντρωθούν N επιβάτες. Οπότε, η κατανομή ισορροπίας της $\{Q(t)\}$ είναι*

$$p_j = \frac{a}{Na} = \frac{1}{N}, \quad 0 \leq j \leq N-1,$$

δηλαδή είναι διακριτή ομοιόμορφη στο σύνολο $\{0, 1, 2, \dots, N-1\}$. Στο σύστημα αυτό οι διαδοχικά αφικνούμενοι πελάτες βλέπουν κατά την άφιξή τους $0, 1, 2, 3, \dots, N-1, 0, 1, 2, \dots, N-1, 0, 1, \dots$ πελάτες κ.ο.κ., οπότε σε κάθε κύκλο λειτουργίας του συστήματος υπάρχει ακριβώς ένας πελάτης που βλέπει $0, 1, \dots, N-1$ πελάτες κατά την άφιξή του. Επομένως τα μακροπρόθεσμα ποσοστά των πελατών που βλέπουν κατά την άφιξή τους j πελάτες, για $j = 0, 1, 2, \dots, N-1$, είναι όλα ίσα και έχουμε

$$a_j = \frac{1}{N}, \quad 0 \leq j \leq N-1.$$

Εδώ, βλέπουμε ότι $(a_j : j \geq 0) = (p_j : j \geq 0)$. Όμως, επειδή οι πελάτες αναχωρούν όλοι μαζί όταν τους παίρνει το βανάκι, μετά από κάθε αναχώρηση πελάτη το σύστημα είναι άδειο, δηλαδή

$$d_0 = 1, \quad \text{και} \quad d_j = 0, \quad j \geq 1.$$

Οπότε, τελικά $(d_j : j \geq 0) \neq (a_j : j \geq 0) = (p_j : j \geq 0)$. ◊

1.4 Ασκήσεις

Άσκηση 1.1 *Πελάτες φθάνουν σε ένα σύστημα εξυπηρέτησης, σύμφωνα με διαδικασία Poisson με ρυθμό λ . Οι χρόνοι εξυπηρέτησης των πελατών είναι Erlang(3, μ) (αποτελούμενοι από 3 εκθετικές φάσεις, κάθε μία εκθετικής*

διάρκειας μ) και γνωστοποιούνται στο σύστημα μόλις φθάνουν οι πελάτες. Στο σύστημα υπάρχουν 4 υπηρέτες. Ο χώρος αναμονής του συστήματος είναι κοινός για όλους τους υπηρέτες και έχει 6 θέσεις, πλέον των θέσεων εξυπηρέτησης (μία μπροστά από κάθε υπηρέτη). Κάθε φορά που ένας αφικνούμενος πελάτης βρίσκει τουλάχιστον έναν ελεύθερο υπηρέτη, διαλέγει έναν στην τύχη και αρχίζει να εξυπηρετείται. Αν ένας αφικνούμενος πελάτης βρει όλους τους υπηρέτες απασχολημένους τότε παραμένει σε μια από τις 6 θέσεις στον χώρο αναμονής του συστήματος. Αν είναι, όμως, κατελήμμενες, τότε αναχωρεί άμεσα από το σύστημα. Κάθε φορά που αναχωρεί πελάτης και υπάρχουν πελάτες στον χώρο αναμονής, ο υπηρέτης που ελευθερώθηκε επιλέγει για εξυπηρέτηση τον πελάτη που έχει δηλώσει τον μικρότερο χρόνο εξυπηρέτησης.

1. Πώς αναφέρεται το συγκεκριμένο σύστημα με βάση την ονοματολογία του Kendall;
2. Ποιος είναι ο μέσος ενδιάμεσος χρόνος αφίξεων και ποιος ο μέσος χρόνος εξυπηρέτησης στο συγκεκριμένο σύστημα;

Άσκηση 1.2 Περιγράψτε τα βασικά στοιχεία (διαδικασία αφίξεων, χρόνους εξυπηρέτησης, αριθμό υπηρετών, χωρητικότητα και πειθαρχία ουράς) για τα συστήματα $GI/M/2/S$ και $M/D/1(LCFS)$

Άσκηση 1.3 Ένα σύστημα εξυπηρέτησης με έναν υπηρέτη λειτουργεί με κύκλους λειτουργίας ως εξής: Στην αρχή κάθε κύκλου λειτουργίας φθάνει μια παρτίδα με προϊόντα προς επεξεργασία. Τα μεγέθη των διαδοχικά αφικνούμενων παρτίδων είναι ανεξάρτητες και ισόνομες θετικές ακέραιες τυχαίες μεταβλητές με συνάρτηση πιθανότητας ($g_j : j \geq 1$). Αμέσως μετά την άφιξη μιας παρτίδας, ο υπηρέτης αρχίζει την επεξεργασία των προϊόντων, ένα-ένα. Ο μέσος χρόνος για την επεξεργασία ενός προϊόντος είναι a χρονικές μονάδες. Μόλις τελειώσει η επεξεργασία ενός προϊόντος αυτό αναχωρεί από το σύστημα. Όταν τελειώσει η επεξεργασία όλων των προϊόντων μιας παρτίδας ο υπηρέτης αργεί για b χρονικές μονάδες. Κατόπιν αρχίζει ένας νέος κύκλος λειτουργίας με την άφιξη μιας νέας παρτίδας προϊόντων κ.ο.κ.

1. Να βρεθεί το μακροπρόθεσμο ποσοστό του χρόνου, p_0 , που ο υπηρέτης δεν παρέχει εξυπηρέτηση.
2. Να βρεθεί το μακροπρόθεσμο ποσοστό του χρόνου, p_j , που υπάρχουν j προϊόντα στο σύστημα, για $j \geq 1$.
3. Να βρεθεί το μακροπρόθεσμο ποσοστό προϊόντων, d_j , που με την αναχώρησή τους αφήνουν j προϊόντα στο σύστημα, για $j \geq 0$.

Άσκηση 1.4 Μια $G/M/1$ ουρά έχει εκθετική κατανομή χρόνων εξυπηρέτησης με παράμετρο μ και κατανομή ισορροπίας σε στιγμές αφίξεων πελατών που δίνεται από τον τύπο $a_j = (1 - \alpha)\alpha^j$, για $j \geq 0$, όπου α γνωστή παράμετρος.

1. Να βρεθεί ο μέσος αριθμός πελατών που βρίσκει ένας αφικνούμενος πελάτης στο σύστημα.
2. Να βρεθεί ο μέσος χρόνος παραμονής ενός πελάτη που βρίσκει j πελάτες στο σύστημα, κατά την άφιξή του.
3. Να βρεθεί ο μέσος χρόνος παραμονής ενός πελάτη στο σύστημα.

Άσκηση 1.5 Μια $M/G/1/1$ ουρά έχει ρυθμό αφίξεων λ και μέσο χρόνο εξυπηρέτησης b . Θέτουμε $\rho = \lambda b$. Να υπολογιστεί η κατανομή ισορροπίας ($p_j : j = 0, 1$) σε συνεχή χρόνο, συναρτήσει του ρ .

1.5 Σχόλια

Η αναγεννητικότητα των υποκείμενων στοχαστικών διαδικασιών είναι μια βασική υπόθεση που διατρέχει το σύνολο σχεδόν της Θεωρίας Ουρών. Τα εισαγωγικά βιβλία Wolff 1989, Φακίνος 2003, Tijms 2008 και Kulkarni 2010 δίνουν έμφαση σε αυτή την οπτική. Επίσης, η οπτική αυτή συνδέεται με τη μελέτη των συστημάτων εξυπηρέτησης εστιάζοντας σε μια πραγματοποίηση τους (sample-path analysis). Το βασικό σύγγραμμα που αναπτύσσει αυτή τη θεώρηση είναι το El-Taha και Stidham 1998.

Άλλα συγγράμματα ακολουθούν μια πιο αναλυτική προσέγγιση. Εκεί η έμφαση δίνεται στη μοντελοποίηση των συστημάτων εξυπηρέτησης μέσω κάποιων στοχαστικών διαδικασιών με ανεπτυγμένη θεωρία (π.χ. ως Μαρκοβιανές διαδικασίες συνεχούς χρόνου) και κατόπιν χρησιμοποιούνται κυρίως αναλυτικές τεχνικές. Ο ενδιαφερόμενος αναγνώστης μπορεί να δει π.χ. τα Kleinrock 1975, Kleinrock 1976 και Cohen 1969.

Οι δυο θεωρήσεις, η κλασική-αναλυτική και η σύγχρονη-πιθανοθεωρητική με έμφαση στην αναγεννητικότητα, είναι συμπληρωματικές και είναι σημαντικό ο ενδιαφερόμενος αναγνώστης να έρθει σε επαφή και με τις δύο.

Άλλα εισαγωγικά συγγράμματα στη Θεωρία Ουρών Αναμονής που κινούνται σε μαθηματικό επίπεδο παρόμοιο με του παρόντος βιβλίου είναι τα Shortle, Thompson, Gross και Harris 2018, Haviv 2013 και Harchol-Balter 2013, καθώς και το κλασικό σύγγραμμα του Asmussen 1987 που κινείται σε πιο προχωρημένο επίπεδο.

Το θέμα της σχέσης των κατανομών ισορροπίας (p_j), (a_j) και (d_j), και γενικότερα των κατανομών ισορροπίας της κατάστασης μιας στοχαστικής διαδικασίας σε συνεχή χρόνο και σε στιγμές γεγονότων είναι ιδιαίτερα σημαντικό, αλλά η αυστηρή του αντιμετώπιση απαιτεί ισχυρή θεωρητική κατάρτιση. Ο ενδιαφερόμενος αναγνώστης παραπέμπεται στα προχωρημένα συγγράμματα των Baccelli και Bremaud 1979 και Sigman 1995 που εξετάζουν σε βάθος αυτό το ζήτημα.

Βιβλιογραφία

- [1] D.G. Kendall. “Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain”. Στο: *The Annals of Mathematical Statistics* 24 (1953), σσ. 338–354.
- [2] R.W. Wolff. *Stochastic Modeling and the Theory of Queues*. Englewood Cliffs, New Jersey: Prentice Hall, 1989. ISBN: 978-0138466923.
- [3] Δ. Φακίνος. *Ουρές Αναμονής: Θεωρία και Ασκήσεις*. Αθήνα: Εκδόσεις Συμμετρία, 2003. ISBN: 978-960-266-206-9.
- [4] H.C. Tijms. *A First Course in Stochastic Models, 2nd Edition*. Chichester: Wiley, 2008. ISBN: 978-0471498803.
- [5] V.G. Kulkarni. *Modeling and Analysis of Stochastic Systems, 2nd Edition*. CRC Press, 2010. ISBN: 978-1439808757.
- [6] M. El-Taha και S.Jr. Stidham. *Sample-path Analysis of Queueing Systems*. Kluwer, 1998. ISBN: 978-0792382102.
- [7] L. Kleinrock. *Queueing Systems. Volume 1: Theory*. Wiley, 1975. ISBN: 978-0471491101.
- [8] L. Kleinrock. *Queueing Systems. Volume 2: Computer Applications*. Wiley, 1976. ISBN: 978-0471491118.
- [9] J.W. Cohen. *The Single Server Queue*. Amsterdam: North-Holland, 1969. ISBN: 978-0720423587.
- [10] J.F. Shortle, J.M. Thompson, D. Gross και C.M. Harris. *Fundamentals of Queueing Theory, 5th Edition*. Wiley, 2018. ISBN: 978-1118943526.
- [11] M. Haviv. *Queues: A Course in Queueing Theory*. Springer, 2013. ISBN: 978-1461467649.

- [12] M. Harchol-Balter. *Performance Modeling and Design of Computer Systems: Queueing Theory in Action*. Cambridge, 2013. ISBN: 978-1107027503.
- [13] S. Asmussen. *Applied Probability and Queues*. Wiley, 1987. ISBN: 978-0471911739.
- [14] F. Baccelli και P. Bremaud. *Elements of Queueing Theory. Palm Martingale Calculus and Stochastic Recurrences, 2nd Edition*. New York: McGraw-Hill, Inc., 1979.
- [15] K. Sigman. *Stationary Marked Point Processes: An Intuitive Approach*. Chapman και Hall/CRC, 1995. ISBN: 978-0412984310.

ΚΕΦΑΛΑΙΟ 2

ΒΑΣΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Στο κεφάλαιο αυτό διατυπώνουμε τέσσερα βασικά αποτελέσματα που αφορούν τα συστήματα εξυπηρέτησης, τα οποία ισχύουν κάτω από πολύ γενικές συνθήκες. Συγκεκριμένα χαρακτηρίζουμε την ευστάθεια μέσω του ρυθμού συνωστισμού ενός συστήματος και διατυπώνουμε την ιδιότητα των μεμονωμένων μεταβάσεων (αφίξεων-αναχωρήσεων), την ιδιότητα PASTA (Poisson-Arrivals-See-Time-Averages) και τον νόμο του Little.

2.1 Ρυθμός συνωστισμού - Ευστάθεια

Το μέσο ποσό εργασίας που εισέρχεται προς διεκπεραίωση σε ένα σύστημα εξυπηρέτησης ανά χρονική μονάδα αναφέρεται ως ο ρυθμός συνωστισμού ρ και ισούται με το γινόμενο του ρυθμού αφίξεων λ επί τον μέσο χρόνο εξυπηρέτησης b :

$$\rho = \lambda b.$$

Το μέγιστο ποσό της εργασίας που μπορεί να διεκπεραιώσει το σύστημα ανά χρονική μονάδα είναι ίσο με το πλήθος των υπηρετών c αφού κάθε υπηρέτης μπορεί να διεκπεραιώσει μια μονάδα εργασίας ανά χρονική μονάδα. Έτσι για να είναι το σύστημα ευσταθές και να μην απειρίζεται η ουρά αναμένουμε διαισθητικά ότι θα πρέπει το μέσο ποσό εργασίας που εισέρχεται ανά χρονική μονάδα να είναι μικρότερο από τη μέγιστη δυνατότητα διεκπεραίωσης του συστήματος ανά χρονική μονάδα.

Θεώρημα 2.1 (Χαρακτηρισμός ευστάθειας) Στο $G/G/c$ σύστημα με απεριοδική (π.χ. συνεχή ή μεικτή) κατανομή για τους ενδιάμεσους χρόνους μεταξύ των αφίξεων και/ή τους χρόνους εξυπηρέτησης ισχύει ένα ακριβώς από τα παρακάτω:

- $\rho < c$ οπότε το σύστημα είναι ευσταθές, δηλαδή υπάρχουν οι κατανομές ισορροπίας (p_j) , (a_j) και (d_j) και είναι $p_j > 0$, $j \geq 0$, και $\sum_{j=0}^{\infty} p_j = 1$ (και όμοια για τις (a_j) και (d_j)).
- $\rho \geq c$ οπότε το σύστημα είναι ασταθές, δηλαδή το πλήθος των πελατών απειρίζεται καθώς $t \rightarrow \infty$ και $p_j = a_j = d_j = 0$, $j \geq 0$.

Η απόδειξη του αποτελέσματος αυτού είναι ιδιαίτερα περίπλοκη. Ο ενδιαφερόμενος αναγνώστης μπορεί να συμβουλευτεί το βιβλίο των Baccelli και Bremaud 1979 όπου αποδεικνύονται θεωρήματα ευστάθειας και άλλα θεωρητικά αποτελέσματα κάτω από γενικές συνθήκες. Βέβαια, το σύγγραμμα αυτό είναι μεταπτυχιακού χαρακτήρα και απαιτεί προχωρημένη γνώση της θεωρητικής θεμελίωσης της Θεωρίας Πιθανοτήτων.

Ένα σημείο του θεωρήματος που μοιάζει κάπως ασύμφωνο με τη διαίσθησή μας είναι το γεγονός ότι για $\rho = c$ το σύστημα είναι ασταθές. Αυτό συμβαίνει λόγω της τυχαιότητας που ενυπάρχει στο σύστημα. Δεδομένου του ότι δεν είναι όλα τα στοιχεία του συστήματος ντετερμινιστικά, όταν ο υπηρέτης παραμένει ανενεργός λόγω κάποιας τυχαίας μείωσης των αφίξεων για κάποιο διάστημα, χάνεται δυναμικότητα εξυπηρέτησης που δεν μπορεί να αναπληρωθεί αργότερα αφού το ρ είναι ακριβώς ίσο με c . Ένα ντετερμινιστικό σύστημα τύπου $D/D/c$ με $\rho = c$ θα είναι ευσταθές. Π.χ. ένα $D/D/1$ με $\rho = 1$ θα έχει πάντα σταθερό αριθμό πελατών (δείτε το παράδειγμα 1.6).

Επίσης, αξίζει να διευκρινιστεί ότι η (τεχνική) υπόθεση του θεωρήματος που αφορά την απεριοδικότητα των χρόνων δεν είναι περιοριστική, αφού ισχύει σε όλα τα συστήματα που εμφανίζονται στην πράξη λόγω του ότι υποτίθενται συνεχείς κατανομές σε αυτά. Η υπόθεση αυτή, όμως, είναι αναγκαία για να έχουμε την ερμηνεία των μακροπρόθεσμων ποσοστών του χρόνου ή των πελατών ως οριακές πιθανότητες, σύμφωνα με αυτά που είπαμε για την αναγεννητικότητα των συστημάτων εξυπηρέτησης και τις εναλλακτικές ερμηνείες των μέτρων απόδοσής τους.

2.2 Ιδιότητα μεμονωμένων μεταβάσεων και ιδιότητα PASTA

Όπως είπαμε οι κατανομές ισορροπίας (p_j), (a_j) και (d_j) γενικά δεν συμπίπτουν. Υπάρχουν όμως δυο περιπτώσεις στις οποίες κάποιες από αυτές συμπίπτουν.

Θεώρημα 2.2 (Ιδιότητα μεμονωμένων μεταβάσεων) Σε συστήματα εξυπηρέτησης στα οποία οι πελάτες έρχονται και αναχωρούν μεμονωμένα, δηλαδή δεν υπάρχουν ομαδικές αφίξεις, ούτε ομαδικές αναχωρήσεις, οι κατανομές ισορροπίας σε στιγμές αφίξεων και σε στιγμές αναχωρήσεων συμπίπτουν: $(a_j) = (d_j)$.

Η ιδιότητα των μεμονωμένων μεταβάσεων μπορεί να αιτιολογηθεί ως εξής: Έστω $A_j(t)$ το πλήθος των αφίξεων στο $(0, t]$ που βρίσκουν j πελάτες και $A(t)$ το συνολικό πλήθος αφίξεων στο $(0, t]$. Τότε, έχουμε $a_j = \lim_{t \rightarrow \infty} \frac{A_j(t)}{A(t)}$. Ομοίως, $d_j = \lim_{t \rightarrow \infty} \frac{D_j(t)}{D(t)}$, όπου $D_j(t)$ το πλήθος των αναχωρήσεων στο $(0, t]$ που αφήνουν πίσω τους j πελάτες και $D(t)$ το συνολικό πλήθος αναχωρήσεων στο $(0, t]$. Οι μακροπρόθεσμοι ρυθμοί αφίξεων και αναχωρήσεων είναι $\lim_{t \rightarrow \infty} \frac{A(t)}{t}$ και $\lim_{t \rightarrow \infty} \frac{D(t)}{t}$.

Έχουμε ότι

$$\lim_{t \rightarrow \infty} \frac{A(t)}{t} = \lim_{t \rightarrow \infty} \frac{D(t)}{t},$$

αφού οι μακροπρόθεσμοι ρυθμοί αφίξεων και αναχωρήσεων θα είναι ίσοι υπό την υπόθεση της ευστάθειας (ο ρυθμός των αφίξεων είναι προφανώς μεγαλύτερος ή ίσος του ρυθμού των αναχωρήσεων, αλλά δεν μπορεί να είναι μεγαλύτερος αν το σύστημα είναι ευσταθές). Επίσης έχουμε ότι $|A_j(t) - D_j(t)| \leq 1$, αφού μεταξύ δυο διαδοχικών αφίξεων που βρίσκουν j πελάτες πρέπει να υπάρχει μια ακριβώς αναχώρηση που αφήνει j πελάτες στο σύστημα και μεταξύ δυο διαδοχικών αναχωρήσεων που αφήνουν j πελάτες πρέπει να υπάρχει μια ακριβώς άφιξη που βρίσκει j πελάτες. Αυτό συμβαίνει διότι κάθε διάσχιση του ορίου των j πελατών από τη διαδικασία του αριθμού των πελατών $\{Q(t)\}$ από κάτω προς τα πάνω (δηλαδή από καταστάσεις μικρότερες ή ίσες του j σε καταστάσεις μεγαλύτερες του j) ακολουθείται από μια διάσχιση του ορίου από πάνω προς τα κάτω, πριν την επόμενη διάσχιση από κάτω προς τα πάνω. Με άλλα λόγια οι διασχίσεις του ορίου των j πελατών συμβαίνουν εναλλάξ από κάτω προς τα πάνω και από πάνω προς τα κάτω. Οπότε έχουμε ότι $\lim_{t \rightarrow \infty} \frac{A_j(t) - D_j(t)}{t} = 0$ και άρα

$$\lim_{t \rightarrow \infty} \frac{A_j(t)}{t} = \lim_{t \rightarrow \infty} \frac{D_j(t)}{t}.$$

Επομένως, είναι

$$a_j = \lim_{t \rightarrow \infty} \frac{A_j(t)}{A(t)} = \lim_{t \rightarrow \infty} \frac{A_j(t)/t}{A(t)/t} = \lim_{t \rightarrow \infty} \frac{D_j(t)/t}{D(t)/t} = \lim_{t \rightarrow \infty} \frac{D_j(t)}{D(t)} = d_j.$$

Θεώρημα 2.3 (Ιδιότητα PASTA (Poisson-Arrivals-See-Time-Averages)) Σε συστήματα εξυπηρέτησης στα οποία οι πελάτες έρχονται σύμφωνα με μια διαδικασία Poisson (δηλαδή οι ενδιάμεσοι χρόνοι μεταξύ των αφίξεων είναι ανεξάρτητοι και ισόνομοι με εκθετική κατανομή), οι κατανομές ισορροπίας σε στιγμές αφίξεων και σε συνεχή χρόνο συμπίπτουν: $(a_j) = (p_j)$. ◊

Η διαισθητική αιτιολόγηση της ιδιότητας PASTA είναι ότι η διαδικασία Poisson μοντελοποιεί την ιδέα των εντελώς τυχαίων αφίξεων στον χρόνο και επομένως η παρατήρηση του αριθμού των πελατών κατά τη στιγμή της άφιξης ενός πελάτη στο σύστημα ισοδυναμεί με την παρατήρηση του αριθμού των πελατών σε μια τυχαία χρονική στιγμή (σε συνεχή χρόνο). Η αιτιολόγηση αυτή μπορεί να γίνει περισσότερο κατανοητή αν σκεφθούμε ως εξής: Έστω $A(t, t+h)$ το πλήθος των αφίξεων σε ένα σύστημα εξυπηρέτησης στο διάστημα $(t, t+h]$. Τότε, για το a_j , δηλαδή την οριακή πιθανότητα σε στιγμή άφιξης να παρατηρηθούν j πελάτες, έχουμε

$$\begin{aligned} a_j &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \Pr[Q(t) = j | A(t, t+h) > 0] \\ &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \frac{\Pr[Q(t) = j, A(t, t+h) > 0]}{\Pr[A(t, t+h) > 0]} \\ &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \frac{\Pr[Q(t) = j] \Pr[A(t, t+h) > 0 | Q(t) = j]}{\Pr[A(t, t+h) > 0]}. \end{aligned}$$

Όμως το ενδεχόμενο $\{A(t, t+h) > 0\}$ είναι ανεξάρτητο του $\{Q(t) = j\}$ λόγω του ότι η διαδικασία αφίξεων είναι Poisson, της οποίας η (μελλοντική) εξέλιξη μετά τη χρονική στιγμή t δεν εξαρτάται από την κατάσταση του συστήματος τη χρονική στιγμή t . Επομένως $\Pr[A(t, t+h) > 0 | Q(t) = j] = \Pr[A(t, t+h) > 0]$ και άρα

$$a_j = \lim_{t \rightarrow \infty} \Pr[Q(t) = j] = p_j.$$

Η αυστηρή απόδειξη της ιδιότητας PASTA έγινε από τον Wolff 1982. Φυσικά μπορούμε να συνδυάσουμε τα δυο αποτελέσματα των μεμονωμένων μεταβάσεων και της PASTA.

Πόρισμα 2.4 (Ιδιότητα μεμονωμένων μεταβάσεων + PASTA) Σε συστήματα εξυπηρέτησης με μεμονωμένες μεταβάσεις (αφίξεις, αναχωρήσεις) και Poisson διαδικασία αφίξεων όλες οι κατανομές ισορροπίας του πλήθους των πελατών συμπίπτουν: $(p_j) = (a_j) = (d_j)$. ◊

Σε κάποιες εφαρμογές είναι χρήσιμη και η παρακάτω γενίκευση της ιδιότητας PASTA.

Θεώρημα 2.5 (Γενικευμένη ιδιότητα PASTA) Έστω $\{X(t) : t \geq 0\}$ μια στοχαστική διαδικασία και $\{N(t) : t \geq 0\}$ μια διαδικασία Poisson με χρόνους γεγονότων $S_n, n \geq 0$. Αν ισχύει ότι για κάθε $t \geq 0$, η $\{X(u) : 0 \leq u \leq t\}$ και η $\{N(t+u) - N(t) : u > 0\}$ είναι ανεξάρτητες (υπόθεση έλλειψης προσοκίας για την $\{N(t)\}$ - lack of anticipation assumption for $\{N(t)\}$), τότε οι κατανομές ισορροπίας της $\{X(t)\}$ όταν αυτή παρατηρείται σε συνεχή χρόνο ή αμέσως πριν από στιγμές γεγονότων της $\{N(t)\}$ συμπίπτουν, δηλαδή για κάθε ενδεχόμενο B έχουμε

$$\lim_{t \rightarrow \infty} \frac{\int_0^t \mathbf{1}_{\{X(u) \in B\}} du}{t} = \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \mathbf{1}_{\{X(S_k^-) \in B\}}}{n}, \quad \text{με πιθανότητα 1.} \quad (2.1)$$

◊

Διαισθητικά, η γενικευμένη ιδιότητα PASTA αναφέρεται στην παρατήρηση μιας στοχαστικής διαδικασίας $\{X(t)\}$ που καταγράφει την εξέλιξη της κατάστασης ενός συστήματος, κατά τις στιγμές των γεγονότων μιας διαδικασίας Poisson $\{N(t)\}$. Η σχέση-αλληλεπίδραση της $\{X(t)\}$ και της $\{N(t)\}$ δεν εξειδικεύεται, αλλά απαιτείται το μέλλον της διαδικασίας των στιγμών παρατήρησης $\{N(t)\}$ να είναι ανεξάρτητο από την παρελθούσα ιστορία και την παρούσα κατάσταση της $\{X(t)\}$. Με άλλα λόγια για να ισχύει το αποτέλεσμα πρέπει οι στιγμές των μελλοντικών παρατηρήσεων του συστήματος να γίνονται εντελώς τυχαία (αυτή την έννοια έχει το ότι η $\{N(t)\}$ είναι Poisson) και να μην καθοδηγούνται από τη μέχρι τώρα ιστορία του (αυτή την έννοια έχει η υπόθεση έλλειψης προσδοκίας). Κάτω από αυτές τις υποθέσεις, έχουμε ότι αυτό που βλέπει ο παρατηρητής που παρατηρεί το σύστημα σε συνεχή χρόνο ταυτίζεται με αυτό που βλέπει ο παρατηρητής που το παρατηρεί αμέσως πριν από τις στιγμές της $\{N(t)\}$: Το μακροπρόθεσμο ποσοστό του χρόνου που η διαδικασία $\{X(t)\}$ περνάει σε ένα σύνολο B ισούται με το μακροπρόθεσμο ποσοστό των στιγμών της $\{N(t)\}$ που η $\{X(t)\}$ βρίσκεται σε καταστάσεις του B αμέσως πριν. Το αποτέλεσμα επικεντρώνεται στην κατάσταση του συστήματος αμέσως πριν τις στιγμές των παρατηρήσεων, ώστε να μην συμπεριλαμβάνονται στις καταστάσεις που παρατηρούνται οι επιδράσεις των παρατηρήσεων. Αυτό είναι σημαντικό διότι οι στιγμές παρατηρήσεων που παράγει η διαδικασία $\{N(t)\}$ συνδέονται στις περισσότερες εφαρμογές με αλλαγές στην κατάσταση της $\{X(t)\}$ (π.χ. οι στιγμές παρατηρήσεων μπορεί να είναι οι αφίξεις σε ένα σύστημα εξυπηρέτησης και η $\{X(t)\}$ η διαδικασία που καταγράφει τον αριθμό των πελατών).

Η συνήθης ιδιότητα PASTA προκύπτει από την γενικευμένη όταν η $\{X(t)\}$ είναι η διαδικασία που καταγράφει το πλήθος των πελατών σε ένα σύστημα εξυπηρέτησης, η $\{N(t)\}$ είναι η Poisson διαδικασία των αφίξεων του συστήματος και ως σύνολα B θεωρούνται τα μονοσύνολα $\{j\}$ που αντιστοιχούν στην παρουσία j πελατών στο σύστημα.

2.3 Νόμος του Little

Ο νόμος του Little είναι ένα πολύ γενικό αποτέλεσμα που συνδέει το μέσο πλήθος πελατών στο σύστημα, $E[Q]$, τον ρυθμό αφίξεων λ και τον μέσο χρόνο παραμονής ενός πελάτη $E[S]$ σε αυτό. Το αποτέλεσμα αποδείχθηκε αρχικά στην εργασία Little 1961. Κατόπιν δημοσιεύτηκαν αρκετές εργασίες που το γενικεύουν και/ή προσφέρουν εναλλακτικές αποδείξεις, π.χ. οι εργασίες Stidham 1972 και Stidham 1974. Το μόνο που προϋποθέτει είναι η αναγεννητικότητα του αριθμού των πελατών στο σύστημα. Συγκεκριμένα, έχουμε

Θεώρημα 2.6 (Νόμος του Little) Έστω ένα σύστημα εξυπηρέτησης με αναγεννητική διαδικασία αριθμού πελατών στο σύστημα $\{Q(t) : t \geq 0\}$, με σημεία αναγέννησης τις στιγμές που φθάνουν πελάτες σε κενό σύστημα. Έστω, επίσης, $\{A(t) : t \geq 0\}$ η διαδικασία των αφίξεων πελατών του συστήματος και $\{S_n : n \geq 1\}$ η ακολουθία των χρόνων παραμονής των πελατών. Θέτουμε

$$E[Q] = \lim_{t \rightarrow \infty} \frac{\int_0^t Q(u) du}{t}, \quad \lambda = \lim_{t \rightarrow \infty} \frac{A(t)}{t}, \quad E[S] = \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n S_k}{n} \quad (2.2)$$

το μέσο πλήθος πελατών, τον ρυθμό αφίξεων και τον μέσο χρόνο παραμονής πελάτη στο σύστημα, αντίστοιχα. Τότε:

$$E[Q] = \lambda E[S]. \quad (2.3)$$

◊

Διαισθητικά, το αποτέλεσμα του Little μπορεί να γίνει κατανοητό θεωρώντας ότι κάθε πελάτης πληρώνει 1 χρηματική μονάδα ανά χρονική μονάδα παραμονής του στο σύστημα. Τότε ο διαχειριστής του συστήματος λαμβάνει $E[Q]$ χρηματικές μονάδες στη μονάδα του χρόνου, αν υποθέσουμε ότι η πληρωμή γίνεται κατά τρόπο «συνεχή». Από την άλλη μεριά η μέση εισπραξη του διαχειριστή στη μονάδα του χρόνου θα πρέπει να είναι η ίδια αν οι πελάτες πληρώνουν «προκαταβολικά», δηλαδή αν με την είσοδό τους στο σύστημα δίνουν όλο το ποσό για την παραμονή τους. Αλλά τότε θα έχουμε κατά μέσο όρο λ πελάτες ανά χρονική μονάδα

και ο καθένας θα πληρώνει $E[S]$ χρηματικές μονάδες, οπότε η συνολική είσπραξη του διαχειριστή θα είναι $\lambda E[S]$ χρηματικές μονάδες στη μονάδα του χρόνου. Αφού τα δυο ποσά πρέπει να είναι ίσα (μακροπρόθεσμο μέσο έσοδο με συνεχή είσπραξη = μακροπρόθεσμο μέσο έσοδο με προκαταβολική είσπραξη) έχουμε τη σχέση $E[Q] = \lambda E[S]$.

Για μια μαθηματική διατύπωση αυτής της ιδέας σκεφτόμαστε ως εξής: Λόγω της αναγεννητικότητας του συστήματος, τα μέτρα απόδοσης που εμφανίζονται στον νόμο του Little είναι καλά ορισμένα και μπορούν να υπολογιστούν αν περιοριστούμε στον πρώτο αναγεννητικό κύκλο Z , δηλ. ξεκινώντας από τη στιγμή που ένας πελάτης φθάνει σε κενό σύστημα ως την επόμενη στιγμή που ένας άλλος πελάτης θα αφιχθεί σε κενό σύστημα. Έχουμε, δηλαδή, από το εργοδικό θεώρημα των αναγεννητικών διαδικασιών ότι:

$$E[Q] = \frac{E\left[\int_0^Z Q(u)du\right]}{E[Z]}, \quad \lambda = \frac{E[A(Z)]}{E[Z]}, \quad E[S] = \frac{E\left[\sum_{k=1}^{A(Z)} S_k\right]}{E[A(Z)]}. \quad (2.4)$$

Έστω $I_k(u)$ η δείκτρια τυχαία μεταβλητή που αντιστοιχεί στον k -οστό πελάτη και παίρνει την τιμή 1 αν αυτός είναι παρών στο σύστημα τη χρονική στιγμή u και 0 διαφορετικά. Τότε είναι φανερό ότι για κάθε $k = 1, 2, \dots, A(Z)$, ο χρόνος παραμονής του k -οστού πελάτη γράφεται ως

$$S_k = \int_0^Z I_k(u)du.$$

Επίσης, για κάθε $u \in (0, Z)$, έχουμε

$$Q(u) = \sum_{k=1}^{A(Z)} I_k(u).$$

Αλλά τότε έχουμε:

$$\int_0^Z Q(u)du = \int_0^Z \sum_{k=1}^{A(Z)} I_k(u)du = \sum_{k=1}^{A(Z)} \int_0^Z I_k(u)du = \sum_{k=1}^{A(Z)} S_k. \quad (2.5)$$

Ουσιαστικά η σχέση αυτή γίνεται φανερή, αν σκεφτεί κανείς ότι κάθε χρονική μονάδα παραμονής κάθε πελάτη που αφίχθη στο διάστημα $(0, Z)$ συνεισφέρει μια μονάδα στο αριστερό μέλος, επομένως το ολοκλήρωμα του αριστερού μέλους ισούται με το συνολικό χρόνο παραμονής όλων των πελατών που αφίχθηκαν στο διάστημα $(0, Z)$, που είναι ακριβώς το δεξιό μέλος. Συνδυάζοντας τα παραπάνω έχουμε:

$$\begin{aligned} E[Q] &= \frac{E\left[\int_0^Z Q(u)du\right]}{E[Z]} = \frac{E[A(Z)]}{E[Z]} \cdot \frac{E\left[\int_0^Z Q(u)du\right]}{E[A(Z)]} \\ &= \frac{E[A(Z)]}{E[Z]} \cdot \frac{E\left[\sum_{k=1}^{A(Z)} S_k\right]}{E[A(Z)]} = \lambda E[S]. \end{aligned}$$

Το αποτέλεσμα του Little μπορεί να εφαρμοστεί και σε υποσυστήματα ενός συστήματος, δίνοντας ενδιαφέροντα αποτελέσματα. Στην περίπτωση αυτή το $E[Q]$ θα αναφέρεται στο μέσο πλήθος πελατών στο συγκεκριμένο υποσύστημα, το λ στο ρυθμό άφιξης στο συγκεκριμένο υποσύστημα και το $E[S]$ στον χρόνο παραμονής ενός πελάτη στο συγκεκριμένο υποσύστημα.

Θεωρώντας ως υποσύστημα τον χώρο αναμονής ενός συστήματος (δηλαδή την ουρά) παίρνουμε τη σχέση

$$E[Q_q] = \lambda E[W], \quad (2.6)$$

δηλαδή ο μέσος αριθμός πελατών στην ουρά ισούται με τον ρυθμό αφίξεων επί τον μέσο χρόνο αναμονής ενός πελάτη μέχρι να αρχίσει η εξυπηρέτησή του.

Ο νόμος του Little εφαρμοζόμενος στον χώρο εξυπηρέτησης δίνει τη σχέση

$$E[Q_s] = \lambda E[B] = \lambda b = \rho, \quad (2.7)$$

δηλαδή ο μέσος αριθμός πελατών στον χώρο εξυπηρέτησης που προφανώς ταυτίζεται με τον μέσο αριθμό απασχολημένων υπηρετών ισούται με τον ρυθμό συνωστισμού του συστήματος. Έτσι έχουμε μια δεύτερη ερμηνεία του ρυθμού συνωστισμού. Όχι μόνο είναι το μέσο ποσό εργασίας που εισέρχεται στο σύστημα ανά χρονική μονάδα αλλά επιπλέον εκφράζει και τον μέσο αριθμό απασχολημένων υπηρετών μια τυχούσα χρονική στιγμή (σε κατάσταση ισορροπίας). Επειδή ο μέσος αριθμός απασχολημένων υπηρετών ισούται με το πλήθος c των υπηρετών επί την πιθανότητα ένας υπηρέτης να είναι απασχολημένος συμπεραίνουμε ότι η οριακή πιθανότητα απασχολημένου υπηρέτη ή, ισοδύναμα, το ποσοστό του χρόνου απασχόλησης ενός υπηρέτη είναι $\frac{\rho}{c}$.

Επίσης, ειδικά για την $G/G/1$ ουρά έχουμε

$$\rho = E[Q_s] = 0 \Pr[Q_s = 0] + 1 \Pr[Q_s = 1] = \Pr[Q_s = 1] = \Pr[Q \geq 1] = 1 - p_0,$$

επομένως στην περίπτωση αυτή έχουμε ότι η οριακή πιθανότητα κενού συστήματος είναι

$$p_0 = 1 - \rho. \quad (2.8)$$

Συνοψίζοντας, θεωρώντας ως υποσύστημα τον χώρο εξυπηρέτησης ενός συστήματος, έχουμε το ακόλουθο αποτέλεσμα.

Πόρισμα 2.7 (Ερμηνείες του ρυθμού συνωστισμού) Σε ένα ευσταθές $G/G/c$ σύστημα ο ρυθμός συνωστισμού ισούται με τον μέσο αριθμό απασχολημένων υπηρετών. Το μακροπρόθεσμο ποσοστό του χρόνου που ένας υπηρέτης είναι απασχολημένος είναι ίσο με $\frac{\rho}{c}$.

Σε ένα ευσταθές $G/G/1$ σύστημα ο ρυθμός συνωστισμού ταυτίζεται επιπλέον με την πιθανότητα κατειλημένου υπηρέτη (δηλ. μη-κενού συστήματος). ◊

2.4 Ασκήσεις

Άσκηση 2.1 Έστω ένα σύστημα εξυπηρέτησης με μεμονωμένες μεταβάσεις (αφίξεις και αναχωρήσεις) και έστω Q_n^-, Q_n^+ οι αριθμοί των πελατών πριν την n -οστή άφιξη και μετά την n -οστή αναχώρηση, αντίστοιχα. Υποθέτουμε ότι το σύστημα είναι αρχικά κενό.

1. Να αποδείξετε ότι για κάθε $n \geq 1$ και $k \geq 0$ ισχύει ότι $Q_n^+ \leq k$ αν και μόνο αν $Q_{n+k+1}^- \leq k$.

2. Να αποδείξετε την ιδιότητα των μεμονωμένων μεταβάσεων, δηλαδή το θεώρημα 2.2.

Άσκηση 2.2 Με χρήση του νόμου του Little και της ιδιότητας PASTA, να βρείτε τα μακροπρόθεσμα ποσοστά του χρόνου κενού και απασχολημένου υπηρέτη, δηλαδή τις πιθανότητες p_0 και p_1 , σε μια $M/G/1/1$ ουρά με ρυθμό αφίξεων λ και μέσο χρόνο εξυπηρέτησης b .

Άσκηση 2.3 Βρείτε τον οριακό μέσο αριθμό πελατών στο σύστημα, $E[Q]$, στην $G/G/\infty$ ουρά με μέσο ενδιάμεσο χρόνο αφίξεων a και μέσο χρόνο εξυπηρέτησης b .

Άσκηση 2.4 Θεωρήστε μια $M/M/c$ ουρά με ρυθμό αφίξεων 5 πελάτες την ώρα και μέσο χρόνο εξυπηρέτησης ανά πελάτη 78 λεπτά.

1. Ποιος είναι ο ελάχιστος αριθμός υπηρετών c που χρειάζεται για να είναι το σύστημα ευσταθές (δηλαδή να μην απειρίζεται η ουρά);
2. Ποιος είναι ο ελάχιστος αριθμός υπηρετών που χρειάζεται αν η εργατική νομοθεσία επιβάλλει κάθε υπηρέτης να είναι απασχολημένος το πολύ το 80% του χρόνου του;

Άσκηση 2.5 Σε ένα σύστημα εξυπηρέτησης η διαδικασία αφίξεων είναι Poisson με ρυθμό $\lambda = 10$ πελάτες το λεπτό και το σύστημα έχει πεπερασμένη χωρητικότητα $k = 9$. Αν γνωρίζουμε ότι το μακροπρόθεσμο ποσοστό του χρόνου που στο σύστημα βρίσκονται $k = 9$ πελάτες είναι $p_k = p_9 = 0.8$ και ο μέσος αριθμός πελατών στο σύστημα είναι $E[Q] = 40$, να βρεθεί ο μέσος χρόνος παραμονής ενός αφιχθέντος και ενός εισερχόμενου πελάτη στο σύστημα.

2.5 Σχόλια

Τα βασικά αποτελέσματα παρουσιάζονται σχεδόν σε όλα τα εισαγωγικά βιβλία της Θεωρίας Ουρών Αναμονής. Τα εισαγωγικά βιβλία Wolff 1989, Φακίνος 2003, Tijms 2008 και Kulkarni 2010 παρουσιάζουν τα αποτελέσματα αυτά με αρκετή λεπτομέρεια. Για μια πολύ συνοπτική παρουσίαση μπορεί κανείς να ανατρέξει στο Adan και Resing 2001.

Αποτελέσματα σχετικά με την ευστάθεια συστημάτων εξυπηρέτησης και τη σχέση των κατανομών ισορροπίας (p_j), (a_j) και (d_j) μπορεί να βρει κανείς στο βιβλίο των Baccelli και Bremaud 1979, το οποίο όμως είναι πολύ προχωρημένου επιπέδου. Στη σχέση των κατανομών ισορροπίας σε συνεχή χρόνο και σε στιγμές γεγονότων αναφέρεται και το βιβλίο Sigman 1995 που είναι επίσης αρκετά θεωρητικό.

Η ιδιότητα PASTA αποδείχθηκε αυστηρά στη θεμελιώδη εργασία του Wolff 1982. Στο βιβλίο Wolff 1989 η ιδιότητα παρουσιάζεται με λεπτομέρεια.

Ο νόμος του Little αποδείχθηκε στη θεμελιώδη εργασία Little 1961. Διάφορες εναλλακτικές αποδείξεις και γενικεύσεις του παρουσιάζονται στις εργασίες Stidham 1972, Stidham 1974, καθώς και στο βιβλίο του Wolff 1989.

Ο νόμος του Little συνδυαζόμενος με την ιδιότητα PASTA δίνει μια ιδιαίτερα επιτυχημένη μέθοδο υπολογισμού του μέσου αριθμού πελατών και του μέσου χρόνου παραμονής ενός πελάτη σε πολλά συστήματα εξυπηρέτησης. Η μέθοδος αυτή αναφέρεται ως ανάλυση μέσης τιμής (Mean Value Analysis - MVA) και παρουσιάζεται στα κεφάλαια 5 και 13 για Μαρκοβιανά και μη-Μαρκοβιανά συστήματα αντίστοιχα.

Βιβλιογραφία

- [1] F. Baccelli και P. Bremaud. *Elements of Queueing Theory. Palm Martingale Calculus and Stochastic Recurrences, 2nd Edition*. New York: McGraw-Hill, Inc., 1979.
- [2] R.W. Wolff. "Poisson arrivals see time averages". Στο: *Operations Research* 30 (1982), σσ. 223–231.
- [3] J.D.C. Little. "A proof of the queueing formula $L = \lambda W$ ". Στο: *Operations Research* 9 (1961), σσ. 383–387.
- [4] S.Jr Stidham. " $L = \lambda W$: A discounted analogue and a new proof". Στο: *Operations Research* 20 (1972), σσ. 1115–1120.
- [5] S.Jr Stidham. "A last word on $L = \lambda W$ ". Στο: *Operations Research* 22 (1974), σσ. 417–421.
- [6] R.W. Wolff. *Stochastic Modeling and the Theory of Queues*. Englewood Cliffs, New Jersey: Prentice Hall, 1989. ISBN: 978-0138466923.
- [7] Δ. Φακίνος. *Ουρές Αναμονής: Θεωρία και Ασκήσεις*. Αθήνα: Εκδόσεις Συμμετρία, 2003. ISBN: 978-960-266-206-9.
- [8] H.C. Tijms. *A First Course in Stochastic Models, 2nd Edition*. Chichester: Wiley, 2008. ISBN: 978-0471498803.
- [9] V.G. Kulkarni. *Modeling and Analysis of Stochastic Systems, 2nd Edition*. CRC Press, 2010. ISBN: 978-1439808757.
- [10] I. Adan και J. Resing. *Queueing Theory*. Eindhoven: Notes available online, 2001.
- [11] K. Sigman. *Stationary Marked Point Processes: An Intuitive Approach*. Chapman και Hall/CRC, 1995. ISBN: 978-0412984310.

ΚΕΦΑΛΑΙΟ 3

ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ ΠΙΘΑΝΟΤΗΤΩΝ

Στο κεφάλαιο αυτό γίνεται μια σύνοψη κάποιων βασικών στοιχείων πιθανοτήτων που συνιστούν απαραίτητα εργαλεία για τη μελέτη συστημάτων στη Θεωρία Ουρών Αναμονής. Ουσιαστικά, επικεντρωνόμαστε σε τρεις έννοιες που παρουσιάζονται προς το τέλος ενός εισαγωγικού μαθήματος στις Πιθανότητες: στη δεσμευμένη μέση τιμή, στις πιθανογεννήτριες και στους μετασχηματισμούς Laplace-Stieltjes. Επίσης, γίνεται μια κάπως εκτεταμένη παρουσίαση της εκθετικής κατανομής που παίζει σημαντικό ρόλο στη μοντελοποίηση ενδιάμεσων χρόνων μεταξύ αφίξεων πελατών και χρόνων εξυπηρέτησης.

3.1 Δεσμευμένη μέση τιμή

Μια διακριτή διδιάστατη τυχαία μεταβλητή (X, Y) προσδιορίζεται από τη συνάρτηση (μάζας) πιθανότητάς της, δηλ. από τη συνάρτηση

$$f_{X,Y}(x, y) = \Pr[X = x, Y = y], \quad x, y \in \mathbb{R}.$$

Παρόμοια, μια (απόλυτα) συνεχής τυχαία μεταβλητή (X, Y) προσδιορίζεται από μια συνάρτηση πυκνότητας πιθανότητάς της, π.χ., από τη συνάρτηση

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x, y) \\ &= \lim_{\delta x \rightarrow 0^+, \delta y \rightarrow 0^+} \frac{\Pr[x < X \leq x + \delta x, y < Y \leq y + \delta y]}{\delta x \delta y}, \quad x, y \in \mathbb{R}, \end{aligned}$$

όπου $F_{X,Y}(x, y) = \Pr[X \leq x, Y \leq y]$ είναι η συνάρτηση κατανομής της (X, Y) .

Για κάθε y με $f_Y(y) > 0$, η δεσμευμένη συνάρτηση πιθανότητας ή συνάρτηση πυκνότητας πιθανότητας της X , δοθέντος ότι $Y = y$ ορίζεται ως

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad x \in \mathbb{R}.$$

Για κάθε σταθερό y , η $(f_{X|Y}(x|y) : x \in \mathbb{R})$ είναι συνάρτηση πιθανότητας ή συνάρτηση πυκνότητας, ανάλογα με το αν η X είναι διακριτή ή συνεχής. Συμβολίζουμε με $F_{X|Y}(x|y)$, $x \in \mathbb{R}$, την αντίστοιχη συνάρτηση κατανομής. Τότε, η αντίστοιχη μέση τιμή είναι η

$$E[X|Y = y] = \int_{-\infty}^{\infty} x dF_{X|Y}(x|y) \quad (3.1)$$

$$= \begin{cases} \sum_x x f_{X|Y}(x) & \text{αν } X \text{ διακριτή,} \\ \int_{-\infty}^{\infty} x f_{X|Y}(x) dx & \text{αν } X \text{ συνεχής.} \end{cases} \quad (3.2)$$

Η $E[X|Y = y]$ είναι ένας αριθμός που διαισθητικά αντιστοιχεί στην καλύτερη εκτίμηση της X υπό την πληροφορία ότι $Y = y$. Επομένως, θέτοντας $m_{X|Y}(y) = E[X|Y = y]$, έχουμε μια συνάρτηση που σε κάθε τιμή y της Y δίνει τη βέλτιστη εκτίμηση για την X . Η τυχαία μεταβλητή $m_{X|Y}(Y)$ είναι η καλύτερη εκτίμηση της X από συνάρτηση της τυχαίας μεταβλητής Y , συμβολίζεται με $E[X|Y]$ και αναφέρεται ως η δεσμευμένη μέση της X δοθείσης της Y .

Οι δυο διαφορετικές εκδοχές της δεσμευμένης μέσης τιμής, η $E[X|Y = y]$ και η $E[X|Y]$, δεν θα πρέπει να συγχέονται. Η πρώτη είναι αριθμός και αναφέρεται ως δεσμευμένη μέση τιμή της X δοθέντος ότι $Y = y$, ενώ η δεύτερη είναι τυχαία μεταβλητή, συνάρτηση της Y και αναφέρεται ως δεσμευμένη μέση τιμή της X δοθείσης της Y . Παρατηρούμε ότι ο ορισμός της $E[X|Y = y]$ για την περίπτωση διακριτής Y γράφεται εναλλακτικά και ως

$$E[X|Y = y] = \frac{E[X1_{\{Y=y\}}]}{\Pr[Y = y]}.$$

Ο ορισμός αυτός επιτρέπει να γενικεύσουμε την ιδέα της δεσμευμένης μέσης τιμής της X δοθέντος του ενδεχομένου $\{Y = y\}$ και να ορίσουμε τη δεσμευμένη μέση τιμή μιας τυχαίας μεταβλητής δοθέντος ενός οποιουδήποτε ενδεχομένου θετικής πιθανότητας. Έτσι, ορίζουμε τη μέση τιμή της X δοθέντος του ενδεχομένου A , με $\Pr[A] > 0$, ως

$$E[X|A] = \frac{E[X1_A]}{\Pr[A]}.$$

Αυτή η τρίτη εκδοχή δεσμευμένης μέσης τιμής (που γενικεύει την πρώτη εκδοχή) πρέπει να ερμηνευθεί διαισθητικά ως η καλύτερη εκτίμηση της X δοθέντος ότι το ενδεχόμενο A πραγματοποιήθηκε.

Η έννοια της δεσμευμένης μέσης τιμής ως προς ενδεχόμενο επιτρέπει τη διατύπωση του θεωρήματος δέσμευσης μέσης τιμής που διευκολύνει τον υπολογισμό μέσων τιμών κατ' αναλογία του θεωρήματος ολικής πιθανότητας για τον υπολογισμό πιθανοτήτων.

Θεώρημα 3.1 (Θεώρημα δέσμευσης μέσης τιμής) Αν $(B_i : i \geq 0)$ είναι μια διαμέριση του δειγματικού χώρου και X μια τυχαία μεταβλητή, τότε

$$E[X] = \sum_{i=0}^{\infty} \Pr[B_i] E[X|B_i].$$

Πράγματι, έχουμε ότι

$$E[X] = E[X \sum_{i=0}^{\infty} 1_{B_i}] = \sum_{i=0}^{\infty} E[X1_{B_i}] = \sum_{i=0}^{\infty} \Pr[B_i] E[X|B_i],$$

με την εναλλαγή αθροίσματος και μέσης τιμής να ισχύει όποτε υπάρχει η $E[|X|]$ (από το θεώρημα Beppo-Levi της μετροθεωρητικής Θεωρίας Πιθανοτήτων).

Το θεώρημα δέσμευσης μέσης τιμής είναι γενίκευση του θεωρήματος ολικής πιθανότητας. Πράγματι, αν θεωρήσουμε ένα ενδεχόμενο A και εφαρμόσουμε το θεώρημα δέσμευσης μέσης τιμής για την $X = 1_A$ παίρνουμε

$$\Pr[A] = E[1_A] = \sum_{i=0}^{\infty} \Pr[B_i] E[1_A|B_i] = \sum_{i=0}^{\infty} \Pr[B_i] \Pr[A|B_i].$$

Δοθείσης μιας διακριτής διδιάστατης τυχαίας μεταβλητής (X, Y) και θεωρώντας ως διαμέριση του δειγματικού χώρου τα σύνολα $B_i = \{Y = y_i\}$, όπου y_0, y_1, \dots οι δυνατές τιμές της Y παίρνουμε ότι

$$E[X] = \sum_{i=0}^{\infty} \Pr[Y = y_i] E[X|Y = y_i].$$

Γενικότερα ισχύει το ακόλουθο.

Θεώρημα 3.2 (Θεώρημα διπλής (ή επαναλαμβανόμενης) μέσης τιμής) Αν (X, Y) είναι διδιάστατη τυχαία μεταβλητή, τότε

$$\begin{aligned} E[X] &= E[E[X|Y]] = \int_{-\infty}^{\infty} E[X|Y = y] dF_Y(y) \\ &= \begin{cases} \sum_y E[X|Y = y] f_Y(y) & \text{αν } Y \text{ διακριτή,} \\ \int_{-\infty}^{\infty} E[X|Y = y] f_Y(y) dy & \text{αν } Y \text{ συνεχής,} \end{cases} \end{aligned}$$

όπου $f_Y(y)$ είναι η συνάρτηση πιθανότητας της Y για τη διακριτή περίπτωση και η συνάρτηση πυκνότητας πιθανότητας της Y για τη συνεχή περίπτωση. \diamond

Τα θεωρήματα δέσμευσης μέσης τιμής και διπλής μέσης τιμής χρησιμοποιούνται ευρύτατα για υπολογισμούς. Ιδιαίτερα σε περιπτώσεις που μας ενδιαφέρει η μέση τιμή ενός τυχαίου αθροίσματος N τυχαίων μεταβλητών, δηλαδή έχουμε κάποια τυχαία μεταβλητή της μορφής $S_N = \sum_{i=1}^N X_i$, τότε είναι βολικό να δεσμεύουμε στον αριθμό N των προσθετέων.

Παράδειγμα 3.3 (Υπολογισμός μέσης τιμής τυχαίου αθροίσματος) Έστω $\{X_i : i \geq 1\}$ ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών με $E[X_i] = \mu_X < \infty$ και $\text{Var}[X_i] = \sigma_X^2 < \infty$, για $i \geq 1$ και N μη-αρνητική, ακέραια τυχαία μεταβλητή, ανεξάρτητη των X_i με $E[N] = \mu_N < \infty$ και $\text{Var}[N] = \sigma_N^2 < \infty$. Έστω, επίσης $S_N = \sum_{i=1}^N X_i$. Ζητείται να υπολογιστεί η μέση τιμή της S_N .

Από το θεώρημα διπλής μέσης τιμής, έχουμε

$$E[S_N] = \sum_{n=0}^{\infty} \Pr[N = n] E[S_N|N = n].$$

Όμως

$$\begin{aligned} E[S_N|N = n] &= E\left[\sum_{i=1}^N X_i|N = n\right] = E\left[\sum_{i=1}^n X_i|N = n\right] \\ &= \sum_{i=1}^n E[X_i|N = n] = \sum_{i=1}^n E[X_i] = n\mu_X, \end{aligned}$$

οπότε

$$E[S_N] = \sum_{n=0}^{\infty} \Pr[N = n] n\mu_X = \mu_N \mu_X. \quad (3.3)$$

\diamond

Μια άλλη περίπτωση που το θεώρημα διπλής μέσης τιμής διευκολύνει τους υπολογισμούς είναι όταν η τυχαία μεταβλητή X της οποίας τη μέση τιμή προσπαθούμε να υπολογίσουμε έχει την ιδιότητα να αναγεννάται, δηλαδή να ξαναρχίζει από την αρχή όταν συμβεί κάποιο συγκεκριμένο ενδεχόμενο.

Παράδειγμα 3.4 (Μέση τιμή γεωμετρικής κατανομής) Έστω μια ακολουθία ανεξάρτητων και ισόνομων 0–1 τυχαίων μεταβλητών $\{X_i : i \geq 1\}$, με $\Pr[X_i = 1] = p$, $i \geq 1$, και $X = \inf\{n \geq 1 : X_n = 1\}$. Δηλαδή, μπορούμε να σκεφτόμαστε ότι επαναλαμβάνεται κάποιο πείραμα τύχης με δυο δυνατές εκβάσεις, 0 (αποτυχία) και 1 (επιτυχία) και μετράμε το πλήθος των επαναλήψεων του πειράματος μέχρι να εμφανιστεί για πρώτη φορά η έκβαση 1. Είναι φανερό ότι $\Pr[X = n] = (1 - p)^{n-1}p$, $n = 1, 2, \dots$, που είναι η συνάρτηση πιθανότητας γεωμετρικής κατανομής. Ο υπολογισμός της μέσης τιμής της $E[X]$ με τον ορισμό γίνεται χρησιμοποιώντας τον τύπο $E[X] = \sum_{n=1}^{\infty} n(1-p)^{n-1}p$, οπότε απαιτεί κάποιους αλγεβρικούς χειρισμούς με χρήση της γεωμετρικής σειράς $\sum_{n=0}^{\infty} t^n = 1/(1-t)$. Επειδή, όμως, η τυχαία μεταβλητή X αναγεννάται, δηλαδή ξαναρχίζει από την αρχή αν $X_1 = 0$, είναι πιο εύκολο να υπολογίσουμε την $E[X]$ δεσμεύοντας στην X_1 . Έχουμε

$$\begin{aligned} E[X] &= \Pr[X_1 = 0]E[X|X_1 = 0] + \Pr[X_1 = 1]E[X|X_1 = 1] \\ &= (1 - p)(1 + E[X]) + p \cdot 1 = 1 + (1 - p)E[X]. \end{aligned}$$

Λύνοντας ως προς $E[X]$ παίρνουμε $E[X] = 1/p$. ◊

3.2 Πιθανογεννήτριες - Μετασχηματισμοί Laplace-Stieltjes

Οι πιθανογεννήτριες και οι μετασχηματισμοί Laplace-Stieltjes χρησιμεύουν για τον εύκολο χειρισμό ακέραιων μη-αρνητικών και μη-αρνητικών τυχαίων μεταβλητών, αντίστοιχα. Στη Θεωρία Ουρών Αναμονής ο αριθμός των πελατών σε ένα σύστημα εξυπηρέτησης ή σε κάποιο υποσύστημά του είναι μη-αρνητική ακέραια τυχαία μεταβλητή, ενώ οι χρόνοι παραμονής, αναμονής και εξυπηρέτησης πελάτη είναι μη-αρνητικές τυχαίες μεταβλητές. Επομένως, η χρήση των πιθανογεννητριών και των μετασχηματισμών Laplace-Stieltjes αναμένεται να είναι αποτελεσματική. Ιδιαίτερα σημαντική είναι η χρήση των δύο αυτών εργαλείων για τη μελέτη αθροισμάτων ανεξάρτητων τυχαίων μεταβλητών. Και επειδή πολλές τυχαίες μεταβλητές που εμφανίζονται στη Θεωρία Ουρών Αναμονής γράφονται ως τέτοια αθροίσματα, η χρήση τους είναι ιδιαίτερα σημαντική. Π.χ., ο χρόνος αναμονής για έναν πελάτη, σε ένα σύστημα που λειτουργεί με την πειθαρχία FCFS, είναι το άθροισμα των χρόνων εξυπηρέτησης των πελατών που βρίσκει στο σύστημα κατά την άφιξή του.

Οι πιθανογεννήτριες ορίζονται μόνο για ακέριες μη-αρνητικές τυχαίες μεταβλητές, ενώ οι μετασχηματισμοί Laplace-Stieltjes μπορούν να χειριστούν γενικότερα μη-αρνητικές τυχαίες μεταβλητές. Οπότε για διακριτές τυχαίες μεταβλητές, όπως το πλήθος των πελατών σε ένα σύστημα, χρησιμοποιούμε πιθανογεννήτριες, ενώ για συνεχείς τυχαίες μεταβλητές, όπως ο χρόνος παραμονής ενός πελάτη, χρησιμοποιούμε τους μετασχηματισμούς Laplace-Stieltjes.

Ορισμός 3.5 (Πιθανογεννήτρια τυχαίας μεταβλητής) Έστω X μη-αρνητική ακέραια τυχαία μεταβλητή με συνάρτηση πιθανότητας $f_X(k) = \Pr[X = k]$, $k \geq 0$. Η πιθανογεννήτρια της X ορίζεται ως

$$P_X(z) = \sum_{k=0}^{\infty} f_X(k)z^k = E[z^X]. \quad (3.4)$$

◊

Είναι σημαντικό κατά τη χρήση των πιθανογεννητριών να έχουμε κατά νου τον διττό χαρακτήρα τους: Μια πιθανογεννήτρια είναι αναλυτικό αντικείμενο, δηλαδή δυναμοσειρά με συντελεστές τις πιθανότητες των διαφόρων τιμών μιας τυχαίας μεταβλητής και από την άλλη είναι πιθανοθεωρητικό αντικείμενο, δηλαδή μέση τιμή μιας συγκεκριμένης συνάρτησης της τυχαίας μεταβλητής. Αυτός ο διττός χαρακτήρας επάγει πολλούς τρόπους χρήσης των πιθανογεννητριών και διευκολύνει την απόδειξη των βασικών ιδιοτήτων τους. Κάποιες βασικές ιδιότητες των πιθανογεννητριών συνοψίζονται στο παρακάτω αποτέλεσμα.

Θεώρημα 3.6 (Ιδιότητες πιθανογεννητριών) Δεδομένης μιας ακέραιας μη-αρνητικής τυχαίας μεταβλητής X συμβολίζουμε με $(f_X(k) : k \geq 0)$ τη συνάρτηση πιθανότητάς της και με $P_X(z)$ την πιθανογεννήτριά της. Ισχύουν οι ακόλουθες ιδιότητες:

1. (Σύγκλιση στον κλειστό μοναδιαίο δίσκο):

Κάθε πιθανογεννήτρια $P_X(z)$ συγκλίνει τουλάχιστον στο $\{z \in \mathbb{C} : |z| \leq 1\}$. Επίσης, $P_X(1) = 1$.

2. (Χαρακτηρισμός τυχαίας μεταβλητής από την πιθανογεννήτριά της):

$$f_X(k) = \frac{P_X^{(k)}(0)}{k!}, k \geq 0.$$

Επομένως, $f_X(k) = f_Y(k)$, $k \geq 0$ αν και μόνο αν $P_X(z) = P_Y(z)$.

3. (Παραγοντικές ροπές μέσω πιθανογεννήτριας):

$$E[X(X-1)(X-2)\cdots(X-r+1)] = P_X^{(r)}(1), \quad r \geq 0.$$

Ιδιαίτερα

$$E[X] = P_X'(1),$$

και

$$\text{Var}[X] = P_X''(1) + P_X'(1) - (P_X'(1))^2.$$

4. (Πιθανογεννήτρια αθροίσματος ανεξάρτητων τυχαίων μεταβλητών):

Αν $\{X_i : 1 \leq i \leq n\}$ είναι ανεξάρτητες ακέραιες μη-αρνητικές τυχαίες μεταβλητές και $S_n = \sum_{i=1}^n X_i$ τότε

$$P_{S_n}(z) = P_{X_1}(z)P_{X_2}(z)\cdots P_{X_n}(z).$$

5. (Πιθανογεννήτρια τυχαίου αθροίσματος ανεξάρτητων και ισόνομων τυχαίων μεταβλητών):

Αν $\{X_i : i \geq 1\}$ είναι ακολουθία ανεξάρτητων και ισόνομων ακέραιων μη-αρνητικών τυχαίων μεταβλητών, N ακέραια μη-αρνητική τυχαία μεταβλητή, ανεξάρτητη των X_i και $S_N = \sum_{i=1}^N X_i$ τότε

$$P_{S_N}(z) = P_N(P_X(z)),$$

όπου $P_X(z)$ είναι η κοινή πιθανογεννήτρια των X_i , $i \geq 1$.

◊

Είναι εύκολο να δούμε γιατί αυτές οι ιδιότητες ισχύουν. Χωρίς να υπεισέλθουμε σε τεχνικές λεπτομέρειες, ας δούμε σε αδρές γραμμές τα αντίστοιχα επιχειρήματα. Καταρχήν,

$$P_X(1) = \sum_{k=0}^{\infty} f_X(k) = 1.$$

Για $|z| \leq 1$, από τον ορισμό της πιθανογεννήτριας ως δυναμοσειράς με θετικούς συντελεστές που αθροίζουν στο 1, έχουμε

$$|P_X(z)| \leq \sum_{k=0}^{\infty} f_X(k)|z|^k \leq \sum_{k=0}^{\infty} f_X(k) = 1.$$

Επομένως, κάθε πιθανογεννήτρια συγκλίνει στον κλειστό μοναδιαίο δίσκο του \mathbb{C} και μάλιστα η τιμή της στο 1 είναι 1. Έτσι έχουμε την πρώτη ιδιότητα του θεωρήματος 3.6. Η ιδιότητα αυτή είναι χρήσιμη για τον προσδιορισμό άγνωστων παραμέτρων που συχνά εμφανίζονται κατά τη λύση ενός προβλήματος με πιθανογεννήτριες.

Η δεύτερη ιδιότητα του θεωρήματος 3.6 είναι άμεση από το ανάπτυγμα Taylor μιας συνάρτησης σε δυναμοσειρά. Ο τύπος $f_X(k) = \frac{P_X^{(k)}(0)}{k!}$, $k \geq 0$, δεν είναι βέβαια εύχρηστος για τον υπολογισμό της συνάρτησης πιθανότητας μιας τυχαίας μεταβλητής, αφού απαιτεί συνεχείς παραγωγίσεις. Και πράγματι, θα δούμε ότι

υπάρχουν σαφώς αποτελεσματικότεροι τρόποι για να υπολογίσουμε μια συνάρτηση πιθανότητας, ξεκινώντας από την αντίστοιχη πιθανογεννήτρια. Όμως, το βασικό συμπέρασμα αυτής της ιδιότητας είναι ότι η γνώση της πιθανογεννήτριας αρκεί για τον υπολογισμό της συνάρτησης πιθανότητας μιας τυχαίας μεταβλητής. Με άλλα λόγια, η πιθανογεννήτρια χαρακτηρίζει την αντίστοιχη τυχαία μεταβλητή και περιέχει όλη την πληροφορία που χρειαζόμαστε για υπολογισμούς σχετικούς με αυτή την τυχαία μεταβλητή.

Η τρίτη ιδιότητα του θεωρήματος 3.6 προκύπτει από τυπική παραγωγή του τύπου-ορισμού της πιθανογεννήτριας και χρειάζεται βέβαια αναλυτικά επιχειρήματα για να ελεγχθεί ότι η παράγωγος μπορεί να «περάσει» μέσα στη σειρά:

$$\begin{aligned} P_X(z) = \sum_{k=0}^{\infty} f_X(k)z^k &\Rightarrow P_X^{(r)}(z) = \sum_{k=0}^{\infty} f_X(k) \frac{d^r}{dz^r} z^k \\ &\Rightarrow P_X^{(r)}(z) = \sum_{k=0}^{\infty} f_X(k)k(k-1)(k-2)\cdots(k-r+1)z^{k-r} \\ &\Rightarrow P_X^{(r)}(1) = \sum_{k=0}^{\infty} f_X(k)k(k-1)(k-2)\cdots(k-r+1) \\ &\Rightarrow P_X^{(r)}(1) = E[X(X-1)(X-2)\cdots(X-r+1)]. \end{aligned}$$

Για $r = 1$, έχουμε

$$E[X] = P_X'(1).$$

Για $r = 2$, έχουμε $P_X''(1) = E[X(X-1)]$, οπότε $Var[X] = E[X^2] - (E[X])^2 = E[X(X-1)] + E[X] - (E[X])^2$ που δίνει τελικά

$$Var[X] = P_X''(1) + P_X'(1) - (P_X'(1))^2.$$

Για την τέταρτη ιδιότητα, έχουμε:

$$P_{S_n}(z) = E[z^{S_n}] = E[z^{X_1+X_2+\cdots+X_n}] = E[z^{X_1}z^{X_2}\cdots z^{X_n}]$$

και από την ανεξαρτησία των X_1, X_2, \dots, X_n παίρνουμε τελικά

$$P_{S_n}(z) = E[z^{X_1}]E[z^{X_2}]\cdots E[z^{X_n}] = P_{X_1}(z)P_{X_2}(z)\cdots P_{X_n}(z).$$

Τέλος, για την πέμπτη ιδιότητα, έχουμε:

$$\begin{aligned} P_{S_N}(z) &= E[z^{S_N}] = \sum_{n=0}^{\infty} \Pr[N = n]E[z^{S_N}|N = n] \\ &= \sum_{n=0}^{\infty} \Pr[N = n]E[z^{S_n}|N = n] = \sum_{n=0}^{\infty} \Pr[N = n]E[z^{S_n}] \\ &= \sum_{n=0}^{\infty} \Pr[N = n]P_{S_n}(z) = \sum_{n=0}^{\infty} \Pr[N = n](P_X(z))^n = P_N(P_X(z)), \end{aligned}$$

όπου στην πρώτη ισότητα χρησιμοποιήσαμε τον ορισμό της πιθανογεννήτριας για την S_N , στη δεύτερη το θεώρημα της διπλής μέσης τιμής, στην τρίτη την πληροφορία της δέσμευσης, στην τέταρτη την ανεξαρτησία της N από τις X_1, X_2, \dots , στην πέμπτη τον ορισμό της πιθανογεννήτριας για την S_n , στην έκτη την ισονομία των X_1, X_2, \dots και την τέταρτη ιδιότητα του θεωρήματος και τέλος στην έβδομη τον ορισμό της πιθανογεννήτριας για την N .

Χρησιμοποιώντας τις ιδιότητες του θεωρήματος 3.6 μπορούμε σε διάφορα προβλήματα να συναγάγουμε τις πιθανογεννήτριες τυχαίων μεταβλητών. Κατόπιν αναλύοντάς τες σε δυνάμεις του z βρίσκουμε τη συνάρτηση

πιθανότητας που είναι συνήθως και αυτό που μας ενδιαφέρει. Για το σκοπό αυτό χρησιμοποιούνται κάποια βασικά αναπτύγματα Taylor:

$$\sum_{k=0}^{\infty} z^k = \frac{1}{1-z}, \quad |z| < 1 \quad (\text{γεωμετρική σειρά}), \quad (3.5)$$

$$\sum_{k=0}^{\infty} \frac{z^k}{k!} = e^z, \quad z \in \mathbb{C} \quad (\text{εκθετική σειρά}), \quad (3.6)$$

$$\sum_{k=0}^n \binom{n}{k} z^k = (1+z)^n, \quad n \in \mathbb{N}_0, \quad z \in \mathbb{C} \quad (\text{διωνυμικό άθροισμα}), \quad (3.7)$$

$$\sum_{k=0}^{\infty} \binom{x}{k} z^k = (1+z)^x, \quad x \in \mathbb{R}, \quad |z| < 1 \quad (\text{γενικευμένη διωνυμική σειρά}), \quad (3.8)$$

$$\sum_{k=0}^{\infty} \binom{k}{n} z^k = \frac{z^n}{(1-z)^{n+1}}, \quad n \in \mathbb{N}_0, \quad |z| < 1 \quad (\text{αρνητική διωνυμική σειρά}). \quad (3.9)$$

Τα τρία πρώτα αναπτύγματα (γεωμετρική σειρά, εκθετική σειρά και διωνυμικό άθροισμα) είναι ευρέως γνωστά και υπάρχουν σχεδόν σε κάθε βιβλίο απειροστικού λογισμού. Στο τρίτο το πεπερασμένο άθροισμα γράφεται και ως σειρά για $k = 0$ έως ∞ , αφού οι διωνυμικοί συντελεστές $\binom{n}{k}$ είναι 0 για $k > n$. Όταν αντί του ακεραίου n έχουμε έναν οποιοδήποτε πραγματικό x έχουμε τη γενίκευση του διωνυμικού αθροίσματος που αντιστοιχεί στο τέταρτο ανάπτυγμα. Το πέμπτο ανάπτυγμα της αρνητικής διωνυμικής σειράς προκύπτει με n φορές παραγωγή της γεωμετρικής σειράς:

$$\begin{aligned} \sum_{k=0}^{\infty} z^k &= \frac{1}{1-z} \stackrel{d^n/dz^n}{\Rightarrow} \sum_{k=n}^{\infty} k(k-1)(k-2)\cdots(k-n+1)z^{k-n} = \frac{n!}{(1-z)^{n+1}} \\ \stackrel{\times z^n}{\Rightarrow} \sum_{k=n}^{\infty} \frac{k!}{(k-n)!} z^k &= \frac{n!z^n}{(1-z)^{n+1}} \Rightarrow \sum_{k=n}^{\infty} \binom{k}{n} z^k = \frac{z^n}{(1-z)^{n+1}}. \end{aligned}$$

Με βάση αυτά τα αναπτύγματα είναι εύκολο να υπολογιστούν οι πιθανογεννήτριες των κλασικών διακριτών κατανομών.

Θεώρημα 3.7 (Πιθανογεννήτριες κλασικών διακριτών κατανομών) Για τις κλασικές διακριτές κατανομές οι πιθανογεννήτριες έχουν ως εξής:

1. Για τη γεωμετρική κατανομή με παράμετρο ρ ($Geom(\rho)$), με συνάρτηση πιθανότητας $f_X(k) = (1-\rho)\rho^k$, $k \geq 0$, η πιθανογεννήτρια είναι η

$$P_X(z) = \frac{1-\rho}{1-\rho z}. \quad (3.10)$$

2. Για την κατανομή Poisson με παράμετρο λ ($Poisson(\lambda)$), με συνάρτηση πιθανότητας $f_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k \geq 0$, η πιθανογεννήτρια είναι η

$$P_X(z) = e^{-\lambda(1-z)}. \quad (3.11)$$

3. Για τη διωνυμική κατανομή με παραμέτρους n, p ($Bin(n, p)$), με συνάρτηση πιθανότητας $f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$, $0 \leq k \leq n$, η πιθανογεννήτρια είναι η

$$P_X(z) = (1-p + pz)^n. \quad (3.12)$$

4. Για την αρνητική διωνυμική κατανομή με παραμέτρους n, ρ ($NegBin(n, \rho)$), με συνάρτηση πιθανότητας $f_X(k) = \binom{k+n-1}{n-1} \rho^n (1-\rho)^k, k \geq 0$, η πιθανογεννήτρια είναι η

$$P_X(z) = \left(\frac{1-\rho}{1-\rho z} \right)^n. \quad (3.13)$$

◊

Οι πιθανογεννήτριες των κλασικών διακριτών κατανομών που αναφέρονται στο θεώρημα 3.7 είναι σε 1-1 αντιστοιχία με τα αναπτύγματα Taylor (3.5), (3.6), (3.7) και (3.9) και προκύπτουν από αυτά.

Η διαδικασία της εύρεσης της συνάρτησης πιθανότητας μιας μη-αρνητικής ακέραιας τυχαίας μεταβλητής από την αντίστοιχη πιθανογεννήτρια αναφέρεται ως αντιστροφή της αντίστοιχης πιθανογεννήτριας. Δεν είναι πάντα δυνατή αναλυτικά και γι αυτό συχνά επιλέγονται αριθμητικές μέθοδοι. Για κάποιες κλάσεις είναι δυνατή. Δυο τέτοιες βασικές περιπτώσεις είναι όταν η πιθανογεννήτρια είναι ρητής μορφής, δηλαδή $P_X(z) = \frac{N(z)}{D(z)}$ ή εκθετικής μορφής με ρητό εκθέτη, $P_X(z) = e^{\frac{N(z)}{D(z)}}$, όπου τα $N(z)$ και $D(z)$ είναι πολυώνυμα.

Η διαδικασία αναλυτικής αντιστροφής μιας ρητής πιθανογεννήτριας περιγράφεται στον παρακάτω αλγόριθμο.

Αλγόριθμος 3.8 (Αντιστροφή ρητής πιθανογεννήτριας) Έστω μια ρητή πιθανογεννήτρια $P_X(z) = \frac{N(z)}{D(z)}$ με $N(z)$ και $D(z)$ πολυώνυμα. Για την αντιστροφή της ακολουθούμε τα εξής:

Βήμα 1ο: Παραγοντοποίηση του $D(z)$.

Βήμα 2ο: Αν υπάρχουν άγνωστες παράμετροι στην πιθανογεννήτρια τότε χρησιμοποιείται η ιδιότητα 1 του θεωρήματος 3.6. Ιδιαίτερα χρησιμοποιείται η εξίσωση $P_X(1) = 1$ που δίνει μια εξίσωση για τις άγνωστες παραμέτρους. Επιπλέον εξισώσεις προκύπτουν απαιτώντας κάθε ρίζα του $D(z)$ στο $\{z \in \mathbb{C} : |z| \leq 1\}$ να είναι και ρίζα του $N(z)$, ώστε να συγκλίνει η $P_X(z)$ στο $\{z \in \mathbb{C} : |z| \leq 1\}$.

Βήμα 3ο: Ανάλυση σε απλά κλάσματα της $P_X(z)$.

Βήμα 4ο: Χρησιμοποίηση της γεωμετρικής και της αρνητικής διωνυμικής σειράς για την ανάπτυξη των απλών κλασμάτων σε δυνάμεις του z .

◊

Παράδειγμα 3.9 (Αντιστροφή ρητής πιθανογεννήτριας) Να βρεθεί η συνάρτηση πιθανότητας ($f_X(k) : k \geq 0$) της τυχαίας μεταβλητής X με πιθανογεννήτρια

$$P_X(z) = \frac{az + b}{z^2 - \frac{5}{2}z + 1}.$$

Λύση: Πρόκειται για μια ρητή πιθανογεννήτρια, οπότε ακολουθούμε τα βήματα που περιγράφηκαν παραπάνω.

Βήμα 1ο: Ο παρονομαστής παραγοντοποιείται ως $z^2 - \frac{5}{2}z + 1 = (z-2)(z-\frac{1}{2})$.

Βήμα 2ο: Η σχέση $P_X(1) = 1$ δίνει την εξίσωση $-2(a+b) = 1$ για τις άγνωστες παραμέτρους a, b . Επίσης η ρίζα $\frac{1}{2}$ του παρονομαστή της πιθανογεννήτριας πρέπει να είναι και ρίζα του αριθμητή ώστε η πιθανογεννήτρια να συγκλίνει στον κλειστό μοναδιαίο δίσκο, οπότε έχουμε και ότι $a\frac{1}{2} + b = 0$. Λύνοντας το σύστημα των δυο αυτών εξισώσεων έχουμε ότι $(a, b) = (-1, \frac{1}{2})$. Επομένως η πιθανογεννήτρια γράφεται ως

$$P_X(z) = \frac{1}{2-z}.$$

Βήμα 3ο: Η πιθανογεννήτρια είναι ήδη σε μορφή απλού κλάσματος.

Βήμα 4ο: Αναπτύσσουμε την πιθανογεννήτρια σε δυνάμεις του z χρησιμοποιώντας τη γεωμετρική σειρά.

$$P_X(z) = \frac{1}{2-z} = \frac{1}{2} \cdot \frac{1}{1-\frac{z}{2}} = \frac{1}{2} \sum_{k=0}^{\infty} \left(\frac{z}{2}\right)^k = \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^{k+1} z^k.$$

$$\text{Άρα } f_X(k) = \Pr[X = k] = \left(\frac{1}{2}\right)^{k+1}, k \geq 0.$$

◊

Παράδειγμα 3.10 (Αντιστροφή ρητής πιθανογεννήτριας) Να βρεθεί η συνάρτηση πιθανότητας ($f_X(k) : k \geq 0$) της τυχαίας μεταβλητής X με πιθανογεννήτρια

$$P_X(z) = \frac{c - 15z}{54 - 63z + 18z^2}.$$

Λύση: Πρόκειται για ρητή πιθανογεννήτρια, οπότε ακολουθούμε τα βήματα που περιγράφηκαν παραπάνω.

Βήμα 1ο: Ο παρονομαστής παραγοντοποιείται ως $54 - 63z + 18z^2 = 18(2 - z)\left(\frac{3}{2} - z\right)$.

Βήμα 2ο: Η σχέση $P_X(1) = 1$ δίνει την εξίσωση $\frac{c-15}{9} = 1$ από την οποία συνάγουμε την άγνωστη παράμετρο $c = 24$. Παρατηρήστε ότι στην περίπτωση αυτή οι ρίζες του παρονομαστή της πιθανογεννήτριας (οι 2 και $\frac{3}{2}$) είναι εκτός του κλειστού μοναδιαίου δίσκου και επομένως δεν δίνουν επιπλέον εξισώσεις.

Βήμα 3ο: Αναλύουμε την πιθανογεννήτρια σε απλά κλάσματα:

$$P_X(z) = \frac{24 - 15z}{18(2 - z)\left(\frac{3}{2} - z\right)} = \frac{A}{2 - z} + \frac{B}{\frac{3}{2} - z}$$

Για να προσδιορίσουμε το A , πολλαπλασιάζουμε την πιθανογεννήτρια με $2 - z$ και κατόπιν θέτουμε $z = 2$, οπότε έχουμε

$$\frac{24 - 15z}{18\left(\frac{3}{2} - z\right)} = A + \frac{B(2 - z)}{\frac{3}{2} - z} \stackrel{z=2}{\Rightarrow} A = \frac{24 - 15 \cdot 2}{18\left(\frac{3}{2} - 2\right)} = \frac{2}{3}.$$

Ομοίως, για να βρούμε το B , πολλαπλασιάζουμε την πιθανογεννήτρια με $\frac{3}{2} - z$ και κατόπιν θέτουμε $z = \frac{3}{2}$, οπότε παίρνουμε $B = \frac{1}{6}$.

Βήμα 4ο: Αναπτύσσουμε την πιθανογεννήτρια σε δυνάμεις του z χρησιμοποιώντας τη γεωμετρική σειρά. Έχουμε:

$$\begin{aligned} P_X(z) &= \frac{2}{3} \cdot \frac{1}{2-z} + \frac{1}{6} \cdot \frac{1}{\frac{3}{2}-z} = \frac{1}{3} \cdot \frac{1}{1-\frac{z}{2}} + \frac{1}{9} \cdot \frac{1}{1-\frac{2z}{3}} \\ &= \frac{1}{3} \sum_{k=0}^{\infty} \left(\frac{z}{2}\right)^k + \frac{1}{9} \sum_{k=0}^{\infty} \left(\frac{2z}{3}\right)^k = \sum_{k=0}^{\infty} \left[\frac{1}{3} \cdot \left(\frac{1}{2}\right)^k + \frac{1}{9} \cdot \left(\frac{2}{3}\right)^k \right] z^k. \end{aligned}$$

$$\text{Άρα } f_X(k) = \Pr[X = k] = \frac{1}{3} \cdot \left(\frac{1}{2}\right)^k + \frac{1}{9} \cdot \left(\frac{2}{3}\right)^k, k \geq 0.$$

◊

Μια εναλλακτική της εύρεσης κλειστού τύπου για την συνάρτηση πιθανότητας μιας τυχαίας μεταβλητής με γνωστή πιθανογεννήτρια είναι η διατύπωση ενός αναδρομικού σχήματος. Αν η πιθανογεννήτρια είναι της ρητής μορφής $P_X(z) = \frac{N(z)}{D(z)}$, με $N(z)$, $D(z)$ πολώνυμα, μπορούμε να πολλαπλασιάσουμε με $D(z)$ και κατόπιν να αναπτύξουμε τη σχέση $D(z)P_X(z) = N(z)$ σε δυνάμεις του z . Η διαδικασία αυτή δίνει εύκολα ένα αναδρομικό σχήμα. Στο παρακάτω παράδειγμα εφαρμόζουμε αυτή την ιδέα στην πιθανογεννήτρια του παραδείγματος 3.10.

Παράδειγμα 3.11 (Συνέχεια του παραδείγματος 3.10) Να διατυπωθεί ένα αναδρομικό σχήμα για τον υπολογισμό της συνάρτησης πιθανότητας ($f_X(k) : k \geq 0$) της τυχαίας μεταβλητής X με πιθανογεννήτρια

$$P_X(z) = \frac{24 - 15z}{54 - 63z + 18z^2}.$$

Λύση: Έχουμε ότι

$$\begin{aligned} (54 - 63z + 18z^2)P_X(z) &= 24 - 15z \\ \Rightarrow 54 \sum_{i=0}^{\infty} f_X(i)z^i - 63 \sum_{i=0}^{\infty} f_X(i)z^{i+1} + 18 \sum_{i=0}^{\infty} f_X(i)z^{i+2} &= 24 - 15z \\ \Rightarrow 54 \sum_{k=0}^{\infty} f_X(k)z^k - 63 \sum_{k=1}^{\infty} f_X(k-1)z^k + 18 \sum_{k=2}^{\infty} f_X(k-2)z^k &= 24 - 15z. \end{aligned}$$

Εξισώνοντας τους συντελεστές του z^k παίρνουμε:

$$\begin{aligned} k = 0 : \quad 54f_X(0) &= 24 \Rightarrow f_X(0) = \frac{4}{9}. \\ k = 1 : \quad 54f_X(1) - 63f_X(0) &= -15 \Rightarrow f_X(1) = \frac{13}{54}. \\ k \geq 2 : \quad 54f_X(k) - 63f_X(k-1) + 18f_X(k-2) &= 0 \\ \Rightarrow f_X(k) &= \frac{7f_X(k-1) - 2f_X(k-2)}{6}, \quad k \geq 2. \end{aligned}$$

◊

Όπως είδαμε παραπάνω, οι πιθανογεννήτριες ορίζονται μόνο για ακέραιες μη-αρνητικές τυχαίες μεταβλητές. Η αντίστοιχη έννοια για τη μελέτη γενικών μη-αρνητικών τυχαίων μεταβλητών είναι ο μετασχηματισμός Laplace-Stieltjes. Για τον ορισμό του μετασχηματισμού αυτού χρειαζόμαστε την έννοια του ολοκληρώματος Riemann-Stieltjes που αποτελεί γενίκευση του ολοκληρώματος Riemann. Το ολοκλήρωμα Riemann-Stieltjes μιας συνάρτησης $f(x)$ ως προς μια συνάρτηση $\phi(x)$ σε ένα διάστημα I ορίζεται για μια αρκετά ευρεία κλάση συναρτήσεων $f(x)$ και $\phi(x)$ και συμβολίζεται με $\int f(x)d\phi(x)$ ή με $\int_a^b f(x)d\phi(x)$, αν a, b είναι τα άκρα του I . Για τους σκοπούς του βιβλίου θα το χρησιμοποιούμε για την περίπτωση που η $f(x)$ είναι συνεχής συνάρτηση και η $\phi(x)$ είναι κατά τμήματα συνεχής και παραγωγίσιμη μεταξύ των σημείων ασυνεχειάς της. Τότε αποδεικνύεται ότι

$$\int_a^b f(x)d\phi(x) = \sum_{x \in [a, b]} f(x)(\phi(x) - \phi(x-)) + \int_a^b f(x)\phi'(x)dx, \quad (3.14)$$

όπου η άθροιση εκτείνεται σε όλα τα $x \in [a, b]$ στα οποία η $\phi(x)$ είναι ασυνεχής. Η σχέση (3.14) μπορεί να εκληφθεί ως ο ορισμός του ολοκληρώματος Riemann-Stieltjes μιας συνάρτησης $f(x)$ ως προς μια συνάρτηση $\phi(x)$, εφόσον οι συναρτήσεις αυτές ανήκουν στις προαναφερθείσες κλάσεις συναρτήσεων. Είναι φανερό ότι για $\phi(x) = x$ το ολοκλήρωμα Riemann-Stieltjes ανάγεται στο γνωστό ολοκλήρωμα Riemann. Επίσης, αν

X είναι μια τυχαία μεταβλητή με συνάρτηση κατανομής $F_X(x)$, μικτή, χωρίς ιδιάζον μέρος, και $g(x)$ είναι μια συνεχής συνάρτηση, τότε:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)dF_X(x) \quad (3.15)$$

$$= \sum_{x \in \mathbb{R}} g(x)(F_X(x) - F_X(x-)) + \int_{-\infty}^{\infty} g(x)F'(x)dx. \quad (3.16)$$

Οπότε το ολοκλήρωμα Riemann-Stieltjes, για τους σκοπούς του βιβλίου, μπορεί να θεωρηθεί ένας βολικός συμβολισμός για την ενιαία αντιμετώπιση διακριτών, συνεχών και μικτών τυχαίων μεταβλητών. Ήδη έχουμε χρησιμοποιήσει αυτόν τον συμβολισμό στην εξίσωση (1.10) για να ενοποιήσουμε την παρουσίαση της μέσης τιμής συνάρτησης τυχαίας μεταβλητής για τη διακριτή και τη συνεχή περίπτωση (βλέπε (1.11)), όπως και για τη δεσμευμένη μέση τιμή στην παράγραφο 3.1 (βλέπε εξίσωση (3.1)). Είμαστε τώρα έτοιμοι να προχωρήσουμε στον ορισμό του μετασχηματισμού Laplace-Stieltjes μιας συνάρτησης και εν συνεχεία του μετασχηματισμού Laplace-Stieltjes μιας τυχαίας μεταβλητής.

Ορισμός 3.12 (Μετασχηματισμός Laplace-Stieltjes συνάρτησης) Έστω $\phi(x)$ συνάρτηση ορισμένη στο $[0, \infty)$, κατά τμήματα συνεχής και παραγωγίσιμη μεταξύ των σημείων ασυνέχειάς της. Ο μετασχηματισμός Laplace-Stieltjes (LS) της $\phi(x)$ στο $[0, \infty)$ ορίζεται ως

$$\int_0^{\infty} e^{-sx}d\phi(x) = \sum_{x \in [0, \infty)} e^{-sx}(\phi(x) - \phi(x-)) + \int_0^{\infty} e^{-sx}\phi'(x)dx,$$

δεχόμενοι συμβατικά ότι η $\phi(x)$ έχει άλμα ασυνέχειας μεγέθους $\phi(0)$ στο σημείο 0 (βάζοντας στον παραπάνω τύπο $\phi(0-) = 0$). ◊

Ορισμός 3.13 (Μετασχηματισμός Laplace-Stieltjes τυχαίας μεταβλητής) Έστω X μη-αρνητική τυχαία μεταβλητή με συνάρτηση κατανομής $F_X(x) = \Pr[X \leq x]$, $x \in \mathbb{R}$. Ο μετασχηματισμός Laplace-Stieltjes (LS) της X ορίζεται ως ο μετασχηματισμός Laplace-Stieltjes της συνάρτησης κατανομής της στο $[0, \infty)$, δηλαδή είναι

$$\tilde{F}_X(s) = \int_0^{\infty} e^{-sx}dF_X(x) = E[e^{-sX}]. \quad (3.17)$$

Όπως και με τις πιθανογεννήτριες, έτσι και με τους μετασχηματισμούς Laplace-Stieltjes είναι σημαντικό να έχουμε κατά νου τον διττό χαρακτήρα τους: Κάθε μετασχηματισμός είναι μια συνάρτηση που ορίζεται μέσω γενικευμένου ολοκληρώματος, αλλά ταυτόχρονα είναι και η μέση τιμή μιας συγκεκριμένης συνάρτησης της τυχαίας μεταβλητής. Ανάλογα με την περίπτωση, χρησιμοποιείται άλλοτε η αναλυτική και άλλοτε η πιθανοθεωρητική έκφραση ενός μετασχηματισμού Laplace-Stieltjes. Κάποιες βασικές ιδιότητες των μετασχηματισμών συνοψίζονται στο παρακάτω αποτέλεσμα.

Θεώρημα 3.14 (Ιδιότητες μετασχηματισμών Laplace-Stieltjes) Δεδομένης μιας μη-αρνητικής τυχαίας μεταβλητής X συμβολίζουμε με $(F_X(x) : x \in \mathbb{R})$ τη συνάρτηση κατανομής της και με $\tilde{F}_X(s)$ τον μετασχηματισμό Laplace-Stieltjes της. Ισχύουν οι ακόλουθες ιδιότητες:

1. (Σύγκλιση στο μη-αρνητικό πραγματικό ημιεπίπεδο):

Κάθε μετασχηματισμός Laplace-Stieltjes $\tilde{F}_X(s)$ συγκλίνει τουλάχιστον στο $\{s \in \mathbb{C} : \Re(s) \geq 0\}$. Επίσης, $\tilde{F}_X(0) = 1$.

2. (Χαρακτηρισμός μη-αρνητικής τυχαίας μεταβλητής από τον μετασχηματισμό Laplace-Stieltjes της):

Δυο τυχαίες μεταβλητές X και Y είναι ισόνομες δηλαδή, $F_X(x) = F_Y(x)$, $x \in \mathbb{R}$ αν και μόνο αν $\tilde{F}_X(s) = \tilde{F}_Y(s)$, $s \in \mathbb{C}$ με $\Re(s) \geq 0$.

3. (Ροπές μέσω μετασχηματισμών Laplace-Stieltjes):

$$E[X^r] = (-1)^r \tilde{F}_X^{(r)}(0), \quad r \geq 0.$$

Ιδιαίτερα,

$$E[X] = -\tilde{F}'_X(0),$$

και

$$\text{Var}[X] = \tilde{F}''_X(0) - (\tilde{F}'_X(0))^2.$$

4. (Μετασχηματισμός Laplace-Stieltjes αθροίσματος ανεξάρτητων τυχαίων μεταβλητών):

Αν $\{X_i : 1 \leq i \leq n\}$ είναι ανεξάρτητες μη-αρνητικές τυχαίες μεταβλητές και $S_n = \sum_{i=1}^n X_i$, τότε

$$\tilde{F}_{S_n}(s) = \tilde{F}_{X_1}(s)\tilde{F}_{X_2}(s) \cdots \tilde{F}_{X_n}(s).$$

5. (Μετασχηματισμός Laplace-Stieltjes τυχαίου αθροίσματος ανεξάρτητων και ισόνομων τυχαίων μεταβλητών):

Αν $\{X_i : i \geq 1\}$ είναι ακολουθία ανεξάρτητων και ισόνομων μη-αρνητικών τυχαίων μεταβλητών, N ακέραια μη-αρνητική τυχαία μεταβλητή, ανεξάρτητη των X_i και $S_N = \sum_{i=1}^N X_i$ τότε

$$\tilde{F}_{S_N}(s) = P_N(\tilde{F}_X(s)),$$

όπου $P_N(z)$ είναι η πιθανογεννήτρια της N και $\tilde{F}_X(s)$ ο κοινός μετασχηματισμός Laplace-Stieltjes των X_i , $i \geq 1$.

◊

Είναι εύκολο να δούμε γιατί αυτές οι ιδιότητες ισχύουν. Τα επιχειρήματα είναι εντελώς ανάλογα με αυτά που χρησιμοποιήσαμε για τις παρόμοιες ιδιότητες των πιθανογεννητριών και γι αυτό η αιτιολόγηση των ιδιοτήτων παραλείπεται. Χρησιμοποιώντας τις ιδιότητες αυτές μπορούμε σε διάφορα προβλήματα να συναγάγουμε τους μετασχηματισμούς Laplace-Stieltjes τυχαίων μεταβλητών. Κατόπιν χρειαζόμαστε και πάλι κάποιες τεχνικές αντιστροφής ώστε να βρούμε τις υποκείμενες συναρτήσεις κατανομών. Στον πίνακα 3.1 βλέπουμε κάποιες βασικές αντιστοιχίες τυχαίων μεταβλητών, συναρτήσεων και μετασχηματισμών Laplace-Stieltjes.

Για την αντιστροφή ρητών μετασχηματισμών Laplace-Stieltjes η μεθοδολογία είναι λίγο-πολύ η ίδια με την αντιστροφή των πιθανογεννητριών. Χωρίς να υπεισέλθουμε σε λεπτομέρειες, ας δούμε μερικά παραδείγματα.

Παράδειγμα 3.15 (Αντιστροφή μετασχηματισμού Laplace-Stieltjes) Να βρεθεί η συνάρτηση κατανομής $F_X(x)$ της μη-αρνητικής τυχαίας μεταβλητής X με μετασχηματισμό Laplace-Stieltjes

$$\tilde{F}_X(s) = \frac{c}{s^2 + 5s + 6}.$$

Λύση: Έχουμε ότι $\tilde{F}_X(0) = 1$, οπότε $c = 6$. Αναλύουμε τον μετασχηματισμό σε απλά κλάσματα. Έχουμε ότι ο παρονομαστής του μετασχηματισμού παραγοντοποιείται ως $(s+2)(s+3)$, οπότε

$$\tilde{F}_X(s) = \frac{6}{(s+2)(s+3)} = \frac{A}{s+2} + \frac{B}{s+3}.$$

Για να προσδιορίσουμε τον A , πολλαπλασιάζουμε τον μετασχηματισμό με $s+2$ και κατόπιν θέτουμε $s = -2$, οπότε έχουμε

$$\frac{6}{s+3} = A + \frac{B(s+2)}{s+3} \stackrel{s=-2}{\Rightarrow} A = 6.$$

Τυχαία μεταβλητή X	Συνάρτηση $\phi(x)$ (ή $F_X(x)$)	Μετασχηματισμός LS $\tilde{\phi}(s)$ (ή $\tilde{F}_X(s)$)
–	$\phi(x) = x$	$\tilde{\phi}(s) = \frac{1}{s}$
–	$\phi(x) = \int_0^\infty \phi_1(x-u)d\phi_2(u)$	$\tilde{\phi}(s) = \tilde{\phi}_1(s)\tilde{\phi}_2(s)$
–	$\phi(x) = a_1\phi_1(x) + a_2\phi_2(x)$	$\tilde{\phi}(s) = a_1\tilde{\phi}_1(s) + a_2\tilde{\phi}_2(s)$
$X = c \geq 0$ με πιθανότητα 1	$F_X(x) = \begin{cases} 0 & \text{αν } x < c \\ 1 & \text{αν } x \geq c \end{cases}$	$\tilde{F}_X(s) = e^{-cs}$
$X \sim \text{Exp}(\lambda)$	$F_X(x) = 1 - e^{-\lambda x}$	$\tilde{F}_X(s) = \frac{\lambda}{\lambda+s}$
$X \sim \text{Erlang}(n, \lambda)$	$F_X(x) = 1 - \sum_{k=0}^{n-1} e^{-\lambda x} \frac{(\lambda x)^k}{k!}$	$\tilde{F}_X(s) = \left(\frac{\lambda}{\lambda+s}\right)^n$
$Z = X + Y, X, Y$ ανεξ.	$F_Z(z) = \int_0^\infty F_Y(z-x)dF_X(x)$	$\tilde{F}_Z(s) = \tilde{F}_X(s)\tilde{F}_Y(s)$
$Z = \begin{cases} X & \text{με πιθ. } p \\ Y & \text{με πιθ. } 1-p \end{cases}$	$F_Z(z) = pF_X(z) + (1-p)F_Y(z)$	$\tilde{F}_Z(s) = p\tilde{F}_X(s) + (1-p)\tilde{F}_Y(s)$

Πίνακας 3.1: Συναρτήσεις και αντίστοιχοι μετασχηματισμοί Laplace-Stieltjes.

Ομοίως, για να βρούμε το B , πολλαπλασιάσουμε τον μετασχηματισμό με $s + 3$ και κατόπιν θέτουμε $s = -3$, οπότε παίρνουμε $B = -6$. Επομένως, έχουμε

$$\tilde{F}_X(s) = \frac{6}{s+2} - \frac{6}{s+3} = 3\frac{2}{s+2} - 2\frac{3}{s+3}$$

Όμως, από τον πίνακα 3.1, βλέπουμε ότι οι $\frac{2}{s+2}$ και $\frac{3}{s+3}$ είναι οι μετασχηματισμοί Laplace-Stieltjes των κατανομών $\text{Exp}(2)$ και $\text{Exp}(3)$, αντίστοιχα και χρησιμοποιώντας την γραμμικότητα του μετασχηματισμού Laplace-Stieltjes (βλ. τρίτη ιδιότητα του πίνακα 3.1) έχουμε ότι

$$F_X(x) = 3(1 - e^{-2x}) - 2(1 - e^{-3x}), \quad x \geq 0.$$

◊

Πέραν της υπολογιστικής τους χρησιμότητας, οι μετασχηματισμοί Laplace-Stieltjes χρησιμοποιούνται και για τις αποδείξεις θεωρητικών αποτελεσμάτων.

Παράδειγμα 3.16 (Γεωμετρικό άθροισμα εκθετικών τυχαίων μεταβλητών) Έστω $\{X_i : i \geq 1\}$ ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών με την $\text{Exp}(\lambda)$ κατανομή και N ανεξάρτητη των $X_i, i \geq 1$, γεωμετρικά κατανομημένη με συνάρτηση πιθανότητας $\Pr[N = n] = (1 - p)^{n-1}p, n \geq 1$. Αν $S_N = \sum_{i=1}^N X_i$, τότε να βρεθεί η κατανομή της S_N .

Λύση: Έχουμε ότι ο κοινός μετασχηματισμός Laplace-Stieltjes των X_i είναι $\tilde{F}_X(s) = \frac{\lambda}{s+\lambda}$, ενώ η πιθανογεννήτρια της N είναι

$$P_N(z) = \sum_{n=1}^\infty (1-p)^{n-1}pz^n = \frac{pz}{1 - (1-p)z}.$$

Από το θεώρημα 3.14, ιδιότητα 5, έχουμε ότι ο μετασχηματισμός $\tilde{F}_{S_N}(s)$ της S_N είναι

$$\tilde{F}_{S_N}(s) = P_N(\tilde{F}_X(s)) = \frac{p\frac{\lambda}{s+\lambda}}{1 - (1-p)\frac{\lambda}{s+\lambda}} = \frac{\lambda p}{s + \lambda p},$$

οπότε συμπεραίνουμε ότι η S_N ακολουθεί την $\text{Exp}(\lambda p)$.

◊

3.3 Η εκθετική κατανομή

Κεντρικό ρόλο στη Θεωρία Ουρών Αναμονής παίζει η εκθετική κατανομή, η οποία χρησιμοποιείται για τη μοντελοποίηση ενδιάμεσων χρόνων μεταξύ αφίξεων πελατών και χρόνων εξυπηρέτησης.

Η εκθετική κατανομή προκύπτει φυσικά ως η κατανομή χρόνων που έχουν την αμνήμονη ιδιότητα. Μια μη-αρνητική τυχαία μεταβλητή X που περιγράφει το χρόνο για την πραγματοποίηση κάποιου γεγονότος λέγεται ότι έχει την αμνήμονη ιδιότητα εάν οποιαδήποτε χρονική στιγμή s που ο χρόνος αυτός δεν έχει εκπνεύσει, η κατανομή του υπολειπόμενου χρόνου $X - s$ μέχρι την πραγματοποίηση του γεγονότος είναι η ίδια με την αρχική κατανομή. Με μαθηματικούς όρους μια μη-αρνητική τυχαία μεταβλητή X έχει την αμνήμονη ιδιότητα αν για κάθε $s > 0$, ισχύει ότι

$$\Pr[X - s > t | X > s] = \Pr[X > t], \quad t > 0. \quad (3.18)$$

Αποδεικνύεται ότι κάθε συνεχής μη-αρνητική τυχαία μεταβλητή που έχει την αμνήμονη ιδιότητα έχει αναγκαστικά εκθετική κατανομή που δίνεται από τον παρακάτω ορισμό.

Ορισμός 3.17 (Εκθετική κατανομή) Μια συνεχής μη-αρνητική τυχαία μεταβλητή X με συνάρτηση πυκνότητας πιθανότητας

$$f_X(x) = \mu e^{-\mu x}, \quad x \geq 0,$$

λέγεται ότι ακολουθεί την εκθετική κατανομή με παράμετρο (ρυθμό) μ . Η αντίστοιχη συνάρτηση κατανομής είναι η

$$F_X(x) = 1 - e^{-\mu x}, \quad x \geq 0,$$

ο μετασχηματισμός Laplace-Stieltjes είναι ο

$$\tilde{F}_X(s) = \frac{\mu}{s + \mu},$$

η μέση τιμή είναι $E[X] = \frac{1}{\mu}$ και η διασπορά είναι $\text{Var}[X] = \frac{1}{\mu^2}$. Γενικότερα, η n -οστή ροπή είναι $E[X^n] = \frac{n!}{\mu^n}$. Επίσης, ο δεσμευμένος ρυθμός βλάβης της X είναι

$$\lambda_X(x) = \lim_{\delta x \rightarrow 0^+} \frac{\Pr[x < X \leq x + \delta x | X > x]}{\delta x} = \mu, \quad x \geq 0.$$

◊

Η εκθετική κατανομή είναι εξαιρετικά πλούσια σε καλές ιδιότητες και γι αυτό επάγει σημαντική ευκολία στους υπολογισμούς μέτρων απόδοσης στα μοντέλα στα οποία υπεισέρχεται.

Θεώρημα 3.18 (Ιδιότητες εκθετικής κατανομής) Η εκθετική κατανομή έχει τις ακόλουθες ιδιότητες:

1. ((Ασθενής) αμνήμονη ιδιότητα): Αν η X ακολουθεί την εκθετική κατανομή, τότε για κάθε $s > 0$, η δεσμευμένη κατανομή της $X - s$ δεδομένου ότι $X > s$ έχει την ίδια κατανομή με την X : $\Pr[X - s > t | X > s] = \Pr[X > t]$, για $s, t > 0$.
2. (Ισχυρή αμνήμονη ιδιότητα): Αν η X ακολουθεί την εκθετική κατανομή, τότε για κάθε θετική τυχαία μεταβλητή Y που είναι ανεξάρτητη της X , η δεσμευμένη κατανομή της $X - Y$ δεδομένου ότι $X > Y$ έχει την ίδια κατανομή με την X : $\Pr[X - Y > t | X > Y] = \Pr[X > t]$, για $t > 0$.
3. (Ιδιότητα κλίμακας): Αν η X ακολουθεί την εκθετική κατανομή με παράμετρο μ και $a > 0$ τότε η aX ακολουθεί την εκθετική κατανομή με παράμετρο $\frac{\mu}{a}$.

4. (Ιδιότητα ελαχίστου): Αν οι X_1, X_2, \dots, X_n είναι ανεξάρτητες εκθετικές τυχαίες μεταβλητές με παραμέτρους $\mu_1, \mu_2, \dots, \mu_n$ αντίστοιχα τότε η $\min(X_1, X_2, \dots, X_n)$ ακολουθεί την εκθετική κατανομή με παράμετρο $\mu_1 + \mu_2 + \dots + \mu_n$.
5. (Ιδιότητα δείκτη ελαχίστου): Αν οι X_1, X_2, \dots, X_n είναι ανεξάρτητες εκθετικές τυχαίες μεταβλητές με παραμέτρους $\mu_1, \mu_2, \dots, \mu_n$ αντίστοιχα και N είναι η τυχαία μεταβλητή του δείκτη ελαχίστου από αυτές (δηλαδή $N = i$ αν και μόνο αν $X_i = \min(X_1, X_2, \dots, X_n)$) τότε $\Pr[N = i] = \frac{\mu_i}{\mu_1 + \mu_2 + \dots + \mu_n}$.
6. (Ιδιότητα ανεξαρτησίας ελαχίστου και δείκτη ελαχίστου): Αν οι X_1, X_2, \dots, X_n είναι ανεξάρτητες εκθετικές τυχαίες μεταβλητές με παραμέτρους $\mu_1, \mu_2, \dots, \mu_n$ αντίστοιχα και N είναι η τυχαία μεταβλητή του δείκτη ελαχίστου από αυτές (δηλαδή $N = i$ αν και μόνο αν $X_i = \min(X_1, X_2, \dots, X_n)$), τότε οι τυχαίες μεταβλητές $\min(X_1, X_2, \dots, X_n)$ και N είναι ανεξάρτητες.
7. (Ιδιότητα σταθερού αθροίσματος): Αν οι X_1, X_2, \dots, X_n είναι ανεξάρτητες και ισόνομες εκθετικές τυχαίες μεταβλητές με παράμετρο μ , τότε η τυχαία μεταβλητή $S_n = \sum_{i=1}^n X_i$ ακολουθεί την κατανομή Erlang(n, μ) με συνάρτηση πυκνότητας πιθανότητας

$$f_{S_n}(x) = \frac{\mu^n}{(n-1)!} x^{n-1} e^{-\mu x}, \quad x \geq 0. \quad (3.19)$$

Η αντίστοιχη συνάρτηση κατανομής είναι η

$$F_{S_n}(x) = 1 - \sum_{i=0}^{n-1} e^{-\mu x} \frac{(\mu x)^i}{i!}, \quad x \geq 0, \quad (3.20)$$

και ο μετασχηματισμός Laplace-Stieltjes είναι

$$\tilde{F}_{S_n}(s) = \left(\frac{\mu}{s + \mu} \right)^n.$$

8. (Ιδιότητα γεωμετρικού αθροίσματος): Έστω X_1, X_2, \dots ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών με την $\text{Exp}(\mu)$ κατανομή και N , ανεξάρτητη των X_i , $i = 1, 2, \dots$, γεωμετρικά κατανομημένη με συνάρτηση πιθανότητας $\Pr[N = n] = (1-p)^{n-1}p$, $n \geq 1$. Αν $S_N = \sum_{i=1}^N X_i$, τότε η τυχαία μεταβλητή S_N ακολουθεί την κατανομή $\text{Exp}(\mu p)$.
9. (Ιδιότητα διαστημάτων): Έστω X_1, X_2, \dots, X_n ανεξάρτητες και ισόνομες εκθετικές τυχαίες μεταβλητές με παράμετρο μ και έστω $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ οι αντίστοιχες διατεταγμένες τυχαίες μεταβλητές και $D_{i:n}$, $i = 1, 2, \dots, n$, τα διαστήματα μεταξύ αυτών, δηλαδή $D_{1:n} = X_{1:n}$ και $D_{i:n} = X_{i:n} - X_{i-1:n}$, $i = 2, 3, \dots, n$. Τότε οι τυχαίες μεταβλητές $D_{i:n}$, $i = 1, 2, \dots, n$, είναι ανεξάρτητες και η $D_{i:n}$ έχει την κατανομή $\text{Exp}(n-i+1)$.

Αυτές οι ιδιότητες απορρέουν διαισθητικά από τη θεμελιώδη αμνήμονη ιδιότητα. Π.χ. η ιδιότητα κλίμακας είναι φανερή αφού αν ένας χρόνος έχει την αμνήμονη ιδιότητα, η αλλαγή της μονάδας μέτρησής του δεν θα την επηρεάσει. Έτσι είναι λογικό ότι αν η X είναι εκθετική, τότε και η aX θα είναι επίσης εκθετική (π.χ. αν η X είναι ο χρόνος σε ώρες, η $60X$ είναι ο χρόνος σε λεπτά). Επιπλέον, η παράμετρος της aX θα είναι $1/E[aX] = 1/(aE[X])$, οπότε αν η X είναι $\text{Exp}(\mu)$, τότε η aX θα είναι $\text{Exp}(\mu/a)$. Ομοίως, αν οι χρόνοι X_1, X_2, \dots, X_n έχουν όλοι την αμνήμονη ιδιότητα, τότε και ο $\min(X_1, X_2, \dots, X_n)$ θα έχει την αμνήμονη ιδιότητα. Πράγματι, αν κάποια στιγμή s , κανείς από τους χρόνους X_1, X_2, \dots, X_n δεν έχει εκπνεύσει, τότε όλοι οι υπολειπόμενοι χρόνοι θα έχουν την ίδια κατανομή με τους αρχικούς και επομένως ο υπολειπόμενος χρόνος για την εκπνοή του πρώτου από τους X_1, X_2, \dots, X_n θα έχει την ίδια κατανομή με την αρχική. Οπότε έχουμε την ιδιότητα του ελαχίστου κ.ο.κ. Και οι αυστηρές αποδείξεις των ιδιοτήτων είναι σχετικά εύκολες, αλλά περιλαμβάνουν αρκετές πράξεις, οπότε αφήνονται ως ασκήσεις.

Πέρα από την ευκολία στους υπολογισμούς που επάγει η εκθετική κατανομή λόγω των ιδιοτήτων της που παρατέθηκαν παραπάνω, η αξία της έγκειται σε δύο επιπλέον λόγους. Ο πρώτος λόγος είναι ότι η $\text{Exp}(\mu)$ είναι η κατανομή μέγιστης εντροπίας (maximum entropy distribution) στην κλάση των συνεχών κατανομών μη-αρνητικών τυχαίων μεταβλητών με μέση τιμή $\frac{1}{\mu}$. Για μια συνεχή κατανομή με συνάρτηση πυκνότητας πιθανότητας $f_X(x)$, $x \in \mathbb{R}$, η εντροπία ορίζεται ως

$$H(X) = - \int_{-\infty}^{\infty} f_X(x) \log f_X(x) dx.$$

Σύμφωνα με την αρχή της μέγιστης εντροπίας, αν δεν είναι τίποτα γνωστό για μια κατανομή, εκτός από το ότι ανήκει σε κάποια συγκεκριμένη κλάση, τότε η κατανομή μέγιστης εντροπίας της συγκεκριμένης κλάσης πρέπει να επιλεγεί ως η λιγότερο μεροληπτική για την παραπέρα μελέτη ενός φαινομένου. Οπότε, στην περίπτωση που θέλουμε να μοντελοποιήσουμε κάποιον χρόνο εξυπηρέτησης ή ενδιάμεσο χρόνο αφίξεων με γνωστή μέση τιμή, αλλά χωρίς να έχουμε κάποιο άλλο στοιχείο, τότε η επιλογή της εκθετικής με την ίδια μέση τιμή, εισάγει τη λιγότερο μεροληπτική μοντελοποίηση του εξεταζόμενου συστήματος εξυπηρέτησης. Ο δεύτερος λόγος είναι ότι η εκθετική κατανομή μπορεί να χρησιμοποιηθεί ως στοιχείο για την κατασκευή πιο σύνθετων κατανομών μη-αρνητικών τυχαίων μεταβλητών. Συγκεκριμένα, κάθε συνεχής μη-αρνητική τυχαία μεταβλητή μπορεί να προσεγγιστεί οσοδήποτε κοντά θέλουμε από μια μίξη αθροισμάτων εκθετικών τυχαίων μεταβλητών, και μάλιστα με ίδια παράμετρο. Η σχετική μέθοδος αναφέρεται ως μέθοδος των φάσεων και αναπτύσσεται σε επόμενο κεφάλαιο του βιβλίου στο πλαίσιο της μελέτης πιο προχωρημένων συστημάτων εξυπηρέτησης.

3.4 Ασκήσεις

Άσκηση 3.1 Σε ένα τηλεοπτικό παιχνίδι, ο παίκτης έχει την ευκαιρία να ανοίξει έναν από 3 φακέλους που του παρουσιάζονται. Ο πρώτος από αυτούς περιέχει 3 ευρώ, τα οποία παίρνει και αποχωρεί από το παιχνίδι. Ο δεύτερος έχει 5 ευρώ και του δίνει το δικαίωμα να επαναλάβει τη διαδικασία (πάλι με 3 φακέλους όπως οι αρχικοί). Ο τρίτος έχει 7 ευρώ και του δίνει το δικαίωμα να επαναλάβει τη διαδικασία (πάλι με 3 φακέλους όπως οι αρχικοί). Αν ο παίκτης δεν γνωρίζει σε κάθε επανάληψη της διαδικασίας το περιεχόμενο των φακέλων και διαλέγει στην τύχη, τότε ποιο είναι το αναμενόμενο συνολικό κέρδος του;

Άσκηση 3.2 Μια κάλπη έχει n κόκκινα και m πράσινα σφαιρίδια. Επιλέγεται στην τύχη ένα κάθε φορά και αφαιρείται από την κάλπη μέχρι που να μείνουν μόνο σφαιρίδια ενός χρώματος. Έστω $T_{n,m}$ ο αριθμός των σφαιριδίων που έχουν μείνει στο τέλος αυτής της διαδικασίας. Διατυπώστε ένα αναδρομικό σχήμα υπολογισμού για τη «διπλή ακολουθία» $\mu_{n,m} = E[T_{n,m}]$ και χρησιμοποιήστε το για να βρείτε το $\mu_{3,4}$.

Άσκηση 3.3 Επιλέγουμε ένα νόμισμα, εκτελούμε n ανεξάρτητες ρίψεις του και έστω X ο αριθμός των ρίψεων στις οποίες θα εμφανιστούν «γράμματα». Αν η πιθανότητα να φέρνει «γράμματα» το νόμισμα είναι τυχαία μεταβλητή με ομοιόμορφη κατανομή στο $(0, 1)$, να βρεθεί η συνάρτηση πιθανότητας της X και η μέση τιμή της.

Άσκηση 3.4 Μια κάλπη περιέχει a λευκά και b μαύρα σφαιρίδια. Αφού τραβήξουμε ένα σφαιρίδιο, το επανατοποθετούμε στην κάλπη αν είναι λευκό. Αν, όμως, είναι μαύρο, τότε βάζουμε στη θέση του ένα λευκό (από κάποια άλλη κάλπη). Έστω X_n ο αριθμός των λευκών σφαιριδίων στην κάλπη, αφού η διαδικασία έχει επαναληφθεί n φορές.

1. Αποδείξτε, χρησιμοποιώντας το Θεώρημα Διπλής Μέσης Τιμής, ότι

$$E[X_{n+1}] = \left(1 - \frac{1}{a+b}\right) E[X_n] + 1.$$

2. Βρείτε έναν κλειστό τύπο για την $E[X_n]$.

Άσκηση 3.5 Κάθε φορά που ρίπτεται ένα νόμισμα, προσγειώνεται ως «κεφαλή» με πιθανότητα p και ως «γράμματα» με πιθανότητα $1 - p$. Να υπολογιστεί ο μέσος αριθμός ρίψεων μέχρι να παρατηρηθεί μια σειρά από r συνεχόμενες κεφαλές. Υπόδειξη: Δεσμεύστε στον αριθμό των ρίψεων μέχρι τα πρώτα γράμματα ή στον αριθμό των ρίψεων μέχρι να παρατηρηθούν $r - 1$ συνεχόμενες κεφαλές και εφαρμόστε το Θεώρημα Διπλής Μέσης Τιμής.

Άσκηση 3.6 Έστω X μια μη-αρνητική ακέραια τυχαία μεταβλητή με πιθανογεννήτρια

$$P_X(z) = \frac{c}{6 - z - z^2}.$$

1. Να προσδιοριστεί η σταθερά c .
2. Να υπολογιστεί η μέση τιμή $E[X]$.
3. Να υπολογιστεί η συνάρτηση πιθανότητας $f_X(x) = \Pr[X = x]$, $x \geq 0$.
4. Να υπολογιστεί η πιθανότητα $\Pr[X \text{ είναι άρτιος}]$.

Άσκηση 3.7 Έστω X μια μη-αρνητική ακέραια τυχαία μεταβλητή με πιθανογεννήτρια

$$P_X(z) = \frac{c - 5z}{2z^2 - 12z + 16}.$$

1. Να προσδιοριστεί η σταθερά c .
2. Να υπολογιστεί η μέση τιμή $E[X]$.
3. Να υπολογιστεί η συνάρτηση πιθανότητας $f_X(x) = \Pr[X = x]$, $x \geq 0$.

Άσκηση 3.8 Έστω X μια μη-αρνητική ακέραια τυχαία μεταβλητή με πιθανογεννήτρια

$$P_X(z) = \frac{cz^2 - 31z + 36}{6(2 - z)^2(3 - z)}.$$

1. Να προσδιοριστεί η σταθερά c .
2. Να υπολογιστεί η μέση τιμή $E[X]$.
3. Να υπολογιστεί η συνάρτηση πιθανότητας $f_X(x) = \Pr[X = x]$, $x \geq 0$.

Άσκηση 3.9 Έστω X, Y και Z ανεξάρτητες εκθετικές τυχαίες μεταβλητές με ρυθμούς λ, λ και μ αντίστοιχα, με $\lambda \neq \mu$. Να βρεθεί ο μετασχηματισμός Laplace-Stieltjes της $X + Y + Z$ και η συνάρτηση πυκνότητας πιθανότητάς της.

Άσκηση 3.10 Έστω X μια μη-αρνητική τυχαία μεταβλητή με μετασχηματισμό Laplace-Stieltjes

$$\tilde{F}_X(s) = \frac{10s + c}{3s^2 + 30s + 75}.$$

1. Να προσδιοριστεί η σταθερά c .
2. Να υπολογιστεί η μέση τιμή $E[X]$.
3. Να υπολογιστεί η συνάρτηση κατανομής $F_X(x)$ της X .

Άσκηση 3.11 Έστω X, Y μη-αρνητικές τυχαίες μεταβλητές με μετασχηματισμούς Laplace-Stieltjes

$$\tilde{F}_X(s) = e^{-2s}, \quad \tilde{F}_Y(s) = \frac{4e^{-2s} + 6e^{-6s}}{10}.$$

Να υπολογιστούν οι συναρτήσεις κατανομής $F_X(x)$ και $F_Y(x)$.

Άσκηση 3.12 Έστω X μια μη-αρνητική τυχαία μεταβλητή με μετασχηματισμό Laplace-Stieltjes

$$\tilde{F}_X(s) = \frac{1}{3}e^{-2s} + \frac{c}{3s + 15}.$$

1. Να προσδιοριστεί η σταθερά c .
2. Να υπολογιστεί η μέση τιμή $E[X]$.
3. Να υπολογιστεί η συνάρτηση κατανομής $F_X(x)$ της X .

Άσκηση 3.13 Να αποδείξετε ότι αν μια συνεχής μη-αρνητική τυχαία μεταβλητή έχει την αμνήμονη ιδιότητα, τότε έχει αναγκαστικά την εκθετική κατανομή. Υπόδειξη: Θέστε $g(t) = \Pr[X > t]$ και κατόπιν αποδείξτε ότι:

1. $g(t + s) = g(t)g(s), \quad t, s > 0,$
2. $g\left(\frac{1}{n}\right) = (g(1))^{1/n}, \quad n \geq 1,$
3. $g\left(\frac{m}{n}\right) = (g(1))^{m/n}, \quad m, n \geq 1,$
4. $g(x) = (g(1))^x, \quad x > 0.$

Άσκηση 3.14 Να αποδείξετε τις ιδιότητες της εκθετικής κατανομής που διατυπώνονται στο θεώρημα 3.18.

Άσκηση 3.15 Ένα σύστημα έχει δυο εξαρτήματα και εκδηλώνει βλάβη μόλις κάποιο από τα εξαρτήματα χαλάσει. Ο χρόνος ζωής του πρώτου εξαρτήματος ακολουθεί την εκθετική κατανομή $\text{Exp}(\lambda)$ (δηλαδή έχει συνάρτηση πυκνότητας πιθανότητας $f_1(t) = \lambda e^{-\lambda t}, \quad t \geq 0$), ενώ ο χρόνος ζωής του δεύτερου εξαρτήματος ακολουθεί την κατανομή $\text{Gamma}(n, \mu)$ (δηλαδή έχει συνάρτηση πυκνότητας πιθανότητας $f_2(t) = \mu^n t^{n-1} e^{-\mu t} / (n-1)!, \quad t \geq 0$). Υποθέτοντας ότι οι χρόνοι ζωής των εξαρτημάτων είναι ανεξάρτητες τυχαίες μεταβλητές, να βρεθεί ο μέσος χρόνος ζωής του συστήματος των δυο εξαρτημάτων (δηλαδή ο μέσος χρόνος μέχρι να εκδηλώσει το σύστημα βλάβη).

3.5 Σχόλια

Σχεδόν όλα τα εισαγωγικά βιβλίων Θεωρίας Πιθανοτήτων αναφέρονται στα θεμελιώδη θέματα που περιγράψαμε σε αυτό το κεφάλαιο. Η θεωρία για τη δεσμευμένη μέση τιμή μαζί με πολλά παραδείγματα παρουσιάζονται με τρόπο εύληπτο στα συγγράμματα των Ross 2019, Bertsekas και Tsitsiklis 2008 και Grimmett και Stirzaker 2020.

Οι πιθανογεννήτριες συναρτήσεις και οι μετασχηματισμοί Laplace-Stieltjes χρησιμοποιούνται ευρύτατα για τη μελέτη στοχαστικών διαδικασιών, ιδιαίτερα Μαρκοβιανών αλυσίδων διακριτού και συνεχούς χρόνου, καθώς και ανανεωτικών διαδικασιών. Τα βιβλία Adan και Resing 2001, Kulkarni 2010 και Tijms 2008 περιέχουν πολλές πληροφορίες και υπολογιστικές τεχνικές σχετικά με το θέμα αυτό. Το θέμα της αντιστροφής των πιθανογεννητριών και των μετασχηματισμών Laplace-Stieltjes, δηλαδή της εύρεσης της αντίστοιχης κατανομής (ή των συναρτήσεων μάζας πιθανότητας ή πυκνότητας πιθανότητας) είναι ιδιαίτερα σημαντικό. Για το θέμα αυτό (όσον αφορά τις πιθανογεννήτριες) μια καλή πηγή είναι το βιβλίο του Wilf 2005, καθώς και η μεταπτυχιακή διπλωματική εργασία της Κάντα 2006. Όπου δεν είναι δυνατή η αναλυτική αντιστροφή, θα

πρέπει η αντιστροφή να γίνει αριθμητικά. Για το σκοπό αυτό υπάρχουν ιδιαίτερα αποτελεσματικοί αλγόριθμοι. Στο βιβλίο του Kao 1997 δίνονται οι υλοποιήσεις σε Matlab δυο βασικών τέτοιων αλγορίθμων (δείτε το Appendix του πρώτου κεφαλαίου) που βασίζονται στις εργασίες Abate και Whitt 1992 και Abate και Whitt 1995 για την αριθμητική αντιστροφή πιθανογεννητριών και μετασχηματισμών Laplace. Για περισσότερες λεπτομέρειες, μπορεί κανείς να δει την εργασία Abate, Choudhury και Whitt 2000, όπου παρουσιάζεται αναλυτικά το θέμα της αντιστροφής μετασχηματισμών (πιθανογεννητριών, Laplace κλπ.) μαζί με εφαρμογές σε στοχαστικά μοντέλα.

Η εκθετική κατανομή που είναι βασική για τη μελέτη όλων των στοχαστικών μοντέλων της Επιχειρησιακής Έρευνας παρουσιάζεται με αρκετή λεπτομέρεια σε πολλά βιβλία πιθανοτήτων και στοχαστικών διαδικασιών. Το βιβλίο Kulkarni 2010 παρουσιάζει λεπτομερείς αποδείξεις για τις ιδιότητες της εκθετικής κατανομής που παρουσιάσαμε σε αυτό το κεφάλαιο.

Βιβλιογραφία

- [1] S. Ross. *Introduction to Probability Models, 12th Edition*. Academic Publishers, 2019. ISBN: 978-0128143469.
- [2] D.P. Bertsekas και J.N. Tsitsiklis. *Introduction to Probability, 2nd Edition*. Athena Scientific, 2008. ISBN: 978-1886529236.
- [3] G. Grimmett και D. Stirzaker. *Probability and Random Processes, 4th Edition*. Oxford, 2020. ISBN: 978-0198847595.
- [4] I. Adan και J. Resing. *Queueing Theory*. Eindhoven: Notes available online, 2001.
- [5] V.G. Kulkarni. *Modeling and Analysis of Stochastic Systems, 2nd Edition*. CRC Press, 2010. ISBN: 978-1439808757.
- [6] H.C. Tijms. *A First Course in Stochastic Models, 2nd Edition*. Chichester: Wiley, 2008. ISBN: 978-0471498803.
- [7] H.S. Wilf. *Generatingfunctionology, 3rd Edition*. AK Peters/CRC, 2005. ISBN: 978-1568812793.
- [8] Σ. Κάντα. *Αλγοριθμικές Τεχνικές στις Μαρκοβιανές Αλυσίδες: Μέθοδοι Πιθανογεννητριών και Εφαρμογές*. Αθήνα: Διπλωματική Εργασία, Μεταπτυχιακό Πρόγραμμα Ειδικευσης στη Στατιστική και Επιχειρησιακή Έρευνα, Τμήμα Μαθηματικών Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών, 2006.
- [9] E.P.C. Kao. *An Introduction to Stochastic Processes*. New York: Duxbury Press, 1997. ISBN: 978-0534255183.
- [10] J. Abate και W. Whitt. “Numerical inversion of probability generating functions”. Στο: *Operations Research Letters* 12 (1992), σσ. 275–281.
- [11] J. Abate και W. Whitt. “Numerical inversion of Laplace transforms of probability distributions”. Στο: *ORSA Journal on Computing* 7 (1995), σσ. 36–43.
- [12] J. Abate, G.L. Choudhury και W. Whitt. “An introduction to numerical transform inversion and its application to probability models”. Στο: *Grassmann, W.K. (ed.) Computational Probability*. Kluwer, 2000. Κεφ. 8, σσ. 257–323. ISBN: 978-0792386179. DOI: https://doi.org/10.1007/978-1-4757-4828-4_8.

Μέρος II

ΑΠΟΤΙΜΗΣΗ ΑΠΟΔΟΣΗΣ: ΜΑΡΚΟΒΙΑΝΑ
ΜΟΝΤΕΛΑ

ΚΕΦΑΛΑΙΟ 4

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΓΙΑ ΜΑΡΚΟΒΙΑΝΑ ΜΟΝΤΕΛΑ

Στο κεφάλαιο αυτό γίνεται μια σύντομη παρουσίαση της θεωρίας των Μαρκοβιανών αλυσίδων συνεχούς χρόνου που χρησιμοποιούνται στη Θεωρία Ουρών Αναμονής. Η σχετική θεωρία αναπτύσσεται με έμφαση στη διατύπωση των ορισμών και των αποτελεσμάτων και στη διαισθητική ερμηνεία τους. Για τις αυστηρές μαθηματικές αποδείξεις των αποτελεσμάτων, ο ενδιαφερόμενος αναγνώστης θα πρέπει να συμβουλευθεί κάποιο βιβλίο με θέμα τις στοχαστικές διαδικασίες (ανεπίξει), π.χ. Ross 1995, Kao 1997, Tijms 2008, Kulkarni 2010 ή Grimmett και Stirzaker 2020.

4.1 Μαρκοβιανές αλυσίδες συνεχούς χρόνου: Ορισμός και μεταβατική συμπεριφορά

Για την περιγραφή της κατάστασης ενός στοχαστικού δυναμικού συστήματος με διακριτό χώρο καταστάσεων σε συνεχή χρόνο, το απλούστερο μοντέλο είναι μια Μαρκοβιανή αλυσίδα συνεχούς χρόνου. Επομένως, στα πλαίσια μιας εισαγωγής στη Θεωρία Ουρών Αναμονής, οι Μαρκοβιανές αλυσίδες συνεχούς χρόνου αποτελούν την κύρια κλάση στοχαστικών διαδικασιών που προτυποποιούν ουρές αναμονής.

Ορισμός 4.1 (Μαρκοβιανή αλυσίδα συνεχούς χρόνου) Μια στοχαστική διαδικασία $\{X(t) : t \geq 0\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου με χώρο καταστάσεων \mathcal{S} αν

- (i) $X(t) \in \mathcal{S}, t \geq 0$, και \mathcal{S} αριθμησιμο,
- (ii) $\Pr[X(t+s) = j | X(u), 0 \leq u < s, X(s) = i] = \Pr[X(t+s) = j | X(s) = i], t, s \geq 0, i, j \in \mathcal{S}$ (Μαρκοβιανή ιδιότητα).

Η πιθανότητα $p_{ij}(s, s+t) = \Pr[X(t+s) = j | X(s) = i]$ αναφέρεται ως πιθανότητα μετάβασης από την i στη j σε χρόνο t τη στιγμή s . Αν δεν εξαρτάται από το s η Μαρκοβιανή αλυσίδα αναφέρεται ως ομογενής. Τότε συμβολίζουμε

την $p_{ij}(s, s+t)$ με $p_{ij}(t)$. Ο πίνακας

$$\mathbf{P}(t) = (p_{ij}(t)) = \begin{pmatrix} p_{00}(t) & p_{01}(t) & p_{02}(t) & \cdots \\ p_{10}(t) & p_{11}(t) & p_{12}(t) & \cdots \\ p_{20}(t) & p_{21}(t) & p_{22}(t) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

αναφέρεται ως πίνακας πιθανοτήτων μετάβασης της Μαρκοβιανής αλυσίδας (έχοντας υποθέσει, δίχως βλάβη της γενικότητας, ότι $\mathcal{S} = \{0, 1, 2, \dots\}$). Είναι φανερό ότι $\mathbf{P}(0) = \mathbf{I}$, αφού

$$p_{ij}(0) = \begin{cases} 1 & \text{αν } i = j, \\ 0 & \text{αν } i \neq j. \end{cases}$$

Η συνάρτηση πιθανότητας $\mathbf{p}(0) = (p_i(0) : i \in \mathcal{S})$ με $p_i(0) = \Pr[X(0) = i]$, $i \in \mathcal{S}$, αναφέρεται ως αρχική κατανομή της Μαρκοβιανής αλυσίδας. \diamond

Η συνάρτηση πιθανότητας $\mathbf{p}(0)$ θεωρείται διάνυσμα-στήλη και αυτός είναι ο κανόνας και για όλα τα άλλα διανύσματα που εμφανίζονται στο παρόν σύγγραμμα. Όταν χρειάζεται να εμφανιστούν ως διανύσματα-γραμμές χρησιμοποιούμε τον ανάστροφο. Π.χ. το διάνυσμα-γραμμή της αρχικής κατανομής της Μαρκοβιανής αλυσίδας συμβολίζεται ως $(\mathbf{p}(0))^T$.

Η Μαρκοβιανή ιδιότητα εκφράζει την απαίτηση το παρελθόν και το μέλλον της διαδικασίας να είναι ανεξάρτητα, δοθέντος του παρόντος. Αυτό σημαίνει ότι η κατάσταση μιας Μαρκοβιανής διαδικασίας περιέχει όλη την πληροφορία που χρειάζεται για να προβλεφθεί πιθανοθεωρητικά η μελλοντική εξέλιξή της, χωρίς να χρειάζεται επιπλέον πληροφορία για το πώς έφθασε μέχρι την παρούσα κατάσταση.

Σε ό,τι ακολουθεί θα περιοριστούμε σε ομογενείς Μαρκοβιανές αλυσίδες. Η $\{X(t)\}$ είναι πλήρως καθορισμένη, αν δίνονται η αρχική κατανομή της $(p_i(0) : i \in \mathcal{S})$ με

$$p_i(0) = \Pr[X(0) = i], \quad i \in \mathcal{S},$$

και οι πίνακες πιθανοτήτων μετάβασης. Ειδικότερα, θα δούμε ότι μπορούμε να προσδιορίσουμε τις πιθανότητες $p_j(t) = \Pr[X(t) = j]$. Η συνάρτηση πιθανότητας $\mathbf{p}(t) = (p_i(t) : i \in \mathcal{S})$ με $p_i(t) = \Pr[X(t) = i]$ αναφέρεται ως η μεταβατική κατανομή της Μαρκοβιανής αλυσίδας τη στιγμή t .

Το πρόβλημα που υπάρχει είναι ότι χρειαζόμαστε άπειρους πίνακες μετάβασης, τους $\mathbf{P}(t) = (p_{ij}(t))$, έναν για κάθε $t \geq 0$. Για να υπολογιστούν οι πίνακες αυτοί χρειαζόμαστε την έννοια των ρυθμών μετάβασης που περιγράφουν την «τοπική συμπεριφορά» μιας Μαρκοβιανής αλυσίδας συνεχούς χρόνου για «μικρά» t . Για μια Μαρκοβιανή αλυσίδα συνεχούς χρόνου αποδεικνύεται ότι υπάρχουν τα όρια

$$q_{ij} = \lim_{h \rightarrow 0^+} \frac{p_{ij}(h)}{h} = p'_{ij}(0), \quad i, j \in \mathcal{S} \text{ με } i \neq j,$$

$$q_i = \lim_{h \rightarrow 0^+} \frac{1 - p_{ii}(h)}{h} = -p'_{ii}(0), \quad i \in \mathcal{S},$$

που αναφέρονται αντίστοιχα ως ο ρυθμός μετάβασης από την κατάσταση i στην κατάσταση j και ως ο ρυθμός εξόδου από την κατάσταση i . Προφανώς ισχύει

$$q_i = \sum_{j \neq i} q_{ij}, \quad i \in \mathcal{S},$$

αφού $\sum_j p_{ij}(h) = 1$. Εναλλακτικά, μπορούμε να σκεφτόμαστε ότι

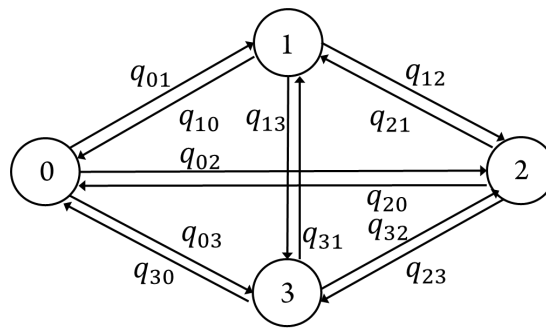
$$p_{ij}(h) = q_{ij}h + o(h), \quad h \rightarrow 0^+, \quad j \neq i,$$

$$p_{ii}(h) = 1 - q_i h + o(h), \quad h \rightarrow 0^+,$$

όπου $o(h)$ κάποια συνάρτηση του h τέτοια ώστε $o(h)/h \rightarrow 0$, καθώς $h \rightarrow 0^+$. Επομένως οι ρυθμοί q_{ij} πράγματι καθορίζουν την «τοπική συμπεριφορά» της $\{X(t)\}$, από την οποία μπορούν να συναχθούν οι $p_{ij}(t)$ για οποιοδήποτε t . Οπότε, μια Μαρκοβιανή αλυσίδα συνεχούς χρόνου μπορεί να περιγραφεί από την αρχική κατανομή της και τον πίνακα ρυθμών μετάβασης

$$Q = (p'_{ij}(0)) = (q_{ij}) = \begin{pmatrix} -q_0 & q_{01} & q_{02} & \cdots \\ q_{10} & -q_1 & q_{12} & \cdots \\ q_{20} & q_{21} & -q_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Ο πίνακας αυτός αναφέρεται και ως απειροστικός ή απειροστός γεννήτορας της Μαρκοβιανής αλυσίδας. Τα εκτός διαγωνίου στοιχεία του είναι μη-αρνητικά, ενώ τα διαγώνια είναι μη-θετικά (ένα διαγώνιο στοιχείο q_{ii} είναι 0 μόνο αν η κατάσταση i είναι απορροφητική, δηλαδή μετά από μια είσοδο σε αυτήν η αλυσίδα παραμένει για πάντα εκεί). Επιπλέον, τα αθροίσματα των γραμμών του πίνακα είναι όλα 0, αφού $\sum_j p_{ij}(t) = 1$. Ο πίνακας Q πολλές φορές παριστάνεται γραφικά με το λεγόμενο διάγραμμα ρυθμών μετάβασης. Το διάγραμμα είναι ένα γράφημα, όπου οι κορυφές (κόμβοι) είναι οι καταστάσεις της Μαρκοβιανής αλυσίδας και οι διατεταγμένες ακμές (βέλη) αντιστοιχούν στις μεταβάσεις. Κάθε διατεταγμένη ακμή έχει μια ετικέτα που δίνει τον ρυθμό της αντίστοιχης μετάβασης. Για παράδειγμα το διάγραμμα ρυθμών μετάβασης μιας Μαρκοβιανής αλυσίδας με 4 καταστάσεις 0, 1, 2 και 3 δίνεται στο σχήμα 4.1.



Σχήμα 4.1: Διάγραμμα ρυθμών μετάβασης μιας Μαρκοβιανής αλυσίδας με χώρο καταστάσεων $\{0, 1, 2, 3\}$.

Σχετικά με τη μεταβατική συμπεριφορά των Μαρκοβιανών αλυσίδων συνεχούς χρόνου, δηλαδή τη συμπεριφορά τους για ένα πεπερασμένο χρόνο t μετά την έναρξη της παρατήρησής τους, έχουμε το ακόλουθο αποτέλεσμα.

Θεώρημα 4.2 (Βασικοί υπολογισμοί σε πεπερασμένη χρονική στιγμή) Έστω Μαρκοβιανή αλυσίδα συνεχούς χρόνου $\{X(t) : t \geq 0\}$, με χώρο καταστάσεων \mathcal{S} , αρχική κατανομή $\mathbf{p}(0) = (p_i(0))$ και πίνακα ρυθμών μετάβασης $Q = (p'_{ij}(0)) = (q_{ij})$. Έστω επίσης $\mathbf{P}(t) = (p_{ij}(t))$. Τότε έχουμε:

- (i) Η πιθανότητα μιας συγκεκριμένης διαδοχής καταστάσεων σε συγκεκριμένες χρονικές στιγμές $0 < t_1 < t_2 < \dots < t_n$ είναι:

$$\begin{aligned} & \Pr[X(0) = i_0, X(t_1) = i_1, X(t_2) = i_2, \dots, X(t_n) = i_n] \\ &= p_{i_0}(0)p_{i_0i_1}(t_1)p_{i_1i_2}(t_2 - t_1) \cdots p_{i_{n-1}i_n}(t_n - t_{n-1}). \end{aligned}$$

- (ii) Οι πιθανότητες μετάβασης t -οστής τάξης ικανοποιούν τις σχέσεις

$$\begin{aligned} p_{ij}(0) &= \delta_{ij}, \\ p_{ij}(t) &= \sum_{r \in \mathcal{S}} p_{ir}(s)p_{rj}(t-s), \quad 0 \leq s \leq t, \end{aligned}$$

που γράφονται σε πίνακική μορφή

$$\begin{aligned}\mathbf{P}(0) &= \mathbf{I}, \\ \mathbf{P}(t) &= \mathbf{P}(s)\mathbf{P}(t-s), \quad 0 \leq s \leq t.\end{aligned}$$

(iii) Οι πιθανότητες μετάβασης t -οστής τάξης είναι η λύση του συστήματος (προδρομικών) εξισώσεων *Charman - Kolmogorov*:

$$\begin{aligned}p_{ij}(0) &= \delta_{ij}, \\ p'_{ij}(t) &= -p_{ij}(t)q_j + \sum_{k \neq j} p_{ik}(t)q_{kj}, \quad t \geq 0,\end{aligned}\tag{4.1}$$

που γράφονται σε πίνακική μορφή

$$\begin{aligned}\mathbf{P}(0) &= \mathbf{I}, \\ \mathbf{P}'(t) &= \mathbf{P}(t)\mathbf{Q}, \quad t \geq 0.\end{aligned}$$

Επομένως,

$$\mathbf{P}(t) = e^{\mathbf{Q}t} = \sum_{n=0}^{\infty} \mathbf{Q}^n \frac{t^n}{n!}, \quad t \geq 0.$$

(iv) Οι μεταβατικές πιθανότητες $p_j(t) = \Pr[X(t) = j]$, $j \in \mathcal{S}$, υπολογίζονται από τη σχέση

$$p_j(t) = \sum_{i \in \mathcal{S}} p_i(0)p_{ij}(t),$$

και επομένως ικανοποιούν το σύστημα των εξισώσεων *Charman - Kolmogorov*

$$p'_j(t) = -p_j(t)q_j + \sum_{k \neq j} p_k(t)q_{kj}, \quad t \geq 0,\tag{4.2}$$

που γράφεται σε πίνακική μορφή

$$(\mathbf{p}'(t))^T = (\mathbf{p}(t))^T \mathbf{Q}, \quad t \geq 0.$$

Επομένως,

$$(\mathbf{p}(t))^T = (\mathbf{p}(0))^T \mathbf{P}(t) = (\mathbf{p}(0))^T e^{\mathbf{Q}t}, \quad t \geq 0.$$

◇

Το (i) αποδεικνύεται εύκολα χρησιμοποιώντας τον πολλαπλασιαστικό τύπο για τον υπολογισμό της πιθανότητας τομής ενδεχομένων και εφαρμόζοντας τη Μαρκοβιανή ιδιότητα. Η σχέση για το $p_{ij}(0)$ στο (ii) είναι προφανής, ενώ η σχέση για το $p_{ij}(t)$, προκύπτει με δέσμευση στην τιμή r της $X(s)$ για κάποιο $s \in (0, t]$ και εφαρμογή του θεωρήματος ολικής πιθανότητας. Όσον αφορά την (iii), χρησιμοποιούμε την (ii) και κοιτάμε την εξέλιξη της Μαρκοβιανής αλυσίδας στο διάστημα $(0, t+h]$, για $h \rightarrow 0^+$. Έχουμε

$$\begin{aligned}p_{ij}(t+h) &= \sum_k p_{ik}(t)p_{kj}(h) \\ &= p_{ij}(t)(1 - q_j h) + \sum_{k \neq j} p_{ik}(t)q_{kj}h + o(h), \quad h \rightarrow 0^+\end{aligned}$$

οπότε

$$\frac{p_{ij}(t+h) - p_{ij}(t)}{h} = -p_{ij}(t)q_j + \sum_{k \neq j} p_{ik}(t)q_{kj} + \frac{o(h)}{h}, \quad h \rightarrow 0^+.$$

Παίρνοντας $h \rightarrow 0^+$ συνάγουμε τις διαφορικές εξισώσεις (4.1). Η πινακική μορφή τους είναι άμεση και η λύση τους προκύπτει από τη θεωρία των γραμμικών διαφορικών εξισώσεων με σταθερούς συντελεστές (δείτε π.χ., Elaydi 2006). Αν εξετάσουμε την εξέλιξη της Μαρκοβιανής αλυσίδας στο διάστημα $(0, t+h]$ για $h \rightarrow 0^+$, αλλά εφαρμόσουμε το θεώρημα ολικής πιθανότητας κοιτώντας την κατάσταση τη στιγμή h τότε προκύπτει αρχικά η εξίσωση

$$p_{ij}(t+h) = \sum_k p_{ik}(h)p_{kj}(t).$$

Προχωρώντας παρόμοια τους υπολογισμούς όπως πιο πάνω προκύπτουν κάποιες διαφορικές εξισώσεις, δυϊκές των (4.1) που αναφέρονται ως οπισθοδρομικές εξισώσεις Charman-Kolmogorov. Η πινακική τους μορφή είναι

$$\begin{aligned} \mathbf{P}(0) &= \mathbf{I}, \\ \mathbf{P}'(t) &= \mathbf{QP}(t), \quad t \geq 0, \end{aligned}$$

και καταλήγουν και πάλι στην ίδια λύση με τις προδρομικές εξισώσεις (4.1).

Για το (iv) χρησιμοποιούμε το θεώρημα ολικής πιθανότητας δεσμεύοντας στην τιμή i της $X(0)$. Με τον ίδιο τρόπο προκύπτουν οι εξισώσεις Charman - Kolmogorov για τις μεταβατικές πιθανότητες, ξεκινώντας από τις αντίστοιχες εξισώσεις για τις πιθανότητες μετάβασης t -οστής τάξης και εφαρμόζοντας το θεώρημα ολικής πιθανότητας.

Μπορούμε, τώρα, να προχωρήσουμε στη μελέτη των χρόνων παραμονής μιας Μαρκοβιανής αλυσίδας συνεχούς χρόνου στις διάφορες καταστάσεις. Αυτή θα μας οδηγήσει σε κριτήρια, εναλλακτικά του ορισμού, για το πότε μια στοχαστική διαδικασία είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου και επιπλέον θα χρειαστεί για τον υπολογισμό πιθανοτήτων και μέσων χρόνων απορρόφησης σε σύνολα καταστάσεων.

Θεώρημα 4.3 (Χρόνος παραμονής σε κατάσταση Μαρκοβιανής αλυσίδας συνεχούς χρόνου) Έστω Μαρκοβιανή αλυσίδα συνεχούς χρόνου $\{X(t) : t \geq 0\}$, με πίνακα ρυθμών μετάβασης $\mathbf{Q} = (q_{ij})$. Ο χρόνος παραμονής T_i σε μια κατάσταση i πριν από κάποια μετάβαση σε κατάσταση $j \neq i$ είναι εκθετικός με παράμετρο q_i . \diamond

Πράγματι έχουμε ότι η δεσμευμένη πιθανότητα

$$\begin{aligned} \Pr[T_i > t+h | T_i > t] &= \Pr[X(t+h) = i | X(t) = i] + o(h) \\ &= p_{ii}(h) + o(h) = 1 - q_i h + o(h), \quad h \rightarrow 0^+ \end{aligned}$$

δεν εξαρτάται από το t , οπότε η T_i έχει την αμνήμονη ιδιότητα και άρα έχει εκθετική κατανομή. Έστω $F_{T_i}(t)$ η συνάρτηση κατανομής και $f_{T_i}(t) = F'_{T_i}(t)$ η συνάρτηση πυκνότητας πιθανότητας της T_i . Τότε, η παραπάνω σχέση δίνει ότι

$$\begin{aligned} \Pr[t < T_i \leq t+h | T_i > t] &= q_i h + o(h), \quad h \rightarrow 0^+ \\ \Rightarrow \frac{\Pr[t < T_i \leq t+h]}{\Pr[T_i > t]} &= q_i h + o(h), \quad h \rightarrow 0^+ \\ \Rightarrow \lim_{h \rightarrow 0^+} \frac{\Pr[t < T_i \leq t+h]}{h \Pr[T_i > t]} &= q_i \\ \Rightarrow \frac{f_{T_i}(t)}{1 - F_{T_i}(t)} &= -\frac{d}{dt}(1 - F_{T_i}(t)) = q_i \\ \Rightarrow \frac{d}{dt}(\log(1 - F_{T_i}(t))) &= -q_i \end{aligned}$$

Επομένως, έχουμε ότι $\log(1 - F_{T_i}(t)) = -q_i t$ που δίνει $F_{T_i}(t) = 1 - e^{-q_i t}$, $t \geq 0$, δηλαδή η T_i είναι $\text{Exp}(q_i)$.

Όταν η $\{X(t)\}$ φύγει από κάποια κατάσταση i , η επόμενη κατάσταση θα είναι η $j \neq i$ με πιθανότητα $p_{ij} = \frac{q_{ij}}{q_i}$. Πράγματι, για την πιθανότητα αυτή έχουμε

$$\begin{aligned} & \lim_{h \rightarrow 0^+} \Pr[X(t+h) = j | X(t) = i, X(t+h) \neq i] \\ &= \lim_{h \rightarrow 0^+} \frac{\Pr[X(t) = i, X(t+h) = j]}{\Pr[X(t) = i, X(t+h) \neq i]} = \lim_{h \rightarrow 0^+} \frac{\Pr[X(t+h) = j | X(t) = i]}{\Pr[X(t+h) \neq i | X(t) = i]} \\ &= \lim_{h \rightarrow 0^+} \frac{q_{ij}h + o(h)}{q_i h + o(h)} = \frac{q_{ij}}{q_i}, \quad j \neq i. \end{aligned}$$

Επομένως, ένας εναλλακτικός τρόπος για να σκεφτόμαστε την εξέλιξη μιας Μαρκοβιανής αλυσίδας συνεχούς χρόνου, με βάση ρυθμούς είναι ότι, όντας σε μια κατάσταση i , μένει σε αυτήν για εκθετικό χρόνο q_i και κατόπιν μεταβαίνει σε κάποια κατάσταση $j \neq i$ με πιθανότητα $\frac{q_{ij}}{q_i}$.

Από τις ιδιότητες εκθετικής κατανομής (βλέπε θεώρημα 3.18), γνωρίζουμε ότι αν T_1, T_2, \dots, T_k είναι ανεξάρτητες εκθετικές τυχαίες μεταβλητές με παραμέτρους $\lambda_1, \lambda_2, \dots, \lambda_k$, αντίστοιχα, τότε η $\min(T_1, T_2, \dots, T_k)$ είναι επίσης εκθετική με παράμετρο $\sum_{i=1}^k \lambda_i$. Επιπλέον, το ενδεχόμενο $\{T_j = \min(T_1, T_2, \dots, T_k)\}$ είναι ανεξάρτητο της τυχαίας μεταβλητής $\min(T_1, T_2, \dots, T_k)$ και ισχύει ότι $\Pr[T_j = \min(T_1, T_2, \dots, T_k)] = \lambda_j / \sum_{i=1}^k \lambda_i$. Με βάση την ιδιότητα αυτή, μπορούμε να συνάγουμε το ακόλουθο κριτήριο για το πότε μια στοχαστική διαδικασία είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου.

Θεώρημα 4.4 (Κριτήριο Μαρκοβιανής αλυσίδας συνεχούς χρόνου) Έστω $\{X(t)\}$ μια στοχαστική διαδικασία με αριθμισμό (διακριτό) χώρο καταστάσεων \mathcal{S} και δοθέντος ότι $X(t) = i$,

- (i) υπάρχουν χρόνοι T_{ik} για $k \in \mathcal{S} \setminus \{i\}$, όπου ο χρόνος T_{ik} έχει την κατανομή $\text{Exp}(q_{ik})$, $k \in \mathcal{S} \setminus \{i\}$,
- (ii) ο χρόνος που θα γίνει η επόμενη μετάβαση είναι $\min_k T_{ik}$ και η κατάσταση j στην οποία πηγαίνει η $\{X(t)\}$ είναι αυτή για την οποία $T_{ij} = \min_k T_{ik}$.

Τότε η $\{X(t)\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου με χώρο καταστάσεων \mathcal{S} και πίνακα ρυθμών μετάβασης $\mathbf{Q} = (q_{ij})$. ◊

Με άλλα λόγια, μια Μαρκοβιανή αλυσίδα συνεχούς χρόνου χαρακτηρίζεται από το ότι οποιαδήποτε κατάσταση είναι συνδεδεμένη με έναν αριθμό γεγονότων, τα οποία συμβαίνουν μετά από εκθετικά κατανομημένο χρόνο και η αλυσίδα μεταβαίνει σε μια νέα κατάσταση όταν συμβεί το γεγονός που συνδέεται με αυτή.

Ένας άλλος τρόπος να σκεφτόμαστε αυτή τη διαδικασία είναι ότι όταν η αλυσίδα βρίσκεται στην κατάσταση i τότε λειτουργούν «ρολόγια-ξυπνητήρια», ένα για κάθε κατάσταση $k \neq i$, που θα «χτυπήσουν» μετά από εκθετικούς χρόνους με παραμέτρους q_{ik} . Όταν χτυπήσει το πρώτο ξυπνητήρι η αλυσίδα πηγαίνει στην κατάσταση που αντιστοιχεί σε αυτό, έστω στην j , οπότε ξεκινάει να λειτουργούν νέα «ρολόγια-ξυπνητήρια», ένα για κάθε κατάσταση $k \neq j$, κ.ο.κ.

Το κριτήριο αυτό μας επιτρέπει να ελέγχουμε εύκολα κατά πόσο μια στοχαστική διαδικασία που περιγράφει την εξέλιξη ενός συστήματος σε συνεχή χρόνο είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου. Στα επόμενα κεφάλαια του βιβλίου παρουσιάζονται πολλά παραδείγματα που δείχνουν τη χρήση αυτού του κριτηρίου στη Θεωρία Ουρών Αναμονής. Κάποια πρώτα παραδείγματα δίνονται αμέσως παρακάτω.

Παράδειγμα 4.5 (Η στοχαστική διαδικασία Poisson) Έστω ότι πελάτες έρχονται σύμφωνα με μια διαδικασία Poisson με ρυθμό λ και $X(t)$ είναι το πλήθος των αφίξεων στο $(0, t]$. Τότε για τη στοχαστική διαδικασία $\{X(t)\}$ έχουμε ότι ο χώρος καταστάσεων είναι ο $\mathcal{S} = \mathbb{N}_0$, δηλαδή αριθμισμός. Επιπλέον, δεδομένου ότι οι χρόνοι μεταξύ διαδοχικών αφίξεων είναι ανεξάρτητες τυχαίες μεταβλητές με την κατανομή $\text{Exp}(\lambda)$, σχετικά με τις μεταβάσεις έχουμε τον πίνακα 4.1.

Κατάσταση	Επόμενη κατάσταση	Χρόνος
$n \geq 0$	$n + 1$	$T_{n,n+1} \sim \text{Exp}(\lambda)$

Πίνακας 4.1: Πίνακας μεταβάσεων στοχαστικής διαδικασίας Poisson.

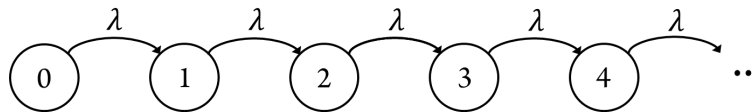
Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	1	$T_{01} \sim \text{Exp}(\lambda)$
1	0	$T_{10} \sim \text{Exp}(\mu)$

Πίνακας 4.2: Πίνακας μεταβάσεων M/M/1/1 ουράς.

Αφού όλοι οι χρόνοι είναι εκθετικοί έχουμε ότι η $\{X(t)\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου με πίνακα ρυθμών μετάβασης

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & \dots \\ 0 & -\lambda & \lambda & \dots \\ 0 & 0 & -\lambda & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

και διάγραμμα ρυθμών μετάβασης που δίνεται στο σχήμα 4.2.



Σχήμα 4.2: Διάγραμμα ρυθμών μετάβασης διαδικασίας Poisson.

Οι εξισώσεις Charpan-Kolmogorov για τη μεταβατική κατανομή είναι

$$\begin{aligned} p'_0(t) &= -\lambda p_0(t), \\ p'_n(t) &= -\lambda p_n(t) + \lambda p_{n-1}(t), \quad n \geq 1. \end{aligned}$$

◊

Παράδειγμα 4.6 (Η M/M/1/1 ουρά) Έστω η M/M/1/1 ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησεων μ , και έστω $X(t)$ το πλήθος των πελατών τη στιγμή t . Τότε για τη στοχαστική διαδικασία $\{X(t)\}$ έχουμε ότι ο χώρος καταστάσεων είναι ο $\mathcal{S} = \{0, 1\}$ και σχετικά με τις μεταβάσεις έχουμε τον πίνακα 4.2.

Αφού όλοι οι χρόνοι είναι εκθετικοί έχουμε ότι η $\{X(t)\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου με πίνακα ρυθμών μετάβασης

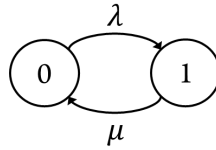
$$Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix},$$

και διάγραμμα ρυθμών μετάβασης που δίνεται στο σχήμα 4.3.

Οι εξισώσεις Charpan-Kolmogorov για τη μεταβατική κατανομή είναι

$$\begin{aligned} p'_0(t) &= -\lambda p_0(t) + \mu p_1(t), \\ p'_1(t) &= -\mu p_1(t) + \lambda p_0(t). \end{aligned}$$

◊



Σχήμα 4.3: Διάγραμμα ρυθμών μετάβασης M/M/1/1 ουράς.

Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	1	$T_{01} \sim \text{Exp}(\lambda)$
$n \geq 1$	$n + 1$	$T_{n,n+1} \sim \text{Exp}(\lambda)$
	$n - 1$	$T_{n,n-1} \sim \text{Exp}(\mu)$

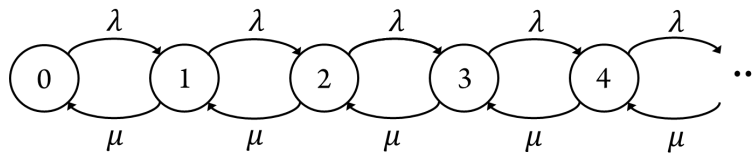
Πίνακας 4.3: Πίνακας μεταβάσεων M/M/1 ουράς.

Παράδειγμα 4.7 (H M/M/1 ουρά) Έστω η M/M/1 ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησεων μ , και έστω $X(t)$ το πλήθος των πελατών τη στιγμή t . Τότε για τη στοχαστική διαδικασία $\{X(t)\}$ έχουμε ότι ο χώρος καταστάσεων είναι ο $\mathcal{S} = \mathbb{N}_0$ και σχετικά με τις μεταβάσεις έχουμε τον πίνακα 4.3.

Αφού όλοι οι χρόνοι είναι εκθετικοί έχουμε ότι η $\{X(t)\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου με πίνακα ρυθμών μετάβασης

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & \dots \\ \mu & -(\lambda + \mu) & \lambda & \dots \\ 0 & \mu & -(\lambda + \mu) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

και διάγραμμα ρυθμών μετάβασης που δίνεται στο σχήμα 4.4.



Σχήμα 4.4: Διάγραμμα ρυθμών μετάβασης M/M/1 ουράς.

Οι εξισώσεις Charpman-Kolmogorov για τη μεταβατική κατανομή είναι

$$\begin{aligned} p'_0(t) &= -\lambda p_0(t) + \mu p_1(t), \\ p'_n(t) &= -(\lambda + \mu)p_n(t) + \lambda p_{n-1}(t) + \mu p_{n+1}(t), \quad n \geq 1. \end{aligned}$$

◊

Παράδειγμα 4.8 (H D/M/1 ουρά) Έστω η D/M/1 ουρά με σταθερούς ενδιάμεσους χρόνους αφίξεων a και ρυθμό εξυπηρέτησεων μ , και έστω $X(t)$ το πλήθος των πελατών τη στιγμή t . Τότε για τη στοχαστική διαδικασία $\{X(t)\}$ έχουμε ότι ο χώρος καταστάσεων είναι ο $\mathcal{S} = \mathbb{N}_0$ και σχετικά με τις μεταβάσεις έχουμε τον πίνακα 4.4.

Αφού έστω k ένας χρόνος δεν είναι εκθετικός έχουμε ότι η $\{X(t)\}$ δεν είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου.

◊

Δοθείσης μιας Μαρκοβιανής αλυσίδας συνεχούς χρόνου και δυο καταστάσεων i, j λέμε ότι η j είναι προ-σπελάσιμη από την i (συμβολικά $i \rightarrow j$), αν υπάρχει $t \geq 0$ τέτοιο ώστε $p_{ij}(t) > 0$. Αυτό αποδεικνύεται ότι είναι ισοδύναμο με την ύπαρξη μονοπατιού

$$i = i_0 \rightarrow i_1 \rightarrow i_2 \rightarrow \dots \rightarrow i_n = j,$$

Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	1	$T_{01} \sim \text{Exp}$
$n \geq 1$	$n + 1$	$T_{n,n+1} \sim \text{Exp}$
	$n - 1$	$T_{n,n-1} \sim \text{Exp}(\mu)$

Πίνακας 4.4: Πίνακας μεταβάσεων D/M/1 ουράς.

με $q_{i_0 i_1} q_{i_1 i_2} \cdots q_{i_{n-1} i_n} > 0$. Αν η i είναι προσπελάσιμη από την j , και, αντίστροφα, η j είναι προσπελάσιμη από την i , λέμε ότι οι i, j επικοινωνούν (συμβολικά $i \leftrightarrow j$). Η επικοινωνία είναι σχέση ισοδυναμίας και επομένως ο χώρος καταστάσεων διαμερίζεται σε κλάσεις επικοινωνίας (αδιαχώριστες κλάσεις), όπου όλες οι καταστάσεις μιας κλάσης επικοινωνούν μεταξύ τους. Μια Μαρκοβιανή αλυσίδα συνεχούς χρόνου λέγεται αδιαχώριστη, αν έχει μόνο μια κλάση επικοινωνίας, δηλαδή όλες οι καταστάσεις επικοινωνούν μεταξύ τους, διαφορετικά λέγεται διαχωρίσιμη. Μια κλάση επικοινωνίας $\mathcal{C} \subseteq \mathcal{S}$ λέγεται κλειστή αν δεν υπάρχουν $i \in \mathcal{C}$ και $j \notin \mathcal{C}$, με $p_{ij} > 0$, αλλιώς λέγεται ανοικτή. Αν μια Μαρκοβιανή αλυσίδα εισέλθει κάποια στιγμή σε μια κλειστή κλάση, λέμε ότι απορροφήθηκε σε αυτήν, αφού είναι αδύνατο να ξαναβγεί. Δοθέντος ενός συνόλου καταστάσεων μιας Μαρκοβιανής αλυσίδας συνεχούς χρόνου, μπορούμε να υπολογίσουμε τις πιθανότητες εισόδου σε αυτό, ξεκινώντας από οποιαδήποτε κατάσταση της Μαρκοβιανής αλυσίδας, καθώς και τους αντίστοιχους μέσους χρόνους πρώτης εισόδου. Στην περίπτωση που το σύνολο καταστάσεων είναι μια κλειστή κλάση επικοινωνίας της Μαρκοβιανής αλυσίδας, τότε η Μαρκοβιανή αλυσίδα δεν θα βγεί από το σύνολο από τη στιγμή που θα εισέλθει σε αυτό, επομένως, μιλάμε για πιθανότητες απορρόφησης και για μέσους χρόνους απορρόφησης. Έχουμε το ακόλουθο αποτέλεσμα:

Θεώρημα 4.9 (Πιθανότητες εισόδου/απορρόφησης και μέσοι χρόνοι πρώτης εισόδου/απορρόφησης)

Έστω $\{X(t) : t \geq 0\}$ Μαρκοβιανή αλυσίδα συνεχούς χρόνου, με χώρο καταστάσεων \mathcal{S} και πίνακα ρυθμών μεταβάσεως $\mathbf{Q} = (q_{ij})$. Έστω, επίσης, $\mathcal{C} \subseteq \mathcal{S}$. Ορίζουμε

$$T_{\mathcal{C}} = \inf\{t \geq 0 : X(t) \in \mathcal{C}\},$$

τον χρόνο πρώτης εισόδου της Μαρκοβιανής αλυσίδας στο \mathcal{C} (δηλ. τον χρόνο απορρόφησης στο \mathcal{C} στην περίπτωση που είναι κλειστή κλάση επικοινωνίας). Επίσης, έστω

$$h_i(\mathcal{C}) = \Pr[T_{\mathcal{C}} < \infty | X(0) = i], \quad i \in \mathcal{S},$$

η πιθανότητα εισόδου/απορρόφησης στο \mathcal{C} , ξεκινώντας από την κατάσταση i και

$$m_i(\mathcal{C}) = E[T_{\mathcal{C}} | X(0) = i], \quad i \in \mathcal{S},$$

ο μέσος χρόνος πρώτης εισόδου/απορρόφησης στο \mathcal{C} , ξεκινώντας από την κατάσταση i . Τότε:

(i) $H \mathbf{h}(\mathcal{C}) = (h_i(\mathcal{C}))$ είναι η ελάχιστη μη-αρνητική λύση του γραμμικού συστήματος

$$x_i = 1, \quad i \in \mathcal{C}, \quad (4.3)$$

$$x_i = \sum_{j \neq i} \frac{q_{ij}}{q_i} x_j, \quad i \in \mathcal{S} \setminus \mathcal{C}. \quad (4.4)$$

(ii) $H \mathbf{m}(\mathcal{C}) = (m_i(\mathcal{C}))$ είναι η ελάχιστη μη-αρνητική λύση του γραμμικού συστήματος

$$y_i = 0, \quad i \in \mathcal{C}, \quad (4.5)$$

$$y_i = \frac{1}{q_i} + \sum_{j \neq i} \frac{q_{ij}}{q_i} y_j, \quad i \in \mathcal{S} \setminus \mathcal{C}. \quad (4.6)$$

Είναι εύκολο να δούμε ότι οι πιθανότητες και οι μέσοι χρόνοι εισόδου ικανοποιούν τις εξισώσεις (4.3)-(4.6), δεσμεύοντας στην πρώτη μετάβαση της Μαρκοβιανής αλυσίδας. Πράγματι, ξεκινώντας από μια κατάσταση $i \in \mathcal{E}$, η είσοδος είναι βέβαια και ο αντίστοιχος μέσος χρόνος είναι μηδενικός. Αν $i \notin \mathcal{E}$, τότε στο πρώτο βήμα η Μαρκοβιανή αλυσίδα θα μεταβεί σε κάποια κατάσταση j και θα πρέπει ξεκινώντας από αυτήν να εισέλθει στο \mathcal{E} . Η πιθανότητα η επόμενη κατάσταση να είναι η j είναι $\frac{q_{ji}}{q_i}$, ενώ ο μέσος χρόνος μετάβασης είναι ο μέσος χρόνος παραμονής στην i . Όπως έχουμε δει, ο χρόνος παραμονής στην i έχει την $\text{Exp}(q_i)$ κατανομή, με μέση τιμή $\frac{1}{q_i}$.

Δοθείσης μιας κατάστασης j μιας Μαρκοβιανής αλυσίδας, ορίζουμε h_j της πιθανότητα επανόδου σε αυτήν και με m_j τον αντίστοιχο μέσο χρόνο επανόδου. Για να υπολογίσουμε αυτές τις ποσότητες, θεωρούμε τη Μαρκοβιανή αλυσίδα που προκύπτει, αν τροποποιήσουμε την αρχική Μαρκοβιανή αλυσίδα, κάνοντας την κατάσταση j απορροφητική. Τότε, μπορούμε να υπολογίσουμε σε αυτή τη νέα αλυσίδα τις πιθανότητες απορρόφησης $h_i(\{j\})$ και τους μέσους χρόνους απορρόφησης $m_i(\{j\})$ και είναι

$$h_j = \sum_{i \neq j} \frac{q_{ji}}{q_j} h_i(\{j\}),$$

$$m_j = \frac{1}{q_j} + \sum_{i \neq j} \frac{q_{ji}}{q_j} m_i(\{j\}).$$

Ορισμός 4.10 (Επαναληπτικότητα - παροδικότητα καταστάσεων) Έστω Μαρκοβιανή αλυσίδα συνεχούς χρόνου $\{X(t) : t \geq 0\}$ και κατάσταση της, j , με πιθανότητα επανόδου h_j και μέσο χρόνο επανόδου m_j .

(i) Η j λέγεται θετικά επαληπτική αν $h_j = 1$ και $m_j < \infty$.

(ii) Η j λέγεται μηδενικά επαληπτική αν $h_j = 1$ και $m_j = \infty$.

(iii) Η j λέγεται παροδική αν $h_j < 1$.

Οι ιδιότητες της θετικής επαναληπτικότητας, της μηδενικής επαναληπτικότητας και της παροδικότητας είναι ιδιότητες των κλάσεων επικοινωνίας, δηλαδή, αν οι καταστάσεις i και j επικοινωνούν, τότε είναι ίδιου τύπου ως προς τη θετική επαναληπτικότητα, τη μηδενική επαναληπτικότητα και την παροδικότητα.

Ισχύει, επίσης, ότι οι καταστάσεις μιας ανοικτής κλάσης επικοινωνίας είναι παροδικές, ενώ οι καταστάσεις μιας πεπερασμένης κλειστής κλάσης επικοινωνίας είναι θετικά επαναληπτικές. Επομένως, η μόνη περίπτωση που απαιτεί περαιτέρω διερεύνηση είναι αυτή της άπειρης κλειστής κλάσης. Εκεί, είναι ανοικτά όλα τα ενδεχόμενα: Μπορεί οι καταστάσεις της να είναι όλες παροδικές ή όλες μηδενικά επαναληπτικές ή όλες θετικά επαναληπτικές.

4.2 Μαρκοβιανές αλυσίδες συνεχούς χρόνου: Οριακή συμπεριφορά

Από τα προηγούμενα, είναι φανερό ότι η μελέτη της μεταβατικής συμπεριφοράς μιας Μαρκοβιανής αλυσίδας συνεχούς χρόνου, δηλαδή η μελέτη των πιθανοτήτων η αλυσίδα να βρίσκεται στις διάφορες δυναμικές καταστάσεις της μια δοσμένη στιγμή t είναι ένα υπολογιστικό απαιτητικό πρόβλημα. Πράγματι, μια τέτοια μελέτη απαιτεί τη λύση του συστήματος των διαφορικών εξισώσεων Chapman-Kolmogorov ή ισοδύναμα τον υπολογισμό της πινακικής εκθετικής συνάρτησης e^{Qt} . Όπως θα δούμε στην παρούσα παράγραφο, η μελέτη της οριακής συμπεριφοράς μιας Μαρκοβιανής αλυσίδας είναι σαφώς απλούστερη. Επιπλέον, από πλευράς εφαρμογών στη Θεωρία Ουρών Αναμονής, αυτό που μας ενδιαφέρει για ένα σύστημα είναι η μακροπρόθεσμη αποτίμησή του στο χρόνο και όχι η εκτίμηση της λειτουργίας του για μικρό χρονικό διάστημα, οπότε η

οριακή συμπεριφορά είναι πράγματι αυτό που έχει σημασία. Τέλος, πρέπει να σημειώσουμε ότι στην πλειονότητα των περιπτώσεων η σύγκλιση της μεταβατικής κατανομής μιας Μαρκοβιανής αλυσίδας στην οριακή κατανομή της είναι πολύ γρήγορη (εκθετική ως προς το χρόνο). Επομένως, ακόμη και για σχετικά μικρούς χρόνους t η εκτίμηση της μεταβατικής κατανομής από την οριακή κατανομή δίνει πολύ καλά αποτελέσματα.

Θα προχωρήσουμε, λοιπόν, στη μελέτη της οριακής συμπεριφοράς των Μαρκοβιανών αλυσίδων συνεχούς χρόνου. Θα περιοριστούμε στην περίπτωση των αδιαχώριστων αλυσίδων, αφού αυτές είναι που κατά κανόνα εμφανίζονται στις εφαρμογές στη Θεωρία Ουρών. Έχουμε το ακόλουθο αποτέλεσμα.

Θεώρημα 4.11 (Οριακή συμπεριφορά αδιαχώριστων Μαρκοβιανών αλυσίδων συνεχούς χρόνου) Έστω $\{X(t) : t \geq 0\}$ μια αδιαχώριστη Μαρκοβιανή αλυσίδα συνεχούς χρόνου, με χώρο καταστάσεων \mathcal{S} και πίνακα ρυθμών μετάβασης $\mathbf{Q} = (q_{ij})$. Η $\{X(t) : t \geq 0\}$ είναι θετικά επαναληπτική, αν και μόνο αν, το σύστημα των εξισώσεων ισορροπίας

$$p_j q_j = \sum_{i \neq j} p_i q_{ij}, \quad j \in \mathcal{S} \quad (4.7)$$

έχει λύση που ικανοποιεί και την εξίσωση κανονικοποίησης

$$\sum_{j \in \mathcal{S}} p_j = 1.$$

Αν υπάρχει τέτοια λύση $\mathbf{p} = (p_j)$, τότε είναι μοναδική. Επιπλέον, όλες οι συντεταγμένες της είναι θετικές και κάθε άλλη λύση $\mathbf{x} = (x_j)$ του συστήματος των εξισώσεων ισορροπίας είναι πολλαπλάσιό της: $\mathbf{x} = c\mathbf{p}$, για κάποιο $c \in \mathbb{R}$. Έχουμε:

(i) Η πιθανότητα p_j εκφράζει το μακροπρόθεσμο ποσοστό του χρόνου που η Μαρκοβιανή αλυσίδα περνάει στην κατάσταση j , δηλαδή,

$$\lim_{t \rightarrow \infty} \frac{\int_0^t 1_{\{X(u)=j\}} du}{t} = p_j, \quad \text{με πιθανότητα 1, } j \in \mathcal{S}.$$

Δηλαδή, η \mathbf{p} είναι η οριακή δειγματική ή οριακή εμπειρική κατανομή της $\{X(t)\}$.

(ii) Η πιθανότητα p_j εκφράζει το μακροπρόθεσμο μέσο ποσοστό του χρόνου που η Μαρκοβιανή αλυσίδα περνάει στην κατάσταση j , δηλαδή,

$$\lim_{t \rightarrow \infty} \frac{E\left[\int_0^t 1_{\{X(u)=j\}} du\right]}{t} = p_j, \quad j \in \mathcal{S}.$$

(iii) Η πιθανότητα p_j συνδέεται με τον ρυθμό εξόδου από την j , q_j , και τον μέσο χρόνο επανόδου στην j , m_j , ως εξής:

$$p_j = \frac{1}{q_j m_j}, \quad j \in \mathcal{S}.$$

(iv) Η πιθανότητα p_j είναι η C-οριακή πιθανότητα (Cesaro οριακή πιθανότητα) της κατάστασης j :

$$\lim_{t \rightarrow \infty} \frac{\int_0^t p_j(u) du}{t} = p_j, \quad j \in \mathcal{S}.$$

Δηλαδή, η \mathbf{p} είναι η C-οριακή κατανομή της $\{X(t)\}$.

(v) Η πιθανότητα p_j είναι η οριακή πιθανότητα της κατάστασης j :

$$\lim_{t \rightarrow \infty} p_j(t) = p_j, \quad j \in \mathcal{S}.$$

Δηλαδή, η \mathbf{p} είναι η οριακή κατανομή της $\{X(t)\}$.

(vi) Αν η Μαρκοβιανή αλυσίδα έχει αρχική κατανομή την \mathbf{p} , τότε κάθε μεταβατική κατανομή της δίνεται πάλι από την \mathbf{p} :

$$\mathbf{p}(0) = \mathbf{p} \Rightarrow \mathbf{p}(t) = \mathbf{p}, \quad t \geq 0.$$

Δηλαδή, η \mathbf{p} είναι η στάσιμη κατανομή της $\{X(t)\}$.

◊

Από το θεώρημα 4.11, βλέπουμε ότι η κατανομή \mathbf{p} φέρει τη σημαντικότερη πληροφορία για τη μακροπρόθεσμη συμπεριφορά μιας Μαρκοβιανής αλυσίδας συνεχούς χρόνου. Αναφέρεται ως οριακή εμπειρική, C-οριακή, οριακή και στάσιμη κατανομή της Μαρκοβιανής αλυσίδας, καθώς και ως κατανομή ισορροπίας. Αν η Μαρκοβιανή αλυσίδα έχει τη στάσιμη κατανομή της ως αρχική, οι μεταβατικές της κατανομές είναι όλες ίσες με τη στάσιμη (βλέπε (vi)) και η Μαρκοβιανή αλυσίδα λέγεται στάσιμη ή λέμε ότι βρίσκεται σε κατάσταση στατιστικής ισορροπίας. Διαισθητικά, αυτό σημαίνει ότι έχει περάσει αρκετός χρόνος από τότε που το στοχαστικό σύστημα άρχισε να λειτουργεί και η επίδραση της αρχικής του κατάστασης έχει χαθεί. Έτσι, παρουσιάζει ομοιόμορφη συμπεριφορά στο χρόνο. Για τα (ii), (iv) και (v) ισχύουν και οι αντίστοιχοι τύποι που αφορούν τις δεσμευμένες ποσότητες ως προς την αρχική κατάσταση. Δηλαδή, έχουμε:

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{E\left[\int_0^t 1_{\{X(u)=j\}} du \mid X(0) = i\right]}{t} &= p_j, \quad i, j \in \mathcal{S}, \\ \lim_{t \rightarrow \infty} \frac{\int_0^t p_{ij}(u) du}{t} &= p_j, \quad i, j \in \mathcal{S}, \\ \lim_{t \rightarrow \infty} p_{ij}(t) &= p_j, \quad i, j \in \mathcal{S}. \end{aligned}$$

Μια διαισθητική ερμηνεία του (iv) είναι ότι η πιθανότητα p_j εκφράζει την πιθανότητα η Μαρκοβιανή αλυσίδα να βρίσκεται στην κατάσταση j μια χρονική στιγμή u που έχει επιλεγεί ομοιόμορφα από το χρονικό διάστημα $(0, t]$, για μεγάλο t . Δηλαδή, ισχύει

$$\lim_{t \rightarrow \infty} \Pr[X(U_t) = j] = p_j, \quad j \in \mathcal{S},$$

όπου η U_t είναι ομοιόμορφη στο $(0, t]$ και ανεξάρτητη της $\{X(t) : t \geq 0\}$. Πράγματι,

$$\begin{aligned} \lim_{t \rightarrow \infty} \Pr[X(U_t) = j] &= \lim_{t \rightarrow \infty} \int_0^\infty \Pr[X(U_t) = j \mid U_t = u] dF_{U_t}(u) \\ &= \lim_{t \rightarrow \infty} \int_0^t \Pr[X(u) = j] \frac{1}{t} du \\ &= \lim_{t \rightarrow \infty} \frac{\int_0^t p_j(u) du}{t} = p_j, \quad j \in \mathcal{S}. \end{aligned} \quad (4.8)$$

Τα περισσότερα από τα συμπεράσματα του θεωρήματος 4.11 μπορούν να δικαιολογηθούν διαισθητικά, παρότι οι αυστηρές αποδείξεις τους δεν είναι απλές. Π.χ., για μια διαισθητική αιτιολόγηση των (i), (ii) ας συμβολίσουμε με p_j το μακροπρόθεσμο ποσοστό του χρόνου που η Μαρκοβιανή αλυσίδα περνάει στην κατάσταση j . Τότε η ποσότητα $p_j q_{ji}$ εκφράζει το μακροπρόθεσμο ρυθμό μεταβάσεων τύπου $j \rightarrow i$. Αυτό είναι φανερό αφού η q_{ji} εκφράζει το ρυθμό των μεταβάσεων προς την i , για κάθε χρονική μονάδα παραμονής στην j , ενώ η p_j είναι το ποσοστό του χρόνου που η αλυσίδα περνάει στην j . Επομένως, η ποσότητα $\sum_{i \neq j} p_j q_{ji}$ εκφράζει το μακροπρόθεσμο ρυθμό αναχωρήσεων από την κατάσταση j , ενώ η ποσότητα $\sum_{i \neq j} p_i q_{ij}$ εκφράζει το μακροπρόθεσμο ρυθμό αφίξεων στην κατάσταση j . Αυτοί οι δυο ρυθμοί θα πρέπει να είναι ίσοι, αφού κάθε άφιξη στην j ακολουθείται από μια αναχώρηση από αυτήν, δηλαδή αφίξεις και αναχωρήσεις σε κάθε κατάσταση j βρίσκονται σε μια 1-1 αντιστοιχία. Οπότε, κατ' ανάγκη ισχύει

$$p_j q_j = \sum_{i \neq j} p_j q_{ji} = \sum_{i \neq j} p_i q_{ij}, \quad j \in \mathcal{S},$$

δηλαδή, τα μακροπρόθεσμα ποσοστά του χρόνου p_j που η Μαρκοβιανή αλυσίδα περνάει στις διάφορες καταστάσεις πρέπει να ικανοποιούν τις εξισώσεις ισορροπίας (4.7).

Για το (iii), παρατηρούμε ότι οι χρόνοι μεταξύ διαδοχικών επισκέψεων σε κάθε συγκεκριμένη κατάσταση j μιας Μαρκοβιανής αλυσίδας είναι ανεξάρτητοι και ισόνομοι, οπότε οι στιγμές επισκέψεων (εισόδων) στην j αντιστοιχούν στα γεγονότα μιας ανανεωτικής διαδικασίας και μάλιστα είναι αναγεννητικές στιγμές για τη Μαρκοβιανή αλυσίδα. Εφαρμόζουμε το Εργοδικό θεώρημα των αναγεννητικών διαδικασιών και έχουμε ότι το ποσοστό του χρόνου που η αλυσίδα περνάει στην j , δηλαδή η p_j , ισούται με τον λόγο της μέσης διάρκειας παραμονής στην j σε έναν ανανεωτικό κύκλο, διά τη μέση διάρκεια ενός ανανεωτικού κύκλου. Στη διάρκεια ενός κύκλου, δηλαδή μεταξύ δυο διαδοχικών εισόδων στην j , η αλυσίδα παραμένει στην j για έναν εκθετικό χρόνο με παράμετρο q_j . Επομένως, η μέση διάρκεια παραμονής στην j σε έναν κύκλο είναι $\frac{1}{q_j}$. Από την άλλη, η μέση διάρκεια του κύκλου είναι m_j , οπότε προκύπτει άμεσα η (iii).

Το (iv) είναι μια απλή αναδιατύπωση του (ii) που προκύπτει αφού μπορούμε να εναλλάξουμε τη μέση τιμή με το ολοκλήρωμα. Για το (v), είναι γνωστό ότι αν το όριο μιας συνάρτησης υπάρχει, τότε υπάρχει και το C-όριο της και είναι ίσα. Για τις Μαρκοβιανές αλυσίδες συνεχούς χρόνου αποδεικνύεται ότι το όριο της $p_j(t)$ υπάρχει, οπότε θα είναι αναγκαστικά ίσο με p_j , λόγω του (iv).

Οι εξισώσεις ισορροπίας (4.7) γράφονται σε πινακική μορφή ως

$$\mathbf{p}^T \mathbf{Q} = \mathbf{0}^T.$$

Επομένως, από τα (iii)-(iv) του θεωρήματος 4.2, έχουμε ότι αν $\mathbf{p}(0) = \mathbf{p}$, τότε

$$\begin{aligned} (\mathbf{p}(t))^T &= (\mathbf{p}(0))^T e^{\mathbf{Q}t} = \mathbf{p}^T e^{\mathbf{Q}t} \\ &= \mathbf{p}^T \sum_{n=0}^{\infty} \mathbf{Q}^n \frac{t^n}{n!} \\ &= \mathbf{p}^T + \sum_{n=1}^{\infty} \mathbf{p}^T \mathbf{Q}^n \frac{t^n}{n!} = \mathbf{p}^T, \quad t \geq 0. \end{aligned}$$

Επομένως, αν $p_j(0) = p_j, j \in \mathcal{S}$, τότε $p_j(t) = p_j, j \in \mathcal{S}, t \geq 0$ και έχουμε το (vi).

Σημειώνουμε, επίσης, ότι όποτε ο χώρος καταστάσεων μιας αδιαχώριστης Μαρκοβιανής αλυσίδας συνεχούς χρόνου είναι πεπερασμένος, τότε υπάρχει πάντοτε η στάσιμη κατανομή.

Μια χρήσιμη παρατήρηση, που σε πολλές περιπτώσεις διευκολύνει τον υπολογισμό της στάσιμης κατανομής, είναι ότι αυτή ικανοποιεί επίσης και τις λεγόμενες εξισώσεις γενικευμένης ισορροπίας. Οι εξισώσεις αυτές απαιτούν ο μακροπρόθεσμος ρυθμός μεταβάσεων που εξέρχονται από ένα σύνολο καταστάσεων \mathcal{A} να ισούται με τον μακροπρόθεσμο ρυθμό μεταβάσεων που εισέρχονται στο ίδιο σύνολο, δηλαδή

$$\sum_{j \in \mathcal{A}} \sum_{i \in \mathcal{A}^c} p_j q_{ji} = \sum_{i \in \mathcal{A}^c} \sum_{j \in \mathcal{A}} p_i q_{ij}, \quad \mathcal{A} \subseteq \mathcal{S}. \quad (4.9)$$

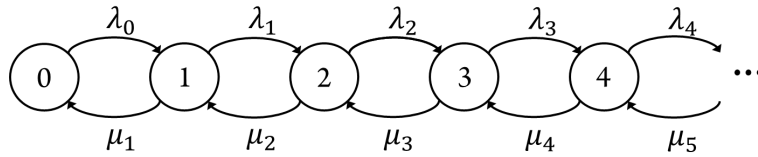
Χρησιμοποιώντας την παρατήρηση αυτή, είναι εύκολο να προσδιορίσουμε τη στάσιμη κατανομή μιας Μαρκοβιανής αλυσίδας συνεχούς χρόνου, $\{X(t)\}$, με χώρο καταστάσεων \mathbb{N}_0 και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda_i & \text{αν } i \geq 0, j = i + 1, \\ \mu_i & \text{αν } i \geq 1, j = i - 1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Μια τέτοια αλυσίδα αναφέρεται ως διαδικασία γέννησης-θανάτου, δεδομένου ότι περιγράφει την εξέλιξη του μεγέθους ενός πληθυσμού που αυξομειώνεται μέσω γεννήσεων (μεταβάσεων τύπου $i \rightarrow i + 1$) και θανάτων (μεταβάσεων τύπου $i \rightarrow i - 1$). Το διάγραμμα ρυθμών μετάβασης μια τέτοιας αλυσίδας δίνεται στο σχήμα 4.5.

Θεωρώντας τα σύνολα $\mathcal{A}_n = \{0, 1, 2, \dots, n - 1\}, n \geq 1$, έχουμε τις εξισώσεις γενικευμένης ισορροπίας

$$\lambda_{n-1} p_{n-1} = \mu_n p_n, \quad n \geq 1,$$



Σχήμα 4.5: Διάγραμμα ρυθμών μετάβασης διαδικασίας γέννησης-θανάτου.

που εξισώνουν τους μακροπρόθεσμους ρυθμούς εξόδου και εισόδου σε κάθε σύνολο. Από αυτές παίρνουμε ότι

$$p_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} p_0, \quad n \geq 1.$$

Από την εξίσωση κανονικοποίησης $\sum_{n=0}^{\infty} p_n = 1$, συνάγουμε ότι η διαδικασία γέννησης-θανάτου είναι ευσταθής, δηλαδή υπάρχει στάσιμη κατανομή, αν και μόνο αν

$$B^{-1} = 1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} < \infty.$$

Τότε, η στάσιμη κατανομή της διαδικασίας γέννησης-θανάτου δίνεται από τον τύπο

$$p_n = \begin{cases} B & \text{αν } n = 0, \\ B \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} & \text{αν } n \geq 1. \end{cases}$$

4.3 Μαρκοβιανές αλυσίδες συνεχούς χρόνου: Αμοιβές και κόστη

Η εξέλιξη ενός στοχαστικού συστήματος που περιγράφεται από μια Μαρκοβιανή αλυσίδα συνεχούς χρόνου συνδέεται, συνήθως, με κάποια δομή κόστους/αμοιβής, με άμεση εφαρμογή στη Θεωρία Ουρών Αναμονής. Πιο συγκεκριμένα, δίνουμε τον ακόλουθο ορισμό.

Ορισμός 4.12 (Δομή κόστους σε Μαρκοβιανή αλυσίδα συνεχούς χρόνου) Έστω $\{X(t) : t \geq 0\}$ Μαρκοβιανή αλυσίδα συνεχούς χρόνου, με χώρο καταστάσεων \mathcal{S} και πίνακα ρυθμών μετάβασης $\mathbf{Q} = (q_{ij})$.

- (i) Μια δομή κόστους παραμονής (holding cost) είναι μια συνάρτηση $c : \mathcal{S} \rightarrow \mathbb{R}$, που δίνει για κάθε κατάσταση j του \mathcal{S} , το κόστος c_j ανά χρονική μονάδα παραμονής της $\{X(t)\}$ σε αυτήν.
- (ii) Μια δομή κόστους μετάβασης (transition cost) είναι μια συνάρτηση $d : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$, που δίνει για κάθε μετάβαση $i \rightarrow j$ του \mathcal{S} , το κόστος μετάβασης d_{ij} της $\{X(t)\}$. \diamond

Το παρακάτω αποτέλεσμα παρέχει έναν εύκολο τρόπο υπολογισμού του μακροπρόθεσμου ρυθμού κόστους.

Θεώρημα 4.13 (Μακροπρόθεσμος ρυθμός κόστους σε Μαρκοβιανή αλυσίδα συνεχούς χρόνου) Έστω $\{X(t) : t \geq 0\}$ Μαρκοβιανή αλυσίδα συνεχούς χρόνου, με χώρο καταστάσεων \mathcal{S} και πίνακα ρυθμών μετάβασης $\mathbf{Q} = (q_{ij})$. Έστω $c : \mathcal{S} \rightarrow \mathbb{R}$ μια δομή κόστους παραμονής ανά χρονική μονάδα και $d : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$ μια δομή κόστους μετάβασης. Έστω επίσης $A(t)$ το πλήθος των μεταβάσεων που έχουν συμβεί στο $(0, t]$ στην $\{X(t)\}$ και $\{X_n : n \geq 0\}$ η αντίστοιχη εμφυτευμένη στοχαστική διαδικασία διακριτού χρόνου σε στιγμές μεταβάσεων της $\{X(t)\}$. Ορίζουμε

$$C(t) = \int_0^t c_{X(u)} du + \sum_{n=0}^{A(t)-1} d_{X_n, X_{n+1}}, \quad t \geq 0,$$

το συνολικό κόστος που συσσωρεύεται μέχρι τη στιγμή t από την εξέλιξη της Μαρκοβιανής αλυσίδας $\{X(t)\}$. Υποθέτουμε, επίσης, ότι η Μαρκοβιανή αλυσίδα είναι αδιαχώριστη, θετικά επαναληπτική με κατανομή ισορροπίας $\mathbf{p} = (p_j : j \in \mathcal{S})$ και

$$\sum_{j \in \mathcal{S}} p_j \left(|c_j| + \sum_{k \neq j} q_{jk} |d_{jk}| \right) < \infty.$$

Τότε:

(i) Για τον μακροπρόθεσμο ρυθμό κόστους έχουμε

$$\lim_{t \rightarrow \infty} \frac{C(t)}{t} = \sum_{j \in \mathcal{S}} p_j c_j + \sum_{j \in \mathcal{S}} \sum_{k \neq j} p_j q_{jk} d_{jk}, \quad \text{με πιθανότητα 1.}$$

(ii) Για τον μακροπρόθεσμο μέσο ρυθμό κόστους έχουμε

$$\lim_{t \rightarrow \infty} \frac{E[C(t)]}{t} = \sum_{j \in \mathcal{S}} p_j c_j + \sum_{j \in \mathcal{S}} \sum_{k \neq j} p_j q_{jk} d_{jk}.$$

◊

Το αποτέλεσμα είναι διαισθητικά εύλογο, αφού η p_j εκφράζει το μακροπρόθεσμο ποσοστό του χρόνου που η Μαρκοβιανή αλυσίδα βρίσκεται στην κατάσταση j , οπότε και το κόστος παραμονής θα είναι c_j για μια χρονική μονάδα. Επομένως ο όρος $\sum_{j \in \mathcal{S}} p_j c_j$ εκφράζει το μέσο κόστος παραμονής ανά χρονική μονάδα, δεσμεύοντας στην κατάσταση που μπορεί να βρίσκεται το σύστημα. Από την άλλη μεριά, η $p_j q_{jk}$ εκφράζει το μακροπρόθεσμο ρυθμό μεταβάσεων από την j στην k . Κάθε τέτοια μετάβαση επάγει κόστος d_{jk} . Επομένως, ο όρος $\sum_{j \in \mathcal{S}} \sum_{k \in \mathcal{S}} p_j q_{jk} d_{jk}$ εκφράζει το μέσο κόστος μεταβάσεων ανά χρονική μονάδα, δεσμεύοντας στην κατάσταση j που μπορεί να βρίσκεται το σύστημα και στην κατάσταση k στην οποία μεταβαίνει.

4.4 Διαδικασία Poisson

Έχουμε ήδη δει ότι η στοχαστική διαδικασία Poisson είναι ειδική περίπτωση τόσο ανανεωτικής διαδικασίας (βλέπε ορισμό 1.1), όσο και Μαρκοβιανής αλυσίδας συνεχούς χρόνου (βλέπε παράδειγμα 4.5). Πιο συγκεκριμένα, μια στοχαστική διαδικασία Poisson με ρυθμό λ είναι μια ανανεωτική διαδικασία με $\text{Exp}(\lambda)$ χρόνους μεταξύ των γεγονότων και επίσης είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου με χώρο καταστάσεων τον \mathbb{N}_0 , αρχική κατανομή $(p_i(0))$ με $p_0(0) = 1$ και πίνακα ρυθμών μετάβασης

$$\mathbf{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & \cdots \\ 0 & -\lambda & \lambda & \cdots \\ 0 & 0 & -\lambda & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Το διάγραμμα των ρυθμών μετάβασής της δίνεται στο σχήμα 4.2.

Η εκθετική κατανομή κληροδοτεί στη διαδικασία Poisson πολύ καλές ιδιότητες που επιτρέπουν κλειστές μορφές για τις βασικές ποσότητες σε πεπερασμένη χρονική στιγμή. Συγκεκριμένα, έχουμε το ακόλουθο αποτέλεσμα.

Θεώρημα 4.14 (Βασικοί υπολογισμοί στοχαστικής διαδικασίας Poisson σε πεπερασμένη χρονική στιγμή)
 Έστω διαδικασία Poisson $\{N(t) : t \geq 0\}$ με χρονικές στιγμές εμφάνισης γεγονότων S_1, S_2, \dots και ενδιάμεσους χρόνους εμφάνισης γεγονότων X_1, X_2, \dots με εκθετική κατανομή με παράμετρο λ . Τότε έχουμε:

(i) Η κατανομή του S_k είναι *Erlang*(k, λ), δηλαδή η συνάρτηση κατανομής είναι

$$F_{S_k}(t) = \Pr[S_k \leq t] = 1 - \sum_{i=0}^{k-1} e^{-\lambda t} \frac{(\lambda t)^i}{i!}, \quad t \geq 0, \quad k \geq 1,$$

και η συνάρτηση πυκνότητας πιθανότητας είναι

$$f_{S_k}(t) = \frac{\lambda^k}{(k-1)!} t^{k-1} e^{-\lambda t}, \quad t \geq 0, \quad k \geq 1.$$

(ii) Η κατανομή της $N(t)$ είναι *Poisson* με παράμετρο λt , οπότε

$$p_k(t) = \Pr[N(t) = k] = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad t \geq 0, \quad k \geq 0.$$

(iii) Η μέση τιμή της $\{N(t)\}$ είναι γραμμική συνάρτηση του t :

$$m(t) = E[N(t)] = \lambda t.$$

◊

Πέραν του ανανεωτικού ορισμού της διαδικασίας *Poisson* και του ορισμού της ως *Μαρκοβιανής αλυσίδας* συνεχούς χρόνου, η στοχαστική διαδικασία *Poisson* μπορεί να οριστεί και με τους ακόλουθους δυο τρόπους (που είναι βέβαια στενά συνδεδεμένοι με τους άλλους δύο ορισμούς):

Ορισμός 4.15 (Ολικός ορισμός της στοχαστικής διαδικασίας *Poisson*) Μια απαριθμήτρια στοχαστική διαδικασία $\{N(t)\}$ λέγεται *στοχαστική διαδικασία Poisson* με ρυθμό λ αν

- (i) έχει ανεξάρτητες προσανξήσεις, δηλαδή για κάθε $0 < t_1 < t_2 < \dots < t_n$ οι τυχαίες μεταβλητές $N(t_1), N(t_2) - N(t_1), \dots, N(t_n) - N(t_{n-1})$ είναι ανεξάρτητες,
- (ii) έχει ομογενείς προσανξήσεις, δηλαδή για κάθε $t, s > 0$, η κατανομή της $N(t+s) - N(s)$ δεν εξαρτάται από το s ,
- (iii) η κατανομή της $N(t)$ είναι *Poisson* με παράμετρο λt , δηλαδή

$$\Pr[N(t) = n] = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n \geq 0.$$

◊

Ορισμός 4.16 (Τοπικός ορισμός της στοχαστικής διαδικασίας *Poisson*) Μια απαριθμήτρια στοχαστική διαδικασία $\{N(t)\}$ λέγεται *στοχαστική διαδικασία Poisson* με ρυθμό λ αν

- (i) έχει ανεξάρτητες προσανξήσεις, δηλαδή για κάθε $0 < t_1 < t_2 < \dots < t_n$ οι τυχαίες μεταβλητές $N(t_1), N(t_2) - N(t_1), \dots, N(t_n) - N(t_{n-1})$ είναι ανεξάρτητες,
- (ii) έχει ομογενείς προσανξήσεις, δηλαδή για κάθε $t, s > 0$, η κατανομή της $N(t+s) - N(s)$ δεν εξαρτάται από το s ,
- (iii) Ισχύει

$$\Pr[N(h) = n] = \begin{cases} 1 - \lambda h + o(h), & \text{αν } n = 0, \\ \lambda h + o(h), & \text{αν } n = 1, \\ o(h), & \text{αν } n \geq 2, \end{cases}$$

για $h \rightarrow 0^+$ (όπου $o(h)$ είναι μια συνάρτηση του h με $\lim_{h \rightarrow 0^+} \frac{o(h)}{h} = 0$).

Η απόδειξη της ισοδυναμίας των τεσσάρων ορισμών είναι κάπως μακροσκελής, οπότε περιοριζόμαστε στο να δώσουμε μόνο το γενικό της περίγραμμα.

Καταρχήν η ισοδυναμία του τοπικού ορισμού με τον Μαρκοβιανό ορισμό είναι άμεση, με βάση το θεώρημα 4.4. Από τον ανανεωτικό ορισμό της στοχαστικής διαδικασίας Poisson είναι εύκολο να δούμε ότι για κάθε χρονική στιγμή s , η $\{N(u) : 0 \leq u \leq s\}$ και η $\{N_s(u) = N(s+u) - N(s) : u \geq 0\}$ είναι ανεξάρτητες και επιπλέον η διαδικασία $\{N_s(u) : u \geq 0\}$ είναι στοχαστική διαδικασία Poisson με ρυθμό λ , όμοια με την αρχική. Δηλαδή, δοθείσης μιας χρονικής στιγμής, το μέλλον της διαδικασίας (που αντιστοιχεί στην $\{N_s(u) : u \geq 0\}$) είναι ανεξάρτητο από το παρελθόν της (που αντιστοιχεί στην $\{N(u) : 0 \leq u \leq s\}$) και επιπλέον οποιαδήποτε χρονική στιγμή η διαδικασία είναι σαν να ξεκινάει από την αρχή. Πραγματικά, η $\{N(u) : 0 \leq u \leq s\}$ καθορίζεται πλήρως από τους ενδιάμεσους χρόνους X_1, X_2, \dots, X_k της αρχικής διαδικασίας και το ενδεχόμενο $\{X_{k+1} > s - X_1 - X_2 - \dots - X_k\}$, για κάποιο k (όποτε ισχύει $N(s) = k$). Η $\{N_s(u) : u \geq 0\}$ καθορίζεται από τον υπολειπόμενο χρόνο για το πρώτο γεγονός που είναι ο $X_{k+1} - (s - X_1 - X_2 - \dots - X_k)$ και τους ενδιάμεσους χρόνους X_{k+2}, X_{k+3}, \dots της αρχικής διαδικασίας. Λόγω της ανεξαρτησίας των ενδιάμεσων χρόνων της αρχικής διαδικασίας και της αμνήμονης ιδιότητας της εκθετικής κατανομής που εξασφαλίζει ότι η κατανομή του $X_{k+1} - (s - X_1 - X_2 - \dots - X_k)$, δεδομένου ότι $\{X_{k+1} > s - X_1 - X_2 - \dots - X_k\}$, είναι $\text{Exp}(\lambda)$ και ανεξάρτητη των X_1, X_2, \dots, X_k και του ενδεχομένου $\{X_{k+1} > s - X_1 - X_2 - \dots - X_k\}$ έχουμε τη ζητούμενη ανεξαρτησία. Επιπλέον, έχουμε ότι η $\{N_s(u) : u \geq 0\}$ έχει ανεξάρτητους και ισόνομους ενδιάμεσους χρόνους με την $\text{Exp}(\lambda)$ κατανομή. Επομένως, έχουμε και ότι ισχύουν οι ιδιότητες (i) και (ii) του ολικού ορισμού της διαδικασίας Poisson, ενώ η ιδιότητα (iii) ισχύει από το θεώρημα 4.14(ii). Επομένως, ο ανανεωτικός ορισμός της στοχαστικής διαδικασίας Poisson συνεπάγεται τον ολικό ορισμό.

Αντίστροφα, έστω ότι η $\{N(t)\}$ ικανοποιεί τον ολικό ορισμό της στοχαστικής διαδικασίας Poisson και έστω X_1, X_2, \dots οι ενδιάμεσοι χρόνοι γεγονότων της $\{N(t)\}$. Θα δικαιολογήσουμε ότι η $\{N(t)\}$ ικανοποιεί τις ιδιότητες του ανανεωτικού ορισμού της στοχαστικής διαδικασίας Poisson. Έχουμε

$$\Pr[X_1 > t] = \Pr[N(t) = 0] = e^{-\lambda t} \frac{(\lambda t)^0}{0!} = e^{-\lambda t}, \quad t \geq 0,$$

που είναι η συνάρτηση επιβίωσης της $\text{Exp}(\lambda)$ κατανομής. Επομένως, η X_1 έχει την $\text{Exp}(\lambda)$ κατανομή. Επίσης, για τη δεσμευμένη συνάρτηση επιβίωσης της X_{k+1} , δοθέντος ότι $X_1 = x_1, X_2 = x_2, \dots, X_k = x_k$, έχουμε

$$\begin{aligned} & \Pr[X_{k+1} > t | X_1 = x_1, X_2 = x_2, \dots, X_k = x_k] \\ &= \Pr[0 \text{ γεγονότα στο } (x_1 + \dots + x_k, x_1 + \dots + x_k + t)] \\ &= \Pr[N(t) = 0] \\ &= e^{-\lambda t}, t \geq 0. \end{aligned} \tag{4.10}$$

Πράγματι, η πρώτη ισότητα γίνεται φανερή μεταφράζοντας τα ενδεχόμενα που αφορούν τις X_1, X_2, \dots, X_{k+1} με όρους της $\{N(t)\}$: Το ενδεχόμενο $\{X_1 = x_1, X_2 = x_2, \dots, X_k = x_k\}$ δίνει την ιστορία της $\{N(t)\}$ στο διάστημα $[0, x_1 + x_2 + \dots + x_k]$. Είναι $N(t) = 0$ για $t \in [0, x_1]$, $N(t) = 1$ για $t \in [x_1, x_1 + x_2]$ κ.ο.κ. Το ενδεχόμενο $\{X_{k+1} > t\}$ είναι ισοδύναμο με το $\{0 \text{ γεγονότα στο } (x_1 + \dots + x_k, x_1 + \dots + x_k + t)\}$, υπό τη δέσμευση. Οπότε, από τις ανεξάρτητες προσauξήσεις της $\{N(t)\}$ (ιδιότητα (i) του ολικού ορισμού) η ιστορία της $\{N(t)\}$ στο διάστημα $[0, x_1 + x_2 + \dots + x_k]$ είναι ανεξάρτητη από το ενδεχόμενο $\{0 \text{ γεγονότα στο } (x_1 + \dots + x_k, x_1 + \dots + x_k + t)\}$ και έχουμε την πρώτη ισότητα. Η δεύτερη ισότητα έπεται άμεσα από την ιδιότητα ομογενών προσauξήσεων της $\{N(t)\}$ (ιδιότητα (ii) του ολικού ορισμού). Τέλος, η τρίτη ισότητα έπεται από το ότι η $N(t)$ έχει την κατανομή $\text{Poisson}(\lambda)$ (ιδιότητα (iii) του ολικού ορισμού). Επομένως, η (4.10) δείχνει ότι η X_{k+1} έχει την $\text{Exp}(\lambda)$ κατανομή και είναι ανεξάρτητη των X_1, X_2, \dots, X_k .

Θα αιτιολογήσουμε, τώρα, την ισοδυναμία του ολικού και του τοπικού ορισμού της στοχαστικής διαδικασίας Poisson. Δεδομένου ότι οι ιδιότητες (i) και (ii) είναι κοινές σε αυτούς τους δύο ορισμούς, επικεντρωνόμαστε στην ιδιότητα (iii).

Έστω, ότι η $\{N(t)\}$ ικανοποιεί τον ολικό ορισμό της στοχαστικής διαδικασίας Poisson. Τότε έχουμε

$$\begin{aligned}\Pr[N(h) = 0] &= e^{-\lambda h} = \sum_{k=0}^{\infty} \frac{(-\lambda h)^k}{k!} = 1 - \lambda h + o(h), \quad h \rightarrow 0^+, \\ \Pr[N(h) = 1] &= e^{-\lambda h} \frac{(\lambda h)^1}{1!} = (1 - \lambda h + o(h))\lambda h = \lambda h + o(h), \quad h \rightarrow 0^+, \\ \Pr[N(h) = n] &= e^{-\lambda h} \frac{(\lambda h)^n}{n!} = o(h), \quad h \rightarrow 0^+, \quad n \geq 2,\end{aligned}$$

που δείχνουν ότι η $\{N(t)\}$ ικανοποιεί την ιδιότητα (iii) του τοπικού ορισμού της στοχαστικής διαδικασίας Poisson. Επομένως ο ολικός ορισμός της διαδικασίας Poisson συνεπάγεται τον τοπικό ορισμό.

Αντίστροφα, έστω ότι η $\{N(t)\}$ ικανοποιεί τον τοπικό ορισμό της στοχαστικής διαδικασίας Poisson. Τότε, ικανοποιεί και τον ορισμό της ως Μαρκοβιανής διαδικασίας συνεχούς χρόνου, οπότε έχουμε τις διαφορικές εξισώσεις Charman-Kolmogorov που είδαμε στο παράδειγμα 4.5. Από την εξίσωση

$$p'_0(t) = -\lambda p_0(t), \quad t \geq 0,$$

μαζί με την αρχική συνθήκη $p_0(0) = 1$ παίρνουμε

$$p_0(t) = e^{-\lambda t}, \quad t \geq 0. \quad (4.11)$$

Για $n \geq 1$, έχουμε την εξίσωση

$$p'_n(t) = -\lambda p_n(t) + \lambda p_{n-1}(t), \quad t \geq 0.$$

Έχουμε

$$\begin{aligned}p'_n(t) + \lambda p_n(t) &= \lambda p_{n-1}(t) \\ \Rightarrow e^{\lambda t} p'_n(t) + \lambda e^{\lambda t} p_n(t) &= \lambda e^{\lambda t} p_{n-1}(t) \\ \Rightarrow (e^{\lambda t} p_n(t))' &= \lambda e^{\lambda t} p_{n-1}(t) \\ \Rightarrow e^{\lambda t} p_n(t) - p_n(0) &= \int_0^t \lambda e^{\lambda u} p_{n-1}(u) du,\end{aligned}$$

η οποία μαζί με την αρχική συνθήκη $p_n(0) = 0$, $n \geq 1$ δίνει

$$p_n(t) = e^{-\lambda t} \int_0^t \lambda e^{\lambda u} p_{n-1}(u) du, \quad t \geq 0. \quad (4.12)$$

Χρησιμοποιώντας, τώρα τις σχέσεις (4.11) και (4.12), αποδεικνύεται εύκολα επαγωγικά ότι $p_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$, $n \geq 0$, δηλαδή ισχύει η ιδιότητα (iii) του ολικού ορισμού της Poisson για τη διαδικασία $\{N(t)\}$. Επομένως, ο τοπικός ορισμός της διαδικασίας Poisson συνεπάγεται τον ολικό ορισμό.

Είδαμε, λοιπόν, ότι οι τέσσερις ορισμοί της στοχαστικής διαδικασίας Poisson είναι ισοδύναμοι και δίνουν κάποιες διαφορετικές όψεις της ιδέας ότι μια στοχαστική διαδικασία Poisson με ρυθμό λ μοντελοποιεί την εκδήλωση γεγονότων στο χρόνο που συμβαίνουν «εντελώς τυχαία» και «ομογενώς» με ρυθμό λ . Το «εντελώς τυχαία» αποδίδεται στον ανανεωτικό ορισμό με την απαίτηση οι χρόνοι μεταξύ των γεγονότων να είναι ανεξάρτητοι και εκθετικοί. Λόγω της αμνήμονης ιδιότητας της εκθετικής κατανομής, αυτό σημαίνει ότι δοθείσης της μέχρι τώρα ιστορίας της διαδικασίας, ο χρόνος που απομένει ως το επόμενο γεγονός είναι εκθετικός με παράμετρο λ . Δηλαδή, το παρελθόν της διαδικασίας δεν προσφέρει κάποια πληροφορία για το μέλλον της. Αυτό άλλωστε είναι και η ουσία του ορισμού της Poisson ως Μαρκοβιανής αλυσίδας συνεχούς χρόνου. Ανάλογα το «εντελώς τυχαία» αποδίδεται στους άλλους δυο ορισμούς με την ιδιότητα των ανεξάρτητων προσαυξήσεων. Η ομογένεια της διαδικασίας στο χρόνο αποδίδεται στον ανανεωτικό ορισμό με το ότι οι χρόνοι μεταξύ των

γεγονότων είναι ισόνομοι, στον Μαρκοβιανό ορισμό με την ισότητα όλων των ρυθμών μετάβασης, ενώ στους άλλους δυο ορισμούς με την ιδιότητα των ομογενών προσauξήσεων.

Πρέπει, εδώ, να τονιστεί ότι η στοχαστική διαδικασία Poisson έχει μια πιο ειδική δομή από μια αυθαίρετη ανανεωτική διαδικασία. Η δομή αυτή επάγει επιπλέον ιδιότητες. Έτσι, π.χ., μια αυθαίρετη ανανεωτική διαδικασία δεν έχει την ιδιότητα των ανεξάρτητων, ούτε των ομογενών προσauξήσεων. Δεν υπάρχουν, δηλαδή, αντίστοιχοι των τελευταίων δυο ορισμών της διαδικασίας Poisson για ανανεωτικές διαδικασίες.

Ένα άλλο σημαντικό αποτέλεσμα που αφορά τη διαδικασία Poisson και δεν έχει αντίστοιχο στις άλλες ανανεωτικές διαδικασίες είναι το παρακάτω αποτέλεσμα.

Θεώρημα 4.17 (Δεσμευμένη κατανομή χρόνων γεγονότων διαδικασίας Poisson) Η από κοινού κατανομή των χρόνων των γεγονότων S_1, S_2, \dots, S_n μιας στοχαστικής διαδικασίας Poisson $\{N(t)\}$, δεδομένου ότι $N(t) = n$, ισούται με την από κοινού κατανομή των διατεταγμένων τυχαίων μεταβλητών από n ανεξάρτητες και ισόνομες τυχαίες μεταβλητές, ομοιόμορφες στο $(0, t]$. Πιο συγκεκριμένα η δεσμευμένη συνάρτηση πυκνότητας πιθανότητας της (S_1, S_2, \dots, S_n) , δεδομένου ότι $N(t) = n$, δίνεται ως

$$f_{(S_1, S_2, \dots, S_n | N(t)=n)}(s_1, s_2, \dots, s_n) = \begin{cases} \frac{n!}{t^n}, & 0 < s_1 < s_2 < \dots < s_n \leq t, \\ 0, & \text{διαφορετικά.} \end{cases} \quad (4.13)$$

◊

Πράγματι, για $0 < s_1 < s_2 < \dots < s_n \leq t$, έχουμε ότι

$$\begin{aligned} & f_{(S_1, S_2, \dots, S_n | N(t)=n)}(s_1, s_2, \dots, s_n) \\ &= \lim_{h_1, h_2, \dots, h_n \rightarrow 0^+} \frac{\Pr[s_1 < S_1 \leq s_1 + h_1, \dots, s_n < S_n \leq s_n + h_n | N(t) = n]}{h_1 h_2 \dots h_n} \\ &= \lim_{h_1, h_2, \dots, h_n \rightarrow 0^+} \frac{\Pr[s_1 < S_1 \leq s_1 + h_1, \dots, s_n < S_n \leq s_n + h_n, N(t) = n]}{h_1 h_2 \dots h_n \Pr[N(t) = n]}. \end{aligned} \quad (4.14)$$

Όμως, το ενδεχόμενο $\{s_1 < S_1 \leq s_1 + h_1, \dots, s_n < S_n \leq s_n + h_n, N(t) = n\}$ είναι ισοδύναμο με το ότι συμβαίνουν 0 γεγονότα στο $(0, s_1]$, 1 γεγονός στο $(s_1, s_1 + h_1]$, 0 γεγονότα στο $(s_1 + h_1, s_2]$, 1 γεγονός στο $(s_2, s_2 + h_2]$, 0 γεγονότα στο $(s_2 + h_2, s_3]$, κ.ο.κ., 1 γεγονός στο $(s_n, s_n + h_n]$, 0 γεγονότα στο $(s_n + h_n, t]$. Επομένως, λόγω των ανεξάρτητων και ομογενών προσauξήσεων, έχουμε ότι

$$\begin{aligned} & \Pr[s_1 < S_1 \leq s_1 + h_1, \dots, s_n < S_n \leq s_n + h_n, N(t) = n] \\ &= \Pr[N(s_1) = 0] \Pr[N(h_1) = 1] \Pr[N(s_2 - s_1 - h_1) = 0] \Pr[N(h_2) = 1] \\ & \quad \dots \Pr[N(h_n) = 1] \Pr[N(t - s_n - h_n) = 0] \\ &= e^{-\lambda s_1} \lambda h_1 e^{-\lambda h_1} e^{-\lambda(s_2 - s_1 - h_1)} \lambda h_2 e^{-\lambda h_2} \dots \lambda h_n e^{-\lambda h_n} e^{-\lambda(t - s_n - h_n)} \\ &= \lambda^n e^{-\lambda t} h_1 h_2 \dots h_n. \end{aligned}$$

Οπότε η (4.14) γίνεται

$$f_{(S_1, S_2, \dots, S_n | N(t)=n)}(s_1, s_2, \dots, s_n) = \lim_{h_1, h_2, \dots, h_n \rightarrow 0^+} \frac{\lambda^n e^{-\lambda t} h_1 h_2 \dots h_n}{h_1 h_2 \dots h_n e^{-\lambda t} \frac{(\lambda t)^n}{n!}}$$

που δίνει τον πρώτο κλάδο της (4.13). Ο δεύτερος κλάδος είναι βέβαια προφανής.

Το θεώρημα 4.17 μας επιτρέπει υπολογισμούς δεσμευμένων πιθανοτήτων της μορφής $\Pr[(S_1, S_2, \dots, S_n) \in A | N(t) = n]$, καθώς και δεσμευμένων μέσων τιμών της μορφής $E[g(S_1, S_2, \dots, S_n) | N(t) = n]$. Πράγματι, αν συμβολίσουμε με

$$(U_{1:n}, U_{2:n}, \dots, U_{n:n})$$

το διάνυσμα των διατεταγμένων παρατηρήσεων από τις ανεξάρτητες και ισόνομες τυχαίες μεταβλητές U_i , $i = 1, 2, \dots, n$, με ομοιόμορφη κατανομή στο $(0, t]$, τότε

$$\begin{aligned}\Pr[(S_1, S_2, \dots, S_n) \in A | N(t) = n] &= \Pr[(U_{1:n}, U_{2:n}, \dots, U_{n:n}) \in A], \\ E[g(S_1, S_2, \dots, S_n) | N(t) = n] &= E[g(U_{1:n}, U_{2:n}, \dots, U_{n:n})].\end{aligned}$$

Ειδικότερα, για τη δεσμευμένη συνάρτηση κατανομής, τη δεσμευμένη συνάρτηση πυκνότητας πιθανότητας και τη δεσμευμένη μέση τιμή της S_i , δεδομένου ότι $N(t) = n$, για $i = 1, 2, \dots, n$, έχουμε αντίστοιχα

$$\begin{aligned}F_{(S_i | N(t)=n)}(s_i) &= F_{U_{i:n}}(s_i) = \sum_{k=i}^n \binom{n}{k} \left(\frac{s_i}{t}\right)^k \left(1 - \frac{s_i}{t}\right)^{n-k}, \quad 0 < s_i \leq t, \\ f_{(S_i | N(t)=n)}(s_i) &= f_{U_{i:n}}(s_i) = \frac{n!}{(i-1)!(n-i)!} \left(\frac{s_i}{t}\right)^{i-1} \frac{1}{t} \left(1 - \frac{s_i}{t}\right)^{n-i}, \quad 0 < s_i \leq t, \\ E[S_i | N(t) = n] &= \frac{it}{n+1}, \quad 1 \leq i \leq n.\end{aligned}$$

Όταν έχουμε απαριθμήτριες διαδικασίες διαφόρων τύπων γεγονότων, μπορούμε να δημιουργήσουμε την απαριθμήτρια διαδικασία όλων των γεγονότων που αναφέρεται ως η υπέρθεσή τους. Αντίστροφα, από μια απαριθμήτρια διαδικασία μπορούμε να δημιουργήσουμε τη διάσπασή της σε απαριθμήτριες γεγονότων συγκεκριμένων τύπων. Κάτω, από κάποιες προϋποθέσεις οι δυο αυτές διαδικασίες, της υπέρθεσης και της διάσπασης, διατηρούν την ιδιότητα μιας απαριθμήτριας διαδικασίας να είναι Poisson. Συγκεκριμένα έχουμε τα ακόλουθα δυο αποτελέσματα.

Θεώρημα 4.18 (Υπέρθεση (superposition) διαδικασιών Poisson) Αν $\{N_1(t)\}, \{N_2(t)\}, \dots, \{N_r(t)\}$ είναι ανεξάρτητες διαδικασίες Poisson με ρυθμούς $\lambda_1, \lambda_2, \dots, \lambda_r$ και η $\{N(t)\}$ είναι η υπέρθεσή τους, με $N(t) = \sum_{j=1}^r N_j(t)$, τότε η $\{N(t)\}$ είναι στοχαστική διαδικασία Poisson με ρυθμό $\lambda = \sum_{i=1}^r \lambda_i$. Επιπλέον, έστω Z_k ο τύπος του k -οστού γεγονότος της υπέρθεσης, δηλαδή το ενδεχόμενο $\{Z_k = i\}$ αντιστοιχεί στο ότι το k -οστό γεγονός της $\{N(t)\}$ προέρχεται από γεγονός της $\{N_i(t)\}$. Τότε, οι Z_k , $k \geq 1$ είναι ανεξάρτητες, ισόνομες και επιπλέον $\Pr[Z_k = i] = \frac{\lambda_i}{\lambda}$, $i = 1, 2, \dots, r$. \diamond

Θεώρημα 4.19 (Διάσπαση (splitting) διαδικασιών Poisson) Έστω $\{N(t)\}$ διαδικασία Poisson ρυθμού λ και Z_1, Z_2, \dots ανεξάρτητες και ισόνομες τυχαίες μεταβλητές, ανεξάρτητες της $\{N(t)\}$, με τιμές στο $\{1, 2, \dots, r\}$ και $\Pr[Z_k = i] = p_i$, $1 \leq i \leq r$. Έστω, επίσης $N_i(t) = \sum_{k=1}^{N(t)} 1_{\{Z_k=i\}}$, $t \geq 0$, $1 \leq i \leq r$. Οι διαδικασίες $\{N_i(t)\}$, $1 \leq i \leq r$ αποτελούν μια τυχαία διάσπαση της $\{N(t)\}$ και είναι ανεξάρτητες διαδικασίες Poisson με αντίστοιχους ρυθμούς λp_i , $1 \leq i \leq r$. \diamond

Τα Θεωρήματα 4.18 και 4.19 αποδεικνύονται χρησιμοποιώντας οποιονδήποτε από τους ισοδύναμους ορισμούς της Poisson. Ειδικότερα με τον τοπικό ορισμό της Poisson προκύπτουν ιδιαίτερα εύκολες αποδείξεις. Δεν υπάρχουν ανάλογα αυτών των θεωρημάτων για γενικές ανανεωτικές διαδικασίες.

4.5 Ασκήσεις

Άσκηση 4.1 Θεωρούμε την τροποποίηση της $M/M/c/c$ ουράς με Poisson διαδικασία αφίξεων ρυθμού λ και $\text{Exp}(\mu)$ χρόνους εξυπηρέτησης, όπου ορισμένοι πελάτες αποχωρούν από το σύστημα αμέσως μόλις αφιχθούν σε αυτό χωρίς να εξυπηρετηθούν (αποθαρρυνόμενοι πελάτες). Συγκεκριμένα, κάθε πελάτης που βρίσκει n άλλα άτομα στο σύστημα κατά την άφιξή του αναχωρεί με πιθανότητα $\frac{n}{c}$, $n = 0, 1, 2, \dots, c$. Να δικαιολογηθεί ότι η διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου, να δοθούν οι ρυθμοί μετάβασης της (είτε αναλυτικά, είτε με τον πίνακα των ρυθμών μετάβασης, είτε με το διάγραμμα ρυθμών μετάβασης), να γραφούν οι εξισώσεις Charman-Kolmogorov για τη μεταβατική κατανομή της και οι εξισώσεις ισορροπίας για την κατανομή ισορροπίας της.

Άσκηση 4.2 Θεωρούμε την τροποποίηση της $M/M/1$ ουράς με ρυθμό αφίξεων λ και εκθετικούς χρόνους εξυπηρέτησης $\text{Exp}(\mu)$, όπου κάθε πελάτης που συμπληρώνει την εξυπηρέτησή του δεν μένει ικανοποιημένος από αυτή με πιθανότητα q ($0 < q < 1$) και την επαναλαμβάνει ευθύς αμέσως κ.ο.κ. Διαδοχικές επαναλήψεις έχουν ανεξάρτητες χρονικές διάρκειες. Να δικαιολογηθεί ότι η διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου, να δοθούν οι ρυθμοί μετάβασής της (είτε αναλυτικά, είτε με τον πίνακα των ρυθμών μετάβασης, είτε με το διάγραμμα ρυθμών μετάβασης), να γραφούν οι εξισώσεις Charman-Kolmogorov για τη μεταβατική κατανομή της και οι εξισώσεις ισορροπίας για την κατανομή ισορροπίας της.

Άσκηση 4.3 Θεωρούμε μια μηχανή η οποία εξυπηρετεί εργασίες σύμφωνα με τη σειρά άφιξής τους (FCFS). Λόγω υψηλών λειτουργικών εξόδων, η μηχανή απενεργοποιείται μόλις το σύστημα αδειάσει. Όταν φθάνει μια νέα εργασία, η μηχανή τίθεται εκ νέου σε λειτουργία, αλλά χρειάζεται κάποιο χρόνο επανεκκίνησης (setup time). Υποθέτουμε ότι οι εργασίες φθάνουν στο σύστημα σύμφωνα με μια διαδικασία Poisson με ρυθμό λ , οι χρόνοι εξυπηρέτησης είναι $\text{Exp}(\mu)$ και υπάρχει ένας υπηρέτης. Οι χρόνοι επανεκκίνησης της μηχανής είναι $\text{Exp}(\theta)$. Θα αναφέρουμε το παραπάνω σύστημα ως η $M/M/1/k$ ουρά με χρόνους επανεκκίνησης, όταν η χωρητικότητα του συστήματος είναι k . Έστω η στοχαστική διαδικασία $\{Q(t), I(t)\}$, όπου $Q(t)$ είναι το πλήθος των πελατών τη στιγμή t και $I(t)$ η κατάσταση της μηχανής (1 αν εξυπηρετεί και 0 αν είναι απενεργοποιημένη).

1. Να δικαιολογηθεί ότι η διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα δεν είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου.
2. Να δικαιολογηθεί ότι η διαδικασία $\{(Q(t), I(t))\}$ είναι Μαρκοβιανή διαδικασία συνεχούς χρόνου.
3. Να δοθούν οι ρυθμοί μετάβασής της $\{(Q(t), I(t))\}$ για $k = 1$ και $k = \infty$ ($M/M/1/1$ και $M/M/1$ ουρές με χρόνους επανεκκίνησης), να γραφούν οι εξισώσεις Charman-Kolmogorov για τη μεταβατική κατανομή της και οι εξισώσεις ισορροπίας για την κατανομή ισορροπίας της.

Άσκηση 4.4 Έστω $\{N(t) : t \geq 0\}$ μια διαδικασία Poisson με ρυθμό λ . Να υπολογιστεί η μέση τιμή

$$E[N(t)(N(t) - 1)(N(t) - 2) \cdots (N(t) - k + 1)],$$

και η συνδιακύμανση

$$\text{Cov}[N(t), N(s)].$$

Άσκηση 4.5 Θεωρούμε $\{N_1(t) : t \geq 0\}$ και $\{N_2(t) : t \geq 0\}$ δυο ανεξάρτητες διαδικασίες Poisson με ρυθμούς λ_1 και λ_2 αντίστοιχα. Έστω A_i να είναι ο αριθμός των γεγονότων στη διαδικασία $\{N_i(t)\}$ πριν το πρώτο γεγονός στην άλλη διαδικασία, $i = 1, 2$.

1. Να υπολογιστούν οι συναρτήσεις πιθανότητας των A_i , $i = 1, 2$.
2. Να εξεταστεί αν οι A_1 και A_2 είναι ανεξάρτητες.

Άσκηση 4.6 Έστω μια διαδικασία Poisson με ρυθμό λ , $0 < s < t$ και n μη αρνητικός ακέραιος. Να υπολογιστούν οι πιθανότητες $\Pr[N(s) = k | N(t) = n]$, $0 \leq k \leq n$. Τι κατανομή είναι η δεσμευμένη κατανομή της $N(s)$ δεδομένου του ότι $N(t) = n$; Μπορείτε να ερμηνεύσετε διαισθητικά το αποτέλεσμα;

Άσκηση 4.7 Έστω δυο ανεξάρτητες διαδικασίες Poisson, $\{N_1(t)\}$ και $\{N_2(t)\}$, με ρυθμούς λ_1 και λ_2 , αντίστοιχα. Έστω, επίσης $\{N(t)\}$ η υπέρθεσή τους. Να υπολογιστούν οι πιθανότητες $\Pr[N_1(t) = k | N(t) = n]$, $0 \leq k \leq n$. Τι κατανομή είναι η δεσμευμένη κατανομή της $N_1(t)$ δεδομένου του ότι $N(t) = n$; Μπορείτε να ερμηνεύσετε διαισθητικά το αποτέλεσμα;

Άσκηση 4.8 Έστω $\{N(t)\}$ μια διαδικασία Poisson με ρυθμό λ και X μια τυχαία μεταβλητή, ανεξάρτητη της $\{N(t)\}$, με κατανομή $\text{Exp}(\mu)$. Έστω N το πλήθος των γεγονότων της $\{N(t)\}$ στο (τυχαίο) διάστημα $[0, X]$. Να υπολογιστεί η συνάρτηση πιθανότητας της N . Τι κατανομή είναι;

Άσκηση 4.9 Έστω $\{N(t)\}$ μια διαδικασία Poisson με ρυθμό λ και S_1, S_2, \dots οι χρόνοι των γεγονότων της. Να υπολογιστεί ο μέσος χρόνος του τελευταίου γεγονότος πριν τη στιγμή t , δηλαδή η $E[S_{N(t)}]$.

Άσκηση 4.10 Έστω $\{N(t) : t \geq 0\}$ μια διαδικασία Poisson με ρυθμό λ και S_1 ο χρόνος του πρώτου γεγονότος της. Να υπολογιστεί η δεσμευμένη μέση τιμή $E[S_1 | N(t) \geq 1]$.

4.6 Σχόλια

Τα εισαγωγικά βιβλία στοχαστικών διαδικασιών αναφέρονται στις Μαρκοβιανές αλυσίδες συνεχούς χρόνου που είναι το αντικείμενο του παρόντος κεφαλαίου. Για μια εκτενή εισαγωγή με πολλά παραδείγματα ο ενδιαφερόμενος αναγνώστης παραπέμπεται στα εξαιρετικά συγγράμματα των Ross 1995, Kao 1997, Tijms 2008, Kulkarni 2010 και Grimmett και Stirzaker 2020.

Βιβλιογραφία

- [1] S. Ross. *Stochastic Processes, 2nd Edition*. Wiley, 1995. ISBN: 978-0471120629.
- [2] E.P.C. Kao. *An Introduction to Stochastic Processes*. New York: Duxbury Press, 1997. ISBN: 978-0534255183.
- [3] H.C. Tijms. *A First Course in Stochastic Models, 2nd Edition*. Chichester: Wiley, 2008. ISBN: 978-0471498803.
- [4] V.G. Kulkarni. *Modeling and Analysis of Stochastic Systems, 2nd Edition*. CRC Press, 2010. ISBN: 978-1439808757.
- [5] G. Grimmett και D. Stirzaker. *Probability and Random Processes, 4th Edition*. Oxford, 2020. ISBN: 978-0198847595.
- [6] S. Elaydi. *An Introduction to Difference Equations, 3rd Edition*. Springer, 2006. ISBN: 978-0387230597.

ΚΕΦΑΛΑΙΟ 5

ΑΝΑΛΥΣΗ ΜΕΣΗΣ ΤΙΜΗΣ ΓΙΑ ΜΑΡΚΟΒΙΑΝΑ ΜΟΝΤΕΛΑ

Η ιδιότητα PASTA σε συνδυασμό με τον νόμο του Little μπορούν να χρησιμοποιηθούν για να υπολογίσουμε με πιθανοθεωρητικούς συλλογισμούς και ελάχιστους υπολογισμούς τον μέσο αριθμό πελατών σε ένα σύστημα εξυπηρέτησης, $E[Q]$, και τον μέσο χρόνο παραμονής ενός πελάτη σε αυτό, $E[S]$, για αρκετά συστήματα, χωρίς να χρειαστεί να υπολογίσουμε τις αντίστοιχες κατανομές. Η ανάλυση αυτή αναφέρεται συχνά ως ανάλυση μέσης τιμής (mean value analysis - MVA) και είναι ένα απλό και κομψό εργαλείο μελέτης συστημάτων εξυπηρέτησης. Στο παρόν κεφάλαιο θα την παρουσιάσουμε στο πλαίσιο Μαρκοβιανών μοντέλων, αναλύοντας μια σειρά από συγκεκριμένα συστήματα με τη μέθοδο αυτή.

Η γενική ιδέα της ανάλυσης μέσης τιμής, είναι ότι για τον προσδιορισμό των δύο άγνωστων ποσοτήτων $E[Q]$ και $E[S]$ χρειάζονται δυο εξισώσεις. Η μία είναι ο νόμος του Little και η δεύτερη προκύπτει υπολογίζοντας τον μέσο χρόνο παραμονής ενός πελάτη, $E[S]$, δεσμεύοντας στο πλήθος των πελατών που βρίσκει κατά την άφιξή του, δηλαδή δεσμεύοντας στην Q^- που έχει την κατανομή (a_j) , για την οποία μιλήσαμε στην παράγραφο 1.3. Επομένως, η δεύτερη σχέση εκφράζει το $E[S]$ συναρτήσει της $E[Q^-]$. Στην περίπτωση συστημάτων με Poisson διαδικασία αφίξεων, η κατανομή του πλήθους των πελατών σε στιγμές αφίξεων συμπίπτει με την κατανομή του πλήθους των πελατών σε συνεχή χρόνο, δηλαδή οι Q^- και Q είναι ισόνομες (ιδιότητα PASTA), οπότε η δεύτερη σχέση εκφράζει τελικά το $E[S]$ συναρτήσει του $E[Q]$.

5.1 Ανάλυση μέσης τιμής στην M/M/1/1 ουρά

Θεωρούμε μια M/M/1/1 ουρά με ρυθμό αφίξεων λ και $\text{Exp}(\mu)$ κατανομή χρόνων εξυπηρέτησης. Ο αριθμός των πελατών στο σύστημα αυτό είναι Μαρκοβιανή αλυσίδα της οποίας το διάγραμμα ρυθμών μετάβασης δίνεται στο σχήμα 4.3. Επιθυμούμε να υπολογίσουμε τις $E[Q]$ και $E[S]$.

Από τον νόμο του Little έχουμε

$$E[Q] = \lambda E[S].$$

Για τον μέσο χρόνο παραμονής ενός πελάτη δεσμεύοντας στον αριθμό των πελατών Q^- που βλέπει ο πελάτης

κατά την άφιξή του έχουμε

$$\begin{aligned} E[S] &= \Pr[Q^- = 0]E[S|Q^- = 0] + \Pr[Q^- = 1]E[S|Q^- = 1] \\ &= a_0 \cdot \frac{1}{\mu} + a_1 \cdot 0 = \frac{a_0}{\mu} = \frac{p_0}{\mu}, \end{aligned}$$

όπου η τελευταία ισότητα ισχύει λόγω της PASTA. Συνδυάζοντας αυτές τις σχέσεις και λαμβάνοντας υπόψη ότι $E[Q] = 0 \cdot p_0 + 1 \cdot p_1 = p_1$, έχουμε

$$p_1 = \lambda \frac{p_0}{\mu} = \rho p_0,$$

όπου $\rho = \frac{\lambda}{\mu}$ είναι ο ρυθμός συνωστισμού. Επίσης, το σύστημα είναι ευσταθές αφού έχει πεπερασμένη χωρητικότητα και άρα

$$p_0 + p_1 = 1.$$

Λύνοντας το σύστημα παίρνουμε

$$p_0 = \frac{1}{1 + \rho}, \quad p_1 = \frac{\rho}{1 + \rho},$$

οπότε

$$E[Q] = \frac{\rho}{1 + \rho}, \quad E[S] = \frac{1}{\mu(1 + \rho)}. \quad (5.1)$$

5.2 Ανάλυση μέσης τιμής στην M/M/1 ουρά

Θεωρούμε μια M/M/1 ουρά με ρυθμό αφίξεων λ και $\text{Exp}(\mu)$ κατανομή χρόνων εξυπηρέτησης. Ο αριθμός των πελατών στο σύστημα αυτό είναι Μαρκοβιανή αλυσίδα της οποίας το διάγραμμα ρυθμών μετάβασης δίνεται στο σχήμα 4.4. Επιθυμούμε να υπολογίσουμε τις $E[Q]$ και $E[S]$.

Από τον νόμο του Little έχουμε

$$E[Q] = \lambda E[S].$$

Για τον μέσο χρόνο παραμονής ενός πελάτη, δεσμεύοντας στον αριθμό των πελατών Q^- που βλέπει ο πελάτης κατά την άφιξή του, έχουμε

$$E[S] = E[E[S|Q^-]] = E\left[\sum_{i=1}^{Q^-+1} B_i\right],$$

όπου B_1 είναι ο υπολειπόμενος χρόνος εξυπηρέτησης του πελάτη που εξυπηρετείται (αν εξυπηρετείται κάποιος), B_2, B_3, \dots, B_{Q^-} οι χρόνοι εξυπηρέτησης των πελατών στον χώρο αναμονής (αν υπάρχουν) και B_{Q^-+1} ο χρόνος εξυπηρέτησης του αφικνούμενου πελάτη. Ο χρόνος B_1 έχει και αυτός την $\text{Exp}(\mu)$ κατανομή, λόγω της αμνήμονης ιδιότητας της εκθετικής κατανομής. Έχουμε, λοιπόν, ότι όλες οι τυχαίες μεταβλητές B_i , $i = 1, 2, \dots, Q^- + 1$ είναι ανεξάρτητες και ισόνομες και επιπλέον η $Q^- + 1$ είναι ανεξάρτητη των B_i (διότι οι χρόνοι εξυπηρέτησης B_i δεν έχουν παίξει κάποιο ρόλο στον καθορισμό της Q^- που εξαρτάται από παρελθόντες χρόνους εξυπηρέτησης μόνο). Επομένως, χρησιμοποιώντας τον τύπο (3.3) που αποδείχθηκε στο παράδειγμα 3.3, έχουμε

$$E[S] = E[Q^- + 1]E[B_1] = \frac{E[Q^-] + 1}{\mu} = \frac{E[Q] + 1}{\mu},$$

όπου η τελευταία ισότητα ισχύει λόγω της ιδιότητας PASTA.

Λύνοντας το σύστημα των $E[S]$ και $E[Q]$ προκύπτει ότι

$$E[Q] = \frac{\rho}{1 - \rho}, \quad E[S] = \frac{1}{\mu(1 - \rho)}, \quad (5.2)$$

όπου $\rho = \frac{\lambda}{\mu}$ είναι ο ρυθμός συνωστισμού.

5.3 Ανάλυση μέσης τιμής στην M/M/1 ουρά με χρόνους επανεκκίνησης

Θεωρούμε μια M/M/1 ουρά με ρυθμό αφίξεων λ και $\text{Exp}(\mu)$ κατανομή χρόνων εξυπηρέτησης. Επιπλέον, υποθέτουμε ότι ο υπηρέτης απενεργοποιείται ακαριαία (περνά σε κατάσταση αναμονής (standby)) κάθε φορά που αναχωρεί ένας πελάτης που αφήνει κενό το σύστημα. Κατόπιν, με την άφιξη ενός πελάτη που βρίσκει το σύστημα κενό, αρχίζει ένας χρόνος επανεκκίνησης του συστήματος (setup time), που ακολουθεί την εκθετική κατανομή με παράμετρο θ . Για το συγκεκριμένο σύστημα, θέλουμε να υπολογίσουμε τις ποσότητες $E[Q]$ και $E[S]$.

Από τον νόμο του Little έχουμε

$$E[Q] = \lambda E[S]. \quad (5.3)$$

Έστω I η τυχαία μεταβλητή που αντιστοιχεί στην κατάσταση του υπηρέτη:

$$I = \begin{cases} 1 & \text{αν ο υπηρέτης παρέχει εξυπηρέτηση,} \\ 0 & \text{αν ο υπηρέτης δεν παρέχει εξυπηρέτηση.} \end{cases}$$

Τότε δεσμεύοντας στην κατάσταση του υπηρέτη που βρίσκει ένας αφικνούμενος πελάτης, I^- , έχουμε τη σχέση

$$E[S] = \Pr[I^- = 0]E[S|I^- = 0] + \Pr[I^- = 1]E[S|I^- = 1]. \quad (5.4)$$

Θα υπολογίσουμε τώρα τις ποσότητες $\Pr[I^- = 0]$, $\Pr[I^- = 1]$, $E[S|I^- = 0]$ και $E[S|I^- = 1]$ συναρτήσει των παραμέτρων του συστήματος.

Για τις $\Pr[I^- = 0]$, $\Pr[I^- = 1]$ εφαρμόζουμε τον νόμο του Little στο χώρο εξυπηρέτησης, θεωρώντας ότι ένας πελάτης βρίσκεται στο χώρο εξυπηρέτησης μόνο εφόσον του παρέχεται εξυπηρέτηση. Έτσι, έχουμε τη σχέση (2.7), που στη συγκεκριμένη περίπτωση γίνεται

$$\Pr[I = 1] = E[I] = E[Q_s] = \lambda E[B] = \lambda \frac{1}{\mu} = \rho,$$

και επομένως $\Pr[I = 0] = 1 - \rho$. Εφαρμόζοντας την γενικευμένη ιδιότητα PASTA (βλέπε θεώρημα 2.5) έχουμε ότι οι πιθανότητες αυτές είναι οι ίδιες με τις αντίστοιχες σε στιγμές αφίξεων πελατών, οπότε

$$\Pr[I^- = 0] = 1 - \rho, \quad (5.5)$$

$$\Pr[I^- = 1] = \rho. \quad (5.6)$$

Για την $E[S|I^- = 0]$, ας θεωρήσουμε έναν αφικνούμενο πελάτη που βρίσκει j πελάτες στο σύστημα κατά την άφιξή του, δεδομένου ότι ο υπηρέτης δεν παρέχει εξυπηρέτηση, τότε αυτός θα πρέπει να περιμένει να περάσει ο υπολειπόμενος χρόνος επανεκκίνησης ώστε να αρχίσει να εξυπηρετεί ο υπηρέτης. Λόγω της αμνήμονης ιδιότητας των χρόνων επανεκκίνησης (αφού έχουν την εκθετική κατανομή), ο μέσος υπολειπόμενος χρόνος επανεκκίνησης είναι $\frac{1}{\theta}$. Κατόπιν θα απαιτηθούν $j + 1$ ολόκληροι χρόνοι εξυπηρέτησης μέχρι να εξυπηρετηθεί ο αφικνούμενος πελάτης, οπότε ο μέσος απαιτούμενος χρόνος είναι $(j + 1)\frac{1}{\mu}$. Επομένως, έχουμε ότι

$$E[S|I^- = 0] = \frac{1}{\theta} + (E[Q^-|I^- = 0] + 1)\frac{1}{\mu}. \quad (5.7)$$

Για την $E[S|I^- = 1]$, έχουμε ομοίως ότι

$$E[S|I^- = 1] = (E[Q^-|I^- = 1] + 1)\frac{1}{\mu}. \quad (5.8)$$

Πράγματι, στην περίπτωση αυτή, ένας αφικνούμενος πελάτης που βρίσκει j πελάτες στο σύστημα κατά την άφιξή του και τον υπηρέτη να παρέχει εξυπηρέτηση, θα μείνει στο σύστημα για $j + 1$ χρόνους εξυπηρέτησης,

έναν υπολειπόμενο και j ολόκληρους. Όμως, λόγω της αμνήμονης ιδιότητας της εκθετικής κατανομής ο υπολειπόμενος χρόνος εξυπηρέτησης του πελάτη που εξυπηρετείται είναι σαν να ξεκινάει εκείνη τη στιγμή και άρα ο μέσος απαιτούμενος χρόνος παραμονής για την εξυπηρέτηση του αφικνούμενου πελάτη είναι $(j+1)\frac{1}{\mu}$.

Αντικαθιστώντας, τώρα, τις (5.5)-(5.8) στην (5.4) παίρνουμε

$$\begin{aligned} E[S] &= \Pr[I^- = 0] \left(\frac{1}{\theta} + (E[Q^- | I^- = 0] + 1) \frac{1}{\mu} \right) + \Pr[I^- = 1] (E[Q^- | I^- = 1] + 1) \frac{1}{\mu} \\ &= \Pr[I^- = 0] \frac{1}{\theta} + \frac{\Pr[I^- = 0]E[Q^- | I^- = 0] + \Pr[I^- = 1]E[Q^- | I^- = 1] + 1}{\mu} \\ &= (1 - \rho) \frac{1}{\theta} + \frac{E[Q^-] + 1}{\mu}. \end{aligned} \quad (5.9)$$

Λόγω της PASTA έχουμε ότι $E[Q^-] = E[Q]$. Οπότε η (5.9) γίνεται

$$E[S] = (1 - \rho) \frac{1}{\theta} + \frac{E[Q] + 1}{\mu}. \quad (5.10)$$

Λύνοντας το σύστημα των (5.3) και (5.10) για τις $E[Q]$ και $E[S]$ παίρνουμε

$$E[Q] = \frac{\lambda}{\theta} + \frac{\rho}{1 - \rho}, \quad E[S] = \frac{1}{\theta} + \frac{1}{\mu(1 - \rho)}. \quad (5.11)$$

Το ενδιαφέρον και κάπως απρόσμενο αποτέλεσμα είναι ότι με την εισαγωγή των χρόνων επανεκκίνησης των πελατών, ο μέσος χρόνος παραμονής ενός πελάτη έχει προσ αυξηθεί κατά $\frac{1}{\theta}$, δηλαδή κατά έναν μέσο χρόνο επανεκκίνησης, σε σχέση με το αντίστοιχο σύστημα χωρίς χρόνους επανεκκίνησης (συγκρίνοντας με τον τύπο (5.2) για τον μέσο χρόνο παραμονής της M/M/1 ουράς). Αυτό μοιάζει παράδοξο, καθώς μόνο ένα κλάσμα των πελατών (ίσο με $1 - \rho$) υπομένει τον χρόνο επανεκκίνησης, δηλαδή μόνο οι πελάτες που βρίσκουν τον υπηρέτη απενεργοποιημένο. Όμως, η εισαγωγή των χρόνων επανεκκίνησης επιβαρύνει έμμεσα και τους πελάτες που βρίσκουν τον υπηρέτη ενεργό, καθώς αυξάνει το συνωστισμό στο σύστημα και αυτό αιτιολογεί γιατί τελικά όλοι επιβαρύνονται χρονικά κατά έναν μέσο χρόνο επανεκκίνησης στην παραμονή τους στο σύστημα.

5.4 Ανάλυση μέσης τιμής στην M/M/1 ουρά με την K-πολιτική ενεργοποίησης

Θεωρούμε μια M/M/1 ουρά με ρυθμό αφίξεων λ και $\text{Exp}(\mu)$ κατανομή χρόνων εξυπηρέτησης, όπου μόλις το σύστημα αδειάσει ο υπηρέτης παύει να παρέχει εξυπηρέτηση και αρχίζει και πάλι να εξυπηρετεί, μόλις μαζευτούν K πελάτες με $K \geq 1$ (για $K = 1$ έχουμε την κλασική M/M/1 ουρά). Ενδιαφερόμαστε για τον υπολογισμό των $E[Q]$ και $E[S]$.

Από τον νόμο του Little έχουμε

$$E[Q] = \lambda E[S]. \quad (5.12)$$

Έστω I η τυχαία μεταβλητή που αντιστοιχεί στην κατάσταση του υπηρέτη:

$$I = \begin{cases} 1 & \text{αν ο υπηρέτης παρέχει εξυπηρέτηση,} \\ 0 & \text{αν ο υπηρέτης δεν παρέχει εξυπηρέτηση.} \end{cases}$$

Τότε, δεσμεύοντας στην κατάσταση του υπηρέτη που βρίσκει ένας αφικνούμενος πελάτης, I^- , έχουμε τη σχέση

$$E[S] = \Pr[I^- = 0]E[S|I^- = 0] + \Pr[I^- = 1]E[S|I^- = 1]. \quad (5.13)$$

Θα υπολογίσουμε, τώρα τις ποσότητες $\Pr[I^- = 0]$, $\Pr[I^- = 1]$, $E[S|I^- = 0]$ και $E[S|I^- = 1]$ συναρτήσει των παραμέτρων του συστήματος.

Για τις $\Pr[I^- = 0]$, $\Pr[I^- = 1]$ εφαρμόζουμε τον νόμο του Little στο χώρο εξυπηρέτησης, θεωρώντας ότι ένας πελάτης βρίσκεται στο χώρο εξυπηρέτησης μόνο εφόσον του παρέχεται εξυπηρέτηση. Έτσι, έχουμε τη σχέση (2.7), που στη συγκεκριμένη περίπτωση γίνεται

$$\Pr[I = 1] = E[I] = E[Q_s] = \lambda E[B] = \lambda \frac{1}{\mu} = \rho,$$

και επομένως $\Pr[I = 0] = 1 - \rho$. Εφαρμόζοντας την γενικευμένη ιδιότητα PASTA (βλέπε θεώρημα 2.5) έχουμε ότι οι πιθανότητες αυτές είναι οι ίδιες με τις αντίστοιχες σε στιγμές αφίξεων πελατών, οπότε

$$\Pr[I^- = 0] = 1 - \rho, \quad (5.14)$$

$$\Pr[I^- = 1] = \rho. \quad (5.15)$$

Ουσιαστικά, μέχρι αυτό το σημείο η εφαρμογή της Ανάλυσης Μέσης Τιμής είναι ακριβώς η ίδια με τη διαδικασία που ακολουθήσαμε για την M/M/1 ουρά με χρόνους επανεκκίνησης, στην παράγραφο 5.3. Από εδώ και πέρα, όμως, οι υπολογισμοί διαφοροποιούνται λόγω του διαφορετικού μηχανισμού ενεργοποίησης του υπηρέτη.

Για την $E[S|I^- = 0]$, ας θεωρήσουμε έναν αφικνούμενο πελάτη που βρίσκει j πελάτες στο σύστημα κατά την άφιξή του, δεδομένου ότι ο υπηρέτης δεν παρέχει εξυπηρέτηση, τότε αυτός θα πρέπει να περιμένει $K - j - 1$ αφίξεις ώστε να μαζευτούν οι K πελάτες που απαιτούνται για την ενεργοποίηση του υπηρέτη. Λόγω του ότι ο μέσος χρόνος μεταξύ δυο διαδοχικών αφίξεων είναι $\frac{1}{\lambda}$, ο μέσος χρόνος για την ενεργοποίηση του υπηρέτη είναι $(K - j - 1)\frac{1}{\lambda}$. Κατόπιν θα απαιτηθούν $j + 1$ ολόκληροι χρόνοι εξυπηρέτησης μέχρι να εξυπηρετηθεί ο αφικνούμενος πελάτης, οπότε ο μέσος απαιτούμενος χρόνος είναι $(j + 1)\frac{1}{\mu}$. Επομένως, έχουμε ότι

$$E[S|I^- = 0] = (K - E[Q^-|I^- = 0] - 1)\frac{1}{\lambda} + (E[Q^-|I^- = 0] + 1)\frac{1}{\mu}. \quad (5.16)$$

Για την $E[S|I^- = 1]$, έχουμε ότι

$$E[S|I^- = 1] = (E[Q^-|I^- = 1] + 1)\frac{1}{\mu}. \quad (5.17)$$

Πράγματι, στην περίπτωση αυτή, ένας αφικνούμενος πελάτης που βρίσκει j πελάτες στο σύστημα κατά την άφιξή του και τον υπηρέτη να παρέχει εξυπηρέτηση, θα μείνει στο σύστημα για $j + 1$ χρόνους εξυπηρέτησης, έναν υπολειπόμενο και j ολόκληρους. Όμως, λόγω της αμνήμονης ιδιότητας της εκθετικής κατανομής ο υπολειπόμενος χρόνος εξυπηρέτησης του πελάτη που εξυπηρετείται είναι σαν να ξεκινάει εκείνη τη στιγμή και άρα ο μέσος απαιτούμενος χρόνος παραμονής για την εξυπηρέτηση του αφικνούμενου πελάτη είναι $(j + 1)\frac{1}{\mu}$.

Αντικαθιστώντας, τώρα, τις (5.14)-(5.17) στην (5.13) παίρνουμε

$$E[S] = (1 - \rho) \frac{K - E[Q^-|I^- = 0] - 1}{\lambda} + \frac{E[Q^-] + 1}{\mu}. \quad (5.18)$$

Λόγω της PASTA, έχουμε ότι $E[Q^-] = E[Q]$ και $E[Q^-|I^- = 0] = E[Q|I = 0]$. Για τον υπολογισμό της $E[Q|I = 0]$, έχουμε ότι αν παρατηρούμε το σύστημα μόνο στις περιόδους που ο υπηρέτης είναι ανενεργός ($I = 0$) τότε το πλήθος των πελατών περνάει διαδοχικά από τις καταστάσεις $0, 1, 2, \dots, K - 1, 0, 1, 2, \dots, K - 1, 0, \dots$ και σε κάθε κατάσταση μένει για $\text{Exp}(\lambda)$ χρόνο. Επομένως, η δεσμευμένη κατανομή του πλήθους των πελατών, δεδομένου ότι $I = 0$, είναι διακριτή ομοιόμορφη στο $\{0, 1, 2, \dots, K - 1\}$, οπότε $E[Q|I = 0] = \frac{K-1}{2}$. Οπότε η (5.18) γίνεται

$$E[S] = (1 - \rho) \frac{K-1}{2\lambda} + \frac{E[Q] + 1}{\mu}. \quad (5.19)$$

Λύνοντας το σύστημα των (5.12) και (5.19) για τις $E[Q]$ και $E[S]$ παίρνουμε

$$E[Q] = \frac{K-1}{2} + \frac{\rho}{1-\rho}, \quad E[S] = \frac{K-1}{2\lambda} + \frac{1}{\mu(1-\rho)}. \quad (5.20)$$

Συγκρίνοντας τον τύπο (5.20) για τον μέσο χρόνο παραμονής πελάτη με τον αντίστοιχο τύπο (5.2) για την συνήθη M/M/1 ουρά, παρατηρούμε ότι υπάρχει μια προσαύξηση κατά $\frac{K-1}{2\lambda}$. Η προσαύξηση αυτή ισούται με τον μέσο αριθμό πελατών $\frac{K-1}{2}$ που χρειάζεται να περιμένει ένας πελάτης που βρίσκεται τον υπηρέτη ανενεργό για να ενεργοποιηθεί επί τον μέσο ενδιάμεσο χρόνο αφίξεων $\frac{1}{\lambda}$. Θα περίμενε κανείς να επιβαρύνονται κατά μέσο όρο με αυτήν την ποσότητα μόνο οι πελάτες που βρίσκουν το σύστημα ανενεργό. Όμως, οι υπολογισμοί δείχνουν ότι όλοι οι πελάτες επιβαρύνονται τελικά κατά την ποσότητα αυτή. Αυτό συμβαίνει διότι η συγκεκριμένη πολιτική ενεργοποίησης αυξάνει τον συνωστισμό του συστήματος και συνακόλουθα τον χρόνο παραμονής όλων των πελατών. Παρατηρούμε, επομένως, και πάλι ένα φαινόμενο σαν αυτό που σημειώσαμε στο τέλος της παραγράφου 5.3 για την M/M/1 ουρά με χρόνους επανεκκίνησης.

5.5 Ασκήσεις

Άσκηση 5.1 Θεωρούμε ένα σύστημα εξυπηρέτησης με έναν υπηρέτη, στο οποίο καταφθάνουν πελάτες δύο τύπων 1 και 2, σύμφωνα με δυο ανεξάρτητες διαδικασίες Poisson με ρυθμούς λ_1 και λ_2 αντίστοιχα. Κάθε πελάτης, ανεξαρτήτως τύπου έχει $\text{Exp}(\mu)$ χρόνο εξυπηρέτησης. Οι πελάτες τύπου 1 έχουν απόλυτη προτεραιότητα έναντι των πελατών τύπου 2, δηλαδή όταν υπάρχουν πελάτες τύπου 1 στο σύστημα ο υπηρέτης εξυπηρετεί αυτούς και αρχίζει να εξυπηρετεί πελάτες τύπου 2 μόνο όταν δεν υπάρχουν πελάτες τύπου 1. Επιπλέον, αν ένας πελάτης τύπου 2 εξυπηρετείται και αφιχθεί πελάτης τύπου 1, ο υπηρέτης διακόπτει την εξυπηρέτηση και πηγαίνει να εξυπηρετήσει τον νεοαφιχθέντα πελάτη τύπου 1. Να βρεθούν οι μέσοι οριακοί αριθμοί πελατών τύπων 1 και 2, $E[Q_1]$ και $E[Q_2]$, αντίστοιχα.

Άσκηση 5.2 Να βρείτε τις οριακές κατανομές (p_n) , (a_n) και (d_n) των αριθμών των πελατών σε συνεχή χρόνο, σε στιγμές αφίξεων και σε στιγμές αναχωρήσεων, αντίστοιχα, σε μια ευσταθή M/M/1 ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ , χρησιμοποιώντας τον νόμο του Little και την ιδιότητα PASTA. Για το σκοπό αυτό θεωρήστε ως «σύστημα» την i θέση του συστήματος εξυπηρέτησης για $i = 1, 2, \dots$

Άσκηση 5.3 Θεωρούμε ένα σύστημα εξυπηρέτησης το οποίο διαθέτει έναν υπηρέτη και άπειρη χωρητικότητα. Η διαδικασία αφίξεων είναι Poisson με ρυθμό λ . Οι πελάτες που βρίσκουν το σύστημα κενό εισέρχονται σε αυτό με πιθανότητα p , ενώ όσοι το βρίσκουν μη-κενό εισέρχονται με πιθανότητα q . Οι χρόνοι εξυπηρέτησης πελατών που βρίσκουν το σύστημα κενό είναι $\text{Exp}(\nu)$, ενώ αυτών που βρίσκουν το σύστημα μη-κενό είναι $\text{Exp}(\mu)$. Το σύστημα αυτό αναφέρεται ως η M/M/1 ουρά με αποθαρρυνόμενους πελάτες και ειδικούς πρώτους χρόνους εξυπηρέτησης σε κάθε περίοδο συνεχούς λειτουργίας. Να βρεθεί ο μέσος αριθμός πελατών στο σύστημα και ο μέσος χρόνος παραμονής ενός πελάτη στο σύστημα.

Άσκηση 5.4 Θεωρούμε ένα σύστημα εξυπηρέτησης το οποίο διαθέτει έναν υπηρέτη και άπειρη χωρητικότητα. Η διαδικασία αφίξεων είναι Poisson με ρυθμό λ , ενώ οι χρόνοι εξυπηρέτησης ακολουθούν την κατανομή $\text{Exp}(\mu)$. Ο υπηρέτης εναλλάσσεται μεταξύ περιόδων λειτουργίας και αργίας. Οι χρόνοι λειτουργίας έχουν την κατανομή $\text{Exp}(\zeta)$, ενώ οι χρόνοι αργίας την κατανομή $\text{Exp}(\theta)$. Οι αφίξεις στο σύστημα γίνονται κανονικά, ανεξάρτητα αν ο υπηρέτης είναι σε περίοδο λειτουργίας ή αργίας. Όμως, εξυπηρέτηση παρέχεται μόνο όταν ο υπηρέτης βρίσκεται σε περίοδο λειτουργίας. Το σύστημα αυτό αναφέρεται ως η M/M/1 ουρά με ενεργές και ανενεργές περιόδους εξυπηρέτησεων. Βρείτε τον οριακό μέσο αριθμό πελατών στο σύστημα, $E[Q]$, και τον μέσο χρόνο παραμονής ενός πελάτη σε αυτό, $E[S]$, χρησιμοποιώντας την ανάλυση μέσης τιμής.

Άσκηση 5.5 Θεωρούμε ένα σύστημα εξυπηρέτησης το οποίο διαθέτει έναν υπηρέτη και μηδενικό χώρο αναμονής. Οι πελάτες φθάνουν σύμφωνα με μια διαδικασία Poisson με ρυθμό λ , ενώ οι χρόνοι εξυπηρέτησης των πελατών

είναι $\text{Exp}(\mu)$. Ένας πελάτης που κατά την άφιξή του βρίσκει τον υπηρέτη ελεύθερο αρχίζει άμεσα να εξυπηρετείται. Αν, όμως, βρει τον υπηρέτη κατειλημμένο, τότε αναχωρεί από το σύστημα και προσπαθεί αργότερα μέχρι τελικά να εξυπηρετηθεί. Οι χρόνοι μεταξύ διαδοχικών επαναπροσπαθειών ενός πελάτη έχουν την κατανομή $\text{Exp}(\gamma)$. Το σύστημα αυτό αναφέρεται ως η $M/M/1/1$ ουρά με (εκθετικές) επαναπροσπάθειες. Βρείτε τον οριακό μέσο αριθμό πελατών στο σύστημα, $E[Q]$, και τον μέσο χρόνο παραμονής ενός πελάτη σε αυτό, $E[S]$, χρησιμοποιώντας την ανάλυση μέσης τιμής.

5.6 Σχόλια

Το βιβλίο Adan και Resing 2001 αναλύει πολλά από τα βασικά μοντέλα συστημάτων εξυπηρέτησης με τη μέθοδο της ανάλυσης μέσης τιμής. Για μια διεξοδική περιγραφή της μεθόδου ο ενδιαφερόμενος αναγνώστης μπορεί να διαβάσει την εργασία των Adan και Wal 2011 και τη μεταπτυχιακή διπλωματική εργασία Τσουπαρόπουλος 2015.

Βιβλιογραφία

- [1] I. Adan και J. Resing. *Queueing Theory*. Eindhoven: Notes available online, 2001.
- [2] I. Adan και J. van der Wal. “Mean value techniques”. Στο: Boucherie, R. and van Dijk, N.M. (eds.) *Queueing Networks: A Fundamental Approach*. Springer, 2011. Κεφ. 13, σσ. 561–586. ISBN: 978-1-4419-6471-7. DOI: https://doi.org/10.1007/978-1-4419-6472-4_13.
- [3] Ν. Τσουπαρόπουλος. *Ανάλυση Μέσης Τιμής σε Συστήματα Εξυπηρέτησης*. Αθήνα: Διπλωματική Εργασία, Μεταπτυχιακό Πρόγραμμα Ειδίκευσης στη Στατιστική και Επιχειρησιακή Έρευνα, Τμήμα Μαθηματικών Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών, 2015.

ΚΕΦΑΛΑΙΟ 6

ΑΠΛΕΣ ΜΑΡΚΟΒΙΑΝΕΣ ΟΥΡΕΣ

Στο κεφάλαιο αυτό μελετάμε τις πλέον απλές ουρές, στις οποίες ο αριθμός των πελατών στο σύστημα είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου τύπου γέννησης-θανάτου, δηλαδή σε κάθε χρονική στιγμή το σύστημα όντας σε μια κατάσταση n μπορεί να μεταβεί μόνο στην $n + 1$ (λόγω άφιξης πελάτη) ή στην $n - 1$ (λόγω αναχώρησης). Παρόλο που η μελέτη αυτών των συστημάτων είναι πολύ απλή, η θέση τους στη Θεωρία Ουρών είναι πολύ σημαντική δεδομένης της μεγάλης εφαρμοσιμότητάς τους.

6.1 Ορισμός και βασικοί υπολογισμοί

Ένα σύστημα εξυπηρέτησης αναφέρεται ως απλή Μαρκοβιανή ουρά αν η στοχαστική διαδικασία $\{Q(t)\}$ που καταγράφει τον αριθμό των πελατών στο σύστημα είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου τύπου γέννησης-θανάτου με ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda_i & \text{αν } i \geq 0, j = i + 1, \\ \mu_i & \text{αν } i \geq 1, j = i - 1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Το διάγραμμα ρυθμών μετάβασης μιας τέτοιας διαδικασίας δίνεται στο σχήμα 4.5. Σε ένα τέτοιο σύστημα οι αφίξεις και οι αναχωρήσεις γίνονται μεμονωμένα και επομένως οι μεταβάσεις της Μαρκοβιανής αλυσίδας είναι μόνο προς γειτονικές καταστάσεις. Οι ρυθμοί αφίξεων και αναχωρήσεων είναι λ_j και μ_j αντίστοιχα, όταν υπάρχουν j πελάτες στο σύστημα.

Αν ενδιαφερόμαστε για την εύρεση της μεταβατικής κατανομής του αριθμού των πελατών στο σύστημα, δηλαδή για τις πιθανότητες $p_n(t)$ για συγκεκριμένο $t \geq 0$ και $n = 0, 1, 2, \dots$, τότε θα πρέπει να λύσουμε το σύστημα των εξισώσεων Chapman-Kolmogorov (4.2), που στη συγκεκριμένη περίπτωση παίρνει τη μορφή

$$p'_0(t) = -\lambda_0 p_0(t) + \mu_1 p_1(t), \quad (6.1)$$

$$p'_n(t) = -(\lambda_n + \mu_n) p_n(t) + \lambda_{n-1} p_{n-1}(t) + \mu_{n+1} p_{n+1}(t), \quad n \geq 1. \quad (6.2)$$

Αν για κάποιο k είναι $\lambda_k = 0$, δηλαδή σταματούν οι αφίξεις, π.χ., λόγω περιορισμένης χωρητικότητας συστήματος, τότε οι παραπάνω εξισώσεις ισχύουν για $n = 0, 1, 2, \dots, k-1$, ενώ η εξίσωση για την $p'_k(t)$ γίνεται

$$p'_k(t) = -\mu_k p_k(t) + \lambda_{k-1} p_{k-1}(t).$$

Γενικά, οι εξισώσεις αυτές δεν επιδέχονται λύση σε κλειστή μορφή για την $p_n(t)$, εκτός από λίγες ειδικές περιπτώσεις. Η συνήθης μεθοδολογία είναι να ανάγονται σε αλγεβρικές εξισώσεις παίρνοντας μετασχηματισμούς Laplace. Π.χ., η εξίσωση (6.2) για $n \geq 1$, πολλαπλασιαζόμενη με e^{-st} και παίρνοντας το ολοκλήρωμά της στο $[0, \infty)$ δίνει

$$\begin{aligned} \int_0^{\infty} e^{-st} p'_n(t) dt &= -(\lambda_n + \mu_n) \int_0^{\infty} e^{-st} p_n(t) dt \\ &+ \lambda_{n-1} \int_0^{\infty} e^{-st} p_{n-1}(t) dt + \mu_{n+1} \int_0^{\infty} e^{-st} p_{n+1}(t) dt, \quad n \geq 1. \end{aligned}$$

Έχουμε, όμως, ότι (προσοχή στην ασυνέχεια της $p_n(t)$ στο $t = 0$ για τον υπολογισμό του μετασχηματισμού Laplace-Stieltjes της!):

$$\begin{aligned} \int_0^{\infty} e^{-st} p'_n(t) dt &= \int_0^{\infty} e^{-st} dp_n(t) - p_n(0) = \tilde{p}_n(s) - p_n(0), \\ \int_0^{\infty} e^{-st} p_n(t) dt &= \left[\frac{e^{-st}}{-s} p_n(t) \right]_{t=0}^{\infty} - \int_0^{\infty} \frac{e^{-st}}{-s} p'_n(t) dt = \frac{\tilde{p}_n(s)}{s}, \end{aligned}$$

οπότε η μετασχηματισμένη εξίσωση Chapman-Kolmogorov παίρνει τη μορφή

$$\tilde{p}_n(s) - p_n(0) = -(\lambda_n + \mu_n) \frac{\tilde{p}_n(s)}{s} + \lambda_{n-1} \frac{\tilde{p}_{n-1}(s)}{s} + \mu_{n+1} \frac{\tilde{p}_{n+1}(s)}{s}, \quad n \geq 1.$$

Με απλούς αλγεβρικούς μετασχηματισμούς, βλέπουμε επομένως ότι το μετασχηματισμένο σύστημα των εξισώσεων Chapman-Kolmogorov παίρνει τη μορφή

$$\begin{aligned} -(s + \lambda_0) \tilde{p}_0(s) + \mu_1 \tilde{p}_1(s) &= -s p_0(0), \\ \lambda_{n-1} \tilde{p}_{n-1}(s) - (s + \lambda_n + \mu_n) \tilde{p}_n(s) + \mu_{n+1} \tilde{p}_{n+1}(s) &= -s p_n(0), \quad n \geq 1, \end{aligned}$$

με την επιπλέον εξίσωση

$$\lambda_{k-1} \tilde{p}_{k-1}(s) - (s + \mu_k) \tilde{p}_k(s) = -s p_k(0),$$

για την περίπτωση που υπάρχει κάποιο k με $\lambda_k = 0$. Το σύστημα αυτό είναι ένα τριδιαγώνιο γραμμικό σύστημα εξισώσεων και μπορεί να λυθεί αναλυτικά όταν έχει μικρή διάσταση, αλλιώς μπορεί να λυθεί αποτελεσματικά με αριθμητικές μεθόδους. Στην περίπτωση που το σύστημα είναι πεπερασμένο (δηλαδή όταν υπάρχει κάποιο k με $\lambda_k = 0$), οι μετασχηματισμοί Laplace-Stieltjes $\tilde{p}_n(s)$ είναι ρητές συναρτήσεις του s και μπορούν να αντιστραφούν χρησιμοποιώντας τη μεθοδολογία που έχουμε δει για την αντιστροφή ρητών μετασχηματισμών. Όμως, σε κάθε περίπτωση, η σχετική διαδικασία είναι υπολογιστικά επίπονη και αχρείαστη, καθώς η κατανομή ισορροπίας δίνει τις σημαντικότερες πληροφορίες για τον αριθμό των πελατών και μάλιστα με πολύ μικρό υπολογιστικό κόστος. Για πληρότητα θα παρουσιάσουμε τη διαδικασία εύρεσης της μεταβατικής κατανομής μόνο για λίγες εξαιρετικά απλές Μαρκοβιανές ουρές και θα επικεντρωθούμε στην εύρεση μόνο της κατανομής ισορροπίας για τις άλλες ουρές.

Χρησιμοποιώντας τα αποτελέσματα για τις διαδικασίες γέννησης-θανάτου που είδαμε στην παράγραφο 4.2, συνάγουμε ότι μια απλή Μαρκοβιανή ουρά είναι ευσταθής (δηλαδή η Μαρκοβιανή αλυσίδα του αριθμού των πελατών είναι θετικά επαναληπτική) αν και μόνο αν

$$B^{-1} = 1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} < \infty.$$

Στην περίπτωση αυτή έχουμε ότι η κατανομή του αριθμού των πελατών στο σύστημα σε συνεχή χρόνο δίνεται από τη σχέση

$$p_n = \begin{cases} B & \text{αν } n = 0, \\ B \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} & \text{αν } n \geq 1. \end{cases} \quad (6.3)$$

Βεβαίως, αν $\lambda_k = 0$ για κάποιο k , τότε η Μαρκοβιανή αλυσίδα απορροφάται στο πεπερασμένο σύνολο $\{0, 1, \dots, k\}$ (για τον μικρότερο τέτοιο k) και στο σύνολο αυτό η Μαρκοβιανή αλυσίδα είναι θετικά επαναληπτική. Επομένως, απλές Μαρκοβιανές ουρές με πεπερασμένη χωρητικότητα είναι πάντα ευσταθείς.

Πέραν του υπολογισμού της μεταβατικής κατανομής ($p_n(t) : n \geq 0$) και της κατανομής ισορροπίας ($p_n : n \geq 0$) του αριθμού των πελατών στο σύστημα σε συνεχή χρόνο, είναι συχνά επιθυμητό να υπολογιστεί η απόδοση του συστήματος που μετρείται με τον μέσο ρυθμό εξυπηρέτησεων ανά χρονική μονάδα (ρυθμό διαπέρασης ή διακίνησης του συστήματος - throughput), καθώς και οι κατανομές ισορροπίας του αριθμού των πελατών σε στιγμές αφίξεων και αναχωρήσεων. Οι υπολογισμοί αυτοί παρουσιάζονται στην παράγραφο 6.2. Επίσης, είναι σημαντικό να υπολογιστεί η κατανομή του χρόνου παραμονής ενός πελάτη στο σύστημα, καθώς και κάποια χαρακτηριστικά του κύκλου απασχόλησης (busy cycle), όπως η μέση διάρκεια της περιόδου συνεχούς λειτουργίας (mean busy period) και η μέση διάρκεια ενός κύκλου απασχόλησης (mean busy cycle). Πρέπει να σημειωθεί, πάντως, ότι ο υπολογισμός της κατανομής του χρόνου παραμονής ενός πελάτη είναι ένα δύσκολο εν γένει πρόβλημα. Η βασική προσέγγιση για την επίλυσή του είναι η θεώρηση ενός συγκεκριμένου πελάτη και η δέσμευση στον αριθμό των πελατών που βλέπει κατά την είσοδό του, ενώ σημαντικό ρόλο παίζει και η πειθαρχία της ουράς. Η ιδέα αυτή δίνει λύσεις σε αρκετές περιπτώσεις, όπως θα δούμε στην ανάλυση διαφόρων συστημάτων που θα ακολουθήσει στις επόμενες παραγράφους. Ο υπολογισμός βασικών χαρακτηριστικών του κύκλου απασχόλησης γίνεται χρησιμοποιώντας την αναγεννητικότητα του αντίστοιχου συστήματος και θα παρουσιαστεί επίσης στο πλαίσιο της ανάλυσης συγκεκριμένων συστημάτων.

6.2 Ο ρυθμός διαπέρασης και οι εμφυτευμένες κατανομές

Ως ρυθμό διαπέρασης (throughput) μ^* ενός συστήματος εξυπηρέτησης εννοούμε τον μακροπρόθεσμο μέσο ρυθμό εξυπηρετούμενων πελατών ανά χρονική μονάδα. Η ποσότητα αυτή έχει αποδοθεί με πολλούς διαφορετικούς όρους στα Ελληνικά, όπως ρυθμός διαμεταγωγής, ρυθμός διακίνησης και ρυθμαπόδοση ενός συστήματος. Ο ρυθμός αυτός, στην περίπτωση που εξυπηρετούνται όλοι οι πελάτες που εισέρχονται (δηλαδή δεν συμβαίνουν υπαναχωρήσεις πελατών), ισούται με τον μακροπρόθεσμο ρυθμό εισερχομένων πελατών ανά χρονική μονάδα λ^* . Για έναν δεισθητικό υπολογισμό του λ^* μπορούμε να σκεφτούμε ως εξής: Έστω $A(t, t+h)$ και $D(t, t+h)$ τα πλήθη αφίξεων και αναχωρήσεων αντίστοιχα σε μια απλή Μαρκοβιανή ουρά στο διάστημα $(t, t+h]$. Τότε έχουμε

$$\begin{aligned} \lambda^* &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \frac{E[A(t, t+h)]}{h} = \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \frac{\Pr[A(t, t+h) = 1] + o(h)}{h} \\ &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \frac{\sum_{n=0}^{\infty} \Pr[Q(t) = n] \Pr[A(t, t+h) = 1 | Q(t) = n]}{h} \\ &= \sum_{n=0}^{\infty} p_n \lambda_n. \end{aligned}$$

Η δεύτερη από τις παραπάνω ισότητες προκύπτει καθώς η πιθανότητα για δύο αφίξεις στο $(t, t+h]$ είναι $o(h)$ για $h \rightarrow 0^+$. Ομοίως, έχουμε

$$\mu^* = \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \frac{E[D(t, t+h)]}{h} = \sum_{n=1}^{\infty} p_n \mu_n.$$

Για την αυστηρή αιτιολόγηση αυτών των σχέσεων, αρκεί να εφαρμόσουμε το θεώρημα 4.13 για τον υπολογισμό του μακροπρόθεσμου ρυθμού κόστους. Συγκεκριμένα, θεωρώντας ότι έχουμε κόστος 1 χρηματική μονάδα για κάθε μετάβαση τύπου $n \rightarrow n + 1$, δηλαδή για κάθε άφιξη, έχουμε τη δομή κόστους με κόστη παραμονής $c_j = 0$ για $j \geq 0$ και κόστη μετάβασης $d_{jk} = 1$ για $j \geq 0$ και $k = j + 1$ και $d_{jk} = 0$, διαφορετικά. Οπότε το θεώρημα 4.13 δίνει

$$\begin{aligned} \lambda^* &= \lim_{t \rightarrow \infty} \frac{E[C(t)]}{t} = \sum_{j \in \mathcal{S}} p_j c_j + \sum_{j \in \mathcal{S}} \sum_{k \neq j} p_j q_{jk} d_{jk} \\ &= \sum_{j=0}^{\infty} p_j q_{j,j+1} = \sum_{j=0}^{\infty} p_j \lambda_j. \end{aligned}$$

Ομοίως, θεωρώντας ότι έχουμε κόστος 1 χρηματική μονάδα για κάθε μετάβαση τύπου $n \rightarrow n - 1$, δηλαδή για κάθε αναχώρηση, συνάγουμε την έκφραση για τον μ^* . Ισχύει, όπως είπαμε $\lambda^* = \mu^*$, που μπορεί να αποδειχθεί και αλγεβρικά αθροίζοντας τις εξισώσεις γενικευμένης ισορροπίας $\lambda_{n-1} p_{n-1} = \mu_n p_n$, για όλες τις καταστάσεις $n \geq 1$.

Οι οριακές κατανομές σε στιγμές αφίξεων και αναχωρήσεων μπορούν να βρεθούν σκεφτόμενοι διαισθητικά ως εξής:

$$\begin{aligned} a_j &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \Pr[Q(t) = j | A(t, t+h) = 1] \\ &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \frac{\Pr[Q(t) = j, A(t, t+h) = 1]}{\Pr[A(t, t+h) = 1]} \\ &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \frac{\Pr[Q(t) = j] \Pr[A(t, t+h) = 1 | Q(t) = j]}{\Pr[A(t, t+h) = 1]} \\ &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \frac{\Pr[Q(t) = j] \Pr[A(t, t+h) = 1 | Q(t) = j] / h}{\Pr[A(t, t+h) = 1] / h} \\ &= \frac{p_j \lambda_j}{\lambda^*}, \end{aligned}$$

και ομοίως

$$\begin{aligned} d_j &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \Pr[Q(t+h) = j | D(t, t+h) = 1] \\ &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \frac{\Pr[Q(t+h) = j, D(t, t+h) = 1]}{\Pr[D(t, t+h) = 1]} \\ &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \frac{\Pr[Q(t) = j+1] \Pr[D(t, t+h) = 1 | Q(t) = j+1]}{\Pr[D(t, t+h) = 1]} \\ &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \frac{\Pr[Q(t) = j+1] \Pr[D(t, t+h) = 1 | Q(t) = j+1] / h}{\Pr[D(t, t+h) = 1] / h} \\ &= \frac{p_{j+1} \mu_{j+1}}{\mu^*}. \end{aligned}$$

Εναλλακτικά, χρησιμοποιώντας την αναγεννητικότητα του συστήματος, όταν είναι ευσταθές, έχουμε ότι η πιθανότητα a_j ισούται με το μακροπρόθεσμο ποσοστό των αφίξεων που βλέπουν j πελάτες, το οποίο ισούται με το λόγο του ρυθμού των αφίξεων που βλέπουν j πελάτες προς τον συνολικό ρυθμό των αφίξεων, οπότε έχουμε άμεσα τη σχέση $a_j = p_j \lambda_j / \lambda^*$. Και, ομοίως, η πιθανότητα d_j είναι το μακροπρόθεσμο ποσοστό των αναχωρήσεων που αφήνουν j πελάτες, το οποίο ισούται με το λόγο του ρυθμού των αναχωρήσεων που αφήνουν j πελάτες προς τον συνολικό ρυθμό των αναχωρήσεων. Έτσι παίρνουμε ότι $d_j = p_{j+1} \mu_{j+1} / \mu^*$.

Βλέπουμε, λοιπόν, ότι στην περίπτωση των απλών Μαρκοβιανών ουρών η ιδιότητα των μεμονωμένων αφίξεων και η ιδιότητα PASTA προκύπτουν άμεσα. Πράγματι οι εξισώσεις ισορροπίας δίνουν $\lambda_j p_j = \mu_{j+1} p_{j+1}$ για κάθε $j \geq 0$, οπότε $\lambda^* = \mu^*$ και $a_j = d_j$, για κάθε $j \geq 0$. Επίσης, όταν η διαδικασία αφίξεων είναι Poisson έχουμε $\lambda_j = \lambda$, για κάθε $j \geq 0$, οπότε έχουμε $\lambda^* = \lambda$ και $a_j = \frac{p_j \lambda_j}{\lambda^*} = \frac{p_j \lambda}{\lambda} = p_j$, για $j \geq 0$.

6.3 Η M/M/1/1 ουρά

Η M/M/1/1 ουρά έχει χρησιμοποιηθεί ως το μοντέλο μιας τηλεφωνικής γραμμής που μπορεί να είναι ελεύθερη ή κατειλημμένη (θεωρώντας ότι δεν υπάρχει δυνατότητα κράτησης μιας κλήσης σε αναμονή). Αν ο ρυθμός αφίξεων είναι λ , ο ρυθμός εξυπηρέτησης είναι μ και $\rho = \frac{\lambda}{\mu}$ είναι ο ρυθμός συνωστισμού, τότε έχουμε ότι η στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι διαδικασία γέννησης-θανάτου με χώρο καταστάσεων $\mathcal{S} = \{0, 1\}$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda & \text{αν } i = 0, j = 1, \\ \mu & \text{αν } i = 1, j = 0, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Το διάγραμμα ρυθμών μετάβασης της $\{Q(t)\}$ δίνεται στο σχήμα 4.3. Λόγω της απλότητας της συγκεκριμένης ουράς είναι σχετικά εύκολο να υπολογίσουμε τη μεταβατική της κατανομή. Το μετασχηματισμένο σύστημα των εξισώσεων Chapman-Kolmogorov είναι:

$$\begin{aligned} -(s + \lambda)\tilde{p}_0(s) + \mu\tilde{p}_1(s) &= -sp_0(0), \\ \lambda\tilde{p}_0(s) - (s + \mu)\tilde{p}_1(s) &= -sp_1(0). \end{aligned}$$

Λόγω της $p_1(t) = 1 - p_0(t)$ έχουμε ότι $\tilde{p}_1(s) = 1 - \tilde{p}_0(s)$, οπότε η πρώτη μετασχηματισμένη εξίσωση Chapman-Kolmogorov δίνει

$$\begin{aligned} \tilde{p}_0(s) &= \frac{\mu + sp_0(0)}{s + \lambda + \mu} = p_0(0) + \frac{\mu - (\lambda + \mu)p_0(0)}{s + \lambda + \mu} \\ &= p_0(0) + \left(\frac{\mu}{\lambda + \mu} - p_0(0)\right) \frac{\lambda + \mu}{s + \lambda + \mu}. \end{aligned}$$

Αντιστρέφοντας τον μετασχηματισμό Laplace-Stieltjes με τη βοήθεια του πίνακα 3.1, έχουμε

$$\begin{aligned} p_0(t) &= p_0(0) + \left(\frac{\mu}{\lambda + \mu} - p_0(0)\right) (1 - e^{-(\lambda + \mu)t}) \\ &= \frac{\mu}{\lambda + \mu} - \left(\frac{\mu}{\lambda + \mu} - p_0(0)\right) e^{-(\lambda + \mu)t}. \end{aligned}$$

Ομοίως:

$$p_1(t) = \frac{\lambda}{\lambda + \mu} + \left(\frac{\mu}{\lambda + \mu} - p_0(0)\right) e^{-(\lambda + \mu)t}.$$

Παίρνοντας το όριο καθώς $t \rightarrow \infty$ μπορούμε να πάρουμε την κατανομή ισορροπίας $(p_0, p_1) = \left(\frac{\mu}{\lambda + \mu}, \frac{\lambda}{\lambda + \mu}\right)$. Βλέπουμε επίσης ότι η σύγκλιση της μεταβατικής κατανομής στην κατανομή ισορροπίας είναι πολύ γρήγορη (εκθετική). Η κατανομή ισορροπίας μπορεί να βρεθεί βέβαια πολύ πιο εύκολα, χωρίς να περάσουμε από τη διαδικασία εύρεσης της μεταβατικής κατανομής. Πράγματι, οι εξισώσεις ισορροπίας είναι

$$\begin{aligned} \lambda p_0 &= \mu p_1, \\ \mu p_1 &= \lambda p_0, \end{aligned}$$

ενώ η εξίσωση κανονικοποίησης δίνει

$$p_0 + p_1 = 1.$$

Παρατηρήστε ότι η μια εξίσωση ισορροπίας είναι περιττή και μπορεί να παραληφθεί. Αυτό ισχύει πάντα: Μία από τις εξισώσεις ισορροπίας μπορεί πάντα να παραληφθεί αφού προκύπτει από τις υπόλοιπες (με άθροισή τους). Λύνοντας το σύστημα παίρνουμε

$$\begin{aligned} p_0 &= \frac{\mu}{\mu + \lambda} = \frac{1}{1 + \rho}, \\ p_1 &= \frac{\lambda}{\mu + \lambda} = \frac{\rho}{1 + \rho}. \end{aligned}$$

Λόγω της ιδιότητας των μεμονωμένων μεταβάσεων και της ιδιότητας PASTA έχουμε

$$d_0 = a_0 = p_0 = \frac{1}{1 + \rho}, \quad d_1 = a_1 = p_1 = \frac{\rho}{1 + \rho}.$$

Το ποσοστό των χαμένων πελατών, δηλαδή των πελατών που τελικά δεν εξυπηρετούνται, είναι το ποσοστό των πελατών που βρίσκουν έναν πελάτη κατά την άφιξή τους, δηλαδή δίνεται από την πιθανότητα a_1 .

Ο μέσος αριθμός πελατών στο σύστημα είναι

$$E[Q] = 0 \cdot p_0 + 1 \cdot p_1 = p_1 = \frac{\rho}{1 + \rho}.$$

Εφαρμόζοντας τον νόμο του Little έχουμε ότι ο μέσος χρόνος παραμονής ενός αφικνούμενου πελάτη στο σύστημα είναι

$$E[S] = \frac{E[Q]}{\lambda} = \frac{1}{\mu(1 + \rho)}.$$

Προσέξτε ότι αυτός ο μέσος χρόνος αναφέρεται σε όλους τους πελάτες που φθάνουν στο σύστημα, ανεξάρτητα αν εισέρχονται σε αυτό ή όχι. Ο υπολογισμός του θα μπορούσε να γίνει και ως εξής:

$$E[S] = a_0 \frac{1}{\mu} + a_1 \cdot 0 = \frac{1}{\mu(1 + \rho)},$$

δηλαδή δεσμεύοντας στον αριθμό των πελατών που βρίσκει ένας αφικνούμενος πελάτης. Αν μας ενδιέφερε ο μέσος χρόνος παραμονής ενός εισερχόμενου πελάτη στο σύστημα, που προφανώς είναι ο μέσος χρόνος εξυπηρέτησής του, ο νόμος του Little δίνει και πάλι το σωστό αποτέλεσμα, αρκεί να χρησιμοποιήσουμε όχι το συνολικό ρυθμό αφίξεων λ , αλλά το ρυθμό εισερχόμενων πελατών $\lambda^* = \lambda a_0$. Πράγματι θα είχαμε

$$E[S^{enter}] = \frac{E[Q]}{\lambda^*} = \frac{1}{\mu}.$$

Για την κατανομή του χρόνου παραμονής ενός αφικνούμενου πελάτη έχουμε

$$F_S(x) = \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} (1 - e^{-\mu x}), \quad x \geq 0,$$

αφού ο χρόνος παραμονής είναι 0 με πιθανότητα a_1 , ή ίσος με τον χρόνο εξυπηρέτησής του που ακολουθεί την $\text{Exp}(\mu)$ κατανομή. Η κατανομή του χρόνου παραμονής ενός εισερχόμενου πελάτη είναι η $\text{Exp}(\mu)$, αφού στην περίπτωση αυτή ο χρόνος παραμονής είναι ακριβώς ο χρόνος εξυπηρέτησης του συγκεκριμένου πελάτη.

Όσον αφορά τη μελέτη του κύκλου απασχόλησης του συστήματος, έχουμε ότι η μέση περίοδος αργίας είναι $E[I] = \frac{1}{\lambda}$, αφού η περίοδος αργίας I ακολουθεί την $\text{Exp}(\lambda)$, ενώ η μέση περίοδος συνεχούς λειτουργίας είναι $E[Y] = \frac{1}{\mu}$, αφού η Y ακολουθεί την $\text{Exp}(\mu)$ κατανομή. Ο μέσος κύκλος απασχόλησης είναι $E[Z] = E[I] + E[Y] = \frac{1}{\lambda} + \frac{1}{\mu}$.

6.4 Η M/M/1 ουρά

Η M/M/1 ουρά είναι το απλούστερο μοντέλο συστήματος εξυπηρέτησης με άπειρο χώρο αναμονής. Αν ο ρυθμός αφίξεων είναι λ , ο ρυθμός εξυπηρέτησης είναι μ και $\rho = \frac{\lambda}{\mu}$ είναι ο ρυθμός συνωστισμού τότε, όπως έχουμε δει στο παράδειγμα 4.7, η στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι τύπου γέννησης-θανάτου με χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda & \text{αν } i \geq 0, j = i + 1, \\ \mu & \text{αν } i \geq 1, j = i - 1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Το διάγραμμα ρυθμών μετάβασης για τη διαδικασία $\{Q(t)\}$ δίνεται στο σχήμα 4.4. Παρά την απλότητα της συγκεκριμένης ουράς δεν είναι εύκολο να υπολογίσουμε τη μεταβατική της κατανομή. Το μετασχηματισμένο σύστημα των εξισώσεων Charman-Kolmogorov είναι:

$$\begin{aligned} -(s + \lambda)\tilde{p}_0(s) + \mu\tilde{p}_1(s) &= -sp_0(0), \\ \lambda\tilde{p}_{n-1}(s) - (s + \lambda + \mu)\tilde{p}_n(s) + \mu\tilde{p}_{n+1}(s) &= -sp_n(0), \quad n \geq 1. \end{aligned}$$

Καθώς πρόκειται για ένα τριδιαγώνιο σύστημα γραμμικών εξισώσεων με ομογενή δομή, μπορεί να λυθεί, αλλά οι προκύπτοντες μετασχηματισμοί Laplace-Stieltjes είναι δύσκολο να αντιστραφούν. Λόγω της πολυπλοκότητας της όλης διαδικασίας, παραλείπουμε την επίλυση του συστήματος και την αντιστροφή των μετασχηματισμών στη γενική περίπτωση. Για την περίπτωση που το σύστημα είναι αρχικά κενό, δηλαδή που $p_0(0) = 1$ και $p_n(0) = 0$, για $n \geq 1$, το μετασχηματισμένο σύστημα των εξισώσεων Charman-Kolmogorov παίρνει τη μορφή

$$-(s + \lambda)\tilde{p}_0(s) + \mu\tilde{p}_1(s) = -s, \quad (6.4)$$

$$\lambda\tilde{p}_{n-1}(s) - (s + \lambda + \mu)\tilde{p}_n(s) + \mu\tilde{p}_{n+1}(s) = 0, \quad n \geq 1. \quad (6.5)$$

Επειδή η τελευταία εξίσωση είναι μια ομογενής εξίσωση διαφορών με σταθερούς συντελεστές, μπορεί να εφαρμοστεί η σχετική θεωρία (δείτε, π.χ., Elaydi 2006) και η γενική της λύση δίνεται από την

$$\tilde{p}_n(s) = Ax_+^n(s) + Bx_-^n(s), \quad n \geq 1, \quad (6.6)$$

όπου A, B σταθερές προς προσδιορισμό και $x_{\pm}(s)$ είναι οι λύσεις της χαρακτηριστικής εξίσωσης

$$\lambda - (s + \lambda + \mu)x + \mu x^2 = 0,$$

με

$$x_{\pm}(s) = \frac{(s + \lambda + \mu) \pm \sqrt{(s + \lambda + \mu)^2 - 4\lambda\mu}}{2\mu}.$$

Επιπλέον

$$x_+(s) + x_-(s) = \frac{s + \lambda + \mu}{\mu}, \quad x_+(s)x_-(s) = \frac{\lambda}{\mu}, \quad |x_+(s)| > |x_-(s)|.$$

Χρησιμοποιώντας το θεώρημα Rouché από τη Μιγαδική Ανάλυση (π.χ. βλέπε Παράγραφο 4.5.2 (The Argument Principle) στο βιβλίο Ahlfors 1979) μπορούμε να δείξουμε ότι $|x_+(s)| > 1$, $|x_-(s)| < 1$. Για τους αναγνώστες που δεν έχουν γνώσεις Μιγαδικής ανάλυσης, η εφαρμογή του θεωρήματος που περιγράφεται στην επόμενη παράγραφο του κειμένου μπορεί να παραληφθεί χωρίς να υπάρχει πρόβλημα για την κατανόηση του υπόλοιπου κεφαλαίου.

Για την εφαρμογή του θεωρήματος Rouché, θέτουμε $f(z) = -(s + \lambda + \mu)z$, $g(z) = \mu z^2 - (s + \lambda + \mu)z + \lambda$, για s με $\Re(s) > 0$ και θεωρούμε την καμπύλη γ που διατρέχει τον μοναδιαίο κύκλο με κέντρο το $z = 0$. Τότε

έχουμε ότι οι συναρτήσεις $f(z)$, $g(z)$ είναι αναλυτικές σε όλο το \mathbb{C} και η καμπύλη είναι ομόλογη με το 0 και απλή (δηλ. συρρικνώνεται στο 0 με συνεχή τρόπο και δεν διασταυρώνεται με τον εαυτό της). Για τα σημεία z που ανήκουν στην καμπύλη γ έχουμε $z = 1$ κι επομένως

$$\begin{aligned} |f(z)| &= |-(s + \lambda + \mu)z| = |s + \lambda + \mu| \\ &> |\lambda + \mu| = \lambda + \mu \\ &\geq |\lambda + \mu z^2| = |f(z) - g(z)|. \end{aligned}$$

Εφαρμόζοντας, τώρα, το θεώρημα Rouché έχουμε ότι οι $f(z)$ και $g(z)$ έχουν τον ίδιο αριθμό ριζών εντός της καμπύλης γ . Καθώς η $f(z) = 0$ έχει μόνο τη ρίζα $z = 0$ συνάγουμε ότι και η εξίσωση $g(z) = 0$ έχει μια ρίζα εντός του μοναδιαίου δίσκου.

Επομένως, η χαρακτηριστική εξίσωση

$$\mu x^2 - (s + \lambda + \mu)x + \lambda = 0,$$

έχει ακριβώς μια ρίζα εντός του μοναδιαίου δίσκου, οπότε, $|x_-(s)| < 1 < |x_+(s)|$.

Όμως,

$$\sum_{n=0}^{\infty} p_n(t) = 1 \Leftrightarrow \sum_{n=0}^{\infty} \tilde{p}_n(s) = 1, \quad \Re(s) > 0.$$

Άρα η $\sum_{n=0}^{\infty} \tilde{p}_n(s) = \sum_{n=0}^{\infty} (Ax_+(s)^n + Bx_-(s)^n)$ συγκλίνει, και αφού $|x_+(s)| > 1$, πρέπει υποχρεωτικά $A = 0$, οπότε,

$$\tilde{p}_n(s) = Bx_-(s)^n, \quad n \geq 0. \quad (6.7)$$

Αντικαθιστώντας την (6.7) στην (6.5) για $n = 1$ και λαμβάνοντας υπόψιν την χαρακτηριστική εξίσωση καταλήγουμε στο ότι $\tilde{p}_0(s) = B$, οπότε,

$$\tilde{p}_n(s) = \tilde{p}_0(s)x_-(s)^n, \quad n \geq 0. \quad (6.8)$$

Όμως,

$$1 = \sum_{n=0}^{\infty} \tilde{p}_n(s) = \tilde{p}_0(s) \sum_{n=0}^{\infty} x_-(s)^n = \frac{\tilde{p}_0(s)}{1 - x_-(s)} \Leftrightarrow \tilde{p}_0(s) = 1 - x_-(s).$$

Επομένως

$$\tilde{p}_n(s) = (1 - x_-(s))x_-(s)^n, \quad n \geq 0, \quad \Re(s) > 0. \quad (6.9)$$

Είναι εύκολο να διαπιστώσουμε ότι η (6.9), λαμβάνοντας υπόψιν την χαρακτηριστική εξίσωση, ικανοποιεί επίσης την (6.4).

Η γνώση του μετασχηματισμού Laplace-Stieltjes της μεταβατικής κατανομής αρκεί για την εύρεση της οριακής κατανομής λαμβάνοντας υπόψιν την ακόλουθη ιδιότητα (initial value theorem for Laplace-Stieltjes transforms):

$$\lim_{s \rightarrow 0} \tilde{p}_n(s) = \lim_{t \rightarrow \infty} p_n(t).$$

Οπότε έχουμε

$$\begin{aligned} \lim_{s \rightarrow 0} x_-(s) &= \lim_{s \rightarrow 0} \frac{s + \lambda + \mu - \sqrt{(s + \lambda + \mu)^2 - 4\lambda\mu}}{2\mu} \\ &= \begin{cases} \lambda/\mu := \rho, & \lambda < \mu, \\ 1, & \lambda \geq \mu. \end{cases} \end{aligned}$$

Άρα,

$$\begin{aligned} p_n &= \lim_{t \rightarrow \infty} p_n(t) = \lim_{s \rightarrow 0} \tilde{p}_n(s) = \lim_{s \rightarrow 0} (1 - x_-(s))x_-(s)^n \\ &= \begin{cases} (1 - \rho)\rho^n, & \rho < 1, \\ 0, & \rho \geq 1. \end{cases} \end{aligned}$$

Η αντιστροφή του μετασχηματισμού $\tilde{p}_n(s)$ που δίνεται από την (6.9) γίνεται χρησιμοποιώντας πίνακες μετασχηματισμών ή κάποιο πακέτο συμβολικών υπολογισμών. Για λόγους πληρότητας, παραθέτουμε τη λύση για τη γενικότερη περίπτωση που το σύστημα εξυπηρέτησης ξεκινά με i πελάτες τη χρονική στιγμή 0. Τότε, έχουμε:

$$p_n(t) = p_{in}(t) = e^{-(\lambda+\mu)t} \left[\rho^{(n-i)/2} I_{n-i}(at) + \rho^{(n-i-1)/2} I_{n+i+1}(at) \right] \\ + (1-\rho) \rho^n e^{-(\lambda+\mu)t} \sum_{j=n+i+2}^{\infty} \rho^{-j/2} I_j(at),$$

με

$$\alpha = 2\mu\sqrt{\rho} \quad \text{και} \quad I_n(x) = \sum_{h=0}^{\infty} \frac{(x/2)^{n+2h}}{(n+h)!h!}, \quad n \geq 1.$$

Βλέπουμε, λοιπόν, ότι η εύρεση της μεταβατικής κατανομής ακόμη και του απλούστερου συστήματος εξυπηρέτησης με απεριόριστη χωρητικότητα είναι αρκετά δύσκολη. Από την άλλη μεριά, η εύρεση της κατανομής ισορροπίας είναι πολύ ευκολότερη και ταυτόχρονα πολύ πιο χρήσιμη με βάση τα θεωρήματα που είδαμε για την οριακή θεωρία των Μαρκοβιανών αλυσίδων και τις δομές αμοιβής-κόστους επί αυτών. Για τον λόγο αυτό επικεντρωθήκαμε στη συνέχεια αυτού του κεφαλαίου, καθώς και στα επόμενα κεφάλαια, στον προσδιορισμό μόνο κατανομών ισορροπίας.

Εφαρμόζοντας τη σχετική θεωρία (δείτε την παράγραφο 6.1), έχουμε

$$B^{-1} = 1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} = \sum_{n=0}^{\infty} \rho^n \\ = \begin{cases} \frac{1}{1-\rho} & \text{αν } \rho < 1 \text{ (ευστάθεια)} \\ \infty & \text{αν } \rho \geq 1 \text{ (αστάθεια)}. \end{cases}$$

Οπότε, για την κατανομή ισορροπίας του αριθμού των πελατών έχουμε

$$p_n = \begin{cases} B & \text{αν } n = 0 \\ \frac{B \lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} & \text{αν } n \geq 1 \end{cases} \\ = (1-\rho) \rho^n, \quad n \geq 0. \quad (6.10)$$

Με άλλα λόγια η κατανομή ισορροπίας (p_n) είναι η $\text{Geom}(\rho)$ στο \mathbb{N}_0 . Λόγω της ιδιότητας των μεμονωμένων μεταβάσεων και της PASTA, έχουμε ότι και οι οριακές κατανομές σε στιγμές αφίξεων και αναχωρήσεων συμπίπτουν με την οριακή κατανομή σε συνεχή χρόνο. Για τον μέσο αριθμό πελατών στο σύστημα έχουμε

$$E[Q] = \sum_{n=0}^{\infty} n p_n = (1-\rho) \sum_{n=0}^{\infty} n \rho^n = \frac{\rho}{1-\rho}.$$

Από τον νόμο του Little έχουμε για τον χρόνο παραμονής ότι

$$E[S] = \frac{E[Q]}{\lambda} = \frac{1}{\mu(1-\rho)}.$$

Οι δυο αυτοί τύποι ταυτίζονται με τις εκφράσεις (5.2) που αποδείξαμε με την ανάλυση μέσης τιμής στην παράγραφο 5.2.

Όσον αφορά την κατανομή $F_S(x)$ του χρόνου παραμονής ενός πελάτη στο σύστημα, ο υπολογισμός γίνεται με δέσμευση στον αριθμό των πελατών που βρίσκει ο πελάτης κατά την άφιξή του, Q^- . Συγκεκριμένα, αν ο

πελάτης βρει $Q^- = n$ πελάτες στο σύστημα κατά την άφιξή του, τότε θα περιμένει συνολικά όσο το άθροισμα $n + 1$ χρόνων εξυπηρέτησης, δηλαδή έναν χρόνο που έχει την Erlang($n + 1, \mu$) κατανομή με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{\mu^{n+1}}{n!} x^n e^{-\mu x}, \quad x \geq 0$$

(βλέπε (3.19)). Έχουμε:

$$\begin{aligned} F_S(x) &= \Pr[S \leq x] = \sum_{n=0}^{\infty} \Pr[Q^- = n] \Pr[S \leq x | Q^- = n] \\ &= \sum_{n=0}^{\infty} (1 - \rho) \rho^n \int_0^x \frac{\mu^{n+1}}{n!} u^n e^{-\mu u} du \\ &= \int_0^x \mu(1 - \rho) e^{-\mu u} \sum_{n=0}^{\infty} \frac{(\mu \rho u)^n}{n!} du \\ &= \int_0^x \mu(1 - \rho) e^{-\mu u} e^{\mu \rho u} du = \int_0^x \mu(1 - \rho) e^{-\mu(1-\rho)u} du. \end{aligned} \quad (6.11)$$

Επομένως, βλέπουμε ότι η S είναι στην περίπτωση αυτή συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας $f_S(x) = \mu(1 - \rho) e^{-\mu(1-\rho)x}$, για $x \geq 0$. Επομένως, έχουμε ότι η S ακολουθεί την $\text{Exp}(\mu(1 - \rho))$ κατανομή.

Εναλλακτικά, ο υπολογισμός θα μπορούσε να γίνει μέσω του αντίστοιχου μετασχηματισμού Laplace-Stieltjes. Μάλιστα, αυτή η μέθοδος πλεονεκτεί καθώς δεν χρειάζεται υπολογισμούς ολοκληρωμάτων, παρά μόνο αλγεβρικούς μετασχηματισμούς και επομένως θα πρέπει εν γένει να προτιμάται για τους υπολογισμούς σε πιο πολύπλοκα συστήματα. Έχουμε ότι

$$\begin{aligned} \tilde{F}_S(s) &= E[e^{-sS}] = \sum_{n=0}^{\infty} \Pr[Q^- = n] E[e^{-sS} | Q^- = n] \\ &= \sum_{n=0}^{\infty} (1 - \rho) \rho^n \left(\frac{\mu}{\mu + s} \right)^{n+1} \\ &= \frac{(1 - \rho)\mu}{\mu + s} \left(1 - \frac{\rho\mu}{\mu + s} \right)^{-1} \\ &= \frac{(1 - \rho)\mu}{(1 - \rho)\mu + s}, \end{aligned} \quad (6.12)$$

όπου για την τρίτη ισότητα χρησιμοποιήθηκε η ιδιότητα PASTA, ενώ για την τέταρτη απαιτείται η σύγκλιση της σειράς που εξασφαλίζεται καθώς για s με $\text{Re}(s) \geq 0$ και $\rho < 1$ έχουμε $\left| \frac{\rho\mu}{\mu + s} \right| < 1$. Ο τελευταίος είναι ο μετασχηματισμός Laplace-Stieltjes της $\text{Exp}(\mu(1 - \rho))$, οπότε συνάγουμε και πάλι ότι η S ακολουθεί την $\text{Exp}(\mu(1 - \rho))$ κατανομή.

Όσον αφορά την εύρεση των μέσων τιμών που αφορούν τον κύκλο απασχόλησης του συστήματος, έχουμε ότι η περίοδος αργίας I είναι $\text{Exp}(\lambda)$, αφού ταυτίζεται με τον υπολειπόμενο ενδιάμεσο χρόνο αφίξεων κατά τη στιγμή της αναχώρησης ενός πελάτη που αφήνει κενό το σύστημα (χρησιμοποιείται κι εδώ η αμνήμονη ιδιότητα της εκθετικής κατανομής). Οπότε $E[I] = \frac{1}{\lambda}$. Επιπλέον, λόγω του αναγεννητικού χαρακτήρα του συστήματος, έχουμε ότι η οριακή πιθανότητα κενού συστήματος που ισούται με το μακροπρόθεσμο ποσοστό του χρόνου που το σύστημα είναι κενό, ισούται επίσης με το λόγο του μέσου χρόνου που το σύστημα είναι κενό σε έναν αναγεννητικό κύκλο προς τη μέση διάρκεια του κύκλου (βλέπε θεώρημα 1.5 και ειδικότερα τη σχέση (1.13)). Ένας αναγεννητικός κύκλος στην περίπτωση του συστήματος αυτού είναι ένας κύκλος απασχόλησης και στη διάρκειά του το σύστημα είναι κενό μόνο κατά την αντίστοιχη περίοδο αργίας. Επομένως

$$p_0 = \frac{E[I]}{E[Z]},$$

όπου $E[Z]$ η μέση διάρκεια ενός κύκλου απασχόλησης. Οπότε

$$E[Z] = \frac{E[I]}{p_0} = \frac{1}{\lambda(1-\rho)}.$$

Η μέση περίοδος συνεχούς λειτουργίας θα είναι

$$E[Y] = E[Z] - E[I] = \frac{1}{\lambda(1-\rho)} - \frac{1}{\lambda} = \frac{1}{\mu(1-\rho)}.$$

6.5 Τροποποιήσεις της M/M/1 ουράς

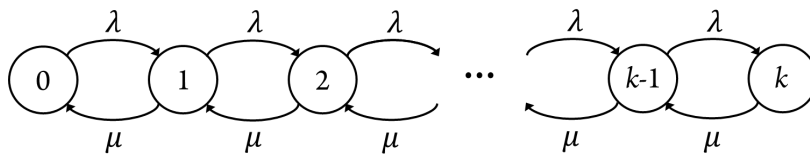
Εδώ θα εξετάσουμε τρεις ενδιαφέρουσες τροποποιήσεις της M/M/1 ουράς που εμφανίζονται στις εφαρμογές.

6.5.1 Η M/M/1/k ουρά

Θεωρούμε την M/M/1/k ουρά με ρυθμό αφίξεων λ , ρυθμό εξυπηρέτησης μ , ρυθμό συνωστισμού $\rho = \frac{\lambda}{\mu}$ και χωρητικότητα k . Η στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι και εδώ διαδικασία τύπου γέννησης-θανάτου με χώρο καταστάσεων $\mathcal{S} = \{0, 1, \dots, k\}$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda & \text{αν } 0 \leq i \leq k-1, j = i+1, \\ \mu & \text{αν } 1 \leq i \leq k, j = i-1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Το διάγραμμα ρυθμών μετάβασης για τη διαδικασία $\{Q(t)\}$ δίνεται στο σχήμα 6.1.



Σχήμα 6.1: Διάγραμμα ρυθμών μετάβασης M/M/1/k ουράς.

Επειδή ο χώρος καταστάσεων είναι πεπερασμένος το σύστημα είναι πάντα ευσταθές και έχουμε

$$\begin{aligned} B^{-1} &= 1 + \sum_{n=1}^k \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} = \sum_{n=0}^k \rho^n \\ &= \begin{cases} \frac{1-\rho^{k+1}}{1-\rho} & \text{αν } \rho \neq 1, \\ k+1 & \text{αν } \rho = 1. \end{cases} \end{aligned}$$

Οπότε για την οριακή κατανομή του αριθμού των πελατών έχουμε

$$p_n = \begin{cases} \frac{(1-\rho)\rho^n}{1-\rho^{k+1}} & \text{αν } 0 \leq n \leq k, \rho \neq 1, \\ \frac{1}{k+1} & \text{αν } 0 \leq n \leq k, \rho = 1. \end{cases}$$

Λόγω της ιδιότητας των μεμονωμένων μεταβάσεων και της PASTA, έχουμε ότι και οι οριακές κατανομές σε στιγμές αφίξεων και αναχωρήσεων συμπίπτουν με την οριακή κατανομή σε συνεχή χρόνο. Αυτό βέβαια ισχύει όταν αναφερόμαστε σε όλους τους αφικνούμενους πελάτες, είτε μπαίνουν είτε όχι στο σύστημα. Ο ρυθμός

διαπέρασης του συστήματος που αναφέρεται στους εισερχόμενους πελάτες στο σύστημα που είναι και αυτοί που εξυπηρετούνται δίνεται ως

$$\lambda^* = \sum_{n=0}^k \lambda_n p_n = \sum_{n=0}^{k-1} \lambda p_n = \lambda(1 - p_k). \quad (6.13)$$

Οπότε, αν ενδιαφερόμαστε για την κατανομή ισορροπίας του αριθμού των πελατών σε στιγμές «πραγματικών αφίξεων», δηλαδή σε στιγμές εισόδων πελατών έχουμε

$$\begin{aligned} a_n^{enter} &= \frac{\lambda_n p_n}{\lambda^*} \\ &= \begin{cases} \frac{p_n}{1-p_k} & \text{αν } 0 \leq n \leq k-1, \\ 0 & \text{αν } n = k, \end{cases} \end{aligned}$$

δηλαδή η PASTA δεν είναι εφαρμόσιμη. Για το μέσο αριθμό πελατών στο σύστημα έχουμε

$$E[Q] = \sum_{n=0}^k n p_n = \begin{cases} \frac{\rho}{1-\rho} \cdot \frac{k\rho^{k+1} - (k+1)\rho^k + 1}{1-\rho^{k+1}} & \text{αν } \rho \neq 1, \\ \frac{k}{2} & \text{αν } \rho = 1. \end{cases} \quad (6.14)$$

Από τον νόμο του Little έχουμε για τον μέσο χρόνο παραμονής με αναφορά σε όλους τους αφικνούμενους πελάτες (είτε εισέρχονται στο σύστημα είτε όχι):

$$E[S] = \frac{E[Q]}{\lambda}.$$

Όμως, αν μας ενδιαφέρει ο μέσος χρόνος παραμονής μόνο των πελατών που εισέρχονται, έχουμε

$$E[S^{enter}] = \frac{E[Q]}{\lambda^*}.$$

Το μακροπρόθεσμο ποσοστό των χαμένων πελατών βρίσκεται ως

$$1 - \frac{\lambda^*}{\lambda} = 1 - \frac{\lambda(1-p_k)}{\lambda} = p_k.$$

Εναλλακτικά, το μακροπρόθεσμο ποσοστό των χαμένων πελατών ισούται με την οριακή πιθανότητα ένας πελάτης να βρει το σύστημα γεμάτο και να αναγκαστεί να φύγει, είναι δηλαδή ίσο με a_k . Οπότε, λόγω της PASTA, συνάγουμε ότι είναι ίσο με p_k .

Όσον αφορά τη μελέτη των βασικών μέτρων (μέσων τιμών) που σχετίζονται με τον κύκλο απασχόλησης του συστήματος, έχουμε ότι και πάλι η περίοδος αργίας I ακολουθεί την κατανομή $\text{Exp}(\lambda)$, οπότε $E[I] = \frac{1}{\lambda}$. Επιπλέον, λόγω του αναγεννητικού χαρακτήρα του συστήματος, έχουμε ότι η οριακή πιθανότητα κενού συστήματος που ισούται με το μακροπρόθεσμο ποσοστό του χρόνου που το σύστημα είναι κενό, ισούται επίσης με το λόγο του μέσου χρόνου σε έναν αναγεννητικό κύκλο που το σύστημα είναι κενό προς τη μέση διάρκεια του κύκλου. Ένας αναγεννητικός κύκλος στην περίπτωση του συστήματος αυτού είναι ένας κύκλος απασχόλησης και στη διάρκειά του το σύστημα είναι κενό μόνο κατά την αντίστοιχη περίοδο αργίας. Επομένως

$$p_0 = \frac{E[I]}{E[Z]},$$

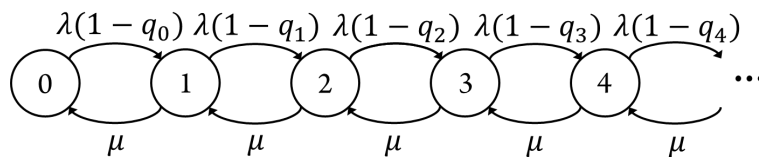
όπου $E[Z]$ η μέση διάρκεια ενός κύκλου απασχόλησης. Οπότε ισχύει ακριβώς η ίδια λογική προσδιορισμού που περιγράφηκε για την M/M/1 ουρά. Λύνουμε την παραπάνω σχέση ως προς $E[Z]$, αφού τα $E[I]$ και p_0 έχουν προσδιοριστεί και κατόπιν προσδιορίζουμε και τη μέση περίοδο συνεχούς λειτουργίας από τη σχέση $E[Y] = E[Z] - E[I]$.

6.5.2 Η M/M/1 ουρά με αποθαρρυνόμενους πελάτες

Θεωρούμε τώρα την M/M/1 ουρά με ρυθμό αφίξεων λ , ρυθμό εξυπηρέτησης μ και ρυθμό συνωστισμού $\rho = \frac{\lambda}{\mu}$, όπου υποθέτουμε ότι οι πελάτες μπορεί να αποθαρρυνθούν να μπουν στο σύστημα, αφού παρατηρήσουν τον υπάρχοντα συνωστισμό σε αυτό. Συγκεκριμένα, υποθέτουμε ότι ένας αφικνούμενος πελάτης που βρίσκει n πελάτες αποχωρεί άμεσα (he/she balks) με πιθανότητα q_n . Ισοδύναμα, το ποσοστό των πελατών που αποχωρούν άμεσα επί αυτών που βρίσκουν το σύστημα με n πελάτες είναι q_n . Και σε αυτή την περίπτωση μπορούμε εύκολα να δούμε ότι η στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι διαδικασία τύπου γέννησης-θανάτου με χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda(1 - q_i) & \text{αν } i \geq 0, j = i + 1, \\ \mu & \text{αν } i \geq 1, j = i - 1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Το διάγραμμα ρυθμών μετάβασης για τη διαδικασία $\{Q(t)\}$ δίνεται στο σχήμα 6.2.



Σχήμα 6.2: Διάγραμμα ρυθμών μετάβασης M/M/1 ουράς με αποθαρρυνόμενους πελάτες.

Στα πραγματικά συστήματα είναι λογικό να υποθέσουμε ότι η q_i είναι αύξουσα ως προς i και ότι τείνει στο 1, καθώς το i τείνει στο άπειρο. Το προηγούμενο σύστημα της M/M/1/k ουράς μπορεί να θεωρηθεί μια ειδική περίπτωση της M/M/1 ουράς με αποθαρρυνόμενους πελάτες που εμπίπτει σε αυτή την κατηγορία (εκεί είναι $q_i = 0$ για $0 \leq i \leq k - 1$ και $q_i = 1$ για $i \geq k$). Στο υπόλοιπο της παραγράφου, θα εξετάσουμε την ειδική περίπτωση όπου $q_i = \frac{i}{i+1}$ που αποτελεί τυπικό παράδειγμα σε πολλά εισαγωγικά συγγράμματα της Θεωρίας Ουρών, καθώς δίνει κομψά αποτελέσματα. Έχουμε

$$B^{-1} = 1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} = \sum_{n=0}^{\infty} \frac{\rho^n}{n!} = e^\rho < \infty,$$

δηλαδή το συγκεκριμένο σύστημα είναι πάντα ευσταθές. Για την κατανομή ισορροπίας του αριθμού των πελατών έχουμε

$$p_n = e^{-\rho} \frac{\rho^n}{n!}, \quad n \geq 0,$$

δηλαδή η (p_n) ακολουθεί την κατανομή Poisson(ρ). Λόγω της ιδιότητας των μεμονωμένων μεταβάσεων και της PASTA, έχουμε ότι και οι οριακές κατανομές σε στιγμές αφίξεων και αναχωρήσεων συμπίπτουν με την κατανομή ισορροπίας σε συνεχή χρόνο. Αυτό βέβαια ισχύει όταν αναφερόμαστε σε όλους τους αφικνούμενους πελάτες, είτε μπαίνουν είτε όχι στο σύστημα. Ο ρυθμός διαπέρασης του συστήματος που αναφέρεται στους εισερχόμενους πελάτες στο σύστημα είναι

$$\begin{aligned} \lambda^* &= \sum_{n=0}^{\infty} \lambda_n p_n = \sum_{n=0}^{\infty} \lambda(1 - q_n) p_n = \sum_{n=0}^{\infty} \frac{\lambda}{n+1} e^{-\rho} \frac{\rho^n}{n!} = \frac{\lambda e^{-\rho}}{\rho} \sum_{n=0}^{\infty} \frac{\rho^{n+1}}{(n+1)!} \\ &= \frac{\lambda e^{-\rho}}{\rho} (e^\rho - 1) = \mu(1 - e^{-\rho}). \end{aligned}$$

Το ίδιο αποτέλεσμα προκύπτει ευκολότερα υπολογίζοντας τον μέσο ρυθμό εξυπηρέτησεων του συστήματος:

$$\mu^* = \sum_{n=1}^{\infty} \mu_n p_n = \sum_{n=1}^{\infty} \mu p_n = \mu(1 - p_0) = \mu(1 - e^{-\rho}).$$

Αν ενδιαφερόμαστε για την κατανομή ισορροπίας του αριθμού των πελατών σε στιγμές «πραγματικών αφίξεων», δηλαδή σε στιγμές εισόδων πελατών έχουμε

$$a_n^{enter} = \frac{\lambda_n p_n}{\lambda^*} = \frac{e^{-\rho}}{1 - e^{-\rho}} \cdot \frac{\rho^{n+1}}{(n+1)!}, \quad n \geq 0,$$

δηλαδή η PASTA δεν είναι εφαρμόσιμη. Ο μέσος αριθμός πελατών στο σύστημα είναι $E[Q] = \rho$. Από τον νόμο του Little έχουμε για το μέσο χρόνο παραμονής με αναφορά σε όλους τους αφικνούμενους πελάτες (είτε εισέρχονται στο σύστημα είτε όχι) ότι

$$E[S] = \frac{E[Q]}{\lambda} = \frac{1}{\mu}.$$

Αυτό μοιάζει κάπως παράδοξο, αφού ο μέσος χρόνος εξυπηρέτησης είναι $\frac{1}{\mu}$. Όμως, δεν είναι παράδοξο αφού η μέση τιμή του χρόνου παραμονής αναφέρεται σε όλους τους αφικνούμενους πελάτες, οπότε κάποιοι από αυτούς (αυτοί που φεύγουν αμέσως) έχουν μηδενικό χρόνο παραμονής. Αν μας ενδιαφέρει ο μέσος χρόνος παραμονής μόνο των πελατών που εισέρχονται, έχουμε

$$E[S^{enter}] = \frac{E[Q]}{\lambda^*} = \frac{\rho}{\mu(1 - e^{-\rho})}.$$

Το ποσοστό των χαμένων πελατών βρίσκεται ως

$$1 - \frac{\lambda^*}{\lambda} = 1 - \frac{\mu(1 - e^{-\rho})}{\lambda} = 1 - \frac{1 - e^{-\rho}}{\rho}.$$

Εναλλακτικά, το ποσοστό των χαμένων πελατών μπορεί να υπολογιστεί ως η οριακή πιθανότητα ένας πελάτης να αποχωρήσει από το σύστημα μόλις αφιχθεί σε αυτό. Οπότε, δεσμεύοντας στον αριθμό των πελατών που βρίσκει φθάνοντας έχουμε ότι είναι

$$\begin{aligned} \sum_{n=0}^{\infty} a_n q_n &= \sum_{n=0}^{\infty} e^{-\rho} \frac{\rho^n}{n!} \cdot \frac{n}{n+1} = \sum_{n=0}^{\infty} e^{-\rho} \frac{\rho^n}{n!} \cdot \left(1 - \frac{1}{n+1}\right) \\ &= 1 - \sum_{n=0}^{\infty} e^{-\rho} \frac{\rho^n}{(n+1)!} = 1 - \frac{e^{-\rho}}{\rho} \sum_{n=0}^{\infty} \frac{\rho^{n+1}}{(n+1)!} \\ &= 1 - \frac{e^{-\rho}}{\rho} (e^{\rho} - 1) = 1 - \frac{1 - e^{-\rho}}{\rho}. \end{aligned}$$

Όσον αφορά τη μελέτη των βασικών μέτρων που σχετίζονται με έναν κύκλο απασχόλησης του συστήματος, αυτή μπορεί να γίνει όπως και στα προηγούμενα παραδείγματα, δηλαδή να ξεκινήσουμε με το ότι η περίοδος αργίας I είναι $\text{Exp}(\lambda)$, οπότε $E[I] = \frac{1}{\lambda}$ και κατόπιν να λύσουμε ως προς τον μέσο κύκλο απασχόλησης τη σχέση $p_0 = \frac{E[I]}{E[Z]}$, αφού το p_0 έχει ήδη υπολογιστεί. Τέλος, η μέση περίοδος συνεχούς λειτουργίας βρίσκεται από τη σχέση $E[Y] = E[Z] - E[I]$.

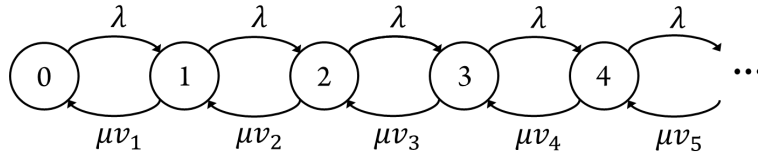
6.5.3 Η M/M/1 ουρά με μεταβλητή ταχύτητα εξυπηρέτησης

Θεωρούμε, τώρα, την M/M/1 ουρά με ρυθμό αφίξεων λ , ρυθμό εξυπηρέτησης μ και ρυθμό συνωστισμού $\rho = \frac{\lambda}{\mu}$, όπου υποθέτουμε ότι ο υπηρέτης δουλεύει με ταχύτητα v_n όταν υπάρχουν n πελάτες στο σύστημα. Αυτό σημαίνει ότι αν ένας ονομαστικός χρόνος εξυπηρέτησης X διανύεται ενώ στο σύστημα υπάρχουν n πελάτες, τότε η αντίστοιχη πραγματική διάρκεια της εξυπηρέτησης είναι X/v_n , δηλαδή η πραγματική διάρκεια εξυπηρέτησης είναι στην περίπτωση αυτή $\text{Exp}(\mu v_n)$ (λόγω της ιδιότητας 3 του θεωρήματος 3.18). Συνεπώς, ο ρυθμός εξυπηρέτησης είναι μv_n στην κατάσταση n . Σε αυτή την περίπτωση η στοχαστική διαδικασία $\{Q(t)\}$

του αριθμού των πελατών στο σύστημα είναι διαδικασία τύπου γέννησης-θανάτου με χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda & \text{αν } i \geq 0, j = i + 1, \\ \mu v_i & \text{αν } i \geq 1, j = i - 1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Το διάγραμμα ρυθμών μετάβασης για τη διαδικασία $\{Q(t)\}$ δίνεται στο σχήμα 6.3.



Σχήμα 6.3: Διάγραμμα ρυθμών μετάβασης M/M/1 ουράς με μεταβλητή ταχύτητα εξυπηρέτησης.

Στα πραγματικά συστήματα είναι λογικό να υποθέσουμε ότι η v_i είναι αύξουσα ως προς i . Εδώ θα εξετάσουμε την ειδική περίπτωση όπου $v_i = i$ που δίνει κομψά αποτελέσματα. Έχουμε

$$B^{-1} = 1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} = \sum_{n=0}^{\infty} \frac{\rho^n}{n!} = e^\rho < \infty,$$

δηλαδή το συγκεκριμένο σύστημα είναι πάντα ευσταθές. Οπότε για την κατανομή ισορροπίας του αριθμού των πελατών έχουμε

$$p_n = e^{-\rho} \frac{\rho^n}{n!}, \quad n \geq 0,$$

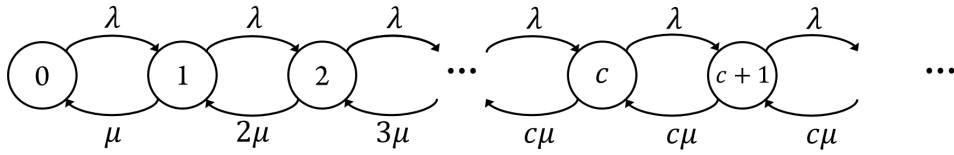
δηλαδή η (p_n) είναι Poisson(ρ). Είναι ενδιαφέρον να παρατηρήσουμε ότι η οριακή κατανομή του αριθμού των πελατών συμπίπτει με την αντίστοιχη κατανομή για το σύστημα με αποθαρρυνόμενους πελάτες που μελετήσαμε πρωτύτερα, στην παράγραφο 6.5.2, καίτοι τα δυο συστήματα είναι πολύ διαφορετικά. Η PASTA είναι εδώ εφαρμόσιμη, όπως και η ιδιότητα των μεμονωμένων μεταβάσεων οπότε οι κατανομές ισορροπίας των αριθμών των πελατών σε στιγμές αφίξεων και αναχωρήσεων συμπίπτουν με την κατανομή ισορροπίας του αριθμού των πελατών σε συνεχή χρόνο: $p_n = a_n = d_n, n \geq 0$. Ο μέσος αριθμός πελατών στο σύστημα είναι $E[Q] = \rho$. Από τον νόμο του Little έχουμε ότι ο μέσος χρόνος παραμονής στο σύστημα $E[S] = \frac{E[Q]}{\lambda} = \frac{1}{\mu}$. Τέλος, η μελέτη του κύκλου απασχόλησης του συστήματος μπορεί να γίνει όπως και στα προηγούμενα παραδείγματα.

6.6 Η M/M/c ουρά

Στην παράγραφο αυτή μελετάμε την M/M/c ουρά με ρυθμό αφίξεων λ , ρυθμό εξυπηρέτησης μ , ρυθμό συνωστισμού $\rho = \frac{\lambda}{c\mu}$, c υπηρέτες και απεριόριστη χωρητικότητα. Το μοντέλο αυτό αναφέρεται στη βιβλιογραφία και ως μοντέλο καθυστερήσεων του Erlang (Erlang delay model). Η στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι και εδώ διαδικασία τύπου γέννησης-θανάτου με χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda & \text{αν } i \geq 0, j = i + 1, \\ \min(i, c)\mu & \text{αν } i \geq 1, j = i - 1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Το διάγραμμα ρυθμών μετάβασης για τη διαδικασία $\{Q(t)\}$ δίνεται στο σχήμα 6.4.



Σχήμα 6.4: Διάγραμμα ρυθμών μετάβασης M/M/c ουράς.

Εδώ έχουμε

$$\begin{aligned} B^{-1} &= 1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} = \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \sum_{n=c}^{\infty} \frac{\rho^n}{c! c^{n-c}} \\ &= \begin{cases} \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \frac{c}{c-\rho} & \text{αν } \rho < c \text{ (ευστάθεια)} \\ \infty & \text{αν } \rho \geq c \text{ (αστάθεια)}. \end{cases} \end{aligned}$$

Οπότε για την κατανομή ισορροπίας του αριθμού των πελατών, όταν το σύστημα είναι ευσταθές, έχουμε

$$p_n = \begin{cases} \frac{B \rho^n}{n!} & \text{αν } 0 \leq n \leq c, \\ \frac{B \rho^n}{c! c^{n-c}} & \text{αν } n \geq c+1. \end{cases}$$

Λόγω της ιδιότητας των μεμονωμένων μεταβάσεων και της PASTA, έχουμε ότι και οι κατανομές ισορροπίας σε στιγμές αφίξεων και αναχωρήσεων συμπίπτουν με την κατανομή ισορροπίας σε συνεχή χρόνο.

Η πιθανότητα να καθυστερήσει ένας πελάτης και να περιμένει κάποιο χρόνο για την έναρξη της εξυπηρέτησής του είναι

$$C(c, \rho) = \sum_{n=c}^{\infty} a_n = \sum_{n=c}^{\infty} p_n = \Pr[Q \geq c] = B \frac{\rho^c}{c!} \frac{c}{c-\rho}, \quad (6.15)$$

η οποία αναφέρεται και ως τύπος καθυστερήσεων του Erlang ή τύπος Erlang C (Erlang C formula).

Η κατανομή ισορροπίας του αριθμού των πελατών στον χώρο αναμονής, δεδομένου ότι όλοι οι υπηρέτες είναι απασχολημένοι βρίσκεται ως εξής:

$$\begin{aligned} \Pr[Q_q = n | Q \geq c] &= \Pr[Q = n + c | Q \geq c] = \frac{B \frac{\rho^{c+n}}{c! c^n}}{\sum_{m=c}^{\infty} B \frac{\rho^m}{c! c^{m-c}}} \\ &= \left(1 - \frac{\rho}{c}\right) \left(\frac{\rho}{c}\right)^n, \quad n \geq 0, \end{aligned}$$

δηλαδή η δεσμευμένη κατανομή της Q_q δεδομένου του ενδεχομένου $\{Q \geq c\}$ (συμβολικά $(Q_q | Q \geq c)$) είναι $\text{Geom}(\rho/c)$ στο \mathbb{N}_0 . Οπότε έχουμε και ότι $E[Q_q | Q \geq c] = \frac{\rho/c}{1-\rho/c}$. Η μέση τιμή του αριθμού των πελατών στον χώρο αναμονής μπορεί τώρα να βρεθεί εύκολα ως εξής:

$$\begin{aligned} E[Q_q] &= \Pr[Q < c] E[Q_q | Q < c] + \Pr[Q \geq c] E[Q_q | Q \geq c] \\ &= \Pr[Q < c] \cdot 0 + \Pr[Q \geq c] \cdot \frac{\rho/c}{1-\rho/c} \\ &= B \frac{\rho^{c+1}}{(c-1)!(c-\rho)^2}. \end{aligned}$$

Εφαρμόζοντας τον νόμο του Little στον χώρο αναμονής παίρνουμε τον μέσο χρόνο αναμονής ενός πελάτη που είναι

$$E[W] = \frac{E[Q_q]}{\lambda} = B \frac{\rho^c}{(c-1)!(c-\rho)^2} \frac{1}{\mu}.$$

Ο μέσος χρόνος παραμονής είναι $E[S] = E[W] + E[B] = E[W] + \frac{1}{\mu}$, οπότε εφαρμόζοντας και πάλι τον νόμο του Little παίρνουμε τον μέσο αριθμό πελατών στο σύστημα

$$E[Q] = \lambda E[S] = B \frac{\rho^{c+1}}{(c-1)!(c-\rho)^2} + \rho.$$

Αν ενδιαφερόμαστε για την κατανομή του χρόνου παραμονής S ενός πελάτη στο σύστημα είναι βολικότερο για το συγκεκριμένο σύστημα να προσδιορίσουμε πρώτα την κατανομή του χρόνου αναμονής W . Ο υπολογισμός αυτός μπορεί να γίνει εντελώς ανάλογα με τον αντίστοιχο υπολογισμό για τον χρόνο παραμονής πελάτη στην $M/M/1$ ουρά, είτε εστιάζοντας στην συνάρτηση κατανομής είτε στον αντίστοιχο μετασχηματισμό Laplace-Stieltjes (βλέπε στην παράγραφο 6.4 τους υπολογισμούς (6.11) και (6.12) αντίστοιχα). Για οικονομία θα παρουσιάσουμε τον υπολογισμό μόνο μέσω μετασχηματισμών Laplace-Stieltjes.

Ο μετασχηματισμός Laplace-Stieltjes $\tilde{F}_W(s)$ του χρόνου αναμονής ενός πελάτη στο σύστημα βρίσκεται εύκολα δεσμεύοντας στο πλήθος των πελατών που βρίσκει ο πελάτης κατά την άφιξή του σε αυτό. Αν ο πελάτης βρει λιγότερους από c πελάτες στο σύστημα τότε ο χρόνος αναμονής του είναι μηδενικός. Αν βρεί $n \geq c$ πελάτες, τότε θα περιμένει να γίνουν $n+1-c$ αναχωρήσεις για να αρχίσει να εξυπηρετείται από κάποιον ελεύθερο υπηρέτη. Οι χρόνοι μεταξύ αυτών των εξυπηρετήσεων είναι $\text{Exp}(c\mu)$, αφού καθένας τους αντιστοιχεί στον ελάχιστο από τους c ανεξάρτητους $\text{Exp}(\mu)$ χρόνους εξυπηρέτησης που τρέχουν παράλληλα. Επομένως έχουμε

$$\begin{aligned} \tilde{F}_W(s) &= E[e^{-sW}] = \sum_{n=0}^{\infty} \Pr[Q^- = n] E[e^{-sW} | Q^- = n] \\ &= \sum_{n=0}^{c-1} \Pr[Q^- = n] \cdot 1 + \sum_{n=c}^{\infty} \Pr[Q^- = n] \cdot \left(\frac{c\mu}{s + c\mu} \right)^{n+1-c} \\ &= \sum_{n=0}^{c-1} B \frac{\rho^n}{n!} + \sum_{n=c}^{\infty} B \frac{\rho^n}{c!c^{n-c}} \left(\frac{c\mu}{s + c\mu} \right)^{n+1-c} \\ &= \sum_{n=0}^{c-1} B \frac{\rho^n}{n!} + B \frac{\rho^c}{c!} \cdot \frac{c\mu}{c\mu - \lambda} \cdot \frac{c\mu - \lambda}{s + c\mu - \lambda} \\ &= \Pr[Q < c] \cdot 1 + \Pr[Q \geq c] \frac{c\mu - \lambda}{s + c\mu - \lambda}. \end{aligned}$$

Στους υπολογισμούς αυτούς χρησιμοποιήθηκε η ιδιότητα PASTA, ενώ η σύγκλιση της σειράς εξασφαλίζεται αφού για s με $\text{Re}(s) \geq 0$ και $\rho < c$ έχουμε $\left| \frac{c\rho\mu}{\mu+s} \right| < 1$. Αυτή η μορφή του μετασχηματισμού Laplace-Stieltjes δείχνει ότι η τυχαία μεταβλητή W είναι μικτή και παίρνει την τιμή 0 με πιθανότητα $\Pr[Q < c]$, ενώ ακολουθεί την $\text{Exp}(c\mu - \lambda)$ κατανομή με πιθανότητα $\Pr[Q \geq c]$. Για τον μετασχηματισμό Laplace-Stieltjes του χρόνου παραμονής S θα έχουμε $\tilde{F}_S(s) = \tilde{F}_W(s) \frac{\mu}{s+\mu}$, καθώς $S = W+B$, με W, B ανεξάρτητες και ο χρόνος εξυπηρέτησης B έχει την $\text{Exp}(\mu)$ κατανομή.

Όσον αφορά τον προσδιορισμό του μέσου κύκλου απασχόλησης του συστήματος $E[Z]$, έχουμε καταρχήν ότι η περίοδος αργίας I έχει την $\text{Exp}(\lambda)$ κατανομή, οπότε $E[I] = \frac{1}{\lambda}$. Επίσης, έχουμε $p_0 = \frac{E[I]}{E[Z]}$, οπότε $E[Z] = \frac{E[I]}{p_0} = \frac{1}{\lambda B}$. Τέλος, η μέση περίοδος συνεχούς λειτουργίας είναι $E[Y] = E[Z] - E[I] = \frac{1}{\lambda B} - \frac{1}{\lambda}$.

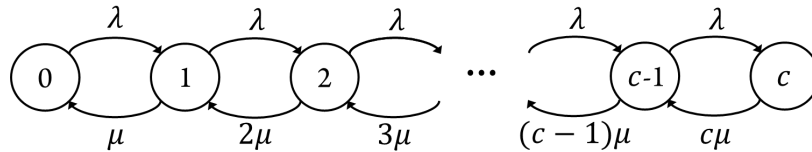
6.7 Η $M/M/c/c$ ουρά και το μοντέλο Engset

Αν στο μοντέλο της $M/M/c$ ουράς της παραγράφου 6.6 δεν υπάρχει χώρος αναμονής για τους πελάτες που δεν μπορούν να αρχίσουν να εξυπηρετούνται άμεσα, τότε έχουμε την $M/M/c/c$ ουρά, που είναι γνωστή και ως το μοντέλο απωλειών του Erlang (Erlang loss model). Η στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των

πελατών στο σύστημα είναι διαδικασία τύπου γέννησης-θανάτου με πεπερασμένο χώρο καταστάσεων $\mathcal{S} = \{0, 1, 2, \dots, c\}$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda & \text{αν } 0 \leq i \leq c-1, j = i+1, \\ i\mu & \text{αν } 1 \leq i \leq c, j = i-1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Το διάγραμμα ρυθμών μετάβασης για τη διαδικασία $\{Q(t)\}$ δίνεται στο σχήμα 6.5.



Σχήμα 6.5: Διάγραμμα ρυθμών μετάβασης M/M/c/c ουράς.

Θέτοντας $\rho = \frac{\lambda}{\mu}$, έχουμε

$$B^{-1} = 1 + \sum_{n=1}^c \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} = \sum_{n=0}^c \frac{\rho^n}{n!},$$

και η κατανομή ισορροπίας του αριθμού των πελατών είναι

$$p_n = \frac{\frac{\rho^n}{n!}}{\sum_{m=0}^c \frac{\rho^m}{m!}}, \quad 0 \leq n \leq c. \quad (6.16)$$

Λόγω της ιδιότητας των μεμονωμένων μεταβάσεων και της PASTA, έχουμε ότι και οι κατανομές ισορροπίας σε στιγμές αφίξεων και αναχωρήσεων συμπίπτουν με την κατανομή ισορροπίας σε συνεχή χρόνο, εφόσον αναφερόμαστε σε όλους τους πελάτες και όχι μόνο στους πραγματικά εισερχόμενους.

Επομένως η πιθανότητα να χαθεί ένας πελάτης αφού θα βρει όλους τους υπηρέτες κατελημμένους είναι

$$B(c, \rho) = a_c = p_c = \frac{\frac{\rho^c}{c!}}{\sum_{m=0}^c \frac{\rho^m}{m!}}, \quad (6.17)$$

η οποία αναφέρεται και ως τύπος απωλειών του Erlang ή Erlang B (Erlang B formula).

Το ενδιαφέρον είναι ότι ο τύπος των απωλειών του Erlang (6.17) και γενικότερα ο τύπος για την κατανομή ισορροπίας (6.16) ισχύει και για το αντίστοιχο M/G/c/c σύστημα, με $\rho = \lambda E[B]$, όπου B ένας τυπικός χρόνος εξυπηρέτησης με γενική κατανομή. Η ιδιότητα αυτή αναφέρεται ως ιδιότητα της μη-ευαισθησίας (insensitivity property) του M/G/c/c συστήματος, εννοώντας ότι η κατανομή του αριθμού των πελατών στο σύστημα εξαρτάται μόνο από τους ρυθμούς αφίξεων και αναχωρήσεων και δεν επηρεάζεται από τη μορφή της κατανομής των χρόνων εξυπηρέτησης. Ελάχιστα συστήματα έχουν την ιδιότητα της μη-ευαισθησίας. Όμως, όποτε αυτή υπάρχει, διευκολύνει την εφαρμογή των θεωρητικών αποτελεσμάτων, αφού αρκεί η εκτίμηση μόνο του ρυθμού εξυπηρέτησης και δεν χρειάζεται εκτίμηση της κατανομής των χρόνων εξυπηρέτησης.

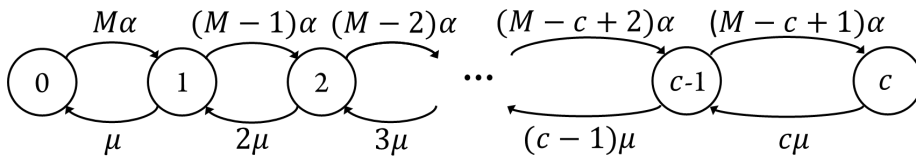
Το μοντέλο της M/M/c/c ουράς προτυποποιεί ένα σύστημα με c υπηρέτες και $\text{Exp}(\mu)$ χρόνους εξυπηρέτησης, χωρίς χώρο αναμονής, όπου ο αριθμός δυνητικών πελατών είναι μεγάλος και καθένας από αυτούς χρησιμοποιεί αραιά το σύστημα. Πράγματι, στην περίπτωση αυτή η διαδικασία των αφίξεων μπορεί να θεωρηθεί ως η υπέρθεση ενός μεγάλου αριθμού ανεξάρτητων ανανεωτικών διαδικασιών (μία για κάθε δυνητικό πελάτη) με μεγάλους μέσους ενδιάμεσους χρόνους αφίξεων (αφού κάθε δυνητικός πελάτης χρησιμοποιεί το σύστημα αραιά) και μια τέτοια υπέρθεση προσεγγίζεται καλά από τη διαδικασία αφίξεων Poisson. Όταν, όμως, ο αριθμός των δυνητικών πελατών είναι σχετικά μικρός, το κατάλληλο μοντέλο είναι το μοντέλο που εισήχθη στην εργασία Engset 1918.

Στο μοντέλο αυτό οι αφίξεις γεννώνται από «απαιτήσεις εξυπηρέτησης» M δυνητικών πελατών με $M \geq c$. Κάθε δυνητικός πελάτης αντιστοιχεί επομένως σε μια πηγή που γεννά απαιτήσεις (πραγματικούς πελάτες) σύμφωνα με μια διαδικασία Poisson με ρυθμό α . Μια απαίτηση που γεννάται όταν είναι απασχολημένοι και οι c υπηρέτες χάνεται. Κάθε δυνητικός πελάτης-πηγή γεννά απαιτήσεις μόνο εφόσον δεν υπάρχει κάποια απαίτηση του ίδιου που να βρίσκεται σε διαδικασία εξυπηρέτησης. Κατά τα άλλα, υπάρχουν c υπηρέτες και οι χρόνοι εξυπηρέτησης είναι $\text{Exp}(\mu)$, ακριβώς όπως και στο πλαίσιο της $M/M/c/c$ ουράς.

Η στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα (σε διαδικασία εξυπηρέτησης) είναι και εδώ διαδικασία τύπου γέννησης-θανάτου με χώρο καταστάσεων $\mathcal{S} = \{0, 1, 2, \dots, c\}$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} (M-i)\alpha & \text{αν } 0 \leq i \leq c-1, j = i+1, \\ i\mu & \text{αν } 1 \leq i \leq c, j = i-1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Το διάγραμμα ρυθμών μετάβασης για τη διαδικασία $\{Q(t)\}$ δίνεται στο σχήμα 6.6.



Σχήμα 6.6: Διάγραμμα ρυθμών μετάβασης του μοντέλου του Engset.

Εδώ έχουμε

$$B^{-1} = 1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} = \sum_{n=0}^c \binom{M}{n} \left(\frac{\alpha}{\mu}\right)^n = \left(\frac{\mu + \alpha}{\mu}\right)^M \sum_{n=0}^c \binom{M}{n} p^n (1-p)^{M-n},$$

όπου

$$p = \frac{\alpha}{\mu + \alpha}.$$

Οπότε, για την κατανομή ισορροπίας του αριθμού των πελατών, όταν το σύστημα είναι ευσταθές, έχουμε

$$p_n = \frac{\binom{M}{n} p^n (1-p)^{M-n}}{\sum_{i=0}^c \binom{M}{i} p^i (1-p)^{M-i}}, \quad 0 \leq n \leq c. \quad (6.18)$$

Αυτή η κατανομή είναι γενικά περικεκομμένη διωνυμική $\text{TruncBin}(M, p, c)$. Στην ειδική περίπτωση $M = c$ είναι διωνυμική $\text{Bin}(c, p)$.

Ο ρυθμός διαπέρασης του συστήματος είναι

$$\begin{aligned} \mu^* &= \sum_{n=1}^c \mu_n p_n = \sum_{n=1}^c \mu n p_n = \frac{\sum_{n=1}^c \mu n \binom{M}{n} p^n (1-p)^{M-n}}{\sum_{i=0}^c \binom{M}{i} p^i (1-p)^{M-i}} \\ &= M p \mu \frac{\sum_{m=0}^{c-1} \binom{M-1}{m} p^m (1-p)^{M-1-m}}{\sum_{i=0}^c \binom{M}{i} p^i (1-p)^{M-i}}. \end{aligned}$$

Το μακροπρόθεσμο ποσοστό των απαιτήσεων εξυπηρέτησης που βλέπουν n κατειλημμένους υπηρέτες κατά τη γέννησή τους (άφιξή τους) ισούται με το λόγο του ρυθμού γεννήσεων εξυπηρέτησης που βρίσκουν n πελάτες ανά χρονική μονάδα προς τον ρυθμό των απαιτήσεων εξυπηρέτησης που γεννώνται ανά χρονική μονάδα, οπότε είναι

$$a_n = \frac{(M-n)\alpha p_n}{\sum_{m=0}^c (M-m)\alpha p_m} = \frac{\binom{M-1}{n} p^n (1-p)^{M-1-n}}{\sum_{m=0}^c \binom{M-1}{m} p^m (1-p)^{M-1-m}}, \quad 0 \leq n \leq c, \quad (6.19)$$

δηλαδή είναι περικεκομμένη διωνυμική $\text{TruncBin}(M-1, p, c)$. Είναι αξιοσημείωτο ότι η κατανομή (a_n) έχει την ίδια μορφή με την κατανομή (p_n) με τη διαφορά ότι η παράμετρος M έχει αντικατασταθεί από $M-1$.

Το ποσοστό των χαμένων απαιτήσεων εξυπηρέτησης (χαμένων πελατών) ισούται με τον λόγο του ρυθμού των χαμένων απαιτήσεων εξυπηρέτησεων ανά χρονική μονάδα προς τον ρυθμό των απαιτήσεων εξυπηρέτησεων που γεννώνται ανά χρονική μονάδα και είναι

$$\frac{(M-c)\alpha p_c}{\sum_{m=0}^c (M-m)\alpha p_m} = \frac{\binom{M-1}{c} p^c (1-p)^{M-1-c}}{\sum_{m=0}^c \binom{M-1}{m} p^m (1-p)^{M-1-m}}. \quad (6.20)$$

Και στην περίπτωση του μοντέλου του Engset ισχύει η ιδιότητα της μη-ευαισθησίας (insensitivity property) για την οποία μιλήσαμε παραπάνω για την $M/G/c/c$ ουρά. Δηλαδή, οι τύποι για τις κατανομές ισορροπίας του αριθμού των κατελημμένων υπηρετών, σε συνεχή χρόνο και σε στιγμές αφίξεων απαιτήσεων εξυπηρέτησης, καθώς και ο τύπος για το μακροπρόθεσμο ποσοστό χαμένων εξυπηρέτησεων (τύποι (6.18), (6.19) και (6.20)) ισχύουν και στην περίπτωση που οι ενδιάμεσοι χρόνοι γεννήσεων απαιτήσεων εξυπηρέτησης και οι χρόνοι εξυπηρέτησης έχουν οποιοσδήποτε κατανομές με μέσες τιμές $\frac{1}{\alpha}$ και $\frac{1}{\mu}$, αντίστοιχα. Αυτό το αποτέλεσμα μη-ευαισθησίας απαιτεί την τεχνική συνθήκη ότι τουλάχιστον μια από αυτές τις κατανομές έχει συνεχές μέρος.

Σημειώνουμε επίσης ότι αν $\alpha = \frac{\lambda}{M}$, τότε οι τύποι του μοντέλου του Engset τείνουν στους αντίστοιχους τύπους του μοντέλου απωλειών του Erlang, καθώς $M \rightarrow \infty$. Αυτό είναι αναμενόμενο, καθώς όπως είπαμε η διαδικασία των αφίξεων προσεγγίζεται ικανοποιητικά από τη διαδικασία Poisson όταν έχουμε μεγάλο αριθμό δυναμικών πελατών που χρησιμοποιούν αραιά το σύστημα.

6.8 Ασκήσεις

Άσκηση 6.1 Θεωρούμε την τροποποίηση της $M/M/1$ ουράς με ρυθμό αφίξεων λ και $\text{Exp}(\mu)$ χρόνους εξυπηρέτησης, όπου ο χρόνος υπομονής κάθε πελάτη που περιμένει στον χώρο αναμονής έχει την κατανομή $\text{Exp}(\nu)$ και μόλις αυτός συμπληρωθεί ο πελάτης αναχωρεί.

1. Αιτιολογήστε ότι η στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου και δώστε το διάγραμμα των ρυθμών μετάβασής της.
2. Για την ειδική περίπτωση $\nu = \mu$, βρείτε την κατανομή ισορροπίας του αριθμού των πελατών σε συνεχή χρόνο.

Άσκηση 6.2 Θεωρούμε την τροποποίηση της $M/M/1$ ουράς με ρυθμό αφίξεων λ και $\text{Exp}(\mu)$ χρόνους εξυπηρέτησης, με αποθαρρυνόμενους πελάτες, όπου κάθε πελάτης που βρίσκει n πελάτες κατά την άφιξή του αναχωρεί με πιθανότητα q_n , με $q_0 = \frac{1}{4}$ και $q_n = \frac{3}{4}$ για $n \geq 1$.

1. Να διατυπωθεί η συνθήκη ευστάθειας (στασιμότητας) για το σύστημα.
2. Να υπολογιστούν οι κατανομές ισορροπίας (p_n) , (a_n) και (d_n) του αριθμού των πελατών σε συνεχή χρόνο, σε στιγμές αφίξεων και αναχωρήσεων αντίστοιχα.
3. Να προσδιοριστεί το ποσοστό των χαμένων πελατών.

Άσκηση 6.3 Θεωρούμε μια απλή Μαρκοβιανή ουρά με ρυθμούς αφίξεων και εξυπηρέτησεων $\lambda_n = \alpha^n \lambda$ ($n \geq 0$) και $\mu_n = n \alpha^n \mu$ ($n \geq 1$) αντίστοιχα, όπου $\alpha \in (0, 1)$, $\lambda, \mu > 0$ γνωστές παράμετροι.

1. Πότε η ουρά είναι ευσταθής; Να βρεθεί η κατανομή ισορροπίας (p_n) του αριθμού των πελατών στο σύστημα. Τι είδους κατανομή είναι;

2. Να βρεθούν οι κατανομές ισορροπίας (a_n) και (d_n) του αριθμού των πελατών σε στιγμές αφίξεων και αναχωρήσεων αντίστοιχα. Τι κατανομές είναι;
3. Να βρεθεί ο μέσος χρόνος παραμονής $E[S]$ ενός πελάτη στο σύστημα, καθώς και οι μέσοι χρόνοι συνεχούς λειτουργίας, αργίας και κύκλου απασχόλησης $E[Y]$, $E[I]$ και $E[Z]$ αντίστοιχα.

Άσκηση 6.4 Θεωρούμε την τροποποίηση της $M/M/c/c$ ουράς με Poisson διαδικασία αφίξεων ρυθμού λ και $\text{Exp}(\mu)$ χρόνους εξυπηρέτησης, όπου ορισμένοι πελάτες αναχωρούν από το σύστημα αμέσως μόλις αφιχθούν, χωρίς να εξυπηρετηθούν (αποθαρρυνόμενοι πελάτες). Συγκεκριμένα, κάθε πελάτης που βρίσκει n άλλα άτομα στο σύστημα κατά την άφιξή του αναχωρεί με πιθανότητα $\frac{n}{c}$, $0 \leq n \leq c$.

1. Να βρεθεί η κατανομή ισορροπίας (p_n) του αριθμού των πελατών στο σύστημα $\{Q(t)\}$. Τι είδους κατανομή είναι;
2. Να βρεθεί το μακροπρόθεσμο ποσοστό των άμεσα αποχωρούντων πελατών (χαμένων πελατών).
3. Να βρεθούν οι κατανομές ισορροπίας (a_n) και (d_n) του αριθμού των πελατών σε στιγμές αφίξεων και αναχωρήσεων αντίστοιχα, που αναφέρονται στους πραγματικά εισερχόμενους πελάτες του συστήματος.
4. Να βρεθεί ο μέσος χρόνος παραμονής ενός πελάτη στο σύστημα, $E[S]$, λαμβάνοντας υπόψη όλους τους αφιχθέντες πελάτες. Να βρεθεί ο μέσος χρόνος παραμονής ενός πελάτη στο σύστημα, $E[S']$, λαμβάνοντας υπόψη μόνο τους τελικά εισερχόμενους πελάτες.
5. Να βρεθεί ο μέσος κύκλος απασχόλησης του συστήματος, $E[Z]$.

Άσκηση 6.5 Θεωρούμε ένα σύστημα εξυπηρέτησης με έναν υπηρέτη και απεριόριστο χώρο αναμονής. Πελάτες δυο τύπων φθάνουν στο σύστημα, «υπομονετικοί» και «βιαστικοί», σύμφωνα με δυο ανεξάρτητες διαδικασίες Poisson με ρυθμούς λ_α και λ_β αντίστοιχα. Οι υπομονετικοί πελάτες εισέρχονται πάντα στο σύστημα, ενώ οι βιαστικοί μόνο όταν το συνολικό πλήθος πελατών που βρίσκουν κατά την άφιξή τους στο σύστημα είναι το πολύ $K-1$, όπου $K \geq 2$ δοσμένος ακέραιος. Οι χρόνοι εξυπηρέτησης των πελατών ανεξαρτήτως τύπου είναι ανεξάρτητες τυχαίες μεταβλητές με κατανομή $\text{Exp}(\mu)$ και η πειθαρχία ουράς είναι η FCFS.

1. Να αποδειχθεί ότι ο αριθμός των πελατών στο σύστημα, $\{Q(t)\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου και να σχεδιαστεί το διάγραμμα των ρυθμών μετάβασής της.
2. Να βρεθεί η κατανομή ισορροπίας (p_n) του αριθμού των πελατών στο σύστημα, σε συνεχή χρόνο.
3. Να βρεθεί ο ρυθμός διαπέρασης, λ^* , του συστήματος, καθώς και το μακροπρόθεσμο ποσοστό των άμεσα αποχωρούντων πελατών.
4. Να βρεθούν οι κατανομές ισορροπίας (a_n^{enter}) και (d_n^{enter}) των αριθμών των πελατών σε στιγμές αφίξεων και αναχωρήσεων, που αναφέρονται στους εισερχόμενους πελάτες.
5. Να βρεθεί ο μέσος κύκλος απασχόλησης του συστήματος.

6.9 Σχόλια

Οι απλές Μαρκοβιανές ουρές είναι τα πιο βασικά μοντέλα συστημάτων εξυπηρέτησης. Όλα τα εισαγωγικά βιβλία της Θεωρίας Ουρών Αναμονής αφιερώνουν αρκετό χώρο στη μελέτη τους. Για μια σύντομη εισαγωγή με πολλά παραδείγματα και υπολογιστικό προσανατολισμό, ο ενδιαφερόμενος αναγνώστης μπορεί να μελετήσει το σύγγραμμα των Adan και Resing 2001. Για περισσότερη λεπτομέρεια και παραδείγματα, μπορεί κανείς να δει τα κλασικά βιβλία Kleinrock 1975, Wolff 1989, Tijms 2008, Kulkarni 2010 και Shortle, Thompson, Gross και Harris 2018, ή τα νεότερα συγγράμματα των Haviv 2013 και Harchol-Balter 2013.

Βιβλιογραφία

- [1] S. Elaydi. *An Introduction to Difference Equations, 3rd Edition*. Springer, 2006. ISBN: 978-0387230597.
- [2] L.V. Ahlfors. *Complex Analysis: An Introduction to the Theory of Analytic Functions of One Complex Variable, 3rd Edition*. New York: McGraw-Hill, Inc., 1979.
- [3] T. Engset. “Die Wahrscheinlichkeitsrechnung zur Bestimmung der Wähleranzahl in automatischen Fernsprechämtern”. Στο: *Elektrotechn. Zeitschrift* 31 (1918), σσ. 304–306.
- [4] I. Adan και J. Resing. *Queueing Theory*. Eindhoven: Notes available online, 2001.
- [5] L. Kleinrock. *Queueing Systems. Volume 1: Theory*. Wiley, 1975. ISBN: 978-0471491101.
- [6] R.W. Wolff. *Stochastic Modeling and the Theory of Queues*. Englewood Cliffs, New Jersey: Prentice Hall, 1989. ISBN: 978-0138466923.
- [7] H.C. Tijms. *A First Course in Stochastic Models, 2nd Edition*. Chichester: Wiley, 2008. ISBN: 978-0471498803.
- [8] V.G. Kulkarni. *Modeling and Analysis of Stochastic Systems, 2nd Edition*. CRC Press, 2010. ISBN: 978-1439808757.
- [9] J.F. Shortle, J.M. Thompson, D. Gross και C.M. Harris. *Fundamentals of Queueing Theory, 5th Edition*. Wiley, 2018. ISBN: 978-1118943526.
- [10] M. Haviv. *Queues: A Course in Queueing Theory*. Springer, 2013. ISBN: 978-1461467649.
- [11] M. Harchol-Balter. *Performance Modeling and Design of Computer Systems: Queueing Theory in Action*. Cambridge, 2013. ISBN: 978-1107027503.

ΚΕΦΑΛΑΙΟ 7

ΓΕΝΙΚΕΣ ΜΑΡΚΟΒΙΑΝΕΣ ΟΥΡΕΣ

Στο κεφάλαιο αυτό μελετάμε ουρές στις οποίες η διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου, όχι όμως γέννησης-θανάτου. Αυτό συμβαίνει όταν οι χρόνοι μεταξύ των αφίξεων είναι εκθετικοί και οι χρόνοι εξυπηρέτησης είναι επίσης εκθετικοί, αλλά οι πελάτες έρχονται και/ή εξυπηρετούνται σε ομάδες. Το πιο εύχρηστο εργαλείο για τη μελέτη της κατανομής ισορροπίας γενικών Μαρκοβιανών ουρών είναι οι πιθανογεννήτριες. Για τη μελέτη των χρόνων παραμονής χρησιμοποιούνται μετασχηματισμοί Laplace-Stieltjes όπως και στις απλές Μαρκοβιανές ουρές.

7.1 Η M/M/1 ουρά με ομαδικές αφίξεις

Θεωρούμε την M/M/1 ουρά με ομαδικές αφίξεις. Το μοντέλο αυτό είναι παρόμοιο με την M/M/1 ουρά, με μόνη διαφορά ότι οι αφίξεις συμβαίνουν κατά ομάδες και τα μεγέθη των διαδοχικά αφικνούμενων ομάδων είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές. Πιο συγκεκριμένα μια αφικνούμενη ομάδα είναι μεγέθους j με πιθανότητα g_j . Έστω, λοιπόν μια τέτοια ουρά, με ρυθμό αφίξεων λ , συνάρτηση πιθανότητας μεγέθους αφικνούμενων ομάδων $(g_j : j \geq 1)$, $\text{Exp}(\mu)$ χρόνους εξυπηρέτησης, 1 υπηρέτη και απεριόριστη χωρητικότητα. Αν εξετάσουμε τη στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα τότε βλέπουμε ότι για τις μεταβάσεις της έχουμε τον πίνακα 7.1.

Αφού όλοι οι χρόνοι είναι εκθετικοί έχουμε ότι η $\{Q(t)\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου με

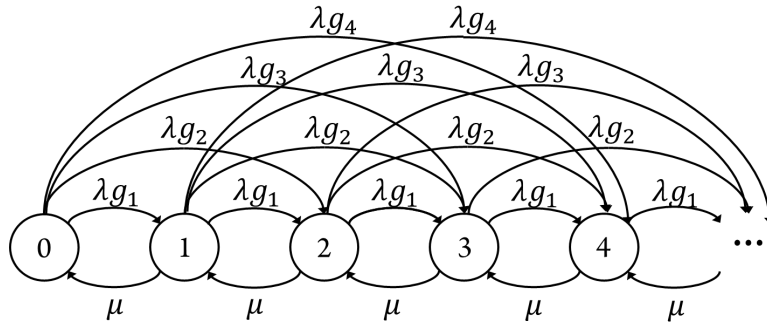
Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	$k \geq 1$	$T_{0k} \sim \text{Exp}(\lambda g_k)$
$n \geq 1$	$n + k, k \geq 1$	$T_{n,n+k} \sim \text{Exp}(\lambda g_k)$
	$n - 1$	$T_{n,n-1} \sim \text{Exp}(\mu)$

Πίνακας 7.1: Πίνακας μεταβάσεων M/M/1 ουράς με ομαδικές αφίξεις.

χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$. Οι ρυθμοί μετάβασης είναι

$$q_{ij} = \begin{cases} \lambda g_k & \text{αν } i \geq 0, j = i + k, \\ \mu & \text{αν } i \geq 1, j = i - 1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Το διάγραμμα των ρυθμών μετάβασης της $\{Q(t)\}$ παρουσιάζεται στο σχήμα 7.1.



Σχήμα 7.1: Διάγραμμα ρυθμών μετάβασης M/M/1 ουράς με ομαδικές αφίξεις.

Έστω $(p_n : n \geq 0)$ η κατανομή ισορροπίας της $\{Q(t)\}$. Οι εξισώσεις ισορροπίας είναι

$$\begin{aligned} \lambda p_0 &= \mu p_1, \\ (\lambda + \mu)p_n &= \sum_{i=0}^{n-1} \lambda g_{n-i} p_i + \mu p_{n+1}, \quad n \geq 1. \end{aligned}$$

Για την επίλυση αυτών των εξισώσεων και τον προσδιορισμό της οριακής κατανομής, θα ακολουθήσουμε τη λεγόμενη μέθοδο των πιθανογεννητριών. Για την εφαρμογή αυτής της μεθόδου εισάγουμε καταρχήν τις πιθανογεννήτριες

$$\begin{aligned} P(z) &= \sum_{n=0}^{\infty} p_n z^n, \quad |z| \leq 1, \\ G(z) &= \sum_{n=1}^{\infty} g_n z^n, \quad |z| \leq 1. \end{aligned}$$

Κατόπιν μετασχηματίζουμε τις εξισώσεις ισορροπίας ώστε να προκύψει κάποια εξίσωση για την άγνωστη πιθανογεννήτρια $P(z)$ την οποία και επιλύουμε. Στο τέλος αντιστρέφουμε την πιθανογεννήτρια, δηλαδή την αναπτύσσουμε σε δυνάμεις του z^n ώστε να προκύψουν οι πιθανότητες ισορροπίας p_n .

Για τον μετασχηματισμό των εξισώσεων ισορροπίας, πολλαπλασιάζουμε την εξίσωση που αναφέρεται στην κατάσταση n με z^n και κατόπιν αθροίζουμε για όλα τα $n \geq 0$. Έτσι έχουμε:

$$\begin{aligned} \lambda p_0 + \sum_{n=1}^{\infty} (\lambda + \mu)p_n z^n &= \mu p_1 + \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} \lambda g_{n-i} p_i z^n + \sum_{n=1}^{\infty} \mu p_{n+1} z^n \\ \Leftrightarrow (\lambda + \mu)P(z) - \mu p_0 &= \lambda \sum_{i=0}^{\infty} p_i z^i \sum_{n=i+1}^{\infty} g_{n-i} z^{n-i} + \mu \sum_{n=0}^{\infty} p_{n+1} z^n \\ \Leftrightarrow (\lambda + \mu)P(z) &= \mu p_0 + \lambda P(z)G(z) + \frac{\mu}{z}(P(z) - p_0) \\ \Leftrightarrow (\lambda + \mu - \lambda G(z) - \frac{\mu}{z})P(z) &= \mu p_0 \left(1 - \frac{1}{z}\right). \end{aligned}$$

Λύνοντας για την $P(z)$ παίρνουμε

$$P(z) = \frac{\mu p_0(1-z)}{\mu - (\lambda + \mu)z + \lambda z G(z)}.$$

Από την εξίσωση κανονικοποίησης έχουμε $\sum_{n=0}^{\infty} p_n = P(1) = 1$. Χρησιμοποιώντας αυτή τη σχέση στην έκφραση για την $P(z)$, παίρνουμε μια απροσδιόριστη μορφή $\frac{0}{0}$. Εφαρμόζοντας τον κανόνα L'Hospital έχουμε

$$1 = P(1) = \frac{-\mu p_0}{-(\lambda + \mu) + \lambda G(1) + \lambda G'(1)} = \frac{\mu p_0}{\lambda + \mu - \lambda - \lambda m'}$$

όπου

$$m = G'(1) = \sum_{i=1}^{\infty} i g_i$$

είναι το μέσο μέγεθος αφικνούμενων ομάδων. Επομένως

$$p_0 = 1 - \frac{\lambda m}{\mu}.$$

Το σύστημα είναι ευσταθές αν και μόνο αν $p_0 > 0$, δηλαδή αν

$$\frac{\lambda m}{\mu} < 1.$$

Η συνθήκη αυτή εξηγείται διαισθητικά, αφού απαιτεί ο ρυθμός αφίξεων πελατών λm να είναι μικρότερος από τον μέγιστο ρυθμό εξυπηρέτησης μ που μπορεί να παρέχει το σύστημα. Τελικά, όταν το σύστημα είναι ευσταθές, έχουμε

$$P(z) = \frac{\mu \left(1 - \frac{\lambda m}{\mu}\right) (1-z)}{\mu - (\lambda + \mu)z + \lambda z G(z)}. \quad (7.1)$$

Από την πιθανογεννήτρια μπορούμε να ανακτήσουμε την κατανομή ισορροπίας (p_n), χρησιμοποιώντας τη σχέση

$$p_n = \frac{P^{(n)}(0)}{n!}, \quad n \geq 0.$$

Επίσης μπορούμε να πάρουμε τις παραγοντικές ροπές του οριακού αριθμού πελατών στο σύστημα:

$$E[Q_{(n)}] = E[Q(Q-1)(Q-2) \cdots (Q-n+1)] = P^{(n)}(1), \quad n \geq 1.$$

Οι σχέσεις αυτές βέβαια έχουν θεωρητική κυρίως αξία. Σε ειδικές περιπτώσεις για την κατανομή του μεγέθους αφικνούμενων ομάδων μπορούμε να ανακτήσουμε την κατανομή ισορροπίας σε κλειστή μορφή.

7.1.1 Η περίπτωση των μεμονωμένων αφίξεων

Στην ειδική περίπτωση των μεμονωμένων αφίξεων, δηλαδή όταν $g_1 = 1$ και $g_j = 0$ για $j \geq 2$ έχουμε ότι $G(z) = z$, και $m = 1$, οπότε η σχέση (7.1) γίνεται

$$P(z) = \frac{\mu \left(1 - \frac{\lambda}{\mu}\right) (1-z)}{\mu - (\lambda + \mu)z + \lambda z^2} = \frac{1 - \frac{\lambda}{\mu}}{1 - \frac{\lambda}{\mu}z},$$

και αναπτύσσοντας σε δυνάμεις του z με βάση τη γεωμετρική σειρά (βλέπε (3.5)) παίρνουμε τη γνωστή γεωμετρική κατανομή του αριθμού των πελατών της M/M/1 ουράς (βλέπε (6.10)).

7.1.2 Η περίπτωση των ομάδων γεωμετρικού μεγέθους

Στην ειδική περίπτωση που η συνάρτηση πιθανότητας αφικνούμενων ομάδων ($g_i : i \geq 1$) είναι η $\text{Geom}(\alpha)$ στο \mathbb{N} , δηλαδή $g_i = (1 - \alpha)\alpha^{i-1}$, $i \geq 1$, έχουμε

$$G(z) = \frac{(1 - \alpha)z}{1 - \alpha z},$$

και $m = G'(1) = \frac{1}{1 - \alpha}$ οπότε

$$P(z) = \frac{\mu \left(1 - \frac{\lambda}{(1 - \alpha)\mu}\right) (1 - z)}{\mu - (\lambda + \mu)z + \lambda z \frac{(1 - \alpha)z}{1 - \alpha z}} = \frac{\left(1 - \frac{\lambda}{(1 - \alpha)\mu}\right) (1 - \alpha z)}{1 - \frac{\lambda + \alpha\mu}{\mu} z}.$$

Χρησιμοποιώντας τη γεωμετρική σειρά (3.5) μπορούμε να αναπτύξουμε το $\left(1 - \frac{\lambda + \alpha\mu}{\mu} z\right)^{-1}$ σε δυνάμεις του z , αφού λόγω της συνθήκης ευστάθειας έχουμε ότι $\frac{\lambda + \alpha\mu}{\mu} < 1$. Οπότε, παίρνουμε

$$\begin{aligned} P(z) &= \left(1 - \frac{\lambda}{(1 - \alpha)\mu}\right) (1 - \alpha z) \sum_{k=0}^{\infty} \left(\frac{\lambda + \alpha\mu}{\mu}\right)^k z^k \\ &= \sum_{n=0}^{\infty} \left(1 - \frac{\lambda}{(1 - \alpha)\mu}\right) \left(\frac{\lambda + \alpha\mu}{\mu}\right)^n z^n \\ &\quad - \sum_{n=1}^{\infty} \left(1 - \frac{\lambda}{(1 - \alpha)\mu}\right) \alpha \left(\frac{\lambda + \alpha\mu}{\mu}\right)^{n-1} z^n. \end{aligned}$$

Από τον ορισμό της πιθανογεννήτριας, ο συντελεστής του z^n στο δεξιό μέλος αυτής της σχέσης είναι ακριβώς η πιθανότητα p_n , για $n \geq 0$. Καθώς η πρώτη σειρά ξεκινάει από $n = 0$, ενώ η δεύτερη από $n = 1$, έχουμε ότι

$$\begin{aligned} p_0 &= \left(1 - \frac{\lambda}{(1 - \alpha)\mu}\right) \left(\frac{\lambda + \alpha\mu}{\mu}\right)^0, \\ p_n &= \left(1 - \frac{\lambda}{(1 - \alpha)\mu}\right) \left(\frac{\lambda + \alpha\mu}{\mu}\right)^n - \left(1 - \frac{\lambda}{(1 - \alpha)\mu}\right) \alpha \left(\frac{\lambda + \alpha\mu}{\mu}\right)^{n-1}, \quad n \geq 1. \end{aligned}$$

Απλοποιώντας αυτές τις παραστάσεις παίρνουμε τελικά

$$\begin{aligned} p_0 &= 1 - \frac{\lambda}{(1 - \alpha)\mu}, \\ p_n &= \left(1 - \frac{\lambda}{(1 - \alpha)\mu}\right) \frac{\lambda}{\mu} \left(\frac{\lambda + \alpha\mu}{\mu}\right)^{n-1}, \quad n \geq 1. \end{aligned}$$

7.1.3 Κατανομές ισορροπίας πλήθους πελατών σε στιγμές αφίξεων

Αν ενδιαφερόμαστε για τις πιθανότητες ισορροπίας του αριθμού πελατών σε στιγμές αφίξεων στην $M/M/1$ ουρά με ομαδικές αφίξεις, τότε είναι σημαντικό να συνειδητοποιήσουμε ότι υπάρχουν δυο ειδών τέτοιες πιθανότητες. Συμβολίζουμε με a_n^{group} την πιθανότητα μια αφικνούμενη ομάδα να βρει n πελάτες στο σύστημα κατά την άφιξή της και με a_n την πιθανότητα ένας αφικνούμενος πελάτης να βρει n πελάτες στο σύστημα κατά την άφιξή του. Λόγω της ιδιότητας PASTA έχουμε ότι $a_n^{\text{group}} = p_n$, όπου p_n είναι η οριακή πιθανότητα n πελατών στο σύστημα σε συνεχή χρόνο, την οποία έχουμε ήδη υπολογίσει. Για να βρούμε την a_n , πρέπει να διευκρινήσουμε τον τρόπο με τον οποίο οι πελάτες μιας αφικνούμενης ομάδας εισέρχονται στο σύστημα.

Θεωρούμε ότι οι πελάτες μπαίνουν ακολουθιακά, αλλά ακαριαία. Οπότε αν μια ομάδα μεγέθους k φθάσει στο σύστημα όταν υπάρχουν n πελάτες σε αυτό, τότε ο πρώτος πελάτης της θα δει n άτομα και θα μπει στην $n + 1$ θέση, ο δεύτερος θα δει $n + 1$ άτομα και θα μπει στην $n + 2$ θέση, ο τρίτος θα δει $n + 2$ άτομα και θα μπει στην $n + 3$ θέση κ.ο.κ. Επομένως έχουμε

$$a_n = \sum_{k=0}^n a_k^{group} \tilde{g}_{n-k+1} = \sum_{k=0}^n p_k \tilde{g}_{n-k+1}, \quad n \geq 0,$$

όπου \tilde{g}_n είναι η πιθανότητα ένας πελάτης να είναι ο n -οστός στην ομάδα του. Χρειάζεται, επομένως, να βρούμε τις πιθανότητες \tilde{g}_n , $n \geq 1$, όταν έχουμε τη συνάρτηση πιθανότητας του μεγέθους αφικνούμενων ομάδων ($g_n : n \geq 1$).

Η πιθανότητα \tilde{g}_n ένας πελάτης να είναι ο n -οστός της ομάδας του μπορεί να ερμηνευθεί και ως το μακροπρόθεσμο ποσοστό των πελατών που είναι n -οστοί στις ομάδες τους. Επομένως, συμβολίζοντας με Y_i , $i \geq 1$ τα μεγέθη των διαδοχικά αφικνούμενων ομάδων στο σύστημα έχουμε

$$\begin{aligned} \tilde{g}_n &= \lim_{N \rightarrow \infty} \frac{\text{πλήθος πελατών που είναι } n\text{-οστοί στις ομάδες } 1, 2, \dots, N}{Y_1 + Y_2 + \dots + Y_N} \\ &= \lim_{N \rightarrow \infty} \frac{\text{πλήθος ομάδων με μέγεθος } \geq n \text{ στις ομάδες } 1, 2, \dots, N}{Y_1 + Y_2 + \dots + Y_N}, \end{aligned}$$

διότι μια ομάδα έχει έναν n -οστό πελάτη αν και μόνο αν το μέγεθός της είναι τουλάχιστον n . Επομένως, θέτοντας $X_i = 1_{\{Y_i \geq n\}}$, $i \geq 1$, έχουμε

$$\tilde{g}_n = \lim_{N \rightarrow \infty} \frac{\sum_{i=1}^N X_i}{\sum_{i=1}^N Y_i} = \lim_{N \rightarrow \infty} \frac{\sum_{i=1}^N X_i/N}{\sum_{i=1}^N Y_i/N} = \frac{E[X_i]}{E[Y_i]},$$

όπου η τελευταία ισότητα προκύπτει από το νόμο των μεγάλων αριθμών (και ισχύει με πιθανότητα 1). Όμως $E[Y_i] = \sum_{k=1}^{\infty} k g_k$ και $E[X_i] = \Pr[Y_i \geq n] = \sum_{k=n}^{\infty} g_k$, οπότε

$$\tilde{g}_n = \frac{\sum_{k=n}^{\infty} g_k}{\sum_{k=1}^{\infty} k g_k}, \quad n \geq 1.$$

7.2 Το Μαρκοβιανό σύστημα εκκαθαρίσεων με ομαδικές αφίξεις

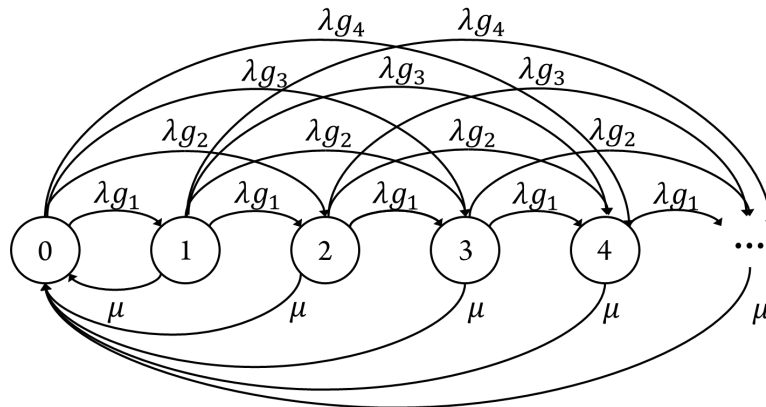
Στο Μαρκοβιανό σύστημα εκκαθαρίσεων (Markovian clearing system) με ομαδικές αφίξεις, οι πελάτες φθάνουν κατά ομάδες σύμφωνα με μια διαδικασία Poisson ρυθμού λ με συνάρτηση πιθανότητας μεγέθους αφικνούμενων ομάδων ($g_j : j \geq 1$). Υπάρχει ένα υπηρέτης που εξυπηρετεί όλους τους παρόντες πελάτες μαζί, οι χρόνοι εξυπηρέτησης είναι ανεξάρτητοι $\text{Exp}(\mu)$ και ο χώρος αναμονής είναι απεριόριστος. Ένα τέτοιο σύστημα είναι κατάλληλο για τη μοντελοποίηση της εξυπηρέτησης σε συστήματα μεταφορών. Π.χ. στην περίπτωση αυτή το σύστημα είναι μια στάση μεταφορικού μέσου και ο υπηρέτης είναι το μεταφορικό μέσο. Ο χρόνος εξυπηρέτησης αντιστοιχεί σε έναν ενδιάμεσο χρόνο μεταξύ δυο διαδοχικών επισκέψεων του μέσου από τη στάση και κάθε φορά που γίνεται επίσκεψη του μέσου η στάση αδειάζει. Για τις μεταβάσεις της στοχαστικής διαδικασίας $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα έχουμε τον πίνακα 7.2.

Αφού όλοι οι χρόνοι είναι εκθετικοί έχουμε ότι η $\{Q(t)\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου με χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$. Οι ρυθμοί μετάβασης είναι

$$q_{ij} = \begin{cases} \lambda g_k & \text{αν } i \geq 0, j = i + k, \\ \mu & \text{αν } i \geq 1, j = 0, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	$k \geq 1$	$T_{0k} \sim \text{Exp}(\lambda g_k)$
$n \geq 1$	$n+k, k \geq 1$	$T_{n,n+k} \sim \text{Exp}(\lambda g_k)$
	0	$T_{n0} \sim \text{Exp}(\mu)$

Πίνακας 7.2: Πίνακας μεταβάσεων Μαρκοβιανού συστήματος εκκαθαρίσεων με ομαδικές αφίξεις.



Σχήμα 7.2: Διάγραμμα ρυθμών μετάβασης Μαρκοβιανού συστήματος εκκαθαρίσεων με ομαδικές αφίξεις.

Το διάγραμμα ρυθμών μετάβασης της $\{Q(t)\}$ δίνεται στο σχήμα 7.2.

Έστω $(p_n : n \geq 0)$ η κατανομή ισορροπίας της $\{Q(t)\}$. Οι εξισώσεις ισορροπίας είναι

$$\lambda p_0 = \sum_{i=1}^{\infty} \mu p_i,$$

$$(\lambda + \mu)p_n = \sum_{i=0}^{n-1} \lambda g_{n-i} p_i, \quad n \geq 1.$$

Για την επίλυση αυτών των εξισώσεων και τον προσδιορισμό της οριακής κατανομής, θα ακολουθήσουμε τη μέθοδο των πιθανογεννητριών. Έστω

$$P(z) = \sum_{n=0}^{\infty} p_n z^n, \quad |z| \leq 1,$$

$$G(z) = \sum_{n=1}^{\infty} g_n z^n, \quad |z| \leq 1.$$

Για τον μετασχηματισμό των εξισώσεων ισορροπίας, πολλαπλασιάζουμε την εξίσωση που αναφέρεται στην κατάσταση n με z^n και κατόπιν αθροίζουμε για όλα τα $n \geq 0$. Έτσι έχουμε:

$$\lambda p_0 + \sum_{n=1}^{\infty} (\lambda + \mu)p_n z^n = \mu \sum_{i=1}^{\infty} p_i + \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} \lambda g_{n-i} p_i z^n$$

$$\Leftrightarrow (\lambda + \mu)P(z) - \mu p_0 = \mu(1 - p_0) + \lambda \sum_{i=0}^{\infty} p_i z^i \sum_{n=i+1}^{\infty} g_{n-i} z^{n-i}$$

$$\Leftrightarrow (\lambda + \mu)P(z) = \mu + \lambda P(z)G(z).$$

Λύνοντας για την $P(z)$ παίρνουμε

$$P(z) = \frac{\mu}{\mu + \lambda - \lambda G(z)}. \quad (7.2)$$

Εδώ το σύστημα είναι πάντα ευσταθές, καθώς επιστρέφει στην κατάσταση 0 κάθε $\text{Exp}(\mu)$ χρονικές μονάδες. Από την πιθανογεννήτρια μπορούμε να ανακτήσουμε την κατανομή ισορροπίας (p_n), χρησιμοποιώντας τη σχέση

$$p_n = \frac{P^{(n)}(0)}{n!}, \quad n \geq 0.$$

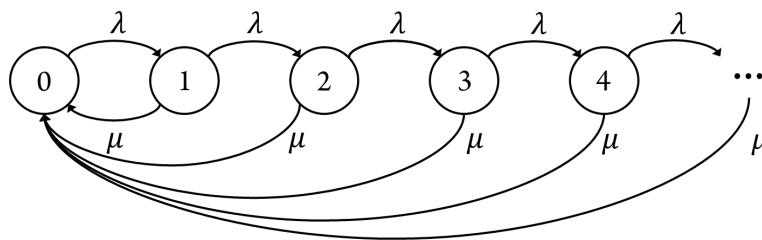
Επίσης μπορούμε να πάρουμε τις παραγοντικές ροπές του οριακού αριθμού πελατών στο σύστημα:

$$E[Q_{(n)}] = E[Q(Q-1)(Q-2)\cdots(Q-n+1)] = P^{(n)}(1), \quad n \geq 1.$$

Σε ειδικές περιπτώσεις για την κατανομή του μεγέθους αφικνούμενων ομάδων μπορούμε να ανακτήσουμε την κατανομή ισορροπίας σε κλειστή μορφή.

7.2.1 Η περίπτωση των μεμονωμένων αφίξεων

Στην ειδική περίπτωση που έχουμε μεμονωμένες αφίξεις, δηλαδή $g_1 = 1$ και $g_j = 0$ για $j \geq 2$ το διάγραμμα των ρυθμών μετάβασης της $\{Q(t)\}$ δίνεται στο σχήμα 7.3.



Σχήμα 7.3: Διάγραμμα ρυθμών μετάβασης Μαρκοβιανού συστήματος εκκαθαρίσεων με μεμονωμένες αφίξεις.

Στην περίπτωση αυτή, έχουμε ότι $G(z) = z$, οπότε η (7.2) γίνεται

$$P(z) = \frac{\mu}{\mu + \lambda - \lambda z}.$$

Αναπτύσσοντας σε δυνάμεις του z , χρησιμοποιώντας τη γεωμετρική σειρά (3.5), έχουμε

$$P(z) = \frac{\mu}{\mu + \lambda} \cdot \frac{1}{1 - \frac{\lambda}{\lambda + \mu} z} = \sum_{n=0}^{\infty} \frac{\mu}{\mu + \lambda} \left(\frac{\lambda}{\mu + \lambda} \right)^n z^n,$$

απ' όπου

$$p_n = \frac{\mu}{\mu + \lambda} \left(\frac{\lambda}{\mu + \lambda} \right)^n, \quad n \geq 0, \quad (7.3)$$

δηλαδή η $(p_n : n \geq 0)$ είναι η κατανομή $\text{Geom}\left(\frac{\lambda}{\lambda + \mu}\right)$ στο \mathbb{N}_0 .

Κατάσταση	Επόμενη κατάσταση	Χρόνος
$n, 0 \leq n \leq r-1$	$n+1$	$T_{n,n+1} \sim \text{Exp}(\lambda)$
$n \geq r$	$n+1$	$T_{n,n+1} \sim \text{Exp}(\lambda)$
	$n-r$	$T_{n,n-r} \sim \text{Exp}(\mu)$

Πίνακας 7.3: Πίνακας μεταβάσεων M/M/1 ουράς με ομαδικές εξυπηρετήσεις.

7.2.2 Η περίπτωση των ομάδων γεωμετρικού μεγέθους

Στην ειδική περίπτωση που η συνάρτηση πιθανότητας αφικνούμενων ομάδων ($g_i : i \geq 1$) είναι η $\text{Geom}(\alpha)$ στο \mathbb{N} , δηλαδή $g_i = (1-\alpha)\alpha^{i-1}$, $i \geq 1$ έχουμε

$$G(z) = \frac{(1-\alpha)z}{1-\alpha z}$$

και $m = G'(1) = \frac{1}{1-\alpha}$ οπότε η (7.2) γίνεται

$$P(z) = \frac{\mu}{\mu + \lambda - \lambda \frac{(1-\alpha)z}{1-\alpha z}} = \frac{\mu}{\lambda + \mu} (1-\alpha z) \frac{1}{1 - \frac{\lambda + \alpha\mu}{\lambda + \mu} z}$$

Αναπτύσσοντας με βάση τη γεωμετρική σειρά (3.5), έχουμε

$$\begin{aligned} P(z) &= \frac{\mu}{\lambda + \mu} (1-\alpha z) \sum_{k=0}^{\infty} \left(\frac{\lambda + \alpha\mu}{\lambda + \mu} \right)^k z^k \\ &= \sum_{n=0}^{\infty} \frac{\mu}{\lambda + \mu} \left(\frac{\lambda + \alpha\mu}{\lambda + \mu} \right)^n z^n - \sum_{n=1}^{\infty} \frac{\mu\alpha}{\lambda + \mu} \left(\frac{\lambda + \alpha\mu}{\lambda + \mu} \right)^{n-1} z^n \end{aligned}$$

και εξισώνοντας τους συντελεστές των δυνάμεων του z παίρνουμε

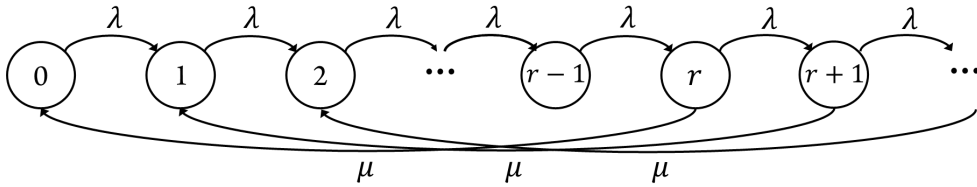
$$\begin{aligned} p_0 &= \frac{\mu}{\lambda + \mu} = 1 - \frac{\lambda}{\lambda + \mu}, \\ p_n &= \frac{\lambda\mu(1-\alpha)}{(\lambda + \mu)^2} \left(\frac{\lambda + \alpha\mu}{\lambda + \mu} \right)^{n-1}, \quad n \geq 1. \end{aligned}$$

7.3 Η M/M/1 ουρά με ομαδικές εξυπηρετήσεις

Θεωρούμε την M/M/1 ουρά με ομαδικές εξυπηρετήσεις σταθερού μεγέθους. Πιο συγκεκριμένα, οι πελάτες φθάνουν στο σύστημα σύμφωνα με μια διαδικασία Poisson με ρυθμό λ , οι χρόνοι εξυπηρέτησης είναι ανεξάρτητοι $\text{Exp}(\mu)$, υπάρχει 1 υπηρέτης που εξυπηρετεί ταυτόχρονα r πελάτες ως μια ομάδα και η χωρητικότητα του συστήματος είναι απεριόριστη. Όταν στο σύστημα υπάρχουν λιγότεροι από r πελάτες, ο υπηρέτης δεν παρέχει εξυπηρέτηση, αλλά περιμένει να συγκεντρωθούν r . Για τις μεταβάσεις της στοχαστικής διαδικασίας $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα έχουμε τον πίνακα 7.3.

Αφού όλοι οι χρόνοι είναι εκθετικοί έχουμε ότι η $\{Q(t)\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου με χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$. Οι ρυθμοί μετάβασης είναι

$$q_{ij} = \begin{cases} \lambda & \text{αν } i \geq 0, j = i+1, \\ \mu & \text{αν } i \geq r, j = i-r, \\ 0 & \text{διαφορετικά.} \end{cases}$$



Σχήμα 7.4: Διάγραμμα ρυθμών μετάβασης M/M/1 ουράς με ομαδικές εξυπηρετήσεις.

Το διάγραμμα των ρυθμών μετάβασης της $\{Q(t)\}$ δίνεται στο σχήμα 7.4.

Έστω $(p_n : n \geq 0)$ η κατανομή ισορροπίας της $\{Q(t)\}$. Οι εξισώσεις ισορροπίας είναι

$$\begin{aligned} \lambda p_0 &= \mu p_r, \\ \lambda p_n &= \lambda p_{n-1} + \mu p_{n+r}, \quad 1 \leq n \leq r-1, \\ (\lambda + \mu)p_n &= \lambda p_{n-1} + \mu p_{n+r}, \quad n \geq r. \end{aligned}$$

Για την επίλυση αυτών των εξισώσεων και τον προσδιορισμό της κατανομής ισορροπίας, θα ακολουθήσουμε τη μέθοδο των πιθανογεννητριών. Έστω $P(z) = \sum_{n=0}^{\infty} p_n z^n$, $|z| \leq 1$, η πιθανογεννήτρια της (p_n) . Πολλαπλασιάζοντας την n -οστή εξίσωση ισορροπίας με z^n και αθροίζοντας παίρνουμε:

$$\begin{aligned} \lambda p_0 + \sum_{n=1}^{r-1} \lambda p_n z^n + \sum_{n=r}^{\infty} (\lambda + \mu) p_n z^n &= \mu p_r + \sum_{n=1}^{\infty} \lambda p_{n-1} z^n + \sum_{n=1}^{\infty} \mu p_{n+r} z^n \\ \Leftrightarrow (\lambda + \mu)P(z) - \mu \sum_{n=0}^{r-1} p_n z^n &= \lambda z \sum_{n=1}^{\infty} p_{n-1} z^{n-1} + \frac{\mu}{z^r} \sum_{n=0}^{\infty} p_{n+r} z^{n+r} \\ \Leftrightarrow (\lambda + \mu)P(z) - \mu \sum_{n=0}^{r-1} p_n z^n &= \lambda z \sum_{k=0}^{\infty} p_k z^k + \frac{\mu}{z^r} \sum_{k=r}^{\infty} p_k z^k \\ \Leftrightarrow (\lambda + \mu)P(z) - \mu \sum_{n=0}^{r-1} p_n z^n &= \lambda z P(z) + \frac{\mu}{z^r} \left(P(z) - \sum_{k=0}^{r-1} p_k z^k \right). \end{aligned}$$

Λύνοντας για την $P(z)$ έχουμε

$$\begin{aligned} P(z) &= \frac{\mu \sum_{n=0}^{r-1} p_n z^n \left(1 - \frac{1}{z^r}\right)}{\lambda + \mu - \lambda z - \frac{\mu}{z^r}} \\ &= \frac{\mu \sum_{n=0}^{r-1} p_n z^n (z^r - 1)}{(\lambda + \mu)z^r - \lambda z^{r+1} - \mu} \\ &= \frac{\sum_{n=0}^{r-1} p_n z^n (z^r - 1)}{(\rho + 1)z^r - \rho z^{r+1} - 1} \end{aligned}$$

όπου $\rho = \lambda/\mu$. Για τον προσδιορισμό, επομένως, της $P(z)$ θα πρέπει να προσδιοριστούν οι r άγνωστες πιθανότητες p_0, p_1, \dots, p_{r-1} .

Έστω $N(z)$ και $D(z)$ ο αριθμητής και ο παρονομαστής του παραπάνω κλάσματος. Γνωρίζουμε ότι η παραπάνω πιθανογεννήτρια συγκλίνει για τιμές του z στον κλειστό μοναδιαίο δίσκο, όταν δηλαδή $|z| \leq 1$ (βλέπε ιδιότητα 1 στο θεώρημα 3.6 που συνοψίζει τις ιδιότητες των πιθανογεννητριών). Παρατηρούμε ότι ο βαθμός του παρονομαστή $D(z) = (1 + \rho)z^r - \rho z^{r+1} - 1$ είναι $r + 1$, και συνεπώς θα έχει $r + 1$ ρίζες, έστω τις z_0, z_1, \dots, z_r . Ας υποθέσουμε ότι έχει μια ρίζα στο εσωτερικό του μοναδιαίου δίσκου και έστω χωρίς βλάβη της γενικότητας ότι είναι η ρίζα z_0 , με $|z_0| < 1$. Τότε η $P(z)$ θα έχει πόλο στο z_0 , αν $N(z_0) \neq 0$, το οποίο είναι άτοπο, εφόσον η

Συνθήκες	ρίζες $\in \{ z < 1\}$	ρίζες $\in \{ z = 1\}$	ρίζες $\in \{ z \leq 1\}$
$a(1) < 1$	N	0	N
$a(1) = 1, A'(1) > 0$	$N - k$	k (απλές, τις k -οστές ρίζες της 1)	N
$a(1) = 1, A'(1) = 0$	$N - k$	k (διπλές, τις k -οστές ρίζες της 1)	$N + k$
$a(1) = 1, A'(1) < 0$	N	k (απλές, τις k -οστές ρίζες της 1)	$N + k$

Πίνακας 7.4: Πίνακας πλήθους ριζών της $A(z) = z^N - a(z)$ για πιθανογεννήτρια $a(z)$ και θετικό ακέραιο N .

πιθανογεννήτρια $P(z)$ συγκλίνει στο $\{z : |z| \leq 1\}$. Συμπεραίνουμε, λοιπόν, ότι η ρίζα z_0 θα πρέπει να είναι και ρίζα του αριθμητή $N(z) = (z^r - 1) \sum_{n=0}^{r-1} p_n z^n$ (βλέπε και σχετική συζήτηση στην παράγραφο 3.2 και ιδιαίτερα τον αλγόριθμο 3.8 σχετικά με την αναλυτική αντιστροφή ρητών πιθανογεννητριών και το παράδειγμα 3.9). Κατ' επέκταση κάθε ρίζα του παρονομαστή $D(z) = (1 + \rho)z^r - \rho z^{r+1} - 1$ που βρίσκεται στον μοναδιαίο δίσκο, δηλαδή κάθε $z_i \in \{z : |z| \leq 1\}$, θα πρέπει να είναι και ρίζα του αριθμητή $N(z) = (z^r - 1) \sum_{n=0}^{r-1} p_n z^n$.

Για να προχωρήσουμε, λοιπόν στην απλοποίηση της μορφής της $P(z)$ πρέπει να βρούμε τις ρίζες του παρονομαστή $D(z)$ που βρίσκονται στον μοναδιαίο δίσκο, δηλαδή πρέπει να προσδιορίσουμε το πλήθος των $z_i \in \{z : |z| \leq 1\}$. Τη λύση στο πρόβλημα μας δίνει το παρακάτω θεώρημα που προκύπτει ως εφαρμογή του θεωρήματος Rouché από τη Μιγαδική Ανάλυση (π.χ. βλέπε Παράγραφο 4.5.2 (The Argument Principle) στο βιβλίο Ahlfors 1979).

Θεώρημα 7.1 (Πόρισμα Θεωρήματος Rouché) Έστω μια ακολουθία $\{a_n\}_{n \in \mathbb{N}_0}$ με $a_n \geq 0$ για την οποία ισχύουν

$$\sum_{n=0}^{\infty} a_n \leq 1, \quad \sum_{n=0}^{\infty} n a_n < \infty \quad \text{και} \quad \sum_{n=0}^{\infty} n^2 a_n < \infty.$$

Έστω N θετικός ακέραιος. Θέτουμε

$$a(z) = \sum_{n=0}^{\infty} a_n z^n, \quad A(z) = z^N - a(z) \quad \text{και}$$

$$k = \text{MK}\Delta\{j - N \mid \text{για τους δείκτες } j \text{ που } a_j \neq 0\},$$

(MKΔ: μέγιστος κοινός διαιρέτης). Τότε, σχετικά με τα πλήθη των ριζών της $A(z)$ στα σύνολα $\{z : |z| < 1\}$, $\{z : |z| = 1\}$ και $\{z : |z| \leq 1\}$ έχουμε τον πίνακα 7.4.

Για μια διεξοδική συζήτηση του θεωρήματος 7.1 και των εφαρμογών του, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο άρθρο των Gail, Hantler και Taylor 2000 (δείτε θεωρήματα 5 και 6). Η απόδειξη που χρησιμοποιεί μεθόδους μιγαδικής ανάλυσης δόθηκε στην εργασία Gail, Hantler και Taylor 1996.

Σκοπός μας είναι τώρα να απλοποιήσουμε την πιθανογεννήτρια $P(z)$, χρησιμοποιώντας το θεώρημα αυτό για να μελετήσουμε το πλήθος των ριζών του παρονομαστή $D(z)$ στον κλειστό μοναδιαίο δίσκο. Βλέπουμε ότι ο παρονομαστής $D(z)$ γράφεται στη μορφή

$$D(z) = (1 + \rho) \left(z^r - \frac{1}{1 + \rho} - \frac{\rho}{1 + \rho} z^{r+1} \right).$$

Για να εφαρμόσουμε το θεώρημα θέτουμε $N = r$ και ορίζουμε την ακολουθία $(a_n)_{n \in \mathbb{N}_0}$ με $a_0 = \frac{1}{1+\rho}$, $a_1 = a_2 = \dots = a_r = 0$, $a_{r+1} = \frac{\rho}{1+\rho}$ και $a_n = 0$ για $n \geq r+2$. Τότε, για την ακολουθία που ορίσαμε, ισχύουν οι υποθέσεις του θεωρήματος:

$$\sum_{n=0}^{\infty} a_n = 1, \quad \sum_{n=0}^{\infty} n a_n < \infty \quad \text{και} \quad \sum_{n=0}^{\infty} n^2 a_n < \infty.$$

Θέτουμε

$$a(z) = \sum_{n=0}^{\infty} a_n z^n = \frac{1}{1+\rho} + \frac{\rho}{1+\rho} z^{r+1}, \quad A(z) = z^r - a(z).$$

Εδώ

$$\begin{aligned} k &= \text{MK}\Delta\{j-r \mid \text{για τους δείκτες } j \text{ που } a_j \neq 0\}, \\ &= \text{MK}\Delta\{j-r \mid \text{για } j=0 \text{ και } j=r+1\} = \text{MK}\Delta\{-r, 1\} = 1. \end{aligned}$$

Για να δούμε σε ποια από τις τέσσερις περιπτώσεις του θεωρήματος εμπίπτει η συγκεκριμένη κατάσταση υπολογίζουμε την $a(1)$ και $A'(1)$. Είναι

$$a(1) = \frac{1}{1+\rho} + \frac{\rho}{1+\rho} = 1$$

και

$$A'(1) = r - (r+1) \frac{\rho}{1+\rho} = \frac{r-\rho}{1+\rho}.$$

Επομένως διακρίνουμε τρεις περιπτώσεις:

(i) $A'(1) > 0 \Leftrightarrow \rho < r$. Στην περίπτωση αυτή εφαρμόζοντας το θεώρημα 7.1 έπεται ότι το πολυώνυμο $D(z)$ έχει $r-1$ ρίζες στο εσωτερικό του μοναδιαίου δίσκου, έστω τις z_1, z_2, \dots, z_{r-1} , δηλαδή $z_i \in \{z : |z| < 1\}$ για $i = 1, \dots, r-1$, έχει 1 ρίζα τη μονάδα, έστω $z_r = 1$, και άρα θα έχει μια ρίζα, έστω z_0 , με $|z_0| > 1$, άρα $D(z) = c_1(z-z_0)(z-z_1) \dots (z-z_{r-1})(z-1)$, όπου c_1 κάποια σταθερά. Επειδή η $P(z)$ συγκλίνει στον κλειστό μοναδιαίο δίσκο, ως πιθανογεννήτρια, συνάγουμε ότι ο αριθμητής $N(z)$ έχει ρίζες τις z_1, z_2, \dots, z_{r-1} , καθώς και την 1. Αλλά οι ρίζες του $z^r - 1$ είναι οι r -οστές ρίζες της μονάδας που έχουν όλες μέτρο 1, οπότε οι ρίζες z_1, z_2, \dots, z_{r-1} είναι αναγκαστικά ρίζες του άλλου παράγοντα του αριθμητή $N(z)$, δηλαδή του $\sum_{n=0}^{r-1} p_n z^n$ που έχει βαθμό $r-1$, όσο δηλαδή και το πλήθος των z_1, z_2, \dots, z_{r-1} . Επομένως έχουμε αναγκαστικά ότι $\sum_{n=0}^{r-1} p_n z^n = c_2(z-z_1)(z-z_2) \dots (z-z_{r-1})$, όπου c_2 κάποια σταθερά, οπότε $N(z) = c_2(z^r - 1)(z-z_1)(z-z_2) \dots (z-z_{r-1})$. Τελικά:

$$\begin{aligned} P(z) &= \frac{N(z)}{D(z)} = \frac{c_2(z^r - 1)(z-z_1)(z-z_2) \dots (z-z_{r-1})}{c_1(z-z_0)(z-z_1)(z-z_2) \dots (z-z_{r-1})(z-1)} \\ &= c \frac{z^r - 1}{(z-z_0)(z-1)}, \end{aligned}$$

όπου $c = c_2/c_1$ μια σταθερά που θα προσδιοριστεί από την εξίσωση κανονικοποίησης $P(1) = 1$. Πράγματι, χρησιμοποιώντας τη σχέση αυτή και εφαρμόζοντας τον κανόνα του L' Hospital παίρνουμε $c = \frac{1-z_0}{r}$. Επομένως η $P(z)$ παίρνει την απλή μορφή

$$P(z) = \frac{(z_0 - 1)(z^r - 1)}{r(z_0 - z)(z - 1)},$$

όπου η z_0 είναι η μοναδική ρίζα του $D(z)$ με μέτρο μεγαλύτερο της μονάδας. Για τον υπολογισμό της z_0 , παρατηρήστε ότι $D(z_0) = 0$ σημαίνει ότι

$$z_0^r - \frac{1}{1+\rho} - \frac{\rho}{1+\rho} z_0^{r+1} = 0,$$

ή ισοδύναμα (διαιρώντας με z_0^{r+1}) αν

$$\frac{1}{z_0} - \frac{1}{1+\rho} \left(\frac{1}{z_0} \right)^{r+1} - \frac{\rho}{1+\rho} = 0.$$

Επομένως η z_0 μπορεί να χαρακτηριστεί ως η αντίστροφη της μοναδικής ρίζας του πολυωνύμου $f(z) = z - \frac{1}{1+\rho} z^{r+1} - \frac{\rho}{1+\rho}$ με μέτρο μικρότερο της μονάδας. Όμως, έχουμε $f(0) < 0$, $f(1) = 0$ και $f'(1) = \frac{\rho-r}{\rho+1} < 0$ οπότε το $f(z)$ έχει πραγματική ρίζα στο $(0, 1)$ που μπορεί να βρεθεί με τη μέθοδο της διχοτόμησης από την Αριθμητική Ανάλυση.

Έχοντας προσδιορίσει την z_0 με αυτό τον τρόπο, μπορούμε να αντιστρέψουμε εύκολα την πιθανογεννήτρια $P(z)$ και να προσδιορίσουμε τη στάσιμη κατανομή (p_n) . Έχουμε:

$$\begin{aligned} P(z) &= \frac{z_0 - 1}{z_0 - z} \cdot \frac{z^r - 1}{r(z - 1)} \\ &= \frac{1 - z_0^{-1}}{1 - z_0^{-1}z} \cdot \frac{z^r - 1}{r(z - 1)}. \end{aligned}$$

Αναγνωρίζοντας ότι ο πρώτος παράγοντας αντιστοιχεί στην πιθανογεννήτρια της $\text{Geom}(z_0^{-1})$ στο \mathbb{N}_0 με συνάρτηση πιθανότητας $a_n = (1 - z_0^{-1})(z_0^{-1})^n$, για $n \geq 0$, και ο δεύτερος παράγοντας αντιστοιχεί στην πιθανογεννήτρια της $\text{Uniform}(\{0, 1, \dots, r-1\})$ με συνάρτηση πιθανότητας $b_n = 1/r$, για $0 \leq n \leq r-1$ και $b_n = 0$, για $n \geq r$, έχουμε ότι η κατανομή ισορροπίας είναι η συνάρτηση πιθανότητας του αθροίσματος των ανεξάρτητων τυχαίων μεταβλητών με συναρτήσεις πιθανότητας (a_n) και (b_n) και συνεπώς δίνεται από τη σχέση

$$\begin{aligned} p_n &= \sum_{k=0}^n b_k a_{n-k} \\ &= \sum_{k=0}^{\min(n, r-1)} \frac{1}{r} (1 - z_0^{-1})(z_0^{-1})^{n-k}. \end{aligned}$$

- (ii) $A'(1) = 0 \Leftrightarrow \rho = r$. Στην περίπτωση αυτή, εφαρμόζοντας το θεώρημα 7.1, έπεται ότι το πολυώνυμο $D(z)$ έχει $r-1$ ρίζες στο εσωτερικό του μοναδιαίου δίσκου, έστω τις z_1, z_2, \dots, z_{r-1} , με $z_i \in \{z : |z| < 1\}$ για $i = 1, \dots, r-1$ και έχει 1 διπλή ρίζα τη μονάδα. Δεδομένου ότι το $D(z)$ είναι βαθμού $r+1$ παραγοντοποιείται αναγκαστικά ως $D(z) = c_1(z - z_1)\dots(z - z_{r-1})(z - 1)^2$. Για να συγκλίνει η $P(z)$ στον κλειστό μοναδιαίο δίσκο θα πρέπει ο αριθμητής $N(z)$ να έχει ρίζες τις z_1, z_2, \dots, z_{r-1} καθώς και διπλή ρίζα τη μονάδα. Οι ρίζες του $z^r - 1$ είναι οι r -οστές ρίζες της μονάδας που έχουν όλες μέτρο 1, οπότε οι ρίζες z_1, z_2, \dots, z_{r-1} είναι αναγκαστικά ρίζες του άλλου παράγοντα του αριθμητή $N(z)$, δηλαδή του $\sum_{n=0}^{r-1} p_n z^n$ που έχει βαθμό $r-1$, όσο δηλαδή και το πλήθος των z_1, z_2, \dots, z_{r-1} . Επομένως έχουμε αναγκαστικά ότι $\sum_{n=0}^{r-1} p_n z^n = c_2(z - z_1)(z - z_2)\dots(z - z_{r-1})$, όπου c_2 κάποια σταθερά, οπότε $N(z) = (z^r - 1) \sum_{n=0}^{r-1} p_n z^n = c_2(z^r - 1)(z - z_1)\dots(z - z_{r-1})$. Όμως το 1 δεν μπορεί να είναι διπλή ρίζα του $N(z)$, οπότε στην περίπτωση αυτή το μόνο που είναι δυνατό είναι το $N(z)$ να είναι ταυτοτικά 0, οπότε και $P(z) = 0$. Δηλαδή στην περίπτωση αυτή το σύστημα δεν είναι ευσταθές.
- (iii) $A'(1) < 0 \Leftrightarrow \rho > r$. Στην περίπτωση αυτή εφαρμόζοντας το θεώρημα 7.1 έπεται ότι το πολυώνυμο $D(z)$ έχει r ρίζες στο εσωτερικό του μοναδιαίου δίσκου, έστω τις z_1, z_2, \dots, z_r και έχει 1 απλή ρίζα τη μονάδα, άρα $D(z) = c_1(z - z_1)\dots(z - z_r)(z - 1)$. Από τον προηγούμενο ισχυρισμό απαιτούμε ο αριθμητής $N(z)$ να έχει ρίζες τις z_1, z_2, \dots, z_r , καθώς και την 1. Θα έπρεπε (ανάλογα με την προηγούμενη περίπτωση) το πολυώνυμο $(r-1)$ -οστού βαθμού $\sum_{n=0}^{r-1} p_n z^n$ να έχει r ρίζες, δηλαδή να είναι ταυτοτικά 0. Οπότε στην περίπτωση αυτή έχουμε $P(z) = 0$ και το σύστημα δεν είναι ευσταθές.

Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	$k \geq 1$	$T_{0k} \sim \text{Exp}(\lambda g_k)$
$n \geq 1$	$n+k, k \geq 1$	$T_{n,n+k} \sim \text{Exp}(\lambda g_k)$
	$n-1$	$T_{n,n-1} \sim \text{Exp}(n\mu)$

Πίνακας 7.5: Πίνακας μεταβάσεων M/M/∞ ουράς με ομαδικές αφίξεις.

Βλέπουμε, λοιπόν, ότι το θεώρημα 7.1 μας επιτρέπει να χαρακτηρίσουμε πότε το σύστημα είναι ευσταθές και επιπλέον να βρούμε την αντίστοιχη κατανομή ισορροπίας. Η προκύπτουσα συνθήκη ευστάθειας είναι φυσιολογική: Για να είναι το σύστημα ευσταθές θα πρέπει ο ρυθμός αφίξεων να είναι μικρότερος από τον μέγιστο ρυθμό εξυπηρέτησης του συστήματος, δηλαδή θα πρέπει $\lambda < r\mu$ ή ισοδύναμα $\rho < r$.

Το θεώρημα 7.1 και γενικότερα το θεώρημα Rouché είναι γενικά πολύ χρήσιμα για τον προσδιορισμό άγνωστων σταθερών που υπεισέρχονται στον τύπο μιας πιθανογεννήτριας που προκύπτει κατά την εφαρμογή της μεθόδου των πιθανογεννητριών συναρτήσεων.

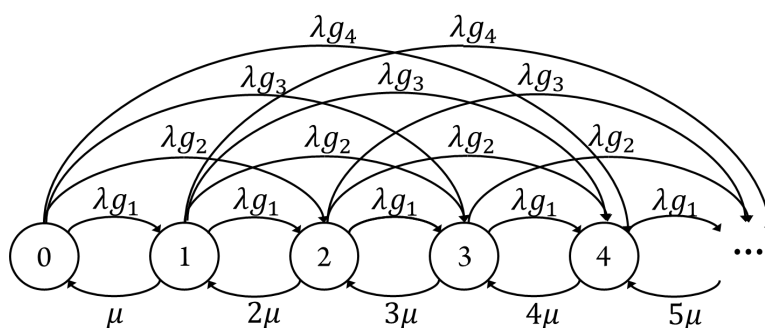
7.4 Η M/M/άπειρο ουρά με ομαδικές αφίξεις

Θεωρούμε την M/M/∞ ουρά με ομαδικές αφίξεις και υποθέτουμε ότι τα μεγέθη των διαδοχικά αφικνούμενων ομάδων είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές. Πιο συγκεκριμένα μια αφικνούμενη ομάδα είναι μεγέθους j με πιθανότητα g_j . Έστω, λοιπόν μια τέτοια ουρά, με ρυθμό αφίξεων ομάδων λ , συνάρτηση πιθανότητας μεγέθους αφικνούμενων ομάδων ($g_j : j \geq 1$), $\text{Exp}(\mu)$ χρόνους εξυπηρέτησης και άπειρους υπηρέτες (επομένως και η χωρητικότητα είναι απεριόριστη). Για τις μεταβάσεις της στοχαστικής διαδικασίας $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα έχουμε τον πίνακα 7.5.

Αφού όλοι οι χρόνοι είναι εκθετικοί έχουμε ότι η $\{Q(t)\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου με χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$. Οι ρυθμοί μετάβασης είναι

$$q_{ij} = \begin{cases} \lambda g_k & \text{αν } i \geq 0, j = i + k, \\ i\mu & \text{αν } i \geq 1, j = i - 1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Το διάγραμμα των ρυθμών μετάβασης της $\{Q(t)\}$ δίνεται στο σχήμα 7.5.



Σχήμα 7.5: Διάγραμμα ρυθμών μετάβασης M/M/∞ ουράς με ομαδικές αφίξεις.

Έστω $(p_n : n \geq 0)$ η κατανομή ισορροπίας της $\{Q(t)\}$. Οι εξισώσεις ισορροπίας είναι

$$\begin{aligned} \lambda p_0 &= \mu p_1, \\ (\lambda + n\mu)p_n &= \sum_{i=0}^{n-1} \lambda g_{n-i} p_i + (n+1)\mu p_{n+1}, \quad n \geq 1. \end{aligned}$$

Για την επίλυση αυτών των εξισώσεων και τον προσδιορισμό της οριακής κατανομής, ακολουθούμε και πάλι τη μέθοδο των πιθανογεννητριών. Έστω οι πιθανογεννήτριες

$$P(z) = \sum_{n=0}^{\infty} p_n z^n, \quad |z| \leq 1,$$

$$G(z) = \sum_{n=1}^{\infty} g_n z^n, \quad |z| \leq 1.$$

Για τον μετασχηματισμό των εξισώσεων ισορροπίας, πολλαπλασιάζουμε την εξίσωση που αναφέρεται στην κατάσταση n με z^n και κατόπιν αθροίζουμε για όλα τα $n \geq 0$. Έτσι έχουμε:

$$\lambda p_0 + \sum_{n=1}^{\infty} (\lambda + n\mu) p_n z^n = \mu p_1 + \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} \lambda g_{n-i} p_i z^n + \sum_{n=1}^{\infty} (n+1) \mu p_{n+1} z^n. \quad (7.4)$$

Όπως στην ανάλυση της M/M/1 ουράς με ομαδικές αφίξεις, έχουμε

$$\sum_{n=1}^{\infty} \sum_{i=0}^{n-1} g_{n-i} p_i z^n = P(z)G(z).$$

Για να προχωρήσουμε, βλέπουμε ότι πρέπει να εκφράσουμε με όρους πιθανογεννητριών τις σειρές $\sum_{n=1}^{\infty} n p_n z^n$ και $\sum_{n=0}^{\infty} (n+1) p_{n+1} z^n$. Παραγωγίζοντας την $P(z)$ παίρνουμε

$$P'(z) = \sum_{n=1}^{\infty} n p_n z^{n-1},$$

οπότε

$$\sum_{n=1}^{\infty} n p_n z^n = z P'(z) \quad \text{και} \quad \sum_{n=0}^{\infty} (n+1) p_{n+1} z^n = P'(z).$$

Επομένως η εξίσωση (7.4) που προέκυψε από το μετασχηματισμό των εξισώσεων ισορροπίας παίρνει τη μορφή:

$$\begin{aligned} \lambda P(z) + \mu z P'(z) &= \lambda P(z)G(z) + \mu P'(z) \\ \Leftrightarrow \lambda(1 - G(z))P(z) &= \mu(1 - z)P'(z), \end{aligned}$$

δηλαδή αυτή τη φορά η μέθοδος των πιθανογεννητριών έδωσε μια διαφορική εξίσωση για την $P(z)$. Αυτό συμβαίνει γενικά όταν υπάρχουν ρυθμοί $q_{i,i-1}$ ανάλογοι του i , όπως εδώ που έχουμε άπειρους υπηρέτες ή παρόμοιες καταστάσεις. Η προκύπτουσα διαφορική εξίσωση είναι ομογενής γραμμική πρώτης τάξης και λύνεται εύκολα:

$$\begin{aligned} \frac{P'(z)}{P(z)} &= \frac{\lambda}{\mu} \cdot \frac{1 - G(z)}{1 - z} \\ \Leftrightarrow (\log P(z))' &= \frac{\lambda}{\mu} \cdot \frac{1 - G(z)}{1 - z} \\ \Leftrightarrow \log P(1) - \log P(z) &= \frac{\lambda}{\mu} \int_z^1 \frac{1 - G(u)}{1 - u} du. \end{aligned}$$

Λύνοντας για την $P(z)$ έχουμε

$$P(z) = \exp\left(-\frac{\lambda}{\mu} \int_z^1 \frac{1 - G(u)}{1 - u} du\right).$$

Από την πιθανογεννήτρια μπορούμε θεωρητικά να ανακτήσουμε την κατανομή ισορροπίας (p_n), χρησιμοποιώντας τη σχέση

$$p_n = \frac{P^{(n)}(0)}{n!}, \quad n \geq 0,$$

και να πάρουμε τις παραγοντικές ροπές του οριακού αριθμού πελατών στο σύστημα:

$$E[Q_{(n)}] = E[Q(Q-1)(Q-2)\cdots(Q-n+1)] = P^{(n)}(1), \quad n \geq 1.$$

Βέβαια οι σχέσεις αυτές στην πράξη μπορούν να χρησιμοποιηθούν μόνο για μικρά n , π.χ. όταν μας ενδιαφέρει η πιθανότητα κενού συστήματος ή η μέση τιμή και η διασπορά του αριθμού των πελατών στο σύστημα. Σε λίγες ειδικές περιπτώσεις για την κατανομή του μεγέθους αφικνούμενων ομάδων (π.χ. γεωμετρικής μορφής όπως στην παράγραφο 7.1.2) μπορούμε να ανακτήσουμε την κατανομή ισορροπίας σε κλειστή μορφή.

7.5 Ασκήσεις

Άσκηση 7.1 Έστω μια $M/M/1$ ουρά με ομαδικές εξυπηρετήσεις σταθερού μεγέθους, όπου οι πελάτες φθάνουν σύμφωνα με μια διαδικασία Poisson με ρυθμό λ , οι χρόνοι εξυπηρέτησης είναι $\text{Exp}(\mu)$, υπάρχει 1 υπηρέτης που εξυπηρετεί ταυτόχρονα r πελάτες ως μια ομάδα και η χωρητικότητα του συστήματος είναι απεριόριστη. Θεωρούμε την τροποποίησή της με τις ίδιες παραμέτρους, όπου ο υπηρέτης δεν περιμένει για τη συμπλήρωση r πελατών, αλλά εξυπηρετεί όταν υπάρχει τουλάχιστον ένας πελάτης στο σύστημα. Αν ένας χρόνος εξυπηρέτησης τελειώσει και υπάρχουν λιγότεροι από r πελάτες στο σύστημα, τότε αναχωρούν όλοι, ενώ αν υπάρχουν περισσότεροι από r , τότε αναχωρούν ακριβώς r .

1. Να γραφούν οι εξισώσεις ισορροπίας για την κατανομή ισορροπίας (p_n) του αριθμού των πελατών στο σύστημα σε συνεχή χρόνο.
2. Αποδείξτε ότι όταν $\lambda < r\mu$ τότε το σύστημα είναι ευσταθές και βρείτε την κατανομή ισορροπίας του. Τι είδους κατανομή είναι;

Άσκηση 7.2 Θεωρούμε την $M/M/1$ ουρά με ομαδικές αφίξεις και μεμονωμένες εξυπηρετήσεις, όπου ομάδες πελατών μεγέθους 2 φθάνουν σύμφωνα με μια στοχαστική διαδικασία Poisson με ρυθμό λ , οι χρόνοι εξυπηρέτησης είναι $\text{Exp}(\mu)$, υπάρχει ένας υπηρέτης και απεριόριστη χωρητικότητα. Έστω $\{Q(t)\}$ η στοχαστική διαδικασία του αριθμού των πελατών στο σύστημα.

1. Να αιτιολογήσετε γιατί η $\{Q(t)\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου και να κάνετε το διάγραμμα των ρυθμών μετάβασής της.
2. Να υπολογίσετε την πιθανογεννήτρια $P(z)$ της κατανομής ισορροπίας (p_n) της $\{Q(t)\}$, συναρτήσει των λ , μ και ρ_0 .
3. Να βρεθεί η συνθήκη ευστάθειας του συστήματος και να υπολογιστεί η ρ_0 .
4. Να βρεθεί συναρτήσει των πιθανοτήτων p_n , η πιθανότητα ένας πελάτης αμέσως μετά την άφιξή του να καταλάβει την n -οστή θέση στο σύστημα (δηλαδή να έχει μπροστά του $n-1$ πελάτες).
5. Για $\lambda = 1$ και $\mu = 6$ να βρείτε έναν γενικό τύπο για την p_n .

Άσκηση 7.3 Θεωρούμε την $M/M/1$ ουρά με ομαδικές αφίξεις και μεμονωμένες εξυπηρετήσεις. Οι ομάδες φθάνουν στο σύστημα με ρυθμό λ και το μέγεθός τους έχει συνάρτηση πιθανότητας $(g_j : j \geq 1)$. Μια ομάδα που βρίσκεται κατά την άφιξή της το σύστημα κενό εισέρχεται με πιθανότητα 1 σε αυτό, ενώ αν βρει έστω κι έναν πελάτη εισέρχεται με πιθανότητα p . Ο ρυθμός εξυπηρέτησης είναι μ . Έστω $\{Q(t)\}$ η στοχαστική διαδικασία του αριθμού των πελατών στο σύστημα.

1. Να αιτιολογήσετε γιατί η $\{Q(t)\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου και να κάνετε το διάγραμμα των ρυθμών μετάβασής της. Γράψτε τις εξισώσεις ισορροπίας για την κατανομή ισορροπίας της, (p_n) .
2. Να υπολογίσετε την πιθανογεννήτρια $P(z)$ της κατανομής ισορροπίας και τη συνθήκη ευστάθειας του συστήματος.
3. Να βρεθεί συναρτήσει των (p_n) και (g_n) , η πιθανότητα ένας πελάτης αμέσως μετά την άφιξή του στο σύστημα να καταλάβει την n -οστή θέση (να έχει μπροστά του $n - 1$ πελάτες).
4. Στην περίπτωση των μεμονωμένων αφίξεων (δηλαδή όταν $g_1 = 1$ και $g_j = 0$ για $j \geq 2$) να βρεθεί η κατανομή ισορροπίας (p_n) .

Άσκηση 7.4 Θεωρούμε την $M/M/1/2$ ουρά με ομαδικές αφίξεις, όπου ο ρυθμός αφίξεων ομάδων είναι λ , ο ρυθμός εξυπηρέτησης είναι μ και κάθε αφικνούμενη ομάδα έχει μέγεθος 1 με πιθανότητα $\frac{1}{2}$ και μέγεθος 2 με πιθανότητα $\frac{1}{2}$. Σε περίπτωση που μια αφικνούμενη ομάδα δεν χωράει να εισέλθει ολόκληρη στο σύστημα, τότε απορρίπτεται ολόκληρη (κανένας πελάτης-μέλος της δεν εισέρχεται).

1. Να γραφούν οι εξισώσεις ισορροπίας για την κατανομή ισορροπίας (p_n) του αριθμού των πελατών στο σύστημα σε συνεχή χρόνο.
2. Να υπολογιστεί η (p_n) .
3. Να βρεθεί το ποσοστό των χαμένων πελατών.
4. Να απαντηθούν τα ίδια ερωτήματα για το τροποποιημένο μοντέλο όπου μια αφικνούμενη ομάδα που δεν χωράει να εισέλθει ολόκληρη στο σύστημα δεν απορρίπτεται, αλλά εισέρχονται όσοι πελάτες της μπορούν να χωρέσουν στο σύστημα.

Άσκηση 7.5 Να υπολογιστεί ο μετασχηματισμός Laplace-Stieltjes του χρόνου παραμονής στο σύστημα ενός πελάτη για την $M/M/1$ ουρά με ομαδικές αφίξεις, με ρυθμό αφίξεων λ , ρυθμό εξυπηρέτησης μ και σταθερό μέγεθος αφικνούμενων ομάδων r .

7.6 Σχόλια

Όλα τα εισαγωγικά βιβλία της Θεωρίας Ουρών Αναμονής παρουσιάζουν τις βασικές επεκτάσεις της $M/M/1$ ουράς με ομαδικές αφίξεις ή ομαδικές εξυπηρετήσεις. Για περισσότερη λεπτομέρεια και παραδείγματα, μπορεί κανείς να δει τα κλασικά βιβλία Kleinrock 1975, Wolff 1989, Tijms 2008, Kulkarni 2010 και Shortle, Thompson, Gross και Harris 2018.

Για τη χρήση των γεννητριών μπορεί κανείς να μελετήσει ιδιαίτερα το βιβλίο Wilf 2005, ενώ για τη χρήση των πιθανογεννητριών σε Μαρκοβιανά μοντέλα μπορεί να ανατρέξει στη μεταπτυχιακή διπλωματική εργασία της Κάντα 2006.

Τα μοντέλα που λύνονται με πιθανογεννήτριες που ανάγονται σε ρητές συναρτήσεις μπορούν να μελετηθούν εναλλακτικά εφαρμόζοντας τη θεωρία γραμμικών εξισώσεων διαφορών με σταθερούς συντελεστές. Στο παρόν σύγγραμμα δεν παρουσιάζουμε αυτή την προσέγγιση, παρόλο που είναι εξίσου αποτελεσματική και στοιχειώδης με τις γεννήτριες. Ο ενδιαφερόμενος αναγνώστης μπορεί να δει την αντίστοιχη θεωρία και πολλά παραδείγματα εφαρμογής της στη Θεωρία Ουρών Αναμονής στο σύγγραμμα Adan και Wal 1998.

Βιβλιογραφία

- [1] L.V. Ahlfors. *Complex Analysis: An Introduction to the Theory of Analytic Functions of One Complex Variable, 3rd Edition*. New York: McGraw-Hill, Inc., 1979.
- [2] H.R. Gail, S.L. Hantler και B.A. Taylor. “Use of characteristics roots for solving infinite state Markov chains”. Στο: *Grassmann, W. (ed) Computational Probability*. Springer, 2000. Κεφ. 7, σσ. 205–255. ISBN: 978-0-7923-8617-9. DOI: https://doi.org/10.1007/978-1-4757-4828-4_7.
- [3] H.R. Gail, S.L. Hantler και B.A. Taylor. “Spectral analysis of M/G/1 and G/M/1 type Markov chains”. Στο: *Advances in Applied Probability* 28 (1996), σσ. 114–165.
- [4] L. Kleinrock. *Queueing Systems. Volume 1: Theory*. Wiley, 1975. ISBN: 978-0471491101.
- [5] R.W. Wolff. *Stochastic Modeling and the Theory of Queues*. Englewood Cliffs, New Jersey: Prentice Hall, 1989. ISBN: 978-0138466923.
- [6] H.C. Tijms. *A First Course in Stochastic Models, 2nd Edition*. Chichester: Wiley, 2008. ISBN: 978-0471498803.
- [7] V.G. Kulkarni. *Modeling and Analysis of Stochastic Systems, 2nd Edition*. CRC Press, 2010. ISBN: 978-1439808757.
- [8] J.F. Shurtle, J.M. Thompson, D. Gross και C.M. Harris. *Fundamentals of Queueing Theory, 5th Edition*. Wiley, 2018. ISBN: 978-1118943526.
- [9] H.S. Wilf. *Generatingfunctionology, 3rd Edition*. AK Peters/CRC, 2005. ISBN: 978-1568812793.
- [10] Σ. Κάντα. *Αλγοριθμικές Τεχνικές στις Μαρκοβιανές Αλυσίδες: Μέθοδοι Πιθανογεννητριών και Εφαρμογές*. Αθήνα: Διπλωματική Εργασία, Μεταπτυχιακό Πρόγραμμα Ειδίκευσης στη Στατιστική και Επιχειρησιακή Έρευνα, Τμήμα Μαθηματικών Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών, 2006.
- [11] I. Adan και J. van der Wal. *Difference and Differential Equations in Stochastic Operations Research*. Eindhoven: Notes available online, 1998.

ΚΕΦΑΛΑΙΟ 8

Η ΜΕΘΟΔΟΣ ΤΩΝ ΦΑΣΕΩΝ

Στο κεφάλαιο αυτό γίνεται μια εισαγωγή στη μέθοδο των φάσεων. Η μέθοδος αυτή χρησιμοποιείται για τη μοντελοποίηση στο πλαίσιο των Μαρκοβιανών αλυσίδων συνεχούς χρόνου κάποιων συστημάτων εξυπηρέτησης με μη-εκθετικούς ενδιάμεσους χρόνους αφίξεων και/ή μη-εκθετικούς χρόνους εξυπηρέτησης. Αρχικά δείχνουμε το πώς αυτή η μέθοδος λειτουργεί στην απλούστερη περίπτωση που οι χρόνοι αυτοί ακολουθούν την κατανομή Erlang. Κατόπιν, περιγράφεται η γενική περίπτωση σε αδρές γραμμές.

8.1 Η μέθοδος των φάσεων για κατανομές Erlang

Η κατανομή Erlang(k, α) ορίζεται ως η κατανομή του αθροίσματος k ανεξάρτητων $\text{Exp}(\alpha)$. Υπό αυτή την έννοια μια μη-αρνητική τυχαία μεταβλητή S (που μοντελοποιεί κάποιον ενδιάμεσο χρόνο αφίξεων ή χρόνο εξυπηρέτησης κ.τ.λ.) με την κατανομή Erlang(k, α) μπορεί να θεωρηθεί ότι αποτελείται από k εκθετικά κατανεμημένες «φάσεις»: $S = X_1 + X_2 + \dots + X_k$, με X_i να ακολουθούν την εκθετική κατανομή $\text{Exp}(\alpha)$, $i = 1, 2, \dots, k$. Βεβαίως, έχουμε $E[S] = \frac{k}{\alpha}$ και $\text{Var}[S] = \frac{k}{\alpha^2}$. Για μια συγκεκριμένη μέση τιμή $\frac{1}{\nu}$ υπάρχουν πολλές κατανομές Erlang που αντιστοιχούν σε αυτή. Συγκεκριμένα, είναι όλες οι Erlang($k, k\nu$). Πράγματι αν ο χρόνος S_k ακολουθεί την κατανομή Erlang($k, k\nu$), τότε $E[S_k] = \frac{k}{k\nu} = \frac{1}{\nu}$. Καθώς το k αυξάνει η διασπορά της S_k γίνεται μικρότερη, καθώς $\text{Var}[S_k] = \frac{k}{k^2\nu^2} = \frac{1}{k\nu^2}$. Έτσι μπορούμε να χρησιμοποιούμε κάποιαν S_k ώστε να μοντελοποιήσουμε χρόνους με μέση τιμή $\frac{1}{\nu}$ και μικρές διασπορές. Αυτό είναι σημαντικό σε συστήματα εξυπηρέτησης που είτε οι χρόνοι εξυπηρέτησης είτε οι ενδιάμεσοι χρόνοι αφίξεων έχουν μικρή διασπορά, πράγμα που συμβαίνει συχνά στις εφαρμογές.

Όταν έχω μια τέτοια κατάσταση όπου οι ενδιάμεσοι χρόνοι αφίξεων ή οι χρόνοι εξυπηρέτησης είναι Erlang, για να διατηρήσω τη Μαρκοβιανή ιδιότητα, θεωρώ ότι οι αντίστοιχοι χρόνοι αναλύονται σε φάσεις και κάθε χρονική στιγμή θα πρέπει να γνωρίζουμε σε ποια φάση βρίσκεται ο αντίστοιχος χρόνος. Π.χ., αν οι αφίξεις γίνονται με ρυθμό λ και οι ενδιάμεσοι χρόνοι αφίξεων είναι Erlang($k, k\lambda$), πρέπει να σκεφτόμαστε ότι κάθε φορά που φθάνει ένας πελάτης στο σύστημα, ένας άλλος ξεκινά να έρχεται και θα φθάσει στο σύστημα αφού περάσει k $\text{Exp}(k\lambda)$ φάσεις. Ομοίως, αν οι εξυπηρετήσεις γίνονται με ρυθμό μ και οι χρόνοι εξυπηρέτησης

Κατάσταση	Επόμενη κατάσταση	Χρόνος
(0, 1)	(0, 2)	$T_{(0,1),(0,2)} \sim \text{Exp}(2\lambda)$
(0, 2)	(1, 1)	$T_{(0,2),(1,1)} \sim \text{Exp}(2\lambda)$
$(n, 1), n \geq 1$	$(n, 2)$	$T_{(n,1),(n,2)} \sim \text{Exp}(2\lambda)$
	$(n-1, 1)$	$T_{(n,1),(n-1,1)} \sim \text{Exp}(\mu)$
$(n, 2), n \geq 1$	$(n+1, 1)$	$T_{(n,2),(n+1,1)} \sim \text{Exp}(2\lambda)$
	$(n-1, 2)$	$T_{(n,2),(n-1,2)} \sim \text{Exp}(\mu)$

Πίνακας 8.1: Πίνακας μεταβάσεων $E_2/M/1$ ουράς.

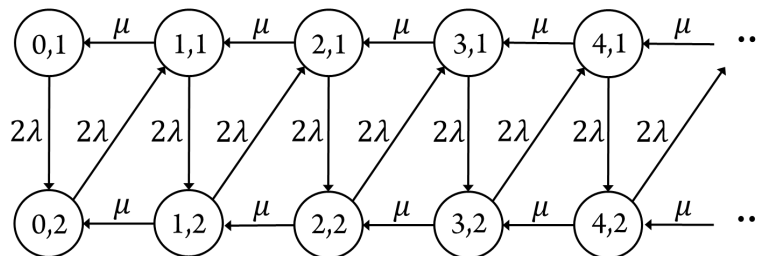
είναι $\text{Erlang}(s, s\mu)$, κάθε πελάτης περνάει από $s \text{Exp}(s\mu)$ φάσεις κατά την εξυπηρέτησή του.

Για να γίνει κατανοητή η μέθοδος, θα δούμε ορισμένα παραδείγματα στα οποία είτε οι ενδιάμεσοι χρόνοι των αφίξεων είτε οι χρόνοι εξυπηρέτησης είναι Erlang.

Παράδειγμα 8.1 (Η $E_2/M/1$ ουρά) Θεωρούμε την $E_2/M/1$ ουρά, όπου οι χρόνοι μεταξύ διαδοχικών αφίξεων είναι ανεξάρτητοι $\text{Erlang}(2, 2\lambda)$ (και επομένως ο ρυθμός των αφίξεων είναι λ), οι χρόνοι εξυπηρέτησης είναι $\text{Exp}(\mu)$, υπάρχει 1 υπηρέτης και απεριόριστη χωρητικότητα. Να μοντελοποιηθεί ως Μαρκοβιανή αλυσίδα συνεχούς χρόνου με τη μέθοδο των φάσεων και να γραφούν οι εξισώσεις ισορροπίας της.

Λύση: Έστω $Q(t)$ το πλήθος των πελατών στο σύστημα τη στιγμή t . Η $\{Q(t)\}$ δεν είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου αφού αν $Q(t) = 0$, ο χρόνος μέχρι την επόμενη μετάβαση που είναι αναγκαστικά προς την κατάσταση 1 είναι ο υπολειπόμενος χρόνος μιας $\text{Erlang}(2, 2\lambda)$ που δεν είναι εκθετικός.

Για να πάρουμε μια Μαρκοβιανή περιγραφή του συστήματος, θεωρούμε ότι κάθε πελάτης περνάει από δύο $\text{Exp}(2\lambda)$ φάσεις για να φθάσει στο σύστημα και θέτουμε $A(t)$ να είναι η φάση που βρίσκεται ο πελάτης σε διαδικασία άφιξης τη στιγμή t . Τότε η διαδικασία $\{(Q(t), A(t))\}$ έχει χώρο καταστάσεων $\{(n, i) : n \geq 0, i = 1, 2\}$ και οι μεταβάσεις δίνονται από τον πίνακα 8.1, ενώ το διάγραμμα μετάβασης παρατίθεται στο σχήμα 8.1.



Σχήμα 8.1: Διάγραμμα ρυθμών μετάβασης $E_2/M/1$ ουράς.

Οι εξισώσεις ισορροπίας είναι

$$\begin{aligned}
 2\lambda p_{(0,1)} &= \mu p_{(1,1)}, \\
 2\lambda p_{(0,2)} &= 2\lambda p_{(0,1)} + \mu p_{(1,2)}, \\
 (2\lambda + \mu)p_{(n,1)} &= 2\lambda p_{(n-1,2)} + \mu p_{(n+1,1)}, \quad n \geq 1, \\
 (2\lambda + \mu)p_{(n,2)} &= 2\lambda p_{(n,1)} + \mu p_{(n+1,2)}, \quad n \geq 1.
 \end{aligned}$$

◊

Παράδειγμα 8.2 (Η $E_k/E_s/1$ ουρά) Θεωρούμε την $E_k/E_s/1$ ουρά που είναι το σύστημα με Erlang- k ανεξάρτητους ενδιάμεσους χρόνους αφίξεων, με Erlang- s ανεξάρτητους εκθετικούς χρόνους εξυπηρέτησης, 1 υπηρέτη και απεριόριστη χωρητικότητα. Υποθέτουμε ότι ο ρυθμός αφίξεων είναι λ και ο ρυθμός εξυπηρέτησης μ . Να μοντελοποιηθεί ως Μαρκοβιανή αλυσίδα συνεχούς χρόνου με τη μέθοδο των φάσεων.

Κατάσταση	Επόμενη κατάσταση	Χρόνος
$(n, i, j), i < k, j < s$	$(n, i + 1, j)$	$T_{(n,i,j),(n,i+1,j)} \sim \text{Exp}(k\lambda)$
	$(n, i, j + 1)$	$T_{(n,i,j),(n,i,j+1)} \sim \text{Exp}(s\mu)$
$(n, k, j), j < s$	$(n + 1, 1, j)$	$T_{(n,k,j),(n+1,1,j)} \sim \text{Exp}(k\lambda)$
	$(n, k, j + 1)$	$T_{(n,k,j),(n,k,j+1)} \sim \text{Exp}(s\mu)$
$(n, i, s), i < k$	$(n, i + 1, s)$	$T_{(n,i,s),(n,i+1,s)} \sim \text{Exp}(k\lambda)$
	$(n - 1, i, 1)$	$T_{(n,i,s),(n-1,i,1)} \sim \text{Exp}(s\mu)$
(n, k, s)	$(n + 1, 1, s)$	$T_{(n,k,s),(n+1,1,s)} \sim \text{Exp}(k\lambda)$
	$(n - 1, k, 1)$	$T_{(n,k,s),(n-1,k,1)} \sim \text{Exp}(s\mu)$

Πίνακας 8.2: Πίνακας μεταβάσεων $E_k/E_s/1$ ουράς.

Κατάσταση	Επόμενη κατάσταση	Χρόνος
$n, 1 \leq n \leq k - 1$	$n + 1$	$T_{n,n+1} \sim \text{Exp}(\alpha)$
k	0	$T_{k,0} \sim \text{Exp}(\alpha)$

Πίνακας 8.3: Πίνακας μεταβάσεων Μαρκοβιανής αλυσίδας αναπαράστασης φάσεων για την κατανομή Erlang(k, α).

Λύση: Για να έχουμε ρυθμό αφίξεων λ και κατανομή ενδιάμεσων χρόνων Erlang- k , θα πρέπει οι ενδιάμεσοι χρόνοι των αφίξεων έχουν την κατανομή Erlang($k, k\lambda$). Ομοίως, θα πρέπει οι χρόνοι εξυπηρέτησης να έχουν την Erlang($s, s\mu$) κατανομή. Για να έχουμε μια Μαρκοβιανή περιγραφή του συστήματος θεωρούμε τη στοχαστική διαδικασία $\{(Q(t), A(t), S(t))\}$, όπου η $Q(t)$ καταγράφει τον αριθμό των πελατών τη στιγμή t , η $A(t)$ τη φάση στην οποία βρίσκεται ο πελάτης που είναι σε διαδικασία άφιξης της στιγμή t και η $S(t)$ τη φάση στην οποία βρίσκεται ο πελάτης που είναι σε διαδικασία εξυπηρέτησης τη στιγμή t . Βεβαίως, η $S(t)$ έχει νόημα μόνο όταν $Q(t) \geq 1$.

Ο χώρος καταστάσεων είναι

$$\mathcal{S} = \{(n, i, j) : n \geq 1, 1 \leq i \leq k, 1 \leq j \leq s\} \cup \{(0, i) : 1 \leq i \leq k\}.$$

Όντας σε μια κατάσταση (n, i, j) , $n \geq 2$, για τις μεταβάσεις έχουμε τον πίνακα 8.2. Ανάλογος είναι και ο πίνακας δυνατών μεταβάσεων από καταστάσεις με $n = 0, 1$.

◊

8.2 Η γενική μέθοδος των φάσεων

Όπως είδαμε στα προηγούμενα, η μέθοδος των φάσεων βασίζεται στην ιδέα ότι η κατανομή Erlang(k, α) μπορεί να εννοηθεί ως η κατανομή του χρόνου για να διέλθει από k καταστάσεις (φάσεις) ένας πελάτης, όπου ο χρόνος παραμονής σε κάθε κατάσταση έχει την $\text{Exp}(\alpha)$ κατανομή. Ισοδύναμα, ένας τέτοιος χρόνος μπορεί να θεωρηθεί ως ο χρόνος απορρόφησης στην κατάσταση 0 μιας Μαρκοβιανής αλυσίδας συνεχούς χρόνου με χώρο καταστάσεων το σύνολο $\{0, 1, 2, \dots, k\}$, που ξεκινάει από την κατάσταση 1 και έχει μεταβάσεις που δίνονται στον πίνακα 8.3.

Με άλλα λόγια, πρόκειται για τον χρόνο απορρόφησης στην κατάσταση 0 μιας Μαρκοβιανής αλυσίδας

συνεχούς χρόνου στο $\{0, 1, 2, \dots, k\}$ με αρχική κατανομή $p^{(0)} = (0, 1, 0, 0, \dots, 0)$ και πίνακα ρυθμών μετάβασης

$$Q = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & -\alpha & \alpha & \dots & 0 & 0 \\ 0 & 0 & -\alpha & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\alpha & \alpha \\ \alpha & 0 & 0 & \dots & 0 & -\alpha \end{pmatrix}.$$

Αυτή η ιδέα επιτρέπει την Μαρκοβιανή αναπαράσταση συστημάτων με Erlang ενδιάμεσους χρόνους αφίξεων και/ή εξυπηρέτησεων, εμφυτεύοντας ουσιαστικά την παραπάνω Μαρκοβιανή αλυσίδα που περιγράφει τους χρόνους αυτούς στη στοχαστική διαδικασία που περιγράφει την εξέλιξη του συστήματος ώστε να καταστεί και αυτή Μαρκοβιανή αλυσίδα συνεχούς χρόνου και να είναι εφαρμόσιμη η σχετική θεωρία. Η διαδικασία αυτή αναπτύχθηκε στην προηγούμενη παράγραφο και μπορεί να γενικευθεί ως εξής.

Ορισμός 8.3 (Κατανομή τύπου φάσεων) Θεωρούμε μια Μαρκοβιανή αλυσίδα συνεχούς χρόνου με χώρο καταστάσεων $\{0, 1, 2, \dots, k\}$ με αρχική κατανομή (τ_0, τ^T) και πίνακα ρυθμών μετάβασης

$$Q = \begin{pmatrix} 0 & \mathbf{0}^T \\ \mathbf{t} & \mathbf{T} \end{pmatrix}.$$

Η κατανομή του χρόνου X έως την απορρόφηση στην απορροφητική κατάσταση 0 λέγεται κατανομή τύπου φάσεων με την αναπαράσταση (τ^T, \mathbf{T}) (συντομογραφικά $PH(\tau^T, \mathbf{T})$). \diamond

Οπότε η $\text{Exp}(\alpha)$ είναι η απλούστερη κατανομή τύπου φάσεων με αναπαράσταση $PH(\tau^T, \mathbf{T})$ με $\tau = (1)$ και $\mathbf{T} = (-\alpha)$, ενώ μια $\text{Erlang}(k, \alpha)$ είναι κατανομή τύπου φάσεων με αναπαράσταση (τ^T, \mathbf{T}) όπου $\tau^T = (1, 0, 0, \dots, 0)$ και

$$\mathbf{T} = \begin{pmatrix} -\alpha & \alpha & \dots & 0 & 0 \\ 0 & -\alpha & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -\alpha & \alpha \\ 0 & 0 & \dots & 0 & -\alpha \end{pmatrix}.$$

Μια τυχαία μεταβλητή που είναι $\text{Exp}(\alpha_i)$ με πιθανότητα p_i για $i = 1, 2, \dots, k$ αναφέρεται ως υπερεκθετική με k εκδοχές και παραμέτρους τις πιθανότητες p_1, p_2, \dots, p_k και τους ρυθμούς $\alpha_1, \alpha_2, \dots, \alpha_k$. Για μια τέτοια κατανομή, χρησιμοποιείται η συντομογραφία $\text{Hyperexp}(p_1, p_2, \dots, p_k; \alpha_1, \alpha_2, \dots, \alpha_k)$. Κάθε τέτοια κατανομή έχει συντελεστή μεταβλητότητας μεγαλύτερο ή ίσο του 1. Για το λόγο αυτό η οικογένεια των υπερεκθετικών κατανομών χρησιμοποιείται συχνά για τη μοντελοποίηση χρόνων με συντελεστή μεταβλητότητας μεγαλύτερο του 1.

Η $\text{Hyperexp}(p_1, p_2, \dots, p_k; \alpha_1, \alpha_2, \dots, \alpha_k)$ είναι κατανομή τύπου φάσεων με αναπαράσταση (τ^T, \mathbf{T}) όπου $\tau^T = (p_1, p_2, p_3, \dots, p_k)$ και

$$\mathbf{T} = \begin{pmatrix} -\alpha_1 & 0 & \dots & 0 & 0 \\ 0 & -\alpha_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -\alpha_{k-1} & 0 \\ 0 & 0 & \dots & 0 & -\alpha_k \end{pmatrix}.$$

Δυο υποοικογένειες της οικογένειας των κατανομών τύπου φάσεων που είναι εξαιρετικά χρήσιμες είναι οι κατανομές Cox και οι μίξεις κατανομών Erlang με κοινό ρυθμό α .

Μια κατανομή Cox είναι ο χρόνος απορρόφησης στο 0 μιας Μαρκοβιανής αλυσίδας συνεχούς χρόνου με χώρο καταστάσεων το $\{0, 1, 2, \dots, k\}$, όπου οι χρόνοι παραμονής στην κατάσταση i είναι $\text{Exp}(\alpha_i)$ για κάθε $i = 1, 2, \dots, k$, η αλυσίδα ξεκινάει από την κατάσταση 1 και κινείται ως εξής: Έχοντας τελειώσει τον χρόνο

παραμονής στην κατάσταση i , η επόμενη κατάσταση είναι η $i + 1$ με πιθανότητα p_i ή 0 με πιθανότητα $1 - p_i$ (με $p_k = 0$). Μια τέτοια κατανομή είναι τύπου φάσεων με αναπαράσταση (τ^T, \mathbf{T}) όπου $\tau^T = (1, 0, 0, \dots, 0)$ και

$$\mathbf{T} = \begin{pmatrix} -\alpha_1 & p_1\alpha_1 & 0 & \cdots & 0 & 0 \\ 0 & -\alpha_2 & p_2\alpha_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & -\alpha_{k-1} & p_{k-1}\alpha_{k-1} \\ 0 & 0 & \cdots & 0 & 0 & -\alpha_k \end{pmatrix}.$$

Μια τέτοια κατανομή αναφέρεται ως Coxian($p_1, p_2, \dots, p_k; \alpha_1, \alpha_2, \dots, \alpha_k$).

Μια μίξη κατανομών Erlang με κοινό ρυθμό α και πιθανότητες-παραμέτρους $p_i, 1 \leq i \leq k$ είναι μια τυχαία μεταβλητή που είναι Erlang(i, α) με πιθανότητα p_i για $1 \leq i \leq k$. Μια τέτοια αναφέρεται με τη συντομογραφία MixedErlang($p_1, p_2, \dots, p_k; \alpha$) και είναι τύπου φάσεων με αναπαράσταση (τ^T, \mathbf{T}) όπου $\tau^T = (p_k, p_{k-1}, p_{k-2}, \dots, p_1)$ και

$$\mathbf{T} = \begin{pmatrix} -\alpha & \alpha & \cdots & 0 & 0 \\ 0 & -\alpha & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -\alpha & \alpha \\ 0 & 0 & \cdots & 0 & -\alpha \end{pmatrix}.$$

Η αξία των οικογενειών των κατανομών Cox, καθώς και των μίξεων κατανομών Erlang είναι ότι και οι δύο αυτές οικογένειες είναι πυκνές στην οικογένεια των κατανομών μη-αρνητικών τυχαίων μεταβλητών. Αυτό σημαίνει ότι για κάθε συνάρτηση κατανομής μη-αρνητικής τυχαίας μεταβλητής μπορεί να βρεθεί ακολουθία Cox κατανομών ή κατανομών μίξεων Erlang που να συγκλίνει κατά σημείο στη συγκεκριμένη συνάρτηση κατανομής για κάθε σημείο συνέχειάς της.

8.3 Προσαρμογή κατανομών σε δεδομένα

Στις εφαρμογές της Θεωρίας Ουρών Αναμονής στην αποτίμηση απόδοσης πραγματικών συστημάτων εξυπηρέτησης, το πρώτο βήμα είναι η επιλογή κατάλληλων κατανομών για την αναπαράσταση των χρόνων μεταξύ των αφίξεων και των χρόνων εξυπηρέτησεων. Για τον σκοπό αυτό συλλέγονται δεδομένα ή χρησιμοποιούνται δεδομένα που έχουν ήδη συλλεχθεί στο παρελθόν κατά τη λειτουργία άλλων συστημάτων. Κατόπιν επιλέγονται θεωρητικές κατανομές (των οποίων οι κατανομές δίνονται από μαθηματικούς τύπους) που προσαρμόζονται καλά στα υπάρχοντα δεδομένα.

Για την επιλογή των κατανομών, μια αρχή είναι η μεγιστοποίηση της εντροπίας κάτω από περιορισμούς που θέτει η γνώση των δεδομένων. Σύμφωνα με την αρχή αυτή, η κατανομή που μεγιστοποιεί την εντροπία εντός μιας κλάσης κατανομών είναι η πλέον αμερόληπτη μεταξύ των κατανομών της κλάσης. Έτσι, όπως έχουμε σημειώσει στην παράγραφο 3.3, η $\text{Exp}(\mu)$ είναι η κατανομή μέγιστης εντροπίας στην κλάση των κατανομών μη-αρνητικών τυχαίων μεταβλητών με μέση τιμή $\frac{1}{\mu}$. Με άλλα λόγια, η επιλογή της $\text{Exp}(\mu)$ για την μοντελοποίηση ενός χρόνου (άρα μη-αρνητικής τυχαίας μεταβλητής) για τον οποίο είναι γνωστή μόνο η μέση του τιμή $\frac{1}{\mu}$ είναι η πλέον αμερόληπτη επιλογή. Λεπτομέρειες για την αρχή της μέγιστης εντροπίας και το πώς αυτή χρησιμοποιείται για την επιλογή κατανομών για την αναπαράσταση χρόνων σε συστήματα εξυπηρέτησης, μπορεί να βρει κανείς στη βιβλιογραφία. Στα πλαίσια του παρόντος συγγράμματος δεν θα επεκταθούμε περισσότερο στην ιδέα αυτή.

Μια άλλη προσέγγιση είναι η προσαρμογή μιας απλής κατανομής τύπου φάσεων για την αναπαράσταση των ενδιάμεσων χρόνων άφιξης ή των χρόνων εξυπηρέτησης ενός συστήματος εξυπηρέτησης. Κατά την προσέγγιση αυτή επιλέγεται μια απλή κατανομή τύπου φάσεων που έχει την ίδια μέση τιμή $E[X]$ και την ίδια τυπική απόκλιση (άρα και συντελεστή μεταβλητότητας c_X) με τα αντίστοιχα δειγματικά μεγέθη, δηλαδή το δειγματικό μέσο και τη δειγματική τυπική απόκλιση που προκύπτουν από τα δεδομένα. Η διαδικασία αυτή περιγράφεται στην παράγραφο 2.5 του Adan και Resing 2001. Συνοπτικά έχει ως εξής:

Για να μοντελοποιήσουμε κατανομές που έχουν μέση τιμή $E[X]$ και συντελεστή μεταβλητότητας $c_X \in (0, 1)$ επιλέγουμε καταρχήν φυσικό αριθμό k τέτοιο ώστε

$$\frac{1}{k} \leq c_X^2 < \frac{1}{k-1}$$

και κατόπιν χρησιμοποιούμε την μίξη Erlang($k-1, \alpha$) και Erlang(k, α) με πιθανότητες p και $1-p$ όπου

$$p = \frac{1}{1+c_X^2} [kc_X^2 - \sqrt{k(1+c_X^2) - k^2c_X^2}], \quad \alpha = \frac{k-p}{E[X]}.$$

Για την περίπτωση κατανομών με μέση τιμή $E[X]$ και συντελεστή μεταβλητότητας $c_X \in [1, \infty)$, χρησιμοποιούμε μια Hyperexpro($p_1, p_2; \alpha_1, \alpha_2$). Επειδή μια κατανομή σε αυτή την κλάση δεν προσδιορίζεται μονοσήμαντα από τις δυο πρώτες ροπές της έχουμε μια σχετική ευελιξία στον καθορισμό των παραμέτρων. Συνήθως χρησιμοποιούμε την λεγόμενη υπερεκθετική κατανομή με ισορροπημένους μέσους, δηλαδή αυτήν που ικανοποιεί επιπλέον τη συνθήκη $\frac{p_1}{\alpha_1} = \frac{p_2}{\alpha_2}$. Στην περίπτωση αυτή οι παράμετροι που εξασφαλίζουν μέση τιμή $E[X]$ και συντελεστή μεταβλητότητας c_X είναι

$$p_1 = \frac{1}{2} \left(1 + \sqrt{\frac{c_X^2 - 1}{c_X^2 + 1}} \right), \quad p_2 = 1 - p_1,$$

και

$$\alpha_1 = \frac{2p_1}{E[X]}, \quad \alpha_2 = \frac{2p_2}{E[X]}.$$

Επίσης, μπορεί να επιτευχθεί προσαρμογή κατάλληλων κατανομών τύπου φάσεων όταν είναι γνωστές ροπές τρίτης ή ανώτερης τάξης, βλέπε π.χ. Johnson 1993 και Whitt 1986.

8.4 Ασκήσεις

Άσκηση 8.1 Βρείτε την κατανομή ισορροπίας του αριθμού των πελατών στην $M/E_s/1/1$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ .

Άσκηση 8.2 Βρείτε την κατανομή ισορροπίας του αριθμού των πελατών στην $M/E_2/2/2$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ .

Άσκηση 8.3 Βρείτε την κατανομή ισορροπίας του αριθμού των πελατών στην $M/E_s/1$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ .

Άσκηση 8.4 Βρείτε την κατανομή ισορροπίας του αριθμού των πελατών στην $E_k/M/1$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ .

Άσκηση 8.5 Κάνετε το διάγραμμα των ρυθμών μετάβασης της για μια Μαρκοβιανή αναπαράσταση της $E_2/E_2/1/1$ ουράς με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ .

8.5 Σχόλια

Η μέθοδος των φάσεων περιγράφεται σε όλα τα κλασικά εισαγωγικά βιβλία της Θεωρίας Ουρών Αναμονής, όπως στα Kleinrock 1975, Wolff 1989, Tijms 2008, Kulkarni 2010 και Shortle, Thompson, Gross και Harris 2018.

Το κλασικό σύγγραμμα Adan και Resing 2001 περιγράφει πώς μπορεί να προσεγγιστεί μια γενική κατανομή που μοντελοποιεί χρόνους (ενδιάμεσους χρόνους αφίξεων, χρόνους εξυπηρέτησης κλπ.), χρησιμοποιώντας κατανομές τύπου φάσεων με μικρό αριθμό φάσεων. Η προσέγγιση αναπτύσσεται χρησιμοποιώντας ιδέες από τις εργασίες των Johnson 1993 και Whitt 1986 και είναι ιδιαίτερα βολική από υπολογιστική σκοπιά.

Για μια πιο προχωρημένη εισαγωγή στη μέθοδο των φάσεων με πολλές θεωρητικές λεπτομέρειες και εφαρμογές, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στα βασικά συγγράμματα πινακο-αναλυτικών μεθόδων στη στοχαστική μοντελοποίηση των Neuts 1981, Neuts 1989 και Latouche και Ramaswami 1999.

Βιβλιογραφία

- [1] I. Adan και J. Resing. *Queueing Theory*. Eindhoven: Notes available online, 2001.
- [2] M.A. Johnson. “An empirical study of queueing approximations based on phase-type approximations”. Στο: *Stochastic Models* 9 (1993), σσ. 531–561.
- [3] W. Whitt. “Approximating a point process by a renewal process I: two basic methods”. Στο: *Operations Research* 30 (1986), σσ. 125–147.
- [4] L. Kleinrock. *Queueing Systems. Volume 1: Theory*. Wiley, 1975. ISBN: 978-0471491101.
- [5] R.W. Wolff. *Stochastic Modeling and the Theory of Queues*. Englewood Cliffs, New Jersey: Prentice Hall, 1989. ISBN: 978-0138466923.
- [6] H.C. Tijms. *A First Course in Stochastic Models, 2nd Edition*. Chichester: Wiley, 2008. ISBN: 978-0471498803.
- [7] V.G. Kulkarni. *Modeling and Analysis of Stochastic Systems, 2nd Edition*. CRC Press, 2010. ISBN: 978-1439808757.
- [8] J.F. Shortle, J.M. Thompson, D. Gross και C.M. Harris. *Fundamentals of Queueing Theory, 5th Edition*. Wiley, 2018. ISBN: 978-1118943526.
- [9] M.F. Neuts. *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*. The Johns Hopkins University Press, 1981. ISBN: 978-0801825606.
- [10] M.F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. CRC Press, 1989. ISBN: 978-0824782832.
- [11] G. Latouche και V. Ramaswami. *Introduction to Matrix Analytic Models in Stochastic Modeling*. ASA-SIAM Series on Statistics και Applied Probability, 1999. ISBN: 978-0898714258.

ΔΙΔΙΑΣΤΑΤΕΣ ΜΑΡΚΟΒΙΑΝΕΣ ΟΥΡΕΣ

Στο κεφάλαιο αυτό εξετάζουμε μοντέλα στα οποία η στοχαστική διαδικασία του αριθμού των πελατών στο σύστημα, $\{Q(t)\}$, δεν είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου, αλλά μπορεί να εμφυτευτεί σε μια Μαρκοβιανή αλυσίδα συνεχούς χρόνου, υπό την έννοια ότι υπάρχει μια στοχαστική διαδικασία $\{I(t)\}$ που καταγράφει κάποια επιπλέον πληροφορία για την κατάσταση του συστήματος ώστε η $\{(Q(t), I(t))\}$ να είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου. Στο προηγούμενο κεφάλαιο είδαμε ουσιαστικά μια πρώτη εφαρμογή αυτής της ιδέας, όπου η $\{I(t)\}$ αναφερόταν σε κάποια επιπλέον πληροφορία (φάση) για τη διαδικασία των αφίξεων και/ή των εξυπηρετήσεων. Στο παρόν κεφάλαιο θα δούμε κάποια άλλα παραδείγματα, όπου η $\{I(t)\}$ αναφέρεται στην κατάσταση του υπηρέτη. Θα δούμε διάφορα παραδείγματα μοντέλων που εμπίπτουν σε αυτή την κατηγορία, καθώς και την επίλυσή τους με τη μέθοδο των πιθανογεννητριών.

Η μέθοδος των πιθανογεννητριών, στην περίπτωση αυτή, διαμορφώνεται ως εξής: Για κάθε τιμή i της $\{I(t)\}$ ορίζεται μια «μερική» πιθανογεννήτρια της κατανομής ισορροπίας της $\{Q(t)\}$ ως

$$P_i(z) = \sum_{n=0}^{\infty} p_{(n,i)} z^n, \quad |z| \leq 1, \quad i \in \mathcal{S}_I,$$

όπου $(p_{(n,i)})$ είναι η κατανομή ισορροπίας της $\{(Q(t), I(t))\}$ και \mathcal{S}_I ο χώρος καταστάσεων της $\{I(t)\}$. Οι εξισώσεις ισορροπίας για την $(p_{(n,i)})$ μετασχηματίζονται σε σύστημα εξισώσεων για τις άγνωστες πιθανογεννήτριες $P_i(z)$, $i \in \mathcal{S}_I$. Επιλύοντας το σύστημα αυτών των εξισώσεων, προσδιορίζουμε τις $P_i(z)$ και κατόπιν την κατανομή ισορροπίας, αναπτύσσοντας τις μερικές πιθανογεννήτριες σε δυνάμεις του z^n . Για τον μετασχηματισμό των εξισώσεων ισορροπίας, πολλαπλασιάζουμε τις εξισώσεις ισορροπίας που αναφέρονται στις καταστάσεις (n, i) για ένα συγκεκριμένο i με τα αντίστοιχα z^n και κατόπιν αθροίζουμε για όλα τα n . Η μέθοδος θα αναπτυχθεί στο συγκεκριμένο πλαίσιο τριών σημαντικών συστημάτων εξυπηρέτησης: Της M/M/1 ουράς με ενεργές και ανενεργές περιόδους εξυπηρέτησεων, της M/M/1 ουράς με χρόνους επανεκκίνησης και της M/M/1/1 ουράς με επαναπροσπάθειες. Τα δυο πρώτα μοντέλα οδηγούν σε αλγεβρικές εξισώσεις για τις μερικές πιθανογεννήτριες, ενώ το τρίτο σε διαφορικές εξισώσεις, οπότε θα γίνει σαφές πώς διαφοροποιείται η μέθοδος στις δύο περιπτώσεις.

Κατάσταση	Επόμενη κατάσταση	Χρόνος
$(n, 0), n \geq 0$	$(n + 1, 0)$	$T_{(n,0),(n+1,0)} \sim \text{Exp}(\lambda)$
	$(n, 1)$	$T_{(n,0),(n,1)} \sim \text{Exp}(\theta)$
$(0, 1)$	$(1, 1)$	$T_{(0,1),(1,1)} \sim \text{Exp}(\lambda)$
	$(0, 0)$	$T_{(0,1),(0,0)} \sim \text{Exp}(\zeta)$
$(n, 1), n \geq 1$	$(n + 1, 1)$	$T_{(n,1),(n+1,1)} \sim \text{Exp}(\lambda)$
	$(n - 1, 1)$	$T_{(n,1),(n-1,1)} \sim \text{Exp}(\mu)$
	$(n, 0)$	$T_{(n,1),(n,0)} \sim \text{Exp}(\zeta)$

Πίνακας 9.1: Πίνακας μεταβάσεων M/M/1 ουράς με ενεργές και ανενεργές περιόδους εξυπηρέτησεων.

9.1 Η M/M/1 ουρά με ενεργές και ανενεργές περιόδους εξυπηρέτησεων

Θεωρούμε την M/M/1 ουρά με ρυθμό αφίξεων λ , ρυθμό εξυπηρέτησεων μ , 1 υπηρέτη και απεριόριστη χωρητικότητα υπό την πειθαρχία ουράς FCFS. Ο υπηρέτης εναλλάσσεται μεταξύ ενεργών και ανενεργών περιόδων εξυπηρέτησεων. Κατά τη διάρκεια των ενεργών περιόδων παρέχει εξυπηρέτηση, ενώ κατά τη διάρκεια των ανενεργών περιόδων παύει να παρέχει εξυπηρέτηση, αλλά το σύστημα συνεχίζει να δέχεται νέες αφίξεις πελατών. Οι διάρκειες των ενεργών και των ανενεργών περιόδων είναι όλες ανεξάρτητες μεταξύ τους. Επιπλέον υποθέτουμε ότι οι διάρκειες των ενεργών περιόδων έχουν την κατανομή $\text{Exp}(\zeta)$, ενώ οι διάρκειες των ανενεργών περιόδων έχουν την κατανομή $\text{Exp}(\theta)$. Το σύστημα αυτό αναφέρεται ως η M/M/1 ουρά με ενεργές και ανενεργές περιόδους εξυπηρέτησεων.

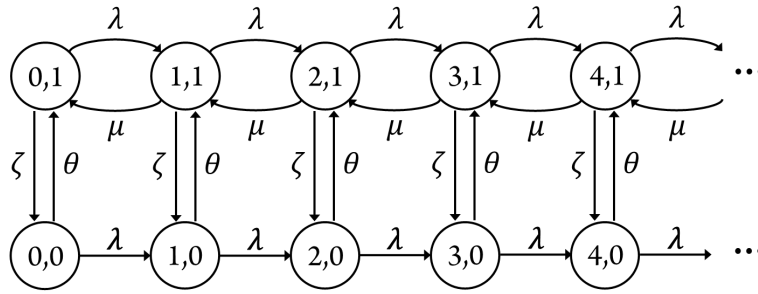
Το σύστημα αυτό είναι πολύ χρήσιμο στις εφαρμογές. Καταρχήν, μπορεί να αναπαριστά την περίπτωση που ο υπηρέτης υπόκειται σε βλάβες. Πιο συγκεκριμένα, αν συμβαίνουν βλάβες λόγω τυχαίων παραγόντων σύμφωνα με διαδικασία Poisson με ρυθμό ζ όποτε ο υπηρέτης λειτουργεί και οι χρόνοι επισκευής ακολουθούν την κατανομή $\text{Exp}(\theta)$, τότε έχουμε μια M/M/1 ουρά με ενεργές και ανενεργές περιόδους εξυπηρέτησεων. Επιπλέον, το μοντέλο μπορεί να περιγράψει καταστάσεις στις οποίες ο υπηρέτης εξυπηρετεί κυκλικά δυο συστήματα εξυπηρέτησης (polling system). Συγκεκριμένα, ας υποθέσουμε ότι οι χρόνοι παραμονής του στο πρωτεύον σύστημα ακολουθούν την κατανομή $\text{Exp}(\zeta)$, στο δευτερεύον σύστημα ακολουθούν την κατανομή $\text{Exp}(\theta)$ και ισχύουν οι προϋποθέσεις ανεξαρτησίας που τέθηκαν παραπάνω. Στην περίπτωση αυτή, αν επικεντρωθούμε στο πρωτεύον σύστημα, τότε έχουμε μια M/M/1 ουρά με ενεργές και ανενεργές περιόδους εξυπηρέτησεων, όπου οι ενεργές περίοδοι αναφέρονται στην παρουσία του υπηρέτη στο πρωτεύον σύστημα και οι ανενεργές περίοδοι στην παρουσία του στο δευτερεύον.

Συμβολίζοντας με $Q(t)$ το πλήθος των πελατών στο σύστημα τη στιγμή t , είναι φανερό ότι η $\{Q(t)\}$ δεν είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου. Πράγματι, δοθέντος ότι $Q(t) = n$, ο χρόνος μέχρι την επόμενη αναχώρηση δεν είναι εκθετικός, αφού η τρέχουσα εξυπηρέτηση υπάρχει περίπτωση να διακοπεί μία ή περισσότερες φορές και να παρεμβληθεί αντίστοιχος αριθμός ανενεργών περιόδων. Επομένως, για να επιτευχθεί μια Μαρκοβιανή περιγραφή του συστήματος, χρειάζεται επίσης να καταγράφεται σε κάθε χρονική στιγμή t η κατάσταση του υπηρέτη $I(t)$ ($I(t) = 1$ αν ο υπηρέτης είναι ενεργός, ενώ $I(t) = 0$ αν είναι ανενεργός). Για τις μεταβάσεις της $\{(Q(t), I(t))\}$ έχουμε τότε τον πίνακα 9.1.

Αφού όλοι οι χρόνοι είναι εκθετικοί, έχουμε ότι η $\{(Q(t), I(t))\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου με διάγραμμα ρυθμών μετάβασης που δίνεται στο σχήμα 9.1.

Έστω $(p_{(n,i)})$ η κατανομή ισορροπίας της. Οι εξισώσεις ισορροπίας της είναι

$$\begin{aligned}
 (\lambda + \theta)p_{(0,0)} &= \zeta p_{(0,1)}, \\
 (\lambda + \theta)p_{(n,0)} &= \lambda p_{(n-1,0)} + \zeta p_{(n,1)}, \quad n \geq 1, \\
 (\lambda + \zeta)p_{(0,1)} &= \mu p_{(1,1)} + \theta p_{(0,0)}, \\
 (\lambda + \mu + \zeta)p_{(n,1)} &= \lambda p_{(n-1,1)} + \mu p_{(n+1,1)} + \theta p_{(n,0)}, \quad n \geq 1.
 \end{aligned}$$



Σχήμα 9.1: Διάγραμμα ρυθμών μετάβασης M/M/1 ουράς με ενεργές και ανενεργές περιόδους εξυπηρέτησεων.

Για να προσδιορίσουμε τη στάσιμη κατανομή, εισάγουμε τις μερικές πιθανογεννήτριες

$$P_0(z) = \sum_{n=0}^{\infty} p_{(n,0)} z^n, \quad |z| \leq 1,$$

$$P_1(z) = \sum_{n=0}^{\infty} p_{(n,1)} z^n, \quad |z| \leq 1.$$

Πολλαπλασιάζοντας κάθε εξίσωση ισορροπίας που αναφέρεται σε κατάσταση $(n, 0)$ με το αντίστοιχο z^n και αθροίζοντας για όλα τα $n \geq 0$ παίρνουμε:

$$\begin{aligned} (\lambda + \theta) \sum_{n=0}^{\infty} p_{(n,0)} z^n &= \lambda \sum_{n=1}^{\infty} p_{(n-1,0)} z^n + \zeta \sum_{n=0}^{\infty} p_{(n,1)} z^n \\ \Leftrightarrow (\lambda + \theta) P_0(z) &= \lambda z P_0(z) + \zeta P_1(z). \end{aligned}$$

Λύνοντας ως προς $P_1(z)$ παίρνουμε

$$P_1(z) = \frac{\lambda + \theta - \lambda z}{\zeta} P_0(z). \quad (9.1)$$

Ομοίως, πολλαπλασιάζοντας κάθε εξίσωση ισορροπίας που αναφέρεται σε κατάσταση $(n, 1)$ με το αντίστοιχο z^n και αθροίζοντας για όλα τα $n \geq 0$ παίρνουμε:

$$\begin{aligned} (\lambda + \mu + \zeta) \sum_{n=0}^{\infty} p_{(n,1)} z^n - \mu p_{(0,1)} &= \lambda \sum_{n=1}^{\infty} p_{(n-1,1)} z^n + \mu \sum_{n=0}^{\infty} p_{(n+1,1)} z^n + \theta \sum_{n=0}^{\infty} p_{(n,0)} z^n \\ \Leftrightarrow (\lambda + \mu + \zeta) P_1(z) - \mu p_{(0,1)} &= \lambda z P_1(z) + \frac{\mu}{z} (P_1(z) - p_{(0,1)}) + \theta P_0(z). \end{aligned} \quad (9.2)$$

Αντικαθιστώντας την $P_1(z)$, από τη σχέση (9.1), στην (9.2), πολλαπλασιάζοντας με ζz και απλοποιώντας το $(z - 1)$ παίρνουμε

$$\begin{aligned} (\lambda + \mu + \zeta)(\lambda + \theta - \lambda z) z P_0(z) - \mu \zeta p_{(0,1)} &= \lambda(\lambda + \theta - \lambda z) z^2 P_0(z) + \mu(\lambda + \theta - \lambda z) P_0(z) - \mu \zeta p_{(0,1)} + \theta \zeta z P_0(z). \end{aligned}$$

Λύνοντας ως προς $P_0(z)$ παίρνουμε

$$P_0(z) = \frac{\mu \zeta p_{(0,1)}}{\lambda^2 z^2 - \lambda(\lambda + \mu + \theta + \zeta)z + \mu(\lambda + \theta)}. \quad (9.3)$$

Για να βρούμε τώρα την κατανομή ισορροπίας πρέπει να αναπτύξουμε την $P_0(z)$ σε δυνάμεις του z^n . Για το σκοπό αυτό πρέπει να παραγοντοποιήσουμε τον παρονομαστή της $P_0(z)$ σε πρωτοβάθμιους όρους ως $c(1 - \rho_1 z)(1 - \rho_2 z)$, κατόπιν να αναλύσουμε την $P_0(z)$ σε απλά κλάσματα και στο τέλος να χρησιμοποιήσουμε το ανάπτυγμα της γεωμετρικής σειράς για την ανάπτυξη της πιθανογεννήτριας σε δυνάμεις του z^n (βλέπε και την παράγραφο 3.2 για περισσότερες λεπτομέρειες). Η διαδικασία έχει αναλυτικά ως εξής:

Έστω $D(z) = \lambda^2 z^2 - \lambda(\lambda + \mu + \theta + \zeta)z + \mu(\lambda + \theta)$ ο παρονομαστής της $P_0(z)$. Τότε $D(z) = z^2 \phi(z^{-1})$, όπου $\phi(z) = \lambda^2 - \lambda(\lambda + \mu + \theta + \zeta)z + \mu(\lambda + \theta)z^2$. Αν ρ_1 και ρ_2 είναι οι ρίζες της $\phi(z)$,

$$\rho_{1,2} = \frac{\lambda(\lambda + \mu + \theta + \zeta) \pm \sqrt{\lambda^2(\lambda + \mu + \theta + \zeta)^2 - 4\lambda^2\mu(\lambda + \theta)}}{2\mu(\lambda + \theta)}, \quad (9.4)$$

τότε

$$D(z) = \mu(\lambda + \theta)z^2 \left(\frac{1}{z} - \rho_1\right) \left(\frac{1}{z} - \rho_2\right) = \mu(\lambda + \theta)(1 - \rho_1 z)(1 - \rho_2 z).$$

Επομένως, η (9.3) γίνεται

$$P_0(z) = \frac{\zeta p_{(0,1)}}{(\lambda + \theta)(1 - \rho_1 z)(1 - \rho_2 z)} = \frac{c_1}{1 - \rho_1 z} + \frac{c_2}{1 - \rho_2 z}, \quad (9.5)$$

όπου c_1, c_2 σταθερές που θα προσδιοριστούν παρακάτω. Επίσης, αντικαθιστώντας την (9.5) στην (9.1) παίρνουμε

$$P_1(z) = \frac{\lambda + \theta - \lambda z}{\zeta} \left(\frac{c_1}{1 - \rho_1 z} + \frac{c_2}{1 - \rho_2 z} \right). \quad (9.6)$$

Αναπτύσσοντας τις $P_0(z)$ και $P_1(z)$ σε δυνάμεις του z^n , χρησιμοποιώντας τη γεωμετρική σειρά (3.5), έχουμε

$$\begin{aligned} P_0(z) &= \sum_{n=0}^{\infty} (c_1 \rho_1^n + c_2 \rho_2^n) z^n, \\ P_1(z) &= \frac{\lambda + \theta}{\zeta} \sum_{k=0}^{\infty} (c_1 \rho_1^k + c_2 \rho_2^k) z^k - \frac{\lambda}{\zeta} \sum_{k=0}^{\infty} (c_1 \rho_1^k + c_2 \rho_2^k) z^{k+1} \\ &= \frac{\lambda + \theta}{\zeta} \sum_{n=0}^{\infty} (c_1 \rho_1^n + c_2 \rho_2^n) z^n - \frac{\lambda}{\zeta} \sum_{n=1}^{\infty} (c_1 \rho_1^{n-1} + c_2 \rho_2^{n-1}) z^n, \end{aligned}$$

απ' όπου συμπεραίνουμε ότι

$$\begin{aligned} p_{(n,0)} &= c_1 \rho_1^n + c_2 \rho_2^n, \quad n \geq 0, \\ p_{(0,1)} &= \frac{\lambda + \theta}{\zeta} (c_1 + c_2), \\ p_{(n,1)} &= c_1 v_1 \rho_1^n + c_2 v_2 \rho_2^n, \quad n \geq 1, \end{aligned}$$

όπου

$$v_i = \frac{(\lambda + \theta)\rho_i - \lambda}{\zeta\rho_i}, \quad i = 1, 2.$$

Το μόνο που απομένει είναι ο υπολογισμός των σταθερών c_1, c_2 . Από την εξίσωση (9.5), κάνοντας ομώνυμα τα κλάσματα του δεξιού μέλους, και κατόπιν εξισώνοντας την αριστερή και τη δεξιά έκφραση παίρνουμε ότι

$$c_1 \rho_2 + c_2 \rho_1 = 0. \quad (9.7)$$

Κατάσταση	Επόμενη κατάσταση	Χρόνος
(0, 0)	(1, 0)	$T_{(0,0),(1,0)} \sim \text{Exp}(\lambda)$
(n, 0), $n \geq 1$	(n + 1, 0)	$T_{(n,0),(n+1,0)} \sim \text{Exp}(\lambda)$
	(n, 1)	$T_{(n,0),(n,1)} \sim \text{Exp}(\theta)$
(1, 1)	(2, 1)	$T_{(1,1),(2,1)} \sim \text{Exp}(\lambda)$
	(0, 0)	$T_{(1,1),(0,0)} \sim \text{Exp}(\mu)$
(n, 1), $n \geq 2$	(n + 1, 1)	$T_{(n,1),(n+1,1)} \sim \text{Exp}(\lambda)$
	(n - 1, 1)	$T_{(n,1),(n-1,1)} \sim \text{Exp}(\mu)$

Πίνακας 9.2: Πίνακας μεταβάσεων M/M/1 ουράς με χρόνους επανεκκίνησης.

Επίσης, από την εξίσωση κανονικοποίησης έχουμε ότι $P_0(1) + P_1(1) = 1$ που δίνει

$$\left(1 + \frac{\theta}{\zeta}\right) \left(\frac{c_1}{1 - \rho_1} + \frac{c_2}{1 - \rho_2}\right) = 1. \quad (9.8)$$

Λύνοντας το σύστημα των εξισώσεων (9.7)-(9.8), παίρνουμε:

$$c_1 = C\rho_1, \quad c_2 = -C\rho_2,$$

με

$$C = \left[\frac{\rho_1(1 + v_1)}{1 - \rho_1} - \frac{\rho_2(1 + v_2)}{1 - \rho_2} \right]^{-1} = \frac{\zeta(1 - \rho_1)(1 - \rho_2)}{(\theta + \zeta)(\rho_1 - \rho_2)}.$$

9.2 Η M/M/1 ουρά με χρόνους επανεκκίνησης

Θεωρούμε την M/M/1 ουρά με ρυθμό αφίξεων λ , ρυθμό εξυπηρέτησεων μ , 1 υπηρέτη και απεριόριστη χωρητικότητα υπό την πειθαρχία ουράς FCFS. Λόγω υψηλών λειτουργικών εξόδων, υποθέτουμε ότι ο υπηρέτης απενεργοποιείται ακαριαία μόλις το σύστημα αδειάσει. Όταν ένας νέος πελάτης αφιχθεί, ο υπηρέτης ξεκινά έναν χρόνο επανεκκίνησης ή προθέρμανσης (setup time) με το πέρας του οποίου ο υπηρέτης αρχίζει να παρέχει εξυπηρέτηση μέχρι το σύστημα να μείνει ξανά κενό. Στη διάρκεια ενός χρόνου επανεκκίνησης το σύστημα δέχεται νέες αφίξεις, αλλά δεν παρέχει εξυπηρέτηση. Οι χρόνοι επανεκκίνησης του υπηρέτη είναι $\text{Exp}(\theta)$. Το σύστημα αυτό αναφέρεται ως η M/M/1 ουρά με χρόνους επανεκκίνησης (όταν το σύστημα μείνει κενό). Έστω $Q(t)$ ο αριθμός των πελατών τη στιγμή t στο σύστημα. Είναι φανερό ότι η $\{Q(t)\}$ δεν είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου, αφού όταν $Q(t) = n \geq 1$, ο χρόνος μέχρι την επόμενη αναχώρηση μπορεί να είναι είτε $\text{Exp}(\mu)$, είτε το άθροισμα δυο ανεξάρτητων $\text{Exp}(\theta)$ και $\text{Exp}(\mu)$, ανάλογα με το αν ο υπηρέτης είναι ενεργός ή όχι, αντίστοιχα. Επομένως, για μια Μαρκοβιανή περιγραφή του συστήματος χρειάζεται επίσης να καταγράφουμε σε κάθε χρονική στιγμή t την κατάσταση του υπηρέτη $I(t)$ ($I(t) = 1$ αν ο υπηρέτης είναι ενεργός, ενώ $I(t) = 0$ αν είναι ανενεργός). Για τις μεταβάσεις της $\{(Q(t), I(t))\}$ έχουμε τότε τον πίνακα 9.2.

Αφού όλοι οι χρόνοι είναι εκθετικοί, έχουμε ότι η $\{(Q(t), I(t))\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου. Το διάγραμμα ρυθμών μετάβασής της δίνεται στο σχήμα 9.2.

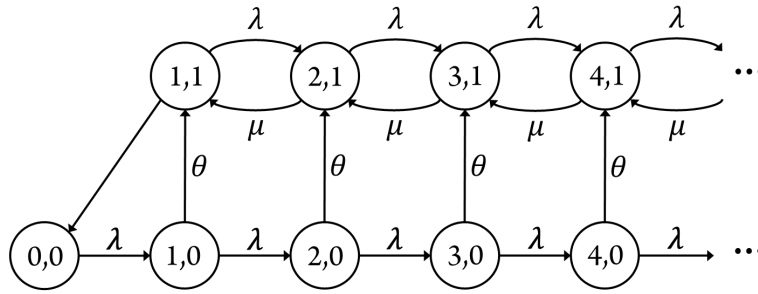
Έστω $(p_{(n,i)})$ η κατανομή ισορροπίας της. Οι εξισώσεις ισορροπίας της είναι

$$\lambda p_{(0,0)} = \mu p_{(1,1)}, \quad (9.9)$$

$$(\lambda + \theta)p_{(n,0)} = \lambda p_{(n-1,0)}, \quad n \geq 1, \quad (9.10)$$

$$(\lambda + \mu)p_{(1,1)} = \theta p_{(1,0)} + \mu p_{(2,1)}, \quad (9.11)$$

$$(\lambda + \mu)p_{(n,1)} = \theta p_{(n,0)} + \mu p_{(n+1,1)} + \lambda p_{(n-1,1)}, \quad n \geq 2. \quad (9.12)$$



Σχήμα 9.2: Διάγραμμα ρυθμών μετάβασης M/M/1 ουράς με χρόνους επανεκκίνησης.

Για να προσδιορίσουμε την κατανομή ισορροπίας της, εισάγουμε τις μερικές πιθανογεννήτριες

$$P_0(z) = \sum_{n=0}^{\infty} p_{(n,0)} z^n, \quad |z| \leq 1,$$

$$P_1(z) = \sum_{n=1}^{\infty} p_{(n,1)} z^n, \quad |z| \leq 1.$$

Πολλαπλασιάζοντας κάθε εξίσωση ισορροπίας που αναφέρεται σε κατάσταση $(n, 0)$ με το αντίστοιχο z^n (βλέπε εξισώσεις (9.9), (9.10)) και αθροίζοντας για όλα τα $n \geq 0$, παίρνουμε

$$\lambda p_{(0,0)} + (\lambda + \theta) \sum_{n=1}^{\infty} p_{(n,0)} z^n = \mu p_{(1,1)} + \lambda \sum_{n=1}^{\infty} p_{(n-1,0)} z^n$$

$$\Leftrightarrow (\lambda + \theta)P_0(z) - \theta p_{(0,0)} = \mu p_{(1,1)} + \lambda z P_0(z)$$

$$\Leftrightarrow (\lambda + \theta)P_0(z) - \theta p_{(0,0)} = \lambda p_{(0,0)} + \lambda z P_0(z).$$

Λύνοντας για την $P_0(z)$ παίρνουμε

$$P_0(z) = \frac{(\lambda + \theta)p_{(0,0)}}{\lambda + \theta - \lambda z}. \tag{9.13}$$

Ομοίως, πολλαπλασιάζοντας κάθε εξίσωση ισορροπίας που αναφέρεται σε κατάσταση $(n, 1)$ με το αντίστοιχο z^n (βλέπε εξισώσεις (9.11) και (9.12)) και αθροίζοντας για όλα τα $n \geq 1$, παίρνουμε

$$(\lambda + \mu) \sum_{n=1}^{\infty} p_{(n,1)} z^n = \theta \sum_{n=1}^{\infty} p_{(n,0)} z^n + \mu \sum_{n=1}^{\infty} p_{(n+1,1)} z^n + \lambda \sum_{n=2}^{\infty} p_{(n-1,1)} z^n$$

$$\Leftrightarrow (\lambda + \mu)P_1(z) = \theta(P_0(z) - p_{(0,0)}) + \frac{\mu}{z}(P_1(z) - p_{(1,1)}z) + \lambda z P_1(z)$$

$$\Leftrightarrow \left(\lambda + \mu - \lambda z - \frac{\mu}{z}\right)P_1(z) = \theta P_0(z) - \theta p_{(0,0)} - \mu p_{(1,1)}.$$

Αντικαθιστώντας το $\mu p_{(1,1)}$ από το $\lambda p_{(0,0)}$, χρησιμοποιώντας την εξίσωση ισορροπίας (9.9) για την $(0, 0)$ και πολλαπλασιάζοντας με z έχουμε

$$(\lambda z + \mu z - \lambda z^2 - \mu)P_1(z) = \theta z P_0(z) - (\lambda + \theta)z p_{(0,0)}.$$

Παραγοντοποιώντας το συντελεστή του $P_1(z)$, αντικαθιστώντας την $P_0(z)$ που δίνεται από την (9.13) και λύνοντας ως προς $P_1(z)$ παίρνουμε

$$P_1(z) = \frac{\lambda(\lambda + \theta)z p_{(0,0)}}{(\lambda + \theta - \lambda z)(\mu - \lambda z)}. \tag{9.14}$$

Το $p_{(0,0)}$ βρίσκεται από την εξίσωση κανονικοποίησης $P_0(1) + P_1(1) = 1$ και είναι

$$p_{(0,0)} = \frac{\theta(\mu - \lambda)}{(\lambda + \theta)\mu}.$$

Αντικαθιστώντας το στους τύπους (9.13) και (9.14) για τις $P_0(z)$ και $P_1(z)$ παίρνουμε

$$P_0(z) = \frac{\theta(\mu - \lambda)}{\mu(\lambda + \theta - \lambda z)}, \quad (9.15)$$

$$P_1(z) = \frac{\lambda\theta(\mu - \lambda)z}{\mu(\lambda + \theta - \lambda z)(\mu - \lambda z)}. \quad (9.16)$$

Για να βρούμε τις πιθανότητες ισορροπίας $p_{(n,0)}$, $n \geq 0$, χρησιμοποιούμε τη γεωμετρική σειρά (3.5) και έχουμε

$$P_0(z) = \frac{\theta(\mu - \lambda)}{\mu(\lambda + \theta)} \cdot \frac{1}{1 - \frac{\lambda}{\lambda + \theta}z} = \sum_{n=0}^{\infty} \frac{\theta(\mu - \lambda)}{\mu(\lambda + \theta)} \left(\frac{\lambda}{\lambda + \theta} \right)^n z^n,$$

απ' όπου και συμπεραίνουμε ότι

$$p_{(n,0)} = \frac{\theta(\mu - \lambda)}{\mu(\lambda + \theta)} \left(\frac{\lambda}{\lambda + \theta} \right)^n, \quad n \geq 0.$$

Για να βρούμε τις πιθανότητες ισορροπίας, $p_{(n,1)}$, $n \geq 1$ διακρίνουμε δυο περιπτώσεις:

Περίπτωση 1: $\lambda + \theta \neq \mu$.

Η μερική πιθανογεννήτρια $P_1(z)$ είναι ρητή, οπότε για την αντιστροφή της ακολουθούμε τη διαδικασία που περιγράφηκε στην παράγραφο 3.2. Η $P_1(z)$ αναλύεται σε απλά κλάσματα ως

$$P_1(z) = \frac{A}{\lambda + \theta - \lambda z} + \frac{B}{\mu - \lambda z}.$$

Οι σταθερές A και B βρίσκονται κατά τα γνωστά, είτε κάνοντας τα κλάσματα ομώνυμα και εξισώνοντας τους συντελεστές του z στον αριθμητή, είτε άμεσα ως εξής:

$$A = \lim_{z \rightarrow \frac{\lambda + \theta}{\lambda}} (\lambda + \theta - \lambda z)P_1(z) = \frac{\theta(\mu - \lambda)(\lambda + \theta)}{\mu(\mu - \lambda - \theta)},$$

$$B = \lim_{z \rightarrow \frac{\mu}{\lambda}} (\mu - \lambda z)P_1(z) = \frac{\theta(\mu - \lambda)}{\lambda + \theta - \mu}.$$

Επομένως έχουμε ότι

$$P_1(z) = \frac{A}{\lambda + \theta} \cdot \frac{1}{1 - \frac{\lambda}{\lambda + \theta}z} + \frac{B}{\mu} \cdot \frac{1}{1 - \frac{\lambda}{\mu}z},$$

και χρησιμοποιώντας τη γεωμετρική σειρά (3.5) συμπεραίνουμε ότι

$$p_{(n,1)} = \frac{\theta(\mu - \lambda)}{\mu(\mu - \lambda - \theta)} \left(\frac{\lambda}{\lambda + \theta} \right)^n - \frac{\theta(\mu - \lambda)}{\mu(\mu - \lambda - \theta)} \left(\frac{\lambda}{\mu} \right)^n, \quad n \geq 1.$$

Περίπτωση 2: $\lambda + \theta = \mu$.

Στην περίπτωση αυτή η $P_1(z)$ γράφεται ως

$$P_1(z) = \frac{\lambda\theta^2 z}{(\lambda + \theta)(\lambda + \theta - \lambda z)^2} = \frac{\lambda\theta^2 z}{(\lambda + \theta)^3} \cdot \frac{1}{\left(1 - \frac{\lambda}{\lambda + \theta}z\right)^2}.$$

Χρησιμοποιώντας τη γενικευμένη διωνυμική σειρά (3.8), παίρνουμε

$$\begin{aligned} p_{(n,1)} &= \frac{\lambda\theta^2}{(\lambda+\theta)^3} \binom{-2}{n-1} (-1)^{n-1} \left(\frac{\lambda}{\lambda+\theta}\right)^{n-1} \\ &= \frac{n\theta^2\lambda^n}{(\lambda+\theta)^{n+2}}, \quad n \geq 1. \end{aligned}$$

Για να βρούμε την κατανομή του χρόνου παραμονής S ενός πελάτη στο σύστημα, μπορούμε να δεμεύσουμε στον αριθμό των πελατών που βρίσκει ο πελάτης κατά την άφιξή του, Q^- , καθώς και στην κατάσταση του υπηρέτη, I^- . Αν ο πελάτης βρει το σύστημα στην κατάσταση $(Q^-, I^-) = (n, 0)$ τότε ο χρόνος παραμονής του ισούται με το άθροισμα του υπολειπόμενου χρόνου ενεργοποίησης που έχει την κατανομή $\text{Exp}(\theta)$, λόγω της αμνήμονης ιδιότητας, και $n+1$ χρόνων εξυπηρέτησης, καθένας εκ των οποίων έχει την κατανομή $\text{Exp}(\mu)$. Ενώ, αν βρει το σύστημα στην κατάσταση $(Q^-, I^-) = (n, 1)$, τότε ο χρόνος παραμονής του ισούται με το άθροισμα του υπολειπόμενου χρόνου εξυπηρέτησης του πελάτη που βρίσκεται σε διαδικασία εξυπηρέτησης και n ακόμη χρόνων εξυπηρέτησης. Λαμβάνοντας υπόψη ότι η κατανομή ισορροπίας της κατάστασης (Q^-, I^-) του συστήματος σε στιγμές αφίξεων πελατών ταυτίζεται με την κατανομή ισορροπίας σε συνεχή χρόνο, λόγω της γενικευμένης ιδιότητας PASTA (βλέπε θεώρημα 2.5) που ισχύει λόγω των αφίξεων Poisson, έχουμε για τον μετασχηματισμό Laplace-Stieltjes του χρόνου παραμονής:

$$\begin{aligned} \tilde{F}_S(s) &= \sum_{i=0}^1 \sum_{n=i}^{\infty} \Pr[Q^- = n, I^- = i] E[e^{-sS} | Q^- = n, I^- = i] \\ &= \sum_{n=0}^{\infty} p_{(n,0)} \frac{\theta}{\theta+s} \left(\frac{\mu}{\mu+s}\right)^{n+1} + \sum_{n=1}^{\infty} p_{(n,1)} \left(\frac{\mu}{\mu+s}\right)^{n+1} \\ &= \frac{\theta}{\theta+s} \cdot \frac{\mu}{\mu+s} P_0 \left(\frac{\mu}{\mu+s}\right) + \frac{\mu}{\mu+s} P_1 \left(\frac{\mu}{\mu+s}\right). \end{aligned} \quad (9.17)$$

Υπολογίζοντας τώρα τις $P_0 \left(\frac{\mu}{\mu+s}\right)$ και $P_1 \left(\frac{\mu}{\mu+s}\right)$ από τις (9.15) και (9.16), αντίστοιχα, και αντικαθιστώντας στην (9.17) παίρνουμε μετά από κάποιες απλοποιήσεις τον μετασχηματισμό Laplace-Stieltjes του χρόνου παραμονής στη μορφή

$$\begin{aligned} \tilde{F}_S(s) &= \frac{\theta^2(\mu-\lambda)}{(\lambda+\theta)(s+\theta)(s+\frac{\theta\mu}{\lambda+\theta})} + \frac{\lambda\theta(\mu-\lambda)}{(\lambda+\theta)(s+\mu-\lambda)(s+\frac{\theta\mu}{\lambda+\theta})} \\ &= \left(1 - \frac{\lambda}{\mu}\right) \cdot \frac{\theta}{s+\theta} \cdot \frac{\frac{\theta\mu}{\lambda+\theta}}{s+\frac{\theta\mu}{\lambda+\theta}} + \frac{\lambda}{\mu} \cdot \frac{\mu-\lambda}{s+\mu-\lambda} \cdot \frac{\frac{\theta\mu}{\lambda+\theta}}{s+\frac{\theta\mu}{\lambda+\theta}}. \end{aligned} \quad (9.18)$$

Επομένως, η κατανομή του χρόνου παραμονής είναι μίξη με πιθανότητες $1 - \frac{\lambda}{\mu}$ και $\frac{\lambda}{\mu}$, αντίστοιχα, του αθροίσματος δυο ανεξάρτητων $\text{Exp}(\theta)$ και $\text{Exp}\left(\frac{\theta\mu}{\lambda+\theta}\right)$ και του αθροίσματος δυο ανεξάρτητων $\text{Exp}(\mu-\lambda)$ και $\text{Exp}\left(\frac{\theta\mu}{\lambda+\theta}\right)$.

9.3 Η M/M/1/1 ουρά με επαναπροσπάθειες

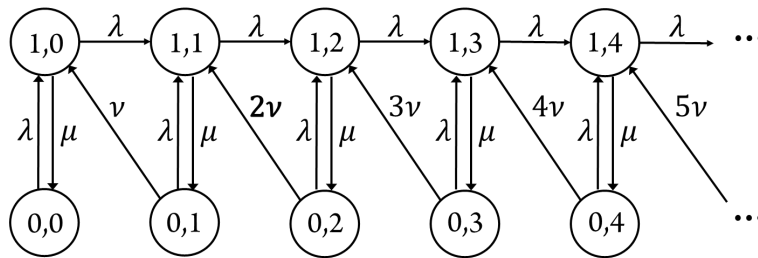
Θεωρούμε την M/M/1/1 ουρά με ρυθμό αφίξεων λ , ρυθμό εξυπηρέτησεων μ , 1 υπηρέτη και χωρητικότητα για 1 πελάτη, όπου οι πελάτες που βρίσκουν τον υπηρέτη απασχολημένο κατά την άφιξή τους μπαίνουν σε «τροχιά επαναπροσπάθειας», δηλαδή ξαναδοκιμάζουν να αποκτήσουν πρόσβαση σε αυτόν μέχρι να εξυπηρευτούν, ανεξάρτητα ο ένας από τον άλλο. Οι χρόνοι μεταξύ διαδοχικών προσπαθειών ενός πελάτη θεωρούνται ανεξάρτητοι και ακολουθούν την κατανομή $\text{Exp}(\nu)$. Η παράμετρος ν αναφέρεται ως ρυθμός επαναπροσπαθειών ανά πελάτη. Έστω $Q(t)$ ο αριθμός των πελατών τη στιγμή t στο σύστημα. Είναι φανερό ότι η $\{Q(t)\}$

Κατάσταση	Επόμενη κατάσταση	Χρόνος
(0, 0)	(1, 0)	$T_{(0,0),(1,0)} \sim \text{Exp}(\lambda)$
(0, n), n ≥ 1	(1, n)	$T_{(0,n),(1,n)} \sim \text{Exp}(\lambda)$
	(1, n - 1)	$T_{(0,n),(1,n-1)} \sim \text{Exp}(n\nu)$
(1, n), n ≥ 0	(1, n + 1)	$T_{(1,n),(1,n+1)} \sim \text{Exp}(\lambda)$
	(0, n)	$T_{(1,n),(0,n)} \sim \text{Exp}(\mu)$

Πίνακας 9.3: Πίνακας μεταβάσεων M/M/1/1 ουράς με επαναπροσπάθειες.

δεν είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου, αφού όταν $Q(t) = 0$, ο χρόνος μέχρι την επόμενη μετάβαση στην κατάσταση 1 είναι $\text{Exp}(\lambda + n\nu)$, όπου n είναι το πλήθος των πελατών σε τροχιά επαναπροσπάθειας (διότι η επόμενη νέα άφιξη στο σύστημα θα γίνει σε χρόνο $\text{Exp}(\lambda)$, ενώ οι n πελάτες σε τροχιά επαναπροσπάθειας επανέρχονται σε $\text{Exp}(\nu)$ χρόνους ο καθένας). Επομένως για μια Μαρκοβιανή περιγραφή του συστήματος χρειάζεται επίσης να καταγράφουμε σε κάθε χρονική στιγμή t τον αριθμό των πελατών σε τροχιά επαναπροσπάθειας, $R(t)$. Για τις μεταβάσεις της $\{(Q(t), R(t))\}$ έχουμε τότε τον πίνακα 9.3.

Αφού όλοι οι χρόνοι είναι εκθετικοί, έχουμε ότι η $\{(Q(t), R(t))\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου. Το διάγραμμα ρυθμών μετάβασής της δίνεται στο σχήμα 9.3.



Σχήμα 9.3: Διάγραμμα ρυθμών μετάβασης M/M/1/1 ουράς με επαναπροσπάθειες.

Έστω $(p_{(i,n)})$ η κατανομή ισορροπίας της. Οι εξισώσεις ισορροπίας της είναι

$$(\lambda + n\nu)p_{(0,n)} = \mu p_{(1,n)}, \quad n \geq 0, \tag{9.19}$$

$$(\lambda + \mu)p_{(1,0)} = \lambda p_{(0,0)} + \nu p_{(0,1)}, \tag{9.20}$$

$$(\lambda + \mu)p_{(1,n)} = \lambda p_{(1,n-1)} + \lambda p_{(0,n)} + (n + 1)\nu p_{(0,n+1)}, \quad n \geq 1. \tag{9.21}$$

Για να προσδιορίσουμε την κατανομή ισορροπίας, εισάγουμε τις μερικές πιθανογεννήτριες

$$P_0(z) = \sum_{n=0}^{\infty} p_{(0,n)} z^n, \quad |z| \leq 1,$$

$$P_1(z) = \sum_{n=0}^{\infty} p_{(1,n)} z^n, \quad |z| \leq 1.$$

Παρατηρήστε ότι και εδώ ο δείκτης άθροισης n αντιστοιχεί στην μεταβλητή που παίρνει άπειρες τιμές, που όμως στην περίπτωση αυτή αντιστοιχεί στην $R(t)$ και όχι στην $Q(t)$, όπως συνέβη στις προηγούμενες παραγράφους. Κατά τα άλλα, όμως, η μέθοδος των πιθανογεννητριών εφαρμόζεται με τον ίδιο τρόπο.

Πολλαπλασιάζοντας κάθε εξίσωση ισορροπίας που αναφέρεται σε κατάσταση $(0, n)$ με το αντίστοιχο z^n

(βλέπε εξίσωση (9.19)) και αθροίζοντας για όλα τα $n \geq 0$ παίρνουμε

$$\begin{aligned} \sum_{n=0}^{\infty} (\lambda + n\nu) p_{(0,n)} z^n &= \sum_{n=0}^{\infty} \mu p_{(1,n)} z^n \\ \Leftrightarrow \lambda P_0(z) + \nu z P_0'(z) &= \mu P_1(z). \end{aligned} \quad (9.22)$$

Ομοίως, πολλαπλασιάζοντας κάθε εξίσωση ισορροπίας που αναφέρεται σε κατάσταση $(1, n)$ με το αντίστοιχο z^n (βλέπε εξισώσεις (9.20), (9.21)) και αθροίζοντας για όλα τα $n \geq 0$, παίρνουμε

$$\begin{aligned} \sum_{n=0}^{\infty} (\lambda + \mu) p_{(1,n)} z^n &= \sum_{n=1}^{\infty} \lambda p_{(1,n-1)} z^n + \sum_{n=0}^{\infty} \lambda p_{(0,n)} z^n + \sum_{n=0}^{\infty} \nu(n+1) p_{(0,n+1)} z^n \\ \Leftrightarrow (\lambda + \mu) P_1(z) &= \lambda z P_1(z) + \lambda P_0(z) + \nu P_0'(z). \end{aligned} \quad (9.23)$$

Διαιρώντας τις εξισώσεις (9.22) και (9.23) με μ και θέτοντας

$$\rho = \frac{\lambda}{\mu},$$

έχουμε το σύστημα 2 εξισώσεων και 2 αγνώστων για τις $P_0(z), P_1(z)$:

$$\rho P_0(z) + \frac{\nu}{\mu} z P_0'(z) = P_1(z), \quad (9.24)$$

$$(1 + \rho - \rho z) P_1(z) = \rho P_0(z) + \frac{\nu}{\mu} P_0'(z). \quad (9.25)$$

Αντικαθιστώντας την (9.24) στην (9.25) έχουμε

$$\begin{aligned} (1 + \rho - \rho z) \rho P_0(z) + (1 + \rho - \rho z) \frac{\nu}{\mu} z P_0'(z) &= \rho P_0(z) + \frac{\nu}{\mu} P_0'(z) \\ \Leftrightarrow \rho^2 (1 - z) P_0(z) &= \frac{\nu}{\mu} (1 - z) (1 - \rho z) P_0'(z), \end{aligned}$$

οπότε

$$(\log P_0(z))' = \frac{P_0'(z)}{P_0(z)} = \frac{\rho^2 \mu}{\nu(1 - \rho z)} = \frac{\lambda \rho}{\nu(1 - \rho z)}.$$

Ολοκληρώνοντας έχουμε:

$$\log P_0(z) - \log P_0(0) = \frac{\lambda \rho}{\nu} \int_0^z \frac{1}{1 - \rho u} du = -\frac{\lambda}{\nu} \log(1 - \rho z).$$

Επομένως,

$$P_0(z) = (1 - \rho z)^{-\frac{\lambda}{\nu}} p_{(0,0)}. \quad (9.26)$$

Αντικαθιστώντας την (9.26) στην (9.24) και κάνοντας μερικές απλοποιήσεις παίρνουμε:

$$P_1(z) = \rho (1 - \rho z)^{-\left(\frac{\lambda}{\nu} + 1\right)} p_{(0,0)}. \quad (9.27)$$

Το $p_{(0,0)}$ βρίσκεται από την εξίσωση κανονικοποίησης $P_0(1) + P_1(1) = 1$ και είναι

$$p_{(0,0)} = (1 - \rho)^{\frac{\lambda}{\nu} + 1}.$$

Για να βρούμε τις πιθανότητες ισορροπίας $p_{(0,n)}, p_{(1,n)}, n \geq 0$ χρησιμοποιούμε τη γενικευμένη διωνυμική σειρά (3.8) για να αναπτύξουμε τις (9.26) και (9.27). Τελικά παίρνουμε

$$\begin{aligned} p_{(0,n)} &= (1-\rho)^{\frac{\lambda}{\nu}+1} \binom{-\lambda/\nu}{n} (-1)^n \rho^n \\ &= (1-\rho)^{\frac{\lambda}{\nu}+1} \frac{\lambda(\lambda+\nu)(\lambda+2\nu)\cdots(\lambda+(n-1)\nu)\rho^n}{n!\nu^n}, \quad n \geq 0, \end{aligned}$$

και

$$p_{(1,n)} = (1-\rho)^{\frac{\lambda}{\nu}+1} \frac{(\lambda+\nu)(\lambda+2\nu)\cdots(\lambda+n\nu)\rho^{n+1}}{n!\nu^n}, \quad n \geq 0.$$

9.4 Ασκήσεις

Άσκηση 9.1 Θεωρούμε δυο συστήματα εξυπηρέτησης O_1 και O_2 , συνδεδεμένα σε σειρά, με έναν υπηρέτη το καθένα, χωρίς χώρο αναμονής. Οι χρόνοι εξυπηρέτησης στα δυο συστήματα είναι εκθετικά κατανομημένοι με παραμέτρους μ_1 και μ_2 αντίστοιχα. Οι αφίξεις στο πρώτο σύστημα O_1 γίνονται σύμφωνα με μια διαδικασία Poisson με ρυθμό λ και κάθε πελάτης που βρίσκει το σύστημα αυτό κατελημμένο αναχωρεί άμεσα. Κάθε πελάτης που αναχωρεί από το σύστημα O_1 πηγαίνει στο O_2 αν αυτό είναι κενό. Διαφορετικά παραμένει στο σύστημα O_1 μέχρι να αδειάσει το σύστημα O_2 . Στην περίπτωση αυτή, το σύστημα O_1 δεν παρέχει εξυπηρέτηση και αδρανοποιείται. Μόλις όμως, το σύστημα O_2 αδειάσει, ο πελάτης που περιμένει στο O_1 μεταφέρεται ακαριαία σε αυτό. Να ορίσετε μια Μαρκοβιανή αλυσίδα συνεχούς χρόνου που να περιγράφει την εξέλιξη του συστήματος στον χρόνο και να παραθέσετε το διάγραμμα των ρυθμών μετάβασης της.

Άσκηση 9.2 Θεωρούμε την $M/M/1$ ουρά με ρυθμό αφίξεων λ , ρυθμό εξυπηρέτησεων μ , 1 υπηρέτη και απεριόριστη χωρητικότητα υπό την πειθαρχία ουράς FCFS. Ο υπηρέτης μπαίνει σε διαδικασία απενεργοποίησης μόλις το σύστημα αδειάσει. Η διάρκεια της διαδικασίας απενεργοποίησης έχει την $\text{Exp}(\zeta)$ κατανομή. Αν αφιχθεί πελάτης κατά τη διάρκεια της διαδικασίας απενεργοποίησης αυτή διακόπτεται ακαριαία και ο πελάτης αρχίζει να εξυπηρετείται. Αν η διαδικασία της απενεργοποίησης ολοκληρωθεί, τότε ο υπηρέτης απενεργοποιείται και δεν μπορεί να παρέχει άμεσα εξυπηρέτηση. Όταν ένας νέος πελάτης αφιχθεί και βρει τον υπηρέτη απενεργοποιημένο, ο υπηρέτης ξεκινά έναν χρόνο επανεκκίνησης με το πέρας του οποίου ο υπηρέτης αρχίζει να παρέχει εξυπηρέτηση μέχρι το σύστημα να μείνει ξανά κενό και να ξεκινήσει η διαδικασία απενεργοποίησης κ.ο.κ. Στη διάρκεια ενός χρόνου επανεκκίνησης το σύστημα δέχεται νέες αφίξεις, αλλά δεν παρέχει εξυπηρέτηση. Οι χρόνοι επανεκκίνησης του υπηρέτη είναι $\text{Exp}(\theta)$. Το σύστημα αυτό αναφέρεται ως η $M/M/1$ ουρά με χρόνους επανεκκίνησης και απενεργοποίησης. Έστω $Q(t)$ ο αριθμός των πελατών στο σύστημα και $I(t)$ η κατάσταση του υπηρέτη (0: απενεργοποιημένος, 1: ενεργοποιημένος), τη στιγμή t .

1. Να αποδείξετε ότι η $\{(Q(t), I(t))\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου.
2. Να κάνετε το διάγραμμα των ρυθμών μετάβασης της.
3. Να προσδιορίσετε την κατανομή ισορροπίας της.
4. Να βρείτε το μακροπρόθεσμο ποσοστό του χρόνου που ο υπηρέτης είναι ενεργοποιημένος.

Άσκηση 9.3 Θεωρούμε την $M/M/1$ ουρά με ρυθμό αφίξεων λ και $\text{Exp}(\mu)$ κατανομή χρόνων εξυπηρέτησης, όπου μόλις το σύστημα αδειάσει ο υπηρέτης παύει να παρέχει εξυπηρέτηση και αρχίζει και πάλι να εξυπηρετεί, μόλις μαζευτούν K πελάτες με $K \geq 1$. Το σύστημα αυτό αναφέρεται ως η $M/M/1$ ουρά με την K -πολιτική ενεργοποίησης. Έστω $Q(t)$ ο αριθμός των πελατών στο σύστημα και $I(t)$ η κατάσταση του υπηρέτη (0: απενεργοποιημένος, 1: ενεργοποιημένος), τη στιγμή t .

1. Να αποδείξετε ότι η $\{(Q(t), I(t))\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου.

2. Να κάνετε το διάγραμμα των ρυθμών μετάβασής της.
3. Να προσδιορίσετε την κατανομή ισορροπίας της.
4. Να βρείτε το μακροπρόθεσμο ποσοστό του χρόνου που ο υπηρέτης είναι ενεργοποιημένος.

Άσκηση 9.4 Θεωρούμε την $M/M/1/1$ ουρά με ρυθμό αφίξεων λ , ρυθμό εξυπηρέτησεων μ , 1 υπηρέτη και χωρητικότητα για 1 πελάτη, όπου οι πελάτες που βρίσκουν τον υπηρέτη απασχολημένο κατά την άφιξή τους μπαίνουν σε «τροχιά αναμονής», δηλαδή αφήνουν τα στοιχεία επικοινωνίας τους και όταν ο υπηρέτης τελειώσει μια εξυπηρέτηση αρχίζει να ανακτά τον παλαιότερο από αυτούς στην τροχιά για να αρχίσει να τον εξυπηρετεί. Οι χρόνοι ανάκτησης ακολουθούν την κατανομή $\text{Exp}(\alpha)$. Αν κατά τη διάρκεια ενός χρόνου ανάκτησης αφιχθεί νέος πελάτης τότε η διαδικασία ανάκτησης εγκαταλείπεται και ο υπηρέτης αρχίζει την εξυπηρέτηση του νεοεισαχθέντος πελάτη. Αν όμως ο χρόνος ανάκτησης ολοκληρωθεί, τότε ο παλαιότερος πελάτης στην τροχιά αναμονής ξεκινάει να εξυπηρετείται. Το σύστημα αυτό αναφέρεται ως η $M/M/1/1$ ουρά με σταθερές επαναπροσπάθειες και η παράμετρος α αναφέρεται ως ο ρυθμός ανάκτησης πελατών από την τροχιά αναμονής/επαναπροσπαθειών. Έστω $Q(t)$ ο αριθμός των πελατών στο σύστημα (δηλαδή στο χώρο εξυπηρέτησης) και $R(t)$ ο αριθμός των πελατών σε τροχιά αναμονής, τη στιγμή t .

1. Να αποδείξετε ότι η $\{(Q(t), R(t))\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου.
2. Να κάνετε το διάγραμμα των ρυθμών μετάβασής της.
3. Να προσδιορίσετε την κατανομή ισορροπίας της.
4. Να βρείτε το μακροπρόθεσμο ποσοστό του χρόνου που ο υπηρέτης είναι απασχολημένος.

Άσκηση 9.5 Θεωρούμε την $M/M/1/k$ ουρά με ρυθμό αφίξεων λ , ρυθμό εξυπηρέτησεων μ , 1 υπηρέτη και χωρητικότητα για k συνολικά πελάτες υπό την πειθαρχία ουράς FCFS. Ο υπηρέτης εναλλάσσεται μεταξύ ενεργών και ανενεργών περιόδων εξυπηρέτησεων. Κατά τη διάρκεια των ενεργών περιόδων παρέχει εξυπηρέτηση, ενώ κατά τη διάρκεια των ανενεργών περιόδων παύει να παρέχει εξυπηρέτηση, αλλά το σύστημα συνεχίζει να δέχεται νέες αφίξεις πελατών (εφόσον δεν είναι γεμάτο). Οι διάρκειες των ενεργών και των ανενεργών περιόδων είναι όλες ανεξάρτητες μεταξύ τους. Επιπλέον υποθέτουμε ότι οι διάρκειες των ενεργών περιόδων έχουν την κατανομή $\text{Exp}(\zeta)$, ενώ οι διάρκειες των ανενεργών περιόδων έχουν την κατανομή $\text{Exp}(\theta)$. Το σύστημα αυτό αναφέρεται ως η $M/M/1/k$ ουρά με ενεργές και ανενεργές περιόδους εξυπηρέτησεων. Έστω $Q(t)$ ο αριθμός των πελατών στο σύστημα και $I(t)$ η κατάσταση του υπηρέτη (0: απενεργοποιημένος, 1: ενεργοποιημένος), τη στιγμή t .

1. Να αποδείξετε ότι η $\{(Q(t), I(t))\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου και να κάνετε το διάγραμμα των ρυθμών μετάβασής της.
2. Να προσδιορίσετε την κατανομή ισορροπίας της.

9.5 Σχόλια

Για την μελέτη διδιάστατων Μαρκοβιανών ουρών χρησιμοποιούνται κυρίως τρεις μεθοδολογικές προσεγγίσεις: πιθανογεννήτριες, πινακο-αναλυτικές μέθοδοι και η θεωρία των εξισώσεων διαφορών.

Η μέθοδος των πιθανογεννητριών, την οποία και αναπτύξαμε αρκετά διεξοδικά στο παρόν κεφάλαιο, δεν διαφέρει κατ' ουσίαν από το πλαίσιο που είδαμε και στο κεφάλαιο 7. Λεπτομέρειες για τη μέθοδο των πιθανογεννητριών μπορεί κανείς να δει στα κλασικά βιβλία Kleinrock 1975, Wolff 1989, Tijms 2008, Kulkarni 2010 και Shortle, Thompson, Gross και Harris 2018. Για μια σύνοψη μεθόδων με πολλά παραδείγματα μπορεί κανείς να ανατρέξει στη μεταπτυχιακή διπλωματική εργασία της Κάντα 2006.

Για τις πινακο-αναλυτικές μεθόδους, ο ενδιαφερόμενος αναγνώστης θα πρέπει να μελετήσει τα κλασικά συγγράμματα των Neuts 1981, Neuts 1989 και Latouche και Ramaswami 1999.

Για μια στοιχειώδη εισαγωγή στη θεωρία των εξισώσεων διαφορών με εφαρμογές στη μελέτη συστημάτων εξυπηρέτησης, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο σύγγραμμα Adan και Wal 1998.

Βιβλιογραφία

- [1] L. Kleinrock. *Queueing Systems. Volume 1: Theory*. Wiley, 1975. ISBN: 978-0471491101.
- [2] R.W. Wolff. *Stochastic Modeling and the Theory of Queues*. Englewood Cliffs, New Jersey: Prentice Hall, 1989. ISBN: 978-0138466923.
- [3] H.C. Tijms. *A First Course in Stochastic Models, 2nd Edition*. Chichester: Wiley, 2008. ISBN: 978-0471498803.
- [4] V.G. Kulkarni. *Modeling and Analysis of Stochastic Systems, 2nd Edition*. CRC Press, 2010. ISBN: 978-1439808757.
- [5] J.F. Shortle, J.M. Thompson, D. Gross και C.M. Harris. *Fundamentals of Queueing Theory, 5th Edition*. Wiley, 2018. ISBN: 978-1118943526.
- [6] Σ. Κάντα. *Αλγοριθμικές Τεχνικές στις Μαρκοβιανές Αλυσίδες: Μέθοδοι Πιθανογεννητριών και Εφαρμογές*. Αθήνα: Διπλωματική Εργασία, Μεταπτυχιακό Πρόγραμμα Ειδίκευσης στη Στατιστική και Επιχειρησιακή Έρευνα, Τμήμα Μαθηματικών Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών, 2006.
- [7] M.F. Neuts. *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*. The Johns Hopkins University Press, 1981. ISBN: 978-0801825606.
- [8] M.F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. CRC Press, 1989. ISBN: 978-0824782832.
- [9] G. Latouche και V. Ramaswami. *Introduction to Matrix Analytic Models in Stochastic Modeling*. ASA-SIAM Series on Statistics και Applied Probability, 1999. ISBN: 978-0898714258.
- [10] I. Adan και J. van der Wal. *Difference and Differential Equations in Stochastic Operations Research*. Eindhoven: Notes available online, 1998.

ΑΝΤΙΣΤΡΕΨΙΜΕΣ ΜΑΡΚΟΒΙΑΝΕΣ ΟΥΡΕΣ

Η αντιστρεψιμότητα είναι μια ιδιότητα που έχουν λίγες στοχαστικές διαδικασίες. Όμως, όταν υπάρχει, διευκολύνει πολύ τους υπολογισμούς που αφορούν την αντίστοιχη στοχαστική διαδικασία. Στο κεφάλαιο αυτό θα δούμε τις έννοιες της αντίστροφης μιας Μαρκοβιανής αλυσίδας συνεχούς χρόνου και της αντιστρέψιμης Μαρκοβιανής αλυσίδας συνεχούς χρόνου, καθώς και το πώς οι έννοιες αυτές μπορούν να χρησιμεύσουν για τη μελέτη κάποιων συστημάτων εξυπηρέτησης.

10.1 Αντίστροφη στοχαστικής διαδικασίας

Η αντίστροφη μιας στοχαστικής διαδικασίας $\{X(t)\}$ ως προς μια χρονική στιγμή τ είναι διαισθητικά η διαδικασία $\{\hat{X}_\tau(t)\}$ που βλέπουμε αν θεωρήσουμε ως χρονική στιγμή 0 τη στιγμή τ και κοιτάξουμε την εξέλιξη της $\{X(t)\}$ αντίστροφα στον χρόνο. Πιο συγκεκριμένα, έχουμε τον ακόλουθο ορισμό.

Ορισμός 10.1 (Αντίστροφη στοχαστικής διαδικασίας) Έστω $\{X(t) : t \in \mathbb{R}\}$ μια στοχαστική διαδικασία και $\tau \in \mathbb{R}$. Η αντίστροφη της $\{X(t)\}$ ως προς το τ είναι η διαδικασία $\{\hat{X}_\tau(t)\}$ με

$$\hat{X}_\tau(t) = X(\tau - t), \quad t \in \mathbb{R}.$$

Για $\tau = 0$, η $\{\hat{X}_0(t)\}$ λέγεται τυπική αντίστροφη της $\{X(t)\}$ και συμβολίζεται απλά με $\{\hat{X}(t)\}$. ◊

Στην περίπτωση που η αρχική στοχαστική διαδικασία $\{X(t)\}$ είναι στάσιμη αδιαχώριστη Μαρκοβιανή αλυσίδα συνεχούς χρόνου (δηλαδή είναι αδιαχώριστη και η μεταβατική κατανομή της για κάθε χρονική στιγμή t είναι η στάσιμη κατανομή της), τότε και η αντίστροφή της είναι στάσιμη αδιαχώριστη Μαρκοβιανή αλυσίδα συνεχούς χρόνου και οι μάλιστα οι ρυθμοί της και η στάσιμη κατανομή της είναι άμεσα υπολογίσιμοι από τα αντίστοιχα στοιχεία της αρχικής.

Θεώρημα 10.2 (Αντίστροφη στάσιμης αδιαχώριστης Μαρκοβιανής αλυσίδας συνεχούς χρόνου) Η αντίστροφη $\{\hat{X}(t)\}$ μιας στάσιμης αδιαχώριστης Μαρκοβιανής αλυσίδας συνεχούς χρόνου $\{X(t)\}$ με χώρο καταστάσεων

\mathcal{S} , ρυθμούς μετάβασης q_{ij} και κατανομή ισορροπίας (p_i) είναι επίσης Μαρκοβιανή αλυσίδα συνεχούς χρόνου με ρυθμούς $\hat{q}_{ij} = \frac{p_j q_{ji}}{p_i}, j \neq i$, και κατανομή ισορροπίας (\hat{p}_i) με $\hat{p}_i = p_i$ για $i \in \mathcal{S}$. \diamond

Μπορούμε να αιτιολογήσουμε εύκολα αυτό το θεώρημα. Πράγματι, όσον αφορά τις πιθανότητες ισορροπίας της αντίστροφης έχουμε

$$\hat{p}_i = \Pr[\hat{X}(t) = i] = \Pr[X(-t) = i] = p_i, \quad i \in \mathcal{S},$$

ενώ όσον αφορά τους ρυθμούς μετάβασης της παίρνουμε

$$\begin{aligned} \hat{q}_{ij} &= \lim_{h \rightarrow 0^+} \frac{\Pr[\hat{X}(t+h) = j | \hat{X}(t) = i]}{h} \\ &= \lim_{h \rightarrow 0^+} \frac{\Pr[X(-t-h) = j | X(-t) = i]}{h}, \quad i, j \in \mathcal{S}, i \neq j. \end{aligned}$$

Χρησιμοποιώντας τον κανόνα του Bayes έχουμε

$$\begin{aligned} \hat{q}_{ij} &= \lim_{h \rightarrow 0^+} \frac{\Pr[X(-t-h) = j] \Pr[X(-t) = i | X(-t-h) = j]}{\Pr[X(-t) = i] h} \\ &= \frac{p_j q_{ji}}{p_i}, \quad i, j \in \mathcal{S}, i \neq j. \end{aligned}$$

Τα αποτελέσματα είναι εύλογα και από διαισθητική σκοπιά. Πράγματι η πιθανότητα p_i είναι το μακροπρόθεσμο ποσοστό του χρόνου που η $\{X(t)\}$ περνάει στην κατάσταση i , ενώ η \hat{p}_i είναι το μακροπρόθεσμο ποσοστό του χρόνου που η ίδια Μαρκοβιανή αλυσίδα περνάει στην κατάσταση i , αν κοιτάμε το χρόνο με αντίστροφη φορά. Αυτά τα δυο ποσοστά είναι ίσα, καθώς η φορά του χρόνου δεν παίζει κάποιο ρόλο, οπότε έπεται ότι $\hat{p}_i = p_i, i \in \mathcal{S}$. Επίσης, η ποσότητα $p_j q_{ji}$ δίνει τον μακροπρόθεσμο ρυθμό μεταβάσεων της $\{X(t)\}$ από την j στην i ανά χρονική μονάδα. Κάθε τέτοια μετάβαση αντιστοιχεί σε μια μετάβαση της $\{\hat{X}(t)\}$ από την i στην j , αφού η $\{\hat{X}(t)\}$ προκύπτει κοιτάζοντας την $\{X(t)\}$ αντίστροφα στο χρόνο. Άρα θα πρέπει ο μακροπρόθεσμος ρυθμός μεταβάσεων της $\{X(t)\}$ από την j στην i ανά χρονική μονάδα να ισούται με τον μακροπρόθεσμο ρυθμό μεταβάσεων της $\{\hat{X}(t)\}$ από την i στην j ανά χρονική μονάδα, δηλαδή $p_j q_{ji} = \hat{p}_i \hat{q}_{ij}$. Λύνοντας ως προς την \hat{q}_{ij} , και χρησιμοποιώντας ότι $\hat{p}_i = p_i$, παίρνουμε $\hat{q}_{ij} = \frac{p_j q_{ji}}{p_i}$.

Για κάποιες περίπλοκες Μαρκοβιανές αλυσίδες συνεχούς χρόνου, όπου ο χώρος καταστάσεων είναι πολυδιάστατος, όπως αυτές που προκύπτουν κατά τη μελέτη δικτύων συστημάτων εξυπηρέτησης, είναι πολλές φορές δυνατό να εικάσει κανείς την κατανομή ισορροπίας τους. Για να αποδείξει κατόπιν ότι η εικαζόμενη κατανομή (p_j) είναι πράγματι η σωστή θα πρέπει κατόπιν να ελέγξει ότι ικανοποιεί το σύστημα των εξισώσεων ισορροπίας

$$p_j q_j = \sum_{i \neq j} p_i q_{ij}, \quad j \in \mathcal{S},$$

όπου q_{ij} είναι οι ρυθμοί μετάβασης της αντίστοιχης Μαρκοβιανής αλυσίδας συνεχούς χρόνου και $q_j = \sum_{i \neq j} q_{ji}$ οι ρυθμοί εξόδου από τις διάφορες καταστάσεις. Αντί αυτού, μπορεί να προχωρήσει και να εικάσει και τη μορφή της αντίστροφης Μαρκοβιανής αλυσίδας (δηλαδή να εικάσει τους ρυθμούς μετάβασης της, \hat{q}_{ij}) και κατόπιν να ελέγξει ότι η εικαζόμενη κατανομή ισορροπίας ικανοποιεί το απλούστερο σύστημα

$$p_j q_{ji} = p_i \hat{q}_{ij}, \quad i \neq j.$$

Πιο συγκεκριμένα, έχουμε το παρακάτω αποτέλεσμα.

Θεώρημα 10.3 (Έλεγχος για την κατανομή ισορροπίας με εικαζόμενη αντίστροφη) Έστω $\{X(t)\}$ μια διαχωρίστη Μαρκοβιανή αλυσίδα συνεχούς χρόνου με χώρο καταστάσεων \mathcal{S} και ρυθμούς μετάβασης q_{ij} . Έστω επίσης $(p_j : j \in \mathcal{S})$ μια συνάρτηση πιθανότητας και $\tilde{q}_{ij}, i \neq j$ μη-αρνητικοί αριθμοί τέτοιοι ώστε

$$p_i q_{ij} = p_j \tilde{q}_{ji}, \quad i \neq j, \quad (10.1)$$

$$\tilde{q}_i \equiv \sum_{j \neq i} \tilde{q}_{ij} = \sum_{j \neq i} q_{ij} \equiv q_i, \quad i \in \mathcal{S}. \quad (10.2)$$

Τότε η $\{X(t)\}$ είναι θετικά επαναληπτική με κατανομή ισορροπίας (p_j) και οι \tilde{q}_{ij} είναι οι ρυθμοί μετάβασης της αντίστροφής της. \diamond

Η απόδειξη αυτού του θεωρήματος είναι άμεση. Πράγματι, για κάθε σταθερό j , αθροίζοντας τις (10.1) για όλα τα $i \neq j$, παίρνουμε

$$\sum_{i \neq j} p_i q_{ij} = p_j \sum_{i \neq j} \tilde{q}_{ji} = p_j \tilde{q}_j = p_j q_j, \quad j \in \mathcal{S},$$

οπότε η συνάρτηση πιθανότητας (p_j) ικανοποιεί τις εξισώσεις ισορροπίας και άρα είναι η κατανομή ισορροπίας της $\{X(t)\}$. Για τους ρυθμούς \hat{q}_{ij} της αντίστροφης, έχουμε από το θεώρημα 10.2 και τις εξισώσεις (10.1) ότι

$$\hat{q}_{ji} = \frac{p_i q_{ij}}{p_j} = \tilde{q}_{ji}.$$

10.2 Αντιστρέψιμες στοχαστικές διαδικασίες

Μια στοχαστική διαδικασία που είναι όμοια με την αντίστροφή της από πιθανοθεωρητική σκοπιά, λέγεται αντιστρέψιμη. Πιο συγκεκριμένα έχουμε τον ακόλουθο ορισμό.

Ορισμός 10.4 (Αντιστρέψιμη στοχαστική διαδικασία) Έστω $\{X(t) : t \in \mathbb{R}\}$ μια στοχαστική διαδικασία και $\{\hat{X}_\tau(t)\}$ με

$$\hat{X}_\tau(t) = X(\tau - t), \quad t \in \mathbb{R},$$

μια αντίστροφή της. Αν οι $\{X(t)\}, \{\hat{X}_\tau(t)\}$ είναι στοχαστικά ισοδύναμες, δηλαδή για κάθε επιλογή χρονικών στιγμών $t_1 < t_2 < \dots < t_n$ οι $(X(t_1), X(t_2), \dots, X(t_n))$ και $(\hat{X}_\tau(t_1), \hat{X}_\tau(t_2), \dots, \hat{X}_\tau(t_n))$ είναι ισόνομες, τότε η $\{X(t)\}$ λέγεται αντιστρέψιμη στοχαστική διαδικασία. \diamond

Θεώρημα 10.5 (Χαρακτηρισμός αντιστρεψιμότητας μέσω λεπτομερούς ισορροπίας) Μια στάσιμη αδιαχώριστη Μαρκοβιανή αλυσίδα συνεχούς χρόνου $\{X(t)\}$ με χώρο καταστάσεων \mathcal{S} , ρυθμούς μετάβασης q_{ij} και κατανομή ισορροπίας (p_i) είναι αντιστρέψιμη αν και μόνο αν η κατανομή (p_i) ικανοποιεί τις εξισώσεις λεπτομερούς ισορροπίας

$$p_i q_{ij} = p_j q_{ji}, \quad i, j \in \mathcal{S}, \quad i \neq j. \quad (10.3)$$

Οι εξισώσεις λεπτομερούς ισορροπίας (10.3) είναι πιο περιοριστικές από τις εξισώσεις πλήρους ισορροπίας (4.7) και τις εξισώσεις γενικευμένης ισορροπίας (4.9), υπό την έννοια του ότι δεν ικανοποιούνται πάντα από την κατανομή ισορροπίας μιας Μαρκοβιανής αλυσίδας συνεχούς χρόνου. Πράγματι, ενώ κάθε Μαρκοβιανή αλυσίδα συνεχούς χρόνου έχει αντίστροφή, μόνο λίγες Μαρκοβιανές αλυσίδες συνεχούς χρόνου είναι αντιστρέψιμες.

Για να δούμε γιατί οι εξισώσεις λεπτομερούς ισορροπίας χαρακτηρίζουν την αντιστρεψιμότητα, αρκεί να θυμηθούμε ότι μια στάσιμη Μαρκοβιανή αλυσίδα συνεχούς χρόνου $\{X(t)\}$ προσδιορίζεται πλήρως από τους ρυθμούς μετάβασής της, q_{ij} . Επομένως η $\{X(t)\}$ είναι αντιστρέψιμη αν και μόνο αν $q_{ij} = \hat{q}_{ij} = \frac{p_j q_{ji}}{p_i}$, για κάθε $i, j \in \mathcal{S}$, δηλαδή αν ισχύουν οι εξισώσεις λεπτομερούς ισορροπίας (10.3).

Οι εξισώσεις λεπτομερούς ισορροπίας επιτρέπουν να συνδέει κανείς άμεσα δυο οποιοσδήποτε πιθανότητες ισορροπίας που αντιστοιχούν σε καταστάσεις που επικοινωνούν κατευθείαν και γι αυτό διευκολύνουν τον υπολογισμό της κατανομής ισορροπίας μιας αντιστρέψιμης Μαρκοβιανής αλυσίδας συνεχούς χρόνου. Είναι σημαντικό, επομένως, να μπορούμε να ελέγξουμε πότε μια Μαρκοβιανή αλυσίδα συνεχούς χρόνου είναι αντιστρέψιμη, χωρίς να χρειάζεται να υπολογίσουμε τη κατανομή ισορροπίας της. Αυτό γίνεται με το ακόλουθο κριτήριο αντιστρεψιμότητας.

Θεώρημα 10.6 (Χαρακτηρισμός αντιστρεψιμότητας Kolmogorov) Μια αδιαχώριστη στάσιμη Μαρκοβιανή αλυσίδα συνεχούς χρόνου είναι αντιστρέψιμη αν και μόνο αν για κάθε κύκλο καταστάσεων $(i_0, i_1, i_2, \dots, i_n, i_0)$, το γινόμενο των ρυθμών μετάβασης είναι το ίδιο ως προς τις δυο φορές διαγραφής του κύκλου, δηλαδή αν

$$q_{i_0 i_1} q_{i_1 i_2} q_{i_2 i_3} \cdots q_{i_{n-1} i_n} q_{i_n i_0} = q_{i_0 i_n} q_{i_n i_{n-1}} q_{i_{n-1} i_{n-2}} \cdots q_{i_2 i_1} q_{i_1 i_0}.$$

◊

Είναι εύκολο να δούμε ότι για να είναι μια Μαρκοβιανή αλυσίδα συνεχούς χρόνου αντιστρέψιμη, αρκεί να περιοριστούμε στον έλεγχο της ισότητας των γινομένων των ρυθμών μετάβασης, μόνο σε απλούς κύκλους καταστάσεων (δηλ. χωρίς επαναλήψεις καταστάσεων) με τουλάχιστον τρεις καταστάσεις. Ως άμεση εφαρμογή του κριτηρίου αντιστρεψιμότητας του Kolmogorov έχουμε ότι κάθε Μαρκοβιανή αλυσίδα συνεχούς χρόνου της οποίας το διάγραμμα ρυθμών μετάβασης είναι αμφίδρομο δέντρο (δηλαδή δεν υπάρχουν κύκλοι και αν υπάρχει θετικός ρυθμός μετάβασης από μια κατάσταση i σε μια κατάσταση j υπάρχει και θετικός ρυθμός μετάβασης από την j στην i) είναι αντιστρέψιμη. Ειδικότερα κάθε Μαρκοβιανή αλυσίδα συνεχούς χρόνου γέννησης-θανάτου είναι αντιστρέψιμη.

Ο τρόπος με τον οποίο ελέγχουμε την αντιστρεψιμότητα μιας Μαρκοβιανής αλυσίδας συνεχούς χρόνου και κατόπιν υπολογίζουμε την κατανομής ισορροπίας της, εφόσον αποδειχθεί ότι είναι αντιστρέψιμη, συνοψίζεται στον παρακάτω αλγόριθμο.

Αλγόριθμος 10.7 (Έλεγχος αντιστρεψιμότητας και προσδιορισμός κατανομής ισορροπίας) Έστω $\{X(t)\}$ μια Μαρκοβιανή αλυσίδα συνεχούς χρόνου.

Βήμα 1ο: Ελέγχουμε αν η $\{X(t)\}$ είναι αντιστρέψιμη με το κριτήριο Kolmogorov. Αν δεν είναι, τότε θα πρέπει να προχωρήσουμε με κάποιον άλλο τρόπο υπολογισμού της κατανομής ισορροπίας (π.χ. εξισώσεις πλήρους ή γενικευμένης ισορροπίας, πιθανογεννήτριες κ.τ.λ.). Αν είναι, τότε προχωρούμε στο επόμενο βήμα.

Βήμα 2ο: Ορίζουμε μια κατάσταση αναφοράς i_0 και για κάθε κατάσταση j βρίσκουμε ένα μονοπάτι που τη συνδέει με την i_0 , έστω το $i_0 \rightarrow i_1 \rightarrow i_2 \rightarrow \cdots \rightarrow i_n \rightarrow j$. Τότε οι εξισώσεις λεπτομερούς ισορροπίας δίνουν για την πιθανότητα ισορροπίας p_j ότι

$$p_j = p_{i_0} \frac{q_{i_0 i_1} q_{i_1 i_2} q_{i_2 i_3} \cdots q_{i_{n-1} i_n} q_{i_n j}}{q_{i_1 i_0} q_{i_2 i_1} q_{i_3 i_2} \cdots q_{i_n i_{n-1}} q_{j i_n}}.$$

Με άλλα λόγια, η πιθανότητα ισορροπίας p_j υπολογίζεται συναρτήσει της p_{i_0} σαν να είχαμε αλυσίδα γέννησης-θανάτου με χώρο καταστάσεων $\{i_0, i_1, \dots, i_n, j\}$.

Βήμα 3ο: Η p_{i_0} υπολογίζεται από την εξίσωση κανονικοποίησης.

◊

Στις επόμενες παραγράφους θα χρησιμοποιήσουμε τον αλγόριθμο αυτό για τον υπολογισμό των κατανομών ισορροπίας διαφόρων συστημάτων εξυπηρέτησης που αναπαριστώνται μέσω αντιστρέψιμων Μαρκοβιανών αλυσίδων συνεχούς χρόνου.

Ένας άλλος τρόπος για να διαπιστώσουμε ότι μια Μαρκοβιανή αλυσίδα είναι αντιστρέψιμη είναι να παρατηρήσουμε ότι κατασκευάζεται από απλούστερες αντιστρέψιμες Μαρκοβιανές αλυσίδες μέσω διαδικασιών που διατηρούν την αντιστρεψιμότητα. Αυτός ο τρόπος είναι ιδιαίτερα χρήσιμος για Μαρκοβιανές αλυσίδες με πολυδιάστατο χώρο καταστάσεων. Ένα τυπικό παράδειγμα αυτής της ιδέας θα δούμε στην παράγραφο 10.5. Δυο τέτοιες διαδικασίες είναι η παράθεση ανεξάρτητων αντιστρέψιμων διαδικασιών και η περικοπή μιας αντιστρέψιμης Μαρκοβιανής αλυσίδας σε κάποιο υποσύνολο του χώρου καταστάσεών της. Συγκεκριμένα, έχουμε τα ακόλουθα δυο αποτελέσματα.

Θεώρημα 10.8 (Παράθεση ανεξάρτητων αντιστρέψιμων διαδικασιών) Έστω οι ανεξάρτητες αντιστρέψιμες στοχαστικές διαδικασίες $\{X_1(t)\}$ και $\{X_2(t)\}$. Τότε η παράθεσή τους, $\{(X_1(t), X_2(t))\}$ είναι αντιστρέψιμη. ◊

Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	1	$T_{01} \sim \text{Exp}(\lambda)$
$n \geq c$	$n + 1$	$T_{n,n+1} \sim \text{Exp}(\lambda)$
	$n - 1$	$T_{n,n-1} \sim \text{Exp}(\mu_1 + \mu_2 + \dots + \mu_c)$
$1 \leq n < c$	$n + 1$	$T_{n,n+1} \sim \text{Exp}(\lambda)$
	$n - 1$;

Πίνακας 10.1: Πίνακας μεταβάσεων M/M/c ουράς με ετερογενείς υπηρέτες.

Θεώρημα 10.9 (Περικοπή αντιστρέψιμης Μαρκοβιανής αλυσίδας) Έστω $\{X(t)\}$ Μαρκοβιανή αλυσίδα συνεχούς χρόνου με χώρο καταστάσεων \mathcal{S} και ρυθμούς $q_{ij}, i \neq j$. Έστω, επίσης, $\mathcal{E} \subseteq \mathcal{S}$. Η περικοπή της $\{X(t)\}$ στο \mathcal{E} ορίζεται ως η Μαρκοβιανή αλυσίδα $\{X_{\mathcal{E}}(t)\}$ με χώρο καταστάσεων \mathcal{E} και τους ίδιους ρυθμούς με την αρχική για $i, j \in \mathcal{E}$. Αν η $\{X(t)\}$ είναι αδιαχώριστη και αντιστρέψιμη με κατανομή ισορροπίας (p_j) και το \mathcal{E} είναι τέτοιο ώστε η $\{X_{\mathcal{E}}(t)\}$ να είναι αδιαχώριστη, τότε η $\{X_{\mathcal{E}}(t)\}$ είναι αντιστρέψιμη με κατανομή ισορροπίας $(p_{\mathcal{E},j})$ με

$$p_{\mathcal{E},j} = \frac{p_j}{\sum_{i \in \mathcal{E}} p_i}, \quad j \in \mathcal{E}. \quad (10.4)$$

◊

Η ισχύς του θεωρήματος 10.8 είναι προφανής, χρησιμοποιώντας μόνο τον ορισμό της αντιστρεψιμότητας. Για την απόδειξη του θεωρήματος 10.9, αρκεί να παρατηρήσουμε ότι αφού η αλυσίδα $\{X(t)\}$ είναι αντιστρέψιμη, η κατανομή ισορροπίας της ικανοποιεί τις εξισώσεις λεπτομερούς ισορροπίας (10.3) για κάθε καταστάσεις $i, j \in \mathcal{S}$ με $i \neq j$, άρα και για κάθε καταστάσεις $i, j \in \mathcal{E}$ με $i \neq j$. Επομένως, αν κανονικοποιηθεί στο \mathcal{E} διαιρώντας τους όρους της με το $\sum_{i \in \mathcal{E}} p_i$, γίνεται συνάρτηση πιθανότητας στο \mathcal{E} που ικανοποιεί τις εξισώσεις ισορροπίας της περικεκομμένης αλυσίδας $\{X_{\mathcal{E}}(t)\}$, οπότε έχουμε ότι η $\{X_{\mathcal{E}}(t)\}$ είναι αντιστρέψιμη με κατανομή ισορροπίας την $(p_{\mathcal{E},j})$ που δίνεται από την (10.4).

10.3 Η M/M/c ουρά με ετερογενείς υπηρέτες

Θεωρούμε την M/M/c ουρά με ετερογενείς υπηρέτες. Το μοντέλο αυτό είναι παρόμοιο με την M/M/c ουρά, με μόνη διαφορά ότι κάθε υπηρέτης έχει διαφορετικό ρυθμό εξυπηρέτησης. Πιο συγκεκριμένα, οι αφίξεις συμβαίνουν σύμφωνα με μια διαδικασία Poisson, υπάρχουν c υπηρέτες και οι χρόνοι εξυπηρέτησης των πελατών που εξυπηρετούνται από τον υπηρέτη i είναι $\text{Exp}(\mu_i), i = 1, 2, \dots, c$. Η χωρητικότητα του συστήματος είναι απεριόριστη. Αν εξετάσουμε τη στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα τότε βλέπουμε ότι για τις μεταβάσεις της έχουμε τον πίνακα 10.1.

Πράγματι, όταν το πλήθος των πελατών στο σύστημα είναι μικρότερο από c , δηλαδή όταν δεν είναι όλοι οι υπηρέτες απασχολημένοι, χρειάζεται να ξέρουμε ποιοι υπηρέτες είναι απασχολημένοι για να προσδιορίσουμε τον χρόνο ως την επόμενη αναχώρηση πελάτη. Επομένως, για να έχουμε μια Μαρκοβιανή περιγραφή του συστήματος, χρειάζεται εκτός από τον αριθμό των πελατών $Q(t)$ στο σύστημα να καταγράφουμε και το σύνολο των υπηρετών που είναι απασχολημένοι κάθε στιγμή t . Βέβαια, για $Q(t) = 0$ το σύνολο των απασχολημένων υπηρετών είναι το κενό ενώ για $Q(t) \geq c$ είναι το σύνολο όλων των υπηρετών, οπότε η καταγραφή του συνόλου των απασχολημένων υπηρετών είναι αναγκαία μόνο όταν $1 \leq Q(t) < c$.

Για να συγκεκριμενοποιήσουμε τη μελέτη μας και να απλοποιήσουμε την παρουσίαση των βασικών ιδεών, θα περιοριστούμε παρακάτω στην περίπτωση των δύο ετερογενών υπηρετών.

Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	(1, 1)	$T_{0,(1,1)} \sim \text{Exp}(\lambda p)$
	(1, 2)	$T_{0,(1,2)} \sim \text{Exp}(\lambda q)$
(1, 1)	2	$T_{(1,1),2} \sim \text{Exp}(\lambda)$
	0	$T_{(1,1),0} \sim \text{Exp}(\mu_1)$
(1, 2)	2	$T_{(1,2),2} \sim \text{Exp}(\lambda)$
	0	$T_{(1,2),0} \sim \text{Exp}(\mu_2)$
2	3	$T_{23} \sim \text{Exp}(\lambda)$
	(1, 1)	$T_{2,(1,1)} \sim \text{Exp}(\mu_2)$
	(1, 2)	$T_{2,(1,2)} \sim \text{Exp}(\mu_1)$
$n \geq 3$	$n + 1$	$T_{n,n+1} \sim \text{Exp}(\lambda)$
	$n - 1$	$T_{n,n-1} \sim \text{Exp}(\mu_1 + \mu_2)$

Πίνακας 10.2: Πίνακας μεταβάσεων M/M/2 ουράς με ετερογενείς υπηρέτες.

10.3.1 Η M/M/2 ουρά με ετερογενείς υπηρέτες

Θεωρούμε την M/M/2 ουρά με ετερογενείς υπηρέτες και υποθέτουμε ότι ένας πελάτης που βρίσκει το σύστημα κενό διαλέγει τον υπηρέτη 1 ή τον υπηρέτη 2 τυχαία, με πιθανότητες p και $q = 1 - p$ αντίστοιχα. Έστω $\{Q(t)\}$ η στοχαστική διαδικασία που καταγράφει τον αριθμό των πελατών. Για να έχουμε μια Μαρκοβιανή περιγραφή του συστήματος, σύμφωνα με τα παραπάνω, πρέπει να καταγράψουμε ποιος υπηρέτης είναι απασχολημένος στην περίπτωση που $Q(t) = 1$. Έστω, λοιπόν, $C(t)$ ο υπηρέτης που είναι απασχολημένος τη στιγμή t . Η $\{(Q(t), C(t))\}$ είναι μια στοχαστική διαδικασία με χώρο καταστάσεων $\mathcal{S} = \{0, (1, 1), (1, 2), 2, 3, \dots\}$. Για τις μεταβάσεις της έχουμε τον πίνακα 10.2.

Η Μαρκοβιανή αλυσίδα συνεχούς χρόνου $\{(Q(t), C(t))\}$ είναι μορφής γέννησης-θανάτου για τις καταστάσεις $2, 3, \dots$. Άρα ο μοναδικός απλός κύκλος καταστάσεων με μήκος μεγαλύτερο του 2 είναι ο κύκλος καταστάσεων $0 \rightarrow (1, 1) \rightarrow 2 \rightarrow (1, 2) \rightarrow 0$. Από το κριτήριο του Κολμογορον, η Μαρκοβιανή αλυσίδα συνεχούς χρόνου θα είναι αντιστρέψιμη αν και μόνο αν το γινόμενο των ρυθμών μετάβασης είναι το ίδιο και για τις δυο φορές διαγραφής του κύκλου, δηλαδή αν

$$q_{0,(1,1)}q_{(1,1),2}q_{2,(1,2)}q_{(1,2),0} = q_{0,(1,2)}q_{(1,2),2}q_{2,(1,1)}q_{(1,1),0}.$$

Αυτή η συνθήκη ικανοποιείται μόνο αν $p = q$, δηλαδή αν οι πελάτες που βρίσκουν το σύστημα κενό επιλέγουν έναν υπηρέτη ισοπίθανα για να εξυπηρετηθούν. Στην περίπτωση αυτή διαλέγουμε μια οποιαδήποτε κατάσταση ως κατάσταση αναφοράς και συνδέουμε τις υπόλοιπες καταστάσεις με αυτή μέσω μονοπατιών. Ας θεωρήσουμε την κατάσταση 0 ως την κατάσταση αναφοράς. Τότε συνάγουμε εύκολα την κατανομή ισορροπίας ως

$$\begin{aligned} p_{(1,1)} &= \frac{\lambda}{2\mu_1} p_0, \\ p_{(1,2)} &= \frac{\lambda}{2\mu_2} p_0, \\ p_n &= \frac{\lambda^2}{2\mu_1\mu_2} \left(\frac{\lambda}{\mu_1 + \mu_2} \right)^{n-2} p_0, \quad n \geq 2. \end{aligned}$$

Η πιθανότητα p_0 υπολογίζεται από την εξίσωση κανονικοποίησης

$$p_0 + p_{(1,1)} + p_{(1,2)} + \sum_{n=2}^{\infty} p_n = 1$$

$$\Leftrightarrow p_0 \left(1 + \frac{\lambda}{2\mu_1} + \frac{\lambda}{2\mu_2} + \sum_{n=2}^{\infty} \frac{\lambda^2}{2\mu_1\mu_2} \left(\frac{\lambda}{\mu_1 + \mu_2} \right)^{n-2} \right) = 1.$$

Για να είναι το σύστημα ευσταθές θα πρέπει η άπειρη σειρά να συγκλίνει δηλαδή $\lambda < \mu_1 + \mu_2$, που εκφράζει και πάλι τη φυσική συνθήκη ο ρυθμός αφίξεων να είναι μικρότερος του μέγιστου ρυθμού εξυπηρέτησης που μπορεί να παρέχει το σύστημα.

Ο υπολογισμός της κατανομής ισορροπίας ολοκληρώνεται χρησιμοποιώντας τη γεωμετρική σειρά (3.5), οπότε παίρνουμε

$$p_0 = \left(1 + \frac{\lambda}{2\mu_1} + \frac{\lambda}{2\mu_2} + \frac{\lambda^2}{2\mu_1\mu_2} \cdot \frac{\mu_1 + \mu_2}{\mu_1 + \mu_2 - \lambda} \right)^{-1}.$$

Στην περίπτωση που $p \neq q$, η Μαρκοβιανή αλυσίδα που περιγράφει την κατάσταση του συστήματος δεν είναι αντιστρέψιμη. Όμως, χρησιμοποιώντας τις εξισώσεις γενικευμένης ισορροπίας έχουμε ότι

$$p_n = \left(\frac{\lambda}{\mu_1 + \mu_2} \right)^{n-2} p_2, \quad n \geq 2.$$

Επομένως, μένει να προσδιορίσουμε μόνο τις πιθανότητες $p_0, p_{(1,1)}, p_{(1,2)}$ και p_2 . Αυτό γίνεται, λύνοντας το σύστημα των εξισώσεων ισορροπίας που αντιστοιχούν στις $p_0, p_{(1,1)}, p_{(1,2)}$ μαζί με την εξίσωση κανονικοποίησης. Έχουμε, επομένως να λύσουμε το σύστημα

$$\begin{aligned} \lambda p_0 &= \mu_1 p_{(1,1)} + \mu_2 p_{(1,2)}, \\ (\lambda + \mu_1) p_{(1,1)} &= \lambda p p_0 + \mu_2 p_2, \\ (\lambda + \mu_2) p_{(1,2)} &= \lambda q p_0 + \mu_1 p_2, \\ p_0 + p_{(1,1)} + p_{(1,2)} + \frac{p_2}{1 - \frac{\lambda}{\mu_1 + \mu_2}} &= 1. \end{aligned}$$

10.4 Η M/M/1 ουρά με ανενεργές περιόδους αφίξεων-εξυπηρέτησεων

Θεωρούμε την M/M/1 ουρά με ρυθμό αφίξεων λ , ρυθμό εξυπηρέτησεων μ , 1 υπηρέτη και απεριόριστη χωρητικότητα υπό την πειθαρχία ουράς FCFS. Ο υπηρέτης εναλλάσσεται μεταξύ ενεργών και νεκρών περιόδων. Κατά τη διάρκεια των ενεργών περιόδων υποδέχεται νέες αφίξεις και παρέχει εξυπηρέτηση, ενώ κατά τη διάρκεια των ανενεργών (νεκρών) περιόδων δεν δέχεται αφίξεις, ούτε παρέχει εξυπηρέτηση, αλλά η κατάσταση του συστήματος παραμένει «παγωμένη». Οι διάρκειες των ενεργών και των νεκρών περιόδων είναι όλες ανεξάρτητες μεταξύ τους. Επιπλέον υποθέτουμε ότι οι διάρκειες των ενεργών περιόδων έχουν την κατανομή $\text{Exp}(\zeta)$, ενώ οι διάρκειες των νεκρών περιόδων έχουν την κατανομή $\text{Exp}(\theta)$. Το σύστημα αυτό αναφέρεται ως η M/M/1 ουρά με ανενεργές περιόδους αφίξεων-εξυπηρέτησεων και ομοιάζει με την M/M/1 ουρά με ενεργές και ανενεργές περιόδους εξυπηρέτησεων που μελετήσαμε στην παράγραφο 9.1. Η διαφορά του συστήματος της παρούσης παραγράφου από το σύστημα της παραγράφου 9.1 είναι ότι στο παρόν σύστημα σταματάνε και οι αφίξεις κατά τη διάρκεια των ανενεργών περιόδων και γι' αυτό μιλάμε για ανενεργές περιόδους αφίξεων-εξυπηρέτησεων ή για νεκρές περιόδους.

Συμβολίζουμε με $Q(t)$ το πλήθος των πελατών στο σύστημα τη στιγμή t και με $I(t)$ την κατάσταση του συστήματος (δηλαδή $I(t) = 1$ αν το σύστημα είναι σε ενεργή περίοδο, ενώ $I(t) = 0$ αν είναι σε νεκρή περίοδο). Για τις μεταβάσεις της $\{(Q(t), I(t))\}$ έχουμε τότε τον πίνακα 10.3.

Κατάσταση	Επόμενη κατάσταση	Χρόνος
$(n, 0), n \geq 0$	$(n, 1)$	$T_{(n,0),(n,1)} \sim \text{Exp}(\theta)$
$(0, 1)$	$(1, 1)$	$T_{(0,1),(1,1)} \sim \text{Exp}(\lambda)$
	$(0, 0)$	$T_{(0,1),(0,0)} \sim \text{Exp}(\zeta)$
$(n, 1), n \geq 1$	$(n + 1, 1)$	$T_{(n,1),(n+1,1)} \sim \text{Exp}(\lambda)$
	$(n - 1, 1)$	$T_{(n,1),(n-1,1)} \sim \text{Exp}(\mu)$
	$(n, 0)$	$T_{(n,1),(n,0)} \sim \text{Exp}(\zeta)$

Πίνακας 10.3: Πίνακας μεταβάσεων M/M/1 ουράς με νεκρές περιόδους.

Αφού όλοι οι χρόνοι είναι εκθετικοί, έχουμε ότι η $\{(Q(t), I(t))\}$ είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου. Επιπλέον, παρατηρούμε ότι το διάγραμμα των ρυθμών μετάβασής της είναι αμφίδρομο δέντρο και επομένως από το κριτήριο του Kolmogorov έχουμε ότι είναι και αντιστρέψιμη. Επομένως η κατανομή ισορροπίας της μπορεί να βρεθεί επιλέγοντας μια κατάσταση αναφοράς και συνδέοντας όλες τις άλλες καταστάσεις με μονοπάτια με αυτήν.

Έστω $(p_{(n,i)})$ η κατανομή ισορροπίας της. Εφαρμόζοντας το δεύτερο βήμα του αλγορίθμου 10.7 και επιλέγοντας ως κατάσταση αναφοράς την $(0, 0)$ έχουμε ότι κάθε κατάσταση $(n, 0), n \geq 0$, συνδέεται με την $(0, 0)$ μέσω του μονοπατιού

$$(0, 0) \rightarrow (0, 1) \rightarrow (1, 1) \rightarrow (2, 1) \rightarrow (3, 1) \rightarrow \dots \rightarrow (n, 1) \rightarrow (n, 0).$$

Επομένως

$$\begin{aligned} p_{(n,0)} &= \frac{q_{(0,0),(0,1)}q_{(0,1),(1,1)}q_{(1,1),(2,1)}q_{(2,1),(3,1)} \cdots q_{(n-1,1),(n,1)}q_{(n,1),(n,0)}}{q_{(0,1),(0,0)}q_{(1,1),(0,1)}q_{(2,1),(1,1)}q_{(3,1),(2,1)} \cdots q_{(n,1),(n-1,1)}q_{(n,0),(n,1)}} p_{(0,0)} \\ &= \frac{\theta \lambda^n \zeta}{\zeta \mu^n \theta} p_{(0,0)} = \rho^n p_{(0,0)}, \quad n \geq 0, \end{aligned}$$

όπου $\rho = \frac{\lambda}{\mu}$. Επίσης, από τις εξισώσεις λεπτομερούς ισορροπίας, έχουμε ότι $p_{(n,1)}q_{(n,1),(n,0)} = p_{(n,0)}q_{(n,0),(n,1)}$, οπότε

$$p_{(n,1)} = \frac{p_{(n,0)}q_{(n,0),(n,1)}}{q_{(n,1),(n,0)}} = \frac{\theta}{\zeta} p_{(n,0)} = \frac{\theta}{\zeta} \rho^n p_{(0,0)}, \quad n \geq 0.$$

Τέλος, από την εξίσωση κανονικοποίησης

$$\sum_{n=0}^{\infty} (p_{(n,0)} + p_{(n,1)}) = 1,$$

και χρησιμοποιώντας τη γεωμετρική σειρά (3.5) συνάγουμε ότι

$$p_{(0,0)} = \frac{\zeta}{\theta + \zeta} (1 - \rho).$$

Οπότε, τελικά η κατανομή ισορροπίας δίνεται από τις σχέσεις

$$\begin{aligned} p_{(n,0)} &= \frac{\zeta}{\theta + \zeta} (1 - \rho) \rho^n, \quad n \geq 0, \\ p_{(n,1)} &= \frac{\theta}{\theta + \zeta} (1 - \rho) \rho^n, \quad n \geq 0. \end{aligned}$$

Παρατηρούμε ότι σε κατάσταση ισορροπίας, οι τυχαίες μεταβλητές $I(t)$ και $Q(t)$ που καταγράφουν την κατάσταση του συστήματος (σε ενεργή περίοδο ή νεκρή περίοδο) και τον αριθμό των πελατών σε αυτό είναι ανεξάρτητες. Πράγματι, η περιθώρια που αντιστοιχεί στην $I(t)$ είναι η $(\text{Pr}[I = 0], \text{Pr}[I = 1]) = \left(\frac{\zeta}{\theta + \zeta}, \frac{\theta}{\theta + \zeta}\right)$, ενώ η περιθώρια που αντιστοιχεί στην $Q(t)$ είναι η $(\text{Pr}[Q(t) = n] : n \geq 0) = ((1 - \rho) \rho^n : n \geq 0)$.

10.5 Δυο παράλληλες M/M/1 ουρές με κοινό χώρο αναμονής

Θεωρούμε δυο συστήματα M/M/1 με ρυθμούς αφίξεων λ_1 και λ_2 , αντίστοιχα, και ρυθμούς εξυπηρέτησης μ_1 και μ_2 , αντίστοιχα, για τα οποία θεωρούμε ότι οι διαδικασίες των αφίξεων και οι διάρκειες των εξυπηρέτησεων είναι ανεξάρτητες. Επιπλέον, υποθέτουμε ότι τα δυο συστήματα μοιράζονται έναν κοινό χώρο αναμονής μέγθους k . Αυτό σημαίνει ότι μια άφιξη που βρίσκει τον υπηρέτη του αντίστοιχου συστήματος απασχολημένο και τον κοινό χώρο αναμονής γεμάτο αναχωρεί άμεσα.

Έστω $Q_1(t)$ και $Q_2(t)$ το πλήθος των πελατών στα δυο συστήματα. Ενδιαφερόμαστε για την κατανομή ισορροπίας $(p_{(i,j)})$ της στοχαστικής διαδικασίας $\{(Q_1(t), Q_2(t))\}$.

Έστω $\{\bar{Q}_1(t)\}$ και $\{\bar{Q}_2(t)\}$ ανεξάρτητες στοχαστικές διαδικασίες που περιγράφουν το πλήθος των πελατών σε δυο συστήματα, όπου το σύστημα i είναι τύπου M/M/1 με ρυθμό αφίξεων λ_i και ρυθμό εξυπηρέτησης μ_i , για $i = 1, 2$. Η στοχαστική διαδικασία $\{(Q_1(t), Q_2(t))\}$ μπορεί να θεωρηθεί ως η περικοπή της διαδικασίας $\{(\bar{Q}_1(t), \bar{Q}_2(t))\}$ με χώρο καταστάσεων το $\mathcal{S} = \mathbb{N}_0^2$, στο σύνολο

$$\mathcal{E} = \{(i, j) \in \mathbb{N}_0^2 : i, j \geq 0 \text{ και } \max(i-1, 0) + \max(j-1, 0) \leq k\}. \quad (10.5)$$

Η κατανομή ισορροπίας $\bar{p}_{(i,j)}$ της $\{(\bar{Q}_1(t), \bar{Q}_2(t))\}$ βρίσκεται άμεσα λόγω της ανεξαρτησίας των συνιστωσών της, χρησιμοποιώντας τον τύπο (6.10) για την κατανομή ισορροπίας μιας M/M/1 ουράς. Επομένως, έχουμε ότι

$$\bar{p}_{(i,j)} = (1 - \rho_1)\rho_1^i(1 - \rho_2)\rho_2^j, \quad (i, j) \in \mathbb{N}_0^2. \quad (10.6)$$

Επιπλέον, οι διαδικασίες $\{\bar{Q}_1(t)\}$ και $\{\bar{Q}_2(t)\}$ είναι αντιστρέψιμες, ως αλυσίδες τύπου γέννησης-θανάτου, από το κριτήριο αντιστρεψιμότητας Kolmogorov (βλέπε θεώρημα 10.6 και τα σχόλια που το ακολουθούν). Εφαρμόζοντας το θεώρημα 10.8 συμπεραίνουμε ότι και η $\{(\bar{Q}_1(t), \bar{Q}_2(t))\}$ είναι αντιστρέψιμη. Αλλά τότε η περικοπή της σε οποιοδήποτε υποσύνολο του χώρου καταστάσεων της, άρα και στο \mathcal{E} που δίνεται από την (10.5), έχει ως κατανομή ισορροπίας την κανονικοποιημένη κατανομή ισορροπίας της αρχικής (από το θεώρημα 10.9). Επομένως, χρησιμοποιώντας τις (10.4) και (10.6) παίρνουμε ότι

$$p_{(i,j)} = \frac{\bar{p}_{(i,j)}}{\sum_{(i',j') \in \mathcal{E}} \bar{p}_{(i',j')}} = \frac{\rho_1^i \rho_2^j}{\sum_{(i',j') \in \mathcal{E}} \rho_1^{i'} \rho_2^{j'}}, \quad (i, j) \in \mathcal{E}. \quad (10.7)$$

10.6 Η διαδικασία αναχωρήσεων σε απλές Μαρκοβιανές ουρές με Poisson διαδικασία αφίξεων

Η διαπίστωση ότι μια Μαρκοβιανή ουρά είναι αντιστρέψιμη επιτρέπει τον εύκολο προσδιορισμό της στάσιμης κατανομής της, αφού αυτή ικανοποιεί το ισχυρότερο σύστημα των εξισώσεων λεπτομερούς ισορροπίας. Πέραν, όμως, της υπολογιστικής ευκολίας που συνεπάγεται η αντιστρεψιμότητα, οδηγεί και σε ποιοτικά συμπεράσματα σχετικά με τη συμπεριφορά μιας ουράς. Το σημαντικότερο από αυτά είναι το Θεώρημα του Burke.

Θεώρημα 10.10 (Θεώρημα του Burke) Η διαδικασία των αναχωρήσεων $\{D(t)\}$ σε μια στάσιμη απλή Μαρκοβιανή ουρά με Poisson διαδικασία αφίξεων με ρυθμό λ , είναι επίσης διαδικασία Poisson με τον ίδιο ρυθμό. Επιπλέον για κάθε επιλεγμένη χρονική στιγμή t , η διαδικασία του πλήθους των πελατών από τη στιγμή t και μετά, δηλαδή η $\{Q(s) : s \geq t\}$, είναι ανεξάρτητη από τη διαδικασία των αναχωρήσεων πριν τη στιγμή t , δηλαδή από την $\{D(s) : s < t\}$. Αυτό ισχύει και για την περίπτωση που η t είναι στιγμή αναχώρησης πελάτη. \diamond

Η ισχύς του θεωρήματος αυτού φαίνεται άμεσα, αρκεί κανείς να συγκρίνει τη διαδικασία του αριθμού των πελατών $\{Q(t)\}$ και την αντίστροφή της $\{\hat{Q}(t)\}$. Πράγματι η διαδικασία αφίξεων της $\{Q(t)\}$ είναι Poisson και επιπλέον η εξέλιξη της διαδικασίας αφίξεων στο $(-t, \infty)$ είναι ανεξάρτητη της $Q(-t)$ για κάθε t . Επομένως, αφού η $\{Q(t)\}$ είναι αντιστρέψιμη, τα ίδια ισχύουν και για την $\{\hat{Q}(t)\}$, δηλαδή η διαδικασία των αφίξεων της $\{\hat{Q}(t)\}$ είναι Poisson και επιπλέον η εξέλιξη της διαδικασίας αφίξεων στο $(-t, \infty)$ είναι ανεξάρτητη της $\hat{Q}(-t)$.

για κάθε t . Μπορούμε τώρα να μεταφράσουμε αυτές τις δυο ιδιότητες με όρους της αρχικής διαδικασίας $\{Q(t)\}$ και να συναγάγουμε το συμπέρασμα. Πράγματι κάθε άφιξη στην $\{\hat{Q}(t)\}$ αντιστοιχεί σε αναχώρηση στην $\{Q(t)\}$. Επομένως η διαδικασία των αφίξεων της $\{\hat{Q}(t)\}$ είναι στοχαστικά ισοδύναμη με τη διαδικασία των αναχωρήσεων της $\{Q(t)\}$. Επομένως η διαδικασία των αναχωρήσεων της $\{Q(t)\}$ είναι Poisson με τον ρυθμό λ . Επίσης αφού η διαδικασία αφίξεων της $\{\hat{Q}(t)\}$ στο $(-t, \infty)$ είναι ανεξάρτητη της $\hat{Q}(-t)$ για κάποιο t , έχουμε ότι η διαδικασία αναχωρήσεων της $\{Q(t)\}$ στο $(-\infty, t)$ είναι ανεξάρτητη της $Q(t)$.

Με βάση το θεώρημα του Burke μπορούμε να υπολογίσουμε την κατανομή ισορροπίας του αριθμού των πελατών σε N το πλήθος $M/M/1$ ουρές τοποθετημένες σε σειρά, όταν η πρώτη ουρά έχει διαδικασία αφίξεων Poisson. Συγκεκριμένα έχουμε το ακόλουθο αποτέλεσμα.

Θεώρημα 10.11 (Κατανομή ισορροπίας σειριακού δικτύου $M/M/1$ ουρών) Έστω N συστήματα εξυπηρέτησης τοποθετημένα σε σειρά. Υποθέτουμε ότι τα συστήματα έχουν 1 υπηρέτη το καθένα και απεριόριστη χωρητικότητα, η διαδικασία των αφίξεων στο πρώτο σύστημα είναι Poisson με ρυθμό λ και οι χρόνοι εξυπηρέτησης είναι $\text{Exp}(\mu_i)$ στο σύστημα i , για $i = 1, 2, \dots, N$. Επιπλέον, η πειθαρχία ουράς σε κάθε σύστημα είναι η FCFS και ένας πελάτης αναχωρώντας από ένα σύστημα πηγαίνει στο επόμενο και όταν αναχωρήσει από το τελευταίο αναχωρεί από το σύστημα. Έστω $\{\mathbf{Q}(t) = (Q_1(t), Q_2(t), \dots, Q_N(t))\}$ η διαδικασία που περιγράφει τους αριθμούς των πελατών στα διάφορα συστήματα του δικτύου. Τότε, η Μαρκοβιανή αλυσίδα $\{\mathbf{Q}(t)\}$ είναι θετικά επαναληπτική (δηλαδή το δίκτυο είναι ευσταθές) αν και μόνο αν $\rho_i = \frac{\lambda}{\mu_i} < 1$, για όλα τα $i = 1, 2, \dots, N$. Η κατανομή ισορροπίας της $\{\mathbf{Q}(t)\}$ στην περίπτωση αυτή είναι

$$p(\mathbf{n}) = p(n_1, n_2, \dots, n_N) = \prod_{i=1}^N (1 - \rho_i) \rho_i^{n_i}, \quad \mathbf{n} = (n_1, n_2, \dots, n_N) \in \mathbb{N}_0^N. \quad (10.8)$$

◊

Η απόδειξη του θεωρήματος αυτού γίνεται με επανειλημμένη χρήση του θεωρήματος του Burke. Πράγματι, το πρώτο σύστημα του σειριακού δικτύου είναι μια $M/M/1$ ουρά, οπότε εφαρμόζοντας το θεώρημα του Burke έχουμε ότι οι αναχωρήσεις από αυτό συνιστούν μια διαδικασία Poisson με ρυθμό λ . Αλλά η διαδικασία των αναχωρήσεων από το πρώτο σύστημα ταυτίζεται με τη διαδικασία των αφίξεων στο δεύτερο σύστημα. Οπότε και το δεύτερο σύστημα είναι μια $M/M/1$ ουρά κ.ο.κ. Επομένως, όλα τα συστήματα είναι $M/M/1$ ουρές. Ειδικότερα το i -οστό σύστημα έχει ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ_i και επομένως είναι ευσταθές αν και μόνο αν $\rho_i < 1$ και η κατανομή ισορροπίας του είναι $(p_i(n_i) : n_i \geq 0)$ με

$$p_i(n_i) = (1 - \rho_i) \rho_i^{n_i}, \quad n_i \geq 0. \quad (10.9)$$

Εξετάζοντας τώρα το σύστημα μια χρονική στιγμή t , έχουμε από το θεώρημα του Burke ότι ο αριθμός των πελατών στο πρώτο σύστημα τη στιγμή t , $Q_1(t)$, είναι ανεξάρτητος από τη διαδικασία των αναχωρήσεων από το σύστημα αυτό πριν τη στιγμή t . Αλλά οι καταστάσεις των επόμενων συστημάτων τη στιγμή t καθορίζονται από τη διαδικασία των αναχωρήσεων του πρώτου συστήματος πριν τη στιγμή t και από χρόνους εξυπηρέτησης στα άλλα συστήματα. Επομένως η $(Q_2(t), Q_3(t), \dots, Q_N(t))$ είναι ανεξάρτητη από την $Q_1(t)$ και έχουμε για την από κοινού κατανομή ισορροπίας ότι θα ισούται με το γινόμενο των περιθωρίων. Η ίδια συλλογιστική ισχύει και για τα επόμενα συστήματα. Δηλαδή η $Q_2(t)$ είναι ανεξάρτητη από την $(Q_3(t), Q_4(t), \dots, Q_N(t))$ κ.ο.κ. Οπότε τελικά έχουμε ότι

$$p(\mathbf{n}) = p(n_1, n_2, \dots, n_N) = \prod_{i=1}^N p_i(n_i), \quad \mathbf{n} = (n_1, n_2, \dots, n_N) \in \mathbb{N}_0^N$$

και χρησιμοποιώντας την (10.9) παίρνουμε την (10.8). Το θεώρημα 10.11 μπορεί να γενικευτεί για την εύρεση της κατανομής ισορροπίας των αριθμών των πελατών στους διάφορους σταθμούς ενός δικτύου N συστημάτων εξυπηρέτησης τοποθετημένων σε σειρά, εφόσον το πρώτο σύστημα έχει διαδικασία αφίξεων Poisson και επιπλέον όλα τα συστήματα λειτουργούν σαν απλές Μαρκοβιανές ουρές αν τα δούμε μεμονωμένα.

10.7 Ασκήσεις

Άσκηση 10.1 Βρείτε την κατανομή ισορροπίας του αριθμού των πελατών στο σύστημα και το ποσοστό των χαμένων πελατών σε μια $M/M/2/2$ ουρά με ετερογενείς υπηρετές, οι οποίοι επιλέγονται ισοπίθανα όταν είναι και οι δυο ελεύθεροι. Δίνονται ο ρυθμός αφίξεων λ και οι ρυθμοί εξυπηρέτησης μ_1 και μ_2 των δυο υπηρετών.

Άσκηση 10.2 Να δοθεί ένα παράδειγμα που να δείχνει ότι η σχέση (10.4) για την κατανομή ισορροπίας της περικοπής μιας Μαρκοβιανής αλυσίδας συνεχούς χρόνου δεν ισχύει γενικά όταν η αρχική αλυσίδα δεν είναι αντιστρέψιμη. Με άλλα λόγια η υπόθεση της αντιστρεψιμότητας είναι ουσιαστική για την ισχύ του θεωρήματος 10.9. Επιπλέον, χρησιμοποιήστε το θεώρημα 10.9 για να βρείτε

1. την κατανομή ισορροπίας της $M/M/1/k$ ουράς, βασιζόμενοι στην κατανομή ισορροπίας της $M/M/1$ ουράς, και
2. την κατανομή ισορροπίας της $M/M/k/k$ ουράς, βασιζόμενοι στην κατανομή ισορροπίας της $M/M/\infty$ ουράς.

Άσκηση 10.3 Έστω μια αδιαχώριστη αντιστρέψιμη Μαρκοβιανή αλυσίδα $\{X(t)\}$ με χώρο καταστάσεων \mathcal{S} και ρυθμούς μετάβασης q_{ij} . Έστω, επίσης, ένα γνήσιο μη-κενό υποσύνολο \mathcal{A} του \mathcal{S} . Θεωρούμε τη Μαρκοβιανή αλυσίδα $\{\bar{X}(t)\}$ με χώρο καταστάσεων \mathcal{S} και ρυθμούς μετάβασης

$$\bar{q}_{ij} = \begin{cases} kq_{ij} & \text{αν } i \in \mathcal{A} \text{ και } j \notin \mathcal{A} \\ q_{ij} & \text{διαφορετικά,} \end{cases}$$

όπου $k > 0$ δοσμένη σταθερά. Δείξτε ότι η αλυσίδα $\{\bar{X}(t)\}$ είναι και αυτή αντιστρέψιμη και βρείτε πώς συνδέεται η κατανομή ισορροπίας της με την κατανομή ισορροπίας της αρχικής.

Άσκηση 10.4 Χρησιμοποιήστε το θεώρημα του Burke για να αποδείξετε ότι ο χρόνος παραμονής ενός πελάτη σε μια $M/M/1$ ουρά σε κατάσταση ισορροπίας είναι ανεξάρτητος από τη διαδικασία των αναχωρήσεων πριν από την αναχώρηση του συγκεκριμένου πελάτη. Ακολούθως, αποδείξτε ότι αν έχουμε δυο FCFS συστήματα εξυπηρέτησης σε σειρά, όπου το πρώτο είναι τύπου $M/M/1$, τότε οι χρόνοι παραμονής του ίδιου πελάτη στα δυο συστήματα είναι ανεξάρτητες τυχαίες μεταβλητές.

Άσκηση 10.5 [Βασική ιδέα αρχικής απόδειξης του θεωρήματος Burke] Θεωρούμε μια απλή Μαρκοβιανή ουρά με ρυθμούς αφίξεων $\lambda_j = \lambda$, $j = 0, 1, 2, \dots$ και ρυθμούς εξυπηρέτησης μ_j , $j = 1, 2, \dots$ με κατανομή ισορροπίας (p_j) . Θεωρούμε ότι ο αρχικός αριθμός των πελατών στο σύστημα $Q(0)$ έχει την κατανομή (p_j) και επομένως η διαδικασία του αριθμού των πελατών στο σύστημα $\{Q(t)\}$ είναι στάσιμη. Έστω τ η στιγμή της πρώτης αναχώρησης από το σύστημα και $Q^+ = Q(\tau^+)$ ο αριθμός των πελατών στο σύστημα αμέσως μετά από την πρώτη αναχώρηση.

Έστω

$$P(z) = \sum_{j=0}^{\infty} p_j z^j$$

η πιθανογεννήτρια της κατανομής ισορροπίας του αριθμού των πελατών στο σύστημα και

$$\bar{P}_j(s, z) = E[e^{-s\tau} z^{Q^+} | Q(0) = j], \quad j \geq 0,$$

ο «μικτός» δεσμευμένος μετασχηματισμός της (τ, Q^+) (μετασχηματισμός Laplace-Stieltjes ως προς την τ και πιθανογεννήτρια ως προς την Q^+), δεδομένου ότι $Q(0) = j$. Έστω, επίσης, $\bar{P}(s, z)$ ο αντίστοιχος «αδέσμευτος» μετασχηματισμός:

$$\bar{P}(s, z) = E[e^{-s\tau} z^{Q^+}] = \sum_{j=0}^{\infty} p_j \bar{P}_j(s, z).$$

1. Δεσμεύοντας στο κατά πόσον η πρώτη μετάβαση είναι άφιξη ή αναχώρηση πελάτη, δείξτε ότι οι μετασχηματισμοί $\tilde{P}_j(s, z)$ ικανοποιούν το αναδρομικό σχήμα

$$(\lambda + \mu_j + s)\tilde{P}_j(s, z) = \mu_j z^{j-1} + \lambda \tilde{P}_{j+1}(s, z), \quad j \geq 0.$$

2. Αποδείξτε ότι

$$\tilde{P}(s, z) = \frac{\lambda}{\lambda + s} P(z)$$

και βρείτε την από κοινού συνάρτηση κατανομής των τ και Q^+ .

3. Χρησιμοποιώντας ανάλογα επιχειρήματα για την εξάρτηση της ακολουθίας των ενδιάμεσων χρόνων μεταξύ αναχωρήσεων $\{\tau_n\}$ και των αντίστοιχων αριθμών πελατών στο σύστημα αμέσως μετά, $\{Q_n^+\}$, αιτιολογήστε ότι η διαδικασία των αναχωρήσεων είναι Poisson και δείτε κατά πόσον μπορείτε να δικαιολογήσετε και τα υπόλοιπα συμπεράσματα του θεωρήματος Burke.

10.8 Σχόλια

Από τα κλασικά εισαγωγικά συγγράμματα στη Θεωρία Ουρών Αναμονής, εκείνο που αφιερώνει περισσότερο χώρο στο θέμα της αντιστρεψιμότητας και των εφαρμογών της είναι το βιβλίο του Wolff 1989.

Η αντιστρεψιμότητα χρησιμοποιείται κυρίως για τη μελέτη δικτύων συστημάτων εξυπηρέτησης. Στο θέμα αυτό επικεντρώνεται το βιβλίο του Kelly 1979 που παρουσιάζει πολλές και ενδιαφέρουσες θεωρητικές λεπτομέρειες και εφαρμογές.

Βιβλιογραφία

- [1] R.W. Wolff. *Stochastic Modeling and the Theory of Queues*. Englewood Cliffs, New Jersey: Prentice Hall, 1989. ISBN: 978-0138466923.
- [2] F.P. Kelly. *Reversibility and Stochastic Networks*. Chichester: Wiley, 1979. ISBN: 978-0471276012.

ΑΠΛΑ ΜΑΡΚΟΒΙΑΝΑ ΔΙΚΤΥΑ ΟΥΡΩΝ

Ένα δίκτυο συστημάτων εξυπηρέτησης είναι ένα σύνολο από συστήματα εξυπηρέτησης τα οποία συνδέονται μεταξύ τους, υπό την έννοια ότι οι πελάτες που αναχωρούν από ένα σύστημα μπορεί να πηγαίνουν σε κάποιο άλλο για να συνεχίσουν την εξυπηρέτησή τους και από εκεί σε ένα άλλο κ.ο.κ. Καθώς τα περισσότερα πραγματικά συστήματα εξυπηρέτησης περιλαμβάνουν πολλά στάδια εξυπηρέτησης, η μαθηματική μελέτη των δικτύων συστημάτων εξυπηρέτησης είναι πολύτιμη για τη μελέτη του συνωστισμού σε πραγματικά συστήματα. Στο πλαίσιο του παρόντος συγγράμματος θα γίνει μόνο μια σύντομη εισαγωγή στη θεωρία των δικτύων συστημάτων εξυπηρέτησης.

11.1 Μαρκοβιανά δίκτυα - Απλά Μαρκοβιανά δίκτυα

Για να περιγραφεί η κατάσταση ενός δικτύου συστημάτων εξυπηρέτησης με N σταθμούς πρέπει κατ' ελάχιστον να καταγράφεται ο αριθμός των πελατών στους διάφορους σταθμούς του δικτύου. Έστω ένα δίκτυο με N σταθμούς εξυπηρέτησης. Θα συμβολίζουμε με $Q_i(t)$ το πλήθος των παρόντων πελατών στο σταθμό i του δικτύου τη χρονική στιγμή t . Θέτουμε $\mathbf{Q}(t) = (Q_1(t), Q_2(t), \dots, Q_N(t))$. Η $\{\mathbf{Q}(t)\}$ είναι η βασική στοχαστική διαδικασία των αριθμών πελατών στους διάφορους σταθμούς του δικτύου.

Ορισμός 11.1 (Μαρκοβιανό δίκτυο - Απλό Μαρκοβιανό δίκτυο) Ένα δίκτυο συστημάτων εξυπηρέτησης λέγεται Μαρκοβιανό, αν η στοχαστική διαδικασία του αριθμού των πελατών στους διάφορους σταθμούς του δικτύου, $\{\mathbf{Q}(t)\}$ με $\mathbf{Q}(t) = (Q_1(t), Q_2(t), \dots, Q_N(t))$, είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου.

Αν επιπλέον οι πελάτες κινούνται μεμονωμένα, τότε έχουμε ένα απλό Μαρκοβιανό δίκτυο. ◊

Σε ό,τι ακολουθεί στο κεφάλαιο αυτό, θα εφαρμόσουμε τα αποτελέσματα της θεωρίας Μαρκοβιανών αλυσίδων συνεχούς χρόνου για να μελετήσουμε διάφορα Μαρκοβιανά δίκτυα. Επειδή οι καταστάσεις των Μαρκοβιανών αλυσίδων τέτοιων δικτύων είναι διανύσματα, δεν είναι βολικό να εμφανίζονται ως «δείκτες», όπως κάναμε στα προηγούμενα κεφάλαια, αλλά θα εμφανίζονται μέσα σε παρενθέσεις. Π.χ., θα γράφουμε $q(\mathbf{n}, \mathbf{m})$ αντί για $q_{\mathbf{n}, \mathbf{m}}$ για τους ρυθμούς μετάβασης από μια κατάσταση \mathbf{n} σε μια κατάσταση \mathbf{m} , $p(\mathbf{n})$ αντί για $p_{\mathbf{n}}$ για τις πιθανότητες ισοροπίας κ.ο.κ.

Ένα δίκτυο λέγεται ανοικτό εάν επιτρέπονται εξωτερικές αφίξεις σε έναν ή περισσότερους σταθμούς του και εάν επιτρέπονται εξωτερικές αναχωρήσεις (δηλαδή αναχωρήσεις εκτός δικτύου) από έναν ή περισσότερους σταθμούς του. Διαφορετικά, λέγεται κλειστό και ο αριθμός των πελατών που κυκλοφορεί σε αυτό είναι σταθερός, έστω M .

Ο χώρος καταστάσεων \mathcal{S} της διαδικασίας $\{\mathbf{Q}(t)\}$ που καταγράφει τον αριθμό των πελατών στους διάφορους σταθμούς ενός δικτύου είναι υποσύνολο του \mathbb{N}_0^N . Για κλειστά δίκτυα με συνολικό αριθμό κυκλοφορούντων πελατών M , ο χώρος καταστάσεων είναι υποσύνολο του

$$\{(n_1, n_2, \dots, n_N) : n_i \geq 0, i = 1, 2, \dots, N \text{ και } \sum_{i=1}^N n_i = M\}.$$

Έστω ότι $\mathbf{Q}(t) = \mathbf{n} = (n_1, n_2, \dots, n_N)$ μια χρονική στιγμή t . Έστω επίσης $\mathbf{e}_i = (0, 0, \dots, 0, 1, 0, \dots, 0)$, $i = 1, 2, \dots, N$ τα μοναδιαία διανύσματα του \mathbb{N}_0^N . Σε ένα απλό Μαρκοβιανό δίκτυο η επόμενη μετάβαση μπορεί να είναι:

- $\mathbf{n} \rightarrow \mathbf{n} + \mathbf{e}_i$: εξωτερική άφιξη στον σταθμό i ,
- $\mathbf{n} \rightarrow \mathbf{n} - \mathbf{e}_j$: αναχώρηση από το σταθμό j προς το εξωτερικό του δικτύου (δυνατή μόνο όταν $n_j \geq 1$),
- $\mathbf{n} \rightarrow \mathbf{n} - \mathbf{e}_j + \mathbf{e}_i$: μετάβαση πελάτη από το σταθμό j στον i (δυνατή μόνο όταν $n_j \geq 1$).

Συμβολίζοντας με 0 έναν εικονικό σταθμό που εκφράζει το εξωτερικό του δικτύου, οι αντίστοιχοι ρυθμοί είναι:

- $\lambda_i(\mathbf{n}) = q(\mathbf{n}, \mathbf{n} + \mathbf{e}_i)$: εξωτερικός ρυθμός αφίξεων στον σταθμό i , όταν η κατάσταση του δικτύου είναι \mathbf{n} , $i = 1, 2, \dots, N$,
- $\mu_{j0}(\mathbf{n}) = q(\mathbf{n}, \mathbf{n} - \mathbf{e}_j)$: εξωτερικός ρυθμός αναχωρήσεων από τον σταθμό j , όταν η κατάσταση του δικτύου είναι \mathbf{n} , $j = 1, 2, \dots, N$,
- $\mu_{ji}(\mathbf{n}) = q(\mathbf{n}, \mathbf{n} - \mathbf{e}_j + \mathbf{e}_i)$: ρυθμός μεταβάσεων από τον j στον i , όταν η κατάσταση του δικτύου είναι \mathbf{n} , $i, j = 1, 2, \dots, N$.

Επίσης θέτουμε

- $\mu_j(\mathbf{n}) = \sum_{i=0}^N \mu_{ji}(\mathbf{n})$: συνολικός ρυθμός εξυπηρέτησης στον j , όταν η κατάσταση του δικτύου είναι \mathbf{n} .

Σε ένα ανοικτό απλό Μαρκοβιανό δίκτυο υπάρχουν τουλάχιστον δυο καταστάσεις \mathbf{n}, \mathbf{m} και σταθμοί i, j έτσι ώστε $\lambda_i(\mathbf{n}) > 0$ και $\mu_{j0}(\mathbf{m}) > 0$.

Ένα σύστημα με πολλούς σταθμούς εξυπηρέτησης μπορεί να παρασταθεί ως ανοικτό ή κλειστό δίκτυο, ανάλογα με τι είδους μοντέλο φαίνεται πιο κατάλληλο για τη μελέτη του. Στα συστήματα εξυπηρέτησης που εμφανίζονται στην πράξη ο αριθμός των (δυνατικών) πελατών είναι πάντα πεπερασμένος και επομένως θα μπορούσαν να παρασταθούν ως κλειστά δίκτυα. Όταν, όμως, ο αριθμός των δυνατικών πελατών είναι μεγάλος, μπορούν να παρασταθούν ως ανοικτά δίκτυα και αυτό είναι πολλές φορές εξαιρετικά βολικό από υπολογιστική σκοπιά. Γενικότερα, όπως θα δούμε παρακάτω, τα ανοικτά δίκτυα είναι πιο εύκολα στη μελέτη τους, κατά τον ίδιο τρόπο που οι ιδεατές Μαρκοβιανές ουρές με άπειρο χώρο καταστάσεων έχουν πιο κομψούς τύπους για τα διάφορα μέτρα απόδοσής τους από τις αντίστοιχες ουρές με πεπερασμένο χώρο καταστάσεων. Ίσως το πλέον τυπικό παράδειγμα αυτής της ιδέας να είναι το μοντέλο του Engset που είδαμε στην παράγραφο 6.7 που προσεγγίζεται από την $M/M/c/c$ ουρά όταν το πλήθος των δυνατικών πελατών είναι μεγάλο.

Το κεντρικό υπολογιστικό πρόβλημα στη θεωρία των Μαρκοβιανών δικτύων αφορά την εύρεση της κατανομής ισορροπίας, $(p(\mathbf{n}))$, της στοχαστικής διαδικασίας $\{\mathbf{Q}(t)\}$ του αριθμού των πελατών του δικτύου, καθώς και των περιθωρίων της, $(p_i(n_i))$, $i = 1, 2, \dots, N$.

11.2 Δίκτυα Jackson

Τα δίκτυα Jackson συγκροτούν μια οικογένεια απλών Μαρκοβιανών δικτύων εξυπηρέτησης που είναι αρκετά ευρεία ώστε να μπορούν να αναπαριστούν πολλά συστήματα εξυπηρέτησης που εμφανίζονται σε πραγματικές εφαρμογές και παράλληλα έχουν εύκολα υπολογίσιμη κατανομή ισορροπίας.

Ορισμός 11.2 (Δίκτυο Jackson) Ένα δίκτυο συστημάτων εξυπηρέτησης λέγεται δίκτυο Jackson αν είναι απλό Μαρκοβιανό δίκτυο με τις επιπλέον ιδιότητες

- (i) $\lambda_i(\mathbf{n}) = \lambda_i$ (δηλαδή η διαδικασία εξωτερικών αφίξεων σε κάθε σταθμό είναι Poisson),
- (ii) $\mu_j(\mathbf{n}) = \mu_j(n_j)$ (δηλαδή ο ρυθμός εξυπηρέτησης σε κάθε σταθμό εξαρτάται μόνο από το πλήθος των πελατών σε αυτόν),
- (iii) $\mu_{ji}(\mathbf{n}) = \mu_j(n_j)p_{ji}$ (δηλαδή οι πελάτες επιλέγουν τους σταθμούς που θα επισκεφθούν σύμφωνα με μια Μαρκοβιανή αλυσίδα διακριτού χρόνου (Μαρκοβιανές διαδρομές πελατών)).

◊

Ένα δίκτυο Jackson λέγεται ανοικτό αν υπάρχουν σταθμοί i, j τέτοιοι ώστε $\lambda_i > 0$ και $\mu_{j0}(n_j) > 0$ για κάποιο $n_j \geq 1$. Σε αντίθετη περίπτωση, δηλαδή όταν $\lambda_i = 0$ και $\mu_{j0}(n_j) = 0$ για όλους τους σταθμούς i και j και όλα τα $n_j \geq 1$, τότε το δίκτυο λέγεται κλειστό.

Για ένα ανοικτό δίκτυο Jackson ο πίνακας Μαρκοβιανών διαδρομών ενός πελάτη ορίζεται ως

$$\mathbf{P} = (p_{ij}) = \begin{pmatrix} 0 & \frac{\lambda_1}{\lambda} & \frac{\lambda_2}{\lambda} & \cdots & \frac{\lambda_{N-1}}{\lambda} & \frac{\lambda_N}{\lambda} \\ p_{10} & p_{11} & p_{12} & \cdots & p_{1,N-1} & p_{1N} \\ p_{20} & p_{21} & p_{22} & \cdots & p_{2,N-1} & p_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{N-1,0} & p_{N-1,1} & p_{N-1,2} & \cdots & p_{N-1,N-1} & p_{N-1,N} \\ p_{N0} & p_{N1} & p_{N2} & \cdots & p_{N,N-1} & p_{NN} \end{pmatrix},$$

όπου $\lambda = \sum_{i=1}^N \lambda_i$ είναι ο συνολικός ρυθμός αφίξεων στο δίκτυο. Ο πίνακας αυτός αντιστοιχεί στον πίνακα πιθανοτήτων μετάβασης από σταθμό σε σταθμό για κάθε συγκεκριμένο πελάτη που παρακολουθούμε εσάει να εισέρχεται στο και να εξέρχεται από το δίκτυο. Για ένα κλειστό δίκτυο Jackson ο αντίστοιχος πίνακας Μαρκοβιανών διαδρομών ενός πελάτη είναι

$$\mathbf{P} = (p_{ij}) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1,N-1} & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2,N-1} & p_{2N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{N-1,1} & p_{N-1,2} & \cdots & p_{N-1,N-1} & p_{N-1,N} \\ p_{N1} & p_{N2} & \cdots & p_{N,N-1} & p_{NN} \end{pmatrix}.$$

Ο πίνακας Μαρκοβιανών διαδρομών και στις δυο περιπτώσεις αντιστοιχεί στον πίνακα πιθανοτήτων μετάβασης μιας Μαρκοβιανής αλυσίδας διακριτού χρόνου, που είναι το διακριτό ανάλογο των Μαρκοβιανών αλυσίδων συνεχούς χρόνου.

Είναι εύκολο να ελέγξει κανείς ότι η διαδικασία συνεχούς χρόνου $\{\mathbf{Q}(t)\}$ των αριθμών των πελατών στους διάφορους σταθμούς ενός δικτύου Jackson είναι αδιαχώριστη αν και μόνο αν η Μαρκοβιανή αλυσίδα διακριτού χρόνου των Μαρκοβιανών διαδρομών ενός πελάτη είναι αδιαχώριστη, δηλαδή για κάθε σταθμούς i και j υπάρχουν σταθμοί i_1, i_2, \dots, i_n ώστε $p_{i_1 i_1} p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_n j} > 0$. Αυτό σημαίνει ότι η $\{\mathbf{Q}(t)\}$ είναι αδιαχώριστη αν και μόνο αν ένας πελάτης μπορεί να περάσει από όλους τους σταθμούς του δικτύου, ξεκινώντας από έναν από αυτούς. Αυτή είναι και η μόνη ενδιαφέρουσα περίπτωση. Αν δεν ισχύει, τότε το δίκτυο μπορεί να σπάσει σε

υποδίκτυα που λειτουργούν ανεξάρτητα και επομένως μπορεί να μελετήσει κανείς κάθε υποδίκτυο χωριστά. Για τον λόγο αυτό, από το σημείο αυτό και μετά θα υποθέτουμε ότι η διαδικασία $\{Q(t)\}$ είναι αδιαχώριστη και όταν θα αναφερόμαστε σε δίκτυο Jackson θα εννοούμε αδιαχώριστο δίκτυο Jackson.

Ο χώρος καταστάσεων \mathcal{S} της διαδικασίας $\{Q(t)\}$ που καταγράφει τον αριθμό των πελατών στους διάφορους σταθμούς ενός ανοικτού δικτύου Jackson είναι το σύνολο

$$\mathcal{S}_0 = \{(n_1, n_2, \dots, n_N) : n_i \geq 0, i = 1, 2, \dots, N\} = \mathbb{N}_0^N.$$

Για κλειστά δίκτυα Jackson με M πελάτες, ο χώρος καταστάσεων είναι το σύνολο

$$\mathcal{S}_c^M = \{(n_1, n_2, \dots, n_N) : n_i \geq 0, i = 1, 2, \dots, N \text{ και } \sum_{i=1}^N n_i = M\}.$$

Οι υποθέσεις (i)-(iii) που τέθηκαν στον ορισμό 11.2 δείχνουν ότι τα δίκτυα Jackson αποτελούν μια ειδική κατηγορία απλών Μαρκοβιανών δικτύων στην οποία ισχύουν κάποιες ιδιότητες «ανεξαρτησίας». Η ιδιότητα (i) αναφέρεται στις διαδικασίες εξωτερικών αφίξεων στους διαφόρους σταθμούς του δικτύου και απαιτεί να είναι όλες ανεξάρτητες διαδικασίες Poisson. Η ιδιότητα (ii) αναφέρεται στις διαδικασίες εξυπηρέτησεων στους διαφόρους σταθμούς του δικτύου και απαιτεί για κάθε σταθμό j , ο ρυθμός εξυπηρέτησης να μην εξαρτάται από το πλήθος των πελατών σε άλλους σταθμούς του δικτύου. Και, τέλος, η ιδιότητα (iii) αναφέρεται στη δρομολόγηση των πελατών και απαιτεί να έχει αυτή τη Μαρκοβιανή ιδιότητα, δηλαδή ο επόμενος σταθμός για έναν πελάτη να επιλέγεται με κάποια πιθανότητα που εξαρτάται από τον σταθμό στον οποίο τελειώνει την εξυπηρέτησή του και όχι από την προηγούμενη ιστορία του στο δίκτυο. Παρότι αυτές οι υποθέσεις είναι αρκετά περιοριστικές σε σχέση με το γενικό πλαίσιο των απλών Μαρκοβιανών δικτύων, ισχύουν σε αρκετά συστήματα που εμφανίζονται στις εφαρμογές της Θεωρίας Ουρών. Αυτός είναι και ο λόγος που τα δίκτυα Jackson μελετήθηκαν πρώτα.

Στις επόμενες παραγράφους θα μελετήσουμε ξεχωριστά τα ανοικτά και τα κλειστά δίκτυα Jackson.

11.3 Ρυθμοί διαπέρασης ανοικτών δικτύων Jackson

Ο ρυθμός διαπέρασης (throughput) Λ_i ενός σταθμού i σε ένα δίκτυο Jackson ορίζεται ως ο συνολικός ρυθμός περατώσεων εξυπηρέτησεων στο σταθμό i , $i = 1, 2, \dots, N$. Λόγω του ότι όλοι οι πελάτες που φθάνουν στο δίκτυο εξυπηρετούνται και δεν αναχωρούν πρόωρα (δεν υπάρχουν δηλαδή αποθαρρυνόμενοι, ούτε ανυπόμονοι πελάτες, ούτε κάποια άλλη διαδικασία αποχώρησης πελατών), ο ρυθμός διαπέρασης Λ_i ισούται επίσης με τον συνολικό ρυθμό αφίξεων και τον συνολικό ρυθμό αναχωρήσεων στο σταθμό i . Οι ρυθμοί διαπέρασης των σταθμών ενός ανοικτού δικτύου Jackson υπολογίζονται εύκολα με τη λύση ενός γραμμικού συστήματος διάστασης $N \times N$, όπως φαίνεται στο ακόλουθο αποτέλεσμα.

Θεώρημα 11.3 (Υπολογισμός ρυθμών διαπέρασης ανοικτού δικτύου Jackson) Οι ρυθμοί διαπέρασης Λ_i , $i = 1, 2, \dots, N$, σε ένα ανοικτό δίκτυο Jackson είναι η μοναδική θετική λύση του συστήματος των εξισώσεων κίνησης

$$\Lambda_i = \lambda_i + \sum_{j=1}^N \Lambda_j p_{ji}, \quad i = 1, 2, \dots, N. \quad (11.1)$$

Επιπλέον οι ρυθμοί ικανοποιούν τη σχέση

$$\sum_{j=1}^N \Lambda_j p_{j0} = \sum_{j=1}^N \lambda_j. \quad (11.2)$$

Το ότι οι ρυθμοί διαπέρασης ικανοποιούν τις εξισώσεις κίνησης είναι άμεσο, αρκεί να ερμηνεύσει κανείς κατάλληλα τα δυο μέλη κάθε εξίσωσης κίνησης. Πράγματι το αριστερό μέλος της εξίσωσης κίνησης για τον σταθμό i είναι ο συνολικός ρυθμός αφίξεων Λ_i στον σταθμό αυτό. Από την άλλη μεριά, ο πρώτος όρος του δεξιού μέλους είναι ο ρυθμός λ_i των εξωτερικών αφίξεων στον σταθμό i , ενώ κάθε προσθετέος του δεύτερου όρου του δεξιού μέλους, δηλαδή η ποσότητα $\Lambda_j p_{ji}$ αντιστοιχεί στον ρυθμό μεταβάσεων πελατών από τον σταθμό j στον σταθμό i . Προσθέτοντας κατά μέλη όλες τις εξισώσεις κίνησης παίρνουμε

$$\sum_{i=1}^N \Lambda_i = \sum_{i=1}^N \lambda_i + \sum_{i=1}^N \sum_{j=1}^N \Lambda_j p_{ji} = \sum_{i=1}^N \lambda_i + \sum_{j=1}^N \Lambda_j (1 - p_{j0}).$$

Απλοποιώντας το $\sum_{i=1}^N \Lambda_i$, παίρνουμε τη σχέση $\sum_{j=1}^N \Lambda_j p_{j0} = \sum_{j=1}^N \lambda_j$ που εξισώνει το ρυθμό αναχωρήσεων από το δίκτυο με το ρυθμό των αφίξεων σε αυτό. Θα μπορούσαμε να σκεφτόμαστε την εξίσωση αυτή ως εξίσωση κίνησης για τον εικονικό σταθμό 0 που παριστάνει το εξωτερικό του δικτύου. Επομένως αποδείξαμε ότι οι ρυθμοί διαπέρασης ικανοποιούν τις εξισώσεις κίνησης (11.1) και (11.2).

Για να δούμε ότι οι εξισώσεις κίνησης έχουν μοναδική λύση, ορίζουμε $\Lambda_0 = \sum_{j=1}^N \lambda_j = \lambda$ και $p_{0j} = \frac{\lambda_j}{\lambda}$. Τότε το σύστημα των εξισώσεων κίνησης για τους σταθμούς $1, 2, \dots, N$ και 0 γράφεται στη μορφή

$$\begin{aligned} \Lambda_i(1 - p_{ii}) &= \sum_{j \neq i} \Lambda_j p_{ji}, \quad i = 1, 2, \dots, N, \\ \Lambda_0 &= \sum_{j \neq 0} \Lambda_j p_{j0}. \end{aligned}$$

Το σύστημα αυτό ταυτίζεται με το σύστημα των εξισώσεων ισορροπίας της Μαρκοβιανής αλυσίδας συνεχούς χρόνου με πίνακα ρυθμών μετάβασης τον πίνακα

$$\mathbf{Q} = (q_{ij}) = \begin{pmatrix} -1 & \frac{\lambda_1}{\lambda} & \frac{\lambda_2}{\lambda} & \cdots & \frac{\lambda_{N-1}}{\lambda} & \frac{\lambda_N}{\lambda} \\ p_{10} & p_{11} - 1 & p_{12} & \cdots & p_{1,N-1} & p_{1N} \\ p_{20} & p_{21} & p_{22} - 1 & \cdots & p_{2,N-1} & p_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{N-1,0} & p_{N-1,1} & p_{N-1,2} & \cdots & p_{N-1,N-1} - 1 & p_{N-1,N} \\ p_{N0} & p_{N1} & p_{N2} & \cdots & p_{N,N-1} & p_{NN} - 1 \end{pmatrix}.$$

Η συγκεκριμένη Μαρκοβιανή αλυσίδα είναι αδιαχώριστη και πεπερασμένη και επομένως θετικά επαναληπτική. Επομένως, το διάνυσμα $(\Lambda_0, \Lambda_1, \dots, \Lambda_N)$ θα είναι πολλαπλάσιο της κατανομής ισορροπίας, δηλαδή $(\Lambda_0, \Lambda_1, \dots, \Lambda_N) = c(p_0, p_1, \dots, p_N)$, όπου (p_0, p_1, \dots, p_N) η μοναδική κατανομή ισορροπίας της αλυσίδας. Όμως $\Lambda_0 = \lambda$, οπότε $c = \frac{\lambda}{p_0}$, οπότε η λύση είναι μοναδική.

Παρατηρούμε, επομένως ότι οι ρυθμοί διαπέρασης των διαφόρων σταθμών ενός ανοικτού δικτύου Jackson υπολογίζονται μόνο από τους εξωτερικούς ρυθμούς αφίξεων και τις πιθανότητες του πίνακα Μαρκοβιανών διαδρομών. Οι συναρτήσεις των ρυθμών εξυπηρέτησης στους διάφορους σταθμούς του δικτύου δεν παίζουν κάποιο ρόλο.

11.4 Κατανομή ισορροπίας ανοικτών δικτύων Jackson

Οι εξισώσεις ισορροπίας για την κατανομή ισορροπίας της διαδικασίας $\{\mathbf{Q}(t)\}$ του αριθμού των πελατών στους διάφορους σταθμούς ενός Μαρκοβιανού δικτύου είναι ιδιαίτερα πολύπλοκες. Αυτό ισχύει ακόμη και στην απλούστερη περίπτωση των απλών Μαρκοβιανών δικτύων ή των δικτύων Jackson. Στη γενική περίπτωση οι εξισώσεις αυτές δεν επιδέχονται λύση σε κλειστή μορφή και για το λόγο αυτό επιστρατεύονται αριθμητικές και προσεγγιστικές μέθοδοι. Στην περίπτωση όμως των ανοικτών δικτύων Jackson, η κατανομή ισορροπίας βρίσκεται σε μια εξαιρετικά απλή κλειστή μορφή, όπως αποδεικνύεται στο παρακάτω θεώρημα.

Θεώρημα 11.4 (Κατανομή ισορροπίας ανοικτού δικτύου Jackson) Έστω δίκτυο Jackson N σταθμών, με ρυθμούς εξωτερικών αφίξεων $\lambda_i, i = 1, 2, \dots, N$, ρυθμούς εξυπηρέτησεων $\mu_j(n_j), n_j \geq 1, j = 1, 2, \dots, N$ και πιθανότητες μετάβασης πελατών $p_{j\bar{i}}, j = 1, 2, \dots, N$ και $i = 0, 1, \dots, N$. Έστω $\Lambda_i, i = 1, 2, \dots, N$ οι ρυθμοί διαπέρασης των σταθμών του δικτύου. Τότε:

- (i) Το δίκτυο είναι ευσταθές αν κάθε σταθμός $i = 1, 2, \dots, N$ μεμονωμένα είναι ευσταθής όταν τροφοδοτείται από μια διαδικασία Poisson με ρυθμό Λ_i και λειτουργεί σύμφωνα με τους ρυθμούς εξυπηρέτησης $\mu_i(n_i)$. Δηλαδή το δίκτυο είναι ευσταθές, αν

$$B_i^{-1} = 1 + \sum_{n_i=1}^{\infty} \frac{\Lambda_i^{n_i}}{\mu_i(1)\mu_i(2)\cdots\mu_i(n_i)} < \infty, \quad i = 1, 2, \dots, N. \quad (11.3)$$

- (ii) Όταν το δίκτυο είναι ευσταθές, τότε η κατανομή ισορροπίας του είναι

$$p(\mathbf{n}) = p(n_1, n_2, \dots, n_N) = \prod_{i=1}^N p_i(n_i), \quad \mathbf{n} \in \mathbb{N}_0^N, \quad (11.4)$$

όπου η $(p_i(n_i) : n_i \geq 0)$ είναι η κατανομή ισορροπίας του σταθμού i όταν θεωρηθεί μεμονωμένα τροφοδοτούμενος από μια διαδικασία Poisson με ρυθμό Λ_i και λειτουργώντας με τους ρυθμούς $\mu_i(n_i)$. Δηλαδή η $p_i(n_i)$ δίνεται από τον τύπο

$$p_i(n_i) = B_i \frac{\Lambda_i^{n_i}}{\mu_i(1)\mu_i(2)\cdots\mu_i(n_i)}, \quad n_i \geq 0, \quad i = 1, 2, \dots, N. \quad (11.5)$$

◊

Η απόδειξη του θεωρήματος αυτού γίνεται δείχνοντας ότι η κατανομή ισορροπίας που δίνεται από τους τύπους (11.4) και (11.5) ικανοποιεί τις εξισώσεις ισορροπίας της $\{\mathbf{Q}(t)\}$. Οι εξισώσεις ισορροπίας για τη στάσιμη κατανομή είναι οι

$$\begin{aligned} & p(\mathbf{n}) \left(\sum_{i=1}^N \lambda_i + \sum_{j=1}^N \mu_j(n_j) 1\{n_j > 0\} \right) \\ &= \sum_{j=1}^N p(\mathbf{n} - \mathbf{e}_j) \lambda_j 1\{n_j > 0\} + \sum_{i=1}^N p(\mathbf{n} + \mathbf{e}_i) \mu_i(n_i + 1) p_{i0} \\ & \quad + \sum_{j=1}^N \sum_{i=1}^N p(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j) \mu_i(n_i + 1) p_{ij} 1\{n_j > 0\}, \quad \mathbf{n} \in \mathcal{S}_0, \end{aligned}$$

που γράφονται ισοδύναμα ως

$$\begin{aligned} & p(\mathbf{n}) \sum_{i=1}^N \lambda_i - \sum_{i=1}^N p(\mathbf{n} + \mathbf{e}_i) \mu_i(n_i + 1) p_{i0} \\ & \quad + \sum_{j=1}^N \left(p(\mathbf{n}) \mu_j(n_j) - p(\mathbf{n} - \mathbf{e}_j) \lambda_j - \sum_{i=1}^N p(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j) \mu_i(n_i + 1) p_{ij} \right) 1\{n_j > 0\} \\ &= 0, \quad \mathbf{n} \in \mathcal{S}_0. \end{aligned} \quad (11.6)$$

Ορίζοντας για κάθε κατάσταση $\mathbf{n} \in \mathbb{N}_0^N$ τις ποσότητες $\Sigma_j = \Sigma_j(\mathbf{n})$, $0 \leq j \leq N$ από τις σχέσεις

$$\Sigma_0 = p(\mathbf{n}) \sum_{i=1}^N \lambda_i - \sum_{i=1}^N p(\mathbf{n} + \mathbf{e}_i) \mu_i(n_i + 1) p_{i0},$$

$$\Sigma_j = p(\mathbf{n}) \mu_j(n_j) - p(\mathbf{n} - \mathbf{e}_j) \lambda_j - \sum_{i=1}^N p(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j) \mu_i(n_i + 1) p_{ij}, \quad 1 \leq j \leq N,$$

η σχέση (11.6) γράφεται $\Sigma_0 + \sum_{j=1}^N \Sigma_j 1\{n_j > 0\} = 0$. Θα αποδείξουμε ότι $\Sigma_j = 0$ για $0 \leq j \leq N$, αντικαθιστώντας την $p(\mathbf{n})$ που δίνεται από την (11.4). Χρησιμοποιώντας τις (11.4) και (11.5), παρατηρούμε ότι

$$\frac{p(\mathbf{n} + \mathbf{e}_i)}{p(\mathbf{n})} = \frac{\Lambda_i}{\mu_i(n_i + 1)}, \quad \frac{p(\mathbf{n})}{p(\mathbf{n} - \mathbf{e}_j)} = \frac{\Lambda_j}{\mu_j(n_j)}, \quad \frac{p(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j)}{p(\mathbf{n} - \mathbf{e}_j)} = \frac{\Lambda_i}{\mu_i(n_i + 1)}.$$

Χρησιμοποιώντας τις σχέσεις αυτές και διαιρώντας με $p(\mathbf{n})$ τη σχέση $\Sigma_0 = 0$ έχουμε

$$\begin{aligned} \Sigma_0 = 0 &\Leftrightarrow \sum_{i=1}^N \lambda_i - \sum_{i=1}^N \frac{p(\mathbf{n} + \mathbf{e}_i)}{p(\mathbf{n})} \mu_i(n_i + 1) p_{i0} = 0 \\ &\Leftrightarrow \sum_{i=1}^N \lambda_i - \sum_{i=1}^N \frac{\Lambda_i}{\mu_i(n_i + 1)} \mu_i(n_i + 1) p_{i0} = 0 \\ &\Leftrightarrow \sum_{i=1}^N \lambda_i - \sum_{i=1}^N \Lambda_i p_{i0} = 0, \end{aligned}$$

που ισχύει λόγω της εξίσωσης κίνησης (11.2) για τον σταθμό 0. Ομοίως, διαιρώντας με $p(\mathbf{n} - \mathbf{e}_j)$ τη σχέση $\Sigma_j = 0$, για $1 \leq j \leq N$ έχουμε

$$\begin{aligned} \Sigma_j = 0 &\Leftrightarrow \frac{p(\mathbf{n})}{p(\mathbf{n} - \mathbf{e}_j)} \mu_j(n_j) - \lambda_j - \sum_{i=1}^N \frac{p(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j)}{p(\mathbf{n} - \mathbf{e}_j)} \mu_i(n_i + 1) p_{ij} = 0 \\ &\Leftrightarrow \frac{\Lambda_j}{\mu_j(n_j)} \mu_j(n_j) - \lambda_j - \sum_{i=1}^N \frac{\Lambda_i}{\mu_i(n_i + 1)} \mu_i(n_i + 1) p_{ij} = 0 \\ &\Leftrightarrow \Lambda_j - \lambda_j - \sum_{i=1}^N \Lambda_i p_{ij} = 0, \end{aligned}$$

που ισχύει από την εξίσωση κίνησης (11.1) για τον σταθμό j με $1 \leq j \leq N$.

Η απόδειξη του θεωρήματος 11.4 αποκαλύπτει ότι η κατανομή ισορροπίας ενός ανοικτού δικτύου Jackson ικανοποιεί για κάθε σταθμό j με $0 \leq j \leq N$ και κατάσταση \mathbf{n} τις πιο περιοριστικές εξισώσεις $\Sigma_j = \Sigma_j(\mathbf{n}) = 0$, που συνεπάγονται τις εξισώσεις πλήρους ισορροπίας (11.6) που γράφονται ως $\Sigma_0 + \sum_{j=1}^N \Sigma_j 1\{n_j > 0\} = 0$. Οι εξισώσεις αυτές γράφονται στη μορφή

$$p(\mathbf{n}) \sum_{i=1}^N \lambda_i = \sum_{i=1}^N p(\mathbf{n} + \mathbf{e}_i) \mu_i(n_i + 1) p_{i0}, \quad (11.7)$$

$$p(\mathbf{n}) \mu_j(n_j) = p(\mathbf{n} - \mathbf{e}_j) \lambda_j + \sum_{i=1}^N p(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j) \mu_i(n_i + 1) p_{ij}, \quad 1 \leq j \leq N. \quad (11.8)$$

Οι εξισώσεις αυτές αναφέρονται ως εξισώσεις μερικής ισορροπίας για την κατανομή ισορροπίας. Η εξίσωση (11.8) απαιτεί ο ρυθμός εξόδου από μια κατάσταση \mathbf{n} λόγω αναχωρήσεως (δηλαδή συμπλήρωσης εξυπηρέτησης) στην ουρά j να ισούται με τον ρυθμό εισόδου στην κατάσταση \mathbf{n} λόγω αφίξεως (εξωτερικής ή εσωτερικής) στην ουρά j . Η εξίσωση (11.7) απαιτεί το ίδιο για τον σταθμό 0 που αντιστοιχεί στο εξωτερικό του δικτύου. Επομένως, μπορεί να ερμηνευθεί ευκολότερα, λέγοντας ότι απαιτεί ο ρυθμός εξόδου από μια κατάσταση \mathbf{n} λόγω (εξωτερικής) άφιξης στο δίκτυο να ισούται με τον ρυθμό εισόδου στην ίδια κατάσταση λόγω αναχωρήσεως από το δίκτυο.

Οι εξισώσεις μερικής ισορροπίας είναι πιο περιοριστικές από τις εξισώσεις (πλήρους) ισορροπίας που απαιτούν ο ρυθμός εξόδου από μια κατάσταση \mathbf{n} να ισούται με τον ρυθμό εισόδου σε αυτήν και λιγότερο περιοριστικές από τις εξισώσεις λεπτομερούς ισορροπίας που απαιτούν ο ρυθμός εξόδου από μια κατάσταση \mathbf{n} λόγω μετάβασής της σε μια άλλη κατάσταση \mathbf{n}' να ισούται με το ρυθμό εισόδου στην \mathbf{n} από την κατάσταση \mathbf{n}' . Ενώ οι εξισώσεις (πλήρους) ισορροπίας ισχύουν σε όλα τα δίκτυα και οι εξισώσεις λεπτομερούς ισορροπίας ισχύουν μόνο στα αντιστρέψιμα Μαρκοβιανά δίκτυα, οι εξισώσεις μερικής ισορροπίας ισχύουν στα δίκτυα Jackson.

Το θεώρημα 11.4 δείχνει πώς η κατανομή ισορροπίας ενός δικτύου Jackson μπορεί να υπολογιστεί πολύ αποδοτικά. Το πρώτο βήμα είναι ο υπολογισμός των ρυθμών διαπέρασης των διάφορων σταθμών του δικτύου που ανάγεται στη λύση του $N \times N$ γραμμικού συστήματος των εξισώσεων κίνησης. Το δεύτερο βήμα είναι ο υπολογισμός της κατανομής ισορροπίας κάθε σταθμού i , με βάση τη θεωρία των απλών Μαρκοβιανών ουρών, σαν ο σταθμός να είχε Poisson διαδικασία αφίξεων με ρυθμό τον αντίστοιχο ρυθμό διαπέρασης του συγκεκριμένου σταθμού. Και τέλος, στο τρίτο βήμα, η κατανομή ισορροπίας του δικτύου υπολογίζεται ως το γινόμενο των αντίστοιχων περιθωρίων κατανομών που έχουν υπολογιστεί στο δεύτερο βήμα.

Για να κατανοήσει κανείς την υπολογιστική απλοποίηση που προσφέρει το θεώρημα, ας υποθέσουμε ότι είχαμε ένα δίκτυο N σταθμών και αποφασίζαμε να βρούμε την κατανομή ισορροπίας του λύνοντας αριθμητικά το σύστημα των εξισώσεων ισορροπίας του, περικόπτοντας τις καταστάσεις από κάποιο αριθμό πελατών M και πάνω για κάθε σταθμό. Τότε, ο αριθμός των καταστάσεων του περικεκομένου δικτύου θα ήταν M^N και επομένως ο υπολογισμός της κατανομής ισορροπίας θα απαιτούσε τη λύση ενός $M^N \times M^N$ γραμμικού συστήματος. Χρησιμοποιώντας το θεώρημα 11.4, θα έπρεπε να λύσουμε ένα $N \times N$ σύστημα για τον προσδιορισμό των ρυθμών διαπέρασης των σταθμών του δικτύου και κατόπιν θα είχαμε να λύσουμε N γραμμικά συστήματα διάστασης $M \times M$, ένα για κάθε σταθμό. Στην πραγματικότητα μάλιστα δεν θα χρειαζόταν καν να λυθούν αυτά τα συστήματα, αφού θα χρησιμοποιούσαμε τον τύπο (6.3) που δίνει την κατανομή ισορροπίας μιας απλής Μαρκοβιανής ουράς. Ακόμα και για μικρές τιμές των N, M η υπολογιστική οικονομία είναι εκπληκτική.

Πέραν της υπολογιστικής ευκολίας, το θεώρημα Jackson 11.4 είναι σημαντικό καθώς συνδέεται με διάφορα αποτελέσματα θεωρητικού χαρακτήρα. Καταρχήν, ο τύπος (11.4) δείχνει ότι όταν ένα δίκτυο Jackson είναι στάσιμο, τότε οι τυχαίες μεταβλητές $Q_j(t)$, $1 \leq j \leq N$, που αντιστοιχούν στα πλήθη των πελατών στους διάφορους σταθμούς του δικτύου για κάθε συγκεκριμένη χρονική στιγμή t είναι ανεξάρτητες τυχαίες μεταβλητές. Προσοχή, όμως: Οι στοχαστικές διαδικασίες $\{Q_j(t)\}$, $1 \leq j \leq N$, που αντιστοιχούν στα πλήθη των πελατών στους διάφορους σταθμούς του δικτύου δεν είναι ανεξάρτητες. Επίσης, παρότι η περιθωρία κατανομή ισορροπίας που αντιστοιχεί σε κάθε σταθμό j υπολογίζεται χωριστά, θεωρώντας τον σταθμό σαν να τροφοδοτείτο από μια στοχαστική διαδικασία Poisson με ρυθμό Λ_j , η διαδικασία συνολικών (εξωτερικών και εσωτερικών) αφίξεων στο σταθμό δεν είναι Poisson. Για να αποσαφηνιστεί αυτό θεωρούμε το επόμενο παράδειγμα μιας $M/M/1$ ουράς με επανάληψη εξυπηρέτησης κάποιων πελατών που τοποθετούνται στο τέλος της ουράς.

Παράδειγμα 11.5 (M/M/1 ουρά με επανάληψη εξυπηρέτησης ενός ποσοστού πελατών) Έστω μια $M/M/1$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ , όπου κάθε πελάτης που τελειώνει την εξυπηρέτησή του την επαναλαμβάνει με πιθανότητα p , τοποθετούμενος στο τέλος της ουράς. Ένα τέτοιο σύστημα είναι ένα τετριμμένο δίκτυο Jackson με $N = 1$ σταθμούς και πιθανότητα δρομολόγησης $p_{11} = p$. Επομένως, η εξίσωση κίνησης είναι $\Lambda_1 = \lambda + \Lambda_1 p$, που δίνει $\Lambda_1 = \frac{\lambda}{1-p}$. Για να βρούμε την κατανομή ισορροπίας αυτού του συστήματος θεωρούμε την $M/M/1$ ουρά με ρυθμό εξυπηρέτησης μ (χωρίς επαναλήψεις εξυπηρέτησης), όταν τροφοδοτείται από μια διαδικασία Poisson με ρυθμό Λ_1 . Χρησιμοποιώντας τον τύπο (6.10) για την κατανομή ισορροπίας της

$M/M/1$ ουράς, έχουμε επομένως ότι η κατανομή ισορροπίας του αριθμού των πελατών είναι

$$p(n) = \left(1 - \frac{\lambda}{(1-p)\mu}\right) \left(\frac{\lambda}{(1-p)\mu}\right)^n, \quad n \geq 0. \quad (11.9)$$

Μπορούμε τώρα να δούμε ότι η διαδικασία των συνολικών (εξωτερικών και εσωτερικών) αφίξεων δεν έχει την ιδιότητα των ανεξάρτητων προσυξήσεων και επομένως δεν είναι Poisson. Πράγματι, έστω $A(t, s)$ το πλήθος των αφίξεων στο διάστημα $(t, s]$. Τότε, έχουμε

$$\begin{aligned} \Pr[A(t, t+h) = 1] &= \lambda h + (1-p(0))\mu p h + o(h) \\ &= \Lambda_1 h + o(h), \quad \text{για } h \rightarrow 0^+, \\ \Pr[A(t, t+h) = 1 | A(t-h, t) = 1] &= \lambda h + \mu p h + o(h) \\ &> \Lambda_1 h + o(h), \quad \text{για } h \rightarrow 0^+. \end{aligned}$$

Επομένως, βλέπουμε ότι η γνώση του ότι έγινε άφιξη λίγο πριν τη στιγμή t , μεγαλώνει την πιθανότητα να συμβεί άφιξη λίγο μετά τη στιγμή t . Ο λόγος είναι ότι μια άφιξη λίγο πριν τη στιγμή t βεβαιώνει ότι ο υπηρέτης εξυπηρετεί τη στιγμή t και επομένως υπάρχει πιθανότητα εσωτερικής άφιξης, αν τελειώσει κάποια εξυπηρέτηση στο επόμενο διάστημα μήκους h με πιθανότητα $(\mu h + o(h))$ και επαναληφθεί με πιθανότητα p . Χωρίς τη γνώση μιας τέτοιας άφιξης, η πιθανότητα ο υπηρέτης να εξυπηρετεί τη στιγμή t είναι $1 - p(0)$ και επομένως η πιθανότητα εσωτερικής άφιξης γίνεται μικρότερη. \diamond

Παρά το ότι η συνολική διαδικασία αφίξεων σε έναν σταθμό δεν είναι Poisson όταν το δίκτυο επιτρέπει τη δρομολόγηση πελατών σε σταθμούς που έχουν ήδη επισκεφθεί, ισχύει για τα δίκτυα Jackson μια ιδιότητα ανάλογη της PASTA. Πιο συγκεκριμένα, η πιθανότητα ένας αφικνούμενος πελάτης στον σταθμό i ενός δικτύου να δει το δίκτυο στην κατάσταση \mathbf{n} (εξαιρώντας από την κατάσταση τον εαυτό του), δίνεται από την πιθανότητα ισορροπίας της κατάστασης \mathbf{n} σε συνεχή χρόνο.

Θεώρημα 11.6 (Ιδιότητα Arrivals See Time Averages (ASTA) για ανοικτά δίκτυα Jackson) Σε ένα ανοικτό δίκτυο Jackson οι κατανομές ισορροπίας των αριθμών των πελατών σε στιγμές αφίξεων σε κάποιο συγκεκριμένο σταθμό (εξαιρουμένου του πελάτη που αφικνείται) και σε συνεχή χρόνο συμπίπτουν. \diamond

Για να αιτιολογηθεί αυτή η ιδιότητα, θα πρέπει καταρχήν να εκφράσουμε την πιθανότητα ισορροπίας σε στιγμή άφιξης πελάτη στον σταθμό i η κατάσταση του δικτύου (εξαιρουμένου του αφικνούμενου πελάτη) να είναι \mathbf{n} . Αν $\{\mathbf{Q}(t)\}$ είναι η διαδικασία του αριθμού των πελατών του δικτύου και $A_i(t, t+h)$ η διαδικασία των συνολικών αφίξεων στον σταθμό i , η πιθανότητα αυτή είναι

$$\begin{aligned} a_i(\mathbf{n}) &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \Pr[\mathbf{Q}(t+h) = \mathbf{n} + \mathbf{e}_i | A_i(t, t+h) > 0] \\ &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \frac{\Pr[\mathbf{Q}(t+h) = \mathbf{n} + \mathbf{e}_i, A_i(t, t+h) > 0]}{\Pr[A_i(t, t+h) > 0]}. \end{aligned} \quad (11.10)$$

Όμως,

$$\begin{aligned} &\Pr[\mathbf{Q}(t+h) = \mathbf{n} + \mathbf{e}_i, A_i(t, t+h) > 0] \\ &= \Pr[\mathbf{Q}(t) = \mathbf{n}, \mathbf{Q}(t+h) = \mathbf{n} + \mathbf{e}_i] + \sum_{j=1}^N \Pr[\mathbf{Q}(t) = \mathbf{n} + \mathbf{e}_j, \mathbf{Q}(t+h) = \mathbf{n} + \mathbf{e}_i] \\ &= \Pr[\mathbf{Q}(t) = \mathbf{n}] \Pr[\mathbf{Q}(t+h) = \mathbf{n} + \mathbf{e}_i | \mathbf{Q}(t) = \mathbf{n}] \\ &\quad + \sum_{j=1}^N \Pr[\mathbf{Q}(t) = \mathbf{n} + \mathbf{e}_j] \Pr[\mathbf{Q}(t+h) = \mathbf{n} + \mathbf{e}_i | \mathbf{Q}(t) = \mathbf{n} + \mathbf{e}_j], \end{aligned}$$

οπότε

$$\begin{aligned}
& \lim_{t \rightarrow \infty} \Pr[\mathbf{Q}(t+h) = \mathbf{n} + \mathbf{e}_i, A_i(t, t+h) > 0] \\
&= p(\mathbf{n})\lambda_i h + \sum_{j=1}^N p(\mathbf{n} + \mathbf{e}_j)\mu_j(n_j+1)hp_{ji} + o(h), \quad h \rightarrow 0^+, \\
&= p(\mathbf{n}) \left(\lambda_i + \sum_{j=1}^N \frac{p(\mathbf{n} + \mathbf{e}_j)}{p(\mathbf{n})} \mu_j(n_j+1)p_{ji} \right) h + o(h), \quad h \rightarrow 0^+, \\
&= p(\mathbf{n}) \left(\lambda_i + \sum_{j=1}^N \frac{\Lambda_j}{\mu_j(n_j+1)} \mu_j(n_j+1)p_{ji} \right) h + o(h), \quad h \rightarrow 0^+, \\
&= p(\mathbf{n})\Lambda_i h + o(h), \quad h \rightarrow 0^+, \tag{11.11}
\end{aligned}$$

χρησιμοποιώντας στην τελευταία ισότητα τις εξισώσεις κίνησης (11.1) και στην προτελευταία τους τύπους (11.4) και (11.5) για την κατανομή ισορροπίας. Αθροίζοντας για όλα τα \mathbf{n} , παίρνουμε

$$\begin{aligned}
& \lim_{t \rightarrow \infty} \Pr[A_i(t, t+h) > 0] \\
&= \sum_{\mathbf{n} \in \mathbb{N}_0^N} p(\mathbf{n})\Lambda_i h + o(h), \quad h \rightarrow 0^+, \\
&= \Lambda_i h + o(h), \quad h \rightarrow 0^+. \tag{11.12}
\end{aligned}$$

Αντικαθιστώντας τις (11.11) και (11.12) στην (11.10), παίρνουμε

$$\begin{aligned}
a_i(\mathbf{n}) &= \lim_{t \rightarrow \infty} \lim_{h \rightarrow 0^+} \Pr[\mathbf{Q}(t+h) = \mathbf{n} + \mathbf{e}_i | A_i(t, t+h) > 0] \\
&= \lim_{h \rightarrow 0^+} \frac{p(\mathbf{n})\Lambda_i h + o(h)}{\Lambda_i h + o(h)} = \lim_{h \rightarrow 0^+} \frac{p(\mathbf{n})\Lambda_i + \frac{o(h)}{h}}{\Lambda_i + \frac{o(h)}{h}} = p(\mathbf{n}), \tag{11.13}
\end{aligned}$$

που αποδεικνύει το συμπέρασμα του θεωρήματος.

Μια άλλη ιδιότητα της κλάσης των δικτύων Jackson είναι ότι είναι κλειστή ως προς τη (χρονική) αντιστροφή. Δηλαδή, η αντίστροφη στοχαστική διαδικασία ενός δικτύου Jackson είναι και αυτή δίκτυο Jackson. Πιο συγκεκριμένα έχουμε το ακόλουθο αποτέλεσμα.

Θεώρημα 11.7 (Αντίστροφη διαδικασία ανοικτού δικτύου Jackson) Η αντίστροφη στοχαστική διαδικασία $\{\hat{\mathbf{Q}}(t)\}$ ενός ανοικτού δικτύου Jackson N σταθμών με συναρτήσεις ρυθμών εξυπηρέτησης $(\mu_j(n_j) : n_j > 0), 1 \leq j \leq N$, ρυθμούς εξωτερικών αφίξεων $\lambda_i, 1 \leq i \leq N$, πιθανότητες δρομολόγησης πελατών $p_{ij}, 1 \leq i \leq N$ και $0 \leq j \leq N$, και διαδικασία αριθμών πελατών $\{\mathbf{Q}(t)\}$ αντιστοιχεί σε ανοικτό δίκτυο Jackson με συναρτήσεις ρυθμών εξυπηρέτησης $(\hat{\mu}_j(n_j) : n_j > 0), 1 \leq j \leq N$, ρυθμούς εξωτερικών αφίξεων $\hat{\lambda}_i, 1 \leq i \leq N$ και πιθανότητες δρομολόγησης πελατών $\hat{p}_{ij}, 1 \leq i \leq N$ και $0 \leq j \leq N$, όπου

$$\begin{aligned}
\hat{\mu}_j(n_j) &= \mu_j(n_j), \quad n_j > 0, \quad 1 \leq j \leq N, \\
\hat{\lambda}_i &= \Lambda_i p_{i0}, \quad 1 \leq i \leq N, \\
\hat{p}_{j0} &= \frac{\lambda_j}{\Lambda_j}, \quad 1 \leq j \leq N, \\
\hat{p}_{ji} &= \frac{\Lambda_i p_{ij}}{\Lambda_j}, \quad 1 \leq j, i \leq N,
\end{aligned}$$

με $\Lambda_j, 1 \leq j \leq N$ να είναι οι ρυθμοί διαπέρασης του αρχικού δικτύου. ◊

Είναι άμεσο να δούμε ότι το αντίστροφο ενός ανοικτού δικτύου Jackson είναι απλό Μαρκοβιανό δίκτυο, αφού οι επιτρεπτές μεταβάσεις από μια κατάσταση \mathbf{n} στο αντίστροφο αντιστοιχούν στην μετακίνηση ενός μόνο πελάτη, είτε από μια ουρά σε άλλη, είτε από μια ουρά στο εξωτερικό του δικτύου, είτε από το εξωτερικό του δικτύου σε κάποια ουρά. Θα πρέπει επιπλέον να ελέγξουμε ότι οι ρυθμοί των μεταβάσεων του αντιστρόφου δικτύου έχουν τη μορφή που αναφέρεται στον ορισμό 11.2. Για τον υπολογισμό των ρυθμών του αντίστροφου δικτύου εφαρμόζουμε το θεώρημα 10.2. Έχουμε:

$$\begin{aligned}\hat{\lambda}_i(\mathbf{n}) &= \frac{p(\mathbf{n} + \mathbf{e}_i)\mu_i(n_i + 1)p_{i0}}{p(\mathbf{n})} = \frac{\Lambda_i\mu_i(n_i + 1)p_{i0}}{\mu_i(n_i + 1)} = \Lambda_i p_{i0}, \quad 1 \leq i \leq N, \\ \hat{\mu}_{ji}(\mathbf{n}) &= \frac{p(\mathbf{n} - \mathbf{e}_j + \mathbf{e}_i)\mu_i(n_i + 1)p_{ij}}{p(\mathbf{n})} = \frac{\frac{p(\mathbf{n} - \mathbf{e}_j + \mathbf{e}_i)}{p(\mathbf{n} - \mathbf{e}_j)}\mu_i(n_i + 1)p_{ij}}{\frac{p(\mathbf{n})}{p(\mathbf{n} - \mathbf{e}_j)}} \\ &= \frac{\frac{\Lambda_i}{\mu_i(n_i + 1)}\mu_i(n_i + 1)p_{ij}}{\frac{\Lambda_j}{\mu_j(n_j)}} = \mu_j(n_j)\frac{\Lambda_i p_{ij}}{\Lambda_j}, \quad 1 \leq j, i \leq N, \\ \hat{\mu}_{j0}(\mathbf{n}) &= \frac{p(\mathbf{n} - \mathbf{e}_j)\lambda_j}{p(\mathbf{n})} = \mu_j(n_j)\frac{\lambda_j}{\Lambda_j}, \quad 1 \leq j \leq N,\end{aligned}\tag{11.14}$$

οπότε βλέπουμε ότι το αντίστροφο δίκτυο είναι και αυτό Jackson με τις παραμέτρους που περιγράφονται στο θεώρημα 11.7.

Βασιζόμενοι τώρα στο θεώρημα 11.7, μπορούμε να συναγάγουμε το ακόλουθο αποτέλεσμα, που αποτελεί γενίκευση του θεωρήματος 10.10 του Burke.

Θεώρημα 11.8 (Επέκταση θεωρήματος Burke στα ανοικτά δίκτυα Jackson) *Οι διαδικασίες των αναχωρήσεων από το σύστημα, $\{D_i(t)\}$, $i = 1, 2, \dots, N$, των σταθμών ενός στάσιμου ανοικτού δικτύου Jackson είναι ανεξάρτητες διαδικασίες Poisson με αντίστοιχους ρυθμούς $\Lambda_i p_{i0}$, $i = 1, 2, \dots, N$. Επιπλέον, για κάθε επιλεγμένη χρονική στιγμή t , η διαδικασία των αριθμών των πελατών στους διάφορους σταθμούς του δικτύου από τη στιγμή t και μετά, δηλαδή η $\{\mathbf{Q}(s) : s \geq t\}$, είναι ανεξάρτητη από τις διαδικασίες των αναχωρήσεων πριν τη στιγμή t , δηλαδή από τις $\{D_i(s) : s < t\}$, $i = 1, 2, \dots, N$.* ◊

Η απόδειξη του θεωρήματος αυτού ακολουθεί την ίδια λογική με την απόδειξη του θεωρήματος του Burke, δηλαδή αρκεί κανείς να δει ότι οι διαδικασίες αναχωρήσεων από το σύστημα των διαφόρων σταθμών του δικτύου αντιστοιχούν στις διαδικασίες εξωτερικών αφίξεων του αντιστρόφου δικτύου, αν τις δει κανείς με την αντίστροφη φορά του χρόνου. Επομένως, επειδή οι διαδικασίες εξωτερικών αφίξεων στο αντίστροφο δίκτυο είναι Poisson και μάλιστα η εξελιξή τους μετά από κάποια στιγμή $-t$ είναι ανεξάρτητη από την κατάσταση του αντιστρόφου δικτύου μέχρι τη στιγμή $-t$, έχουμε τα συμπεράσματα του θεωρήματος.

11.5 Σχετικοί ρυθμοί διαπέρασης κλειστών δικτύων Jackson

Ο ρυθμός διαπέρασης (throughput) Λ_i^{TH-c} ενός σταθμού i σε ένα κλειστό δίκτυο Jackson ορίζεται ακριβώς όπως και για τα ανοικτά δίκτυα. Είναι, δηλαδή, ο συνολικός ρυθμός περατώσεων εξυπηρητήσεων στον σταθμό i , $i = 1, 2, \dots, N$ ή ισοδύναμα ο συνολικός ρυθμός αφίξεων στον σταθμό αυτό (δεδομένου ότι εξυπηρετούνται όλοι οι πελάτες που φθάνουν). Οι ρυθμοί διαπέρασης των σταθμών ενός κλειστού δικτύου Jackson ικανοποιούν ένα γραμμικό $N \times N$ σύστημα, ανάλογο με το σύστημα των εξισώσεων κίνησης που είχαμε στα ανοικτά δίκτυα. Ουσιαστικά πρόκειται για το σύστημα που προκύπτει όταν οι ρυθμοί εξωτερικών αφίξεων είναι όλοι 0.

Θεώρημα 11.9 (Εξισώσεις κίνησης για ρυθμούς διαπέρασης κλειστού δικτύου Jackson) Οι ρυθμοί διαπέρασης Λ_i^{TH-c} , $i = 1, 2, \dots, N$, σε ένα κλειστό δίκτυο Jackson ικανοποιούν το σύστημα των εξισώσεων κίνησης

$$\Lambda_i = \sum_{j=1}^N \Lambda_j p_{ji}, \quad i = 1, 2, \dots, N. \quad (11.15)$$

Το σύστημα των εξισώσεων κίνησης δεν έχει μοναδική λύση. Οι συντεταγμένες της μοναδικής λύσης του συστήματος που ικανοποιεί επιπλέον την εξίσωση

$$\sum_{i=1}^N \Lambda_i = 1,$$

αναφέρονται ως σχετικοί ρυθμοί διαπέρασης των διαφόρων σταθμών. Ο σχετικός ρυθμός διαπέρασης του σταθμού i συμβολίζεται με Λ_i^c . Κάθε λύση $(\Lambda_1, \Lambda_2, \dots, \Lambda_N)$ του συστήματος των εξισώσεων κίνησης είναι πολλαπλάσιο της λύσης $(\Lambda_1^c, \Lambda_2^c, \dots, \Lambda_N^c)$. ◊

Το ότι οι ρυθμοί διαπέρασης ικανοποιούν το σύστημα των εξισώσεων κίνησης αιτιολογείται ακριβώς όπως το αντίστοιχο αποτέλεσμα για ανοικτά δίκτυα Jackson. Το σύστημα γράφεται στη μορφή

$$\Lambda_i(1 - p_{ii}) = \sum_{j \neq i} \Lambda_j p_{ji}, \quad i = 1, 2, \dots, N,$$

και επομένως μπορεί να ερμηνευθεί ως το σύστημα των εξισώσεων ισορροπίας μιας Μαρκοβιανής αλυσίδας συνεχούς χρόνου με πίνακα ρυθμών μετάβασης τον πίνακα

$$\mathbf{Q} = (q_{ij}) = \begin{pmatrix} p_{11} - 1 & p_{12} & \cdots & p_{1,N-1} & p_{1N} \\ p_{21} & p_{22} - 1 & \cdots & p_{2,N-1} & p_{2N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{N-1,1} & p_{N-1,2} & \cdots & p_{N-1,N-1} - 1 & p_{N-1,N} \\ p_{N1} & p_{N2} & \cdots & p_{N,N-1} & p_{NN} - 1 \end{pmatrix}.$$

Η συγκεκριμένη Μαρκοβιανή αλυσίδα είναι αδιαχώριστη και πεπερασμένη και επομένως θετικά επαναληπτική. Η κατανομή ισορροπίας της που επιπλέον ικανοποιεί την εξίσωση κανονικοποίησης είναι ακριβώς το διάνυσμα των σχετικών ρυθμών διαπέρασης των σταθμών. Κάθε άλλη λύση του συστήματος, άρα και το διάνυσμα των (απόλυτων) ρυθμών διαπέρασης είναι πολλαπλάσιο της κατανομής ισορροπίας. Παρατηρούμε, επομένως, ότι σε αντίθεση με τα ανοικτά δίκτυα, οι ρυθμοί διαπέρασης των διαφόρων σταθμών ενός κλειστού δικτύου Jackson δεν μπορούν να υπολογιστούν μόνο από τους εξωτερικούς ρυθμούς αφίξεων και τις πιθανότητες του πίνακα Μαρκοβιανών διαδρομών. Πάντως, οι σχετικοί ρυθμοί αρκούν για τον υπολογισμό της κατανομής ισορροπίας ενός δικτύου, όπως θα δούμε στην επόμενη παράγραφο. Όμως, και οι ρυθμοί διαπέρασης μπορούν να υπολογιστούν, αλλά η διαδικασία είναι υπολογιστικά απαιτητική. Θα επανέρθουμε στο ζήτημα αυτό στην επόμενη παράγραφο.

11.6 Κατανομή ισορροπίας κλειστών δικτύων Jackson

Οι εξισώσεις ισορροπίας της Μαρκοβιανής αλυσίδας $\{\mathbf{Q}(t)\}$ του αριθμού των πελατών ενός κλειστού δικτύου Jackson έχουν ακριβώς την ίδια μορφή με τις εξισώσεις ισορροπίας για ένα ανοικτό δίκτυο Jackson αν θέσουμε σε αυτές την τιμή 0 για όλους τους εξωτερικούς ρυθμούς αφίξεων και για όλες τις πιθανότητες αναχώρησης πελάτη από το δίκτυο. Επομένως, η κατανομή ισορροπίας ενός κλειστού δικτύου θα είναι πολλαπλάσιο της αντίστοιχης κατανομής ισορροπίας ενός ανοικτού δικτύου στο οποίο θα έχουμε θέσει $\lambda_i = 0$, $p_{i0} = 0$ για $i = 1, 2, \dots, N$ και θα έχουμε κρατήσει ίδιες τις συναρτήσεις ρυθμών εξυπηρέτησης και τις υπόλοιπες πιθανότητες δρομολόγησης πελατών. Για να είναι όμως συνάρτηση πιθανότητας, θα πρέπει να κανονικοποιηθεί ώστε οι

πιθανότητες όλων των καταστάσεων να αθροίζονται στο 1. Βεβαίως, σε αντίθεση με τα ανοικτά δίκτυα Jackson, αυτό είναι πάντα δυνατό, καθώς ο χώρος καταστάσεων ενός κλειστού δικτύου Jackson είναι πεπερασμένος. Με βάση αυτά έχουμε το ακόλουθο αποτέλεσμα.

Θεώρημα 11.10 (Κατανομή ισορροπίας κλειστού δικτύου Jackson) Έστω κλειστό δίκτυο Jackson N σταθμών, με ρυθμούς εξυπηρέτησεων $\mu_j(n_j)$, $n_j \geq 1$, $j = 1, 2, \dots, N$ και πιθανότητες μετάβασης πελατών p_{ij} , $j, i = 1, 2, \dots, N$. Έστω Λ_j , $j = 1, 2, \dots, N$ οι συντεταγμένες μιας λύσης του συστήματος των εξισώσεων κίνησης (11.15). Τότε το δίκτυο είναι ευσταθές και η κατανομή ισορροπίας του είναι

$$p(\mathbf{n}) = p(n_1, n_2, \dots, n_N) = B_M \prod_{i=1}^N \frac{\Lambda_i^{n_i}}{\mu_i(1)\mu_i(2)\cdots\mu_i(n_i)}, \quad \mathbf{n} \in \mathcal{S}_c^M, \quad (11.16)$$

όπου

$$\mathcal{S}_c^M = \{(n_1, n_2, \dots, n_N) : n_i \geq 0, i = 1, 2, \dots, N \text{ και } \sum_{i=1}^N n_i = M\}$$

και

$$B_M = \left(\sum_{\mathbf{n} \in \mathcal{S}_c^M} \prod_{i=1}^N \frac{\Lambda_i^{n_i}}{\mu_i(1)\mu_i(2)\cdots\mu_i(n_i)} \right)^{-1}. \quad (11.17)$$

◇

Η απόδειξη ακολουθεί κατά γράμμα την απόδειξη του θεωρήματος 11.4 για τα ανοικτά δίκτυα Jackson: Οι εξισώσεις ισορροπίας για τη στάσιμη κατανομή είναι οι

$$p(\mathbf{n}) \sum_{j=1}^N \mu_j(n_j) 1_{\{n_j > 0\}} = \sum_{j=1}^N \sum_{i=1}^N p(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j) \mu_i(n_i + 1) p_{ij} 1_{\{n_j > 0\}}, \quad \mathbf{n} \in \mathcal{S}_c^M$$

που γράφονται ισοδύναμα ως

$$\sum_{j=1}^N \left(p(\mathbf{n}) \mu_j(n_j) - \sum_{i=1}^N p(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j) \mu_i(n_i + 1) p_{ij} \right) 1_{\{n_j > 0\}} = 0, \quad \mathbf{n} \in \mathcal{S}_c^M. \quad (11.18)$$

Ορίζοντας για κάθε κατάσταση $\mathbf{n} \in \mathcal{S}_c^M$ τις ποσότητες $\Sigma_j = \Sigma_j(\mathbf{n})$, $1 \leq j \leq N$ από τις σχέσεις

$$\Sigma_j = p(\mathbf{n}) \mu_j(n_j) - \sum_{i=1}^N p(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j) \mu_i(n_i + 1) p_{ij}, \quad 1 \leq j \leq N,$$

η σχέση (11.18) γράφεται $\sum_{j=1}^N \Sigma_j 1_{\{n_j > 0\}} = 0$. Θα αποδείξουμε ότι $\Sigma_j = 0$, για $1 \leq j \leq N$, αντικαθιστώντας την $p(\mathbf{n})$ που δίνεται από την (11.16). Πράγματι, από την (11.16) έχουμε ότι

$$\frac{p(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j)}{p(\mathbf{n} - \mathbf{e}_j)} = \frac{\Lambda_i}{\mu_i(n_i + 1)}.$$

Διαιρώντας με $p(\mathbf{n} - \mathbf{e}_j)$ τη σχέση $\Sigma_j = 0$, για $1 \leq j \leq N$ έχουμε

$$\begin{aligned} \Sigma_j = 0 &\Leftrightarrow \frac{p(\mathbf{n})}{p(\mathbf{n} - \mathbf{e}_j)} \mu_j(n_j) - \sum_{i=1}^N \frac{p(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j)}{p(\mathbf{n} - \mathbf{e}_j)} \mu_i(n_i + 1) p_{ij} = 0 \\ &\Leftrightarrow \frac{\Lambda_j}{\mu_j(n_j)} \mu_j(n_j) - \sum_{i=1}^N \frac{\Lambda_i}{\mu_i(n_i + 1)} \mu_i(n_i + 1) p_{ij} = 0 \\ &\Leftrightarrow \Lambda_j - \sum_{i=1}^N \Lambda_i p_{ij} = 0, \end{aligned}$$

που ισχύει από την εξίσωση κίνησης (11.15) για τον σταθμό j , με $1 \leq j \leq N$.

Θα πρέπει εδώ να παρατηρήσουμε ότι το θεώρημα 11.10 αφήνει κάποια περιθώρια «ελευθερίας» στη διαδικασία εύρεσης της στάσιμης κατανομής. Πράγματι το πρώτο βήμα είναι η εύρεση μιας λύσης του συστήματος των εξισώσεων κίνησης (11.15) και όπως είδαμε το σύστημα αυτό έχει άπειρες λύσεις, όλες πολλαπλάσια του διανύσματος των σχετικών ρυθμών διαπέρασης. Επομένως, στην εφαρμογή του θεωρήματος πρέπει να επιλέξουμε μία από αυτές τις λύσεις και συνήθως επιλέγουμε το διάνυσμα των σχετικών ρυθμών διαπέρασης. Βέβαια, όποια λύση και να επιλέξουμε το αποτέλεσμα θα είναι το ίδιο, καθώς όταν πολλαπλασιάζουμε όλα τα Λ_i με τον ίδιο παράγοντα, διαιρείται η αντίστοιχη σταθερή κανονικοποίησης B_M με τον ίδιο παράγοντα και η τιμή των πιθανοτήτων ισορροπίας $p(\mathbf{n})$ που δίνεται από τον τύπο (11.16) παραμένει η ίδια.

Ένα σημαντικό σημείο θεωρητικής σημασίας που το θεώρημα 11.10 διαφέρει από το αντίστοιχο θεώρημα 11.4 για τα ανοικτά δίκτυα είναι ότι στα κλειστά δίκτυα οι αριθμοί των πελατών στους διάφορους σταθμούς του δικτύου δεν είναι ανεξάρτητες τυχαίες μεταβλητές. Η εξάρτηση εδώ «κρύβεται» στον χώρο καταστάσεων και είναι φανερή αφού το άθροισμα αυτών των τυχαίων μεταβλητών είναι σταθερό και ίσο με τον αριθμό πελατών M .

Από πρακτική σκοπιά, ο προσδιορισμός της κατανομής ισορροπίας ενός κλειστού δικτύου Jackson είναι μια υπολογιστικά δαπανηρή διαδικασία, πολύ πιο «ακριβή» από τον αντίστοιχο υπολογισμό για ανοικτά δίκτυα. Η υπολογιστική δυσκολία έγκειται στον υπολογισμό της σταθεράς κανονικοποίησης B_M που εμφανίζεται στον τύπο της κατανομής ισορροπίας. Όπως βλέπουμε, αυτή δίνεται ως ο αντίστροφος ενός αθροίσματος τόσων όρων όσες και οι δυνατές καταστάσεις του δικτύου. Και εδώ πρέπει να σημειώσουμε ότι το πλήθος των δυνατών καταστάσεων του δικτύου είναι το πλήθος των ακεραίων μη-αρνητικών λύσεων της εξίσωσης $n_1 + n_2 + \dots + n_N = M$ που είναι ίσο με $\binom{M+N-1}{M}$, ένας αριθμός πολύ μεγάλος ακόμα και για μικρές τιμές των N και M .

Τα υπολογιστικά προβλήματα για τον προσδιορισμό της σταθεράς B_M επηρεάζουν και τον υπολογισμό άλλων σημαντικών ποσοτήτων που σχετίζονται με την αποτίμηση απόδοσης ενός κλειστού δικτύου Jackson. Για παράδειγμα, για τον ρυθμό διαπέρασης Λ_i^{TH-c} ενός κλειστού δικτύου έχουμε

$$\begin{aligned} \Lambda_i^{TH-c} &= \sum_{\mathbf{n} \in \mathcal{S}_c^M: n_i > 0} p(\mathbf{n}) \mu_i(n_i) = \sum_{\mathbf{m} \in \mathcal{S}_c^{M-1}} p(\mathbf{m} + \mathbf{e}_i) \mu_i(m_i + 1) \\ &= \sum_{\mathbf{m} \in \mathcal{S}_c^{M-1}} p(\mathbf{m}) \frac{\Lambda_i}{\mu_i(m_i + 1)} \mu_i(m_i + 1) = \Lambda_i \sum_{\mathbf{m} \in \mathcal{S}_c^{M-1}} p(\mathbf{m}) \\ &= \Lambda_i B_M \sum_{\mathbf{m} \in \mathcal{S}_c^{M-1}} \prod_{i=1}^N \frac{\Lambda_i^{m_i}}{\mu_i(1) \mu_i(2) \dots \mu_i(m_i)} = \frac{\Lambda_i B_M}{B_{M-1}}. \end{aligned}$$

Επίσης, σε ένα κλειστό δίκτυο $M/M/1$ ουρών (οπότε $\mu_i(n_i) = \mu_i$), το ποσοστό του χρόνου που ο υπηρέτης του συστήματος i είναι απασχολημένος είναι

$$\begin{aligned} \rho_i^{TH-c} &= \sum_{\mathbf{n} \in \mathcal{S}_c^M: n_i > 0} p(\mathbf{n}) = \sum_{\mathbf{m} \in \mathcal{S}_c^{M-1}} p(\mathbf{m} + \mathbf{e}_i) \\ &= \sum_{\mathbf{m} \in \mathcal{S}_c^{M-1}} p(\mathbf{m}) \frac{\Lambda_i}{\mu_i(m_i + 1)} = \frac{\Lambda_i}{\mu_i} \sum_{\mathbf{m} \in \mathcal{S}_c^{M-1}} p(\mathbf{m}) \\ &= \frac{\Lambda_i}{\mu_i B_{M-1}}. \end{aligned}$$

Λόγω της σημασίας του υπολογισμού της σταθεράς B_M για ένα κλειστό δίκτυο Jackson έχουν αναπτυχθεί υπολογιστικές μέθοδοι που είναι πολύ αποτελεσματικότερες από τον υπολογισμό του αθροίσματος του ορισμού του B_M . Η υπολογιστική πολυπλοκότητα είναι της τάξης του NM^2 . Για σημαντικές κλάσεις κλειστών δικτύων Jackson υπάρχουν ακόμα πιο αποτελεσματικοί αλγόριθμοι. Για λεπτομέρειες ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο σύγγραμμα Kelly 1979, παράγραφο 2.3 (ιδιαίτερα στις ασκήσεις της παραγράφου 2.3).

Για τα κλειστά δίκτυα Jackson ισχύουν ανάλογα αποτελέσματα με αυτά που είδαμε για τα ανοικτά δίκτυα (π.χ. ανάλογα των θεωρημάτων 11.6 και 11.7). Ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο σύγγραμμα του Wolff 1989, κεφάλαιο 6.

11.7 Ασκήσεις

Άσκηση 11.1 Θεωρούμε δυο συστήματα εξυπηρέτησης O_1 και O_2 . Η ουρά O_i έχει Poisson εξωτερική διαδικασία αφίξεων με ρυθμό λ_i , έναν υπηρέτη και εκθετικούς χρόνους εξυπηρέτησης με παράμετρο μ_i , για $i = 1, 2$. Συνδέουμε τις O_1 και O_2 έτσι ώστε κάθε πελάτης που φεύγει από την O_i να πηγαίνει στην άλλη ουρά με πιθανότητα p_i ή να αναχωρεί οριστικά από το σύστημα με πιθανότητα $q_i = 1 - p_i$. Να βρεθεί ικανή και αναγκαία συνθήκη για την ευστάθεια (στασιμότητα) του δικτύου και να υπολογιστεί η αντίστοιχη κατανομή ισορροπίας των αριθμών των πελατών στους σταθμούς του δικτύου.

Άσκηση 11.2 Θεωρούμε ένα δίκτυο πέντε συστημάτων εξυπηρέτησης O_1, O_2, \dots, O_5 . Για κάθε $i = 1, 2, 3, 4$ το σύστημα O_i έχει Poisson εξωτερική διαδικασία αφίξεων με ρυθμό $i\lambda$ και έναν υπηρέτη με εκθετικούς χρόνους εξυπηρέτησης με ρυθμό μ_i . Κάθε πελάτης που ολοκληρώνει την εξυπηρέτησή του στο σύστημα O_i πηγαίνει στο σύστημα O_{i+1} με πιθανότητα $\frac{1}{i}$ ή επαναλαμβάνει την εξυπηρέτησή του με πιθανότητα $\frac{i-1}{i}$ για $i = 1, 2, 3, 4$. Το σύστημα O_5 έχει Poisson διαδικασία εξωτερικών αφίξεων με ρυθμό 5λ και άπειρους υπηρέτες με εκθετικούς χρόνους εξυπηρέτησης με ρυθμό μ_5 και κάθε πελάτης που ολοκληρώνει την εξυπηρέτησή του σε αυτό αναχωρεί από το δίκτυο.

1. Να βρεθεί ικανή και αναγκαία συνθήκη για την ευστάθεια (στασιμότητα) του παραπάνω Μαρκοβιανού δικτύου.
2. Να βρεθεί η κατανομή ισορροπίας των αριθμών των πελατών στους σταθμούς του δικτύου, όταν αυτό είναι ευσταθές.
3. Να υπολογιστεί ο μέσος αριθμός πελατών στο δίκτυο.
4. Να υπολογιστεί ο μέσος χρόνος παραμονής ενός πελάτη στο δίκτυο.
5. Να βρεθεί η κατανομή του αριθμού των επισκέψεων ενός πελάτη στο σύστημα O_i , για $i = 1, 2, 3, 4$.

Άσκηση 11.3 Θεωρούμε ένα δίκτυο τριών συστημάτων O_1, O_2 και O_3 . Το σύστημα O_1 έχει Poisson εξωτερική διαδικασία αφίξεων ρυθμού λ και έναν υπηρέτη με εκθετικούς χρόνους εξυπηρέτησης με ρυθμό μ . Κάθε πελάτης που ολοκληρώνει την εξυπηρέτησή του στο σύστημα O_1 πηγαίνει στο O_2 ή στο O_3 με πιθανότητες $\frac{1}{3}$ και $\frac{2}{3}$, αντίστοιχα. Τα συστήματα O_2 και O_3 έχουν άπειρους υπηρέτες και εκθετικούς χρόνους εξυπηρέτησης με ρυθμό μ . Κάθε πελάτης που ολοκληρώνει την εξυπηρέτησή του στο σύστημα O_i , $i = 2, 3$ αναχωρεί με πιθανότητα $\frac{1}{2}$ από το δίκτυο ή επιστρέφει με πιθανότητα $\frac{1}{2}$ στο ίδιο σύστημα O_i .

1. Να βρεθεί ικανή και αναγκαία συνθήκη για την ευστάθεια (στασιμότητα) του παραπάνω Μαρκοβιανού δικτύου.
2. Να βρεθεί η κατανομή ισορροπίας των αριθμών των πελατών στους σταθμούς του δικτύου, όταν αυτό είναι ευσταθές.
3. Να υπολογιστεί ο μέσος συνολικός αριθμός πελατών στο δίκτυο και ο μέσος αριθμός πελατών στο σύστημα O_2 .
4. Να υπολογιστεί ο μέσος αριθμός εξυπηρέτησεων που θα λάβει ένας πελάτης μέχρι να αναχωρήσει από το δίκτυο.

5. Να υπολογιστεί ο μέσος χρόνος παραμονής ενός πελάτη στο δίκτυο.

Άσκηση 11.4 Θεωρούμε ένα δίκτυο τριών συστημάτων O_1, O_2 και O_3 . Για κάθε $i = 1, 2, 3$, το σύστημα O_i έχει εξωτερική διαδικασία αφίξεων Poisson με ρυθμό $\lambda_i = i$ και i υπηρέτες με εκθετικούς χρόνους εξυπηρέτησης μ_i . Κάθε πελάτης που ολοκληρώνει την εξυπηρέτησή του στο σύστημα O_i αναχωρεί από το δίκτυο με πιθανότητα $\frac{i}{i+1}$, ενώ διαφορετικά πηγαίνει στην O_{i+1} , όπου $O_4 = O_1$.

1. Να βρεθεί ικανή και αναγκαία συνθήκη για την ευστάθεια (στασιμότητα) του παραπάνω Μαρκοβιανού δικτύου.
2. Να βρεθεί η κατανομή ισορροπίας των αριθμών των πελατών στους σταθμούς του δικτύου, όταν αυτό είναι ευσταθές.
3. Να υπολογιστεί η κατανομή του αριθμού των επισκέψεων ενός πελάτη στο σταθμό O_2 , δεδομένου ότι ο πελάτης εισήλθε στο δίκτυο από τον σταθμό O_2 , καθώς και το μέσο πλήθος επισκέψεων σε αυτόν.

Άσκηση 11.5 Θεωρήστε ένα κλειστό δίκτυο Jackson με N σταθμούς, με ρυθμούς εξυπηρέτησεων $\mu_j(n_j)$, $j = 1, 2, \dots, N$ και πιθανότητες μετάβασης πελατών p_{ji} , $j, i = 1, 2, \dots, N$. Έστω Λ_i , $i = 1, 2, \dots, N$ οι συντεταγμένες μιας λύσης του συστήματος των εξισώσεων κίνησης (11.15). Για κάθε δυνατικό αριθμό πελατών M , έστω B_M η σταθερά κανονικοποίησης για την κατανομή ισορροπίας του αριθμού πελατών του δικτύου που δίνεται από τη σχέση (11.17). Ορίζουμε τις γεννήτριες συναρτήσεις

$$\Phi_j(z) = \sum_{n_j=0}^{\infty} \frac{\Lambda_j^{n_j}}{\mu_j(1)\mu_j(2)\cdots\mu_j(n_j)} z^{n_j}, \quad 1 \leq j \leq N,$$

$$B(z) = \sum_{M=0}^{\infty} B_M^{-1} z^M.$$

1. Δείξτε ότι

$$B(z) = \prod_{j=1}^N \Phi_j(z)$$

και επομένως η σταθερά B_M μπορεί να υπολογιστεί πολλαπλασιάζοντας τις συναρτήσεις $\Phi_j(z)$, αφού έχουν όλες περικοπεί στους πρώτους $M+1$ όρους τους. Ο αριθμός των απαιτούμενων βημάτων είναι τάξης NM^2 , και επομένως η μέθοδος αυτή είναι υπολογιστικά πολύ πιο αποτελεσματική από τον ευθύ υπολογισμό του αθροίσματος του ορισμού της B_M (τύπος (11.17)).

2. Αυτή η μέθοδος μπορεί να εφαρμοστεί και για τον υπολογισμό των περιθώριων κατανομών ισορροπίας των αριθμών των πελατών στους διάφορους σταθμούς. Για τον σκοπό αυτό, αποδείξτε ότι αν β_n είναι ο συντελεστής του z^M στη γεννήτρια

$$\frac{\Lambda_i^n z^n}{\mu_i(1)\mu_i(2)\cdots\mu_i(n)} \prod_{j \neq i} \Phi_j(z),$$

τότε η πιθανότητα ισορροπίας ο σταθμός i να έχει n πελάτες είναι $\beta_n B_M$.

3. Στην περίπτωση που κάθε σταθμός είναι τύπου $M/M/1$ να δείξετε πώς διαμορφώνονται οι συναρτήσεις $\Phi_j(z)$ και πώς απλοποιείται η διαδικασία υπολογισμού της σταθεράς B_M . Επιπλέον, δείξτε ότι στην περίπτωση αυτή η πιθανότητα ο σταθμός i να περιέχει τουλάχιστον n πελάτες είναι

$$\frac{\Lambda_i^n B_M}{\mu_i^n B_{M-n}}.$$

11.8 Σχόλια

Τα δίκτυα ουρών αναμονής αποτελούν μια ιδιαίτερα ενδιαφέρουσα και δύσκολη υποπεριοχή της Θεωρίας Ουρών Αναμονής. Καθώς τα περισσότερα συστήματα εξυπηρέτησης απαιτούν οι πελάτες να περάσουν από περισσότερα του ενός σημεία εξυπηρέτησης, τα δίκτυα ουρών αναφέρονται παντού. Όμως, οι υποκείμενες στοχαστικές διαδικασίες είναι ιδιαίτερα δύσκολο να μελετηθούν και τις περισσότερες φορές η εύρεση κλειστών τύπων για τα βασικά μέτρα απόδοσής τους είναι αδύνατη. Μια εξαίρεση σε αυτόν τον κανόνα είναι τα δίκτυα Jackson που παρουσιάστηκαν στο παρόν κεφάλαιο.

Για πολλά χρόνια το ενδιαφέρον των ερευνητών εστιάστηκε στη μελέτη γενικεύσεων των δικτύων Jackson που επίσης παρουσιάζουν κάποιο είδος κατανομής ισορροπίας μορφής γινομένου όπως αυτή που είδαμε στο θεώρημα 11.4. Έγινε αντιληπτό ότι η μορφή γινομένου για την κατανομή ισορροπίας συνδέεται με κάποιες γενικεύσεις της αντιστρεψιμότητας. Τα πρώτα αποτελέσματα αυτής της θεώρησης συνοψίζονται στο κλασικό σύγγραμμα του Kelly 1979. Από τότε σημειώθηκε τεράστια πρόοδος η οποία κατέληξε στη δημιουργία μιας θεωρίας που χαρακτηρίζει τα δίκτυα τα οποία έχουν κατανομή ισορροπίας μορφής γινομένου και προσδιορίζει κάποιες σημαντικές ιδιότητές τους. Τα σχετικά αποτελέσματα συνοψίζονται στο Chao, Miyazawa και Pinedo 1999. Άλλα βιβλία που επικεντρώνονται στο ίδιο θέμα είναι τα Dijk 1993 και Serfozo 1999.

Βέβαια, η μελέτη των δικτύων συστημάτων εξυπηρέτησης είναι μια πολύ ευρεία περιοχή με πολλές άλλες διαστάσεις. Αυτές παρουσιάζονται σε πολλά αξιόλογα εισαγωγικά βιβλία, όπως των Bolch, Greiner, Meer και Trivedi 1998, Gelenbe και Pujolle 1998 και Chen και Yao 2001.

Βιβλιογραφία

- [1] F.P. Kelly. *Reversibility and Stochastic Networks*. Chichester: Wiley, 1979. ISBN: 978-0471276012.
- [2] R.W. Wolff. *Stochastic Modeling and the Theory of Queues*. Englewood Cliffs, New Jersey: Prentice Hall, 1989. ISBN: 978-0138466923.
- [3] X. Chao, M. Miyazawa και M. Pinedo. *Queueing Networks: Customers, Signals and Product Form Solutions*. Wiley, 1999. ISBN: 978-0471985706.
- [4] N.M. van Dijk. *Queueing Networks and Product Forms: A Systems Approach*. Chichester: Wiley, 1993. ISBN: 978-0471928488.
- [5] R. Serfozo. *Introduction to Stochastic Networks*. Springer, 1999. ISBN: 978-0387987736.
- [6] G. Bolch, S. Greiner, H. de Meer και K.S. Trivedi. *Queueing Networks and Markov Chains*. Wiley, 1998. ISBN: 978-0471193666.
- [7] E. Gelenbe και G. Pujolle. *Introduction to Queueing Networks, 2nd Edition*. Chichester: Wiley, 1998. ISBN: 978-0471962945.
- [8] H. Chen και D.D. Yao. *Fundamentals of Queueing Networks*. Springer, 2001. ISBN: 978-1475753011.

Μέρος III

ΑΠΟΤΙΜΗΣΗ ΑΠΟΔΟΣΗΣ:
ΜΗ-ΜΑΡΚΟΒΙΑΝΑ ΜΟΝΤΕΛΑ

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΓΙΑ ΜΗ-ΜΑΡΚΟΒΙΑΝΑ ΜΟΝΤΕΛΑ

Ένα σύστημα εξυπηρέτησης λέγεται μη-Μαρκοβιανό, αν ο αριθμός των πελατών σε αυτό δεν είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου, ούτε μπορεί να εμφυτευτεί σε κάποια Μαρκοβιανή αλυσίδα συνεχούς χρόνου με τον τρόπο που είδαμε στα κεφάλαια 8, 9 και 10. Σε αυτές τις περιπτώσεις, χρησιμοποιούνται κάποιες άλλες κατηγορίες στοχαστικών διαδικασιών για να προχωρήσει η μελέτη των συστημάτων. Οι δυο πιο βασικές από αυτές είναι οι Μαρκοβιανές διαδικασίες διακριτού χρόνου και οι ανανεωτικές διαδικασίες. Στο υπόλοιπο του κεφαλαίου θα γίνει μια ανασκόπηση αυτών των διαδικασιών.

12.1 Μαρκοβιανές αλυσίδες διακριτού χρόνου

Πολλά μη-Μαρκοβιανά συστήματα εξυπηρέτησης παρουσιάζουν τη Μαρκοβιανή ιδιότητα, δηλαδή την ανεξαρτησία του μέλλοντος από το παρελθόν δεδομένου του παρόντος, αν εξεταστούν σε κατάλληλα επιλεγμένες χρονικές στιγμές, συνήθως σε στιγμές αφίξεων ή περατώσεων εξυπηρετήσεων. Επομένως αξίζει να μελετήσουμε διεξοδικότερα τις εμφυτευμένες στοχαστικές διαδικασίες σε στιγμές αφίξεων ή αναχωρήσεων που είναι διαδικασίες διακριτού χρόνου με διακριτό χώρο καταστάσεων. Το πιο απλό μοντέλο για την περιγραφή της κατάστασης ενός στοχαστικού δυναμικού συστήματος με διακριτό χώρο καταστάσεων σε διακριτό χρόνο είναι μια Μαρκοβιανή αλυσίδα διακριτού χρόνου, που είναι το ανάλογο διακριτού χρόνου των Μαρκοβιανών αλυσίδων συνεχούς χρόνου. Πιο συγκεκριμένα, έχουμε τον ακόλουθο ορισμό.

Ορισμός 12.1 (Μαρκοβιανή αλυσίδα διακριτού χρόνου) *Μια στοχαστική διαδικασία $\{X_n : n \geq 0\}$ είναι Μαρκοβιανή αλυσίδα διακριτού χρόνου με χώρο καταστάσεων \mathcal{S} αν*

(i) $X_n \in \mathcal{S}$, $n \geq 0$, και \mathcal{S} αριθμήσιμο,

(ii) $\Pr[X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i] = \Pr[X_{n+1} = j | X_n = i]$, $n \geq 0$, $i_0, \dots, i_{n-1}, i, j \in \mathcal{S}$
(Μαρκοβιανή ιδιότητα).

Αν, επιπλέον, οι πιθανότητες $\Pr[X_{n+1} = j | X_n = i]$ δεν εξαρτώνται από το n , η Μαρκοβιανή αλυσίδα λέγεται ομογενής. Συμβολίζουμε τότε τις πιθανότητες αυτές με p_{ij} και ο πίνακας $\mathbf{P} = (p_{ij})$ αναφέρεται ως πίνακας πιθανοτήτων μετάβασης της Μαρκοβιανής αλυσίδας.

Η συνάρτηση πιθανότητας $\pi^{(0)} = (\pi_i^{(0)})$ με $\pi_i^{(0)} = \Pr[X_0 = i]$ αναφέρεται ως αρχική κατανομή της Μαρκοβιανής αλυσίδας. \diamond

Σε ό,τι ακολουθεί θα περιοριστούμε σε ομογενείς Μαρκοβιανές αλυσίδες. Η συμπεριφορά μιας Μαρκοβιανής αλυσίδας προσδιορίζεται πλήρως από την αρχική της κατανομή και τον πίνακα πιθανοτήτων μετάβασης της. Σε πρώτο επίπεδο, μας ενδιαφέρουν οι υπολογισμοί που αφορούν την εξέλιξη μιας Μαρκοβιανής αλυσίδας σε πεπερασμένο χρόνο. Ιδιαίτερα, μας ενδιαφέρει ο προσδιορισμός των πιθανοτήτων μετάβασης n -οστής τάξης, $p_{ij}^{(n)} = \Pr[X_n = j | X_0 = i]$, καθώς και ο προσδιορισμός των πιθανοτήτων $\pi_j^{(n)} = \Pr[X_n = j]$. Ο πίνακας $\mathbf{P}^{(n)} = (p_{ij}^{(n)})$ αναφέρεται ως πίνακας πιθανοτήτων μετάβασης n -οστής τάξης της Μαρκοβιανής αλυσίδας, ενώ η συνάρτηση πιθανότητας $\pi^{(n)} = (\pi_i^{(n)})$ με $\pi_i^{(n)} = \Pr[X_n = i]$ αναφέρεται ως η μεταβατική κατανομή της Μαρκοβιανής αλυσίδας τη στιγμή n . Έχουμε το ακόλουθο αποτέλεσμα:

Θεώρημα 12.2 (Βασικοί υπολογισμοί σε πεπερασμένη χρονική στιγμή) Έστω Μαρκοβιανή αλυσίδα διακριτού χρόνου $\{X_n : n \geq 0\}$ με χώρο καταστάσεων \mathcal{S} , αρχική κατανομή $\pi^{(0)} = (\pi_i^{(0)})$ και πίνακα πιθανοτήτων μετάβασης $\mathbf{P} = (p_{ij})$. Τότε έχουμε

(i) Η πιθανότητα πραγματοποίησης ενός συγκεκριμένου μονοπατιού $i_0 \rightarrow i_1 \rightarrow \dots \rightarrow i_n$ είναι

$$\Pr[X_0 = i_0, X_1 = i_1, \dots, X_n = i_n] = \pi_{i_0}^{(0)} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n}.$$

(ii) Οι πιθανότητες μετάβασης n -οστής τάξης υπολογίζονται αναδρομικά από τις σχέσεις

$$\begin{aligned} p_{ij}^{(0)} &= \delta_{ij} = \begin{cases} 1, & \text{αν } i = j, \\ 0, & \text{αν } i \neq j, \end{cases} \\ p_{ij}^{(1)} &= p_{ij}, \\ p_{ij}^{(n)} &= \sum_{r \in \mathcal{S}} p_{ir}^{(k)} p_{rj}^{(n-k)}, \quad n \geq 2, 1 \leq k \leq n-1, \end{aligned}$$

που γράφονται σε πίνακική μορφή

$$\begin{aligned} \mathbf{P}^{(0)} &= \mathbf{I}, \\ \mathbf{P}^{(1)} &= \mathbf{P}, \\ \mathbf{P}^{(n)} &= \mathbf{P}^{(k)} \mathbf{P}^{(n-k)}, \quad n \geq 2, 1 \leq k \leq n-1. \end{aligned}$$

Επομένως,

$$\mathbf{P}^{(n)} = \mathbf{P}^n, \quad n \geq 0.$$

(iii) Οι μεταβατικές πιθανότητες υπολογίζονται από τη σχέση

$$\pi_j^{(n)} = \sum_{i \in \mathcal{S}} \pi_i^{(0)} p_{ij}^{(n)},$$

που γράφεται σε πίνακική μορφή ως

$$(\pi^{(n)})^T = (\pi^{(0)})^T \mathbf{P}^n = (\pi^{(0)})^T \mathbf{P}^n, \quad n \geq 0.$$

Το (i) αποδεικνύεται εύκολα χρησιμοποιώντας τον πολλαπλασιαστικό τύπο για τον υπολογισμό της πιθανότητας τομής ενδεχομένων και εφαρμόζοντας τη Μαρκοβιανή ιδιότητα. Οι σχέσεις για τις $p_{ij}^{(0)}$ και $p_{ij}^{(1)}$ είναι προφανείς, ενώ η σχέση για τις $p_{ij}^{(n)}$, $n \geq 2$, προκύπτει με δέσμευση στην X_k από το θεώρημα ολικής πιθανότητας. Ομοίως και το (iii) προκύπτει δεσμεύοντας στην X_0 , χρησιμοποιώντας το θεώρημα ολικής πιθανότητας.

Δοθείσης μιας Μαρκοβιανής αλυσίδας διακριτού χρόνου και δυο καταστάσεων i, j λέμε ότι η j είναι προσπελάσιμη από την i (συμβολικά $i \rightarrow j$), αν υπάρχει $n \geq 0$ τέτοιο ώστε $p_{ij}^{(n)} > 0$. Αν η i είναι προσπελάσιμη από την j , και, αντίστροφα, η j είναι προσπελάσιμη από την i , λέμε ότι οι i, j επικοινωνούν (συμβολικά $i \leftrightarrow j$). Η επικοινωνία είναι σχέση ισοδυναμίας και επομένως ο χώρος καταστάσεων διαμερίζεται σε κλάσεις επικοινωνίας (αδιαχώριστες κλάσεις), όπου όλες οι καταστάσεις μιας κλάσης επικοινωνούν μεταξύ τους. Μια Μαρκοβιανή αλυσίδα διακριτού χρόνου λέγεται αδιαχώριστη, αν έχει μόνο μια κλάση επικοινωνίας, δηλαδή όλες οι καταστάσεις επικοινωνούν μεταξύ τους, διαφορετικά λέγεται διαχωρίσιμη. Μια κλάση επικοινωνίας $\mathcal{C} \subseteq \mathcal{S}$ λέγεται κλειστή αν δεν υπάρχουν $i \in \mathcal{C}$ και $j \notin \mathcal{C}$, με $p_{ij} > 0$, αλλιώς λέγεται ανοικτή. Αν μια Μαρκοβιανή αλυσίδα εισέλθει κάποια στιγμή σε μια κλειστή κλάση, λέμε ότι απορροφήθηκε σε αυτήν, αφού είναι αδύνατο να ξαναβγει. Δοθέντος ενός συνόλου καταστάσεων μιας Μαρκοβιανής αλυσίδας διακριτού χρόνου, μπορούμε να υπολογίσουμε τις πιθανότητες εισόδου σε αυτό, ξεκινώντας από οποιαδήποτε κατάσταση της Μαρκοβιανής αλυσίδας, καθώς και τους αντίστοιχους μέσους χρόνους πρώτης εισόδου. Στην περίπτωση που το σύνολο καταστάσεων είναι μια κλειστή κλάση επικοινωνίας, η Μαρκοβιανή αλυσίδα δεν θα βγει από το σύνολο από τη στιγμή που θα εισέλθει σε αυτό, και για το λόγο αυτό μιλάμε για πιθανότητες απορρόφησης και μέσους χρόνους απορρόφησης. Έχουμε το ακόλουθο αποτέλεσμα.

Θεώρημα 12.3 (Πιθανότητες εισόδου/απορρόφησης και μέσοι χρόνοι πρώτης εισόδου/απορρόφησης)
Έστω $\{X_n : n \geq 0\}$ Μαρκοβιανή αλυσίδα διακριτού χρόνου, με χώρο καταστάσεων \mathcal{S} και πίνακα πιθανοτήτων μετάβασης $\mathbf{P} = (p_{ij})$. Έστω, επίσης, $\mathcal{C} \subseteq \mathcal{S}$. Ορίζουμε

$$T_{\mathcal{C}} = \inf\{n \geq 0 : X_n \in \mathcal{C}\},$$

τον χρόνο πρώτης εισόδου ή απορρόφησης της Μαρκοβιανής αλυσίδας στο \mathcal{C} . Επίσης, έστω

$$h_i(\mathcal{C}) = \Pr[T_{\mathcal{C}} < \infty | X_0 = i], \quad i \in \mathcal{S},$$

η πιθανότητα εισόδου/απορρόφησης στο \mathcal{C} , ξεκινώντας από την κατάσταση i , και

$$m_i(\mathcal{C}) = E[T_{\mathcal{C}} | X_0 = i], \quad i \in \mathcal{S},$$

ο μέσος χρόνος πρώτης εισόδου/απορρόφησης στο \mathcal{C} , ξεκινώντας από την κατάσταση i . Τότε:

(i) $H \mathbf{h}(\mathcal{C}) = (h_i(\mathcal{C}))$ είναι η ελάχιστη μη-αρνητική λύση του γραμμικού συστήματος

$$x_i = 1, \quad i \in \mathcal{C}, \quad (12.1)$$

$$x_i = \sum_{j \in \mathcal{S}} p_{ij} x_j, \quad i \in \mathcal{S} \setminus \mathcal{C}. \quad (12.2)$$

(ii) $H \mathbf{m}(\mathcal{C}) = (m_i(\mathcal{C}))$ είναι η ελάχιστη μη-αρνητική λύση του γραμμικού συστήματος

$$y_i = 0, \quad i \in \mathcal{C}, \quad (12.3)$$

$$y_i = 1 + \sum_{j \in \mathcal{S}} p_{ij} y_j, \quad i \in \mathcal{S} \setminus \mathcal{C}. \quad (12.4)$$

Είναι άμεσο να δούμε ότι οι πιθανότητες και οι μέσοι χρόνοι απορρόφησης ικανοποιούν τις παραπάνω εξισώσεις (12.1)-(12.4), δεσμεύοντας στην πρώτη μετάβαση της Μαρκοβιανής αλυσίδας. Πράγματι, ξεκινώντας από μια κατάσταση $i \in \mathcal{E}$, η απορρόφηση είναι βέβαια και ο μέσος χρόνος απορρόφησης είναι μηδενικός. Αν $i \notin \mathcal{E}$, τότε στο πρώτο βήμα η Μαρκοβιανή αλυσίδα θα μεταβεί σε κάποια κατάσταση j και θα πρέπει ξεκινώντας από αυτήν να απορροφηθεί στο \mathcal{E} . Το γιατί οι πιθανότητες και οι μέσοι χρόνοι απορρόφησης είναι η ελάχιστη λύση των παραπάνω εξισώσεων απαιτεί περισσότερη εργασία.

Δοθείσης μιας κατάστασης j μιας Μαρκοβιανής αλυσίδας, ορίζουμε h_j την πιθανότητα επανόδου σε αυτήν και με m_j τον αντίστοιχο μέσο χρόνο επανόδου. Για να υπολογίσουμε αυτές τις ποσότητες, θεωρούμε τη Μαρκοβιανή αλυσίδα που προκύπτει αν τροποποιήσουμε την αρχική Μαρκοβιανή αλυσίδα, κάνοντας την κατάσταση j απορροφητική. Τότε, μπορούμε να υπολογίσουμε σε αυτήν τη νέα αλυσίδα τις πιθανότητες απορρόφησης $h_i(\{j\})$ και τους μέσους χρόνους απορρόφησης $m_i(\{j\})$ και είναι

$$h_j = \sum_{i \in \mathcal{S}} p_{ji} h_i(\{j\}),$$

$$m_j = 1 + \sum_{i \in \mathcal{S}} p_{ji} m_i(\{j\}).$$

Ορισμός 12.4 (Επαναληπτικότητα - παροδικότητα καταστάσεων) Έστω Μαρκοβιανή αλυσίδα $\{X_n : n \geq 0\}$ και κατάστασή της, j , με πιθανότητα επανόδου h_j και μέσο χρόνο επανόδου m_j .

- (i) H_j λέγεται θετικά επαληπτική, αν $h_j = 1$ και $m_j < \infty$.
- (ii) H_j λέγεται μηδενικά επαληπτική, αν $h_j = 1$ και $m_j = \infty$.
- (iii) H_j λέγεται παροδική, αν $h_j < 1$.

◊

Οι ιδιότητες της θετικής επαληπτικότητας, της μηδενικής επαληπτικότητας και της παροδικότητας είναι ιδιότητες των κλάσεων επικοινωνίας, δηλαδή, αν οι καταστάσεις i και j επικοινωνούν τότε είναι ίδιου τύπου ως προς την θετική επαληπτικότητα, μηδενική επαληπτικότητα και παροδικότητα.

Ισχύει επίσης ότι οι καταστάσεις μιας ανοικτής κλάσης επικοινωνίας είναι παροδικές, ενώ οι καταστάσεις μιας πεπερασμένης κλειστής κλάσης επικοινωνίας είναι θετικά επαληπτικές. Επομένως, η μόνη περίπτωση που απαιτεί περαιτέρω διερεύνηση είναι αυτή της άπειρης κλειστής κλάσης. Εκεί, είναι ανοικτά όλα τα ενδεχόμενα: Μπορεί οι καταστάσεις της να είναι όλες παροδικές ή όλες μηδενικά επαληπτικές ή όλες θετικά επαληπτικές.

Μια άλλη ιδιότητα μιας κατάστασης είναι η περιοδικότητα/απεριοδικότητα. Έχουμε τον ακόλουθο ορισμό:

Ορισμός 12.5 (Περιοδικότητα καταστάσεων) Έστω Μαρκοβιανή αλυσίδα $\{X_n : n \geq 0\}$ και κατάστασή της, j , με

$$d_j = \text{MK}\Delta\{n : p_{jj}^{(n)} > 0\}.$$

H_j λέγεται απεριοδική αν $d_j = 1$.

H_j λέγεται περιοδική με περίοδο d_j αν $d_j > 1$.

◊

Αν $p_{jj} > 0$ τότε έχουμε προφανώς ότι η j είναι απεριοδική. Αλλά η συνθήκη αυτή είναι ικανή και όχι αναγκαία. Επίσης, οι ιδιότητες της απεριοδικότητας και της περιοδικότητας είναι ιδιότητες των κλάσεων επικοινωνίας, δηλαδή, αν οι καταστάσεις i και j επικοινωνούν τότε $d_i = d_j$.

Προχωράμε, τώρα, στην παρουσίαση της οριακής συμπεριφοράς των Μαρκοβιανών αλυσίδων διακριτού χρόνου. Δεδομένου ότι οι περισσότερες εφαρμογές αφορούν αδιαχώριστες αλυσίδες, θα περιοριστούμε στην περίπτωση αυτή, που είναι και η πιο σημαντική. Έχουμε το ακόλουθο αποτέλεσμα.

Θεώρημα 12.6 (Οριακή συμπεριφορά Μαρκοβιανών αλυσίδων διακριτού χρόνου) Έστω $\{X_n : n \geq 0\}$ μια αδιαχώριστη Μαρκοβιανή αλυσίδα διακριτού χρόνου, με χώρο καταστάσεων \mathcal{S} και πίνακα πιθανοτήτων μετάβασης $\mathbf{P} = (p_{ij})$. Η $\{X_n : n \geq 0\}$ είναι θετικά επαναληπτική, αν και μόνο αν, το σύστημα των εξισώσεων ισορροπίας

$$\pi_j = \sum_{i \in \mathcal{S}} \pi_i p_{ij}, \quad j \in \mathcal{S}, \quad (12.5)$$

έχει λύση που ικανοποιεί και την εξίσωση κανονικοποίησης

$$\sum_{j \in \mathcal{S}} \pi_j = 1.$$

Αν υπάρχει τέτοια λύση $\pi = (\pi_j)$, τότε είναι μοναδική. Επιπλέον, όλες της οι συντεταγμένες είναι θετικές και κάθε άλλη λύση $\mathbf{x} = (x_j)$ του συστήματος των εξισώσεων ισορροπίας είναι πολλαπλάσιό της: $\mathbf{x} = c\pi$, για κάποιο $c \in \mathbb{R}$. Έχουμε:

(i) Η πιθανότητα π_j εκφράζει το μακροπρόθεσμο ποσοστό του χρόνου που η Μαρκοβιανή αλυσίδα περνάει στην κατάσταση j , δηλαδή,

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=0}^{n-1} 1_{\{X_k=j\}}}{n} = \pi_j, \quad \text{με πιθανότητα } 1, \quad j \in \mathcal{S}.$$

Δηλαδή, η π είναι η οριακή δειγματική ή οριακή εμπειρική κατανομή της $\{X_n\}$.

(ii) Η πιθανότητα π_j εκφράζει το μακροπρόθεσμο μέσο ποσοστό του χρόνου που η Μαρκοβιανή αλυσίδα περνάει στην κατάσταση j , δηλαδή,

$$\lim_{n \rightarrow \infty} \frac{E[\sum_{k=0}^{n-1} 1_{\{X_k=j\}}]}{n} = \pi_j, \quad j \in \mathcal{S}.$$

(iii) Η πιθανότητα π_j ισούται με τον αντίστροφο του μέσου χρόνου επανόδου στην κατάσταση j :

$$\pi_j = \frac{1}{m_j}, \quad j \in \mathcal{S}.$$

(iv) Η πιθανότητα π_j είναι η C-οριακή πιθανότητα (Cesaro οριακή πιθανότητα) της κατάστασης j :

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=0}^{n-1} \pi_j^{(k)}}{n} = \pi_j, \quad j \in \mathcal{S}.$$

Δηλαδή, η π είναι η C-οριακή κατανομή της $\{X_n\}$.

(v) Αν η Μαρκοβιανή αλυσίδα είναι και απεριοδική, τότε η πιθανότητα π_j είναι η οριακή πιθανότητα της κατάστασης j :

$$\lim_{n \rightarrow \infty} \pi_j^{(n)} = \pi_j, \quad j \in \mathcal{S}.$$

Δηλαδή, η π είναι η οριακή κατανομή της $\{X_n\}$.

(vi) Αν η Μαρκοβιανή αλυσίδα έχει αρχική κατανομή την π , τότε κάθε μεταβατική κατανομή της δίνεται πάλι από την π :

$$\pi^{(0)} = \pi \Rightarrow \pi^{(n)} = \pi, \quad n \geq 0.$$

Δηλαδή, η π είναι η στάσιμη κατανομή της $\{X_n\}$.

Από το θεώρημα 12.6 γίνεται φανερό ότι η κατανομή π φέρει τη σημαντικότερη πληροφορία για τη μακροπρόθεσμη συμπεριφορά μιας Μαρκοβιανής αλυσίδας διακριτού χρόνου. Αναφέρεται ως οριακή εμπειρική, C-οριακή, οριακή και στάσιμη κατανομή της Μαρκοβιανής αλυσίδας, καθώς και ως κατανομή ισορροπίας. Αν η Μαρκοβιανή αλυσίδα έχει τη στάσιμη κατανομή της ως αρχική, οι μεταβατικές της κατανομές είναι όλες ίσες με τη στάσιμη (βλέπε (vi)) και η Μαρκοβιανή αλυσίδα λέγεται στάσιμη ή λέμε ότι βρίσκεται σε κατάσταση στατιστικής ισορροπίας. Διαισθητικά, αυτό σημαίνει ότι έχει περάσει αρκετός χρόνος που το στοχαστικό σύστημα λειτουργεί και η επίδραση της αρχικής του κατάστασης έχει χαθεί και παρουσιάζει ομοιόμορφη συμπεριφορά στο χρόνο. Για τα (ii), (iv) και (v) ισχύουν και οι αντίστοιχοι τύποι που αφορούν τις δεσμευμένες ποσότητες ως προς την αρχική κατάσταση. Δηλαδή, έχουμε:

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{E[\sum_{k=0}^{n-1} 1_{\{X_k=j\}} | X_0 = i]}{n} &= \pi_j, \quad i, j \in \mathcal{S}, \\ \lim_{n \rightarrow \infty} \frac{\sum_{k=0}^{n-1} p_{ij}^{(k)}}{n} &= \pi_j, \quad i, j \in \mathcal{S}, \\ \lim_{n \rightarrow \infty} p_{ij}^{(n)} &= \pi_j, \quad i, j \in \mathcal{S}.\end{aligned}$$

Μια διαισθητική ερμηνεία του (iv) είναι ότι η στάσιμη πιθανότητα π_j εκφράζει την πιθανότητα η Μαρκοβιανή αλυσίδα να βρίσκεται στην κατάσταση j μια χρονική στιγμή k που έχει επιλεγεί ομοιόμορφα από το χρονικό διάστημα $\{0, 1, 2, \dots, n-1\}$, για μεγάλο n . Δηλαδή, ισχύει

$$\lim_{n \rightarrow \infty} \Pr[X_{K_n} = j] = \pi_j, \quad j \in \mathcal{S},$$

όπου η K_n είναι τυχαία μεταβλητή Uniform($\{0, 1, 2, \dots, n-1\}$) (δηλαδή διακριτή ομοιόμορφη στο $\{0, 1, 2, \dots, n-1\}$) και ανεξάρτητη της $\{X_n : n \geq 0\}$. Πράγματι

$$\begin{aligned}\lim_{n \rightarrow \infty} \Pr[X_{K_n} = j] &= \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} \Pr[K_n = k] \Pr[X_{K_n} = j | K_n = k] \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} \frac{1}{n} \Pr[X_k = j] \\ &= \lim_{n \rightarrow \infty} \frac{\sum_{k=0}^{n-1} \pi_j^{(k)}}{n} = \pi_j, \quad j \in \mathcal{S}.\end{aligned}$$

Τα περισσότερα από τα συμπεράσματα του Θεωρήματος 12.6 μπορούν να δικαιολογηθούν διαισθητικά, παρότι οι αυστηρές αποδείξεις τους δεν είναι απλές. Π.χ. για μια διαισθητική αιτιολόγηση των (i), (ii), ας συμβολίσουμε με π_j το μακροπρόθεσμο ποσοστό του χρόνου που η Μαρκοβιανή αλυσίδα περνάει στην κατάσταση j . Τότε η ποσότητα $\pi_j p_{ji}$ εκφράζει το μακροπρόθεσμο ρυθμό μεταβάσεων τύπου $j \rightarrow i$, δηλαδή,

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=0}^{n-1} 1_{\{X_k=j, X_{k+1}=i\}}}{n} = \pi_j p_{ji}, \quad j, i \in \mathcal{S}.$$

Αυτό είναι φανερό αφού η p_{ji} εκφράζει το μακροπρόθεσμο ποσοστό των φορών που η αλυσίδα μετακινείται προς την i , όταν βρίσκεται στην j . Επομένως, η ποσότητα $\sum_{i \in \mathcal{S}} \pi_j p_{ji}$ εκφράζει το μακροπρόθεσμο ρυθμό αναχωρήσεων από την κατάσταση j , ενώ η ποσότητα $\sum_{i \in \mathcal{S}} \pi_i p_{ij}$ εκφράζει το μακροπρόθεσμο ρυθμό αφίξεων στην κατάσταση j . Αυτοί οι δυο ρυθμοί θα πρέπει να είναι ίσοι, αφού κάθε άφιξη στην j ακολουθείται από μια αναχώρηση από αυτήν, δηλαδή αφίξεις και αναχωρήσεις στην j βρίσκονται σε μια 1-1 αντιστοιχία. Οπότε πρέπει να ισχύει

$$\pi_j = \sum_{i \in \mathcal{S}} \pi_j p_{ji} = \sum_{i \in \mathcal{S}} \pi_i p_{ij}, \quad j \in \mathcal{S},$$

δηλαδή τα μακροπρόθεσμα ποσοστά του χρόνου π_j που η Μαρκοβιανή αλυσίδα περνάει στις διάφορες καταστάσεις πρέπει να ικανοποιούν τις εξισώσεις ισορροπίας.

Για το (iii), παρατηρούμε ότι οι χρόνοι μεταξύ διαδοχικών επισκέψεων σε κάθε συγκεκριμένη κατάσταση j μιας Μαρκοβιανής αλυσίδας είναι ανεξάρτητοι και ισόνομοι, οπότε οι στιγμές επισκέψεων στην j αντιστοιχούν στα γεγονότα μιας ανανεωτικής διαδικασίας και μάλιστα είναι αναγεννητικές στιγμές για τη Μαρκοβιανή αλυσίδα. Εφαρμόζοντας το εργοδικό θεώρημα των αναγεννητικών διαδικασιών (θεώρημα 1.5), έχουμε ότι ο μακροπρόθεσμος ρυθμός επισκέψεων στην j , δηλαδή η π_j , ισούται με τον αντίστροφο του μέσου ενδιάμεσου χρόνου της ανανεωτικής διαδικασίας, που στην περίπτωση μας είναι ο μέσος χρόνος επανόδου στη j , m_j . Ουσιαστικά, η σχέση αυτή λέει ότι η συχνότητα των επισκέψεων στην j είναι το αντίστροφο μιας περιόδου μεταξύ δυο διαδοχικών επισκέψεων στην j .

Το (iv) είναι μια απλή αναδιατύπωση του (ii), που προκύπτει αφού μπορούμε να εναλλάξουμε τη μέση τιμή με το άθροισμα. Για το (v), είναι γνωστό ότι αν το όριο μιας ακολουθίας υπάρχει, τότε υπάρχει και το C-όριο της και είναι ίσα. Επομένως, αν το όριο της $\pi_j^{(n)}$ υπάρχει, τότε θα είναι αναγκαστικά ίσο με π_j , λόγω του (iv). Η απεριοδικότητα είναι ικανή και αναγκαία συνθήκη για να υπάρχει αυτό το όριο.

Τέλος, υπολογίζοντας τις πιθανότητες $\pi_j^{(n+1)} = \Pr[X_{n+1} = j]$, δεσμεύοντας στην X_n , έχουμε

$$\pi_j^{(n+1)} = \sum_{i \in \mathcal{S}} \pi_i^{(n)} p_{ij}, \quad j \in \mathcal{S}. \quad (12.6)$$

Οπότε, αν το όριο $\lim_{n \rightarrow \infty} \pi_j^{(n)}$ υπάρχει θα έχουμε

$$\lim_{n \rightarrow \infty} \pi_j^{(n+1)} = \lim_{n \rightarrow \infty} \sum_{i \in \mathcal{S}} \pi_i^{(n)} p_{ij} = \sum_{i \in \mathcal{S}} \lim_{n \rightarrow \infty} \pi_i^{(n)} p_{ij}, \quad j \in \mathcal{S},$$

που δείχνει ότι οι οριακές πιθανότητες $\lim_{n \rightarrow \infty} \pi_j^{(n)}$ ικανοποιούν το σύστημα των εξισώσεων ισορροπίας (12.5).

Επίσης, οι εξισώσεις (12.6), σε συνδυασμό με τις εξισώσεις ισορροπίας (12.5), δείχνουν ότι, αν $\pi_i^{(0)} = \pi_i$, $i \in \mathcal{S}$, τότε $\pi_i^{(1)} = \pi_i$, $i \in \mathcal{S}$, και επαγωγικά έχουμε $\pi_i^{(n)} = \pi_i$, $i \in \mathcal{S}$, για $n \geq 0$, οπότε έχουμε το (vi).

Σημειώνουμε, επίσης, ότι, αν ο χώρος καταστάσεων μιας αδιαχώριστης Μαρκοβιανής αλυσίδας διακριτού χρόνου είναι πεπερασμένος, τότε υπάρχει πάντοτε η στάσιμη κατανομή. Αυτό συμβαίνει, διότι, όπως έχουμε αναφέρει, οι κλειστές πεπερασμένες κλάσεις επικοινωνίας είναι πάντα θετικά επαναληπτικές.

Μια άλλη χρήσιμη παρατήρηση είναι ότι η στάσιμη κατανομή π ικανοποιεί, επίσης, και τις λεγόμενες εξισώσεις γενικευμένης ισορροπίας που απαιτούν ο μακροπρόθεσμος ρυθμός μεταβάσεων που εξέρχονται από ένα σύνολο καταστάσεων \mathcal{A} να ισούται με τον μακροπρόθεσμο ρυθμό μεταβάσεων που εισέρχονται στο ίδιο σύνολο, δηλαδή

$$\sum_{j \in \mathcal{A}} \sum_{i \in \mathcal{A}^c} \pi_j p_{ji} = \sum_{i \in \mathcal{A}^c} \sum_{j \in \mathcal{A}} \pi_i p_{ij}, \quad \mathcal{A} \subseteq \mathcal{S}.$$

Η εξέλιξη ενός στοχαστικού συστήματος που περιγράφεται από μια Μαρκοβιανή αλυσίδα διακριτού χρόνου συνδέεται, συνήθως, με κάποια δομή κόστους/αμοιβής. Π.χ., η παραμονή σε κάθε κατάσταση ή η μετάβαση από κατάσταση σε κατάσταση μπορεί να επάγει κόστος. Πιο συγκεκριμένα, δίνουμε τον ακόλουθο ορισμό.

Ορισμός 12.7 (Δομή κόστους σε Μαρκοβιανή αλυσίδα διακριτού χρόνου) Έστω $\{X_n : n \geq 0\}$ Μαρκοβιανή αλυσίδα διακριτού χρόνου, με χώρο καταστάσεων \mathcal{S} και πίνακα πιθανοτήτων μετάβασης $\mathbf{P} = (p_{ij})$.

- (i) Μια δομή κόστους παραμονής είναι μια συνάρτηση $c : \mathcal{S} \rightarrow \mathbb{R}$, που δίνει για κάθε κατάσταση j του \mathcal{S} , το κόστος $c(j)$ μιας επίσκεψης (δηλαδή, χρονικής μονάδας παραμονής) της $\{X_n\}$ σε αυτήν.
- (ii) Μια δομή κόστους μετάβασης είναι μια συνάρτηση $d : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$, που δίνει για κάθε μετάβαση $i \rightarrow j$ του \mathcal{S} , το κόστος μετάβασης $d(i, j)$ που επάγει η $\{X_n\}$.

Το παρακάτω αποτέλεσμα παρέχει έναν εύκολο τρόπο υπολογισμού του μακροπρόθεσμου ρυθμού κόστους.

Θεώρημα 12.8 (Μακροπρόθεσμος ρυθμός κόστους σε Μαρκοβιανή αλυσίδα διακριτού χρόνου) Έστω $\{X_n : n \geq 0\}$ Μαρκοβιανή αλυσίδα διακριτού χρόνου, με χώρο καταστάσεων \mathcal{S} και πίνακα πιθανοτήτων μετάβασης $\mathbf{P} = (p_{ij})$. Έστω $c : \mathcal{S} \rightarrow \mathbb{R}$ μια δομή κόστους παραμονής και $d : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$ μια δομή κόστους μετάβασης. Ορίζουμε

$$C(n) = \sum_{k=0}^{n-1} (c(X_k) + d(X_k, X_{k+1})), \quad n \geq 0,$$

το συνολικό κόστος που συσσωρεύεται μέχρι την n -οστή μετάβαση της Μαρκοβιανής αλυσίδας. Υποθέτουμε ότι η $\{X_n\}$ είναι αδιαχώριστη και θετικά επαναληπτική με κατανομή ισορροπίας $\pi = (\pi_j : j \in \mathcal{S})$ και

$$\sum_{j \in \mathcal{S}} \pi_j \left(|c_j| + \sum_{k \in \mathcal{S}} p_{jk} |d(j, k)| \right) < \infty.$$

Τότε:

(i) Για τον μακροπρόθεσμο ρυθμό κόστους έχουμε

$$\lim_{n \rightarrow \infty} \frac{C(n)}{n} = \sum_{j \in \mathcal{S}} \pi_j c(j) + \sum_{j \in \mathcal{S}} \sum_{k \in \mathcal{S}} \pi_j p_{jk} d(j, k), \quad \text{με πιθανότητα 1.}$$

(ii) Για τον μακροπρόθεσμο μέσο ρυθμό κόστους έχουμε

$$\lim_{n \rightarrow \infty} \frac{E[C(n)]}{n} = \sum_{j \in \mathcal{S}} \pi_j c(j) + \sum_{j \in \mathcal{S}} \sum_{k \in \mathcal{S}} \pi_j p_{jk} d(j, k).$$

Το αποτέλεσμα αυτό είναι διαισθητικά εύλογο, αφού η π_j εκφράζει το μακροπρόθεσμο ποσοστό του χρόνου που η Μαρκοβιανή αλυσίδα βρίσκεται στην κατάσταση j , οπότε και το κόστος θα είναι $c(j)$ για μια χρονική μονάδα. Επομένως ο όρος $\sum_{j \in \mathcal{S}} \pi_j c(j)$ εκφράζει το μέσο κόστος παραμονής ανά χρονική μονάδα, δεσμεύοντας στην κατάσταση που μπορεί να βρίσκεται το σύστημα. Από την άλλη μεριά, η $\pi_j p_{jk}$ εκφράζει τον μακροπρόθεσμο ρυθμό μεταβάσεων από την j στην k . Κάθε τέτοια μετάβαση επάγει κόστος $d(j, k)$. Επομένως, ο όρος $\sum_{j \in \mathcal{S}} \sum_{k \in \mathcal{S}} \pi_j p_{jk} d(j, k)$ εκφράζει το μέσο κόστος μεταβάσεων ανά χρονική μονάδα, δεσμεύοντας στην κατάσταση j που μπορεί να βρίσκεται το σύστημα και στην κατάσταση k στην οποία μεταβαίνει.

12.2 Ανανεωτικές διαδικασίες

Οι ανανεωτικές διαδικασίες ορίστηκαν στο κεφάλαιο 1 (βλ. ορισμό 1.1), καθώς αποτελούν το πιο απλό μοντέλο για τη μελέτη γεγονότων που συμβαίνουν τυχαία στο χρόνο με κάποια μορφή πιθανοθεωρητικής περιοδικότητας. Για τον σκοπό αυτό χρησιμοποιούνται στα πλαίσια της Θεωρίας Ουρών για τη μοντελοποίηση και μελέτη των διαδικασιών αφίξεων πελατών, όταν οι χρόνοι μεταξύ διαδοχικών αφίξεων είναι ανεξάρτητοι και ισόνομοι, όπως είναι η τυπική υπόθεση στις περισσότερες εφαρμογές. Επίσης, χρησιμοποιούνται για τη μελέτη των διαδικασιών αναχωρήσεων στη διάρκεια μιας περιόδου συνεχούς λειτουργίας, όταν οι χρόνοι εξυπηρέτησης είναι ανεξάρτητοι και ισόνομοι. Όταν οι ενδιάμεσοι χρόνοι μιας ανανεωτικής διαδικασίας έχουν

την εκθετική κατανομή, τότε έχουμε τη διαδικασία Poisson, που είναι και Μαρκοβιανή αλυσίδα συνεχούς χρόνου, και η οποία έχει πλούσια δομή που αντιστοιχεί στα πολύ κομψά και χρήσιμα θεωρήματα που είδαμε στην παράγραφο 4.4 του κεφαλαίου 4. Η διαδικασία Poisson ως ειδική περίπτωση ανανεωτικής διαδικασίας αρκεί για την αποτίμηση απόδοσης Μαρκοβιανών μοντέλων ουρών αναμονής, όπως είδαμε στο μέρος II. Όμως, στο παρόν μέρος III, θα θεωρήσουμε μοντέλα με γενικούς ενδιάμεσους χρόνους αφίξεων ή γενικούς χρόνους εξυπηρέτησης, οπότε είναι απαραίτητο να δούμε και κάποια στοιχεία της γενικότερης θεωρίας των ανανεωτικών διαδικασιών.

Έστω, λοιπόν, μια ακολουθία μη-αρνητικών ανεξάρτητων και ισόνομων τυχαίων μεταβλητών X_1, X_2, \dots που αντιστοιχούν στους ενδιάμεσους χρόνους μεταξύ κάποιων γεγονότων (συμβάντων) και $S_0 = 0, S_n = \sum_{i=1}^n X_n, n \geq 1$, οι χρόνοι των συμβάντων για τους οποίους ισχύει $S_0 \leq S_1 \leq S_2 \leq \dots$. Η στοχαστική διαδικασία $\{S_n : n \geq 0\}$ αναφέρεται ως ανανεωτική ακολουθία με χρόνους γεγονότων $S_n, n \geq 0$, και ενδιάμεσους χρόνους γεγονότων $X_n, n \geq 1$. Η απαριθμήτρια διαδικασία $\{N(t) : t \geq 0\}$ της $\{S_n : n \geq 0\}$ που ορίζεται από τη σχέση

$$N(t) = \sup\{n \geq 0 : S_n \leq t\}, \quad t \geq 0,$$

(δηλαδή, η $N(t)$ μετράει το πλήθος των γεγονότων που έχουν συμβεί στο $(0, t]$) αναφέρεται ως (απαριθμήτρια) ανανεωτική διαδικασία.

Είναι φανερό ότι η πιθανοθεωρητική συμπεριφορά της ανανεωτικής διαδικασίας καθορίζεται πλήρως από την κοινή κατανομή των ενδιάμεσων χρόνων της, $F_X(t) = \Pr[X_i \leq t]$. Πράγματι, με βάση την κατανομή αυτή, μπορούμε εύκολα να υπολογίσουμε τις κατανομές του χρόνου του n -οστού γεγονότος, S_n , και του αριθμού των γεγονότων μέχρι τη χρονική στιγμή t , $N(t)$, καθώς και την ανανεωτική συνάρτηση $m(t) = E[N(t)]$ για σταθερά n και t . Έχουμε το ακόλουθο αποτέλεσμα.

Θεώρημα 12.9 (Βασικοί ανανεωτικοί υπολογισμοί σε πεπερασμένη χρονική στιγμή) Έστω ανανεωτική διαδικασία $\{N(t) : t \geq 0\}$ με χρόνους γεγονότων S_1, S_2, \dots και ενδιάμεσους χρόνους γεγονότων X_1, X_2, \dots με κατανομή $F_X(t) = \Pr[X_i \leq t]$. Τότε έχουμε:

(i) Η συνάρτηση κατανομής του χρόνου του k -οστού γεγονότος είναι

$$F_{S_k}(t) = \Pr[S_k \leq t] = F_X^{*k}(t), \quad k \geq 1, \quad t \geq 0,$$

όπου $F_X^{*k}(t) = (F_X * F_X * \dots * F_X)(t)$ είναι η k -οστή συνέλιξη της $F_X(t)$, με $(G * F)(t) = \int_{-\infty}^{\infty} G(t-x)dF(x)$.

(ii) Η συνάρτηση πιθανότητας του αριθμού των γεγονότων τη στιγμή t είναι

$$p_k(t) = \Pr[N(t) = k] = F_X^{*k}(t) - F_X^{*k+1}(t), \quad k \geq 1, \quad t \geq 0,$$

με τη σύμβαση $F_X^{*0}(t) = 1, t \geq 0$.

(iii) Η ανανεωτική συνάρτηση είναι

$$m(t) = \sum_{k=1}^{\infty} F_X^{*k}(t), \quad t \geq 0.$$

◊

Το (i) είναι προφανές αφού η τυχαία μεταβλητή S_k είναι το άθροισμα k ανεξάρτητων και ισόνομων τυχαίων μεταβλητών με κατανομή $F_X(t)$, και η συνέλιξη $(G * F)(t)$ δυο κατανομών $G(t)$ και $F(t)$ είναι η συνάρτηση κατανομής του αθροίσματος δυο ανεξάρτητων τυχαίων μεταβλητών με τις κατανομές αυτές.

Για το (ii), αρκεί να παρατηρήσουμε την ισότητα ενδεχομένων $\{N(t) = k\} = \{S_k \leq t < S_{k+1}\} = \{S_k \leq t\} \setminus \{S_{k+1} \leq t\}$, να πάρουμε τις αντίστοιχες πιθανότητες και να χρησιμοποιήσουμε το (i). Πράγματι, το ενδεχόμενο

να έχουν συμβεί k γεγονότα ως τη στιγμή t ισοδυναμεί με το ότι το k -οστό γεγονός συνέβει το πολύ ως και τη στιγμή t , ενώ το $(k + 1)$ -οστό γεγονός θα συμβεί μετά από αυτή. Οπότε,

$$\begin{aligned} p_k(t) &= \Pr[N(t) = k] = \Pr[S_k \leq t < S_{k+1}] \\ &= \Pr[S_k \leq t] - \Pr[S_{k+1} \leq t] = F_X^{*k}(t) - F_X^{*(k+1)}(t). \end{aligned}$$

Τέλος, το (iii) προκύπτει παρατηρώντας ότι $N(t) = \sum_{k=1}^{\infty} 1_{\{S_k \leq t\}}$ και χρησιμοποιώντας το (i):

$$m(t) = E[N(t)] = E \left[\sum_{k=1}^{\infty} 1_{\{S_k \leq t\}} \right] = \sum_{k=1}^{\infty} \Pr[S_k \leq t] = \sum_{k=1}^{\infty} F_X^{*k}(t).$$

Επειδή η συνέλιξη κατανομών είναι άβολη από υπολογιστική σκοπιά, συχνά στις εφαρμογές δουλεύουμε με τους μετασχηματισμούς Laplace-Stieltjes των αντίστοιχων ποσοτήτων. Μετασχηματίζοντας τους τύπους του Θεωρήματος 12.9, παίρνουμε το ακόλουθο αποτέλεσμα.

Θεώρημα 12.10 (Μετασχηματισμοί Laplace-Stieltjes βασικών ανανεωτικών ποσοτήτων) *Οι μετασχηματισμοί Laplace-Stieltjes των $F_{S_k}(t)$, $p_k(t)$ και $m(t)$ δίνονται αντίστοιχα από τους τύπους*

$$\begin{aligned} \tilde{F}_{S_k}(s) &= \int_0^{\infty} e^{-st} dF_{S_k}(t) = (\tilde{F}_X(s))^k, \quad k \geq 1, \\ \tilde{p}_k(s) &= \int_0^{\infty} e^{-st} dp_k(t) = (1 - \tilde{F}_X(s))(\tilde{F}_X(s))^k, \quad k \geq 0, \\ \tilde{m}(s) &= \int_0^{\infty} e^{-st} dm(t) = \frac{\tilde{F}_X(s)}{1 - \tilde{F}_X(s)}. \end{aligned}$$

◊

Πράγματι, έχουμε

$$\begin{aligned} \tilde{F}_{S_k}(s) &= \widetilde{F_X^{*k}}(s) = (\tilde{F}_X(s))^k, \\ \tilde{p}_k(s) &= \widetilde{F_X^{*k}}(s) - \widetilde{F_X^{*(k+1)}}(s) = (\tilde{F}_X(s))^k - (\tilde{F}_X(s))^{k+1} = (1 - \tilde{F}_X(s))(\tilde{F}_X(s))^k, \\ \tilde{m}(s) &= \sum_{k=1}^{\infty} \widetilde{F_X^{*k}}(s) = \sum_{k=1}^{\infty} (\tilde{F}_X(s))^k = \frac{\tilde{F}_X(s)}{1 - \tilde{F}_X(s)}. \end{aligned}$$

Όσον αφορά την οριακή συμπεριφορά της $\{N(t)\}$, έχουμε τα αποτελέσματα που περιγράφονται παρακάτω, στα θεωρήματα 12.11 και 12.12. Το θεώρημα 12.11 αναφέρεται στην κατανομή του συνολικού πλήθους των ανανεώσεων μιας ανανεωτικής διαδικασίας στο διάστημα $(0, \infty)$. Το θεώρημα 12.12 δίνει κάποια αντίστοιχα των κλασικών αποτελεσμάτων της Θεωρίας Πιθανοτήτων, του Νόμου των Μεγάλων Αριθμών και του Κεντρικού Οριακού Θεωρήματος, στα πλαίσια της θεωρίας των ανανεωτικών διαδικασιών.

Θεώρημα 12.11 (Κατανομή συνολικού πλήθους ανανεώσεων) *Έστω $N(\infty)$ το πλήθος των ανανεώσεων της $\{N(t)\}$ στο $(0, \infty)$ και $F_X(\infty) = \lim_{t \rightarrow \infty} F_X(t) = \Pr[X_n < \infty]$.*

(i) *Αν $F_X(\infty) = 1$, τότε $\Pr[N(\infty) = \infty] = 1$.*

(ii) *Αν $F_X(\infty) < 1$, τότε $\Pr[N(\infty) < \infty] = 1$ και $\Pr[N(\infty) = k] = (1 - F_X(\infty))F_X(\infty)^k$, $k \geq 0$.*

◊

Πράγματι, αν οι ενδιάμεσοι χρόνοι είναι πεπερασμένοι με πιθανότητα 1, τότε είναι βέβαιο ότι μετά από κάθε ανανέωση ακολουθεί και άλλη σε πεπερασμένο χρόνο. Επομένως, το συνολικό πλήθος των ανανεώσεων σε άπειρο χρονικό ορίζοντα είναι άπειρο με πιθανότητα 1. Αν, όμως, υπάρχει θετική πιθανότητα $1 - F_X(\infty)$ ένας ενδιάμεσος χρόνος να είναι άπειρος, τότε μετά από κάθε ανανέωση ακολουθεί και άλλη μόνο με πιθανότητα $F_X(\infty)$. Στην περίπτωση αυτή, το συνολικό πλήθος των ανανεώσεων είναι γεωμετρικά κατανομημένο με τη συνάρτηση πιθανότητας που δίνεται στο (ii).

Θεώρημα 12.12 (Οριακά θεωρήματα στην ανανεωτική θεωρία) Έστω ανανεωτική διαδικασία $\{N(t) : t \geq 0\}$, με ενδιάμεσους χρόνους $X_k, k \geq 1$, με $E[X_k] = \mu, \text{Var}[X_k] = \sigma^2, k \geq 1$, και ανανεωτική συνάρτηση $m(t)$. Τότε:

(i) **Νόμος μεγάλων αριθμών:** Για τον μακροπρόθεσμο ρυθμό (μακροπρόθεσμη συχνότητα) ανανεώσεων έχουμε

$$\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \frac{1}{\mu}, \text{ με πιθανότητα } 1.$$

(ii) **Στοιχειώδες ανανεωτικό θεώρημα:** Για τον μακροπρόθεσμο μέσο ρυθμό (μακροπρόθεσμη μέση συχνότητα) ανανεώσεων έχουμε

$$\lim_{t \rightarrow \infty} \frac{m(t)}{t} = \frac{1}{\mu}.$$

(iii) **Κεντρικό οριακό θεώρημα:** Αν $\mu < \infty$ και $\sigma^2 \in (0, \infty)$, έχουμε

$$\lim_{t \rightarrow \infty} \Pr \left(\frac{N(t) - \frac{t}{\mu}}{\sqrt{\frac{\sigma^2 t}{\mu^3}}} \leq x \right) = \Phi(x), \quad x \in \mathbb{R},$$

όπου $\Phi(x)$ η συνάρτηση κατανομής της τυποποιημένης κανονικής.

◊

Μια θεμελιώδης τεχνική για τη μελέτη των ανανεωτικών διαδικασιών και των εφαρμογών τους είναι ο ανανεωτικός συλλογισμός. Συγκεκριμένα, έστω ότι έχουμε μια συνάρτηση $h(t)$ που αναφέρεται στην εξέλιξη της ανανεωτικής διαδικασίας στο $(0, t]$ (π.χ., η $h(t)$ μπορεί να είναι κάποια πιθανότητα ή μέση τιμή που εξαρτάται από το t , όπως η ανανεωτική συνάρτηση $m(t)$ ή ο μέσος χρόνος $E[S_{N(t)+1} - t]$ που απαιτείται για το επόμενο ανανεωτικό γεγονός τη στιγμή t). Ο ανανεωτικός συλλογισμός συνίσταται στη δέσμευση στον χρόνο $S_1 = u$ του πρώτου ανανεωτικού γεγονότος για τον υπολογισμό της ποσότητας $h(t)$. Αυτή η τεχνική οδηγεί σε ολοκληρωτικές εξισώσεις συγκεκριμένου τύπου για την $h(t)$ που αναφέρονται ως ανανεωτικές εξισώσεις. Για να γίνει κατανοητή η τεχνική δίνουμε δυο παραδείγματα.

Παράδειγμα 12.13 (Η ανανεωτική εξίσωση για την ανανεωτική συνάρτηση) Έστω $h(t) = m(t)$ η ανανεωτική συνάρτηση που αντιστοιχεί σε μια ανανεωτική διαδικασία $\{N(t)\}$. Έστω S_1 ο χρόνος του πρώτου γεγονότος της $\{N(t)\}$ και $F_X(t)$ η κατανομή των ενδιάμεσων χρόνων της $\{N(t)\}$. Δεσμεύοντας στον S_1 , έχουμε

$$h(t) = m(t) = \int_0^\infty E[N(t)|S_1 = u] dF_X(u). \quad (12.7)$$

Όμως, όταν $u \leq t$, τότε η δεσμευμένη κατανομή της $N(t)$, δεδομένου του ενδεχομένου $\{S_1 = u\}$ είναι η ίδια με την κατανομή της $1 + N(t - u)$. Συμβολικά γράφουμε $(N(t)|S_1 = u) \stackrel{d}{=} 1 + N(t - u)$, για $u \leq t$, ώστε να δηλώσουμε την ισότητα κατά κατανομή (το d πάνω από την ισότητα παραπέμπει στη λέξη *distribution*). Επομένως, έχουμε

$$E[N(t)|S_1 = u] = \begin{cases} 0, & \text{αν } t < u, \\ 1 + E[N(t - u)], & \text{αν } t \geq u. \end{cases}$$

Αντικαθιστώντας στη (12.7), παίρνουμε

$$\begin{aligned} h(t) &= \int_0^t dF_X(u) + \int_0^t h(t-u)dF_X(u) \\ &= F_X(t) + \int_0^t h(t-u)dF_X(u). \end{aligned}$$

Η εξίσωση αυτή έχει τη μορφή

$$h(t) = d(t) + \int_0^t h(t-u)dF_X(u) = d(t) + (h * F_X)(t),$$

όπου η $h(t)$ είναι η υπό μελέτη (άγνωστη) ποσότητα, ενώ η $d(t)$ είναι μια γνωστή συνάρτηση και η $F_X(t)$ είναι η συνάρτηση κατανομής των ενδιάμεσων χρόνων της υποκείμενης ανανεωτικής διαδικασίας. Τέτοιες (ολοκληρωτικές) εξισώσεις αναφέρονται ως ανανεωτικές εξισώσεις.

Στην προκειμένη περίπτωση, έχουμε $d(t) = F_X(t)$, αλλά στη γενική περίπτωση μιας ανανεωτικής εξίσωσης η $d(t)$ δεν είναι κατ' ανάγκη κάποια συνάρτηση κατανομής, αλλά μια γνωστή συνάρτηση που έχει υπολογιστεί. ◊

Παράδειγμα 12.14 (Η ανανεωτική εξίσωση για τον μέσο υπολειπόμενο χρόνο ανανέωσης) Έστω $\{N(t)\}$ ανανεωτική διαδικασία και $h(t) = E[S_{N(t)+1} - t]$ ο μέσος υπολειπόμενος χρόνος ανανέωσης τη στιγμή t , δηλαδή ο μέσος χρόνος που απαιτείται για το επόμενο ανανεωτικό γεγονός από τη στιγμή t . Έστω S_1 ο χρόνος του πρώτου γεγονότος της $\{N(t)\}$ και $F_X(t)$ η κατανομή των ενδιάμεσων χρόνων της $\{N(t)\}$. Δεσμεύοντας στον S_1 , έχουμε

$$h(t) = E[S_{N(t)+1} - t] = \int_0^\infty E[S_{N(t)+1} - t | S_1 = u] dF_X(u). \quad (12.8)$$

Όμως, όταν $u \leq t$, η δεσμευμένη κατανομή του $S_{N(t)+1} - t$ δεδομένου του ενδεχομένου $\{S_1 = u\}$ είναι η ίδια με την κατανομή του $S_{N(t-u)+1} - (t-u)$. Δηλαδή, έχουμε $(S_{N(t)+1} - t | S_1 = u) \stackrel{d}{=} S_{N(t-u)+1} - (t-u)$, για $u \leq t$. Αυτό συμβαίνει, διότι κάθε φορά που συμβαίνει ένα ανανεωτικό γεγονός, ο υπολειπόμενος χρόνος ανανέωσης επανεκκινεί. Με άλλα λόγια, μετά από ένα ανανεωτικό γεγονός, ο υπολειπόμενος χρόνος ανανέωσης έχει την ίδια συμπεριφορά, σαν όλα να ξεκινούν από την αρχή. Επομένως, έχουμε

$$E[S_{N(t)+1} - t | S_1 = u] = \begin{cases} u - t, & \text{αν } t < u, \\ E[S_{N(t-u)+1} - (t-u)], & \text{αν } t \geq u. \end{cases}$$

Αντικαθιστώντας στην (12.8), παίρνουμε

$$h(t) = \int_t^\infty (u-t)dF_X(u) + \int_0^t h(t-u)dF_X(u).$$

Η εξίσωση αυτή έχει τη μορφή

$$h(t) = d(t) + \int_0^t h(t-u)dF_X(u) = d(t) + (h * F_X)(t),$$

είναι, δηλαδή, μια ανανεωτική εξίσωση (με την έννοια που είπαμε στο προηγούμενο παράδειγμα). Στην προκειμένη περίπτωση έχουμε

$$\begin{aligned} d(t) &= \int_t^\infty (u-t)dF_X(u) = \int_t^\infty \int_t^u dy dF_X(u) \\ &= \int_t^\infty \int_y^\infty dF_X(u) dy = \int_t^\infty (1 - F_X(y)) dy. \end{aligned}$$

Τα δυο αυτά παραδείγματα δείχνουν πώς η τεχνική του ανανεωτικού συλλογισμού οδηγεί σε ανανεωτικές εξισώσεις. Υπάρχουν βέβαια ανανεωτικές εξισώσεις για πολλές συναρτήσεις $h(t)$ που συνδέονται με την εξέλιξη μιας ανανεωτικής διαδικασίας. Έχοντας τώρα μια εξίσωση για μια τέτοια συνάρτηση, μας ενδιαφέρει η λύση της, δηλαδή η εύρεση κάποιου τύπου για την $h(t)$, καθώς και η οριακή συμπεριφορά της λύσης για $t \rightarrow \infty$.

Όσον αφορά τη λύση μιας ανανεωτικής εξίσωσης έχουμε το ακόλουθο αποτέλεσμα.

Θεώρημα 12.15 (Λύση ανανεωτικής εξίσωσης) Έστω η ανανεωτική εξίσωση

$$h(t) = d(t) + \int_0^t h(t-u)dF_X(u) = d(t) + (h * F_X)(t), \quad t \geq 0.$$

Η ανανεωτική εξίσωση έχει μοναδική λύση, που δίνεται από τον τύπο

$$h(t) = d(t) + \int_0^t d(t-u)dm_X(u) = d(t) + (d * m_X)(t),$$

όπου $m_X(t)$ είναι η ανανεωτική συνάρτηση που αντιστοιχεί σε ανανεωτική διαδικασία με κατανομή ενδιάμεσων χρόνων γεγονότων $F_X(t)$.

Η απόδειξη αυτού του θεωρήματος στη γενική περίπτωση έχει κάποια τεχνικά σημεία, αλλά σε τυπικό επίπεδο μπορεί να γίνει πολύ εύκολα μετασχηματίζοντας την ανανεωτική εξίσωση, χρησιμοποιώντας τον μετασχηματισμό Laplace-Stieltjes. Πράγματι, αν συμβολίσουμε με $\tilde{h}(s)$, $\tilde{d}(s)$ και $\tilde{F}_X(s)$ τους μετασχηματισμούς Laplace-Stieltjes των $h(t)$, $d(t)$ και $F_X(t)$, αντίστοιχα, έχουμε

$$\begin{aligned} h(t) &= d(t) + (h * F_X)(t) \\ \Rightarrow \tilde{h}(s) &= \tilde{d}(s) + \tilde{h}(s)\tilde{F}_X(s) \\ \Rightarrow \tilde{h}(s) &= \frac{\tilde{d}(s)}{1 - \tilde{F}_X(s)} \\ \Rightarrow \tilde{h}(s) &= \tilde{d}(s) + \tilde{d}(s) \frac{\tilde{F}_X(s)}{1 - \tilde{F}_X(s)}. \end{aligned}$$

Όμως, από το θεώρημα 12.10, ο μετασχηματισμός Laplace-Stieltjes της ανανεωτικής συνάρτησης είναι $\tilde{m}_X(s) = \frac{\tilde{F}_X(s)}{1 - \tilde{F}_X(s)}$, οπότε παίρνουμε

$$\begin{aligned} \tilde{h}(s) &= \tilde{d}(s) + \tilde{d}(s)\tilde{m}_X(s) \\ \Rightarrow h(t) &= d(t) + (d * m_X)(t). \end{aligned}$$

Η ανανεωτική εξίσωση και η λύση της έχουν επίσης την εξής «φυσική» ερμηνεία: Έστω ότι κάθε γεγονός επάγει μια επίδραση, της οποίας η ένταση είναι $d(t)$, t χρονικές μονάδες μετά την εκδήλωσή του και έστω $h(t)$ η ένταση της συνολικής επίδρασης από όλα τα γεγονότα μέχρι τη χρονική στιγμή t . Θεωρούμε, επίσης, ότι τη χρονική στιγμή 0 έχει συμβεί γεγονός (το γεγονός-0), του οποίου η επίδραση λαμβάνεται υπόψη στην $h(t)$. Τότε, ο ανανεωτικός συλλογισμός δίνει για την $h(t)$ την ανανεωτική εξίσωση

$$h(t) = d(t) + \int_0^t h(t-u)dF_X(u),$$

η οποία εκφράζει ότι η ένταση της συνολικής επίδρασης τη στιγμή t ισούται με την ένταση της επίδρασης του αρχικού γεγονότος της στιγμής 0 συν την ένταση της συνολικής επίδρασης των υπόλοιπων γεγονότων (από το γεγονός-1 και μετά που συνέβη κάποια στιγμή u στο $(0, t]$). Όμως, η $h(t)$ μπορεί να μετρηθεί εναλλακτικά, προσθέτοντας τις εντάσεις των γεγονότων, δεσμεύοντας στις στιγμές που συνέβησαν. Έτσι, έχουμε

$$\begin{aligned} h(t) &= d(t) + \sum_{k=1}^{\infty} \int_0^t d(t-u)dF_{S_k}(u) = d(t) + \sum_{k=1}^{\infty} \int_0^t d(t-u)dF_X^{*k}(u) \\ &= d(t) + \int_0^t d(t-u)dm_X(u). \end{aligned}$$

Όπως είπαμε, συχνά μας ενδιαφέρει το $\lim_{t \rightarrow \infty} h(t)$, όπου $h(t)$ είναι η λύση μιας ανανεωτικής εξίσωσης. Το όριο αυτό υπάρχει κάτω από κάποιες συνθήκες κανονικότητας και υπολογίζεται με έναν σχετικά εύκολο τρόπο. Το σχετικό θεώρημα αναφέρεται στη βιβλιογραφία ως βασικό ανανεωτικό θεώρημα (key renewal theorem). Η διατύπωσή του εξαρτάται από μια ιδιότητα της κατανομής των ενδιάμεσων χρόνων (περιοδικότητα - απεριοδικότητα). Συγκεκριμένα, δίνουμε τον ακόλουθο ορισμό αυτής της ιδιότητας.

Ορισμός 12.16 (Περιοδική μη-αρνητική τυχαία μεταβλητή) Μια μη-αρνητική, γνήσια τυχαία μεταβλητή X (δηλαδή $\Pr[0 \leq X < \infty] = 1$) λέγεται περιοδική, αν υπάρχει $p > 0$ ώστε $\sum_{k=0}^{\infty} \Pr[X = kp] = 1$. Ο μεγαλύτερος αριθμός p με αυτή την ιδιότητα αναφέρεται ως περίοδος της X . Αν δεν υπάρχει τέτοιος αριθμός p , η X λέγεται απεριοδική. Για την αντίστοιχη κατανομή της τυχαίας μεταβλητής χρησιμοποιούμε τους ίδιους όρους (περιοδική - απεριοδική). \diamond

Οι συνεχείς και οι μικτές κατανομές (αυτές, δηλαδή, που έχουν κάποιο συνεχές και κάποιο διακριτό μέρος) είναι απεριοδικές. Από τις διακριτές κατανομές, κάποιες είναι περιοδικές και κάποιες όχι. Π.χ., οι κλασικές διακριτές κατανομές, όπως η διωνυμική, η γεωμετρική, η Poisson κλπ. είναι περιοδικές και μάλιστα με περίοδο 1. Από την άλλη μεριά, μια διακριτή που παίρνει με θετική πιθανότητα μόνο τις τιμές 1 και $\sqrt{2}$ είναι απεριοδική. Μια διακριτή που παίρνει με θετική πιθανότητα όλες τις τιμές $\frac{1}{n}$, $n \geq 1$, είναι επίσης απεριοδική.

Είμαστε τώρα σε θέση να διατυπώσουμε το βασικό ανανεωτικό θεώρημα.

Θεώρημα 12.17 (Βασικό ανανεωτικό θεώρημα) Έστω $h(t)$ η μοναδική λύση της ανανεωτικής εξίσωσης

$$h(t) = d(t) + \int_0^t h(t-u) dF_X(u) = d(t) + (h * F_X)(t), \quad t \geq 0.$$

Έστω, επίσης, ότι η $d(t)$ γράφεται ως διαφορά δυο μη-αρνητικών, φραγμένων, φθινουσών συναρτήσεων και

$$\int_0^{\infty} |d(u)| du < \infty.$$

Τότε

(i) Αν η $F_X(t)$ είναι απεριοδική με μέση τιμή $\mu > 0$,

$$\lim_{t \rightarrow \infty} h(t) = \frac{\int_0^{\infty} d(u) du}{\mu}.$$

(ii) Αν η $F_X(t)$ είναι περιοδική με περίοδο p και μέση τιμή $\mu > 0$,

$$\lim_{t \rightarrow \infty} h(tp + x) = \frac{p \sum_{t=0}^{\infty} d(tp + x)}{\mu}, \quad x \in \mathbb{R}.$$

Σε πολλές εφαρμογές του θεωρήματος ελέγχεται ότι η $d(t)$ είναι μη-αρνητική, φραγμένη και φθίνουσα με την απόλυτη τιμή της να έχει πεπερασμένο ολόκληρωμα στο $(0, \infty)$. Τότε γράφεται τετριμμένα ως διαφορά της ίδιας και της μηδενικής συνάρτησης (που πληροί επίσης τις ίδιες συνθήκες) και το βασικό ανανεωτικό θεώρημα είναι άμεσα εφαρμόσιμο.

Η απόδειξη του βασικού ανανεωτικού θεωρήματος είναι ιδιαίτερα τεχνική. Όμως, είναι σχετικά εύκολο να δούμε γιατί ένα τέτοιο αποτέλεσμα είναι εύλογο να ισχύει. Ας υποθέσουμε ότι η $F_X(t)$ είναι απεριοδική. Η $h(t)$, ως λύση της ανανεωτικής εξίσωσης, δίνεται από τη σχέση

$$h(t) = d(t) + \int_0^t d(t-u) dm_X(u).$$

Έχουμε άμεσα ότι $\lim_{t \rightarrow \infty} d(t) = 0$, αφού $\int_0^{\infty} d(u)du < \infty$, και επομένως

$$\lim_{t \rightarrow \infty} h(t) = \lim_{t \rightarrow \infty} \int_0^t d(t-u)dm_X(u).$$

Επιλέγοντας ένα $\epsilon \in (0, t)$, το παραπάνω ολοκλήρωμα σπάει σε δυο κομμάτια $\int_0^{\epsilon} d(t-u)dm_X(u)$ και $\int_{\epsilon}^t d(t-u)dm_X(u)$. Η τεχνική απόδειξη επιλέγει το ϵ κατάλληλα ώστε το $\int_0^{\epsilon} d(t-u)dm_X(u)$ να τείνει στο 0, καθώς $t \rightarrow \infty$. Πράγματι, αυτό είναι εύλογο, αφού για μικρά u (στο $[0, \epsilon]$) το $d(t-u)$ θα είναι κοντά στο 0, αφού $\lim_{t \rightarrow \infty} d(t) = 0$. Επίσης για μεγάλα u (στο (ϵ, t)) η $m_X(u) \simeq \frac{u}{\mu}$, από το στοιχειώδες ανανεωτικό θεώρημα (θεώρημα 12.12(ii)), οπότε

$$\int_{\epsilon}^t d(t-u)dm_X(u) \simeq \frac{1}{\mu} \int_{\epsilon}^t d(t-u)du = \frac{1}{\mu} \int_0^{t-\epsilon} d(y)dy \rightarrow \frac{\int_0^{\infty} d(y)dy}{\mu},$$

καθώς $t \rightarrow \infty$.

Ως παράδειγμα εφαρμογής του βασικού ανανεωτικού θεωρήματος, συνεχίζουμε το παράδειγμα 12.14.

Παράδειγμα 12.18 (Οριακός μέσος υπολειπόμενος χρόνος ανανέωσης) Θεωρούμε μια ανανεωτική διαδικασία $\{N(t)\}$, με απεριοδικούς ενδιάμεσους χρόνους γεγονότων. Ένας ενδιάμεσος χρόνος γεγονότων X έχει κατανομή $F_X(t)$, με μέση τιμή $\mu \in (0, \infty)$ και διασπορά $\sigma^2 \in (0, \infty)$. Όπως είδαμε στο παράδειγμα 12.14, ο μέσος υπολειπόμενος χρόνος ανανέωσης τη στιγμή t , $h(t)$, ικανοποιεί την ανανεωτική εξίσωση

$$h(t) = d(t) + (h * F_X)(t), \quad t \geq 0,$$

με

$$d(t) = \int_t^{\infty} (1 - F_X(y))dy, \quad t \geq 0.$$

Η $d(t)$ είναι μη-αρνητική και φθίνουσα αφού $1 - F_X(y) \geq 0$, $y \geq 0$. Επίσης, είναι φραγμένη αφού

$$d(t) \leq \int_0^{\infty} (1 - F_X(y))dy = \mu < \infty, \quad t \geq 0.$$

Τέλος, έχουμε

$$\begin{aligned} \int_0^{\infty} |d(u)|du &= \int_0^{\infty} d(u)du \\ &= \int_0^{\infty} \int_u^{\infty} (1 - F_X(y))dydu \\ &= \int_0^{\infty} \int_u^{\infty} \int_y^{\infty} dF_X(x)dydu \\ &= \int_0^{\infty} \int_0^x \int_0^y dudydF_X(x) \\ &= \int_0^{\infty} \int_0^x ydydF_X(x) \\ &= \int_0^{\infty} \frac{x^2}{2}dF_X(x) \\ &= \frac{E[X^2]}{2} = \frac{\mu^2 + \sigma^2}{2} < \infty. \end{aligned}$$

Επομένως, το βασικό ανανεωτικό θεώρημα είναι εφαρμόσιμο και έχουμε

$$\begin{aligned}\lim_{t \rightarrow \infty} h(t) &= \frac{\int_0^{\infty} d(u)du}{\mu} \\ &= \frac{\mu^2 + \sigma^2}{2\mu}.\end{aligned}\quad (12.9)$$

Το αποτέλεσμα αυτό λέει ότι αν αφήσουμε να περάσει αρκετός χρόνος ώστε το σύστημα να βρεθεί σε κάποια κατάσταση «ισορροπίας», τότε έχει χαθεί η αρχική επίδραση της ύπαρξης γεγονότος τη στιγμή 0, ο αναμενόμενος υπολειπόμενος χρόνος για το επόμενο γεγονός είναι $\frac{\mu}{2} + \frac{\sigma^2}{2\mu}$, δηλαδή μεγαλύτερος από μισό αναμενόμενο ενδιάμεσο χρόνο. Αυτό μοιάζει καταρχήν παράδοξο και μάλιστα αναφέρεται στη βιβλιογραφία ως ανανεωτικό παράδοξο. Την κατάσταση αυτή θα τη συζητήσουμε εκτενώς παρακάτω στο παράδειγμα 12.25. \diamond

Όπως και για τις Μαρκοβιανές διαδικασίες, μπορούμε να εφοδιάσουμε μια ανανεωτική διαδικασία με μια δομή κόστους. Πιο συγκεκριμένα, έστω $\{N(t)\}$ μια ανανεωτική διαδικασία με χρόνους γεγονότων S_1, S_2, \dots , ενδιάμεσους χρόνους γεγονότων X_1, X_2, \dots και κατανομή ενδιάμεσων χρόνων $F_X(t)$. Θεωρούμε επίσης διαδικασία $\{C(t)\}$, την οποία ονομάζουμε διαδικασία κόστους και η οποία μπορεί να είναι εξαιρετικά γενική. Για κάθε συγκεκριμένη σταθερή στιγμή t , σκεφτόμαστε την τυχαία μεταβλητή $C(t)$ ως το κόστος που έχει συσσωρευτεί στο διάστημα $(0, t]$ στο υπό εξέταση σύστημα. Το κόστος αυτό μπορεί να μην είναι μονότονη συνάρτηση του t για μια πραγματοποίηση της διαδικασίας. Το κόστος μπορεί να σχετίζεται με κάποιο τρόπο με τα γεγονότα της $\{N(t)\}$ και να συσσωρεύεται με συνεχή τρόπο ή με άλματα στις στιγμές των γεγονότων. Το κόστος μπορεί να παίρνει και αρνητικές τιμές, οπότε τότε μπορεί να ερμηνεύεται ως αμοιβή.

Η διαδικασία κόστους $\{C(t)\}$ λέγεται συμβατή με την ανανεωτική διαδικασία $\{N(t)\}$, αν οι διδιάστατες τυχαίες μεταβλητές (X_n, C_n) με $C_n = C(S_n) - C(S_{n-1})$ είναι ανεξάρτητες και ισόνομες για $n \geq 1$ με κοινή συνάρτηση κατανομής $F_{X,C}(x, y)$. Αυτή η συνθήκη σημαίνει ότι το κόστος C_n που συσσωρεύεται σε έναν ανανεωτικό κύκλο δεν εξαρτάται από το τι συμβαίνει στους άλλους ανανεωτικούς κύκλους, αλλά μπορεί βέβαια να εξαρτάται από το τι συμβαίνει στο τρέχον ανανεωτικό κύκλο. Είναι σημαντικό για τις εφαρμογές οι C_n και X_n να μπορεί να είναι εξαρτημένες. Δίνουμε έτσι τον ακόλουθο ορισμό:

Ορισμός 12.19 (Ανανεωτική διαδικασία κόστους) Μια στοχαστική διαδικασία $\{C(t)\}$ που είναι συμβατή με μια ανανεωτική διαδικασία $\{N(t)\}$ λέγεται ανανεωτική διαδικασία κόστους. Η συνάρτηση κατανομής $F_{X,C}(x, y)$ των τυχαίων μεταβλητών (X_n, C_n) που δίνουν τις διάρκειες των ανανεωτικών κύκλων και τα αντίστοιχα κόστη αναφέρεται ως γεννώσα συνάρτηση κατανομής της $\{C(t)\}$. \diamond

Παράδειγμα 12.20 (Αλματική ανανεωτική διαδικασία κόστους) Αν έχουμε μια ανανεωτική διαδικασία $\{N(t)\}$ με ενδιάμεσους χρόνους X_1, X_2, \dots , μια ακολουθία ανεξάρτητων, ισόνομων τυχαίων μεταβλητών Y_1, Y_2, \dots , ανεξάρτητων της $\{N(t)\}$, και μια συνάρτηση $g(x, y)$ τότε η διαδικασία $\{C(t)\}$ με

$$C(t) = \sum_{i=1}^{N(t)} g(X_i, Y_i), \quad t \geq 0,$$

αναφέρεται ως αλματική ανανεωτική διαδικασία κόστους. Στην περίπτωση αυτή

$$C_n = C(S_n) - C(S_{n-1}) = g(X_n, Y_n),$$

οπότε οι τυχαίες μεταβλητές $(X_n, C_n) = (X_n, g(X_n, Y_n))$, $n \geq 1$ είναι ανεξάρτητες και επομένως η $C(t)$ είναι πράγματι ανανεωτική διαδικασία κόστους.

Η φυσική ερμηνεία της αλματικής ανανεωτικής διαδικασίας κόστους είναι ότι το κόστος συσσωρεύεται με άλματα που συμβαίνουν τις στιγμές των γεγονότων της υποκείμενης ανανεωτικής διαδικασίας $\{N(t)\}$. Το κόστος που

επάγεται τη στιγμή S_n που τελειώνει ο n -οστός ανανεωτικός κύκλος εξαρτάται από τη διάρκεια X_n του ανανεωτικού κύκλου και τυχαίους παράγοντες που εκφράζονται μέσω της τυχαίας μεταβλητής Y_n . Η συνάρτηση $g(x, y)$ εκφράζει τη σύνδεση του κόστους που επάγεται στο τέλος του n -οστού ανανεωτικού κύκλου με τη διάρκειά του και τους τυχαίους παράγοντες που υπεισήλθαν σε αυτόν. ◊

Παράδειγμα 12.21 (Ανανεωτική διαδικασία) Μια ανανεωτική διαδικασία μπορεί να θεωρηθεί αλματική ανανεωτική διαδικασία κόστους ορίζοντας $Y_i = 0$ (ή γενικά κάποια αυθαίρετη τιμή) και $g(x, y) = 1$. Στην περίπτωση αυτή $C_n = 1$, $n \geq 1$ και $C(t) = N(t)$. ◊

Παράδειγμα 12.22 (Σύνθετη ανανεωτική διαδικασία) Η περίπτωση της αλματικής ανανεωτικής διαδικασίας κόστους με $f(x, y) = y$ αναφέρεται ως σύνθετη ανανεωτική διαδικασία. Στην περίπτωση αυτή $C_n = Y_n$, $n \geq 1$ και $C(t) = \sum_{i=1}^{N(t)} Y_i$. Η διαδικασία αυτή είναι χρήσιμη για να περιγράψουμε τη συσσώρευση κόστους λόγω εκδήλωσης γεγονότων καθένα από τα οποία επάγει κόστος ανεξάρτητα από τα υπόλοιπα, όταν τα κόστη των γεγονότων είναι ισόνομα.

Μια ειδική περίπτωση προκύπτει όταν η $\{N(t)\}$ είναι διαδικασία Poisson που μοντελοποιεί τις απαιτήσεις αποζημίωσης που φθάνουν σε μια ασφαλιστική εταιρεία και Y_1, Y_2, \dots είναι τα χρηματικά ποσά των απαιτήσεων. Ομοίως, η $\{N(t)\}$ μπορεί να είναι μια διαδικασία που μοντελοποιεί τις αφίξεις παραγγελιών σε μια αποθήκη και Y_1, Y_2, \dots είναι τα μεγέθη των παραγγελιών.

Μια άλλη περίπτωση προκύπτει όταν οι Y_1, Y_2, \dots είναι ακέραιες, μη-αρνητικές. Στην περίπτωση αυτή η $\{N(t)\}$ μπορεί να μοντελοποιεί τη διαδικασία αφίξεων ομάδων πελατών σε ένα σύστημα εξυπηρέτησης και Y_1, Y_2, \dots είναι τα μεγέθη των ομάδων. ◊

Οι ανανεωτικές διαδικασίες κόστους είναι το κατάλληλο μοντέλο για να περιγραφούν στοχαστικά περιοδικά φαινόμενα κατά τα οποία συσσωρεύεται κόστος. Κατόπιν το ενδιαφέρον εστιάζεται στον υπολογισμό του μακροπρόθεσμου μέσου ρυθμού κόστους (δηλαδή του μέσου κόστους ανά χρονική μονάδα). Το βασικό αποτέλεσμα για τον υπολογισμό αυτό είναι το παρακάτω στοιχειώδες ανανεωτικό θεώρημα με κόστη που ουσιαστικά λέει ότι ο υπολογισμός αυτός μπορεί να γίνει επικεντρώνοντας την προσοχή μας σε έναν τυπικό ανανεωτικό κύκλο. Πιο συγκεκριμένα έχουμε το εξής:

Θεώρημα 12.23 (Στοιχειώδες ανανεωτικό θεώρημα με κόστη) Έστω $\{C(t)\}$ ανανεωτική διαδικασία κόστους, με τυπικό ζεύγος διάρκειας ανανεωτικού κύκλου και αντίστοιχης αμοιβής (X, C) με κατανομή $F_{X,C}(x, y)$. Υποθέτουμε επίσης $E[X] < \infty$ και $E[C] < \infty$. Τότε:

(i) Για τον μακροπρόθεσμο ρυθμό κόστους έχουμε

$$\lim_{t \rightarrow \infty} \frac{C(t)}{t} = \frac{E[C]}{E[X]}, \text{ με πιθανότητα } 1.$$

(ii) Για τον μακροπρόθεσμο μέσο ρυθμό κόστους έχουμε

$$\lim_{t \rightarrow \infty} \frac{E[C(t)]}{t} = \frac{E[C]}{E[X]}.$$

Τα αποτελέσματα αυτά έχουν πληθώρα εφαρμογών. Χρησιμοποιούνται σε πολλές περιπτώσεις ως εναλλακτικά του βασικού ανανεωτικού θεωρήματος και δίνουν ευκολότερα τα αποτελέσματα (στην πλειονότητα των περιπτώσεων). Ενδεικτικά, θα δούμε δυο παραδείγματα. ◊

Παράδειγμα 12.24 (Εναλλασσόμενη ανανεωτική διαδικασία) Θεωρούμε μια μηχανή η οποία έχει χρόνους ζωής (λειτουργίας) O_1, O_2, \dots που εναλλάσσονται με χρόνους αντικατάστασης (αργίας) D_1, D_2, \dots . Οι τυχαίες μεταβλητές (O_n, D_n) , $n \geq 1$, θεωρούνται ανεξάρτητες και ισόνομες με κατανομή $F_{O,D}(x, y)$. Θέλουμε να προσδιορίσουμε το μακροπρόθεσμο μέσο ποσοστό του χρόνου λειτουργίας της μηχανής.

Είναι φανερό ότι ξεκινώντας από μια μηχανή που έχει μόλις αρχίσει να λειτουργεί, οι ανανεωτικοί κύκλοι αρχίζουν κάθε φορά που η μηχανή ξεκινά έναν νέο χρόνο λειτουργίας. Επομένως, ο χρόνος του n -οστού ανανεωτικού γεγονότος, S_n , είναι ο χρόνος που τελειώνει ο n -στός χρόνος αντικατάστασης της μηχανής, οπότε η μηχανή ξαναμπαίνει σε λειτουργία. Επίσης, αφού μας ενδιαφέρει το μακροπρόθεσμο μέσο ποσοστό του χρόνου λειτουργίας της μηχανής, είναι φυσικό να ορίσουμε ως $C(t)$ το συνολικό χρόνο λειτουργίας της μηχανής στο $(0, t]$. Επομένως έχουμε ότι

$$\begin{aligned} X_n &= S_n - S_{n-1} = O_n + D_n, \quad n \geq 1, \\ C_n &= C(S_n) - C(S_{n-1}) = O_n, \quad n \geq 1. \end{aligned}$$

Επομένως, έχουμε άμεσα ότι οι (X_n, C_n) , $n \geq 1$, είναι ανεξάρτητες και ισόνομες, αφού οι (O_n, D_n) , $n \geq 1$, είναι ανεξάρτητες και ισόνομες. Συνεπώς, το στοιχειώδες ανανεωτικό θεώρημα με κόστη είναι εφαρμόσιμο και έχουμε ότι το μακροπρόθεσμο μέσο ποσοστό του χρόνου λειτουργίας της μηχανής είναι

$$\lim_{t \rightarrow \infty} \frac{E[C(t)]}{t} = \frac{E[C]}{E[X]} = \frac{E[O]}{E[O] + E[D]}.$$

◊

Παράδειγμα 12.25 (Οριακοί παρελθόντες, υπολειπόμενοι και t -εξαρτώμενοι χρόνοι ανανέωσης) Θεωρούμε μια ανανεωτική διαδικασία $\{N(t)\}$, με ενδιάμεσους χρόνους γεγονότων X_1, X_2, \dots με κατανομή $F_X(x)$, και χρόνους γεγονότων $S_0 = 0, S_1, S_2, \dots$. Ορίζουμε

$$\begin{aligned} A(t) &= t - S_{N(t)}, \quad t \geq 0, \\ R(t) &= S_{N(t)+1} - t, \quad t \geq 0, \\ T(t) &= S_{N(t)+1} - S_{N(t)}, \quad t \geq 0, \end{aligned}$$

την ηλικία (παρελθόντα ή αναδρομικό χρόνο ανανέωσης - το $A(t)$ παραπέμπει στον όρο *age process*), τον υπολειπόμενο χρόνο ανανέωσης (προδρομικό χρόνο ανανέωσης - το $R(t)$ παραπέμπει στον όρο *remaining renewal time process* ή *residual renewal time process*) και τον t -εξαρτώμενο (το $T(t)$ παραπέμπει στο *total renewal time process*), αντίστοιχα.

Η τυχαία μεταβλητή $A(t)$ μετράει το χρόνο από το τελευταίο γεγονός της ανανεωτικής διαδικασίας πριν τη στιγμή t μέχρι τη στιγμή t . Η τυχαία μεταβλητή $R(t)$ μετράει το χρόνο από τη στιγμή t μέχρι το επόμενο γεγονός της ανανεωτικής διαδικασίας. Τέλος, η τυχαία μεταβλητή $T(t)$ μετράει το μήκος του ενδιάμεσου χρόνου ανανέωσης που περιλαμβάνει τη χρονική στιγμή t .

Οι τρεις αυτές διαδικασίες παίζουν σημαντικό ρόλο στη μελέτη ουρών αναμονής που περιλαμβάνουν ανανεωτικές διαδικασίες. Π.χ., σε ένα σύστημα εξυπηρέτησης όπου οι αφίξεις συμβαίνουν σύμφωνα με μια ανανεωτική διαδικασία, η $R(t)$ μετράει το χρόνο από τη στιγμή t μέχρι την επόμενη άφιξη πελάτη.

Σκοπός μας είναι η μελέτη της μακροπρόθεσμης συμπεριφοράς αυτών των διαδικασιών. Ιδιαίτερα, θα επικεντρωθούμε στη μελέτη της $\{R(t)\}$, που είναι και η πιο σημαντική στις εφαρμογές. Αρχικά, θα προσδιορίσουμε τον μακροπρόθεσμο αναμενόμενο δειγματικό μέσο του υπολειπόμενου χρόνου ανανέωσης, δηλαδή την ποσότητα

$$\lim_{t \rightarrow \infty} \frac{E \left[\int_0^t R(u) du \right]}{t}.$$

Η ποσότητα αυτή είναι κατά κάποιο τρόπο η απάντηση στο ερώτημα: «Αν επιλεγεί τυχαία μια χρονική στιγμή, πόσος χρόνος απομένει μέχρι το επόμενο ανανεωτικό γεγονός;». Ένας άλλος τρόπος για να μοντελοποιηθεί το ίδιο

ερώτημα είναι να προσδιορίσουμε την ποσότητα $\lim_{t \rightarrow \infty} E[R(t)]$. Ο τρόπος που γίνεται αυτό αναπτύχθηκε στα Παραδείγματα 12.14 και 12.18 με τη διατύπωση κατάλληλης ανανεωτικής εξίσωσης και την εύρεση του ορίου της λύσης της με εφαρμογή του βασικού ανανεωτικού θεωρήματος.

Για να μελετήσουμε το πρόβλημα στο παρόν πλαίσιο ορίζουμε μια διαδικασία κόστους $C(t) = \int_0^t R(u)du$. Επομένως έχουμε ότι

$$\begin{aligned} C_n &= \int_{S_{n-1}}^{S_n} R(u)du = \int_{S_{n-1}}^{S_n} (X_n - (u - S_{n-1}))du \\ &= \int_0^{X_n} (X_n - u)du = \frac{X_n^2}{2}. \end{aligned}$$

Επομένως, αν $E[X_n] = \mu$ και $\text{Var}[X_n] = \sigma^2$, τότε εφαρμόζοντας το στοιχειώδες ανανεωτικό θεώρημα με κόστη έχουμε

$$\lim_{t \rightarrow \infty} \frac{E\left[\int_0^t R(u)du\right]}{t} = \frac{E[C_n]}{E[X_n]} = \frac{E[X_n^2]}{2E[X_n]} = \frac{\mu^2 + \sigma^2}{2\mu}, \quad (12.10)$$

που συμφωνεί με την (12.9).

Το γεγονός ότι ο μακροπρόθεσμος αναμενόμενος δειγματικός μέσος του υπολειπόμενου χρόνου ανανέωσης δίνεται από αυτόν το τύπο και επομένως είναι ίσος με $\frac{\mu}{2} + \frac{\sigma^2}{2\mu}$ και όχι με $\frac{\mu}{2}$ αναφέρεται ως ανανεωτικό παράδοξο. Πράγματι αν τα γεγονότα της ανανεωτικής διαδικασίας συμβαίνουν κατά μέσο όρο κάθε μ χρονικές μονάδες και επιλέξουμε μια χρονική στιγμή στην τύχη, φαίνεται λογικό ότι θα περιμένουμε $\frac{\mu}{2}$ χρονικές μονάδες για να δούμε το επόμενο γεγονός. Η αφελής αυτή διαίσθηση είναι σωστή μόνο όταν η ανανεωτική διαδικασία είναι προσδιοριστική (ντετερμινιστική), δηλαδή όταν τα ανανεωτικά γεγονότα συμβαίνουν ακριβώς κάθε μ χρονικές μονάδες (οπότε $\sigma = 0$). Όταν, όμως, υπάρχει τυχαιότητα και οι ενδιάμεσοι χρόνοι των γεγονότων δεν είναι ίσοι με μ , η τυχαία επιλεγμένη χρονική στιγμή παρατήρησης είναι πιθανότερο να πέσει σε μεγάλο ενδιάμεσο χρόνο ανανέωσης και επομένως και ο υπολειπόμενος χρόνος ανανέωσης θα είναι μεγάλος. Αυτή ακριβώς η παρατήρηση δείχνει ότι ο μακροπρόθεσμος αναμενόμενος δειγματικός μέσος του υπολειπόμενου χρόνου ανανέωσης πρέπει να είναι μεγαλύτερος του $\frac{\mu}{2}$ και ο παραπάνω υπολογισμός δίνει την ακριβή του τιμή. Για να γίνει ακόμα πιο κατανοητή αυτή η σκέψη, σκεφτείτε π.χ. την ακραία περίπτωση οι ενδιάμεσοι χρόνοι να είναι 0 με πιθανότητα $\frac{1}{2}$ και 2μ με πιθανότητα $\frac{1}{2}$. Τότε μια τυχαία επιλεγμένη χρονική στιγμή είναι αδύνατο να πέφτει σε ενδιάμεσο χρόνο 0, αλλά θα πέφτει πάντα σε ενδιάμεσο χρόνο 2μ , οπότε ο μέσος υπολειπόμενος χρόνος θα είναι λογικό να είναι $\frac{2\mu}{2} = \mu$. Πράγματι στην περίπτωση αυτή η διασπορά των ενδιάμεσων χρόνων είναι $\sigma^2 = \frac{1}{2}(0 - \mu)^2 + \frac{1}{2}(2\mu - \mu)^2 = \mu^2$ και ο τύπος δίνει $\frac{\mu^2 + \sigma^2}{2\mu} = \mu$.

Στη συνέχεια, προχωράμε στην οριακή μέση δειγματική συνάρτηση κατανομής του υπολειπόμενου χρόνου ανανέωσης, $F_R(x)$, που ορίζεται ως

$$F_R(x) = \lim_{t \rightarrow \infty} \frac{E\left[\int_0^t 1_{\{R(u) \leq x\}} du\right]}{t}, \quad x \in \mathbb{R}.$$

Με άλλα λόγια, η ποσότητα $F_R(x)$ εκφράζει το μακροπρόθεσμο μέσο ποσοστό του χρόνου που ο υπολειπόμενος χρόνος ανανέωσης είναι το πολύ x .

Για να μελετήσουμε το πρόβλημα ορίζουμε μια διαδικασία κόστους

$$C(t) = \int_0^t 1_{\{R(u) \leq x\}} du, \quad t \geq 0,$$

που για κάθε t καταγράφει τον χρόνο που ο υπολειπόμενος χρόνος ανανέωσης ήταν το πολύ x , στο διάστημα $(0, t]$. Επομένως, το κόστος που συσσωρεύεται σε έναν ανανεωτικό κύκλο ισούται με το διάστημα που ο υπολειπόμενος

χρόνος ανανέωσης είναι το πολύ x . Είναι φανερό ότι αν η διάρκεια του κύκλου, X_n , είναι μικρότερη ή ίση του x , τότε καθόλη τη διάρκεια του κύκλου ο υπολειπόμενος χρόνος ανανέωσης είναι το πολύ x και έχουμε $C_n = X_n$. Αν όμως $X_n > x$, τότε μόνο στις τελευταίες x χρονικές μονάδες του κύκλου ο υπολειπόμενος χρόνος ανανέωσης είναι το πολύ x και έχουμε $C_n = x$. Επομένως, είναι

$$C_n = \min(X_n, x).$$

Εφαρμόζοντας το στοιχειώδες ανανεωτικό θεώρημα με κόστη έχουμε

$$\begin{aligned} F_R(x) &= \lim_{t \rightarrow \infty} \frac{E \left[\int_0^t 1_{\{R(u) \leq x\}} du \right]}{t} = \frac{E[C_n]}{E[X_n]} \\ &= \frac{E[\min(X_n, x)]}{E[X_n]} = \frac{\int_0^\infty \Pr[\min(X_n, x) > t] dt}{\mu} \\ &= \frac{\int_0^x (1 - F_X(t)) dt}{\mu}. \end{aligned} \quad (12.11)$$

Η κατανομή αυτή αναφέρεται ως κατανομή ισορροπίας της συνάρτησης κατανομής $F_X(x)$. Η σύνδεση, λοιπόν, των δυο κατανομών μέσα στο πλαίσιο της ανανεωτικής θεωρίας είναι η εξής: Αν η $F_X(x)$ είναι η κατανομή των ενδιάμεσων χρόνων μιας ανανεωτικής διαδικασίας, η αντίστοιχη κατανομή ισορροπίας $F_R(x)$ είναι η κατανομή του χρόνου μέχρι το επόμενο γεγονός της ανανεωτικής διαδικασίας, αν την κοιτάξουμε μια τυχαία χρονική στιγμή.

Αντί του στοιχειώδους ανανεωτικού θεωρήματος με κόστη, θα μπορούσαμε να αιτιολογήσουμε τον υπολογισμό της $F_R(x)$ με βάση το εργοδικό θεώρημα των αναγεννητικών διαδικασιών (θεώρημα 1.5). Άλλωστε, τα δυο αυτά θεωρήματα είναι στενά συνδεδεμένα και αποτελούν διαφορετικές διατυπώσεις της ίδιας μαθηματικής ιδέας. Εναλλακτικά, αντί της $F_R(x)$ μπορεί κανείς να υπολογίσει το $\lim_{t \rightarrow \infty} \Pr[R(t) \leq x]$ και να καταλήξει στην ίδια έκφραση με την (12.11). Αυτό γίνεται σχηματίζοντας κατάλληλη ανανεωτική εξίσωση για την $\Pr[R(t) \leq x]$ και κατόπιν βρίσκοντας το όριο της λύσης της με βάση το βασικό ανανεωτικό θεώρημα. Είναι μάλιστα πιο βολικό να εργαστεί μελετώντας την πιθανότητα του συμπληρωματικού ενδεχομένου, δηλαδή την $\Pr[R(t) > x]$ (βλέπε άσκηση 12.14).
◊

12.3 Ασκήσεις

Άσκηση 12.1 Δίνεται η Μαρκοβιανή αλυσίδα διακριτού χρόνου, με χώρο καταστάσεων $\mathcal{S} = \{1, 2, 3\}$ και πίνακα πιθανοτήτων μετάβασης

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

Να υπολογιστούν οι πιθανότητες μετάβασης n -οστής τάξης $p_{11}^{(n)}$ για κάθε $n \geq 0$.

Άσκηση 12.2 Έστω η Μαρκοβιανή αλυσίδα διακριτού χρόνου $\{X_n\}$, με χώρο καταστάσεων $\mathcal{S} = \{1, 2\}$ και πίνακα πιθανοτήτων μετάβασης

$$\mathbf{P} = \begin{pmatrix} \alpha & 1 - \alpha \\ 1 - \beta & \beta \end{pmatrix},$$

όπου $\alpha, \beta \in (0, 1)$. Έστω $\tilde{T}_j = \min\{n \geq 1 : X_n = j\}$, για κάθε $j \in \mathcal{S}$.

1. Να υπολογιστούν οι πιθανότητες μετάβασης n -οστής τάξης $p_{ij}^{(n)}$, για κάθε $i, j \in \mathcal{S}$ και $n \geq 0$.
2. Να υπολογιστούν οι πιθανότητες $\tilde{f}_{ij}^{(n)} = \Pr[\tilde{T}_j = n | X_0 = i]$, για κάθε $i, j \in \mathcal{S}$ και $n \geq 1$.

3. Να υπολογιστούν οι μέσες τιμές $\bar{m}_{ij} = E[\tilde{T}_j | X_0 = i]$, για κάθε $i, j \in \mathcal{S}$.

Άσκηση 12.3 Θεωρούμε ανεξάρτητες ρίψεις ενός νομίσματος με πιθανότητα κορώνας p και πιθανότητα γραμμάτων $q = 1 - p$ σε κάθε ρίψη. Οι ρίψεις επαναλαμβάνονται μέχρι να εμφανιστούν για πρώτη φορά είτε r συνεχόμενες κορώνες είτε m συνεχόμενα γράμματα. Να υπολογιστεί η πιθανότητα να εμφανιστούν r συνεχόμενες κορώνες πριν εμφανιστούν m συνεχόμενα γράμματα.

Άσκηση 12.4 Έστω $\{X_n\}$ ο απλός τυχαίος περίπατος στο σύνολο $\{0, 1, 2, \dots, N\}$ με απορροφητικά φράγματα στο 0 και στο N . Δηλαδή, οι μη-μηδενικές πιθανότητες μετάβασης είναι οι $p_{00} = p_{NN} = 1$ και $p_{i,i+1} = p, p_{i,i-1} = q$ για $i = 1, 2, \dots, N - 1$. Έστω T ο χρόνος μέχρι την απορρόφηση, δηλαδή $T = \min\{n \geq 0 : X_n = 0 \text{ ή } N\}$.

1. Αποδείξτε ότι, για $q \neq p$, η πιθανότητα απορρόφησης στο 0 είναι

$$\Pr[X_T = 0 | X_0 = i] = \frac{\left(\frac{q}{p}\right)^N - \left(\frac{q}{p}\right)^i}{\left(\frac{q}{p}\right)^N - 1}, \quad 0 \leq i \leq N.$$

2. Βρείτε την $\Pr[X_T = 0 | X_0 = i], 0 \leq i \leq N$, όταν $q = p$.

3. Αποδείξτε ότι, για $q \neq p$, ο μέσος χρόνος απορρόφησης είναι

$$E[T | X_0 = i] = \frac{i}{q - p} - \frac{N}{q - p} \cdot \frac{1 - (q/p)^i}{1 - (q/p)^N}, \quad 0 \leq i \leq N.$$

4. Βρείτε την $E[T | X_0 = i], 0 \leq i \leq N$, όταν $q = p$.

Άσκηση 12.5 Να βρείτε τις κλάσεις επικοινωνίας και να κατατάξετε ως προς την περιοδικότητα/απεριοδικότητα και τη θετική επαναληπτικότητα/μηδενική επαναληπτικότητα/παροδικότητα τις καταστάσεις των Μαρκοβιανών αλυσίδων διακριτού χρόνου με χώρο καταστάσεων $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$ και πίνακες πιθανοτήτων μετάβασης

$$\mathbf{P}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{6} & 0 & \frac{1}{2} & 0 & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix} \quad \text{και} \quad \mathbf{P}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{8} & \frac{3}{8} & 0 \\ \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Άσκηση 12.6 Δίνεται η αδιαχώριστη Μαρκοβιανή αλυσίδα διακριτού χρόνου με χώρο καταστάσεων τους μη-αρνητικούς ακεραίους και τις ακόλουθες μη-μηδενικές πιθανότητες μετάβασης:

$$\begin{aligned} p_{0i} &= p_i, \quad i \geq 0, \\ p_{i,i-1} &= 1, \quad i \geq 1. \end{aligned}$$

Έστω ότι $p_i > 0$ για κάθε $i \geq 0$ και $\sum_{i=0}^{\infty} p_i = 1$. Τι είναι η αλυσίδα ως προς την περιοδικότητα: απεριοδική ή περιοδική; Να διατυπώσετε ικανές κι αναγκαίες συνθήκες για τα p_i ώστε η αλυσίδα να είναι θετικά επαναληπτική, μηδενικά επαναληπτική ή παροδική. Ποια είναι η στάσιμη κατανομή της, όταν είναι θετικά επαναληπτική;

Άσκηση 12.7 Δυο δοχεία A και B περιέχουν συνολικά N σφαιρίδια. Εκτελούμε το ακόλουθο πείραμα τύχης: Σε κάθε βήμα ένα σφαιρίδιο από τα N επιλέγεται τυχαία και μετά τοποθετείται στο A ή στο B με πιθανότητες p και q αντίστοιχα. Έστω X_n ο αριθμός των σφαιριδίων στο δοχείο A μετά το n -οστό βήμα. Να βρεθούν τα μακροπρόθεσμα ποσοστά του χρόνου (δηλαδή των βημάτων) που το δοχείο A θα περιέχει i σφαιρίδια για $i = 0, 1, 2, \dots, N$.

Άσκηση 12.8 Θεωρούμε μια συνήθη άδεια 8×8 σκακιέρα, κι έναν Βασιλιά, ο οποίος αρχίζει να κινείται σε αυτήν τυχαία, επιλέγοντας σε κάθε βήμα του ένα από τα επιτρεπόμενα τετράγωνα ισοπίθανα.

1. Να βρεθεί ο μέσος αριθμός κινήσεων που θα χρειαστεί για να επιστρέψει στην κάτω δεξιά γωνία για πρώτη φορά, ξεκινώντας από αυτή.
2. Να λυθεί το ίδιο πρόβλημα για έναν Ίππο και για έναν Αξιωματικό.

Άσκηση 12.9 Έστω ανανεωτική διαδικασία $\{N(t)\}$ με κατανομή ενδιάμεσων χρόνων $Erlang(r, \lambda)$, δηλαδή, με συνάρτηση πυκνότητας πιθανότητας

$$f_X(t) = \frac{\lambda^r}{(r-1)!} t^{r-1} e^{-\lambda t}, \quad t \geq 0,$$

και συνάρτηση κατανομής

$$F_X(t) = 1 - \sum_{j=0}^{r-1} e^{-\lambda t} \frac{(\lambda t)^j}{j!}, \quad t \geq 0.$$

Να υπολογιστούν η συνάρτηση κατανομής του χρόνου του k -οστού γεγονότος, $F_{S_k}(t)$ και η συνάρτηση πιθανότητας του αριθμού των γεγονότων τη στιγμή t , $(p_k(t) : k \geq 0)$. Να αποδειχθεί ότι η ανανεωτική συνάρτηση δίνεται από τον τύπο

$$m(t) = \sum_{k=1}^{\infty} \left\{ 1 - \sum_{j=0}^{kr-1} e^{-\lambda t} \frac{(\lambda t)^j}{j!} \right\}, \quad t \geq 0.$$

Να υπολογιστούν, επίσης, οι αντίστοιχοι μετασχηματισμοί Laplace - Stieltjes $\tilde{F}_{S_k}(s)$, $(\tilde{p}_k(s) : k \geq 0)$ και $\tilde{m}(s)$.

Άσκηση 12.10 Έστω ανανεωτική διαδικασία $\{N(t)\}$ με κατανομή ενδιάμεσων χρόνων $Hyperexp(p, 1-p, \lambda, \mu)$, δηλαδή, μίξη δυο κατανομών $Exp(\lambda)$ και $Exp(\mu)$, με πιθανότητες p και $1-p$ αντίστοιχα. Η συνάρτηση πυκνότητας πιθανότητας των ενδιάμεσων χρόνων είναι, επομένως,

$$f_X(t) = p\lambda e^{-\lambda t} + (1-p)\mu e^{-\mu t}, \quad t \geq 0,$$

και η συνάρτηση κατανομής είναι

$$F_X(t) = p(1 - e^{-\lambda t}) + (1-p)(1 - e^{-\mu t}), \quad t \geq 0.$$

Να υπολογιστούν οι μετασχηματισμοί Laplace - Stieltjes $\tilde{F}_{S_k}(s)$, $(\tilde{p}_k(s) : k \geq 0)$ και $\tilde{m}(s)$, της συνάρτησης κατανομής του χρόνου του k -οστού γεγονότος, $F_{S_k}(t)$, της συνάρτησης πιθανότητας του αριθμού των γεγονότων τη στιγμή t , $(p_k(t) : k \geq 0)$ και της ανανεωτικής συνάρτησης $m(t)$, αντίστοιχα. Να αποδειχθεί ότι η ανανεωτική συνάρτηση είναι της μορφής

$$m(t) = A + Bt + Ce^{-(\lambda(1-p)+\mu p)t}, \quad t \geq 0,$$

και να υπολογιστούν οι σταθερές A , B και C .

Άσκηση 12.11 Έστω ανανεωτική διαδικασία $\{N(t)\}$, όπου ένας ενδιάμεσος χρόνος ανανέωσης είναι 0 με πιθανότητα p , και έχει την $Exp(\lambda)$ κατανομή με πιθανότητα $1-p$. Δηλαδή, η συνάρτηση κατανομής του είναι μικτή, με μάζα πιθανότητας p στο 0 και συνάρτηση πυκνότητας πιθανότητας $(1-p)\lambda e^{-\lambda t}$, $t \geq 0$. Η συνάρτηση κατανομής του είναι

$$F_X(t) = p + (1-p)(1 - e^{-\lambda t}), \quad t \geq 0.$$

Να υπολογιστούν οι μετασχηματισμοί Laplace - Stieltjes $\tilde{F}_{S_k}(s)$, $(\tilde{p}_k(s) : k \geq 0)$ και $\tilde{m}(s)$, της συνάρτησης κατανομής του χρόνου του k -οστού γεγονότος, $F_{S_k}(t)$, της συνάρτησης πιθανότητας του αριθμού των γεγονότων τη στιγμή t , $(p_k(t) : k \geq 0)$ και της ανανεωτικής συνάρτησης $m(t)$, αντίστοιχα. Να αποδειχθεί ότι η ανανεωτική συνάρτηση είναι της μορφής

$$m(t) = A + Bt, \quad t \geq 0,$$

και να υπολογιστούν οι σταθερές A και B .

Άσκηση 12.12 Έστω μια ανανεωτική διαδικασία $\{N(t)\}$ με κατανομή ενδιάμεσων χρόνων $F_X(t)$, με μέση τιμή μ και έστω $h(t) = E[S_{N(t)+1}]$, $t \geq 0$. Να διατυπωθεί μια ανανεωτική εξίσωση για την $h(t)$ και, λύνοντάς την, να αποδειχθεί ότι

$$h(t) = \mu(1 + m(t)), \quad t \geq 0,$$

όπου $m(t)$ η ανανεωτική συνάρτηση.

Άσκηση 12.13 Έστω μια ανανεωτική διαδικασία $\{N(t)\}$ με κατανομή ενδιάμεσων χρόνων $F_X(t)$ και έστω $h(t) = E[N(t)(N(t)-1)]$, $t \geq 0$. Να διατυπωθεί μια ανανεωτική εξίσωση για την $h(t)$ και, λύνοντάς την, να αποδειχθεί ότι

$$h(t) = 2(m * m)(t), \quad t \geq 0,$$

όπου $m(t)$ η ανανεωτική συνάρτηση.

Άσκηση 12.14 Έστω μια ανανεωτική διαδικασία $\{N(t)\}$, με χρόνους γεγονότων S_1, S_2, \dots , με συνεχή κατανομή ενδιάμεσων χρόνων $F_X(t)$ και ανανεωτική συνάρτηση $m(t)$. Έστω $R(t) = S_{N(t)+1} - t$ ο υπολειπόμενος χρόνος ανανέωσης τη στιγμή t . Θεωρούμε κάποιο σταθερό $x \geq 0$ και έστω $h(t) = \Pr[R(t) > x]$, $t \geq 0$. Να διατυπωθεί μια ανανεωτική εξίσωση για την $h(t)$ και, λύνοντάς την, να αποδειχθεί ότι

$$h(t) = \Pr[R(t) > x] = 1 - F_X(x+t) + \int_0^t (1 - F_X(x+t-u)) dm(u), \quad t, x \geq 0.$$

Να αποδειχθεί, επίσης, ότι

$$\lim_{t \rightarrow \infty} h(t) = \lim_{t \rightarrow \infty} \Pr[R(t) > x] = \frac{\int_x^\infty (1 - F_X(u)) du}{\mu}, \quad x \geq 0.$$

Έστω $A(t) = t - S_{N(t)}$ ο παρελθών χρόνος ανανέωσης τη στιγμή t . Δικαιολογήστε την ισότητα ενδεχομένων $\{A(t) > x\} = \{R(t-x) > x\}$ και, χρησιμοποιώντας την, βρείτε το $\lim_{t \rightarrow \infty} \Pr[A(t) > x]$. Ομοίως, δείξτε ότι

$$\lim_{t \rightarrow \infty} \Pr[A(t) > x, R(t) > y] = \frac{\int_{x+y}^\infty (1 - F_X(u)) du}{\mu}, \quad x, y \geq 0.$$

Άσκηση 12.15 Έστω μια ανανεωτική διαδικασία $\{N(t)\}$, με συνεχή κατανομή ενδιάμεσων χρόνων $F_X(t)$ με μέση τιμή $\mu \in (0, \infty)$ και διασπορά $\sigma^2 \in [0, \infty)$. Έστω, επίσης, $m(t)$ η ανανεωτική συνάρτηση και $h(t) = m(t) - \frac{t}{\mu}$, $t \geq 0$. Να αποδειχθεί ότι η $h(t)$ ικανοποιεί την ανανεωτική εξίσωση

$$h(t) = d(t) + \int_0^t h(t-u) dF_X(u) = d(t) + (h * F_X)(t), \quad t \geq 0,$$

με

$$\begin{aligned} d(t) &= \int_0^t \left(1 - \frac{u}{\mu}\right) dF_X(u) - \int_t^\infty \frac{t}{\mu} dF_X(u) \\ &= \frac{1}{\mu} \int_t^\infty (1 - F_X(u)) du - (1 - F_X(t)). \end{aligned}$$

Επίσης, αποδείξτε ότι

$$\lim_{t \rightarrow \infty} h(t) = \lim_{t \rightarrow \infty} \left(m(t) - \frac{t}{\mu}\right) = \frac{\sigma^2 - \mu^2}{2\mu^2},$$

δηλαδή η ευθεία $\frac{1}{\mu}t + \frac{\sigma^2 - \mu^2}{2\mu^2}$ είναι πλάγια ασύμπτωτη της $m(t)$.

12.4 Σχόλια

Τα εισαγωγικά βιβλία στοχαστικών διαδικασιών αναφέρονται στις Μαρκοβιανές αλυσίδες διακριτού χρόνου και στις ανανεωτικές διαδικασίες που είναι τα αντικείμενα του παρόντος κεφαλαίου. Για μια εκτενή εισαγωγή με πολλά παραδείγματα ο ενδιαφερόμενος αναγνώστης παραπέμπεται στα εξαιρετικά συγγράμματα των Ross 1995, Tijms 2008, Kulkarni 2010 και Grimmett και Stirzaker 2020.

Βιβλιογραφία

- [1] S. Ross. *Stochastic Processes, 2nd Edition*. Wiley, 1995. ISBN: 978-0471120629.
- [2] H.C. Tijms. *A First Course in Stochastic Models, 2nd Edition*. Chichester: Wiley, 2008. ISBN: 978-0471498803.
- [3] V.G. Kulkarni. *Modeling and Analysis of Stochastic Systems, 2nd Edition*. CRC Press, 2010. ISBN: 978-1439808757.
- [4] G. Grimmett και D. Stirzaker. *Probability and Random Processes, 4th Edition*. Oxford, 2020. ISBN: 978-0198847595.

ΑΝΑΛΥΣΗ ΜΕΣΗΣ ΤΙΜΗΣ ΓΙΑ ΜΗ-ΜΑΡΚΟΒΙΑΝΑ ΜΟΝΤΕΛΑ

Στο κεφάλαιο αυτό βλέπουμε πώς η μεθοδολογία της ανάλυσης μέσης τιμής που παρουσιάσαμε στο κεφάλαιο 5 επεκτείνεται για τον υπολογισμό του μέσου αριθμού πελατών, $E[Q]$, και του μέσου χρόνου παραμονής ενός πελάτη, $E[S]$, σε ένα μη-Μαρκοβιανό σύστημα εξυπηρέτησης. Η βασική ιδέα παραμένει η ίδια: Χρειαζόμαστε ένα σύστημα δυο εξισώσεων για τον προσδιορισμό αυτών των δύο αγνώστων. Η μια εξίσωση θα προέρχεται από τον νόμο του Little και η δεύτερη θα προκύπτει από τον υπολογισμό του μέσου χρόνου παραμονής, $E[S]$, ενός πελάτη, δεσμεύοντας στον αριθμό των πελατών Q^- που βρίσκει κατά την άφιξή του στο σύστημα. Για να λειτουργεί η μέθοδος, θα πρέπει ο αριθμός των πελατών σε στιγμές αφίξεων, Q^- , να έχει ίδια κατανομή με τον αριθμό των πελατών σε συνεχή χρόνο, Q , πράγμα που εξασφαλίζεται από την ιδιότητα PASTA στην περίπτωση που η διαδικασία των αφίξεων είναι Poisson. Για τον λόγο αυτό, η ανάλυση μέσης τιμής για μη-Μαρκοβιανά μοντέλα λειτουργεί μόνο για μοντέλα με Poisson αφίξεις και γενικούς χρόνους εξυπηρέτησης, δηλαδή σε μοντέλα τύπου M/G/1 και στις γενικεύσεις τους.

Η μελέτη των μοντέλων τύπου M/G/1 και των γενικεύσεών τους έχει σημαντική αξία, καθώς σε πολλά πραγματικά συστήματα, είναι λογικό να υποθεθεί ότι η διαδικασία αφίξεων είναι Poisson, αφού η διαδικασία αυτή προκύπτει όταν υπάρχει μια δυνητικά πολύ μεγάλη «δεξαμενή» πελατών και κάθε ένας πελάτης από αυτήν χρησιμοποιεί το σύστημα σε αραιά διάστημα. Στη γλώσσα των στοχαστικών διαδικασιών, το αντίστοιχο αποτέλεσμα είναι ότι η υπέρθεση ανεξάρτητων ανανεωτικών διαδικασιών τείνει στη διαδικασία Poisson με ρυθμό λ , εφόσον το πλήθος των διαδικασιών τείνει στο άπειρο με τέτοιο τρόπο, ώστε ο μέσος χρόνος μεταξύ διαδοχικών ανανεώσεων κάθε ανανεωτικής διαδικασίας να τείνει στο ∞ και επιπλέον ο μέσος χρόνος μεταξύ διαδοχικών ανανεώσεων της υπέρθεσης να τείνει στο $\frac{1}{\lambda}$. Από την άλλη μεριά, συχνά έχουμε αρκετά ακριβή πληροφορία για τους χρόνους εξυπηρέτησης των πελατών και όχι μόνο τη μέση διάρκειά τους, οπότε είναι χρήσιμο να τους μοντελοποιήσουμε με μια γενική κατανομή και όχι κατ' ανάγκη με την εκθετική κατανομή, που υπαγορεύει η αρχή της μέγιστης εντροπίας (βλέπε παράγραφο 3.3). Επομένως, το M/G/1 σύστημα εξυπηρέτησης και οι γενικεύσεις του είναι εύλογα μοντέλα για την αναπαράσταση πολλών συστημάτων εξυπηρέτησης που εμφανίζονται στις εφαρμογές. Στις επόμενες παραγράφους θα δούμε πώς ακριβώς προσαρμόζεται η ανάλυση μέσης τιμής για τέτοια μοντέλα.

13.1 Ανάλυση μέσης τιμής στην M/G/1/1 ουρά

Θεωρούμε μια M/G/1/1 ουρά με ρυθμό αφίξεων λ και γενικούς χρόνους εξυπηρέτησης με κατανομή $F_B(x)$ με μέση τιμή b . Συμβολίζουμε με $\rho = \lambda b$ τον ρυθμό συνωστισμού. Από τον νόμο του Little έχουμε

$$E[Q] = \lambda E[S].$$

Θεωρώντας τον χρόνο παραμονής, S , ενός πελάτη που φθάνει στο σύστημα και βρίσκει Q^- πελάτες σε αυτό έχουμε

$$E[S] = \Pr[Q^- = 0]E[S|Q^- = 0] + \Pr[Q^- = 1]E[S|Q^- = 1] = a_0 \cdot b + a_1 \cdot 0 = p_0 b,$$

όπου η τελευταία ισότητα ισχύει λόγω της PASTA. Συνδυάζοντας αυτές τις σχέσεις και λαμβάνοντας υπόψη ότι $E[Q] = 0 \cdot p_0 + 1 \cdot p_1 = p_1$, έχουμε $p_1 = \lambda b p_0 = \rho p_0$. Επιπλέον, έχουμε $p_0 + p_1 = 1$. Λύνοντας το σύστημα παίρνουμε

$$p_0 = \frac{1}{1 + \rho}, \quad p_1 = \frac{\rho}{1 + \rho},$$

οπότε

$$E[Q] = \frac{\rho}{1 + \rho}, \quad E[S] = \frac{1}{\mu(1 + \rho)}, \quad (13.1)$$

που ταυτίζονται με τις σχέσεις (5.1) που έχουμε βρει για την M/M/1/1 ουρά.

13.2 Ανάλυση μέσης τιμής στην M/G/1 ουρά

Θεωρούμε μια M/G/1 ουρά με ρυθμό αφίξεων λ και γενική κατανομή χρόνων εξυπηρέτησης $F_B(x)$. Θα συμβολίζουμε έναν γενικό χρόνο εξυπηρέτησης με B . Έστω $b = E[B]$, $b_2 = E[B^2]$, $\sigma_B^2 = \text{Var}[B]$ και $\rho = \lambda b$ ο ρυθμός συνωστισμού. Για την ανάλυση μέσης τιμής, μια σχέση προκύπτει από τον νόμο του Little, οπότε έχουμε

$$E[Q] = \lambda E[S].$$

Για να βρούμε μια δεύτερη σχέση, υπολογίζουμε τον μέσο χρόνο παραμονής ενός επιλεγμένου πελάτη, δεσμεύοντας στον αριθμό των πελατών Q^- που βλέπει κατά την άφιξή του. Έτσι έχουμε

$$S = \sum_{i=1}^{Q^-+1} B_i,$$

όπου B_1, B_2, \dots είναι οι χρόνοι εξυπηρέτησης των παρόντων πελατών, κατά την άφιξη του επιλεγμένου πελάτη. Συγκεκριμένα, ο B_1 αναφέρεται στον πελάτη που βρίσκεται ήδη σε εξυπηρέτηση κατά την άφιξη του επιλεγμένου πελάτη και επομένως πρόκειται για υπολειπόμενο χρόνο εξυπηρέτησης, ενώ οι B_2, B_3, \dots, B_{Q^-} στους χρόνους εξυπηρέτησης των πελατών που βρίσκονται στον χώρο αναμονής και επομένως είναι ολόκληροι χρόνοι εξυπηρέτησης. Τέλος, ο B_{Q^-+1} αντιστοιχεί στο χρόνο εξυπηρέτησης του επιλεγμένου πελάτη.

Δεσμεύοντας στην Q^- , όπως κάναμε στην ανάλυση της μέσης τιμής για την M/M/1 ουρά, έχουμε ότι

$$E[S] = \sum_{n=0}^{\infty} \Pr[Q^- = n] \sum_{i=1}^{n+1} E[B_i | Q^- = n].$$

Για όλα τα B_2, B_3, \dots έχουμε ότι είναι ανεξάρτητα από την Q^- και άρα $E[B_i | Q^- = n] = E[B_i] = b$. Όμως, η ανάλυση που κάναμε για την M/M/1 ουρά δεν μπορεί να προχωρήσει πανομοιότυπα, διότι ο B_1 είναι υπολειπόμενος χρόνος εξυπηρέτησης και εξαρτάται από την Q^- , αφού δεν ισχύει η αμνήμονη ιδιότητα δεδομένου

του ότι η $B(x)$ δεν είναι η εκθετική κατανομή. Για να ξεπεράσουμε αυτό το πρόβλημα, αντί να δεσμεύσουμε στην ακριβή τιμή της Q^- δεσμεύουμε στο κατά πόσον είναι ίση με 0 ή αυστηρά θετική, δηλαδή έχουμε

$$E[S] = \Pr[Q^- = 0]E[B_1|Q^- = 0] + \Pr[Q^- > 0] \left(E[B_1|Q^- > 0] + E \left[\sum_{i=2}^{Q^-+1} B_i | Q^- > 0 \right] \right). \quad (13.2)$$

Εδώ, χρησιμοποιώντας τον τύπο (2.8) για την πιθανότητα κενού συστήματος της G/G/1 ουράς και την ιδιότητα PASTA (βλέπε θεώρημα 2.3) έχουμε $\Pr[Q^- = 0] = a_0 = p_0 = 1 - \rho$ και $\Pr[Q^- > 0] = 1 - a_0 = 1 - p_0 = \rho$. Επίσης, $E[B_1|Q^- = 0] = b$, διότι στην περίπτωση που ο επιλεγμένος πελάτης βρίσκει το σύστημα κενό, ο B_1 είναι ο δικός του χρόνος εξυπηρέτησης και όχι ένας υπολειπόμενος χρόνος εξυπηρέτησης και άρα η μέση τιμή του είναι b . Επίσης, $E[B_i|Q^- > 0] = E[B_i] = b$, για $i = 2, 3, \dots$, καθώς οι B_2, B_3, \dots είναι ανεξάρτητοι από την Q^- . Επομένως, η (13.2) γίνεται

$$\begin{aligned} E[S] &= (1 - \rho)b + \rho (E[B_1|Q^- > 0] + E[Q^-|Q^- > 0]b) \\ &= (1 - \rho)b + \rho E[B_1|Q^- > 0] + \Pr[Q^- > 0]E[Q^-|Q^- > 0]b. \end{aligned} \quad (13.3)$$

Μένει να βρούμε την $E[B_1|Q^- > 0]$. Η πληροφορία $Q^- > 0$ μας λέει ότι ο επιλεγμένος πελάτης φθάνει κατά τη διάρκεια μιας περιόδου συνεχούς λειτουργίας του συστήματος. Αν παρατηρούμε το σύστημα μόνο κατά τη διάρκεια των περιόδων συνεχούς λειτουργίας του συστήματος, τότε οι στιγμές που ολοκληρώνονται οι εξυπηρετήσεις είναι οι χρόνοι γεγονότων μιας ανανεωτικής διαδικασίας με ενδιάμεσους χρόνους τους χρόνους εξυπηρέτησης του συστήματος. Αυτό συμβαίνει επειδή ο υπηρέτης εξυπηρετεί συνεχώς αν κοιτάζουμε το σύστημα μόνο κατά τη διάρκεια των περιόδων συνεχούς λειτουργίας. Επομένως, ο υπολειπόμενος χρόνος εξυπηρέτησης αν κοιτάζουμε το σύστημα σε μια τυχαία χρονική στιγμή έχει την ίδια κατανομή με την οριακή κατανομή του υπολειπόμενου χρόνου ανανέωσης, R_B , αυτής της ανανεωτικής διαδικασίας (όπου το B εισάγεται ως δείκτης για να θυμίζει τους ενδιάμεσους χρόνους της ανανεωτικής διαδικασίας στην οποία αναφερόμαστε). Αυτή είναι η κατανομή ισορροπίας της ανανεωτικής διαδικασίας που δίνεται στην περίπτωση μας από τον τύπο

$$F_{R_B}(x) = \frac{\int_0^x (1 - F_B(t)) dt}{b}, \quad (13.4)$$

εφαρμόζοντας τη σχέση (12.11). Η μέση τιμή είναι η

$$E[R_B] = \frac{b_2}{2b} = \frac{b^2 + \sigma_B^2}{2b} \quad (13.5)$$

(εφαρμόζοντας τη σχέση (12.10) ή την (12.9)). Επειδή οι αφίξεις κατά τη διάρκεια των περιόδων συνεχούς απασχόλησης γίνονται σύμφωνα με διαδικασία Poisson, από την γενικευμένη ιδιότητα PASTA (θεώρημα 2.5), η κατανομή του υπολειπόμενου χρόνου εξυπηρέτησης αν κοιτάζουμε το σύστημα σε μια τυχαία στιγμή άφιξης ταυτίζεται με την κατανομή σε τυχαία χρονική στιγμή. Έτσι έχουμε ότι

$$E[B_1|Q^- > 0] = E[R_B] = \frac{b^2 + \sigma_B^2}{2b}. \quad (13.6)$$

Επομένως, αντικαθιστώντας την (13.6) στην (13.3) έχουμε

$$E[S] = (1 - \rho)b + \rho E[R_B] + \Pr[Q^- > 0]E[Q^-|Q^- > 0]b. \quad (13.7)$$

Όμως, $\Pr[Q^- > 0]E[Q^-|Q^- > 0] = E[Q^- \mathbf{1}_{\{Q^- > 0\}}] = E[Q^-] = E[Q]$, οπότε πολλαπλασιάζοντας με λ τη σχέση (13.7), και χρησιμοποιώντας τον νόμο του Little έχουμε

$$E[Q] = (1 - \rho)\rho + \lambda\rho E[R_B] + \rho E[Q].$$

Λύνοντας ως προς $E[Q]$, παίρνουμε τον μέσο αριθμό πελατών στο σύστημα

$$E[Q] = \rho + \frac{\rho}{1-\rho} \lambda E[R_B],$$

και διαιρώντας με λ παίρνουμε και τον μέσο χρόνο παραμονής πελάτη στο σύστημα

$$E[S] = b + \frac{\rho}{1-\rho} E[R_B].$$

13.3 Ανάλυση μέσης τιμής στην M/G/1 ουρά με πολλαπλές διακοπές υπηρετή

Θεωρούμε την τροποποίηση μιας M/G/1 ουράς με ρυθμό αφίξεων λ και γενική κατανομή χρόνων εξυπηρέτησης $F_B(x)$, όπου κάθε φορά που το σύστημα μένει κενό ο υπηρετής ξεκινά μια περίοδο διακοπών. Όταν λήξει ο χρόνος διακοπών, ο υπηρετής επιστρέφει και αν το σύστημα εξακολουθεί να είναι άδειο αρχίζει μια νέα περίοδο διακοπών κ.ο.κ. (γι αυτό και μιλάμε για πολλαπλές διακοπές), ενώ αν στο σύστημα υπάρχουν πελάτες σε αναμονή ξεκινά να εξυπηρετεί. Οι διάρκειες των περιόδων διακοπών ακολουθούν μια γενική κατανομή $F_V(x)$. Συμβολίζουμε με B και V έναν γενικό χρόνο εξυπηρέτησης και έναν γενικό χρόνο διακοπών, αντίστοιχα. Επίσης, συμβολίζουμε με R_B και R_V τους αντίστοιχους οριακούς υπολειπόμενους χρόνους ανανέωσης σε ανανεωτικές διαδικασίες με ενδιάμεσους χρόνους με κατανομή $F_B(x)$ και $F_V(x)$, αντίστοιχα. Θέτουμε $b = E[B]$, $\sigma_B^2 = Var[B]$, $v = E[V]$, και $\sigma_V^2 = Var[V]$. Η ανάλυση μέσης τιμής γι αυτό το μοντέλο που αναφέρεται ως η M/G/1 ουρά με πολλαπλές διακοπές υπηρετή ακολουθεί τη συλλογιστική που είδαμε για την απλή M/G/1 ουρά.

Θεωρώντας έναν επιλεγμένο πελάτη, θέτοντας S τον χρόνο παραμονής του και Q^- το πλήθος των πελατών που βρίσκει κατά την άφιξή του στο σύστημα έχουμε ότι

$$S = 1_{\{Q^-=0\}} R_V + \sum_{i=1}^{Q^-+1} B_i,$$

όπου $B_1, B_2, \dots, B_{Q^-+1}$ είναι οι χρόνοι εξυπηρέτησης των παρόντων πελατών, αμέσως μετά την άφιξη του επιλεγμένου πελάτη και R_V ο υπολειπόμενος χρόνος διακοπών του υπηρετή. Επομένως, κατά τον υπολογισμό της μέσης τιμής $E[S]$, παίρνουμε τη σχέση

$$\begin{aligned} E[S] &= \Pr[Q^- = 0] E[R_V + B_1 | Q^- = 0] \\ &+ \Pr[Q^- > 0] \left(E[B_1 | Q^- > 0] + E \left[\sum_{i=2}^{Q^-+1} B_i | Q^- > 0 \right] \right), \end{aligned} \quad (13.8)$$

αντί της (13.2) που έχουμε στην κλασική M/G/1 ουρά. Από εκεί και πέρα η ανάλυση είναι πανομοιότυπη, οπότε έχουμε τη σχέση

$$E[S] = (1-\rho)(E[R_V] + b) + \rho E[R_B] + \Pr[Q^- > 0] E[Q^- | Q^- > 0] b, \quad (13.9)$$

που είναι η αντίστοιχη της (13.7) που είδαμε στην κλασική M/G/1 ουρά. Πολλαπλασιάζοντας με λ και χρησιμοποιώντας τον νόμο του Little, παίρνουμε

$$E[Q] = (1-\rho)(\lambda E[R_V] + \rho) + \lambda \rho E[R_B] + \rho E[Q].$$

Λύνοντας ως προς $E[Q]$, παίρνουμε τον μέσο αριθμό πελατών στο σύστημα

$$E[Q] = \lambda E[R_V] + \rho + \frac{\rho}{1-\rho} \lambda E[R_B],$$

και διαιρώντας με λ παίρνουμε και τον μέσο χρόνο παραμονής πελάτη στο σύστημα

$$E[S] = E[R_V] + b + \frac{\rho}{1-\rho} E[R_B],$$

με

$$E[R_B] = \frac{b^2 + \sigma_B^2}{2b}, \quad E[R_V] = \frac{v^2 + \sigma_V^2}{2v}.$$

13.4 Ανάλυση μέσης τιμής στην M/G/1 ουρά με την K-πολιτική ενεργοποίησης

Θεωρούμε την τροποποίηση μιας M/G/1 ουράς με ρυθμό αφίξεων λ και γενική κατανομή χρόνων εξυπηρέτησης $F_B(x)$, όπου κάθε φορά που το σύστημα μένει κενό ο υπηρέτης απενεργοποιείται και ενεργοποιείται ξανά όταν συσσωρευθούν K πελάτες με $K \geq 1$ (για $K = 1$ έχουμε την κλασική M/G/1 ουρά). Θα συμβολίζουμε έναν γενικό χρόνο εξυπηρέτησης με B . Έστω $b = E[B]$, $b_2 = E[B^2]$, $\sigma_B^2 = \text{Var}[B]$ και $\rho = \lambda b$ ο ρυθμός συνωστισμού. Θα αναφέρουμε αυτό το σύστημα ως την M/G/1 ουρά με την K-πολιτική ενεργοποίησης.

Για την ανάλυση μέσης τιμής, θεωρούμε έναν επιλεγμένο πελάτη και θέτουμε S τον χρόνο παραμονής του και Q^- το πλήθος των πελατών που βρίσκει κατά την άφιξή του στο σύστημα. Επιπλέον, συμβολίζουμε με I^- την κατάσταση του υπηρέτη που βρίσκει κατά την άφιξή του με

$$I^- = \begin{cases} 1 & \text{αν ο υπηρέτης παρέχει εξυπηρέτηση,} \\ 0 & \text{αν ο υπηρέτης δεν παρέχει εξυπηρέτηση.} \end{cases}$$

Αν ο πελάτης φθάσει και βρει τον υπηρέτη απενεργοποιημένο, δεδομένου ότι δεν έχουν συσσωρευθεί ακόμη K πελάτες από την τελευταία ενεργοποίηση, θα χρειαστεί να περιμένει και τον υπολειπόμενο χρόνο ενεργοποίησης, R_A , πριν αρχίσουν οι εξυπηρετήσεις των παρόντων πελατών. Επομένως, για τον χρόνο παραμονής έχουμε τη σχέση

$$S = 1_{\{I^-=0\}} R_A + \sum_{i=1}^{Q^-+1} B_i,$$

όπου B_1, B_2, \dots είναι οι χρόνοι εξυπηρέτησης των παρόντων πελατών, αμέσως μετά την άφιξη του επιλεγμένου πελάτη.

Επομένως, κατά τον υπολογισμό της μέσης τιμής $E[S]$, παίρνουμε τη σχέση

$$E[S] = \Pr[I^- = 0]E[R_A] + E\left[\sum_{i=1}^{Q^-+1} B_i\right]. \quad (13.10)$$

Για να προχωρήσει η ανάλυση θα πρέπει καταρχήν να υπολογίσουμε την πιθανότητα $\Pr[I^- = 0]$ και τη μέση τιμή $E[R_A]$.

Για την $\Pr[I^- = 0]$, εφαρμόζουμε τον νόμο του Little στο χώρο εξυπηρέτησης, θεωρώντας ότι ένας πελάτης βρίσκεται στο χώρο εξυπηρέτησης μόνο εφόσον του παρέχεται εξυπηρέτηση. Έτσι, έχουμε τη σχέση (2.7), που στη συγκεκριμένη περίπτωση γίνεται

$$\Pr[I = 1] = E[I] = E[Q_s] = \lambda E[B] = \lambda b = \rho$$

και επομένως $\Pr[I = 0] = 1 - \rho$. Εφαρμόζοντας τη γενικευμένη ιδιότητα PASTA (θεώρημα 2.5) έχουμε ότι οι πιθανότητες αυτές είναι οι ίδιες με τις αντίστοιχες σε στιγμές αφίξεων πελατών, οπότε

$$\Pr[I^- = 0] = 1 - \rho, \quad (13.11)$$

$$\Pr[I^- = 1] = \rho. \quad (13.12)$$

Για την $E[R_A]$, ας θεωρήσουμε έναν αφικνούμενο πελάτη που βρίσκει Q^- πελάτες στο σύστημα κατά την άφιξή του, δεδομένου ότι ο υπηρέτης δεν παρέχει εξυπηρέτηση. Τότε αυτός θα πρέπει να περιμένει $K - Q^- - 1$ αφίξεις ώστε να μαζευτούν οι K πελάτες που απαιτούνται για την ενεργοποίηση του υπηρέτη. Λόγω του ότι ο μέσος χρόνος μεταξύ δυο διαδοχικών αφίξεων είναι $\frac{1}{\lambda}$, ο μέσος χρόνος για την ενεργοποίηση του υπηρέτη είναι $(K - Q^- - 1)\frac{1}{\lambda}$. Επομένως, έχουμε ότι

$$E[R_A] = (K - E[Q^-|I^- = 0] - 1)\frac{1}{\lambda}. \quad (13.13)$$

Όμως, λόγω της PASTA, έχουμε ότι $E[Q^-|I^- = 0] = E[Q|I = 0]$. Για τον υπολογισμό της $E[Q|I = 0]$, έχουμε ότι αν παρατηρούμε το σύστημα μόνο στις περιόδους που ο υπηρέτης είναι ανενεργός ($I = 0$) τότε το πλήθος των πελατών περνάει διαδοχικά από τις καταστάσεις $0, 1, 2, \dots, K-1, 0, 1, 2, \dots, K-1, 0, \dots$ και σε κάθε κατάσταση μένει για $\text{Exp}(\lambda)$ χρόνο. Επομένως, η δεσμευμένη κατανομή του πλήθους των πελατών, δεδομένου ότι $I = 0$, είναι διακριτή ομοιόμορφη στο $\{0, 1, 2, \dots, K-1\}$, οπότε $E[Q|I = 0] = \frac{K-1}{2}$. Οπότε η (13.13) γίνεται

$$E[R_A] = \frac{K-1}{2} \cdot \frac{1}{\lambda}. \quad (13.14)$$

Αντικαθιστώντας τις (13.11) και (13.14) στην (13.10), παίρνουμε

$$E[S] = (1-\rho) \frac{K-1}{2\lambda} + E \left[\sum_{i=1}^{Q^-+1} B_i \right]. \quad (13.15)$$

Ο υπολογισμός του $E \left[\sum_{i=1}^{Q^-+1} B_i \right]$, γίνεται από εδώ και πέρα όπως ο αντίστοιχος υπολογισμός για την κλασική M/G/1 ουρά. Ακριβώς με τη συλλογιστική που αναπτύχθηκε εκεί (βλέπε αιτιολόγηση της (13.3)) έχουμε

$$\begin{aligned} E \left[\sum_{i=1}^{Q^-+1} B_i \right] &= \Pr[I^- = 0] E \left[\sum_{i=1}^{Q^-+1} B_i | I^- = 0 \right] \\ &\quad + \Pr[I^- = 1] \left(E[B_1 | I^- = 1] + E \left[\sum_{i=2}^{Q^-+1} B_i | I^- = 1 \right] \right) \\ &= (1-\rho)b + \rho E[R_B] + E[Q]b. \end{aligned} \quad (13.16)$$

Αντικαθιστώντας τώρα την (13.16) στην (13.15) έχουμε

$$E[S] = (1-\rho) \frac{K-1}{2\lambda} + (1-\rho)b + \rho E[R_B] + E[Q]b. \quad (13.17)$$

Πολλαπλασιάζοντας με λ και χρησιμοποιώντας τον νόμο του Little, παίρνουμε

$$E[Q] = (1-\rho) \frac{K-1}{2} + (1-\rho)\rho + \rho\lambda E[R_B] + E[Q]\rho.$$

Λύνοντας ως προς $E[Q]$, παίρνουμε τον μέσο αριθμό πελατών στο σύστημα

$$E[Q] = \frac{K-1}{2} + \rho + \frac{\rho}{1-\rho} \lambda E[R_B]$$

και διαιρώντας με λ παίρνουμε και τον μέσο χρόνο παραμονής πελάτη στο σύστημα

$$E[S] = \frac{K-1}{2\lambda} + b + \frac{\rho}{1-\rho} E[R_B].$$

13.5 Ασκήσεις

Άσκηση 13.1 Θεωρούμε ένα σύστημα εξυπηρέτησης με έναν υπηρέτη και C κλάσεις πελατών $1, 2, \dots, C$. Οι πελάτες της κλάσης i φθάνουν στο σύστημα σύμφωνα με μια διαδικασία Poisson με ρυθμό $\lambda^{(i)}$ και έχουν χρόνους εξυπηρέτησης με κατανομή $F_{B^{(i)}}(x)$. Συμβολίζουμε έναν γενικό χρόνο εξυπηρέτησης για την κλάση i με $B^{(i)}$ και θέτουμε $b^{(i)} = E[B^{(i)}]$, $b_2^{(i)} = E[(B^{(i)})^2]$, $\sigma_B^{2(i)} = \text{Var}[B^{(i)}]$ και $\rho^{(i)} = \lambda^{(i)} b^{(i)}$. Οι πελάτες κάθε κλάσης i έχουν απόλυτη προτεραιότητα (preemptive priority) έναντι των πελατών των κλάσεων j με $j > i$, με την έννοια ότι προηγούνται

στο να εξυπηρετηθούν από τους πελάτες των κλάσεων αυτών και μάλιστα αφίξεις πελατών υψηλότερης προτεραιότητας επάγουν τη διακοπή της εξυπηρέτησης πελατών χαμηλότερης προτεραιότητας. Οι πελάτες της ίδιας κλάσης εξυπηρετούνται σύμφωνα με την πειθαρχία FCFS. Το σύστημα αυτό αναφέρεται ως η $M/G/1$ ουρά με απόλυτες προτεραιότητες. Να βρεθούν οι μέσοι αριθμοί πελατών των διαφόρων κλάσεων του συστήματος και οι μέσοι χρόνοι παραμονής των πελατών των διαφόρων κλάσεων.

Άσκηση 13.2 Θεωρούμε ένα σύστημα εξυπηρέτησης με έναν υπηρέτη και C κλάσεις πελατών $1, 2, \dots, C$. Οι πελάτες της κλάσης i φθάνουν στο σύστημα σύμφωνα με μια διαδικασία Poisson με ρυθμό $\lambda^{(i)}$ και έχουν χρόνους εξυπηρέτησης με κατανομή $F_{B^{(i)}}(x)$. Συμβολίζουμε έναν γενικό χρόνο εξυπηρέτησης για την κλάση i με $B^{(i)}$ και θέτουμε $b^{(i)} = E[B^{(i)}]$, $b_2^{(i)} = E[(B^{(i)})^2]$, $\sigma_B^{2(i)} = \text{Var}[B^{(i)}]$ και $\rho^{(i)} = \lambda^{(i)}b^{(i)}$. Οι πελάτες κάθε κλάσης i έχουν απλή προτεραιότητα (non-preemptive priority) έναντι των πελατών των κλάσεων j με $j > i$, με την έννοια ότι προηγούνται στο να επιλεγούν για την έναρξη της εξυπηρέτησής τους από τους πελάτες των κλάσεων αυτών, αλλά αφίξεις πελατών υψηλότερης προτεραιότητας δεν συνεπάγονται τη διακοπή πιθανής εξυπηρέτησης πελάτη χαμηλότερης προτεραιότητας που βρίσκεται σε εξέλιξη. Οι πελάτες που ανήκουν στην ίδια κλάση εξυπηρετούνται σύμφωνα με την πειθαρχία FCFS. Το σύστημα αυτό αναφέρεται ως η $M/G/1$ ουρά με απλές προτεραιότητες. Να βρεθούν οι μέσοι αριθμοί πελατών των διαφόρων κλάσεων του συστήματος και οι μέσοι χρόνοι παραμονής των πελατών των διαφόρων κλάσεων.

Άσκηση 13.3 Θεωρούμε την $M/G/1$ ουρά με ρυθμό αφίξεων λ , γενική κατανομή χρόνων εξυπηρέτησης $F_B(x)$ με μέση τιμή b και την πειθαρχία LCFS-PR (Last-Come-First-Served - Preemptive-Resume). Σύμφωνα με την πειθαρχία αυτή, κάθε χρονική στιγμή λειτουργίας του συστήματος εξυπηρετείται ο πελάτης που έχει αφιχθεί πιο πρόσφατα. Επομένως, με το που φθάνει ένας πελάτης αρχίζει αμέσως να εξυπηρετείται διακόπτοντας τον πελάτη που βρίσκεται σε διαδικασία εξυπηρέτησης (αν υπάρχει κάποιος). Ο πελάτης του οποίου διεκόπη η εξυπηρέτηση θα συνεχίσει την εξυπηρέτησή του από το σημείο διακοπής και μετά (δηλαδή όχι από την αρχή), αφού έχει τελειώσει την εξυπηρέτησή του ο πελάτης που τον διέκοψε και όλοι όσοι ήρθαν μεταγενέστερα από αυτόν. Βρείτε τον οριακό μέσο αριθμό πελατών στο σύστημα, $E[Q]$, και τον μέσο χρόνο παραμονής ενός πελάτη σε αυτό, $E[S]$.

Άσκηση 13.4 Θεωρούμε την $M/G/1$ ουρά με ρυθμό αφίξεων λ , γενική κατανομή χρόνων εξυπηρέτησης $F_B(x)$ και την πειθαρχία LAS (Least-Attained-Service) του λιγότερου παρασχεθέντος χρόνου εξυπηρέτησης. Σύμφωνα με την πειθαρχία αυτή, ο υπηρέτης κάθε χρονική στιγμή δίνει προτεραιότητα στον πελάτη που έχει λάβει τη μικρότερη ποσότητα εξυπηρέτησης μέχρι εκείνη τη χρονική στιγμή. Βρείτε τον οριακό μέσο αριθμό πελατών στο σύστημα, $E[Q]$, και τον μέσο χρόνο παραμονής ενός πελάτη σε αυτό, $E[S]$.

Άσκηση 13.5 Θεωρούμε ένα σύστημα εξυπηρέτησης το οποίο διαθέτει έναν υπηρέτη και μηδενικό χώρο αναμονής. Οι πελάτες φθάνουν σύμφωνα με μια διαδικασία Poisson με ρυθμό λ , ενώ οι χρόνοι εξυπηρέτησης των πελατών έχουν μια γενική κατανομή $F_B(x)$. Ένας πελάτης που κατά την άφιξή του βρίσκει τον υπηρέτη ελεύθερο αρχίζει άμεσα να εξυπηρετείται. Αν, όμως, βρει τον υπηρέτη κατειλημένο, τότε αναχωρεί από το σύστημα και προσπαθεί αργότερα μέχρι τελικά να εξυπηρετηθεί. Οι χρόνοι μεταξύ διαδοχικών επαναπροσπαθειών ενός πελάτη έχουν την κατανομή $\text{Exp}(\gamma)$. Το σύστημα αυτό αναφέρεται ως η $M/G/1/1$ ουρά με εκθετικές επαναπροσπάθειες. Βρείτε τον οριακό μέσο αριθμό πελατών στο σύστημα, $E[Q]$, και τον μέσο χρόνο παραμονής ενός πελάτη σε αυτό, $E[S]$.

13.6 Σχόλια

Στο βιβλίο I. Adan και J. Resing 2001 μελετώνται τα βασικά μοντέλα συστημάτων εξυπηρέτησης με τη μέθοδο της ανάλυσης μέσης τιμής. Για την εφαρμογή της μεθόδου σε περισσότερα μοντέλα, ο ενδιαφερόμενος αναγνώστης μπορεί να διαβάσει την εργασία των I. Adan και Wal 2011 και τη μεταπτυχιακή διπλωματική εργασία Τσουπαρόπουλος 2015.

Για πιο εξειδικευμένες χρήσεις τις μεθόδου της ανάλυσης μέσης τιμής μπορεί κανείς να ανατρέξει στις εργασίες Winands, I.J.B.F. Adan και Houtum 2006 και Artalejo και J.A.C. Resing 2010.

Βιβλιογραφία

- [1] I. Adan και J. Resing. *Queueing Theory*. Eindhoven: Notes available online, 2001.
- [2] I. Adan και J. van der Wal. “Mean value techniques”. Στο: *Boucherie, R. and van Dijk, N.M. (eds.) Queueing Networks: A Fundamental Approach*. Springer, 2011. Κεφ. 13, σσ. 561–586. ISBN: 978-1-4419-6471-7. DOI: https://doi.org/10.1007/978-1-4419-6472-4_13.
- [3] Ν. Τσουπαρόπουλος. *Ανάλυση Μέσης Τιμής σε Συστήματα Εξυπηρέτησης*. Αθήνα: Διπλωματική Εργασία, Μεταπτυχιακό Πρόγραμμα Ειδίκευσης στη Στατιστική και Επιχειρησιακή Έρευνα, Τμήμα Μαθηματικών Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών, 2015.
- [4] E.M.M. Winands, I.J.B.F. Adan και G.J.J.A.N. van Houtum. “Mean value analysis for polling systems”. Στο: *Queueing Systems: Theory and Applications* 54.1 (2006), σσ. 35–44.
- [5] J.R. Artalejo και J.A.C. Resing. “Mean value analysis of single server retrial queues”. Στο: *Asian-Pacific Journal of Operational Research* 27 (2010), σσ. 335–345.

ΚΕΦΑΛΑΙΟ 14

ΚΛΑΣΙΚΕΣ ΜΗ-ΜΑΡΚΟΒΙΑΝΕΣ ΟΥΡΕΣ

Στο κεφάλαιο αυτό εξετάζουμε συστήματα εξυπηρέτησης που διατηρούν την υπόθεση της εκθετικότητας μόνο για τους ενδιάμεσους χρόνους των αφίξεων ή μόνο για τους χρόνους εξυπηρέτησης, αλλά όχι και για τους δύο. Τα συστήματα αυτά παρέχουν τη δυνατότητα καλύτερης αναπαράστασης πραγματικών συστημάτων εξυπηρέτησης, αλλά έχουν το μειονέκτημα ότι ο αριθμός των πελατών σε αυτά δεν είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου και επομένως η μελέτη τους καθίσταται δυσκολότερη. Μια προσέγγιση για τη μελέτη τους είναι η μέθοδος των φάσεων που είδαμε στο κεφάλαιο 8. Μια απλούστερη, όμως, διαδικασία είναι η ανάλυσή τους σε συγκεκριμένες στιγμές κατά τις οποίες «ανακτούν» τη Μαρκοβιανή ιδιότητα. Οι στιγμές αυτές είναι ακριβώς οι στιγμές της λήξης των χρόνων εξυπηρέτησης για τα M/G/1 συστήματα και οι στιγμές έναρξης των ενδιάμεσων χρόνων αφίξεων για τα G/M/1 συστήματα. Μελετώντας ένα σύστημα σε τέτοιες στιγμές, βλέπουμε ότι ο αριθμός των πελατών αντιστοιχεί σε μια Μαρκοβιανή αλυσίδα διακριτού χρόνου η οποία μπορεί να μελετηθεί με βάση τα αποτελέσματα που έχουν παρουσιαστεί στο κεφάλαιο 12. Στις επόμενες παραγράφους μελετάμε κάποια σημαντικά κλασικά συστήματα που εμπίπτουν σε αυτή την κατηγορία.

14.1 Η M/G/1 ουρά

Στην παρούσα παράγραφο, θεωρούμε ένα M/G/1 σύστημα, όπου οι πελάτες έρχονται σύμφωνα με μια διαδικασία Poisson ρυθμού λ , οι χρόνοι εξυπηρέτησης έχουν κάποια γενική κατανομή $F_B(x)$, υπάρχει ένας υπηρέτης και άπειρη χωρητικότητα. Επιπλέον, θεωρούμε ότι η πειθαρχία της ουράς είναι η FCFS. Συμβολίζουμε με B έναν τυπικό χρόνο εξυπηρέτησης. Θέτουμε $b = E[B]$, $b_2 = E[B^2]$, $\sigma_B^2 = Var[B]$ και $\tilde{F}_B(s) = E[e^{-sB}]$. Θέτουμε $\rho = \lambda b$ τον ρυθμό συνωστισμού. Στις επόμενες παραγράφους θα δούμε πώς μπορούν να υπολογιστούν τα βασικότερα μέτρα απόδοσης αυτού του συστήματος που σχετίζονται με τον αριθμό των πελατών σε αυτό, με τον χρόνο παραμονής ενός πελάτη σε αυτό, καθώς και με τα χαρακτηριστικά του κύκλου απασχόλησής του.

14.1.1 Ο αριθμός των πελατών στο σύστημα

Έστω $\{Q(t)\}$ η διαδικασία του αριθμού των πελατών στο σύστημα σε συνεχή χρόνο και $\{Q_n^-\}$ και $\{Q_n^+\}$ οι εμφυτευμένες διαδικασίες σε στιγμές αφίξεων και αναχωρήσεων, αντίστοιχα. Θα συμβολίζουμε με Q , Q^- και Q^+ τις αντίστοιχες τυχαίες μεταβλητές σε κατάσταση ισορροπίας και με (p_n) , (a_n) και (d_n) τις συναρτήσεις πιθανότητάς τους.

Η διαδικασία $\{Q(t)\}$ δεν είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου, λόγω των γενικών χρόνων εξυπηρέτησης. Για να είχαμε μια Μαρκοβιανή περιγραφή της κατάστασης του συστήματος, θα έπρεπε να συμπεριλαμβάναμε στην κατάσταση του συστήματος σε μια χρονική στιγμή t όχι μόνο τον αριθμό των πελατών σε αυτό, $Q(t)$, αλλά και τον διανυθέντα χρόνο εξυπηρέτησης, έστω $A_B(t)$ για τον πελάτη που εξυπηρετείται. Τότε, η διαδικασία $\{(Q(t), A_B(t))\}$ είναι Μαρκοβιανή. Όμως, ο χώρος καταστάσεων είναι το $\mathbb{N}_0 \times [0, \infty)$, οπότε χάνεται ο διακριτός χαρακτήρας της. Το σύστημα μπορεί να μελετηθεί με βάση αυτή τη διαδικασία, αλλά η σχετική θεωρία (Μέθοδος Συμπληρωματικών Μεταβλητών - Supplementary Variables Method, δείτε, π.χ., Cox 1955) εκφεύγει των ορίων ενός εισαγωγικού συγγράμματος. Κοιτάζοντας, όμως, το σύστημα αμέσως μετά τις στιγμές αναχώρησης, βλέπουμε ότι η $\{A_B(t)\}$ μηδενίζεται και επομένως ανακτάται ο διακριτός χαρακτήρας του χώρου καταστάσεων. Με άλλα λόγια η διαδικασία $\{Q_n^+\}$ είναι Μαρκοβιανή αλυσίδα διακριτού χρόνου. Αυτό φαίνεται και από τη σχέση

$$Q_{n+1}^+ = \max(Q_n^+ - 1, 0) + A_{n+1}, \quad (14.1)$$

όπου A_{n+1} είναι ο αριθμός των αφίξεων που συμβαίνουν κατά τη διάρκεια εξυπηρέτησης του $(n + 1)$ -οστού πελάτη. Η σχέση αυτή προκύπτει ως εξής:

- Αν ο n -οστός πελάτης αφήσει πίσω του 0 πελάτες, τότε ο $(n + 1)$ -οστός πελάτης θα αφήσει πίσω του όσους πελάτες συσσωρεύθηκαν κατά την εξυπηρέτησή του.
- Αν ο n -οστός πελάτης αφήσει πίσω του μη-κενό σύστημα, τότε ο $(n + 1)$ -οστός πελάτης θα αφήσει πίσω του όσους πελάτες άφησε ο προηγούμενος μείον έναν (τον εαυτό του) συν όσους συσσωρεύθηκαν κατά την εξυπηρέτησή του.

Από τη σχέση (14.1), βλέπουμε ότι ο πίνακας πιθανοτήτων μετάβασης της Μαρκοβιανής αλυσίδας $\{Q_n^+\}$ είναι

$$\mathbf{P} = \begin{pmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \cdots \\ \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \cdots \\ 0 & \alpha_0 & \alpha_1 & \alpha_2 & \cdots \\ 0 & 0 & \alpha_0 & \alpha_1 & \cdots \\ 0 & 0 & 0 & \alpha_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (14.2)$$

όπου $\alpha_j, j \geq 0$, είναι η πιθανότητα άφιξης j πελατών κατά τη διάρκεια του χρόνου εξυπηρέτησης ενός πελάτη (δηλαδή η συνάρτηση πιθανότητας των τυχαίων μεταβλητών A_{n+1} που εμφανίζονται στη σχέση (14.1)).

Δεσμεύοντας στη διάρκεια του χρόνου εξυπηρέτησης και λαμβάνοντας υπόψη ότι οι αφίξεις συμβαίνουν κατά τη διάρκειά του σύμφωνα με διαδικασία Poisson ρυθμού λ συνάγουμε ότι

$$\alpha_j = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^j}{j!} dF_B(t), \quad j \geq 0. \quad (14.3)$$

Η κατανομή ισορροπίας $(d_n : n \geq 0)$ της $\{Q_n^+\}$ βρίσκεται λύνοντας το σύστημα των εξισώσεων ισορροπίας της. Λόγω της μορφής του πίνακα πιθανοτήτων μετάβασης (14.2), οι εξισώσεις έχουν τη μορφή

$$d_n = d_0 \alpha_n + \sum_{i=1}^{n+1} d_i \alpha_{n+1-i}, \quad n \geq 0. \quad (14.4)$$

Για την επίλυση αυτών των εξισώσεων και τον προσδιορισμό της κατανομής ισορροπίας, θα ακολουθήσουμε τη μέθοδο των πιθανογεννητριών. Για τον σκοπό αυτό εισάγουμε τις πιθανογεννήτριες

$$P_D(z) = \sum_{n=0}^{\infty} d_n z^n, \quad |z| \leq 1,$$

και

$$\begin{aligned} \sum_{n=0}^{\infty} \alpha_n z^n &= \sum_{n=0}^{\infty} \int_0^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} dF_B(t) z^n \\ &= \int_0^{\infty} e^{-\lambda(1-z)t} dF_B(t) = \tilde{F}_B(\lambda(1-z)), \quad |z| \leq 1. \end{aligned} \quad (14.5)$$

Για τον μετασχηματισμό των εξισώσεων ισορροπίας σε μια εξίσωση για την πιθανογεννήτρια $P_D(z)$, πολλαπλασιάσουμε την εξίσωση (14.4) με z^n και αθροίζουμε για όλα τα $n \geq 0$. Έτσι έχουμε:

$$\begin{aligned} \sum_{n=0}^{\infty} d_n z^n &= d_0 \sum_{n=0}^{\infty} \alpha_n z^n + \sum_{n=0}^{\infty} \sum_{i=1}^{n+1} d_i \alpha_{n+1-i} z^n \\ \Leftrightarrow P_D(z) &= d_0 \tilde{F}_B(\lambda(1-z)) + \frac{1}{z} \sum_{i=1}^{\infty} d_i z^i \sum_{n=i-1}^{\infty} \alpha_{n+1-i} z^{n+1-i} \\ \Leftrightarrow P_D(z) &= d_0 \tilde{F}_B(\lambda(1-z)) + \frac{1}{z} (P_D(z) - d_0) \tilde{F}_B(\lambda(1-z)). \end{aligned}$$

Λύνοντας ως προς $P_D(z)$ παίρνουμε

$$P_D(z) = \frac{d_0 \tilde{F}_B(\lambda(1-z))(1-z)}{\tilde{F}_B(\lambda(1-z)) - z}.$$

Χρησιμοποιώντας την εξίσωση κανονικοποίησης $\sum_{n=0}^{\infty} d_n = P_D(1) = 1$, παίρνουμε την απροσδιόριστη μορφή $\frac{0}{0}$ στο δεξιό μέλος της εξίσωσης. Εφαρμόζοντας τον κανόνα L'Hospital έχουμε ότι

$$1 = P_D(1) = \frac{-d_0}{-\lambda \tilde{F}'_B(0) - 1}.$$

Όμως $b = E[B] = -\tilde{F}'_B(0)$, απ' όπου έχουμε ότι $d_0 = 1 - \rho$. Εναλλακτικά, θα μπορούσαμε να καταλήξουμε στο ίδιο αποτέλεσμα, χρησιμοποιώντας την ιδιότητα των μεμονωμένων μεταβάσεων (θεώρημα 2.2) που δίνει ότι $d_0 = p_0$ και το γενικό αποτέλεσμα (2.8) για την πιθανότητα κενού συστήματος στην G/G/1 ουρά που δίνει $p_0 = 1 - \rho$. Επομένως, έχουμε τελικά

$$P_D(z) = \frac{(1-\rho) \tilde{F}_B(\lambda - \lambda z)(1-z)}{\tilde{F}_B(\lambda - \lambda z) - z}. \quad (14.6)$$

Η πιθανογεννήτρια (14.6) αντιστοιχεί στον αριθμό των πελατών όχι μόνο σε στιγμές αναχωρήσεων, αλλά και σε στιγμές αφίξεων και σε συνεχή χρόνο, λόγω της ιδιότητας των μεμονωμένων μεταβάσεων και της ιδιότητας PASTA. Ο τύπος (14.6) αναφέρεται ως τύπος Pollaczek-Khinchin για την πιθανογεννήτρια του αριθμού των πελατών της M/G/1 ουράς.

Όταν ο μετασχηματισμός Laplace-Stieltjes των χρόνων εξυπηρέτησης είναι ρητή συνάρτηση του s , όπως στην περίπτωση που οι χρόνοι εξυπηρέτησης είναι μίξη κατανομών Erlang, η πιθανογεννήτρια εκφράζεται ως ρητή συνάρτηση του z και τότε είναι δυνατή η ακριβής αντιστροφή της ώστε να βρεθεί τύπος για την κατανομή ισορροπίας του μήκους ουράς. Πιο συγκεκριμένα, χρησιμοποιείται ο αλγόριθμος 3.8. Η απλούστερη

περίπτωση προκύπτει όταν οι χρόνοι εξυπηρέτησης έχουν την εκθετική κατανομή. Πράγματι, έστω ότι η $F_B(x)$ είναι η κατανομή $\text{Exp}(\mu)$. Τότε, έχουμε $\tilde{F}_B(s) = \frac{\mu}{s+\mu}$, οπότε η (14.6) γίνεται:

$$P_D(z) = \frac{(1-\rho) \frac{\mu}{\lambda-\lambda z+\mu} (1-z)}{\frac{\mu}{\lambda-\lambda z+\mu} - z} = \frac{(1-\rho)\mu(1-z)}{\mu - (\lambda - \lambda z + \mu)z} = \frac{1-\rho}{1-\rho z}.$$

Χρησιμοποιώντας τη γεωμετρική σειρά (3.5) παίρνουμε την κατανομή ισορροπίας $d_n = (1-\rho)\rho^n$, $n \geq 0$ που συμπίπτει με την (6.10).

14.1.2 Ο χρόνος παραμονής πελάτη

Έστω, τώρα, S μια τυχαία μεταβλητή που αντιστοιχεί στον χρόνο παραμονής ενός πελάτη στο σύστημα, όταν αυτό βρίσκεται σε κατάσταση ισορροπίας. Στα Μαρκοβιανά μοντέλα ουρών είδαμε ότι η κατανομή του χρόνου αυτού υπολογίζεται δεσμεύοντας στον αριθμό των πελατών που βρίσκει ένας αφικνούμενος πελάτης στο σύστημα (βλέπε π.χ. τη σχέση (6.12) στην παράγραφο 6.4 και τη σχέση (9.17) στην παράγραφο 9.2). Η προσέγγιση αυτή μοιάζει φυσική αφού ο χρόνος παραμονής ενός επιλεγμένου πελάτη είναι το άθροισμα των χρόνων εξυπηρέτησης των πελατών που βρίσκει κατά την άφιξή του συμπεριλαμβανομένου και του εαυτού του. Όμως, ο όρος του αθροίσματος που αντιστοιχεί στον πελάτη που ήδη εξυπηρετείται κατά την άφιξη του επιλεγμένου πελάτη δεν είναι ένας ολόκληρος χρόνος εξυπηρέτησης, αλλά ένας υπολειπόμενος χρόνος. Όταν οι χρόνοι εξυπηρέτησης ακολουθούν την εκθετική κατανομή (όπως συμβαίνει στα Μαρκοβιανά μοντέλα), αυτό δεν συνιστά πρόβλημα, αφού ο υπολειπόμενος χρόνος εξυπηρέτησης έχει την ίδια κατανομή με έναν ολόκληρο χρόνο εξυπηρέτησης. Αυτός είναι και ο λόγος που η ιδέα αυτή λειτουργεί καλά στα Μαρκοβιανά μοντέλα. Όμως, στο πλαίσιο των γενικών χρόνων εξυπηρέτησης, η κατανομή του υπολειπόμενου χρόνου εξυπηρέτησης του πελάτη που εξυπηρετείται δεδομένου ότι ο επιλεγμένος αφικνούμενος πελάτης βρίσκει έναν συγκεκριμένο αριθμό πελατών είναι δύσκολο να υπολογιστεί. Για το λόγο αυτό χρησιμοποιούνται κάποιες άλλες μεθοδολογικές προσεγγίσεις που αποδεικνύονται πιο αποτελεσματικές.

14.1.2.1 Συνδέοντας τον χρόνο παραμονής πελάτη με τον αριθμό αφίξεων κατά τη διάρκειά του

Έστω ένας επιλεγμένος πελάτης. Τότε ο αριθμός των πελατών που αφήνει πίσω στο σύστημα κατά την αναχώρησή του είναι ακριβώς ο αριθμός των πελατών που αφίχθησαν κατά τη διάρκεια παραμονής του. Αλλά ο τελευταίος είναι ίσος με τον αριθμό των γεγονότων μιας διαδικασίας Poisson με ρυθμό λ σε ένα διάστημα μήκους S και επομένως έχει πιθανογεννήτρια ίση με

$$\sum_{n=0}^{\infty} \int_0^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} dF_S(t) z^n = \int_0^{\infty} e^{-\lambda(1-z)t} dF_S(t) = \tilde{F}_S(\lambda(1-z)), \quad (14.7)$$

με τον ακριβώς ίδιο υπολογισμό που οδήγησε στην (14.5). Επομένως, από την ισότητα των πιθανογεννητριών του αριθμού των πελατών που αφήνει πίσω στο σύστημα κατά την αναχώρησή του ο επιλεγμένος πελάτης με τον αριθμό των πελατών που αφίχθησαν κατά τη διάρκεια παραμονής του, έχουμε $P_D(z) = \tilde{F}_S(\lambda(1-z))$. Επομένως, κάνοντας την αλλαγή μεταβλητής $s = \lambda(1-z)$ και χρησιμοποιώντας την (14.6) παίρνουμε

$$\tilde{F}_S(s) = \frac{(1-\rho)\tilde{F}_B(s)s}{\lambda\tilde{F}_B(s) + s - \lambda}. \quad (14.8)$$

Ο τύπος αυτός αναφέρεται ως τύπος Pollaczek-Khinchin για τον μετασχηματισμό Laplace-Stieltjes του χρόνου παραμονής της $M/G/1$ ουράς. Λαμβάνοντας υπόψη ότι ο χρόνος παραμονής, S , ενός πελάτη γράφεται ως το άθροισμα του χρόνου αναμονής, W και του χρόνου εξυπηρέτησής του, B και οι χρόνοι αυτοί είναι ανεξάρτητοι, έχουμε ότι $\tilde{F}_S(s) = \tilde{F}_W(s)\tilde{F}_B(s)$. Επομένως, χρησιμοποιώντας την (14.8) παίρνουμε τον μετασχηματισμό

Laplace-Stieltjes του χρόνου αναμονής της M/G/1 ουράς που είναι

$$\tilde{F}_W(s) = \frac{(1-\rho)s}{\lambda\tilde{F}_B(s) + s - \lambda}. \quad (14.9)$$

Ο τύπος αυτός αναφέρεται ως τύπος Pollaczek-Khinchin για τον μετασχηματισμό Laplace-Stieltjes του χρόνου αναμονής της M/G/1 ουράς.

Στην περίπτωση που οι χρόνοι εξυπηρέτησης είναι $\text{Exp}(\mu)$, έχουμε ότι ο τύπος (14.8) γίνεται

$$\tilde{F}_S(s) = \frac{(1-\rho)\frac{\mu}{s+\mu}s}{\lambda\frac{\mu}{s+\mu} + s - \lambda} = \frac{(1-\rho)\mu}{s + (1-\rho)\mu}, \quad (14.10)$$

που ταυτίζεται με τον τύπο (6.12) που έχουμε δει στην παράγραφο 6.4, κατά τη μελέτη της M/M/1 ουράς, και ο οποίος δείχνει ότι ο χρόνος παραμονής είναι $\text{Exp}((1-\rho)\mu)$.

14.1.2.2 Συνδέοντας τον χρόνο παραμονής πελάτη με τον χρόνο παραμονής του προηγούμενου

Ένας άλλος τρόπος για να προσδιορίσουμε τον μετασχηματισμό Laplace-Stieltjes του χρόνου παραμονής ενός πελάτη στην M/G/1 ουρά βασίζεται στη σύνδεση των χρόνων παραμονής δυο διαδοχικών πελατών.

Ας θεωρήσουμε, λοιπόν, έναν επιλεγμένο πελάτη, έστω τον n -οστό πελάτη που φθάνει στο σύστημα και τον προηγούμενο αφιχθέντα πελάτη (δηλαδή τον $(n-1)$ -οστό). Συμβολίζουμε με S_n , W_n και B_n τους χρόνους παραμονής, αναμονής και εξυπηρέτησης του n -οστού πελάτη και με S_{n-1} , W_{n-1} και B_{n-1} τους αντίστοιχους χρόνους του προηγούμενου. Έστω επίσης A_{n-1} ο ενδιάμεσος χρόνος αφίξεων μεταξύ $(n-1)$ -οστού και n -οστού πελάτη. Έχουμε τότε την προφανή σχέση

$$W_n = \max(S_{n-1} - A_{n-1}, 0). \quad (14.11)$$

Πράγματι ξεκινώντας να μετράμε από τη στιγμή της άφιξης του $(n-1)$ -οστού πελάτη, ο $(n-1)$ -οστός πελάτης θα φύγει από το σύστημα σε χρόνο S_{n-1} , ενώ ο n -οστός πελάτης θα έρθει στο σύστημα σε χρόνο A_{n-1} . Αν $A_{n-1} > S_{n-1}$, ο n -οστός πελάτης θα φθάσει μετά την αναχώρηση του προηγούμενου, οπότε θα βρει κενό σύστημα και θα αρχίσει αμέσως να εξυπηρετείται, οπότε $W_n = 0$. Διαφορετικά, αν $A_{n-1} \leq S_{n-1}$, θα περιμένει χρόνο $W_n = S_{n-1} - A_{n-1}$ πριν αρχίσει την εξυπηρέτησή του (δηλαδή τον χρόνο από την άφιξή του μέχρι την αναχώρηση του προηγούμενου πελάτη). Χρησιμοποιώντας τη σχέση $S_{n-1} = W_{n-1} + B_{n-1}$, η σχέση (14.11) γράφεται ως αναδρομική σχέση για τους χρόνους αναμονής πελατών:

$$W_n = \max(W_{n-1} + B_{n-1} - A_{n-1}, 0), \quad (14.12)$$

η οποία αναφέρεται ως εξίσωση του Lindley. Αφού ενδιαφερόμαστε για τον προσδιορισμό της κατανομής του χρόνου παραμονής, μπορούμε να μεταφράσουμε τη σχέση αυτή σε αναδρομική σχέση για τους χρόνους παραμονής διαδοχικά αφικνούμενων πελατών. Πράγματι, έχουμε

$$S_n = \max(S_{n-1} - A_{n-1}, 0) + B_n, \quad (14.13)$$

όπου οι τυχαίες μεταβλητές S_{n-1} , A_{n-1} και B_n είναι ανεξάρτητες. Επομένως, για την οριακή τυχαία μεταβλητή S που αντιστοιχεί στον χρόνο παραμονής πελάτη όταν το σύστημα βρίσκεται σε κατάσταση ισορροπίας, έχουμε ότι οι τυχαίες μεταβλητές S και $\max(S - A, 0) + B$ είναι ισόνομες. Επομένως οι αντίστοιχοι μετασχηματισμοί Laplace-Stieltjes είναι ίσοι κι έχουμε

$$\tilde{F}_S(s) = E[e^{-sS}] = E[e^{-s(\max(S-A,0)+B)}] = E[e^{-s\max(S-A,0)}]\tilde{F}_B(s). \quad (14.14)$$

Υπολογίζουμε, τώρα, τον μετασχηματισμό $E[e^{-s \max(S-A,0)}]$. Έχουμε:

$$\begin{aligned}
 E[e^{-s \max(S-A,0)}] &= \int_0^\infty \int_0^\infty e^{-s \max(x-y,0)} dF_A(y) dF_S(x) \\
 &= \int_0^\infty \int_0^x e^{-s(x-y)} \lambda e^{-\lambda y} dy dF_S(x) + \int_0^\infty \int_x^\infty \lambda e^{-\lambda y} dy dF_S(x) \\
 &= \int_0^\infty e^{-sx} \lambda \frac{1 - e^{-(s-\lambda)x}}{\lambda - s} dF_S(x) + \int_0^\infty e^{-\lambda x} dF_S(x) \\
 &= \frac{\lambda}{\lambda - s} (\tilde{F}_S(s) - \tilde{F}_S(\lambda)) + \tilde{F}_S(\lambda) = \frac{s\tilde{F}_S(\lambda) - \lambda\tilde{F}_S(s)}{s - \lambda}. \tag{14.15}
 \end{aligned}$$

Αντικαθιστώντας την (14.15) στην (14.14) και λύνοντας ως προς $\tilde{F}_S(s)$ παίρνουμε:

$$\tilde{F}_S(s) = \frac{\tilde{F}_S(\lambda)\tilde{F}_B(s)s}{\lambda\tilde{F}_B(s) + s - \lambda}. \tag{14.16}$$

Όμως

$$\tilde{F}_S(\lambda) = \int_0^\infty e^{-\lambda x} dF_S(x) = \int_0^\infty \Pr[A > x] dF_S(x) = \Pr[A > S] = 1 - \rho, \tag{14.17}$$

αφού η $\Pr[A > S]$ ισούται με την πιθανότητα ένας αφικνούμενος πελάτης να βρει κενό σύστημα (δείτε και την αιτιολόγηση στην αρχή της παρούσας υποπαραγράφου σχετικά με το ενδεχόμενο ο n -οστός πελάτης να βρει κενό σύστημα που ταυτίζεται με το ενδεχόμενο $\{A_{n-1} > S_{n-1}\}$). Επομένως, οι (14.16) και (14.17) συνεπάγονται τον τύπο (14.8) των Pollaczek-Khinchin για τον μετασχηματισμό Laplace-Stieltjes του χρόνου παραμονής της M/G/1 ουράς.

Παρότι ο δεύτερος τρόπος απόδειξης με βάση την αναδρομική σχέση του Lindley μοιάζει πιο περίπλοκος από τον πρώτο τρόπο, θα πρέπει να σημειώσουμε ότι ο πρώτος τρόπος χρησιμοποιεί την πιθανογεννήτρια του αριθμού των πελατών σε στιγμές αφίξεων και κατά συνέπεια η προσέγγιση του Lindley είναι πιο άμεση.

14.1.3 Η διάρκεια και ο αριθμός των πελατών ενός κύκλου απασχόλησης

Στην παράγραφο αυτή θα προσδιορίσουμε την κατανομή της διάρκειας και του αριθμού των πελατών που εξυπηρετούνται σε έναν κύκλο απασχόλησης. Όπως έχουμε δει στην παράγραφο 1.2, ο κύκλος απασχόλησης ή λειτουργίας ενός συστήματος ορίζεται ως το διάστημα από την άφιξη ενός πελάτη που βρίσκει το σύστημα κενό, μέχρι την επόμενη άφιξη πελάτη που θα βρει το σύστημα κενό. Κάθε τέτοιος κύκλος αποτελείται από δύο μέρη: την περίοδο συνεχούς λειτουργίας που το σύστημα είναι συνεχώς απασχολημένο και την περίοδο αργίας που το σύστημα παραμένει κενό. Συμβολίζοντας με Z , Y και I τις διάρκειες ενός κύκλου απασχόλησης και των αντίστοιχων περιόδων συνεχούς λειτουργίας και αργίας έχουμε ότι $Z = Y + I$. Πρέπει να σημειωθεί εδώ ότι οι τυχαίες μεταβλητές Y και I δεν είναι γενικά ανεξάρτητες σε ένα G/G/k σύστημα. Αυτό συμβαίνει διότι η περίοδος αργίας είναι ο υπολειπόμενος χρόνος μέχρι την επόμενη άφιξη και επομένως γενικά εξαρτάται από το πώς έληξε η προηγούμενη περίοδος συνεχούς λειτουργίας, δηλαδή εξαρτάται από τη διάρκειά της.

Όπως είδαμε στο κεφάλαιο 6, είναι εύκολο να υπολογίσουμε τις μέσες τιμές $E[Z]$, $E[Y]$ και $E[I]$ για απλές Μαρκοβιανές ουρές. Όμως, ο υπολογισμός των αντίστοιχων κατανομών είναι γενικά δύσκολος. Στην παρούσα παράγραφο θα δούμε πώς μπορούμε να υπολογίσουμε τις κατανομές αυτές για την M/G/1 ουρά.

Καταρχήν, μπορούμε να παρατηρήσουμε ότι στην περίπτωση της M/G/1 ουράς και γενικότερα στην περίπτωση ουρών με Poisson διαδικασίες αφίξεων, οι τυχαίες μεταβλητές Y και I είναι ανεξάρτητες. Αυτό οφείλεται στην αμνήμονη ιδιότητα της εκθετικής κατανομής. Πράγματι, η περίοδος αργίας που είναι ο υπολειπόμενος χρόνος για την επόμενη άφιξη είναι ανεξάρτητος από την παρελθούσα ιστορία των αφίξεων και επομένως και από την περίοδο συνεχούς λειτουργίας που μόλις προηγήθηκε. Επιπλέον, η περίοδος αργίας θα έχει την εκθετική κατανομή. Άρα στην περίπτωση αυτή αρκεί να επικεντρωθούμε στην εύρεση της κατανομής της διάρκειας της περιόδου συνεχούς λειτουργίας.

Θεωρούμε, λοιπόν, μια M/G/1 ουρά με τις παραμέτρους που θέσαμε στην αρχή της παραγράφου. Είναι φανερό ότι η διάρκεια μιας περιόδου συνεχούς λειτουργίας δεν εξαρτάται από την πειθαρχία ουράς, καθώς αυτή επηρεάζει το ποιος είναι ο πελάτης που κάθε φορά εξυπηρετείται, αλλά όχι αν το σύστημα είναι κενό ή όχι. Υποθέτουμε ότι η πειθαρχία ουράς είναι η LCFS. Θεωρούμε έναν πελάτη που βρίσκει το σύστημα κενό, έστω τον πελάτη C , και ο οποίος εγκαινιάζει μια περίοδο συνεχούς λειτουργίας διάρκειας Y . Έστω K το πλήθος των πελατών που φθάνουν κατά τη διάρκεια εξυπηρέτησης του πελάτη C , και έστω C_1, C_2, \dots, C_K οι πελάτες αυτοί. Όλοι αυτοί οι πελάτες θα εξυπηρετηθούν μέσα στην περίοδο Y με αντίστροφη σειρά και πιθανά όχι συνεχόμενα. Πράγματι, μόλις αναχωρήσει ο αρχικός πελάτης C αρχίζει μια υποπερίοδος Y_1 κατά την οποία εξυπηρετείται ο C_K καθώς και όλοι οι νέοι πελάτες που έφθασαν κατά την εξυπηρέτησή του και οι πελάτες που έφθασαν κατά την εξυπηρέτηση αυτών κ.ο.κ. Δηλαδή ο χρόνος Y_1 από τη στιγμή της έναρξης της εξυπηρέτησης του C_K μέχρι την έναρξη της εξυπηρέτησης του C_{K-1} είναι ο χρόνος για να αδειάσει η ουρά από το C_K και όλους όσους ήρθαν μετά την έναρξη της εξυπηρέτησής του ώστε να ασχοληθεί ο υπηρέτης με τον C_{K-1} . Επομένως η Y_1 έχει την ίδια κατανομή με μια περίοδο συνεχούς λειτουργίας του αρχικού συστήματος. Το ίδιο ισχύει και για τους άλλους πελάτες $C_{K-1}, C_{K-2}, \dots, C_1$, δηλαδή κάθε ένας αρχίζει μια υποπερίοδο του χρόνου συνεχούς λειτουργίας Y κατά την οποία εξυπηρετείται αυτός και όσοι πελάτες έφθασαν μετά από αυτόν κατά τη διάρκεια του χρόνου εξυπηρέτησής του κ.ο.κ. μέχρι να έρθει η σειρά του αμέσως προηγούμενα αφιχθέντα πελάτη από τους $C_{K-1}, C_{K-2}, \dots, C_1$. Κάθε τέτοια υποπερίοδος έχει την ίδια κατανομή με μια περίοδο συνεχούς λειτουργίας και επιπλέον όλες αυτές οι υποπερίοδοι είναι ανεξάρτητες τυχαίες μεταβλητές και ανεξάρτητες από τη διάρκεια του χρόνου εξυπηρέτησης του πελάτη C . Έχουμε τη σχέση

$$Y = B + \sum_{i=1}^K Y_i, \quad (14.18)$$

όπου B ο χρόνος εξυπηρέτησης του C και Y_1, Y_2, \dots, Y_K οι υποπερίοδοι που αντιστοιχούν στους C_K, C_{K-1}, \dots, C_1 .

Έστω $\tilde{F}_Y(s)$ ο μετασχηματισμός Laplace-Stieltjes της Y . Έχουμε

$$\tilde{F}_Y(s) = E[e^{-sY}] = \int_0^\infty E[e^{-s(B+\sum_{i=1}^K Y_i)} | B = x] dF_B(x). \quad (14.19)$$

Όμως,

$$\begin{aligned} E[e^{-s(B+\sum_{i=1}^K Y_i)} | B = x] &= e^{-sx} \sum_{k=0}^{\infty} \Pr[K = k | B = x] E[e^{-s \sum_{i=1}^k Y_i}] \\ &= e^{-sx} \sum_{k=0}^{\infty} e^{-\lambda x} \frac{(\lambda x)^k}{k!} (\tilde{F}_Y(s))^k \\ &= e^{-x(s+\lambda-\lambda\tilde{F}_Y(s))}. \end{aligned} \quad (14.20)$$

Αντικαθιστώντας την (14.20) στην (14.19) παίρνουμε

$$\begin{aligned} \tilde{F}_Y(s) &= E[e^{-sY}] = \int_0^\infty e^{-x(s+\lambda-\lambda\tilde{F}_Y(s))} dF_B(x) \\ &= \tilde{F}_B(s + \lambda - \lambda\tilde{F}_Y(s)). \end{aligned} \quad (14.21)$$

Η σχέση αυτή είναι μια εξίσωση για την $\tilde{F}_Y(s)$ η επίλυση της οποίας μας δίνει τον μετασχηματισμό Laplace-Stieltjes της διάρκειας του χρόνου συνεχούς λειτουργίας του συστήματος.

Με παρόμοια λογική μπορούμε να συναγάγουμε την πιθανογεννήτρια του αριθμού των πελατών, R , που εξυπηρετούνται μέσα σε έναν κύκλο συνεχούς λειτουργίας. Η αντίστοιχη της (14.18) είναι τώρα η

$$R = 1 + \sum_{i=1}^K R_i, \quad (14.22)$$

όπου K είναι και πάλι ο αριθμός των πελατών που φθάνουν κατά την εξυπηρέτηση του πρώτου πελάτη της περιόδου, C , και R_1, R_2, \dots, R_K οι αριθμοί των πελατών που εξυπηρετούνται κατά τη διάρκεια των υποπεριοδών Y_1, Y_2, \dots, Y_K που αντιστοιχούν στους C_K, C_{K-1}, \dots, C_1 .

Έστω $P_R(z), P_K(z)$ οι πιθανογεννήτριες των R και K , αντίστοιχα. Τότε, από τη σχέση (14.22) συνάγουμε ότι

$$P_R(z) = zP_K(P_R(z)). \quad (14.23)$$

Όμως, αν B ο χρόνος εξυπηρέτησης του πρώτου πελάτη της περιόδου συνεχούς λειτουργίας, C , έχουμε

$$P_K(z) = E[z^K] = E[E[z^K|B]] = E[e^{-\lambda B(1-z)}] = \tilde{F}_B(\lambda - \lambda z), \quad (14.24)$$

λόγω του ότι η κατανομή της K δεδομένου του ότι $B = x$ είναι Poisson(λx) με πιθανογεννήτρια $e^{-\lambda x(1-z)}$. Αντικαθιστώντας την (14.24) στην (14.23) έχουμε την εξίσωση

$$P_R(z) = z\tilde{F}_B(\lambda - \lambda P_R(z)) \quad (14.25)$$

για την $P_R(z)$.

Ειδικά για την M/M/1 ουρά με Exp(μ) χρόνους εξυπηρέτησης, έχουμε ότι $\tilde{F}_B(s) = \frac{\mu}{s+\mu}$, οπότε οι εξισώσεις (14.21) και (14.25) γίνονται αντίστοιχα:

$$\tilde{F}_Y(s) = \frac{\mu}{s + \lambda - \lambda\tilde{F}_Y(s) + \mu}, \quad (14.26)$$

$$P_R(z) = \frac{\mu z}{\lambda - \lambda P_R(z) + \mu}. \quad (14.27)$$

Και οι δυο αυτές εξισώσεις ανάγονται άμεσα σε εξισώσεις δευτέρου βαθμού και επομένως οι λύσεις τους μπορούν να υπολογιστούν ακριβώς. Έχουμε, τελικά

$$\tilde{F}_Y(s) = \frac{\lambda + \mu + s - \sqrt{(\lambda + \mu + s)^2 - 4\lambda\mu}}{2\lambda}, \quad (14.28)$$

$$\begin{aligned} P_R(z) &= \frac{\lambda + \mu - \sqrt{(\lambda + \mu)^2 - 4\lambda\mu z}}{2\lambda} \\ &= \frac{1 + \rho}{2\rho} \left[1 - \sqrt{1 - \frac{4\rho z}{(1 + \rho)^2}} \right]. \end{aligned} \quad (14.29)$$

Από τις ρίζες των εξισώσεων επελέγησαν αυτές που αντιστοιχούν στο αρνητικό πρόσημο της τετραγωνικής ρίζας. Αυτό έγινε επειδή αυτές ικανοποιούν τις απαιτήσεις $|\tilde{F}_Y(s)| \leq 1$, για $\Re(s) \geq 0$, και $|P_R(z)| \leq 1$, για $|z| \leq 1$, που είναι αναγκαίες για μετασχηματισμούς Laplace-Stieltjes και πιθανογεννήτριες αντίστοιχα.

Η αντιστροφή των παραπάνω τύπων είναι δυνατή σε ακριβή μορφή, οπότε μπορούμε να συναγάγουμε τη συνάρτηση πυκνότητας πιθανότητας $f_Y(x)$ της διάρκειας μιας περιόδου συνεχούς λειτουργίας και τη συνάρτηση πιθανότητας $f_R(n)$ του αριθμού των πελατών που εξυπηρετούνται σε αυτήν. Πιο συγκεκριμένα, έχουμε

$$f_Y(x) = \frac{e^{-(\lambda+\mu)x} I_1(2x\sqrt{\lambda\mu})}{\sqrt{\rho}x}, \quad x \geq 0, \quad (14.30)$$

$$f_R(n) = \frac{1}{n} \binom{2n-2}{n-1} \rho^{n-1} (1+\rho)^{1-2n}, \quad n \geq 0, \quad (14.31)$$

όπου $I_1(x) = \sum_{m=0}^{\infty} \frac{(x/2)^{2m+1}}{(m+1)!m!}$ είναι η τροποποιημένη συνάρτηση Bessel πρώτου είδους, τάξης 1.

14.2 Η G/M/1 ουρά

Στην παράγραφο αυτή θα μελετήσουμε ένα σύστημα G/M/1, όπου οι πελάτες έρχονται σύμφωνα με μια αναγωγική διαδικασία με κατανομή ενδιάμεσων χρόνων μεταξύ των αφίξεων $F_A(x)$, οι χρόνοι εξυπηρέτησης είναι $\text{Exp}(\mu)$, υπάρχει ένας υπηρέτης και απεριόριστη χωρητικότητα. Η πειθαρχία ουράς είναι η FCFS. Συμβολίζουμε με A έναν τυπικό ενδιάμεσο χρόνο αφίξεων και θέτουμε $a = E[A]$, $a_2 = E[A^2]$, $\sigma_A^2 = \text{Var}[A]$ και $\tilde{F}_A(s) = E[e^{-sA}]$. Επίσης, συμβολίζουμε με $\rho = \frac{1}{\mu a}$ τον ρυθμό συνωστισμού. Στις επόμενες παραγράφους θα υπολογίσουμε τα βασικότερα μέτρα απόδοσης που αφορούν αυτό το σύστημα, και συγκεκριμένα τον αριθμό των πελατών και τον χρόνο παραμονής ενός πελάτη. Η γενική ιδέα είναι ότι το σύστημα είναι με κάποια έννοια δυϊκό του M/G/1 συστήματος και επομένως πολλές ιδέες από την ανάλυση εκείνου του συστήματος λειτουργούν και για το G/M/1 με κατάλληλη προσαρμογή.

14.2.1 Ο αριθμός των πελατών στο σύστημα

Έστω $\{Q(t)\}$ η διαδικασία του αριθμού των πελατών στο σύστημα σε συνεχή χρόνο και $\{Q_n^-\}$ και $\{Q_n^+\}$ οι εμφυτευμένες διαδικασίες σε στιγμές αφίξεων και αναχωρήσεων, αντίστοιχα. Θα συμβολίζουμε με Q , Q^- και Q^+ τις αντίστοιχες τυχαίες μεταβλητές σε κατάσταση ισορροπίας και με (p_n) , (a_n) και (d_n) τις συναρτήσεις πιθανότητάς τους.

Η διαδικασία $\{Q(t)\}$ δεν είναι Μαρκοβιανή αλυσίδα συνεχούς χρόνου, λόγω των γενικών ενδιάμεσων χρόνων αφίξεων. Κοιτάζοντας, όμως, το σύστημα μόλις πριν τις στιγμές των αφίξεων, βλέπουμε ότι ανακτάται ο Μαρκοβιανός χαρακτήρας του. Συγκεκριμένα, η διαδικασία $\{Q_n^-\}$ είναι Μαρκοβιανή αλυσίδα διακριτού χρόνου. Αυτό φαίνεται και από τη σχέση

$$Q_{n+1}^- = \max(Q_n^- + 1 - B_{n+1}, 0), \quad (14.32)$$

όπου B_{n+1} είναι ο αριθμός των εξυπηρέτησεων που συμβαίνουν κατά τη διάρκεια του $(n+1)$ -οστού ενδιάμεσου χρόνου αφίξεων, από την άφιξη του n -οστού μέχρι την άφιξη του $(n+1)$ -οστού πελάτη. Από τη σχέση (14.32), βλέπουμε ότι ο πίνακας πιθανοτήτων μετάβασης της Μαρκοβιανής αλυσίδας $\{Q_n^-\}$ είναι ο

$$\mathbf{P} = \begin{pmatrix} 1 - \beta_0 & \beta_0 & 0 & 0 & \cdots \\ 1 - \beta_0 - \beta_1 & \beta_1 & \beta_0 & 0 & \cdots \\ 1 - \beta_0 - \beta_1 - \beta_2 & \beta_2 & \beta_1 & \beta_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (14.33)$$

όπου $\beta_j, j \geq 0$, είναι η πιθανότητα αναχώρησης j πελατών κατά τη διάρκεια ενός ενδιάμεσου χρόνου αφίξεων υπό την προϋπόθεση ότι ο υπηρέτης είναι συνεχώς απασχολημένος κατά τη διάρκειά του. Με άλλα λόγια, η $(\beta_j : j \geq 0)$ είναι η συνάρτηση πιθανότητας των τυχαίων μεταβλητών B_{n+1} που εμφανίζονται στη σχέση (14.32)).

Δεσμεύοντας στη διάρκεια του ενδιάμεσου χρόνου αφίξεων και λαμβάνοντας υπόψη ότι οι αναχωρήσεις συμβαίνουν κατά τη διάρκειά του σύμφωνα με διαδικασία Poisson ρυθμού μ εφόσον ο υπηρέτης είναι συνεχώς απασχολημένος, συνάγουμε ότι

$$\beta_j = \int_0^\infty e^{-\mu t} \frac{(\mu t)^j}{j!} dF_A(t), \quad j \geq 0. \quad (14.34)$$

Επιπλέον για την πιθανογεννήτρια των β_j έχουμε

$$\begin{aligned} \sum_{j=0}^{\infty} \beta_j z^j &= \sum_{j=0}^{\infty} \int_0^\infty e^{-\mu t} \frac{(\mu t)^j}{j!} dF_A(t) z^j = \int_0^\infty e^{-(\mu - \mu z)t} dF_A(t) \\ &= \tilde{F}_A(\mu - \mu z). \end{aligned} \quad (14.35)$$

Η κατανομή ισορροπίας $(a_n : n \geq 0)$ της $\{Q_n^-\}$ βρίσκεται λύνοντας το σύστημα των εξισώσεων ισορροπίας της. Λόγω της μορφής του πίνακα πιθανοτήτων μετάβασης (14.33), οι εξισώσεις έχουν τη μορφή

$$a_0 = \sum_{i=0}^{\infty} a_i \left(1 - \sum_{j=0}^i \beta_j \right), \quad (14.36)$$

$$a_n = \sum_{i=0}^{\infty} a_{n-1+i} \beta_i, \quad n \geq 1. \quad (14.37)$$

Οι εξισώσεις για $n \geq 1$ παρουσιάζουν μια «ομογένεια» ως προς n υπό την έννοια ότι η εξίσωση για την a_{n+1} προκύπτει από την εξίσωση για την a_n αντικαθιστώντας το n με $n+1$ σε όλους τους όρους. Επομένως, έχουν λύσεις γεωμετρικής μορφής $a_n = c\sigma^n$, $n \geq 0$, για κάποια κατάλληλη παράμετρο σ . Για να βρούμε την παράμετρο σ αντικαθιστούμε τη λύση γεωμετρικής μορφής σε κάποια από τις εξισώσεις ισορροπίας για $n \geq 1$, π.χ., στην εξίσωση για την a_1 . Έχουμε

$$c\sigma = \sum_{i=0}^{\infty} c\sigma^i \beta_i,$$

που γράφεται ισοδύναμα ως

$$\sigma = \tilde{F}_A(\mu - \mu\sigma), \quad (14.38)$$

λαμβάνοντας υπόψη την (14.35). Το θέμα είναι κατά πόσον η εξίσωση (14.38) έχει λύση με $\sigma \in (0, 1)$. Τότε θα έχουμε ότι η εξίσωση κανονικοποίησης $\sum_{n=0}^{\infty} a_n = 1$ δίνει $\sum_{n=0}^{\infty} c\sigma^n = 1$, οπότε $c = 1 - \sigma$ και τελικά

$$a_n = (1 - \sigma)\sigma^n, \quad n \geq 0. \quad (14.39)$$

Θα δείξουμε ότι κάτω από τη συνθήκη ευστάθειας

$$\rho = \frac{1}{\mu a} < 1, \quad (14.40)$$

η εξίσωση (14.38) έχει πράγματι μοναδική λύση στο $(0, 1)$. Έχουμε ότι η συνάρτηση $\tilde{F}_A(\mu - \mu z) = \sum_{i=0}^{\infty} \beta_i z^i$ είναι μη-αρνητική, αύξουσα και κυρτή για $z \in (0, 1)$, αφού όλοι οι συντελεστές β_i της δυναμοσειράς είναι μη-αρνητικοί. Επομένως, η συνάρτηση $f(z) = \tilde{F}_A(\mu - \mu z) - z$ είναι κυρτή. Εξετάζουμε την παράγωγό της στο $z = 1$, που είναι η $f'(1) = -\tilde{F}'_A(0)\mu - 1 = \mu a - 1 = \frac{1}{\rho} - 1 > 0$, λόγω της συνθήκης ευστάθειας. Επομένως, για $\epsilon > 0$ αρκετά μικρό, έχουμε $f(1 - \epsilon) < f(1) = \tilde{F}_A(0) - 1 = 0$. Επίσης $f(0) = \tilde{F}_A(\mu) = \beta_0 > 0$. Αφού $f(0)f(1 - \epsilon) < 0$, από το θεώρημα Bolzano έχουμε ότι η $f(z)$ έχει ρίζα $\sigma \in (0, 1 - \epsilon) \subseteq (0, 1)$.

Πρέπει να σημειωθεί ότι η σ είναι η μοναδική ρίζα της $z = \tilde{F}_A(\mu - \mu z)$ στο $(0, 1)$. Πράγματι, αν υπήρχε και δεύτερη ρίζα, τότε θα είχαμε άτοπο από την κυρτότητα της $f(z)$. Εναλλακτικά μπορούμε να επιχειρηματολογήσουμε ότι αν υπήρχε και δεύτερη ρίζα $\sigma' \in (0, 1)$ τότε και η συνάρτηση πιθανότητας $((1 - \sigma')\sigma'^n : n \geq 0)$ θα ήταν κατανομή ισορροπίας της $\{Q_n^-\}$, που είναι άτοπο από τη μοναδικότητα της κατανομής ισορροπίας μιας αδιαχώριστης θετικά επαναληπτικής Μαρκοβιανής αλυσίδας διακριτού χρόνου.

Από την ιδιότητα των μεμονωμένων μεταβάσεων έχουμε ότι η κατανομή ισορροπίας του αριθμού των πελατών σε στιγμές αναχωρήσεων ταυτίζεται με την κατανομή ισορροπίας σε στιγμές αφίξεων και επομένως

$$d_n = (1 - \sigma)\sigma^n, \quad n \geq 0.$$

Για να βρούμε την κατανομή ισορροπίας (p_n) της $\{Q(t)\}$, βρίσκουμε μια σύνδεση της (p_n) με την (a_n) . Έχουμε ότι η ποσότητα $\frac{1}{a}$ εκφράζει τον μακροπρόθεσμο ρυθμό αφίξεων στο σύστημα και η a_{n-1} είναι η πιθανότητα μια άφιξη να βρίσκει $n-1$ πελάτες. Επομένως, το γινόμενο $\frac{1}{a}a_{n-1}$ είναι ο μακροπρόθεσμος ρυθμός μεταβάσεων της στοχαστικής διαδικασίας $\{Q(t)\}$ από την κατάσταση $n-1$ στην κατάσταση n . Αυτός πρέπει να ισούται με τον μακροπρόθεσμο ρυθμό μεταβάσεων από την n στην $n-1$, καθώς μεταξύ δυο διαδοχικών μεταβάσεων

$n - 1 \rightarrow n$ παρεμβάλεται μια μετάβαση $n \rightarrow n - 1$ (είναι το ίδιο επιχείρημα που χρησιμοποιήσαμε για την αιτιολόγηση της ιδιότητας των μεμονωμένων μεταβάσεων). Ο μακροπρόθεσμος ρυθμός μεταβάσεων $n \rightarrow n - 1$ ισούται με το μακροπρόθεσμο ποσοστό του χρόνου που το σύστημα βρίσκεται στην κατάσταση n , p_n , επί τον δεσμευμένο ρυθμό των αναχωρήσεων δεδομένου ότι βρίσκεται στην κατάσταση n . Λόγω της Poisson διαδικασίας αναχωρήσεων όσο το σύστημα δεν είναι κενό, αυτός ο δεσμευμένος ρυθμός είναι μ . Οπότε έχουμε τη σχέση

$$\frac{1}{a} a_{n-1} = p_n \mu, \quad n \geq 1,$$

και χρησιμοποιώντας την (14.39) έχουμε την κατανομή ισορροπίας του αριθμού των πελατών σε συνεχή χρόνο να δίνεται από τις σχέσεις

$$p_0 = 1 - \rho, \quad (14.41)$$

$$p_n = \rho(1 - \sigma)\sigma^{n-1}, \quad n \geq 1. \quad (14.42)$$

Η εξίσωση (14.38) στην περίπτωση που οι ενδιάμεσοι χρόνοι αφίξεων είναι $\text{Exp}(\lambda)$ γίνεται

$$\sigma = \frac{\lambda}{\mu - \mu\sigma + \lambda}$$

της οποίας η μοναδική λύση στο $(0, 1)$ είναι η $\sigma = \rho$. Οπότε στην περίπτωση της $M/M/1$ ουράς έχουμε ότι οι εξισώσεις (14.41) και (14.42) δίνουν $p_n = (1 - \rho)\rho^n$, $n \geq 0$, που είναι το αποτέλεσμα που είδαμε στη σχέση (6.10).

14.2.2 Ο χρόνος παραμονής πελάτη

Έστω, τώρα, S μια τυχαία μεταβλητή που αντιστοιχεί στον χρόνο παραμονής ενός πελάτη στο σύστημα, όταν αυτό βρίσκεται σε κατάσταση ισορροπίας. Λόγω της αμνήμονης ιδιότητας των χρόνων εξυπηρέτησης, μπορούμε να υπολογίσουμε την κατανομή του χρόνου αυτού με την ίδια τεχνική που εφαρμόσαμε στα Μαρκοβιανά μοντέλα ουρών, δηλαδή δεσμεύοντας στον αριθμό των πελατών που βρίσκει ένας αφικνούμενος πελάτης στο σύστημα. Έχουμε:

$$\begin{aligned} \tilde{F}_S(s) &= E[e^{-sS}] = \sum_{n=0}^{\infty} \Pr[Q^- = n] E[e^{-sS} | Q^- = n] \\ &= \sum_{n=0}^{\infty} (1 - \sigma)\sigma^n \left(\frac{\mu}{\mu + s} \right)^{n+1} \\ &= \frac{(1 - \sigma)\mu}{\mu + s} \left(1 - \frac{\sigma\mu}{\mu + s} \right)^{-1} \\ &= \frac{(1 - \sigma)\mu}{(1 - \sigma)\mu + s}. \end{aligned} \quad (14.43)$$

Ο τελευταίος είναι ο μετασχηματισμός Laplace-Stieltjes της $\text{Exp}(\mu(1 - \sigma))$, οπότε συνάγουμε ότι η S ακολουθεί την $\text{Exp}(\mu(1 - \sigma))$ κατανομή.

14.3 Η M/G/άπειρο ουρά

Θεωρούμε την $M/G/\infty$ ουρά, όπου οι πελάτες έρχονται σύμφωνα με μια διαδικασία Poisson ρυθμού λ , οι χρόνοι εξυπηρέτησης έχουν κάποια γενική κατανομή $F_B(x)$ και υπάρχουν άπειροι υπηρέτες, οπότε κάθε πελάτης αρχίζει και εξυπηρετείται με την άφιξή του στο σύστημα. Θέτουμε A_i τον χρόνο άφιξης του i -οστού πελάτη και με B_i τον χρόνο εξυπηρέτησής του.

Ενδιαφερόμαστε για την κατανομή του αριθμού των πελατών στο σύστημα τη στιγμή t , $Q(t)$, υποθέτοντας ότι τη χρονική στιγμή 0 το σύστημα είναι κενό ($Q(0) = 0$). Θέτουμε $A(t)$ τον αριθμό των αφίξεων στο $(0, t]$ και για κάθε $i \geq 1$ θέτουμε

$$I_i(t) = \begin{cases} 1 & \text{αν ο } i \text{ πελάτης βρίσκεται στο σύστημα τη στιγμή } t, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Έχουμε προφανώς τη σχέση

$$Q(t) = \sum_{i=1}^{A(t)} 1_{\{I_i(t)=1\}},$$

οπότε

$$\begin{aligned} p_n(t) &= \Pr[Q(t) = n] = \Pr\left[\sum_{i=1}^{A(t)} 1_{\{I_i(t)=1\}} = n\right] \\ &= \sum_{m=n}^{\infty} \Pr[A(t) = m] \Pr\left[\sum_{i=1}^{A(t)} 1_{\{I_i(t)=1\}} = n \mid A(t) = m\right]. \end{aligned}$$

Όμως, το ενδεχόμενο $\{I_i(t) = 1\}$ αντιστοιχεί στο να είναι παρών ο i -οστός πελάτης τη στιγμή t που είναι το ίδιο με το $\{A_i + B_i > t\}$, οπότε έχουμε

$$p_n(t) = \sum_{m=n}^{\infty} \Pr[A(t) = m] \Pr\left[\sum_{i=1}^m 1_{\{A_i+B_i>t\}} = n \mid A(t) = m\right].$$

Όμως από το θεώρημα 4.17 έχουμε ότι η δεσμευμένη κατανομή της (A_1, A_2, \dots, A_m) δοθέντος ότι $A(t) = m$ είναι η κατανομή των διατεταγμένων τυχαίων μεταβλητών $U_{1:m}, U_{2:m}, \dots, U_{m:m}$ από m ανεξάρτητες και ισόνομες Uniform($[0, t]$), U_1, U_2, \dots, U_m . Λαμβάνοντας υπόψη και ότι οι B_i είναι ανεξάρτητες του ενδεχομένου $\{A(t) = m\}$ έχουμε

$$p_n(t) = \sum_{m=n}^{\infty} \Pr[A(t) = m] \Pr\left[\sum_{i=1}^m 1_{\{U_{i:m}+B_i>t\}} = n\right].$$

Καθώς εμφανίζονται όλες οι διατεταγμένες τυχαίες μεταβλητές $U_{i:m}$, για $i = 1, 2, \dots, m$ και οι B_i είναι ανεξάρτητες των $U_{i:m}$ και ισόνομες, το άθροισμα $\sum_{i=1}^m 1_{\{U_{i:m}+B_i>t\}}$ μπορεί να γραφεί $\sum_{i=1}^m 1_{\{U_i+B_i>t\}}$ με κατάλληλη μετονομασία των B_i ώστε οι δείκτες τους να συμφωνούν με τις U_i . Οπότε, έχουμε

$$p_n(t) = \sum_{m=n}^{\infty} \Pr[A(t) = m] \Pr\left[\sum_{i=1}^m 1_{\{U_i+B_i>t\}} = n\right]. \quad (14.44)$$

Οι τυχαίες μεταβλητές $1_{\{U_i+B_i>t\}}$, $i = 1, 2, \dots, m$ είναι ανεξάρτητες και ισόνομες δοκιμές Bernoulli με πιθανότητα επιτυχίας

$$p(t) = \Pr[U_i + B_i > t] = \int_0^{\infty} \Pr[B_i > t - u] dF_U(u) \quad (14.45)$$

$$= \frac{1}{t} \int_0^t (1 - F_B(t - u)) du. \quad (14.46)$$

Επομένως, η τυχαία μεταβλητή $\sum_{i=1}^m 1_{\{U_i+B_i>t\}}$ είναι διωνυμική με παραμέτρους m και $p(t)$. Επομένως, έχουμε

$$\Pr\left[\sum_{i=1}^m 1_{\{U_i+B_i>t\}} = n\right] = \binom{m}{n} p(t)^n (1 - p(t))^{m-n}, \quad 0 \leq n \leq m, \quad (14.47)$$

και αντικαθιστώντας στην (14.44) έχουμε τελικά

$$\begin{aligned}
 p_n(t) &= \sum_{m=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^m}{m!} \binom{m}{n} p(t)^n (1-p(t))^{m-n} \\
 &= e^{-\lambda t} \frac{(\lambda t p(t))^n}{n!} \sum_{m=n}^{\infty} \frac{(\lambda t (1-p(t)))^{m-n}}{(m-n)!} \\
 &= e^{-\lambda t} \frac{(\lambda t p(t))^n}{n!} \sum_{i=0}^{\infty} \frac{(\lambda t (1-p(t)))^i}{i!} \\
 &= e^{-\lambda t p(t)} \frac{(\lambda t p(t))^n}{n!}. \tag{14.48}
 \end{aligned}$$

Επομένως, βλέπουμε ότι το πλήθος των πελατών στο σύστημα τη στιγμή t , $Q(t)$, έχει την κατανομή Poisson με παράμετρο

$$\lambda t p(t) = \lambda \int_0^t (1 - F_B(t-u)) du.$$

Παρατηρούμε ότι

$$\lim_{t \rightarrow \infty} \lambda \int_0^t (1 - F_B(t-u)) du = \lambda \int_0^{\infty} (1 - F_B(x)) dx = \lambda b = \rho,$$

όπου $b = E[B_i]$. Βλέπουμε, λοιπόν, ότι η κατανομή ισορροπίας του αριθμού των πελατών στο σύστημα είναι Poisson(ρ) και εξαρτάται μόνο από τη μέση τιμή των χρόνων εξυπηρέτησης και όχι από την κατανομή τους. Έχουμε, δηλαδή, και πάλι μια ιδιότητα μη-ευαισθησίας, όπως αυτές που είδαμε ότι ισχύουν για τα μοντέλα M/G/c/c και Engset που συζητήθηκαν στην παράγραφο 6.7.

14.4 Το σύστημα με Poisson αφίξεις και ανανεωτική διαδικασία εκκαθαρίσεων

Στην παράγραφο αυτή θα μελετήσουμε ένα σύστημα που γενικεύει το Μαρκοβιανό σύστημα εκκαθαρίσεων με μεμονωμένες αφίξεις που είδαμε στην παράγραφο 7.2.1. Υποθέτουμε ότι οι πελάτες φθάνουν σύμφωνα με μια διαδικασία Poisson, $\{A(t)\}$ με ρυθμό λ . Το σύστημα εκκαθαρίζεται στα σημεία μιας ανανεωτικής διαδικασίας $\{B(t)\}$, με ενδιάμεσους χρόνους B_1, B_2, \dots με κατανομή $F_B(t)$ και ανανεωτική συνάρτηση $m_B(t) = E[B(t)]$. Ένα τέτοιο σύστημα είναι κατάλληλο για τη μοντελοποίηση συστημάτων μεταφορών στα πλαίσια της Θεωρίας Ουρών. Ουσιαστικά, το σύστημα αντιστοιχεί σε μια στάση ενός μεταφορικού μέσου όπου οι επιβάτες καταφθάνουν σύμφωνα με τη διαδικασία $\{A(t)\}$ και τα μεταφορικά μέσα διέρχονται τις στιγμές των γεγονότων της $\{B(t)\}$ οπότε επιβιβάζονται σε αυτά όλοι οι πελάτες που έχουν συσσωρευθεί στη στάση. Ως συνήθως, συμβολίζουμε με $\{Q(t)\}$ τη διαδικασία του αριθμού των πελατών στο σύστημα. Ενδιαφερόμαστε για τον προσδιορισμό της κατανομής της $Q(t)$, για κάθε χρονική στιγμή t , δηλαδή για την $(p_n(t) : n \geq 0)$ με $p_n(t) = \Pr[Q(t) = n, n \geq 0]$, υποθέτοντας ότι το σύστημα είναι αρχικό κενό (δηλ. $Q(0) = 0$).

Θα προσδιορίσουμε τη μεταβατική κατανομή $(p_n(t))$ χρησιμοποιώντας επιχειρήματα από την ανανεωτική θεωρία.

Δεσμεύοντας στον χρόνο B_1 της πρώτης εκκαθάρισης έχουμε ότι:

$$p_n(t) = \int_0^{\infty} \Pr[Q(t) = n | Q(0) = 0, B_1 = x] dF_B(x). \tag{14.49}$$

Αν η πρώτη εκκαθάριση συμβαίνει σε χρόνο $x > t$ τότε η εξέλιξη της $\{Q(t)\}$ στο $[0, t]$ είναι η ίδια όπως της διαδικασίας αφίξεων $\{A(t)\}$. Αν, όμως, η πρώτη εκκαθάριση συμβαίνει σε χρόνο $x \leq t$ τότε ο αριθμός των πελατών γίνεται ακαριαία 0 τη στιγμή x και η διαδικασία $\{Q(t)\}$ επανεκκινείται (δηλαδή το σημείο της εκκαθάρισης x είναι αναγεννητικό σημείο για την $\{Q(t)\}$). Επομένως,

$$\Pr[Q(t) = n | Q(0) = 0, B_1 = x] = \begin{cases} e^{-\lambda t} \frac{(\lambda t)^n}{n!} & \text{αν } x > t, \\ p_n(t-x) & \text{αν } x \leq t. \end{cases} \tag{14.50}$$

Συνδυάζοντας τις σχέσεις (14.49) και (14.50) έχουμε

$$p_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} (1 - F_B(t)) + \int_0^t p_n(t-x) dF_B(x). \quad (14.51)$$

Αυτή είναι μια ανανεωτική εξίσωση και η λύση της είναι (βλέπε θεώρημα 12.15)

$$p_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} (1 - F_B(t)) + \int_0^t e^{-\lambda(t-x)} \frac{(\lambda(t-x))^n}{n!} (1 - F_B(t-x)) dm_B(x). \quad (14.52)$$

Στην περίπτωση που η ανανεωτική διαδικασία $\{B(t)\}$ είναι διαδικασία Poisson με ρυθμό μ έχουμε ότι $F_B(t) = 1 - e^{-\mu t}$, $m_B(t) = \mu t$, οπότε ο τύπος (14.52) μπορεί να απλοποιηθεί περαιτέρω ως εξής:

$$\begin{aligned} p_n(t) &= e^{-(\lambda+\mu)t} \frac{(\lambda t)^n}{n!} + \frac{\mu \lambda^n}{n!} \int_0^t e^{-(\lambda+\mu)(t-x)} (t-x)^n dx \\ &= e^{-(\lambda+\mu)t} \frac{(\lambda t)^n}{n!} + \frac{\mu \lambda^n}{(\lambda+\mu)^{n+1}} \int_0^t \frac{(\lambda+\mu)^{n+1}}{n!} e^{-(\lambda+\mu)u} u^n du \\ &= e^{-(\lambda+\mu)t} \frac{(\lambda t)^n}{n!} + \frac{\mu \lambda^n}{(\lambda+\mu)^{n+1}} \left(1 - \sum_{i=0}^n e^{-(\lambda+\mu)t} \frac{((\lambda+\mu)t)^i}{i!} \right). \end{aligned} \quad (14.53)$$

Για τις οριακές πιθανότητες στην περίπτωση αυτή έχουμε

$$\begin{aligned} p_n &= \lim_{t \rightarrow \infty} p_n(t) \\ &= \lim_{t \rightarrow \infty} \left(e^{-(\lambda+\mu)t} \frac{(\lambda t)^n}{n!} + \frac{\mu \lambda^n}{(\lambda+\mu)^{n+1}} \int_0^t \frac{(\lambda+\mu)^{n+1}}{n!} e^{-(\lambda+\mu)u} u^n du \right) \\ &= \frac{\mu \lambda^n}{(\lambda+\mu)^{n+1}}, \quad n \geq 0 \end{aligned} \quad (14.54)$$

που συμπίπτει με την (7.3).

Γενικότερα, η μέθοδος αυτή μπορεί να επεκταθεί και στην περίπτωση που η διαδικασία των αφίξεων είναι μια σύνθετη ή μη-ομογενής διαδικασία Poisson. Επίσης, οι πιθανότητες $p_n(t)$ μπορούν να υπολογιστούν σε κλειστή μορφή στην περίπτωση που οι χρόνοι μεταξύ των εκκαθαρίσεων ακολουθούν υπερεκθετικές, μίξεις Erlang ή Coxian κατανομές. Σε αυτές τις περιπτώσεις οι αντίστοιχες ανανεωτικές συναρτήσεις εκφράζονται σε πιο εύκολα διαχειρίσιμες μορφές για τους υπολογισμούς.

14.5 Ασκήσεις

Άσκηση 14.1 Θεωρούμε την $M/G/1$ ουρά με ρυθμό αφίξεων λ και Erlang(2, 2μ) κατανομή χρόνων εξυπηρέτησης (δηλαδή την $M/E_2/1$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ - βλέπε κεφάλαιο 8).

1. Να υπολογιστεί η πιθανογεννήτρια του αριθμού πελατών χρησιμοποιώντας τον τύπο Pollaczek-Khinchin (14.6).
2. Αν $\lambda = 1$ και $\mu = 3$ να βρείτε τύπο για την κατανομή ισορροπίας του αριθμού των πελατών, αντιστρέφοντας την πιθανογεννήτρια του αριθμού των πελατών.
3. Αν $\lambda = 1$ και $\mu = 3$ να βρείτε τύπο για την κατανομή του χρόνου παραμονής ενός πελάτη στο σύστημα, αντιστρέφοντας τον αντίστοιχο μετασχηματισμό Laplace-Stieltjes.

Άσκηση 14.2 Θεωρούμε την $M/G/1$ ουρά με ρυθμό αφίξεων $\lambda = 1$ και υπερεκθετική κατανομή χρόνων εξυπηρέτησης. Συγκεκριμένα, ο χρόνος εξυπηρέτησης ενός πελάτη ακολουθεί την $\text{Exp}(1)$ κατανομή με πιθανότητα $\frac{1}{4}$ και την $\text{Exp}(2)$ κατανομή με πιθανότητα $\frac{3}{4}$.

1. Να υπολογιστεί ο μετασχηματισμός Laplace-Stieltjes ενός χρόνου εξυπηρέτησης.
2. Να υπολογιστεί η πιθανογεννήτρια του αριθμού πελατών χρησιμοποιώντας τον τύπο Pollaczek-Khinchin (14.6).
3. Να βρείτε τύπο για την κατανομή ισορροπίας του αριθμού των πελατών, αντιστρέφοντας την πιθανογεννήτρια του αριθμού των πελατών.
4. Να βρείτε τύπο για την κατανομή του χρόνου παραμονής ενός πελάτη στο σύστημα, αντιστρέφοντας τον αντίστοιχο μετασχηματισμό Laplace-Stieltjes.

Άσκηση 14.3 Θεωρούμε την $M/G/1$ ουρά με ρυθμό αφίξεων λ και κατανομή χρόνων εξυπηρέτησης $F_B(x)$, όπου κάθε πελάτης που συμπληρώνει την εξυπηρέτησή του την επαναλαμβάνει άμεσα με πιθανότητα $q \in [0, 1]$. Υποθέτουμε ότι οι διαδοχικές επαναλήψεις των εξυπηρετήσεων είναι ανεξάρτητες και ισόνομες.

1. Να υπολογιστεί η πιθανογεννήτρια της κατανομής ισορροπίας του αριθμού των πελατών στο σύστημα.
2. Να υπολογιστεί ο μετασχηματισμός Laplace-Stieltjes του χρόνου παραμονής ενός πελάτη στο σύστημα.

Άσκηση 14.4 Θεωρούμε την τροποποίηση της $M/G/1$ ουράς με ρυθμό αφίξεων λ και κατανομή χρόνων εξυπηρέτησης $F_B(x)$, όπου ο πρώτος πελάτης κάθε περιόδου συνεχούς λειτουργίας έχει διαφορετική κατανομή χρόνων εξυπηρέτησης, έστω την $F_{B_0}(x)$.

1. Να υπολογιστεί η πιθανογεννήτρια της κατανομής ισορροπίας του αριθμού των πελατών στο σύστημα.
2. Να υπολογιστεί ο μετασχηματισμός Laplace-Stieltjes του χρόνου παραμονής ενός πελάτη στο σύστημα.

Άσκηση 14.5 Θεωρούμε την $M/G/1$ ουρά με πολλαπλές διακοπές υπηρετή, δηλαδή το μοντέλο που είδαμε στην παράγραφο 13.3. Αυτή είναι η τροποποίηση της $M/G/1$ ουράς με ρυθμό αφίξεων λ και γενική κατανομή χρόνων εξυπηρέτησης $F_B(x)$, όπου κάθε φορά που το σύστημα μένει κενό ο υπηρέτης ξεκινά μια περίοδο διακοπών. Όταν λήξει ο χρόνος διακοπών, ο υπηρέτης επιστρέφει και αν το σύστημα εξακολουθεί να είναι άδειο αρχίζει μια νέα περίοδο διακοπών κ.ο.κ., ενώ αν στο σύστημα υπάρχουν πελάτες σε αναμονή ξεκινά να εξυπηρετεί. Οι διάρκειες των περιόδων διακοπών ακολουθούν μια γενική κατανομή $F_V(x)$. Να βρεθεί η πιθανογεννήτρια της κατανομής ισορροπίας του αριθμού των πελατών στο σύστημα.

Άσκηση 14.6 Έστω η $M/G/1$ ουρά με ομαδικές αφίξεις, όπου οι πελάτες φθάνουν σε ομάδες σύμφωνα με μια διαδικασία Poisson με ρυθμό λ και κάθε ομάδα είναι μεγέθους j με πιθανότητα $g_j, j \geq 1$. Οι πελάτες εξυπηρετούνται ένας-ένας από τον μοναδικό υπηρέτη του συστήματος και οι χρόνοι εξυπηρέτησης είναι ανεξάρτητοι και ισόνομοι με κατανομή $F_B(x)$.

1. Να υπολογιστεί η πιθανογεννήτρια του συνολικού αριθμού αφίξεων πελατών κατά τη διάρκεια ενός χρόνου εξυπηρέτησης.
2. Να υπολογιστεί η πιθανογεννήτρια της κατανομής ισορροπίας του αριθμού των πελατών στο σύστημα.

Άσκηση 14.7 Θεωρούμε το σύστημα εξυπηρέτησης $D/M/1$, όπου οι ενδιάμεσοι χρόνοι είναι ίσοι με 4 χρονικές μονάδες και ο μέσος χρόνος εξυπηρέτησης είναι 2 χρονικές μονάδες. Να βρεθεί η κατανομή του χρόνου παραμονής ενός πελάτη στο σύστημα.

Άσκηση 14.8 Θεωρούμε την $G/M/1$ ουρά, όπου οι ενδιάμεσοι χρόνοι των αφίξεων είναι $\text{Exp}(3)$ με πιθανότητα $\frac{13}{24}$ και $\text{Exp}(2)$ με πιθανότητα $\frac{11}{24}$, ενώ οι χρόνοι εξυπηρέτησης είναι $\text{Exp}(6)$.

1. Να βρεθούν οι κατανομές ισορροπίας του αριθμού των πελατών σε στιγμές αφίξεων, αναχωρήσεων και σε συνεχή χρόνο.

2. Να βρεθεί η κατανομή του χρόνου παραμονής ενός πελάτη στο σύστημα.

Άσκηση 14.9 Να βρείτε την κατανομή ισορροπίας του αριθμού των πελατών σε στιγμές αφίξεων, αναχωρήσεων και σε συνεχή χρόνο για την $E_2/M/1$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησεων μ .

Άσκηση 14.10 Να βρείτε την κατανομή ισορροπίας του αριθμού των πελατών σε στιγμές αφίξεων για την $G/M/1/2$ ουρά με κατανομή ενδιάμεσων χρόνων αφίξεων $F_A(x)$ και $\text{Exp}(\mu)$ χρόνους εξυπηρέτησης.

14.6 Σχόλια

Όλα τα εισαγωγικά βιβλία της Θεωρίας Ουρών Αναμονής αφιερώνουν αρκετό χώρο στη μελέτη των κλασικών μη-Μαρκοβιανών μοντέλων. Για μια σύντομη εισαγωγή με πολλά παραδείγματα και υπολογιστικό προσανατολισμό, ο ενδιαφερόμενος αναγνώστης μπορεί να μελετήσει το σύγγραμμα των Adan και Resing 2001. Για περισσότερη λεπτομέρεια και παραδείγματα, μπορεί κανείς να δει τα κλασικά βιβλία Kleinrock 1975, Wolff 1989, Tijms 2008, Kulkarni 2010 και Shortle, Thompson, Gross και Harris 2018, ή τα νεώτερα συγγράμματα των Haviv 2013 και Harchol-Balter 2013.

Μια μεθοδολογία που έχει υπολογιστικά πλεονεκτήματα είναι η προσέγγιση ενός μη-Μαρκοβιανού μοντέλου από ένα Μαρκοβιανό μοντέλο με βάση τη μέθοδο των φάσεων. Η ιδέα αυτή περιγράφεται με μεγάλη λεπτομέρεια στα συγγράμματα των Neuts 1981, Neuts 1989 και Latouche και Ramaswami 1999.

Ένα θέμα που δεν έχουμε θίξει στο παρόν κεφάλαιο είναι το θέμα των προσεγγίσεων και των φραγμάτων. Για μια σύντομη εισαγωγή στο θέμα αυτό ο ενδιαφερόμενος αναγνώστης μπορεί να κοιτάξει στο τελευταίο κεφάλαιο του Wolff 1989.

Βιβλιογραφία

- [1] D.R. Cox. "The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables". Στο: *Mathematical Proceedings of the Cambridge Philosophical Society* 51 (1955), σσ. 433–441.
- [2] I. Adan και J. Resing. *Queueing Theory*. Eindhoven: Notes available online, 2001.
- [3] L. Kleinrock. *Queueing Systems. Volume 1: Theory*. Wiley, 1975. ISBN: 978-0471491101.
- [4] R.W. Wolff. *Stochastic Modeling and the Theory of Queues*. Englewood Cliffs, New Jersey: Prentice Hall, 1989. ISBN: 978-0138466923.
- [5] H.C. Tijms. *A First Course in Stochastic Models, 2nd Edition*. Chichester: Wiley, 2008. ISBN: 978-0471498803.
- [6] V.G. Kulkarni. *Modeling and Analysis of Stochastic Systems, 2nd Edition*. CRC Press, 2010. ISBN: 978-1439808757.
- [7] J.F. Shortle, J.M. Thompson, D. Gross και C.M. Harris. *Fundamentals of Queueing Theory, 5th Edition*. Wiley, 2018. ISBN: 978-1118943526.
- [8] M. Haviv. *Queues: A Course in Queueing Theory*. Springer, 2013. ISBN: 978-1461467649.
- [9] M. Harchol-Balter. *Performance Modeling and Design of Computer Systems: Queueing Theory in Action*. Cambridge, 2013. ISBN: 978-1107027503.
- [10] M.F. Neuts. *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*. The Johns Hopkins University Press, 1981. ISBN: 978-0801825606.
- [11] M.F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. CRC Press, 1989. ISBN: 978-0824782832.

- [12] G. Latouche και V. Ramaswami. *Introduction to Matrix Analytic Models in Stochastic Modeling*. ASA-SIAM Series on Statistics και Applied Probability, 1999. ISBN: 978-0898714258.

Μέρος IV

ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ, ΕΛΕΓΧΟΣ ΚΑΙ ΣΤΡΑΤΗΓΙΚΗ ΣΥΜΠΕΡΙΦΟΡΑ

ΒΕΛΤΙΣΤΟΣ ΣΧΕΔΙΑΣΜΟΣ ΚΑΙ ΣΤΡΑΤΗΓΙΚΗ ΣΥΜΠΕΡΙΦΟΡΑ

Στα προηγούμενα κεφάλαια είδαμε κάποιες μεθόδους για την αποτίμηση της απόδοσης συγκεκριμένων συστημάτων εξυπηρέτησης, βασιζόμενοι στην παραδοχή ότι οι οντότητες που εμπλέκονται (πελάτες, υπηρέτες, διαχειριστής) είναι παθητικές, δηλαδή δεν λαμβάνουν αποφάσεις. Στο κεφάλαιο αυτό, θα μελετήσουμε προβλήματα βέλτιστης λήψης αποφάσεων σε συστήματα εξυπηρέτησης, και πιο συγκεκριμένα, θα θεωρήσουμε ότι κάποιες από τις οντότητες είναι στρατηγικές, δηλαδή παίρνουν αποφάσεις για μεγιστοποίηση της ωφέλειάς τους. Η περιοχή αυτή συνθέτει ιδέες από τη Θεωρία Ουρών Αναμονής, από τον Μαθηματικό Προγραμματισμό και από τη Θεωρία Παιγνίων.

Χωρίς να υπεισέλθουμε σε μεγαλύτερη λεπτομέρεια αυτή τη στιγμή, τα κυριότερα προβλήματα που τίθενται είναι τα εξής:

- Εύρεση στρατηγικών ισορροπίας, δηλαδή στρατηγικών που κανείς από τους αποφασίζοντες - παίκτες δεν έχει κίνητρο να αλλάξει μονομερώς.
- Εύρεση κοινωνικά βέλτιστων στρατηγικών, δηλαδή στρατηγικών που μεγιστοποιούν κάποιο μέτρο συνολικής κοινωνικής ωφέλειας.
- Ποσοτικοποίηση της απόκλισης των στρατηγικών ισορροπίας και των κοινωνικά βέλτιστων στρατηγικών. Ένα κλασικό μέτρο εδώ είναι η λεγόμενη «Τιμή της Αναρχίας» (Price of Anarchy).
- Επιβολή μηχανισμών ρύθμισης, ώστε κοινωνικά βέλτιστες ή έστω κοινωνικά ικανοποιητικές στρατηγικές να προκύψουν ως στρατηγικές ισορροπίας. Κλασικά εργαλεία για τον σκοπό αυτό είναι η κατάλληλη τιμολόγηση των υπηρεσιών, η ύπαρξη προτεραιοτήτων στην εξυπηρέτηση των πελατών, η κατάλληλη επιλογή πειθαρχίας ουράς και ο έλεγχος της πληροφορίας που παρέχεται στους πελάτες σχετικά με την κατάσταση του συστήματος.

15.1 Βασικές έννοιες από τη Θεωρία Παιγνίων

Ένα παιχνίδι προσδιορίζεται από ένα σύνολο παικτών $\mathcal{N} = \{1, 2, \dots, n\}$, σύνολα πλάνων δράσης \mathcal{A}_i , ένα για κάθε παίκτη $i = 1, 2, \dots, n$, και συναρτήσεις πληρωμής \mathcal{U}_i , μία για κάθε παίκτη $i = 1, 2, \dots, n$. Το σύνολο \mathcal{A}_i περιέχει όλα τα δυνατά πλάνα δράσης του παίκτη i , που του υπαγορεύουν πώς να παίξει κατά τη διάρκεια του παιχνιδιού, ανάλογα με την κατάσταση στην οποία θα βρεθεί. Κάθε στοιχείο του \mathcal{A}_i αναφέρεται και ως καθαρή στρατηγική του i . Μια κατανομή πιθανότητας πάνω στο \mathcal{A}_i αναφέρεται ως μεικτή στρατηγική του i και έχει την έννοια, ότι ο παίκτης i επιλέγει μία από τις καθαρές στρατηγικές του για να παίξει το παιχνίδι με τις πιθανότητες που υπαγορεύει η μεικτή στρατηγική του. Το σύνολο των μεικτών στρατηγικών του i , συμβολίζεται με \mathcal{S}_i . Ένα προφίλ στρατηγικών \mathbf{s} είναι μια διατεταγμένη n -άδα στρατηγικών, μία για κάθε παίκτη, δηλαδή

$$\mathbf{s} = (s_1, s_2, \dots, s_n), \quad s_i \in \mathcal{S}_i, \quad i = 1, 2, \dots, n.$$

Δεδομένου του προφίλ \mathbf{s} , γράφουμε $\mathbf{s} = (s_i, \mathbf{s}_{-i})$, εννοώντας ότι το $(n-1)$ -διάστατο διάνυσμα \mathbf{s}_{-i} περιέχει τις στρατηγικές του \mathbf{s} πλην αυτής που αντιστοιχεί στο παίκτη i . Η συνάρτηση πληρωμής \mathcal{U}_i παίρνει τιμές στο \mathbb{R} και έχει πεδίο ορισμού το σύνολο των προφίλ στρατηγικών όλων των παικτών. Η ποσότητα $\mathcal{U}_i(\mathbf{s}) = \mathcal{U}_i(s_i, \mathbf{s}_{-i})$ δίνει την πληρωμή που θα λάβει ο παίκτης i , αν υιοθετηθεί το προφίλ στρατηγικών \mathbf{s} , δηλαδή αν ο i ακολουθήσει τη στρατηγική s_i και οι υπόλοιποι παίκτες ακολουθήσουν τις στρατηγικές του \mathbf{s}_{-i} .

Η συνάρτηση $\mathcal{U}_i(\mathbf{s}) = \mathcal{U}_i(s_i, \mathbf{s}_{-i})$ είναι γραμμική ως προς s_i , υπό την έννοια ότι αν η s_i είναι μίξη των στρατηγικών $s_i^k, k = 1, 2, \dots, r$ με αντίστοιχες πιθανότητες $\alpha_k, k = 1, 2, \dots, r$ με $\sum_{k=1}^r \alpha_k = 1$ τότε

$$\mathcal{U}_i(s_i, \mathbf{s}_{-i}) = \sum_{k=1}^r \alpha_k \mathcal{U}_i(s_i^k, \mathbf{s}_{-i}).$$

Αν s_i^1 και s_i^2 είναι δυο στρατηγικές του παίκτη i , λέμε ότι η s_i^1 κυριαρχεί τελείως ασθενώς της s_i^2 αν για κάθε προφίλ στρατηγικών των υπολοίπων παικτών, \mathbf{s}_{-i} , ισχύει

$$\mathcal{U}_i(s_i^1, \mathbf{s}_{-i}) \geq \mathcal{U}_i(s_i^2, \mathbf{s}_{-i}).$$

Αν η ανισότητα ισχύει γνήσια για ένα τουλάχιστον προφίλ \mathbf{s}_{-i} , λέμε ότι η s_i^1 κυριαρχεί ασθενώς της s_i^2 , και αν η ανισότητα ισχύει γνήσια για όλα τα προφίλ \mathbf{s}_{-i} , τότε λέμε ότι η s_i^1 κυριαρχεί ισχυρά της s_i^2 .

Έστω ένας παίκτης i . Δοθέντος ενός προφίλ στρατηγικών \mathbf{s}_{-i} των υπολοίπων παικτών, η στρατηγική s_i^* του παίκτη i λέγεται βέλτιστη απάντηση στο \mathbf{s}_{-i} , αν για κάθε στρατηγική s_i του i ισχύει

$$\mathcal{U}_i(s_i^*, \mathbf{s}_{-i}) \geq \mathcal{U}_i(s_i, \mathbf{s}_{-i}),$$

δηλαδή η s_i^* βελτιστοποιεί την $f(s_i) = \mathcal{U}_i(s_i, \mathbf{s}_{-i})$. Το σύνολο των βέλτιστων απαντήσεων στο \mathbf{s}_{-i} , συμβολίζεται με $BR(\mathbf{s}_{-i})$.

Μια στρατηγική s_i του i λέγεται τελείως ασθενώς (αντίστοιχα ασθενώς, ή ισχυρά) κυριαρχούσα, αν κυριαρχεί τελείως ασθενώς (αντίστοιχα ασθενώς, ή ισχυρά) κάθε άλλης στρατηγικής του i . Μια στρατηγική ενός παίκτη είναι τελείως ασθενώς (αντίστοιχα ασθενώς, ή ισχυρά) κυριαρχούσα αν ο παίκτης αποκομίζει ακολουθώντας την μεγαλύτερη ή ίση ωφέλεια (αντίστοιχα μεγαλύτερη ή ίση ωφέλεια με τουλάχιστον μια φορά μεγαλύτερη, ή μεγαλύτερη ωφέλεια) από όση θα αποκόμιζε ακολουθώντας κάποια άλλη στρατηγική, ό,τι και να κάνουν οι υπόλοιποι παίκτες.

Ένα προφίλ στρατηγικών $\mathbf{s}^e = (s_1^e, s_2^e, \dots, s_n^e)$ λέγεται σημείο στρατηγικής ισορροπίας, αν για κάθε παίκτη i , η στρατηγική s_i^e είναι βέλτιστη απάντηση έναντι της \mathbf{s}_{-i}^e . Με άλλα λόγια ένα προφίλ στρατηγικών είναι σημείο στρατηγικής ισορροπίας, αν κανένας παίκτης δεν έχει κίνητρο να αλλάξει στρατηγική μονομερώς, υπό την έννοια ότι αν μετακινηθεί μόνος του προς κάποια άλλη στρατηγική, η ωφέλεια του αποκλείεται να αυξηθεί (αλλά ενδέχεται να μειωθεί).

15.2 Στρατηγική αλληλεπίδραση μεταξύ των πελατών σε συστήματα ουρών αναμονής

Για την ανάλυση της στρατηγικής συμπεριφοράς των πελατών σε συστήματα ουρών αναμονής, οι έννοιες της Θεωρίας Παιγνίων που εκτέθηκαν παραπάνω είναι πολύ χρήσιμες. Δεν μπορούν, όμως, να μεταφερθούν άμεσα διότι υπάρχουν δυο θεμελιώδη προβλήματα. Το πρώτο πρόβλημα είναι ότι οι παίκτες, δηλαδή οι δυνητικοί πελάτες ενός συστήματος είναι άπειροι το πλήθος. Το δεύτερο πρόβλημα είναι ότι οι παίκτες δεν παίρνουν ταυτόχρονα τις αποφάσεις τους, αφού φθάνουν στο σύστημα διαδοχικά σε άπειρο ορίζοντα. Τα προβλήματα αυτά ξεπερνιούνται δημιουργώντας ανάλογες έννοιες, εκμεταλλευόμενοι την ομοιογένεια των πελατών. Δηλαδή σε πρώτο επίπεδο θεωρούμε ότι όλοι οι πελάτες είναι όμοιοι και κατόπιν μπορούμε να επεκτείνουμε το πλαίσιο και στην περίπτωση ετερογενών πελατών, υποθέτοντας ότι υπάρχουν διάφορες κλάσεις πελατών και οι πελάτες κάθε κλάσης είναι όμοιοι (ομογενείς).

Στην περίπτωση, λοιπόν, ομογενών πελατών, ένα παιχνίδι μεταξύ των πελατών προσδιορίζεται από το κοινό σύνολο των στρατηγικών τους, \mathcal{S} , και μια συνάρτηση πληρωμής $\mathcal{U}(s, s')$, που δίνει την πληρωμή ενός πελάτη που χρησιμοποιεί τη στρατηγική s , όταν όλοι οι άλλοι χρησιμοποιούν τη στρατηγική s' .

Η συνάρτηση $\mathcal{U}(s, s')$ είναι γραμμική ως προς s , υπό την έννοια ότι αν η s είναι η μίξη των στρατηγικών s^k , $k = 1, 2, \dots, r$ με αντίστοιχες πιθανότητες α_k , $k = 1, 2, \dots, r$, με $\sum_{k=1}^r \alpha_k = 1$, τότε

$$\mathcal{U}(s, s') = \sum_{k=1}^r \alpha_k \mathcal{U}(s^k, s').$$

Αν s^1 και s^2 είναι δυο στρατηγικές ενός πελάτη, τότε λέμε ότι η s^1 κυριαρχεί τελείως ασθενώς της s^2 , αν για κάθε στρατηγική s' των άλλων πελατών ισχύει

$$\mathcal{U}(s^1, s') \geq \mathcal{U}(s^2, s').$$

Αν η ανισότητα ισχύει γνήσια για μια τουλάχιστον στρατηγική s' λέμε ότι η s^1 κυριαρχεί ασθενώς της s^2 . Τέλος, αν η ανισότητα ισχύει γνήσια για όλες τις στρατηγικές s' , τότε λέμε ότι η s^1 κυριαρχεί ισχυρά της s^2 .

Έστω ένας επιλεγμένος πελάτης. Δοθείσης μιας στρατηγικής s' που ακολουθούν όλοι οι άλλοι πελάτες, η στρατηγική s^* του επιλεγμένου πελάτη λέγεται βέλτιστη απάντηση στην s' , αν για κάθε στρατηγική s του επιλεγμένου πελάτη ισχύει

$$\mathcal{U}(s^*, s') \geq \mathcal{U}(s, s'),$$

δηλαδή η s^* βελτιστοποιεί της $f(s) = \mathcal{U}(s, s')$. Το σύνολο των βέλτιστων απαντήσεων στο s' , συμβολίζεται με $BR(s')$.

Μια στρατηγική s ενός επιλεγμένου πελάτη λέγεται τελείως ασθενώς (αντίστοιχα ασθενώς, ή ισχυρά) κυριαρχούσα, αν κυριαρχεί τελείως ασθενώς (αντίστοιχα ασθενώς, ή ισχυρά) κάθε άλλης στρατηγικής του επιλεγμένου πελάτη. Μια στρατηγική ενός επιλεγμένου πελάτη είναι τελείως ασθενώς (αντίστοιχα ασθενώς, ή ισχυρά) κυριαρχούσα, αν ο επιλεγμένος πελάτης αποκομίζει ακολουθώντας την μεγαλύτερη ή ίση ωφέλεια (αντίστοιχα μεγαλύτερη ή ίση ωφέλεια με τουλάχιστον μια φορά μεγαλύτερη, ή μεγαλύτερη ωφέλεια) από όση θα αποκόμιζε ακολουθώντας κάποια άλλη στρατηγική, ανεξάρτητα από το τι θα έκαναν οι υπόλοιποι πελάτες.

Μια στρατηγική s^e λέγεται (συμμετρική) στρατηγική ισορροπίας, αν είναι βέλτιστη απάντηση έναντι του εαυτού της. Δηλαδή, η s^e είναι στρατηγική ισορροπίας αν

$$\mathcal{U}(s^e, s^e) \geq \mathcal{U}(s, s^e), \quad s \in \mathcal{S},$$

ή, ισοδύναμα, όταν $s^e \in BR(s^e)$. Με άλλα λόγια μια στρατηγική είναι στρατηγική ισορροπίας, αν κανένας πελάτης δεν έχει κίνητρο να την αλλάξει μονομερώς, όταν όλοι οι άλλοι την ακολουθούν.

Ένα βασικό ερώτημα αφορά τον υπολογισμό της συνάρτησης πληρωμής $\mathcal{U}(s, s')$ σε ένα στρατηγικό πρόβλημα απόφασης πελατών στο πλαίσιο ενός συγκεκριμένου συστήματος εξυπηρέτησης. Η βασική ιδέα είναι ότι αν έχουμε έναν επιλεγμένο πελάτη που ακολουθεί την στρατηγική s , όταν οι υπόλοιποι ακολουθούν τη

στρατηγική s' , τότε αυτός δεν επηρεάζει την γενική συμπεριφορά του συστήματος. Η γενική συμπεριφορά του συστήματος προσδιορίζεται από τη στρατηγική s' που ακολουθούν οι υπόλοιποι πελάτες, μια και η επίδραση του επιλεγμένου πελάτη είναι αμελητέα. Στον υπολογισμό της συμπεριφοράς του συστήματος υποτίθεται, επίσης, ότι το σύστημα εξετάζεται για μεγάλο χρονικό διάστημα και επομένως τα διάφορα μέτρα υπολογίζονται υποθέτοντας ότι το σύστημα βρίσκεται σε κατάσταση στατιστικής ισορροπίας (στασιμότητας).

Επομένως, η διαδικασία εύρεσης των στρατηγικών ισορροπίας σε ένα συγκεκριμένο πρόβλημα με στρατηγικούς πελάτες γίνεται σε βήματα ως εξής:

Βήμα 1ο: Μελέτη της στάσιμης συμπεριφοράς ενός συστήματος κάτω από κάθε στρατηγική s' των πελατών.

Βήμα 2ο: Υπολογισμός της συνάρτησης πληρωμής $\mathcal{U}(s, s')$ ενός επιλεγμένου πελάτη που ακολουθεί την στρατηγική s , όταν οι άλλοι ακολουθούν τη στρατηγική s' .

Βήμα 3ο: Εύρεση βέλτιστης απάντησης του επιλεγμένου πελάτη έναντι μιας στρατηγικής των άλλων:

$$BR(s') = \{s \in \mathcal{S} : \mathcal{U}(s, s') \geq \mathcal{U}(\hat{s}, s'), \hat{s} \in \mathcal{S}\}.$$

Βήμα 4ο: Εύρεση στρατηγικών με την ιδιότητα $s^e \in BR(s^e)$.

15.3 Συμπεριφορές απόφυγε-το-πλήθος και ακολούθησε-το-πλήθος

Ας υποθέσουμε ότι σε ένα πρόβλημα στρατηγικής συμπεριφοράς πελατών ο χώρος των στρατηγικών είναι ολικά διατεταγμένος, δηλαδή υπάρχει σχέση διάταξης, ώστε για κάθε δυο στρατηγικές s και s' να ισχύει $s \leq s'$ ή $s \geq s'$. Υποθέτουμε, επίσης, ότι κάθε στρατηγική s έχει μοναδική βέλτιστη απάντηση, την οποία συμβολίζουμε με $BR(s)$ (κανονικά το $BR(s)$ είναι το σύνολο των βέλτιστων απαντήσεων, αλλά αφού εδώ πρόκειται για μονοσύνολο το ταυτίζουμε με την μοναδική στρατηγική που περιέχει).

Αν η συνάρτηση $BR(s)$ είναι αύξουσα ως προς s , λέμε ότι η στρατηγική συμπεριφορά των πελατών στο εν λόγω σύστημα είναι του τύπου ακολούθησε-το-πλήθος (Follow-The-Crowd (FTC)). Διαισθητικά, αυτό σημαίνει ότι αν οι πελάτες υιοθετήσουν μια καινούργια στρατηγική που είναι «μεγαλύτερη» τότε και η βέλτιστη αντίδραση ενός επιλεγμένου πελάτη είναι η μετακίνηση σε μια καινούργια στρατηγική που είναι «μεγαλύτερη» της αρχικής του. Οπότε, ο επιλεγμένος πελάτης αλλάζει στρατηγική προς την ίδια κατεύθυνση με τους άλλους πελάτες. Στην περίπτωση αυτή συνήθως υπάρχουν πολλές στρατηγικές ισορροπίας. Πράγματι οι στρατηγικές ισορροπίας αντιστοιχούν στα σημεία που το γράφημα της συνάρτησης $y = BR(s)$ τέμνει την ευθεία $y = s$. Δεδομένου ότι η $BR(s)$ είναι αύξουσα, τα αντίστοιχα σημεία τομής μπορεί να είναι περισσότερα από ένα.

Αν η συνάρτηση $BR(s)$ είναι φθίνουσα ως προς s , τότε η στρατηγική συμπεριφορά των πελατών αναφέρεται ως τύπου απόφυγε-το-πλήθος (Avoid-The-Crowd (ATC)). Στην περίπτωση αυτή, αν οι πελάτες υιοθετήσουν μια καινούργια στρατηγική που είναι «μεγαλύτερη» από την προηγούμενη επιλογή τους, τότε η βέλτιστη αντίδραση ενός επιλεγμένου πελάτη είναι να κινηθεί σε αντίθετη κατεύθυνση. Σε αυτή την περίπτωση, υπάρχει μοναδική στρατηγική ισορροπίας που βρίσκεται στο σημείο τομής του γραφήματος της φθίνουσας συνάρτησης $y = BR(s)$ με την ευθεία $y = s$.

Τα πιο κλασικά προβλήματα στρατηγικής συμπεριφοράς πελατών είναι αυτά στα οποία οι πελάτες αντιμετωπίζουν το δίλημμα της εισόδου/αποχώρησης κατά την άφιξή τους σε ένα σύστημα. Στις περισσότερες τέτοιες περιπτώσεις η στρατηγική συμπεριφορά είναι τύπου ATC. Πράγματι, αν ένας επιλεγμένος πελάτης μάθει ότι οι άλλοι πελάτες γίνονται περισσότερο πρόθυμοι να εισέλθουν στο σύστημα, τότε αυτός γίνεται περισσότερο απρόθυμος να εισέλθει, καθώς αναμένει μεγαλύτερο συνωστισμό σε αυτό και επομένως μεγαλύτερα κόστη αναμονής.

Βεβαίως υπάρχουν και περιπτώσεις προβλημάτων στρατηγικής συμπεριφοράς στις οποίες η συμπεριφορά είναι τύπου FTC. Π.χ., όταν οι πελάτες αντιμετωπίζουν το δίλημμα της αγοράς δικαιώματος προτεραιότητας, αν ένας πελάτης μάθει ότι οι άλλοι πελάτες γίνονται περισσότερο πρόθυμοι να αγοράσουν ένα τέτοιο δικαίωμα, τότε και αυτός γίνεται περισσότερο πρόθυμος να αγοράσει το δικαίωμα, αλλιώς θα εξυπηρετηθεί αργότερα από τους πελάτες που έχουν το δικαίωμα και επομένως θα υποστεί μεγαλύτερη αναμονή.

15.4 Στρατηγικές κατωφλίου

Ας υποθέσουμε ότι οι πελάτες ενός συστήματος εξυπηρέτησης παρατηρούν το πλήθος των πελατών στο σύστημα πριν επιλέξουν μία από δύο αποφάσεις A_1 και A_2 (π.χ., μπορεί να παρατηρούν το πλήθος των πελατών στο σύστημα κατά τη στιγμή της άφιξής τους και να αποφασίζουν αν θα μπουκ ή όχι). Στην περίπτωση αυτή, και εφόσον δεν υπάρχει επιπλέον πληροφορία που να δίνεται στους πελάτες για την κατάσταση του συστήματος, μια γενική στρατηγική έχει τη μορφή $\mathbf{q} = (q_0, q_1, q_2 \dots)$, όπου q_i είναι η πιθανότητα ένας πελάτης να πάρει την απόφαση A_1 , όταν το πλήθος των πελατών είναι i .

Η στρατηγική $\mathbf{q} = (1, 1, \dots, 1, 0, 0, \dots)$ με 1 στις πρώτες n θέσεις και 0 στις υπόλοιπες λέγεται (καθαρή) στρατηγική κατωφλίου με κατώφλι n και υπαγορεύει τη λήψη της απόφασης A_1 εφόσον ο πελάτης παρατηρήσει $0, 1, 2, \dots, n - 1$ πελάτες πριν πάρει την απόφαση και τη λήψη της απόφασης A_2 , διαφορετικά. Με άλλα λόγια, υπό τη στρατηγική κατωφλίου n , ο πελάτης παίρνει την απόφαση A_1 αν το πλήθος των πελατών συμπεριλαμβανομένου του εαυτού του είναι το πολύ n .

Η στρατηγική $\mathbf{q} = (1, 1, \dots, 1, p, 0, 0, \dots)$ με 1 στις πρώτες n θέσεις, $p \in [0, 1]$ στην $n + 1$ θέση και 0 στις υπόλοιπες λέγεται μεικτή (ή τυχαίοποιημένη) στρατηγική κατωφλίου με κατώφλι $x = n + p$. Υπό αυτή τη στρατηγική, ένας πελάτης λαμβάνει την απόφαση A_1 εφόσον παρατηρήσει $0, 1, \dots, n - 1$ πελάτες πριν πάρει την απόφαση του και την απόφαση A_2 εφόσον παρατηρήσει $n + 1, n + 2, \dots$ πελάτες. Αν παρατηρήσει ακριβώς n πελάτες, τότε παίρνει την απόφαση A_1 με πιθανότητα p και την απόφαση A_2 με τη συμπληρωματική πιθανότητα $1 - p$. Για $p = 0$ ή 1 , δηλαδή x ακέραιο, η μεικτή στρατηγική κατωφλίου $x = n + p$ ανάγεται στην αντίστοιχη καθαρή στρατηγική κατωφλίου.

Σε αρκετές εφαρμογές, αποδεικνύεται ότι οι στρατηγικές ισορροπίας ή οι κοινωνικά βέλτιστες στρατηγικές είναι τύπου κατωφλίου. Ακόμη και αν αυτό δεν ισχύει, συχνά η αναζήτηση στρατηγικών ισορροπίας και/ή κοινωνικά βέλτιστων στρατηγικών περιορίζεται στην κλάση των στρατηγικών κατωφλίου ή στην κλάση των μεικτών στρατηγικών κατωφλίου.

Στην περίπτωση που έχουμε συμπεριφορά FTC και οι βέλτιστες απαντήσεις είναι τύπου κατωφλίου, τότε εμφανίζονται τυπικά πολλές καθαρές στρατηγικές ισορροπίας τύπου κατωφλίου και ανάμεσα σε κάθε δυο διαδοχικές από αυτές παρεμβάλλεται μια μεικτή στρατηγική ισορροπίας τύπου κατωφλίου. Αν έχουμε συμπεριφορά ATC και οι βέλτιστες απαντήσεις είναι τύπου κατωφλίου τότε υπάρχει μοναδική στρατηγική ισορροπίας που μπορεί να είναι είτε καθαρή στρατηγική κατωφλίου, είτε μεικτή στρατηγική κατωφλίου.

15.5 Πλαίσιο κοινωνικής βελτιστοποίησης και βελτιστοποίησης μονοπωλίου

Σε ένα περιβάλλον όπου συνυπάρχουν ως οντότητες οι πελάτες και ο διαχειριστής ενός συστήματος, οι πελάτες επιθυμούν να βελτιστοποιήσουν το καθαρό πλεόνασμά τους, δηλαδή την αναμενόμενη αμοιβή/ωφέλεια από την εξυπηρέτηση μείον την πλήρη τιμή που περιλαμβάνει τόσο την τιμή της υπηρεσίας που μπορεί να επιβάλει ο διαχειριστής όσο και τα αναμενόμενα κόστη αναμονής. Συνήθως, οι πελάτες θεωρούνται ουδέτεροι ως προς τον κίνδυνο, δηλαδή επιθυμούν να μεγιστοποιήσουν το αναμενόμενο πλεόνασμά τους, χωρίς να λαμβάνουν υπόψη τη διασπορά του. Από την άλλη μεριά, ο διαχειριστής του συστήματος μπορεί να δρα ως κοινωνικός σχεδιαστής (social planner) ή ως μονοπώλειο.

Όταν ο διαχειριστής ενός συστήματος δρα ως κοινωνικός σχεδιαστής, υποθέτουμε ότι επιθυμεί να μεγιστοποιήσει τον συνολικό κοινωνικό πλούτο ανά χρονική μονάδα, που ορίζεται ως η αναμενόμενη συνολική ωφέλεια ανά χρονική μονάδα όλων των εμπλεκόμενων οντοτήτων (και των πελατών και του ιδίου). Στην περίπτωση αυτή, οι πληρωμές μεταξύ των οντοτήτων δεν εμφανίζονται στη συνάρτηση της κοινωνικής ωφέλειας, αφού αφαιρούνται από κάποιες οντότητες και προστίθενται σε άλλες. Έτσι, π.χ., η επιβολή ενός τέλους εισόδου σε ένα σύστημα εξυπηρέτησης δεν θα εμφανιστεί στη συνάρτηση της κοινωνικής ωφέλειας, αφού θα αφαιρεθεί από την ωφέλεια των πελατών και θα προστεθεί στην ωφέλεια του διαχειριστή συνεισφέροντας τελικά 0 στη συνολική κοινωνική ωφέλεια. Όμως, η επιβολή του τέλους επηρεάζει τη συμπεριφορά των πελατών που δρουν ιδιοτελώς προσπαθώντας να μεγιστοποιήσουν την ωφέλειά τους και επομένως εμμέσως επηρεάζει τη

συνολική κοινωνική ωφέλεια. Σε γενικές γραμμές, η συνολική κοινωνική ωφέλεια ανά χρονική μονάδα ισούται με την συνολική ωφέλεια των πελατών από την εξυπηρέτηση μείον τα κόστη λειτουργίας του συστήματος και τα κόστη αναμονής των πελατών.

Όταν ο διαχειριστής ενός συστήματος δρα ως μονοπώλειο, υποθέτουμε ότι επιθυμεί να μεγιστοποιήσει τη δική του ωφέλεια (συνολικά έσοδα ή κέρδος) ανά χρονική μονάδα. Στην περίπτωση αυτή, η αντικειμενική συνάρτηση που βελτιστοποιεί ο διαχειριστής ισούται με τα έσοδα από τους πελάτες μέσω των τελών/τιμών που επιβάλλει ή με το κέρδος του που ισούται με τα έσοδα μείον τα κόστη λειτουργίας του συστήματος.

15.6 Στρατηγικές εισόδου στη μη-παρατηρήσιμη M/M/1 ουρά

Θεωρούμε ένα σύστημα εξυπηρέτησης τύπου M/M/1, όπου οι πελάτες φθάνουν σύμφωνα με μια διαδικασία Poisson ρυθμού λ , οι χρόνοι εξυπηρέτησης έχουν κατανομή $\text{Exp}(\mu)$, υπάρχει 1 υπηρέτης, άπειρη χωρητικότητα και η πειθαρχία ουράς είναι η FCFS. Για απλότητα στην παρουσίαση υποθέτουμε ότι $\lambda < \mu$ και επομένως το σύστημα είναι ευσταθές, ακόμα κι αν όλοι οι αφικνούμενοι πελάτες εισέρχονται σε αυτό. Κάθε πελάτης λαμβάνει μέση αμοιβή r για την εξυπηρέτησή του, ενώ υφίσταται κόστος αναμονής με ρυθμό c , όσο βρίσκεται στο σύστημα (το οποίο συσσωρεύεται είτε βρίσκεται στον χώρο αναμονής είτε εξυπηρετείται). Ο διαχειριστής του συστήματος επιβάλλει ένα τέλος εισόδου p (τιμή της παρεχόμενης εξυπηρέτησης) για κάθε πελάτη. Υποθέτουμε ότι $p \in [0, r)$.

Το δίλημμα που αντιμετωπίζουν οι πελάτες αφορά το αν θα εισέλθουν στο σύστημα ή όχι. Θεωρούμε ότι δεν μπορούν να παρατηρήσουν τον αριθμό των πελατών στο σύστημα πριν λάβουν την απόφασή τους, αλλά οι παράμετροι του συστήματος λ, μ, r, c, p είναι κοινή γνώση για όλους.

Οι καθαρές στρατηγικές ενός πελάτη είναι στην περίπτωση αυτή δύο: Είσοδος (απόφαση 1) ή αποχώρηση (απόφαση 0). Μια μεικτή στρατηγική προσδιορίζεται στην περίπτωση αυτή από έναν αριθμό $q \in [0, 1]$ που αντιστοιχεί στην πιθανότητα εισόδου.

Σκοπός μας είναι σε πρώτη φάση ο προσδιορισμός των στρατηγικών ισορροπίας των πελατών. Για τον σκοπό αυτό θα ακολουθήσουμε τη γενική διαδικασία που περιγράφηκε νωρίτερα.

Αρχικά, προσδιορίζουμε τη στάσιμη συμπεριφορά του συστήματος κάτω από μια στρατηγική των πελατών. Έστω ότι οι πελάτες ακολουθούν τη στρατηγική q . Τότε, το σύστημα συμπεριφέρεται ως μια M/M/1 ουρά με ρυθμό αφίξεων λq και ρυθμό εξυπηρέτησης μ . Τότε, ο μέσος χρόνος παραμονής ενός πελάτη που θα αποφασίσει να εισέλθει σε αυτό το σύστημα είναι $\frac{1}{\mu - \lambda q}$ (λόγω της (5.2)).

Έστω, τώρα, ένας επιλεγμένος πελάτης που ακολουθεί τη στρατηγική q' όταν οι υπόλοιποι ακολουθούν τη στρατηγική q . Τότε η ωφέλειά του είναι

$$\mathcal{U}(q', q) = (1 - q') \cdot 0 + q' \left(r - p - \frac{c}{\mu - \lambda q} \right).$$

Επομένως, ο επιλεγμένος πελάτης για να βρει βέλτιστη απάντηση στη στρατηγική q των άλλων έχει να λύσει το πρόβλημα

$$\max_{q' \in [0, 1]} \mathcal{U}(q', q).$$

Η συνάρτηση $\mathcal{U}(q', q)$ είναι γραμμική ως προς q' , οπότε για το σύνολο των βέλτιστων απαντήσεων έναντι της q , $BR(q)$, έχουμε

$$BR(q) = \begin{cases} \{0\}, & \text{αν } r - p - \frac{c}{\mu - \lambda q} < 0 \\ [0, 1], & \text{αν } r - p - \frac{c}{\mu - \lambda q} = 0 \\ \{1\}, & \text{αν } r - p - \frac{c}{\mu - \lambda q} > 0, \end{cases}$$

ή λύνοντας ως προς q (υπό την προϋπόθεση ότι $q \in [0, 1]$ και $\lambda q < \mu$)

$$BR(q) = \begin{cases} \{0\}, & \text{αν } q > \bar{q}_e \\ [0, 1], & \text{αν } q = \bar{q}_e \\ \{1\}, & \text{αν } q < \bar{q}_e, \end{cases}$$

με

$$\bar{q}_e = \frac{1}{\lambda} \left(\mu - \frac{c}{r-p} \right).$$

Μπορούμε να προχωρήσουμε, τώρα, στην εύρεση των στρατηγικών ισορροπίας για το συγκεκριμένο πρόβλημα. Ας θυμηθούμε ότι μια στρατηγική q είναι στρατηγική ισορροπίας, αν και μόνο αν $q \in BR(q)$. Επομένως, έχουμε

$$\begin{aligned} q_e = 0 & \text{ είναι στρατηγική ισορροπίας} \\ \Leftrightarrow 0 & \in BR(0) \\ \Leftrightarrow 0 & \geq \bar{q}_e \\ \Leftrightarrow r - p & \leq \frac{c}{\mu}. \end{aligned}$$

Επίσης,

$$\begin{aligned} q_e \in (0, 1) & \text{ είναι στρατηγική ισορροπίας} \\ \Leftrightarrow q_e & \in BR(q_e) \\ \Leftrightarrow q_e & = \bar{q}_e \\ \Leftrightarrow q_e & = \frac{1}{\lambda} \left(\mu - \frac{c}{r-p} \right), \end{aligned}$$

υπό την προϋπόθεση ότι $\bar{q}_e \in (0, 1)$ που ισχύει αν και μόνο αν

$$\frac{c}{\mu} < r - p < \frac{c}{\mu - \lambda}.$$

Τέλος, έχουμε

$$\begin{aligned} q_e = 1 & \text{ είναι στρατηγική ισορροπίας} \\ \Leftrightarrow 1 & \in BR(1) \\ \Leftrightarrow 1 & \leq \bar{q}_e \\ \Leftrightarrow r - p & \geq \frac{c}{\mu - \lambda}. \end{aligned}$$

Επομένως, αποδείξαμε το εξής αποτέλεσμα:

Θεώρημα 15.1 Στο πρόβλημα της εισόδου/αποχώρησης των πελατών στη μη-παρατηρήσιμη M/M/1 ουρά, υπάρχει πάντα στρατηγική ισορροπίας, $q_e = q_e(p)$, που είναι μοναδική και δίνεται από τον τύπο

$$q_e = q_e(p) = \begin{cases} 0, & r - p \leq \frac{c}{\mu}, \\ \frac{1}{\lambda} \left(\mu - \frac{c}{r-p} \right), & \frac{c}{\mu} < r - p < \frac{c}{\mu - \lambda}, \\ 1, & r - p \geq \frac{c}{\mu - \lambda}. \end{cases}$$

Προχωράμε, τώρα, στη λύση του προβλήματος της κοινωνικής βελτιστοποίησης. Ο διαχειριστής του συστήματος, όταν δρα ως κοινωνικός σχεδιαστής, θέλει να επιλέξει μια κοινωνικά βέλτιστη στρατηγική, q_{soc} που να μεγιστοποιεί τον κοινωνικό πλούτο ανά χρονική μονάδα. Αν ο διαχειριστής αφήνει να εισέλθει στο σύστημα κάθε πελάτης με πιθανότητα q , ανεξάρτητα από τους υπόλοιπους πελάτες, τότε η διαδικασία αφίξεων στο σύστημα είναι Poisson με ρυθμό λq και το σύστημα είναι μια M/M/1 ουρά με αυτόν τον ρυθμό των αφίξεων και ρυθμό εξυπηρέτησης μ . Αν επιβληθεί τιμή p για την είσοδο των πελατών, τότε η ωφέλεια του διαχειριστή ανά χρονική μονάδα θα είναι $\lambda q p$, ενώ η συνολική ωφέλεια των πελατών θα είναι $\lambda q \left(r - p - \frac{c}{\mu - \lambda q} \right)$. Επομένως, η συνολική κοινωνική ωφέλεια ανά χρονική μονάδα θα είναι

$$S_{soc}^{(un)}(q) = \lambda q \left(r - p - \frac{c}{\mu - \lambda q} \right) + \lambda q p = \lambda q \left(r - \frac{c}{\mu - \lambda q} \right),$$

που είναι ανεξάρτητη του p , όπως έχουμε πει παραπάνω (αφού στον κοινωνικό πλούτο δεν εμφανίζονται οι εσωτερικές πληρωμές). Ο «εκθέτης» (un) στη συνάρτηση $S_{soc}^{(un)}(q)$ δείχνει ότι αναφερόμαστε στο μη-παρατηρήσιμο (unobservable) μοντέλο. Είναι

$$\begin{aligned} \frac{d}{dq} S_{soc}^{(un)}(q) &= \lambda \left(r - \frac{c\mu}{(\mu - \lambda q)^2} \right), \\ \frac{d^2}{dq^2} S_{soc}^{(un)}(q) &= -\frac{2c\mu\lambda^2}{(\mu - \lambda q)^3} < 0. \end{aligned}$$

Επομένως, η $S_{soc}^{(un)}(q)$ είναι κοίλη για $q \in [0, \frac{\mu}{\lambda}]$ και μεγιστοποιείται στο \bar{q}_{soc} που μηδενίζει την $\frac{d}{dq} S_{soc}^{(un)}(q)$ και δίνεται από τον τύπο

$$\bar{q}_{soc} = \frac{1}{\lambda} \left(\mu - \sqrt{\frac{c\mu}{r}} \right). \quad (15.1)$$

Εφόσον $\bar{q}_{soc} \in [0, 1]$, συνάγουμε ότι η κοινωνικά βέλτιστη στρατηγική (δηλ. η πιθανότητα εισόδου) δίνεται από τον τύπο (15.1), διαφορετικά το μέγιστο της κοινωνικής ωφέλειας επιτυγχάνεται στο 0 ή στο 1. Συγκεκριμένα, έχουμε το παρακάτω αποτέλεσμα.

Θεώρημα 15.2 Στο πρόβλημα της εισόδου/αποχώρησης των πελατών στη μη-παρατηρήσιμη M/M/1 ουρά, η κοινωνικά βέλτιστη στρατηγική είναι μοναδική και δίνεται από τον τύπο

$$q_{soc} = \begin{cases} 0, & r \leq \frac{c}{\mu}, \\ \frac{1}{\lambda} \left(\mu - \sqrt{\frac{c\mu}{r}} \right), & \frac{c}{\mu} < r < \frac{c\mu}{(\mu - \lambda)^2}, \\ 1, & r \geq \frac{c\mu}{(\mu - \lambda)^2}. \end{cases}$$

Αν η τιμή εισόδου που επιβάλλει ο διαχειριστής είναι μηδενική ($p = 0$) τότε ισχύει ότι $q_e(0) \geq q_{soc}$ για όλες τις τιμές των παραμέτρων λ, μ, r και c . Επομένως, βλέπουμε ότι, χωρίς την επιβολή κάποιας τιμής εισόδου, οι πελάτες χρησιμοποιούν το σύστημα περισσότερο από όσο είναι κοινωνικά βέλτιστο.

Προχωράμε, τώρα, στο πρόβλημα της τιμολόγησης του μονοπωλίου. Στην περίπτωση αυτή, ο διαχειριστής δρα με σκοπό τη βελτιστοποίηση της συνολικής του ωφέλειας ανά χρονική μονάδα. Θέτοντας μια τιμή εισόδου, p , οι πελάτες υιοθετούν την αντίστοιχη στρατηγική ισορροπίας $q_e(p)$ που περιγράψαμε παραπάνω. Για να επάγει μια πιθανότητα εισόδου $q \in (0, 1)$ ο διαχειριστής πρέπει να επιβάλλει τιμή $p(q) = r - \frac{c}{\mu - \lambda q}$. Για $q = 1$, θα πρέπει να επιβάλλει τη μέγιστη δυνατή τιμή που είναι η $p(1) = r - \frac{c}{\mu - \lambda}$. Όταν επάγει πιθανότητα εισόδου $q = 0$, επιβάλλοντας μια μεγάλη τιμή το κέρδος του είναι 0. Βλέπουμε, λοιπόν, ότι σε κάθε περίπτωση η συνάρτηση του κέρδους ανά χρονική μονάδα είναι

$$S_{prof}^{(un)}(q) = \lambda q p(q) = \lambda q \left(r - \frac{c}{\mu - \lambda q} \right),$$

ταυτίζεται δηλαδή με τη συνάρτηση της συνολικής κοινωνικής ωφέλειας ανά χρονική μονάδα. Επομένως, η βέλτιστη πιθανότητα εισόδου για το διαχειριστή του συστήματος όταν δρα ως μονοπώλειο είναι η κοινωνικά βέλτιστη πιθανότητα εισόδου. Λόγω της ομοιογένειας των πελατών, οι αντικειμενικές συναρτήσεις της κοινωνικής ωφέλειας και του κέρδους του μονοπωλίου συμπίπτουν και ο διαχειριστής του συστήματος, επιβάλλοντας την τιμή

$$p_{prof} = r - \frac{c}{\mu - \lambda q_{soc}},$$

μεγιστοποιεί την κοινωνική ωφέλεια, αλλά ταυτόχρονα την καρπούται ολόκληρη, αφήνοντας μηδενικό περιθώριο καθαρής ωφέλειας στους πελάτες.

Αντικαθιστώντας την κοινωνικά βέλτιστη πιθανότητα εισόδου στον τύπο της p_{prof} , έχουμε ότι

$$p_{prof} = \begin{cases} r - \frac{c}{\mu - \lambda}, & \text{αν } \lambda < \mu - \sqrt{\frac{c\mu}{r}} \\ r - \sqrt{\frac{rc}{\mu}}, & \text{αν } \lambda \geq \mu - \sqrt{\frac{c\mu}{r}}, \end{cases}$$

που είναι φθίνουσα και τελικά σταθερή συνάρτηση του λ . Αυτό μοιάζει παράδοξο, αφού το λ μπορεί να ερμηνευθεί ως η ζήτηση της εξυπηρέτησης και επομένως αύξηση στη ζήτηση επάγει μείωση στην τιμή της! Όμως, αυτό μπορεί να αιτιολογηθεί διότι στη συγκεκριμένη περίπτωση η αύξηση στη ζήτηση επάγει μείωση στην ποιότητα του αγαθού της εξυπηρέτησης που οφείλεται στην αύξηση του μέσου χρόνου αναμονής, και επομένως οι πελάτες γίνονται λιγότερο πρόθυμοι να το αγοράσουν.

15.7 Στρατηγικές εισόδου στην παρατηρήσιμη M/M/1 ουρά

Θεωρούμε ένα σύστημα εξυπηρέτησης M/M/1 με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ , όπου κάθε πελάτης αντιμετωπίζει το δίλημμα εισόδου/αποχώρησης, αφού παρατηρήσει το πλήθος των πελατών που υπάρχει στο σύστημα, και γνωρίζοντας ότι η μέση αμοιβή από την εξυπηρέτησή του είναι r , ενώ συσσωρεύει κόστος αναμονής με ρυθμό c ανά χρονική μονάδα.

Ο διαχειριστής του συστήματος επιβάλλει ένα τέλος εισόδου p , το οποίο είναι επίσης γνωστό σε κάθε πελάτη.

Μια μεικτή στρατηγική ενός πελάτη στην περίπτωση αυτή δίνεται από μια ακολουθία $\mathbf{q} = (q_0, q_1, q_2, \dots)$, όπου $q_n \in [0, 1]$, $n = 0, 1, 2, \dots$, είναι η πιθανότητα εισόδου στο σύστημα όταν ένας αφικνούμενος πελάτης βρίσκει n άτομα στο σύστημα (χωρίς να μετρά τον εαυτό του).

Αρχικά, θέλουμε να προσδιορίσουμε τις στρατηγικές ισορροπίας των πελατών. Όταν οι πελάτες ακολουθούν μια στρατηγική \mathbf{q} το σύστημα συμπεριφέρεται ως μια M/M/1 ουρά με μεταβλητό ρυθμό αφίξεων $\lambda_n = \lambda q_n$ και οποιοδήποτε μέτρο απόδοσης μπορεί να υπολογιστεί εύκολα με βάση τη σχετική μεθοδολογία (βλέπε κεφάλαιο 6). Το σημαντικό, όμως, είναι ότι ο μέσος χρόνος παραμονής ενός επιλεγμένου πελάτη στο σύστημα, δεδομένου ότι βρίσκει n πελάτες μπροστά του, δεν εξαρτάται από την στρατηγική \mathbf{q} του πληθυσμού των πελατών. Αυτό συμβαίνει διότι οι στρατηγικές των μελλοντικών αφίξεων δεν επηρεάζουν τον επιλεγμένο πελάτη λόγω της FCFS πειθαρχίας ουράς, αλλά ούτε και οι στρατηγικές των προηγούμενων πελατών επηρεάζουν τον επιλεγμένο πελάτη. Πράγματι, οι στρατηγικές των προηγούμενων πελατών δεν βασίστηκαν σε γνώση του χρόνου εξυπηρέτησής τους, οπότε το γεγονός ότι βρίσκονται n πελάτες στην ουρά δεν προσφέρει κάποια πληροφορία για τους χρόνους εξυπηρέτησης που δεν έχουν ακόμη ξεκινήσει. Όσον αφορά τον υπολειπόμενο χρόνο του χρόνου εξυπηρέτησης που βρίσκεται σε εξέλιξη, αυτός είναι σαν να ξεκινάει εκείνη τη στιγμή, λόγω της αμνήμονης ιδιότητας της εκθετικής κατανομής. Επομένως, παρότι ο αριθμός των πελατών δεν είναι ανεξάρτητος με τον παρελθόντα χρόνο της τρέχουσας εξυπηρέτησης, είναι ανεξάρτητος με τον υπολειπόμενο. Έτσι, ο χρόνος παραμονής ενός επιλεγμένου πελάτη που βρίσκει n άτομα κατά την άφιξή του και αποφασίζει να εισέλθει στο σύστημα ισούται με το άθροισμα $n + 1$ ανεξάρτητων και ισόνομων τυχαίων μεταβλητών με κατανομή $\text{Exp}(\mu)$. Επομένως, ο μέσος χρόνος παραμονής είναι $\frac{n+1}{\mu}$.

Ας θεωρήσουμε ότι ένας επιλεγμένος πελάτης ακολουθεί τη στρατηγική $\mathbf{q}' = (q'_0, q'_1, q'_2, \dots)$, όταν οι υπόλοιποι ακολουθούν τη στρατηγική $\mathbf{q} = (q_0, q_1, q_2, \dots)$, και βλέπει n πελάτες, καθώς φθάνει στο σύστημα. Τότε, η ωφέλειά του είναι

$$\mathcal{U}(\mathbf{q}', \mathbf{q}|n) = (1 - q'_n) \cdot 0 + q'_n \left(r - p - \frac{c(n+1)}{\mu} \right).$$

Επομένως, το σύνολο των βέλτιστων απαντήσεων $BR(\mathbf{q}|n)$ ενός επιλεγμένου πελάτη έναντι της στρατηγικής \mathbf{q} των άλλων, όταν βρίσκει n πελάτες στο σύστημα είναι

$$BR(\mathbf{q}|n) = \begin{cases} \{0\}, & \text{αν } r - p - \frac{c(n+1)}{\mu} < 0 \\ [0, 1], & \text{αν } r - p - \frac{c(n+1)}{\mu} = 0 \\ \{1\}, & \text{αν } r - p - \frac{c(n+1)}{\mu} > 0, \end{cases}$$

ή λύνοντας ως προς n

$$BR(\mathbf{q}|n) = \begin{cases} \{0\}, & \text{αν } \frac{\mu(r-p)}{c} - 1 < n \\ [0, 1], & \text{αν } \frac{\mu(r-p)}{c} - 1 = n \\ \{1\}, & \text{αν } \frac{\mu(r-p)}{c} - 1 > n. \end{cases}$$

Επομένως, έχουμε αποδείξει το εξής αποτέλεσμα.

Θεώρημα 15.3 Στο πρόβλημα εισόδου/αποχώρησης των πελατών στην παρατηρήσιμη M/M/1 ουρά, η στρατηγική κατωφλίου n_e με

$$n_e = n_e(p) = \left\lfloor \frac{\mu(r-p)}{c} \right\rfloor,$$

σύμφωνα με την οποία ένας πελάτης εισέρχεται στο σύστημα εφόσον το πλήθος των πελατών στο σύστημα συμπεριλαμβανομένου του ίδιου είναι το πολύ n_e είναι κυριαρχούσα στρατηγική (και επομένως και στρατηγική ισορροπίας).

Ο διαχειριστής του συστήματος, όταν δρα ως κοινωνικός σχεδιαστής, επιδιώκει να επιβάλλει μια κοινωνικά βέλτιστη στρατηγική \mathbf{q}_{soc} , που να μεγιστοποιεί τη συνολική κοινωνική ωφέλεια ανά χρονική μονάδα, δηλαδή την ποσότητα

$$S_{soc}^{(obs)}(\mathbf{q}) = S_{soc}^{(obs)}(q_0, q_1, \dots) = \lambda^*(\mathbf{q})r - cE_{\mathbf{q}}[Q], \quad (15.2)$$

όπου $\lambda^*(\mathbf{q})$ ο ρυθμός διαπέρασης του συστήματος, υπό τη στρατηγική \mathbf{q} , και $E_{\mathbf{q}}[Q]$ ο μέσος αριθμός πελατών σε κατάσταση ισορροπίας.

Κάτω από μια στρατηγική εισόδου $\mathbf{q} = (q_0, q_1, q_2, \dots)$, το σύστημα συμπεριφέρεται ως μια M/M/1 ουρά με μεταβλητό ρυθμό αφίξεων $\lambda_n = \lambda q_n$ και σταθερό ρυθμό εξυπηρέτησης $\mu_n = \mu$. Η στάσιμη κατανομή του αριθμού των πελατών θα είναι τότε (p_n) , όπου

$$p_n = \begin{cases} B, & \text{αν } n = 0 \\ B\rho^n q_0 q_1 \cdots q_{n-1}, & \text{αν } n \geq 1, \end{cases}$$

όπου $\rho = \lambda/\mu$ και

$$B = \left(1 + \sum_{n=1}^{\infty} \rho^n q_0 q_1 \cdots q_{n-1} \right)^{-1}.$$

Η λύση του προβλήματος αυτού γίνεται τότε με στοχαστικό δυναμικό προγραμματισμό, οπότε και αποδεικνύεται ότι η βέλτιστη στρατηγική είναι τύπου κατωφλίου. Εμείς θα επικεντρωθούμε στο πρόβλημα της κοινωνικής βελτιστοποίησης που μπορεί να επιτύχει ο κοινωνικός σχεδιαστής, επιβάλλοντας ένα τέλος (τιμή) εισόδου που είναι κοινό για όλους τους πελάτες.

Ας υποθέσουμε ότι ο διαχειριστής επιβάλλει τέλος εισόδου p . Τότε οι πελάτες υιοθετούν τη στρατηγική κατωφλίου $n_e(p) = \left\lfloor \frac{\mu(r-p)}{c} \right\rfloor$. Το σύστημα συμπεριφέρεται τότε ως μια M/M/1/ $n_e(p)$ ουρά. Ας υποθέσουμε, για συντομία, ότι $\rho \neq 1$. Τότε, ισχύουν τα αποτελέσματα που έχουμε συναγάγει στην παράγραφο 6.5.1, με $k = n_e(p)$. Χρησιμοποιώντας τους τύπους (6.13) και (6.14) για τον ρυθμό διαπέρασης και το μέσο μήκος ουράς, αντίστοιχα, και αντικαθιστώντας στην (15.2) έχουμε ότι η κοινωνική ωφέλεια, όταν επιβληθεί η στρατηγική εισόδου κατωφλίου με κατώφλι n , μέσω κατάλληλης τιμής είναι

$$S_{soc}^{(obs)}(n) = \lambda r \frac{1 - \rho^n}{1 - \rho^{n+1}} - c \left[\frac{\rho}{1 - \rho} - \frac{(n+1)\rho^{n+1}}{1 - \rho^{n+1}} \right].$$

Με απλές πράξεις, είναι εύκολο να δούμε ότι

$$S_{soc}^{(obs)}(n) - S_{soc}^{(obs)}(n-1) = \frac{\lambda r(1-\rho)^2 \rho^{n-1}}{(1-\rho^{n+1})(1-\rho^n)} + \frac{c((n+1)\rho - \rho^{n+1} - n\rho^n)}{(1-\rho^{n+1})(1-\rho^n)}.$$

Για $\rho < 1$ έχουμε

$$\begin{aligned} S_{soc}^{(obs)}(n) - S_{soc}^{(obs)}(n-1) \geq 0 &\Leftrightarrow \lambda r(1-\rho)^2 \geq c\rho(n + \rho^{n+1} - (n+1)\rho) \\ &\Leftrightarrow \frac{r\mu}{c} \geq \frac{n + \rho^{n+1} - (n+1)\rho}{(1-\rho)^2}. \end{aligned}$$

Έστω $g(n)$ η ποσότητα στο δεξιό μέλος της τελευταίας ανισότητας. Τότε

$$\begin{aligned} g(n) &= \frac{n + \rho^{n+1} - (n+1)\rho}{(1-\rho)^2} \\ &= \frac{1}{(1-\rho)^2} (n(1-\rho) - \rho(1-\rho^n)) \\ &= \frac{1}{1-\rho} \left(n - \sum_{k=1}^n \rho^k \right) \\ &= \frac{1}{1-\rho} \sum_{k=1}^n (1-\rho^k), \end{aligned} \tag{15.3}$$

η οποία είναι προφανώς γνησίως αύξουσα ως προς n . Επομένως, υπάρχει μοναδικό n_{soc} τέτοιο ώστε $g(n) \leq \frac{r\mu}{c}$, για $n \leq n_{soc}$, ενώ $g(n) > \frac{r\mu}{c}$, για $n > n_{soc}$. Οπότε, $S_{soc}^{(obs)}(n) - S_{soc}^{(obs)}(n-1) \geq 0$ για $n \leq n_{soc}$, ενώ $S_{soc}^{(obs)}(n) - S_{soc}^{(obs)}(n-1) < 0$ για $n > n_{soc}$. Επομένως, η $S_{soc}^{(obs)}(n)$ είναι μονοκόρυφη με μέγιστο στην n_{soc} . Έχουμε, λοιπόν, αποδείξει το παρακάτω αποτέλεσμα:

Θεώρημα 15.4 Στο πρόβλημα της εισόδου/αποχώρησης των πελατών στην παρατηρήσιμη M/M/1 ουρά, η κοινωνικά βέλτιστη στρατηγική κατωφλίου έχει κατώφλι n_{soc} που δίνεται από τον τύπο

$$n_{soc} = \max \left\{ n : g(n) \leq \frac{r\mu}{c} \right\},$$

όπου η $g(n)$ δίνεται από την (15.3). Το κατώφλι αυτό επάγεται από τον διαχειριστή του συστήματος, θέτοντας τέλος εισόδου p_{soc} ώστε $n_{soc} = \left\lfloor \frac{\mu(r-p_{soc})}{c} \right\rfloor$, δηλαδή οποιοδήποτε $p_{soc} \in \left(r - \frac{c(n_{soc}-1)}{\mu}, r - \frac{cn_{soc}}{\mu} \right]$.

Επιπλέον ισχύει ότι $n_{soc} \leq n_e(0)$, δηλαδή, αν δεν επιβληθεί τέλος εισόδου, τότε το ατομικά βέλτιστο κατώφλι εισόδου για τους πελάτες είναι μεγαλύτερο ή ίσο από το κοινωνικά βέλτιστο. Πράγματι, έχουμε

$$\begin{aligned} g(n) - n &= \frac{1}{1-\rho} \sum_{k=1}^n (1-\rho^k) - \frac{1}{1-\rho} n(1-\rho) \\ &= \frac{\rho}{1-\rho} \sum_{k=1}^n (1-\rho^{k-1}) \geq 0, \end{aligned}$$

οπότε $g(n_{soc}) \geq n_{soc}$. Όμως, $g(n_{soc}) \leq \frac{r\mu}{c}$, από τον ορισμό του n_{soc} , οπότε $n_{soc} \leq \frac{r\mu}{c}$ που τελικά δίνει $n_{soc} \leq \lfloor \frac{r\mu}{c} \rfloor = n_e(0)$.

Βλέπουμε, επομένως, ότι χωρίς την επιβολή κάποιου τέλους εισόδου, οι πελάτες είναι πρόθυμοι να μπαίνουν πιο εύκολα απ' ό,τι είναι κοινωνικά βέλτιστο. Αυτό γίνεται διότι αγνοούν τις «εξωτερικότητες» (αρνητικές επιδράσεις) που επάγουν με την εισοδό τους στους μελλοντικούς πελάτες. Το ίδιο φαινόμενο παρατηρήθηκε και στο αντίστοιχο πρόβλημα της μη-παρατηρήσιμης M/M/1 ουράς.

Περνάμε τώρα στο πρόβλημα του μονοπωλίου. Στην περίπτωση αυτή, θεωρούμε ότι ο διαχειριστής επιβάλλει κάποιο τέλος εισόδου με σκοπό τη βελτιστοποίηση της δικής του ωφέλειας. Αν επιβληθεί τέλος p , τότε οι πελάτες υιοθετούν την αντίστοιχη στρατηγική κατωφλίου $n_e(p)$ και το κέρδος γίνεται $\lambda^* p$. Για να βρούμε ποίο είναι το κατώφλι n_{prof} που βελτιστοποιεί το κέρδος, εκφράζουμε το κέρδος συναρτήσει του επιβληθέντος κατωφλίου n . Για να υιοθετήσουν κατώφλι n οι πελάτες θα πρέπει να τεθεί τέλος εισόδου p τέτοιο ώστε $\lfloor \frac{(r-p)\mu}{c} \rfloor = n$. Στο πρόβλημα του μονοπωλίου θα πρέπει να τεθεί η μέγιστη τιμή που επάγει το συγκεκριμένο κατώφλι, επομένως η $p = r - \frac{cn}{\mu}$. Τότε το κέρδος του μονοπωλίου είναι:

$$\begin{aligned} S_{prof}^{(obs)}(n) &= \lambda \frac{1-\rho^n}{1-\rho^{n+1}} \left(r - \frac{cn}{\mu} \right) \\ &= \lambda r \frac{1-\rho^n}{1-\rho^{n+1}} \left(1 - \frac{n}{v_e} \right), \end{aligned}$$

όπου $v_e = \frac{r\mu}{c}$. Η σχέση αυτή δείχνει ότι για $n > n_e(0) = \lfloor \frac{r\mu}{c} \rfloor$ ισχύει $S_{prof}(n) < 0$, αφού ο διαχειριστής πρέπει να θέσει αρνητικό τέλος εισόδου (δηλαδή πρέπει να επιδοτήσει την είσοδο) για να επάγει κατώφλι μεγαλύτερο από το $n_e(0)$. Επομένως, είναι βέβαιο ότι για το κατώφλι n_{prof} της βελτιστοποίησης του κέρδους του μονοπωλίου ισχύει $n_{prof} \leq n_e(0)$.

Για να δούμε πού βελτιστοποιείται η συνάρτηση $S_{prof}^{(obs)}(n)$, θεωρούμε τον λόγο $S_{prof}^{(obs)}(n)/S_{prof}^{(obs)}(n-1)$. Έχουμε

$$\frac{S_{prof}^{(obs)}(n)}{S_{prof}^{(obs)}(n-1)} = \frac{(1-\rho^n)^2(v_e-n)}{(1-\rho^{n+1})(1-\rho^{n-1})(v_e-n+1)}.$$

Για $\rho < 1$ έχουμε

$$\begin{aligned} \frac{S_{prof}^{(obs)}(n)}{S_{prof}^{(obs)}(n-1)} \geq 1 &\Leftrightarrow \frac{1-\rho^n}{1-\rho^{n+1}}(v_e-n) \geq \frac{1-\rho^{n-1}}{1-\rho^n}(v_e-n+1) \\ &\Leftrightarrow \frac{(1-\rho^n)^2 - (1-\rho^{n-1})(1-\rho^{n+1})}{(1-\rho^{n+1})(1-\rho^n)}(v_e-n) \geq \frac{1-\rho^{n-1}}{1-\rho^n} \\ &\Leftrightarrow v_e-n \geq \frac{(1-\rho^{n-1})(1-\rho^{n+1})}{\rho^{n-1}(1-\rho)^2} \\ &\Leftrightarrow \frac{r\mu}{c} \geq n + \frac{(1-\rho^{n-1})(1-\rho^{n+1})}{\rho^{n-1}(1-\rho)^2}. \end{aligned}$$

Η συνάρτηση $h(n)$ με

$$h(n) = n + \frac{(1 - \rho^{n-1})(1 - \rho^{n+1})}{\rho^{n-1}(1 - \rho)^2} \quad (15.4)$$

αποδεικνύεται ότι είναι αύξουσα ως προς n , οπότε υπάρχει μοναδικό n_{prof} τέτοιο ώστε $h(n) \leq \frac{r\mu}{c}$, για $n \leq n_{prof}$, ενώ $h(n) > \frac{r\mu}{c}$, για $n > n_{prof}$. Οπότε, η $S_{prof}^{(obs)}(n)$ είναι μονοκόρυφη με μέγιστο στην n_{prof} . Έτσι, έχουμε το παρακάτω αποτέλεσμα.

Θεώρημα 15.5 Στο πρόβλημα της εισόδου/αποχώρησης των πελατών στην παρατηρήσιμη M/M/1 ουρά, η βέλτιστη στρατηγική κατώφλιου για το μονοπώλιο έχει κατώφλι n_{prof} που δίνεται από τον τύπο

$$n_{prof} = \max \left\{ n : h(n) \leq \frac{r\mu}{c} \right\},$$

όπου η $h(n)$ δίνεται από την (15.4). Το κατώφλι αυτό επάγεται από το διαχειριστή του συστήματος, θέτοντας τέλος εισόδου $p_{prof} = r - \frac{cn_{prof}}{\mu}$.

Αποδεικνύεται ότι $n_{prof} \leq n_{soc} \leq n_e(0)$, η οποία σχέση αναφέρεται ως ανισότητα του Naor.

15.8 Πληροφορία και στρατηγική συμπεριφορά

Με απλούς όρους, το πρόβλημα του ελέγχου της πληροφορίας που παρέχεται στους πελάτες ενός συστήματος εξυπηρέτησης μπορεί να διατυπωθεί ως εξής:

Τι είδους πληροφορία πρέπει να παρέχουμε στους πελάτες ενός συστήματος εξυπηρέτησης ώστε να επιτυγχάνεται ο μέγιστος ρυθμός διαπέρασης, ή η μέγιστη κοινωνική ωφέλεια ή το μέγιστο κέρδος που παράγεται από αυτό;

Σε προβλήματα βέλτιστου σχεδιασμού και ελέγχου που υπάρχει μοναδικός αποφασίζων, η παροχή περισσότερης πληροφορίας είναι πάντα ευνοϊκή, αλλά αυτό δεν συμβαίνει στην περίπτωση αλληλεπίδρασης περισσότερων του ενός αποφασίζόντων στο πλαίσιο μια παιγνιοθεωρητικής ανάλυσης ενός συστήματος εξυπηρέτησης. Π.χ., η παροχή περισσότερης πληροφορίας σε κάποιον πελάτη μπορεί να επηρεάζει θετικά τη δική του ωφέλεια, αλλά να μειώνει τη συνολική κοινωνική ωφέλεια. Αυτό συμβαίνει γιατί οι πελάτες κατά τη λήψη των αποφάσεών τους επικεντρώνονται στη δική τους ωφέλεια και αγνοούν τις επιδράσεις-εξωτερικότητες (externalities) που επάγουν στους άλλους πελάτες.

Η πιο σημαντική πληροφορία σε ένα σύστημα εξυπηρέτησης αφορά το πλήθος των πελατών σε αυτό. Επομένως, οι ακραίες εκδοχές στα συστήματα εξυπηρέτησης υπό τη στρατηγική οπτική είναι τα παρατηρήσιμα συστήματα (όπου οι πελάτες παίρνουν τις αποφάσεις τους αφού παρατηρήσουν τον αριθμό των πελατών στο σύστημα) και τα μη-παρατηρήσιμα συστήματα (όπου οι αποφάσεις λαμβάνονται με βάση μόνο τις παραμέτρους του συστήματος). Πράγματι, όπως είδαμε, οι πρώτες μελέτες που υιοθέτησαν τη στρατηγική οπτική αφορούσαν αυτές τις δυο ακραίες εκδοχές: Ο Naor 1969 μελέτησε την παρατηρήσιμη M/M/1 ουρά, ενώ οι Edelson και Hildebrand 1975 θεώρησαν το μη-παρατηρήσιμο ανάλογο μοντέλο. Οι Hassin 1986 και Chen και Frank 2004 συνέκριναν αυτές τις δυο εκδοχές και έδειξαν ότι σε κάποιες περιπτώσεις είναι καλύτερο να αποκαλύπτεται το πλήθος των πελατών στο σύστημα και σε κάποιες να αποκρύπτεται. Επομένως, δεν υπάρχει «καθαρός» νικητής-μοντέλο ως προς την πληροφορία, αλλά έχει σημασία να βλέπουμε κάθε φορά προσεκτικά τις παραμέτρους. Για τον λόγο αυτό έχουν προταθεί πολλές διαφορετικές κλάσεις μοντέλων που γεφυρώνουν την παρατηρήσιμη με τη μη-παρατηρήσιμη εκδοχή:

- Συστήματα με ατελή (imperfect) δομή παρατήρησης, όπου οι πελάτες λαμβάνουν μόνο μερική πληροφορία σχετικά με το πλήθος των παρόντων πελατών (π.χ., κατά πόσο υπερβαίνει ένα κρίσιμο κατώφλι ή όχι). Δείτε, π.χ., τις εργασίες των Economou και Kanta 2008, Guo και Zipkin 2009 και Hassin και Koshman 2017.

- Συστήματα με καθυστερημένη (delayed) δομή παρατήρησης, όπου οι πελάτες παρατηρούν το πλήθος των πελατών με κάποια καθυστέρηση. Δείτε, π.χ., τις εργασίες των Burnetas, Economou και Vasiliadis 2017, και Hassin και Roet-Green 2020.
- Συστήματα με μικτή (mixed) δομή παρατήρησης, όπου μόνο ένα ποσοστό των πελατών μπορούν να παρατηρήσουν το πλήθος των πελατών. Δείτε, π.χ., την εργασία των Hu, Li και Wang 2018.
- Συστήματα με εναλλασσόμενη (alternating) δομή παρατήρησης, όπου το σύστημα εναλλάσσεται μεταξύ παρατηρήσιμων και μη-παρατηρήσιμων περιόδων. Δείτε, π.χ., την εργασία των Dimitrakopoulos, Economou και Leonardos 2021.
- Συστήματα με μη-τυπική (non-standard) δομή παρατήρησης, όπου οι πελάτες παρατηρούν άλλα χαρακτηριστικά του συστήματος αντί για (ή επιπρόσθετα με) το πλήθος πελατών. Δείτε, π.χ., τις εργασίες των Burnetas και Economou 2007, Economou και Manou 2013, και Haviv 2014.

Αυτό που είναι ενδιαφέρον στα αποτελέσματα που αφορούν τέτοια «ενδιάμεσα μοντέλα» είναι ότι το βέλτιστο επίπεδο πληροφορίας δεν αντιστοιχεί σε κάποια από τις δυο ακραίες περιπτώσεις (παρατηρήσιμη και μη-παρατηρήσιμη).

Εκτός από το πρακτικό τους ενδιαφέρον, τα προβλήματα σχετικά με το επίπεδο πληροφόρησης που πρέπει να επιλεγεί σε ένα συγκεκριμένο σύστημα θέτουν δύσκολα θεωρητικά ερωτήματα στον ερευνητή, καθώς η μελέτη ενός συστήματος απαιτεί την εκτίμηση της απόδοσής του κάτω από όλες τις δυνατές στρατηγικές των πελατών.

15.9 Ασκήσεις

Άσκηση 15.1 Θεωρούμε επιβάτες που φθάνουν σε έναν σταθμό μεταφορικών μέσων ενός τερματικού σταθμού αεροδρομίου σύμφωνα με μια διαδικασία Poisson με ρυθμό λ και οι οποίοι αντιμετωπίζουν το πρόβλημα επιλογής μεταφορικού μέσου. Οι επιλογές που έχουν είναι δύο: Είτε να χρησιμοποιήσουν μικρά λεωφορεία (shuttles) είτε τρένα. Τα τρένα φθάνουν σύμφωνα με μια ανανεωτική διαδικασία με κατανομή ενδιάμεσων χρόνων $F_B(x)$ και έχουν απεριόριστη χωρητικότητα. Τα μικρά λεωφορεία αναχωρούν το ένα πίσω από το άλλο, μόλις συμπληρωθούν όλες οι θέσεις τους. Καθένα από αυτά έχει χωρητικότητα N . Η μετακίνηση και με τα δύο μέσα είναι δωρεάν. Σκοπός των επιβατών είναι η ελαχιστοποίηση του μέσου χρόνου αναμονής τους στο σταθμό.

1. Να προσδιοριστούν οι στρατηγικές ισορροπίας των πελατών στη μη-παρατηρήσιμη περίπτωση, όπου οι επιβάτες αποφασίζουν σε ποιο μέσο θα επιβιβαστούν βασιζόμενοι μόνο στη γνώση των παραμέτρων του συστήματος.
2. Να προσδιοριστούν οι στρατηγικές ισορροπίας των πελατών στην παρατηρήσιμη περίπτωση, όπου οι επιβάτες αποφασίζουν σε ποιο μέσο θα επιβιβαστούν αφού παρατηρήσουν τον αριθμό επιβατών σε αναμονή στα μικρά λεωφορεία.
3. Να λυθεί το πρόβλημα της κοινωνικής βελτιστοποίησης στην παρατηρήσιμη και στη μη παρατηρήσιμη περίπτωση.

Άσκηση 15.2 Θεωρούμε ένα σύστημα εξυπηρέτησης τύπου $M/M/1$, όπου οι πελάτες φθάνουν σύμφωνα με μια διαδικασία Poisson με ρυθμό λ , οι χρόνοι εξυπηρέτησης έχουν την $\text{Exp}(\mu)$ κατανομή, υπάρχει 1 υπηρέτης και απεριόριστος χώρος αναμονής. Υποθέτουμε ότι οι πελάτες, κατά την άφιξή τους και χωρίς να παρατηρήσουν τον αριθμό των πελατών που υπάρχει στο σύστημα μπορούν να αγοράσουν ένα κουπόνι προτεραιότητας σε τιμή p ή να μην το αγοράσουν. Επιπλέον γνωρίζουν ότι επιβαρύνονται με κόστος με ρυθμό c όσο βρίσκονται στο σύστημα. Ο υπηρέτης εξυπηρετεί με FCFS τους πελάτες που έχουν αγοράσει το κουπόνι προτεραιότητας και αν δεν υπάρχουν τέτοιοι πελάτες τότε εξυπηρετεί αυτούς που δεν έχουν αγοράσει το κουπόνι, και πάλι με FCFS. Στην περίπτωση

που εξυπηρετείται ένας πελάτης που δεν έχει αγοράσει το κουπόνι και αφιχθεί πελάτης που αγοράζει το κουπόνι, η εξυπηρέτηση του πελάτη θα διακοπεί ώστε ο υπηρέτης να εξυπηρετήσει τον πελάτη με κουπόνι προτεραιότητας. Όλοι οι πελάτες που φθάνουν στο σύστημα εισέρχονται σε αυτό και δεν (υπ)αναχωρούν μέχρι να εξυπηρετηθούν. Η στρατηγική ενός πελάτη είναι η πιθανότητα αγοράς κουπονιού προτεραιότητας.

1. Να βρεθεί η βέλτιστη απάντηση ενός επιλεγμένου πελάτη έναντι μιας στρατηγικής των υπολοίπων και να εξεταστεί αν έχουμε κατάσταση FTC ή ATC.
2. Να βρεθούν οι στρατηγικές ισορροπίας.
3. Να βρεθούν οι κοινωνικά βέλτιστες στρατηγικές.

Άσκηση 15.3 Θεωρούμε ένα σύστημα εξυπηρέτησης το οποίο διαθέτει έναν υπηρέτη και άπειρη χωρητικότητα. Η διαδικασία αφίξεων είναι Poisson με παράμετρο λ , ενώ οι χρόνοι εξυπηρέτησης έχουν την κατανομή $\text{Exp}(\mu)$. Κάθε φορά που το σύστημα αδειάζει ο υπηρέτης απενεργοποιείται. Με την άφιξη ενός πελάτη σε κενό σύστημα, ο υπηρέτης μπαίνει σε διαδικασία ενεργοποίησης. Ο χρόνος που χρειάζεται να ενεργοποιηθεί (οπότε και θα αρχίσει να παρέχει κανονικά εξυπηρέτηση) ακολουθεί την κατανομή $\text{Exp}(\theta)$. Κατά τη διάρκεια του χρόνου αυτού οι αφίξεις συνεχίζονται κανονικά. Ας υποθέσουμε ότι κάθε πελάτης προσδοκά ωφέλεια r αν εξυπηρετηθεί από το σύστημα και υφίσταται κόστος με ρυθμό c για κάθε χρονική μονάδα παραμονής του στο σύστημα. Το δίλημά του είναι κατά πόσο θα πρέπει να εισέλθει στο σύστημα ή όχι. Υποθέτουμε ότι όποιος πελάτης μπει στο σύστημα δεν έχει τη δυνατότητα να αναχωρήσει πριν την εξυπηρέτησή του.

1. Να βρείτε τις στρατηγικές ισορροπίας για την περίπτωση μη-παρατηρήσιμου συστήματος.
2. Να βρείτε τις στρατηγικές ισορροπίας για την περίπτωση παρατηρήσιμου συστήματος. Στην περίπτωση παρατηρήσιμου συστήματος, θεωρούμε ότι ο πελάτης λαμβάνει γνώση κατά την άφιξή του τόσο για τον αριθμό των παρόντων πελατών, όσο και για την κατάσταση του υπηρέτη (ανενεργός/ενεργοποιημένος).

Άσκηση 15.4 Θεωρούμε ένα σύστημα εξυπηρέτησης το οποίο διαθέτει έναν υπηρέτη και άπειρη χωρητικότητα. Η διαδικασία αφίξεων είναι Poisson με ρυθμό λ , ενώ οι χρόνοι εξυπηρέτησης ακολουθούν την κατανομή $\text{Exp}(\mu)$. Ο υπηρέτης εναλλάσσεται μεταξύ περιόδων λειτουργίας και αργίας. Οι χρόνοι λειτουργίας έχουν την κατανομή $\text{Exp}(\zeta)$, ενώ οι χρόνοι αργίας την κατανομή $\text{Exp}(\theta)$. Οι αφίξεις στο σύστημα γίνονται κανονικά, ανεξάρτητα αν ο υπηρέτης είναι σε περίοδο λειτουργίας ή αργίας. Όμως, εξυπηρέτηση παρέχεται μόνο όταν ο υπηρέτης βρίσκεται σε περίοδο λειτουργίας. Ας υποθέσουμε ότι κάθε πελάτης προσδοκά ωφέλεια r αν εξυπηρετηθεί από το σύστημα και υφίσταται κόστος με ρυθμό c για κάθε χρονική μονάδα παραμονής του στο σύστημα. Το δίλημά του είναι κατά πόσο θα πρέπει να εισέλθει στο σύστημα ή όχι. Υποθέτουμε ότι όποιος πελάτης μπει στο σύστημα δεν έχει τη δυνατότητα να αναχωρήσει πριν την εξυπηρέτησή του.

1. Να βρείτε τις στρατηγικές ισορροπίας για την περίπτωση μη-παρατηρήσιμου συστήματος.
2. Να βρείτε τις στρατηγικές ισορροπίας για την περίπτωση παρατηρήσιμου συστήματος. Στην περίπτωση αυτή, θεωρούμε ότι ο πελάτης λαμβάνει γνώση κατά την άφιξή του τόσο για τον αριθμό των παρόντων πελατών, όσο και για την κατάσταση του υπηρέτη (ανενεργός/ενεργοποιημένος).

Άσκηση 15.5 Θεωρούμε μια $M/M/1$ ουρά με ρυθμό αφίξεων λ και $\text{Exp}(\mu)$ κατανομή χρόνων εξυπηρέτησης, όπου μόλις το σύστημα αδειάζει ο υπηρέτης παύει να παρέχει εξυπηρέτηση και αρχίζει και πάλι να εξυπηρετεί, μόλις μαζευτούν K πελάτες με $K \geq 1$ (για $K = 1$ έχουμε την κλασική $M/M/1$ ουρά). Ας υποθέσουμε ότι κάθε πελάτης προσδοκά ωφέλεια r αν εξυπηρετηθεί από το σύστημα και υφίσταται κόστος με ρυθμό c για κάθε χρονική μονάδα παραμονής του στο σύστημα. Το δίλημά του είναι κατά πόσο θα πρέπει να εισέλθει στο σύστημα ή όχι. Υποθέτουμε ότι όποιος πελάτης μπει στο σύστημα δεν έχει τη δυνατότητα να αναχωρήσει πριν την εξυπηρέτησή του.

1. Να βρείτε τις στρατηγικές ισορροπίας για την περίπτωση μη-παρατηρήσιμου συστήματος.
2. Να βρείτε τις στρατηγικές ισορροπίας για την περίπτωση παρατηρήσιμου συστήματος. Στην περίπτωση αυτή, θεωρούμε ότι ο πελάτης λαμβάνει γνώση κατά την άφιξή του τόσο για τον αριθμό των παρόντων πελατών, όσο και για την κατάσταση του υπηρέτη (ανενεργός/ενεργοποιημένος).

15.10 Σχόλια

Η στρατηγική οπτική στη Θεωρία Ουρών Αναμονής εισήχθη στην εργασία του Naor 1969 και οι σχετικές μελέτες πολλαπλασιάστηκαν κατά τη διάρκεια των τελευταίων 50 ετών. Στη μονογραφία των Hassin και Havin 2003 μπορεί να βρει κανείς μια ωραία σύνοψη της βασικής μεθοδολογίας και των αποτελεσμάτων στην περιοχή αυτή. Τα βιβλία Stidham 2009 και Hassin Hassin 2016 περιέχουν επιπλέον υλικό καθώς και τη σύνοψη περισσότερων σχετικών μοντέλων.

Κάτω από τη στρατηγική οπτική, οι παράγοντες ενός συστήματος εξυπηρέτησης είναι αποφασίζουσες οντότητες που έχουν στόχο τη μεγιστοποίηση της ωφέλειάς τους. Επομένως, οι ιδέες της Θεωρίας Παιγνίων είναι εφαρμόσιμες και οι αλληλεπιδράσεις που λαμβάνουν χώρα στα αντίστοιχα συστήματα εξυπηρέτησης μοντελοποιούνται ως μη-συνεργατικά παίγνια. Ο σκοπός των ερευνητών είναι να προσδιορίσουν τη στρατηγική συμπεριφορά που θα υιοθετηθεί σε αυτά τα συστήματα και να επινοήσουν μηχανισμούς ώστε οι οντότητες να υιοθετήσουν επιθυμητές στρατηγικές, υπό το πρίσμα ενός κοινωνικού σχεδιαστή ή μιας επιχείρησης (βλέπε π.χ. Havin και Oz 2018).

Για τη ρύθμιση ενός συστήματος εξυπηρέτησης, μπορεί κανείς να εφαρμόσει γενικές οικονομικές ιδέες, όπως η κατάλληλη τιμολόγηση των υπηρεσιών που προσφέρει ένα σύστημα εξυπηρέτησης. Επιπλέον, υπάρχουν κάποια εργαλεία που είναι διαθέσιμα μόνο για συστήματα εξυπηρέτησης, όπως η ανάθεση/πώληση προτεραιοτήτων στους πελάτες και ο έλεγχος της πληροφορίας που παρέχεται σε αυτούς.

Ο έλεγχος της πληροφορίας είναι μια ιδιαίτερα ενεργή περιοχή. Για μια εισαγωγή και τη σχετική βιβλιογραφία ο ενδιαφερόμενος αναγνώστης παραπέμπεται στην εργασία του Economou 2021, καθώς και στο κεφάλαιο 4 του Hassin 2016, και την εργασία της Ibrahim 2018.

Βιβλιογραφία

- [1] P. Naor. “The regulation of queue size by levying tolls”. Στο: *Econometrica* 37 (1969), σσ. 15–24.
- [2] N. M. Edelson και K. Hildebrand. “Congestion tolls for Poisson queueing processes”. Στο: *Econometrica* 43 (1975), σσ. 81–92.
- [3] R. Hassin. “Consumer information in markets with random products quality: The case of queues and balking”. Στο: *Econometrica* 54 (1986), σσ. 1185–1195.
- [4] H. Chen και M. Frank. “Monopoly pricing when customers queue”. Στο: *IIE Transactions* 36 (2004), σσ. 569–581.
- [5] A. Economou και S. Kanta. “Optimal balking strategies and pricing for the single server Markovian queue with compartmented waiting space”. Στο: *Queueing Systems* 59 (2008), σσ. 237–269.
- [6] P. Guo και P. Zipkin. “The effects of the availability of waiting-time information on a balking queue”. Στο: *European Journal of Operational Research* 198 (2009), σσ. 199–209.
- [7] R. Hassin και A. Koshman. “Profit maximization in the M/M/1 queue”. Στο: *Operations Research Letters* 45 (2017), σσ. 436–441.
- [8] A. Burnetas, A. Economou και G. Vasiliadis. “Strategic customer behavior in a queueing system with delayed observations”. Στο: *Queueing Systems* 86 (2017), σσ. 389–418.
- [9] R. Hassin και R. Roet-Green. “On queue-length information when customers travel to a queue”. Στο: *Manufacturing & Service Operations Management* 23 (2020), σσ. 989–1004.
- [10] M. Hu, Y. Li και J. Wang. “Efficient ignorance: Information heterogeneity in a queue”. Στο: *Management Science* 64 (2018), σσ. 2650–2671.

- [11] Y. Dimitrakopoulos, A. Economou και S. Leonardos. “Strategic customer behavior in a queueing system with alternating information structure”. Στο: *European Journal of Operational Research* 291 (2021), σσ. 1024–1040.
- [12] A. Burnetas και A. Economou. “Equilibrium customer strategies in a single server Markovian queue with setup times”. Στο: *Queueing Systems* 56 (2007), σσ. 213–228.
- [13] A. Economou και A. Manou. “Equilibrium balking strategies for a clearing queueing system in alternating environment”. Στο: *Annals of Operations Research* 208 (2013), σσ. 489–514.
- [14] M. Haviv. “Regulating an M/G/1 queue when customers know their demand”. Στο: *Performance Evaluation* 77 (2014), σσ. 57–71.
- [15] R. Hassin και M. Haviv. *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*. Kluwer, 2003. ISBN: 978-1402072031.
- [16] S.Jr. Stidham. *Optimal Design of Queueing Systems*. CRC Press, 2009. ISBN: 978-1584880769.
- [17] R. Hassin. *Rational Queueing*. CRC Press, 2016. ISBN: 978-1498745277.
- [18] M. Haviv και B. Oz. “Self-regulation of an unobservable queue”. Στο: *Management Science* 64 (2018), σσ. 2380–2389.
- [19] A. Economou. “The Impact of Information Structure on Strategic Behavior in Queueing Systems”. Στο: *Anisimov V. and Limnios, N. (eds.) Queueing Theory 2*. John Wiley και Sons, Ltd, 2021. Κεφ. 4, σσ. 137–169. ISBN: 978-1119755234. DOI: <https://doi.org/10.1002/9781119755234.ch4>.
- [20] R. Ibrahim. “Sharing delay information in service systems: a literature survey”. Στο: *Queueing Systems* 89 (2018), σσ. 49–79. DOI: 10.1007/s11134-018-9577-y.

ΚΛΑΣΙΚΑ ΠΡΟΒΛΗΜΑΤΑ ΒΕΛΤΙΣΤΟΥ ΣΧΕΔΙΑΣΜΟΥ ΚΑΙ ΒΕΛΤΙΣΤΟΥ ΕΛΕΓΧΟΥ

Στο προηγούμενο κεφάλαιο είδαμε το πλαίσιο για τη μελέτη της στρατηγικής συμπεριφοράς των πελατών και για τον βέλτιστο σχεδιασμό της λειτουργίας ενός συστήματος. Για τον σκοπό αυτό χρησιμοποιήθηκε ως βασικό παράδειγμα το πρόβλημα της εισόδου/αποχώρησης των πελατών ή αλλιώς το πρόβλημα της βέλτιστης αποδοχής των πελατών. Στο κεφάλαιο αυτό περιγράφουμε κάποια άλλα κλασικά προβλήματα βέλτιστου σχεδιασμού και βέλτιστου ελέγχου από τη Θεωρία Ουρών Αναμονής. Τα προβλήματα αυτά έχουν μελετηθεί σε μεγάλη έκταση στη βιβλιογραφία και σε πολύ γενικότερα πλαίσια από αυτά που επιτρέπει η παρούσα εισαγωγή. Για τον λόγο αυτό, στα βασικά προβλήματα περιγράφουμε τις λύσεις και παρέχουμε τις αποδείξεις, περιοριζόμενοι στα απλούστερα δυνατά πλαίσια, δηλαδή χρησιμοποιώντας απλές Μαρκοβιανές ουρές και τα αποτελέσματα που έχουμε ήδη αναπτύξει γι' αυτές. Στα υπόλοιπα προβλήματα περιγράφουμε μόνο τα βασικά αποτελέσματα που έχουν αποδειχθεί στη βιβλιογραφία και παραπέμπουμε στις σχετικές εργασίες, χωρίς να παρουσιάζουμε μαθηματικές αποδείξεις.

16.1 Ένας γρήγορος ή πολλοί αργοί υπηρέτες;

Το ερώτημα εδώ είναι πώς πρέπει να σχεδιαστεί ένα σύστημα, όταν υπάρχει η δυνατότητα είτε για πολλούς υπηρέτες με κάποια δυναμικότητα είτε για λιγότερους υπηρέτες με ίδια συνολική δυναμικότητα.

Στην απλούστερη εκδοχή του το ερώτημα τίθεται ως εξής: Έστω ότι πελάτες φθάνουν σε ένα σύστημα σύμφωνα με μια διαδικασία Poisson με ρυθμό λ . Ποιο είναι προτιμότερο; Ένα σύστημα με 1 υπηρέτη και $\text{Exp}(2\mu)$ χρόνους εξυπηρέτησης ή ένα σύστημα με 2 υπηρέτες και $\text{Exp}(\mu)$ χρόνους εξυπηρέτησης; Πρόκειται, δηλαδή, για τη σύγκριση ενός M/M/1 συστήματος με ρυθμό εξυπηρέτησης 2μ για τον μοναδικό υπηρέτη έναντι ενός M/M/2 συστήματος με ρυθμό εξυπηρέτησης μ ανά υπηρέτη, όταν και τα δύο έχουν τον ίδιο ρυθμό αφίξεων λ . Θέτουμε $\rho = \frac{\lambda}{2\mu}$, τον κοινό ρυθμό συνωστισμού των δυο συστημάτων.

Για τη σύγκριση των δύο συστημάτων, θα αναφερόμαστε στα μέτρα του πρώτου συστήματος με τον εκθέτη (1), ενώ θα χρησιμοποιούμε τον εκθέτη (2) για το δεύτερο σύστημα. Έστω, λοιπόν, $(p_n^{(1)})$ και $(p_n^{(2)})$ οι κατανομές

ισορροπίας του αριθμού των πελατών στα δυο συστήματα. Τότε, έχουμε

$$p_n^{(1)} = (1 - \rho)\rho^n, \quad n \geq 0, \quad (16.1)$$

$$p_n^{(2)} = \begin{cases} \frac{2-\rho}{2+\rho} & \text{αν } n = 0, \\ \frac{2-\rho}{2+\rho} \frac{\rho^n}{2^{n-1}} & \text{αν } n \geq 1. \end{cases} \quad (16.2)$$

Από εδώ έχουμε ότι για τους μέσους αριθμούς πελατών στα δυο συστήματα ισχύει

$$E[Q^{(1)}] < E[Q^{(2)}], \quad (16.3)$$

ενώ για τους μέσους αριθμούς πελατών στα δυο συστήματα στους αντίστοιχους χώρους αναμονής έχουμε

$$E[Q_q^{(1)}] > E[Q_q^{(2)}]. \quad (16.4)$$

Με βάση τον νόμο του Little συνάγουμε ότι για τους μέσους χρόνους παραμονής πελάτη στα δυο συστήματα ισχύει

$$E[S^{(1)}] < E[S^{(2)}], \quad (16.5)$$

ενώ για τους μέσους χρόνους αναμονής στα δυο συστήματα έχουμε

$$E[W^{(1)}] > E[W^{(2)}]. \quad (16.6)$$

Το συμπέρασμα είναι ότι το σύστημα με έναν γρήγορο υπηρέτη υπερέχει του συστήματος με δυο αργούς υπηρέτες αν το ζητούμενο είναι η ελαχιστοποίηση του χρόνου παραμονής πελάτη, ενώ το σύστημα με τους δυο αργούς υπηρέτες είναι καλύτερο όσον αφορά τον χρόνο αναμονής.

Βλέπουμε, λοιπόν, ότι ανάλογα με την εφαρμογή πρέπει να επιλεγεί και ο σχεδιασμός του συστήματος. Σε ένα σύστημα στο οποίο ο χρόνος εξυπηρέτησης είναι ευχάριστος και η όποια δυσαρέσκεια προκύπτει από την αναμονή, το M/M/2 σύστημα φαίνεται να υπερέχει. Από την άλλη μεριά σε συστήματα όπου το ζητούμενο του πελάτη είναι η ελαχιστοποίηση του χρόνου παραμονής του, το M/M/1 σύστημα είναι προτιμότερο.

Ένα άλλο ερώτημα είναι πόσο γενικά είναι αυτά τα αποτελέσματα. Είναι σχετικά εύκολο να δούμε ότι τα αποτελέσματα συνεχίζουν να ισχύουν αν παραμείνουμε στο Μαρκοβιανό πλαίσιο. Συγκεκριμένα, έστω ότι έχουμε να συγκρίνουμε ένα σύστημα M/M/c⁽¹⁾ με ένα σύστημα M/M/c⁽²⁾, όπου και τα δυο συστήματα έχουν τον ίδιο ρυθμό αφίξεων λ και ίδιους συνολικούς ρυθμούς εξυπηρέτησης (δηλαδή c⁽¹⁾μ⁽¹⁾ = c⁽²⁾μ⁽²⁾, όπου μ⁽ⁱ⁾ ο ρυθμός εξυπηρέτησης ανά υπηρέτη για i = 1, 2) και έστω ότι c⁽¹⁾ < c⁽²⁾. Τότε ισχύουν γενικά οι σχέσεις (16.3), (16.4), (16.5) και (16.6) που είδαμε να ισχύουν για την περίπτωση c⁽¹⁾ = 1 και c⁽²⁾ = 2.

Αν φύγουμε από το Μαρκοβιανό πλαίσιο, η ανάλυση γίνεται αρκετά δυσκολότερη. Πάντως, η βασική σχέση (16.3) αποδεικνύεται ότι ισχύει ακόμη κι αν η διαδικασία των αφίξεων δεν είναι Poisson, αρκεί οι χρόνοι εξυπηρέτησης να είναι αρκετά κανονικοί. Πράγματι, αν ο τετραγωνικός συντελεστής μεταβλητότητας c_B² = $\frac{Var[B]}{(E[B])^2}$ των χρόνων εξυπηρέτησης B είναι μικρότερος ή ίσος της μονάδας η σχέση (16.3) εξακολουθεί να ισχύει. Για περισσότερες λεπτομέρειες για το πρόβλημα αυτό και επιπλέον αναφορές, μπορεί κανείς να συμβουλευτεί το βιβλίο του Wolff 1989, και συγκεκριμένα τις παραγράφους 5.9 και 11.7.

16.2 Κοινή ουρά ή ουρά ανά υπηρέτη;

Το επόμενο κλασικό ερώτημα είναι το κατά πόσον είναι καλύτερο στην περίπτωση που υπάρχουν πολλοί παράλληλοι υπηρέτες να υπάρχει μια ξεχωριστή ουρά για τον καθένα ή μια ενιαία ουρά για όλους. Γενικά η διαδικασία της ενοποίησης ξεχωριστών ουρών αναμονής σε μία που τροφοδοτεί πολλούς παράλληλους υπηρέτες αναφέρεται ως pooling στη διεθνή βιβλιογραφία.

Για να δούμε συγκεκριμένα αν πλεονεκτεί η κοινή ουρά έναντι πολλαπλών ουρών, μία για κάθε υπηρέτη, θεωρούμε ένα Μαρκοβιανό πλαίσιο. Για τον σκοπό αυτό θεωρούμε ότι έχουμε c ανεξάρτητες M/M/1 ουρές

με ρυθμό αφίξεων $\frac{\lambda}{c}$ στην κάθε μία και ρυθμό εξυπηρέτησης μ . Αναφερόμαστε στο συνολικό σύστημα με τον εκθέτη (1) για τα μέτρα απόδοσής του. Αν κατευθύνουμε όλες τις ροές αφίξεων πελατών σε μια κοινή ουρά που τροφοδοτεί τους c παράλληλους υπηρέτες, τότε έχουμε μια $M/M/c$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ ανά υπηρέτη. Θα αναφερόμαστε σε αυτό το σύστημα με τον εκθέτη (2). Χρησιμοποιώντας τους τύπους για τους μέσους αριθμούς πελατών στις $M/M/1$ και $M/M/c$ ουρές έχουμε, ότι για τους μέσους χρόνους αναμονής στα δυο συστήματα ισχύει

$$E[W^{(1)}] > E[W^{(2)}]. \quad (16.7)$$

Δηλαδή, η κοινή ουρά μειώνει τον μέσο χρόνο αναμονής των πελατών. Μάλιστα το πλεονέκτημα αυτό της κοινής ουράς γίνεται ιδιαίτερα ισχυρό όταν το σύστημα είναι ιδιαίτερα συνωστισμένο με ρυθμό συνωστισμού $\rho = \frac{\lambda}{\mu}$ κοντά στο 1.

Πρέπει να σημειωθεί ότι το αποτέλεσμα αυτό γενικεύεται και αποδεικνύεται ότι η υπεροχή της κοινής ουράς είναι ένα πολύ γενικό αποτέλεσμα που ισχύει ανεξάρτητα από το ποια είναι η διαδικασία των αφίξεων και ποια η κατανομή των χρόνων εξυπηρέτησης. Από την άλλη μεριά, η κοινή ουρά έχει κάποια σοβαρά μειονεκτήματα. Π.χ., η ύπαρξη μιας πολύ μεγάλης ουράς έστω και με πολλούς παράλληλους υπηρέτες τείνει να επηρεάζει αρνητικά την ψυχολογία των πελατών. Επίσης, η κοινή ουρά δυσκολεύει την αξιολόγηση των υπηρετών και μειώνει τα κίνητρά τους να εργαστούν με γρήγορο ρυθμό, καθώς οι πελάτες χρεώνουν την οποιαδήποτε αργοπορία τους γενικά στο σύστημα και όχι σε κάποιον συγκεκριμένο υπηρέτη και αντίστροφα πιστώνουν την γρηγοράδα στην εξυπηρέτησή τους στο σύστημα και όχι σε κάποιον συγκεκριμένο υπηρέτη. Επομένως, ενώ το pooling φαίνεται να υπερέχει από μαθηματική σκοπιά, θα πρέπει να εφαρμόζεται με προσοχή.

16.3 Γρήγορος ακριβός υπηρέτης ή αργός οικονομικός υπηρέτης;

Ένα άλλο πρόβλημα που εμφανίζεται κατά τον σχεδιασμό ενός συστήματος εξυπηρέτησης είναι η επιλογή του βέλτιστου ρυθμού εξυπηρέτησης ώστε τα συνολικά έξοδα παροχής εξυπηρέτησης και αναμονής να ελαχιστοποιούνται. Τυπικά, αν επιλεγεί ένας υψηλός ρυθμός εξυπηρέτησης το κόστος παροχής της εξυπηρέτησης είναι μεγάλο, ενώ τα κόστη αναμονής μικρά. Αντίθετα, ένας χαμηλός ρυθμός εξυπηρέτησης συνεπάγεται χαμηλό κόστος παροχής εξυπηρέτησης, αλλά επάγει μεγάλα κόστη αναμονής. Επομένως, ποιος ρυθμός εξισορροπεί καλύτερα αυτά τα δύο κόστη;

Για να μελετήσουμε μια απλή και συγκεκριμένη περίπτωση του προβλήματος, θεωρούμε μια $M/M/1$ ουρά με δοσμένο ρυθμό αφίξεων λ . Ο ρυθμός εξυπηρέτησης είναι μεταβλητή απόφασης που θέλουμε να καθοριστεί έτσι ώστε να μειωθεί το συνολικό κόστος λειτουργίας του συστήματος. Το σύστημα έχει δύο είδη κόστους. Αφενός έχει κόστος εξυπηρέτησης s ανά χρονική μονάδα, για κάθε μονάδα ρυθμού εξυπηρέτησης που παρέχει και αφετέρου έχει κόστος αναμονής c ανά χρονική μονάδα και ανά πελάτη. Δηλαδή, αν ο ρυθμός εξυπηρέτησης που έχει επιλεγεί είναι μ τότε το κόστος εξυπηρέτησης είναι $s\mu$ ανά χρονική μονάδα, και αν υπάρχουν n πελάτες στο σύστημα μια χρονική μονάδα τότε το κόστος αναμονής για αυτή τη χρονική μονάδα είναι cn .

Επομένως, αν ο ρυθμός εξυπηρέτησης τεθεί μ , τότε το μέσο κόστος ανά χρονική μονάδα είναι

$$C(\mu) = s\mu + cE_{\mu}[Q] = s\mu + c\frac{\lambda}{\mu - \lambda}, \quad \mu > \lambda. \quad (16.8)$$

Το πρόβλημα βελτιστοποίησης είναι το

$$\begin{aligned} \min_{\mu} (s\mu + cE_{\mu}[Q]) &= \min_{\mu} \left(s\mu + \frac{c\lambda}{\mu - \lambda} \right) \\ &\text{υπό τον περιορισμό} \\ &\mu > \lambda. \end{aligned} \quad (16.9)$$

Παρατηρούμε ότι

$$C'(\mu) = s - \frac{c\lambda}{(\mu - \lambda)^2},$$

$$C''(\mu) = \frac{2c\lambda}{(\mu - \lambda)^3} > 0, \quad \mu > \lambda.$$

Επομένως, η $C(\mu)$ είναι κυρτή στο (λ, ∞) , με

$$\lim_{\mu \rightarrow \lambda^+} C(\mu) = \lim_{\mu \rightarrow \infty} C(\mu) = \infty,$$

οπότε το ολικό ελάχιστο βρίσκεται λύνοντας την εξίσωση $C'(\mu) = 0$, που δίνει το βέλτιστο ρυθμό εξυπηρέτησης

$$\mu^* = \lambda + \sqrt{\frac{\lambda c}{s}}, \quad (16.10)$$

που επάγει το ελάχιστο αναμενόμενο κόστος ανά χρονική μονάδα

$$C(\mu^*) = s\lambda + 2\sqrt{\lambda cs}. \quad (16.11)$$

Το πρόβλημα της επιλογής ρυθμού εξυπηρέτησης που εξετάσαμε στην παράγραφο αυτή είναι ένα πρόβλημα σχεδιασμού υπό την έννοια ότι ο βέλτιστος ρυθμός εξυπηρέτησης επιλέγεται μια φορά και για πάντα και δεν προσαρμόζεται στην κατάσταση του συστήματος. Με την ορολογία του προηγούμενου κεφαλαίου θα λέγαμε ότι λύσαμε το πρόβλημα της επιλογής ρυθμού εξυπηρέτησης σε ένα μη-παρατηρήσιμο σύστημα, όπου οι πελάτες δεν είναι στρατηγικοί (πρόβλημα εύρεσης κοινωνικά βέλτιστης στρατηγικής σε μη-παρατηρήσιμο σύστημα). Όμως, συνήθως, ο διαχειριστής του συστήματος έχει τη δυνατότητα προσαρμογής του ρυθμού εξυπηρέτησης ανάλογα με τον συνωστισμό που παρατηρείται στο σύστημα. Π.χ., σε ένα σούπερ μάρκετ, οι υπάλληλοι που εξυπηρετούν στα ταμεία αυξομειώνονται ανάλογα με τον αριθμό των πελατών που βρίσκονται σε αναμονή. Επίσης, στο τηλεφωνικό κέντρο μιας εταιρείας, σε περίπτωση μεγάλου φόρτου κλήσεων, καλούνται να εξυπηρετήσουν πελάτες και υπάλληλοι που εργάζονται σε άλλα τμήματα.

Το πρόβλημα της προσαρμογής του ρυθμού εξυπηρέτησης ανάλογα με τον συνωστισμό του συστήματος είναι αρκετά δυσκολότερο στην επίλυσή του από το πρόβλημα της επιλογής ρυθμού εξυπηρέτησης που είδαμε παραπάνω. Ακόμα και στην απλούστερη εκδοχή του χρειάζεται πιο προχωρημένες μαθηματικές μεθόδους ώστε να λυθεί. Στην περίπτωση Μαρκοβιανών συστημάτων, η βασική προσέγγιση είναι μέσω της Θεωρίας των Μαρκοβιανών Διαδικασιών Αποφάσεων (Στοχαστικός Δυναμικός Προγραμματισμός).

Καθώς οι μέθοδοι του Στοχαστικού Δυναμικού Προγραμματισμού εκφεύγουν της παρούσας εισαγωγής, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο βιβλίο Cassandras και Lafortune 2008, στο κεφάλαιο 9, για μια σύντομη εισαγωγή στη μεθοδολογία του Στοχαστικού Δυναμικού Προγραμματισμού με βασικές εφαρμογές στη Θεωρία Ουρών Αναμονής. Στο βιβλίο αυτό περιγράφεται αναλυτικά και η λύση του προβλήματος της προσαρμογής του ρυθμού εξυπηρέτησης ανάλογα με το συνωστισμό του συστήματος, στο απλούστερο δυνατό πλαίσιο. Συγκεκριμένα, μελετάται μια M/M/1 ουρά, όπου ο υπηρέτης μπορεί να λειτουργεί με ρυθμό $\mu \in \{\mu_1, \mu_2\}$ με $\mu_1 > \mu_2$, όπου το κόστος εξυπηρέτησης ανά χρονική μονάδα για ρυθμό μ_i είναι s_i , για $i = 1, 2$, με $s_1 > s_2$. Το κόστος αναμονής ανά πελάτη και χρονική μονάδα είναι c . Τότε, αποδεικνύεται ότι η βέλτιστη πολιτική προσαρμογής για την ελαχιστοποίηση του μέσου κόστους ανά χρονική μονάδα είναι τύπου κατωφλίου, δηλαδή ο υπηρέτης πρέπει να δουλεύει με τον αργό ρυθμό μ_2 όσο ο αριθμός των πελατών είναι μικρότερος ή ίσος από κάποιον κρίσιμο αριθμό n^* , ενώ πρέπει να δουλεύει με τον γρήγορο ρυθμό μ_1 , όταν ο αριθμός των πελατών είναι μεγαλύτερος του n^* . Επιπλέον, ο κρίσιμος αριθμός n^* χαρακτηρίζεται με βάση τις παραμέτρους λειτουργίας του συστήματος λ, μ_1, μ_2 και τις οικονομικές παραμέτρους s_1, s_2 και c .

16.4 Προς ποιον υπηρέτη να κατευθυνθούν οι πελάτες;

Το γενικό πρόβλημα δρομολόγησης πελατών σε διαφορετικούς υπηρέτες απαιτεί την ύπαρξη τουλάχιστον δύο συστημάτων εξυπηρέτησης (υπηρετών). Πελάτες φθάνουν σύμφωνα με κάποια διαδικασία αφίξεων και το ερώτημα είναι σε ποιο σύστημα πρέπει να δρομολογηθούν προς εξυπηρέτηση ώστε να ελαχιστοποιείται το κόστος λειτουργίας του συστήματος.

Στην πλέον απλή περίπτωση μπορούμε να θεωρήσουμε ότι έχουμε πελάτες που φθάνουν σύμφωνα με μια διαδικασία Poisson με ρυθμό λ και δυο υπηρέτες με $\text{Exp}(\mu_1)$ και $\text{Exp}(\mu_2)$ χρόνους εξυπηρέτησης, με $\mu_1 > \mu_2$. Το ζητούμενο είναι να αποφασίσουμε την πιθανότητα p να δρομολογηθεί ένας πελάτης στο σύστημα που έχει ρυθμό εξυπηρέτησης μ_1 , έτσι ώστε να ελαχιστοποιείται ο μέσος αριθμός πελατών που είναι παρόντες στο σύστημα.

Ουσιαστικά έχουμε να λύσουμε το πρόβλημα βελτιστοποίησης

$$\begin{aligned} \min_p \left(\frac{\lambda p}{\mu_1 - \lambda p} + \frac{\lambda(1-p)}{\mu_2 - \lambda(1-p)} \right) \\ \text{υπό τους περιορισμούς} \\ p \in [0, 1], \\ \lambda p < \mu_1, \\ \lambda(1-p) < \mu_2. \end{aligned} \quad (16.12)$$

Έστω

$$C(p) = \frac{\lambda p}{\mu_1 - \lambda p} + \frac{\lambda(1-p)}{\mu_2 - \lambda(1-p)}$$

η αντικειμενική συνάρτηση του προβλήματος. Προφανώς το βέλτιστο επιτυγχάνεται στο σημείο $p = 0$, στο σημείο $p = 1$ ή στο σημείο που $p = \bar{p}$ που μηδενίζεται η $C'(p)$. Είναι εύκολο να δούμε ότι

$$C'(p) = \frac{\lambda\mu_1}{(\mu_1 - \lambda p)^2} - \frac{\lambda\mu_2}{(\mu_2 - \lambda(1-p))^2}$$

και επομένως το κρίσιμο σημείο \bar{p} είναι τέτοιο ώστε

$$\frac{\mu_1}{(\mu_1 - \lambda\bar{p})^2} = \frac{\mu_2}{(\mu_2 - \lambda(1-\bar{p}))^2} = \alpha.$$

Αλλά τότε συνάγουμε ότι

$$\lambda\bar{p} = \mu_1 - \frac{\sqrt{\mu_1}}{\sqrt{\alpha}}, \quad \lambda(1-\bar{p}) = \mu_2 - \frac{\sqrt{\mu_2}}{\sqrt{\alpha}},$$

οπότε προσθέτοντάς τες και λύνοντας για το $\sqrt{\alpha}$ παίρνουμε

$$\sqrt{\alpha} = \frac{\sqrt{\mu_1} + \sqrt{\mu_2}}{\mu_1 + \mu_2 - \lambda}$$

και τελικά αντικαθιστώντας έχουμε ότι το κρίσιμο σημείο \bar{p} δίνεται από τη σχέση

$$\bar{p} = \frac{1}{\lambda} \left(\mu_1 - \frac{\sqrt{\mu_1}}{\sqrt{\mu_1} + \sqrt{\mu_2}} (\mu_1 + \mu_2 - \lambda) \right). \quad (16.13)$$

Επίσης, έχουμε ότι η συνάρτηση $C(p)$ είναι κυρτή, αφού μπορούμε εύκολα να ελέγξουμε ότι $C''(p) > 0$. Παρατηρήστε, επίσης ότι $C(0) = \frac{\lambda}{\mu_2 - \lambda} > \frac{\lambda}{\mu_1 - \lambda} = C(1)$, οπότε το βέλτιστο δεν επιτυγχάνεται ποτέ στο $p = 0$. Για να δούμε αν το βέλτιστο επιτυγχάνεται στο $p = 1$ ή στο $p = \bar{p}$, αρκεί να ελέγξουμε το πρόσημο της $C'(1)$.

Αν $C'(1) \leq 0$, τότε το βέλτιστο επιτυγχάνεται στο 1, διαφορετικά επιτυγχάνεται στο \bar{p} . Έχουμε εύκολα ότι $C'(1) \leq 0$ αν και μόνο αν $\lambda \leq \mu_1 - \sqrt{\mu_1\mu_2}$, οπότε τελικά η βέλτιστη πιθανότητα είναι

$$p^* = \begin{cases} 1 & \lambda \leq \mu_1 - \sqrt{\mu_1\mu_2}, \\ \frac{1}{\lambda} \left(\mu_1 - \frac{\sqrt{\mu_1}}{\sqrt{\mu_1} + \sqrt{\mu_2}} (\mu_1 + \mu_2 - \lambda) \right) & \lambda > \mu_1 - \sqrt{\mu_1\mu_2}. \end{cases} \quad (16.14)$$

Όταν οι υπηρέτες είναι περισσότεροι από δύο, μπορούμε να διατυπώσουμε ένα πρόβλημα βελτιστοποίησης παρόμοιο με το (16.12) και να το λύσουμε με τη μέθοδο των πολλαπλασιαστών Lagrange. Επειδή, στα πλαίσια αυτής της εισαγωγής δεν θέλουμε να επεκταθούμε στη χρήση αυτής της μεθόδου, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στη σχετική εργασία των Bell και S. Stidham 1983. Στην ίδια εργασία, μπορεί κανείς να δει πώς εξετάζεται το πρόβλημα της στρατηγικής συμπεριφοράς των πελατών που αντιμετωπίζουν το ερώτημα προς ποιον υπηρέτη να κατευθυνθούν και πώς υπολογίζεται η στρατηγική ισορροπίας των πελατών.

Όπως περιγράφηκε το παραπάνω πρόβλημα, πρόκειται για ένα πρόβλημα σχεδιασμού, καθώς η δρομολόγηση των πελατών δεν λαμβάνει υπόψη της την παρούσα κατάσταση του συστήματος, δηλαδή το πόσοι πελάτες περιμένουν σε κάθε ουρά για να εξυπηρετηθούν. Όταν ο διαχειριστής του συστήματος βλέπει τον αριθμό των πελατών σε αναμονή σε κάθε ουρά, μπορεί να προσαρμόσει ανάλογα τις αποφάσεις του και έχουμε ένα πρόβλημα βέλτιστου ελέγχου. Τότε, μια πολιτική του απεικονίζει σε κάθε δυνατό διάνυσμα αριθμών πελατών στις διάφορες ουρές την ουρά στην οποία θα κατευθυνθεί ένας αφικνούμενος πελάτης. Το πρόβλημα είναι αρκετά δύσκολο ακόμη και στην απλούστερη εκδοχή του (με μόνο δυο υπηρέτες και Μαρκοβιανό πλαίσιο) και λύνεται χρησιμοποιώντας Μαρκοβιανές Διαδικασίες Αποφάσεων. Ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο βιβλίο των Cassandras και Lafortune 2008, στο κεφάλαιο 9. Η μορφή της βέλτιστης πολιτικής είναι σχετικά απλή και περιγράφεται από μια τεθλασμένη γραμμή αλλαγής (switching curve) που διαχωρίζει το τεταρτημόριο των ζευγών μη-αρνητικών ακεραίων σε δύο μέρη, όπου για καταστάσεις στο ένα μέρος οι πελάτες κατευθύνονται στην ουρά 1 και στο άλλο μέρος κατευθύνονται στην ουρά 2. Πιο συγκεκριμένα, υπάρχει μια συνάρτηση $s(n)$, ώστε αν η κατάσταση του συστήματος είναι (n_1, n_2) (με n_i να είναι ο αριθμός των πελατών στην ουρά i , $i = 1, 2$), τότε ένας αφικνούμενος πελάτης κατευθύνεται στην ουρά 2 αν $n_2 \leq s(n_1)$, ή στην ουρά 1 αν $n_2 > s(n_1)$.

16.5 Ασκήσεις

Άσκηση 16.1 Στο πρόβλημα της παραγράφου 16.1 της σύγκρισης ενός γρήγορου με δυο όμοιους αργούς υπηρέτες ίδιας συνολικής δυναμικότητας, υπολογίστε τους μέσους χρόνους $E[S^{(i)}]$, $E[W^{(i)}]$ για $i = 1, 2$, και αποδείξτε τις σχέσεις (16.5) και (16.6).

Άσκηση 16.2 Λύστε το πρόβλημα της επιλογής ρυθμού εξυπηρέτησης της παραγράφου 16.3, όταν το σύνολο εφικτών τιμών του ρυθμού εξυπηρέτησης είναι κάποιο διάστημα $[\mu_L, \mu_U]$. Επίσης, λύστε το πρόβλημα όταν το σύνολο εφικτών τιμών είναι το πεπερασμένο σύνολο $\{\mu_1, \mu_2, \dots, \mu_k\}$.

Άσκηση 16.3 Στο πρόβλημα της επιλογής ρυθμού εξυπηρέτησης, υποθέστε ότι το κόστος αναμονής ενός πελάτη που παραμένει για χρόνο S στο σύστημα δεν είναι cS (γραμμικό), αλλά cS^a με $a > 0$ γνωστή σταθερά. Να διατυπωθεί το πρόβλημα βελτιστοποίησης που πρέπει να επιλυθεί σε αυτή την περίπτωση και να λυθεί. Για ποιές τιμές του a , είναι η αντικειμενική συνάρτηση (του κόστους ανά χρονική μονάδα) κυρτή ως προς μ ;

Άσκηση 16.4 Να βρείτε τη στρατηγική ισορροπίας των πελατών στο πρόβλημα δρομολόγησης σε δυο ουρές της παραγράφου 16.4. Συγκρίνετε τη στρατηγική αυτή με την (κοινωνικά) βέλτιστη πολιτική δρομολόγησης που βρέθηκε στην παράγραφο αυτή.

Άσκηση 16.5 Θεωρήστε το πρόβλημα της επιλογής ρυθμού αφίξεων και ρυθμού εξυπηρέτησης για την $M/M/1$ ουρά, όπου οι μεταβλητές απόφασης είναι ο ρυθμός αφίξεων λ και ο ρυθμός εξυπηρέτησης μ , όταν υπάρχει κέρδος

r ανά εξυπηρετούμενο πελάτη, κόστος s ανά μονάδα ρυθμού εξυπηρέτησης και χρονική μονάδα και κόστος c ανά παρόντα πελάτη και χρονική μονάδα. Διατυπώστε το πρόβλημα της ελαχιστοποίησης του αναμενόμενου κόστους ανά χρονική μονάδα και επιλύστε το. Τι παρατηρείτε;

16.6 Σχόλια

Τα προβλήματα βέλτιστου σχεδιασμού και βέλτιστου ελέγχου συστημάτων εξυπηρέτησης έχουν μελετηθεί αρκετά εκτενώς στα πλαίσια της Θεωρίας Ουρών.

Το κλασικό εισαγωγικό βιβλίο του Wolff 1989 περιέχει σύντομες αναφορές στα περισσότερα κλασικά προβλήματα σχεδιασμού που είδαμε στο παρόν κεφάλαιο. Το βιβλίο του S.Jr. Stidham 2009 αφιερώνεται εξ ολοκλήρου στην παρουσίαση προβλημάτων βέλτιστου σχεδιασμού και παρουσιάζει τις βασικές μεθόδους και αποτελέσματα που έχουν αναπτυχθεί.

Η βιβλιογραφία που αφορά τον βέλτιστο έλεγχο συστημάτων εξυπηρέτησης είναι εξαιρετικά πλούσια. Το εισαγωγικό βιβλίο των Cassandras και Lafortune 2008 παρουσιάζει τα βασικά προβλήματα βέλτιστου ελέγχου που είδαμε στο παρόν κεφάλαιο με αρκετά στοιχειώδη τρόπο, ενσωματώνοντας μια εισαγωγή στις απαραίτητες έννοιες και τεχνικές από τον Στοχαστικό Δυναμικό Προγραμματισμό (Μαρκοβιανές Διαδικασίες Αποφάσεων). Ένα σημείο εκκίνησης για τον αναγνώστη που θα ήθελε να δει τις αρχικές πηγές αυτής της περιοχής είναι οι εργασίες S.Jr Stidham 1985 και S.Jr. Stidham και Weber 1993. Μια παρουσίαση της θεωρίας του Στοχαστικού Δυναμικού Προγραμματισμού με έμφαση στην εφαρμογή του στον βέλτιστο έλεγχο συστημάτων εξυπηρέτησης δίνεται στο αρκετά θεωρητικό βιβλίο Sennott 1998. Μια μεθοδολογία που έχει αναπτυχθεί για την εφαρμογή του Στοχαστικού Δυναμικού Προγραμματισμού στα συστήματα εξυπηρέτησης με καλά αποτελέσματα αναπτύσσεται στην εργασία του Koole 2007.

Βιβλιογραφία

- [1] R.W. Wolff. *Stochastic Modeling and the Theory of Queues*. Englewood Cliffs, New Jersey: Prentice Hall, 1989. ISBN: 978-0138466923.
- [2] C.G. Cassandras και S. Lafortune. *Introduction to Discrete Event Systems, 2nd Edition*. Springer, 2008.
- [3] C. Bell και S. Stidham. “Individual versus social optimization in the allocation of customers to alternative servers”. Στο: *Management Science* 29 (1983), σσ. 831–839.
- [4] S.Jr. Stidham. *Optimal Design of Queueing Systems*. CRC Press, 2009. ISBN: 978-1584880769.
- [5] S.Jr Stidham. “Optimal control of admission to a queueing system”. Στο: *IEEE Transactions on Automatic Control* 30.8 (1985), σσ. 705–713.
- [6] S.Jr. Stidham και R. Weber. “A survey of Markov decision models for control of networks of queues”. Στο: *Queueing Systems* 13 (1993), σσ. 291–314.
- [7] L.I. Sennott. *Stochastic Dynamic Programming and the Control of Queueing Systems*. Wiley-Interscience, 1998. ISBN: 978-0471161202.
- [8] G. Koole. “Monotonicity in Markov Reward and Decision Chains: Theory and Applications”. Στο: *Foundations and Trends in Stochastic Systems* 1.1 (2007), σσ. 1–76.

ΕΠΙΛΟΓΟΣ

ΙΣΤΟΡΙΑ, ΕΦΑΡΜΟΓΕΣ ΚΑΙ ΚΑΤΕΥΘΥΝΣΕΙΣ

Ιστορία

Η έναρξη της Θεωρίας Ουρών ταυτίζεται με τη δημοσίευση της πρώτης εργασίας του Agner Krarup Erlang (1878-1929) για συστήματα αναμονής (Erlang 1909), στην οποία μελέτησε την ουρά M/D/1 (αν και φυσικά χωρίς αυτό το όνομα, μια και δεν υπήρχε ακόμη η ονοματολογία του David George Kendall (1918-2007) που εισήχθη αρκετά αργότερα στην εργασία Kendall 1953. Αυτό που κάνει την πρώτη εργασία του Erlang τόσο θεμελιώδη είναι η αναγνώριση του συνωστισμού ως στοχαστικού φαινομένου που πρέπει να αναλυθεί με πιθανοθεωρητικά επιχειρήματα. Αν όλα τα χαρακτηριστικά ενός συστήματος είναι ντετερμινιστικά, δηλαδή οι αφίξεις συμβαίνουν σε τακτά διαστήματα και οι χρόνοι εξυπηρέτησης είναι σταθεροί, τότε δεν παρατηρείται συνωστισμός: Σε μια ευσταθή ουρά τύπου D/D/k ο αριθμός των πελατών σε αναμονή είναι πάντα 0.

Στις επόμενες εργασίες του, ο Erlang μελέτησε πιο περίπλοκα συστήματα. Π.χ., στην εργασία Erlang 1917, εισήγαγε τον τύπο απωλειών του, (6.17) (τύπο απωλειών του Erlang ή Erlang B (Erlang B formula)). Λίγο αργότερα, στην εργασία Erlang 1920, ο Erlang μελέτησε την ουρά M/D/k. Τα αποτελέσματα αυτά είναι ιδιαίτερα εντυπωσιακά, αν σκεφτεί κανείς ότι ο Erlang δεν είχε στη διάθεσή του μια ώριμη θεωρία στοχαστικών διαδικασιών, όπως αυτή που αναπτύχθηκε λίγα χρόνια μετά με τις συνεισφορές του Markov και του Kolmogorov.

Η προσέγγιση και τα αποτελέσματα του Erlang ενέπνευσαν μια σειρά επιστημόνων που δούλευαν την εποχή εκείνη σε αυτό που τελικά ονομάστηκε «Θεωρία Τηλεπικοινωνιακού Συνωστισμού» (Teletraffic Theory). Οι επιστήμονες αυτοί είχαν προσληφθεί στα ταχυδρομικά γραφεία της εποχής που τότε ήταν υπεύθυνα και για το τηλεφωνικό δίκτυο.

Ο επόμενος χρονικά πολύ σημαντικός επιστήμονας που εργάστηκε στη Θεωρία Ουρών Αναμονής ήταν ο Felix Pollaczek, που απέκτησε το πτυχίο Μαθηματικού στη Βιέννη και το δίπλωμα Ηλεκτρολόγου Μηχανικού στο Μπρνο. Ακολούθως, ο Pollaczek πήρε το διδακτορικό του στη Θεωρία Αριθμών στο Βερολίνο. Αυτός ο συνδυασμός σπουδών του έδωσε τελικά εξαιρετικά εφόδια για να χρησιμοποιήσει κλασικές μαθηματικές τεχνικές στη Θεωρία Τηλεπικοινωνιακού Συνωστισμού. Σε μια από τις πρώτες του εργασίες στη Θεωρία Ουρών (Pollaczek 1930) παρουσίασε τη μελέτη της M/G/1 ουράς. Ο Pollaczek εργάστηκε με αναλυτικές τεχνικές μέχρι τα βαθιά του γεράματα. Ανήγαγε όσο πιο σύντομα μπορούσε τα προβλήματα της Θεωρίας Ουρών σε ολοκληρωτικές εξισώσεις τις οποίες κατόπιν έλυνε, με αναλυτικά εργαλεία, χωρίς να κατέχει πιθανοθεωρητικές μεθόδους. Είναι πραγματικά εντυπωσιακό το πόσο μακριά μπόρεσε να πάει με τις μεθόδους αυτές.

Οι εργασίες του Pollaczek βρήκαν μεγάλη απήχηση στην κοινότητα της Θεωρίας Τηλεπικοινωνιακού Συνωστισμού, η οποία απαρτιζόταν κυρίως από μηχανικούς. Όμως ήταν δύσκολο να μελετηθούν, λόγω της μαθηματικής πολυπλοκότητάς τους. Για καλή τύχη της Θεωρίας Ουρών, η δουλειά του έπεσε στην αντίληψη του σημαντικού μαθηματικού Aleksandr Yakovlevich Khinchin, ο οποίος γνώριζε τη δουλειά του Markov και ο οποίος ήταν σε επαφή με τον νεαρό τότε Kolmogorov. Ο Khinchin ξαναέγραψε τη λύση της $M/G/1$ ουράς με πιθανοθεωρητικούς όρους και τελικά οι βασικοί τύποι που προέκυψαν αναφέρονται ως τύποι Pollaczek-Khinchin, θυμίζοντας τόσο την πρωτοπορία του Pollaczek στην ανακάλυψή τους, όσο και τη συνεισφορά του Khinchin, που διέγινωσε την πιθανοθεωρητική φύση του προβλήματος και το συνέδεσε με το νέο τότε αντικείμενο των Εφαρμοσμένων Πιθανοτήτων.

Η σύνδεση της Θεωρίας Ουρών με το πεδίο των Εφαρμοσμένων Πιθανοτήτων προσείλκυσε το ενεργό ενδιαφέρον της μαθηματικής κοινότητας. Κατά τη διάρκεια του Β' παγκοσμίου πολέμου, όταν μαθηματικοί από πολλές χώρες ήρθαν αντιμέτωποι με προβλήματα που σχετιζόνταν με την εξέλιξη του πολέμου, γεννήθηκε η Επιχειρησιακή Έρευνα και η Θεωρία Ουρών εντάχθηκε σε αυτό το γενικότερο πεδίο ως η ραχοκοκαλιά του στοχαστικού μέρους της.

Ένας από τους πρώτους μαθηματικούς που ασχολήθηκε εντατικά με τη Θεωρία Ουρών ήταν ο David Kendall. Επηρεασμένος από τις εξελίξεις του Β' Παγκοσμίου Πολέμου και την επίλυση προβλημάτων αποφυγής συνωστισμού που προέκυπταν σε προβλήματα επιχειρήσεων και ανεφοδιασμού, μελέτησε τις εργασίες των Erlang, Khinchin, Crommelin, Palm, Jensen και άλλων. Συνόψισε πολλά από τα σχετικά αποτελέσματα στην εργασία Kendall 1951 που παρουσίασε σε μια από τις συναντήσεις της Βασιλικής Στατιστικής Εταιρείας (του Ηνωμένου Βασιλείου). Σε αυτές τις συναντήσεις συνηθίζεται οι δυο κριτές της εργασίας να αρχίζουν μια συζήτηση η οποία κατόπιν δημοσιεύεται αυτολεξεί με την εργασία. Με τον τρόπο αυτό, η εργασία Kendall (1951) δεν αποτελεί μόνο μια περιήληψη των βασικών αποτελεσμάτων στη Θεωρία Ουρών εκείνης της εποχής, αλλά επιπλέον συνοδεύεται από τις απόψεις σημαντικών επιστημόνων εκείνης της εποχής. Πέραν αυτού, όμως, αυτή η εργασία και η ανάγνωσή της ενώπιον της Βασιλικής Στατιστικής Εταιρείας προσείλκυσε το ενδιαφέρον και άλλων επιστημόνων στο νέο τότε γνωστικό πεδίο. Μεταξύ αυτών ήταν και ο Dennis Lindley, που ξεκίνησε τη συζήτηση με τη διάσημη αναδρομική σχέση του, (14.12), για τους χρόνους αναμονής των πελατών στην $G/G/1$ ουρά.

Η εργασία του Kendall επικεντρώνεται στην $M/G/1$ ουρά, δείχνοντας πώς τα αποτελέσματα των Pollaczek και Khinchin μπορούν να συναχθούν από τη μελέτη της εμφυτευμένης Μαρκοβιανής αλυσίδας σε στιγμές αναχωρήσεων πελατών. Σε μια επόμενη εργασία, στην Kendall 1953, έδειξε πώς μπορεί να μελετηθεί η $G/M/k$ ουρά με ένα δυϊκό επιχείρημα, χρησιμοποιώντας και πάλι τη μεθοδολογία της εμφυτευμένης Μαρκοβιανής αλυσίδας, αλλά σε στιγμές αφίξεων πελατών.

Ενώ η μέθοδος της εμφυτευμένης Μαρκοβιανής αλυσίδας επιτρέπει τη μελέτη συστημάτων που είτε οι ενδιαμέσοι χρόνοι αφίξεων, είτε οι χρόνοι εξυπηρέτησης είναι εκθετικοί, η εξίσωση του Lindley επιτρέπει τη μελέτη της $G/G/1$ ουράς, καθώς αρκεί να υποθεθεί ότι οι ενδιαμέσοι χρόνοι των αφίξεων και οι χρόνοι των εξυπηρέτησεων συνιστούν ακολουθίες ανεξάρτητων και ισόνομων τυχαίων μεταβλητών. Έτσι η εξίσωση του Lindley αποτέλεσε την αρχή μιας εξαιρετικά πλούσιας βιβλιογραφίας που επικεντρώθηκε στη μελέτη της $G/G/1$ ουράς. Ο J.W. Cohen μελέτησε σε βάθος αυτό το θεμελιώδες μοντέλο και συνόψισε τα αποτελέσματα προγενέστερων ερευνητών στο εξαιρετικό σύγγραμμά του, Cohen 1969.

Μισό αιώνα μετά τη θεμελιώδη εργασία του Erlang που σηματοδότησε την έναρξή της, η Θεωρία Ουρών ήταν ένα ώριμο επιστημονικό πεδίο στο οποίο εργάζονταν εντατικά πολλοί αξιόλογοι επιστήμονες. Στο σημείο εκείνο είχαν ήδη σχηματιστεί πάρα πολλές αποκλίνουσες κατευθύνσεις στην ανάπτυξή της, πράγμα που καθιστά αδύνατη την παρουσίασή τους στα πλαίσια μιας σύντομης εισαγωγής. Για τον λόγο αυτό θα περιοριστούμε σε τρεις κατευθύνσεις που αποδείχθηκαν ιδιαίτερα γόνιμες και οι οποίες παρουσιάστηκαν, έστω και συνοπτικά, στο παρόν σύγγραμμα.

Η πρώτη κατεύθυνση αφορά την ανάπτυξη υπολογιστικών μεθόδων για τη μελέτη των Μαρκοβιανών αλυσίδων συνεχούς χρόνου που προκύπτουν από τη μοντελοποίηση Ουρών Αναμονής. Πολύ νωρίς έγινε κατανοητό ότι οι Ουρές Αναμονής οδηγούν σε Μαρκοβιανές αλυσίδες ιδιαίτερων δομών που μπορούν να αξιοποιηθούν για τη μελέτη τους. Πρωτοπόρος σε αυτή την περιοχή ήταν ο Marcel Neuts, ο οποίος επικεντρώθηκε

στη μελέτη των Μαρκοβιανών αλυσίδων που προκύπτουν από τη μέθοδο των φάσεων και των γενικεύσεων της. Οι μέθοδοι που προέκυψαν αναφέρονται ως πινακοαναλυτικές (matrix-analytic) και αποτελούν ένα πολύ ισχυρό εργαλείο για τη μελέτη συστημάτων αναμονής. Ο ενδιαφερόμενος αναγνώστης μπορεί να τις μελετήσει ανατρέχοντας στα βιβλία Neuts 1981, Neuts 1989 και Latouche και Ramaswami 1999.

Μια δεύτερη κατεύθυνση είναι αυτή της μελέτης δικτύων συστημάτων εξυπηρέτησης. Τα δίκτυα εξυπηρέτησης έχουν τεράστιο αριθμό καταστάσεων ακόμη και για μικρό αριθμό σταθμών και πελατών που κυκλοφορούν σε αυτά (βλέπε και σχετική συζήτηση στο κεφάλαιο 11), οπότε η μελέτη τους είναι γενικά εξαιρετικά δύσκολη. Όμως, ο James R. Jackson ανακάλυψε μια κλάση τους που επιδέχεται αναλυτική λύση και δημοσίευσε τα σχετικά αποτελέσματα στις εργασίες του Jackson 1957 and Jackson 1963. Τα αποτελέσματα αυτά προσείλκυσαν το ενδιαφέρον σημαντικών ερευνητών και η σχετική θεωρία προχώρησε ταχύτατα δίνοντας εξαιρετικά αποτελέσματα. Ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει στα κλασικά συγγράμματα Kelly 1979 και Whittle 1986, είτε στα νεώτερα συγγράμματα των Serfozo 1999 και Chao, Miyazawa και Pinedo 1999, που περιγράφουν με μεγάλη λεπτομέρεια τα αποτελέσματα και τις μεθόδους της περιοχής αυτής.

Τέλος, μια τρίτη κατεύθυνση είναι η μελέτη των συστημάτων εξυπηρέτησης από μια οικονομική οπτική. Η εργασία του Naor 1969 θεωρείται ως το σημείο έναρξης της οικονομικής Θεωρίας Ουρών, που βλέπει τους πελάτες και τους άλλους παράγοντες ενός συστήματος εξυπηρέτησης ως δρώσες οντότητες που λαμβάνουν αποφάσεις με σκοπό τη μεγιστοποίηση της ωφέλειάς τους. Η οπτική αυτή παρουσιάστηκε στο κεφάλαιο 15. Οι Hassin και Hanin 2003 παρουσίασαν τα σημαντικότερα κλασικά μεθοδολογικά εργαλεία και αποτελέσματα που αφορούν την οικονομική οπτική μελέτης των ουρών. Ο S.Jr. Stidham 2009 παρουσίασε μια αναλυτική ανασκόπηση των προβλημάτων βέλτιστου σχεδιασμού συστημάτων εξυπηρέτησης, λαμβάνοντας υπόψη και τη στρατηγική φύση των πελατών και των άλλων οντοτήτων αυτών των συστημάτων. Σχετικά πρόσφατα, ο Hassin 2016 συνόψισε τα αποτελέσματα της τελευταίας δεκαπενταετίας στην περιοχή αυτή. Ιδιαίτερο ενδιαφέρον στη μελέτη των ουρών αναμονής από την οικονομική οπτική είναι η μελέτη του ιδανικού επιπέδου πληροφορίας που πρέπει να παρέχεται στους πελάτες. Για μια σύντομη και στοιχειώδη εισαγωγή στο ζήτημα αυτό, ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει στις εργασίες Economou 2021 και Economou 2022.

Για μια αρκετά εκτένεστερη παρουσίαση της ιστορίας της Θεωρίας Ουρών Αναμονής, μπορεί κανείς να ανατρέξει στην εργασία S.Jr Stidham 2002 που δημοσιεύτηκε στον επετειακό τόμο 50 του περιοδικού Operations Research, ο οποίος περιείχε άρθρα σημαντικών επιχειρησιακών ερευνητών για την ιστορία της Επιχειρησιακής Έρευνας. Στον τόμο αυτό μπορεί να βρει και άλλα άρθρα ιστορικού ενδιαφέροντος για την εξέλιξη της Επιχειρησιακής Έρευνας και της Θεωρίας Ουρών μέσα σε αυτήν.

Εφαρμογές

Στις επόμενες παραγράφους αναφερόμαστε συνοπτικά σε τρεις περιοχές που η Θεωρία Ουρών έχει εφαρμοστεί πολύ αποτελεσματικά.

Εφαρμογές σε τηλεφωνικά κέντρα και υπολογιστικά συστήματα

Μια σημαντική περιοχή εφαρμογής της Θεωρίας Ουρών είναι η διαχείριση τηλεφωνικών κέντρων και υπολογιστικών συστημάτων. Ιστορικά, άλλωστε, η Θεωρία Ουρών ξεκίνησε από την προσπάθεια υπολογισμών για τον συνωστισμό σε τηλεφωνικά δίκτυα.

Ένα τηλεφωνικό κέντρο μπορεί να θεωρηθεί ως ένα σύστημα εξυπηρέτησης με μια ουρά και πολλούς παράλληλους υπηρέτες. Η διαδικασία αφίξεων μπορεί συχνά να μοντελοποιηθεί ως μη-ομογενής Poisson, αφού ο ρυθμός αφίξεων κλήσεων εξαρτάται από την ώρα της ημέρας και άλλους παράγοντες (π.χ. κατά πόσο η μέρα είναι εργάσιμη ή αργία, αν υπάρχει κάποιο γεγονός που πυροδοτεί επιπλέον κλήσεις όπως κάποια βλάβη κλπ.). Οι χρόνοι εξυπηρέτησης έχουν γενικά μια αυθαίρετη κατανομή. Ο ρυθμός των αφίξεων και η κατανομή των χρόνων εξυπηρέτησης μπορούν να εκτιμηθούν από δεδομένα προηγούμενων χρήσεων του κέντρου και

κατόπιν μπορούν να χρησιμοποιηθούν ως δεδομένα εισόδου για τη μελέτη ενός μαθηματικού μοντέλου του συστήματος.

Τα βασικά ερωτήματα που τίθενται αφορούν τους υπολογισμούς του μέσου χρόνου αναμονής ενός πελάτη μέχρι να αρχίσει να εξυπηρετείται ή της πιθανότητας να βρει ένας πελάτης όλους τους υπηρέτες απασχολημένους και να χρειαστεί να περιμένει ή της πιθανότητας ένας πελάτης να περιμένει πάνω από ένα αποδεκτό όριο. Η Θεωρία Ουρών μπορεί να απαντήσει σε αυτά τα ερωτήματα και κατόπιν οι απαντήσεις μπορούν να χρησιμοποιηθούν για τη βελτίωση της λειτουργίας του συστήματος.

Ο τελικός στόχος της χρήσης της Θεωρίας Ουρών στον έλεγχο ενός τηλεφωνικού κέντρου είναι ο προγραμματισμός των βαρδιών του προσωπικού. Φυσικά, ένας τέτοιος προγραμματισμός απαιτεί και άλλα εργαλεία Επιχειρησιακής Έρευνας εκτός από τον Μαθηματικό Προγραμματισμό.

Για μια εισαγωγή στο θέμα της βελτιστοποίησης της λειτουργίας των τηλεφωνικών κέντρων και της θέσης της Θεωρίας Ουρών σε αυτά, ο ενδιαφερόμενος αναγνώστης μπορεί να συμβουλευτεί το βιβλίο του Koole 2013.

Στα υπολογιστικά συστήματα, η Θεωρία Ουρών απαντάει σε παρόμοια ερωτήματα, όπως των υπολογισμών του μέσου χρόνου διεκπεραίωσης μιας εργασίας, του ρυθμού διεκπεραίωσης εργασιών ανά χρονική μονάδα κλπ.

Εφαρμογές σε συστήματα παροχής υπηρεσιών υγείας

Τα τελευταία χρόνια υπάρχει ένας σημαντικός όγκος εργασιών που αναφέρονται σε εφαρμογές της Θεωρίας Ουρών στη διαχείριση συστημάτων παροχής υπηρεσιών υγείας. Π.χ., έχουν μελετηθεί μοντέλα που αξιολογούν τις διάφορες πολιτικές διαχείρισης των κλινών ενός νοσοκομείου στον χρόνο αναμονής των ασθενών, στο ποσοστό των πελατών που φεύγουν χωρίς να εξυπηρετηθούν και στον φόρτο του προσωπικού του νοσοκομείου.

Γενικά η Θεωρία Ουρών επικεντρώνεται στη μελέτη του χρόνου αναμονής των ασθενών και του φόρτου εργασίας του προσωπικού, στον σχεδιασμό του συστήματος και ιδιαίτερα των συστημάτων ραντεβού των ασθενών με το ιατρικό προσωπικό. Ο στόχος είναι η βελτίωση των παρεχόμενων υπηρεσιών υγείας μέσω της μείωσης των καθυστερήσεων που υφίστανται οι ασθενείς. Μια σημαντική πτυχή αυτού του στόχου είναι η βελτίωση της ροής των πελατών, έτσι ώστε να μην υφίστανται αχρείαστες αναμονές καθώς κινούνται στο σύστημα παροχής υπηρεσιών υγείας. Για τον σκοπό αυτό είναι σημαντικό να συγχρονίζονται οι διάφορες υπηρεσίες με τη ζήτηση που προκύπτει από τους ασθενείς. Αυτό ισχύει τόσο για τις ιατρικές υπηρεσίες και πράξεις, όσο και για βοηθητικές υπηρεσίες που όμως είναι απαραίτητες, όπως η μεταφορά των ασθενών, οι υπηρεσίες καθαριότητας των χώρων κλπ. Η εμπειρία έχει δείξει ότι η αποτελεσματική διαχείριση των υπηρεσιών υγείας μπορεί να βελτιώσει δραματικά τα ιατρικά αποτελέσματα, την ικανοποίηση των ασθενών και επιπλέον να μειώσει αισθητά το κόστος παροχής υπηρεσιών υγείας. Για μια πρώτη επαφή με το αντικείμενο αυτό, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στην εργασία Green 2006.

Εφαρμογές σε συστήματα παραγωγής, αποθεμάτων και διαχείρισης εφοδιαστικών αλυσίδων

Η Θεωρία Ουρών παίζει κεντρικό ρόλο συνδυαζόμενη με άλλα πεδία της Επιχειρησιακής Έρευνας και της Διοίκησης Λειτουργιών στη βελτιστοποίηση συστημάτων παραγωγής και αποθεμάτων, καθώς και στη διαχείριση εφοδιαστικών αλυσίδων. Η αλληλεπίδραση της Θεωρίας Ουρών με τη Θεωρία Ελέγχου Αποθεμάτων πάει πολλά χρόνια πίσω (π.χ. βλέπε το κλασικό σύγγραμμα του Morse 1958 και άρθρα σημαντικών προσωπικοτήτων που μελέτησαν εφαρμογές των ουρών σε προβλήματα αποθεμάτων, όπως του Karush 1957 και του Graves 1982).

Ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει σε κάποιο από τα κλασικά συγγράμματα της Διοίκησης Λειτουργιών για να δει την πληθώρα των προβλημάτων που μπορούν να αντιμετωπιστούν με τη βοήθεια της Θεωρίας Ουρών, π.χ. στο Nahmias και Lennon Olsen 2015 ή στο Hopp και Spearman 2011.

Κατευθύνσεις

Από τη σύντομη ιστορική αναδρομή που είδαμε στη Θεωρία Ουρών και την πληθώρα των εφαρμογών της σε διάφορα πεδία, μπορούμε τώρα να ολοκληρώσουμε την εισαγωγή με μια ματιά στη σημερινή κατάσταση και στις μελλοντικές κατευθύνσεις αυτού του επιστημονικού πεδίου.

Η Θεωρία Ουρών παρέχει ένα πολύ γενικό πλαίσιο μελέτης συστημάτων εισόδου-εξόδου διακριτών μονάδων (πελατών) που μπορεί να αποτελούνται από έναν ή περισσότερους σταθμούς και τα οποία υπόκεινται σε τυχαιότητα. Επομένως, καταρχήν δίνει τη δυνατότητα μοντελοποίησης και μελέτης συστημάτων εξυπηρέτησης που μελετά η Διοικητική Επιστήμη και η Επιχειρησιακή Έρευνα. Από την άλλη μεριά, απαιτεί την ανάπτυξη ιδιαίτερων πιθανοθεωρητικών τεχνικών και εργαλείων βελτιστοποίησης και επομένως αποτελεί αντικείμενο της επιστήμης των Μαθηματικών. Όταν οι διακριτές μονάδες ενός συστήματος συμπεριφέρονται λογικά με σκοπό να βελτιστοποιήσουν την ωφέλειά τους, τότε η αλληλεπίδρασή τους μπορεί να θεωρηθεί ως ένα παίγνιο και επομένως η Θεωρία Ουρών συνδέεται με τη Θεωρία Παιγνίων και γενικότερα με ιδέες της Οικονομικής Επιστήμης. Τέλος, τα θέματα του συνωστισμού που μελετάει η Θεωρία Ουρών έχουν εφαρμογές σε προβλήματα Μηχανικών που σχετίζονται με τηλεπικοινωνιακά δίκτυα, τηλεφωνικά δίκτυα, συγκοινωνιακά δίκτυα κ.α. Επομένως, η Θεωρία Ουρών απασχολεί την κοινότητα των Μηχανικών, καθώς και των Επιστημόνων Πληροφορικής.

Από τα παραπάνω γίνεται φανερό ότι η Θεωρία Ουρών είναι σήμερα ένα ιδιαίτερα ζωντανό διεπιστημονικό αντικείμενο που συνεχίζει να αναπτύσσεται. Τα βασικά μοντέλα της είναι κομψά μαθηματικά αντικείμενα που έχουν μελετηθεί εκτενώς και συνεχίζουν να μελετώνται. Από την άποψη των εφαρμογών, φαίνεται ότι είναι σημαντικό να συνεχιστεί η οικονομική μελέτη των συστημάτων αναμονής, σε όσο το δυνατόν ευρύτερο πλαίσιο. Επίσης, είναι σημαντικό να αναπτυχθούν υπολογιστικές μέθοδοι για τη μελέτη πολύπλοκων συστημάτων αναμονής που δεν επιδέχονται αναλυτικές λύσεις και για τον σκοπό αυτό είναι επιθυμητή η ανάπτυξη αριθμητικών μεθόδων και μεθόδων προσομοίωσης.

Η κύρια κριτική που ασκείται στη Θεωρία Ουρών είναι ότι η πραγματικότητα είναι πολύ πιο περίπλοκη και οι υποθέσεις που συνήθως τίθενται στα μοντέλα (Poisson διαδικασίες, απλές κατανομές χρόνων εξυπηρέτησης, όμοιοι υπηρέτες, ομογενείς πελάτες κ.α.) είναι εξαιρετικά περιοριστικές και τελικά τα μαθηματικά αποτελέσματα που προκύπτουν δεν μπορούν να εφαρμοστούν σε πραγματικά προβλήματα. Ασφαλώς τα μοντέλα αποτελούν απλοποιήσεις των αντίστοιχων πραγματικών καταστάσεων και επομένως τα αποτελέσματα και οι διαισθήσεις που προκύπτουν από αυτά θα πρέπει να χρησιμοποιούνται με προσοχή. Ισχύει, όμως, ο παραλλαγμένος αφορισμός του σπουδαίου στατιστικού George Box: Όλα τα μοντέλα είναι λάθος, μερικά όμως είναι χρήσιμα. Και τα μοντέλα της Θεωρίας Ουρών έχουν αποδειχθεί εξαιρετικά χρήσιμα.

Τα μοντέλα των ουρών αναμονής έχουν χρησιμοποιηθεί ευρύτατα για πάνω από 100 χρόνια και θα συνεχίσουν να χρησιμοποιούνται. Θα πρέπει, όμως, η θεωρία να αγκαλιάζει νέες επιστημονικές προσεγγίσεις, καθώς τα διαθέσιμα εργαλεία εξελίσσονται. Φαίνεται ότι στο άμεσο μέλλον η πρόκληση είναι να δημιουργηθεί μια Θεωρία Ουρών οδηγούμενη από τα δεδομένα (Data-driven Queueing Theory). Προς την κατεύθυνση αυτή υπάρχουν ήδη κάποιες δημοσιεύσεις, ενώ από τις 21 έως τις 23 Σεπτεμβρίου 2021 έγινε το πρώτο σχετικό συνέδριο (Workshop on Data Driven Queueing Systems) με 18 ομιλίες και πάνω από 300 συμμετέχοντες από 25 χώρες (βλέπε <https://www.ddqc.io/> όπου περιέχονται και τα βίντεο των ομιλιών).

Το 2022 δημοσιεύτηκε ο επετειακός τόμος 100 του περιοδικού Queueing Systems που περιλαμβάνει περίπου 100 τρισελίδα άρθρα από ερευνητές στην περιοχή της Θεωρίας Ουρών Αναμονής. Τα άρθρα αυτά περιγράφουν τις προκλήσεις και τα σημαντικότερα ανοικτά προβλήματα που φαίνεται ότι θα απασχολήσουν την ερευνητική κοινότητα της Θεωρίας Ουρών Αναμονής τα επόμενα χρόνια.

Βιβλιογραφία

- [1] A.K. Erlang. "Sandsynlighedsregning og Telefonsamtaler (Probability calculation and telephone conversations)". Στο: *Nyt Tidsskrift for Matematik* 20.B (1909), σσ. 33–39.

- [2] D.G. Kendall. “Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain”. Στο: *The Annals of Mathematical Statistics* 24 (1953), σσ. 338–354.
- [3] A.K. Erlang. “Løsning af nogle Problemer fra Sandsynlighedsregningen af Betydning for de automatiske Telefoncentraler (Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges)”. Στο: *Elektroteknikeren* 13 (1917), σσ. 5–13.
- [4] A.K. Erlang. “Telefon-Ventetider. Et Stykke Sandsynlighedsregning (Telephone Waiting Times. A Bit of Probability Calculation)”. Στο: *Matematisk Tidsskrift B* 31 (1920), σσ. 25–42.
- [5] F. Pollaczek. “Ueber eine Aufgabe der Wahrscheinlichkeitstheorie”. Στο: *Math. Z.* 32 (1930), σσ. 64–100.
- [6] D.G. Kendall. “Some problems in the theory of queues”. Στο: *Journal of the Royal Statistical Society, Series B* 13 (1951), 151–173, discussion 173–185.
- [7] J.W. Cohen. *The Single Server Queue*. Amsterdam: North-Holland, 1969. ISBN: 978-0720423587.
- [8] M.F. Neuts. *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*. The Johns Hopkins University Press, 1981. ISBN: 978-0801825606.
- [9] M.F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. CRC Press, 1989. ISBN: 978-0824782832.
- [10] G. Latouche και V. Ramaswami. *Introduction to Matrix Analytic Models in Stochastic Modeling*. ASA-SIAM Series on Statistics και Applied Probability, 1999. ISBN: 978-0898714258.
- [11] J.R. Jackson. “Networks of waiting lines”. Στο: *Operations Research* 5 (1957), σσ. 518–521.
- [12] J.R. Jackson. “Jobshop-like queueing systems”. Στο: *Management Science* 10 (1963), σσ. 131–142.
- [13] F.P. Kelly. *Reversibility and Stochastic Networks*. Chichester: Wiley, 1979. ISBN: 978-0471276012.
- [14] P. Whittle. *Systems in Stochastic Equilibrium*. Chichester: Wiley, 1986. ISBN: 978-0471908876.
- [15] R. Serfozo. *Introduction to Stochastic Networks*. Springer, 1999. ISBN: 978-0387987736.
- [16] X. Chao, M. Miyazawa και M. Pinedo. *Queueing Networks: Customers, Signals and Product Form Solutions*. Wiley, 1999. ISBN: 978-0471985706.
- [17] P. Naor. “The regulation of queue size by levying tolls”. Στο: *Econometrica* 37 (1969), σσ. 15–24.
- [18] R. Hassin και M. Haviv. *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*. Kluwer, 2003. ISBN: 978-1402072031.
- [19] S.Jr. Stidham. *Optimal Design of Queueing Systems*. CRC Press, 2009. ISBN: 978-1584880769.
- [20] R. Hassin. *Rational Queueing*. CRC Press, 2016. ISBN: 978-1498745277.
- [21] A. Economou. “The Impact of Information Structure on Strategic Behavior in Queueing Systems”. Στο: *Anisimov V. and Limnios, N. (eds.) Queueing Theory 2*. John Wiley και Sons, Ltd, 2021. Κεφ. 4, σσ. 137–169. ISBN: 978-1119755234. DOI: <https://doi.org/10.1002/9781119755234.ch4>.
- [22] A. Economou. “How much information should be given to the strategic customers of a queueing system?” Στο: *Queueing Systems* 100.3-4 (2022), σσ. 421–423. DOI: 10.1007/s11134-022-09741-2.
- [23] S.Jr Stidham. “Analysis, design and control of queueing systems”. Στο: *Operations Research* 50 (2002), σσ. 197–216.
- [24] G. Koole. *Call Center Optimization*. MG Books, 2013. ISBN: 978-9082017908.

- [25] L. Green. “Queueing analysis in healthcare”. Στο: *Hall, R.W. (ed.) Patient flow: Reducing Delay in Healthcare Delivery*. Springer, 2006. Κεφ. 10, σσ. 281–307. ISBN: 978-1461495116.
- [26] P.M. Morse. *Queues, Inventories and Maintenance; the Analysis of Operational Systems with Variable Demand and Supply*. Wiley, 1958.
- [27] W. Karush. “A queueing model for an inventory problem”. Στο: *Operations Research* 5 (1957), σσ. 693–703.
- [28] S.C. Graves. “The applications of queueing theory to continuous perishable inventory systems”. Στο: *Management Science* 28 (1982), σσ. 400–406.
- [29] S. Nahmias και T. Lennon Olsen. *Production and Operations Analysis, 7th Edition*. Long Groves, Illinois: Waveland Press, Inc., 2015. ISBN: 978-1478623069.
- [30] W.J. Hopp και M.L. Spearman. *Factory Physics, 3rd Edition*. Waveland Press Inc., 2011. ISBN: 978-1577667391.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Βιβλιογραφία

- [1] Σ. Κάντα. *Αλγοριθμικές Τεχνικές στις Μαρκοβιανές Αλυσίδες: Μέθοδοι Πιθανογεννητριών και Εφαρμογές*. Αθήνα: Διπλωματική Εργασία, Μεταπτυχιακό Πρόγραμμα Ειδίκευσης στη Στατιστική και Επιχειρησιακή Έρευνα, Τμήμα Μαθηματικών Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών, 2006.
- [2] Ν. Τσουπαρόπουλος. *Ανάλυση Μέσης Τιμής σε Συστήματα Εξυπηρέτησης*. Αθήνα: Διπλωματική Εργασία, Μεταπτυχιακό Πρόγραμμα Ειδίκευσης στη Στατιστική και Επιχειρησιακή Έρευνα, Τμήμα Μαθηματικών Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών, 2015.
- [3] Δ. Φακίνος. *Ουρές Αναμονής: Θεωρία και Ασκήσεις*. Αθήνα: Εκδόσεις Συμμετρία, 2003. ISBN: 978-960-266-206-9.
- [4] J. Abate, G.L. Choudhury και W. Whitt. “An introduction to numerical transform inversion and its application to probability models”. Στο: *Grassmann, W.K. (ed.) Computational Probability*. Kluwer, 2000. Κεφ. 8, σσ. 257–323. ISBN: 978-0792386179. DOI: https://doi.org/10.1007/978-1-4757-4828-4_8.
- [5] J. Abate και W. Whitt. “Numerical inversion of probability generating functions”. Στο: *Operations Research Letters* 12 (1992), σσ. 275–281.
- [6] J. Abate και W. Whitt. “Numerical inversion of Laplace transforms of probability distributions”. Στο: *ORSA Journal on Computing* 7 (1995), σσ. 36–43.
- [7] I. Adan και J. Resing. *Queueing Theory*. Eindhoven: Notes available online, 2001.
- [8] I. Adan και J. van der Wal. *Difference and Differential Equations in Stochastic Operations Research*. Eindhoven: Notes available online, 1998.
- [9] I. Adan και J. van der Wal. “Mean value techniques”. Στο: *Boucherie, R. and van Dijk, N.M. (eds.) Queueing Networks: A Fundamental Approach*. Springer, 2011. Κεφ. 13, σσ. 561–586. ISBN: 978-1-4419-6471-7. DOI: https://doi.org/10.1007/978-1-4419-6472-4_13.
- [10] L.V. Ahlfors. *Complex Analysis: An Introduction to the Theory of Analytic Functions of One Complex Variable, 3rd Edition*. New York: McGraw-Hill, Inc., 1979.
- [11] J.R. Artalejo και J.A.C. Resing. “Mean value analysis of single server retrial queues”. Στο: *Asian-Pacific Journal of Operational Research* 27 (2010), σσ. 335–345.
- [12] S. Asmussen. *Applied Probability and Queues*. Wiley, 1987. ISBN: 978-0471911739.
- [13] F. Baccelli και P. Bremaud. *Elements of Queueing Theory. Palm Martingale Calculus and Stochastic Recurrences, 2nd Edition*. New York: McGraw-Hill, Inc., 1979.
- [14] C. Bell και S. Stidham. “Individual versus social optimization in the allocation of customers to alternative servers”. Στο: *Management Science* 29 (1983), σσ. 831–839.
- [15] D.P. Bertsekas και J.N. Tsitsiklis. *Introduction to Probability, 2nd Edition*. Athena Scientific, 2008. ISBN: 978-1886529236.
- [16] G. Bolch, S. Greiner, H. de Meer και K.S. Trivedi. *Queueing Networks and Markov Chains*. Wiley, 1998. ISBN: 978-0471193666.
- [17] A. Burnetas και A. Economou. “Equilibrium customer strategies in a single server Markovian queue with setup times”. Στο: *Queueing Systems* 56 (2007), σσ. 213–228.
- [18] A. Burnetas, A. Economou και G. Vasiliadis. “Strategic customer behavior in a queueing system with delayed observations”. Στο: *Queueing Systems* 86 (2017), σσ. 389–418.
- [19] C.G. Cassandras και S. Lafortune. *Introduction to Discrete Event Systems, 2nd Edition*. Springer, 2008.

- [20] X. Chao, M. Miyazawa και M. Pinedo. *Queueing Networks: Customers, Signals and Product Form Solutions*. Wiley, 1999. ISBN: 978-0471985706.
- [21] H. Chen και M. Frank. “Monopoly pricing when customers queue”. Στο: *IIE Transactions* 36 (2004), σσ. 569–581.
- [22] H. Chen και D.D. Yao. *Fundamentals of Queueing Networks*. Springer, 2001. ISBN: 978-1475753011.
- [23] J.W. Cohen. *The Single Server Queue*. Amsterdam: North-Holland, 1969. ISBN: 978-0720423587.
- [24] D.R. Cox. “The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables”. Στο: *Mathematical Proceedings of the Cambridge Philosophical Society* 51 (1955), σσ. 433–441.
- [25] Y. Dimitrakopoulos, A. Economou και S. Leonardos. “Strategic customer behavior in a queueing system with alternating information structure”. Στο: *European Journal of Operational Research* 291 (2021), σσ. 1024–1040.
- [26] A. Economou. “The Impact of Information Structure on Strategic Behavior in Queueing Systems”. Στο: *Anisimov V. and Limnios, N. (eds.) Queueing Theory 2*. John Wiley και Sons, Ltd, 2021. Κεφ. 4, σσ. 137–169. ISBN: 978-1119755234. DOI: <https://doi.org/10.1002/9781119755234.ch4>.
- [27] A. Economou. “How much information should be given to the strategic customers of a queueing system?” Στο: *Queueing Systems* 100.3-4 (2022), σσ. 421–423. DOI: 10.1007/s11134-022-09741-2.
- [28] A. Economou και S. Kanta. “Optimal balking strategies and pricing for the single server Markovian queue with compartmented waiting space”. Στο: *Queueing Systems* 59 (2008), σσ. 237–269.
- [29] A. Economou και A. Manou. “Equilibrium balking strategies for a clearing queueing system in alternating environment”. Στο: *Annals of Operations Research* 208 (2013), σσ. 489–514.
- [30] N. M. Edelson και K. Hildebrand. “Congestion tolls for Poisson queueing processes”. Στο: *Econometrica* 43 (1975), σσ. 81–92.
- [31] M. El-Taha και S.Jr. Stidham. *Sample-path Analysis of Queueing Systems*. Kluwer, 1998. ISBN: 978-0792382102.
- [32] S. Elaydi. *An Introduction to Difference Equations, 3rd Edition*. Springer, 2006. ISBN: 978-0387230597.
- [33] T. Engset. “Die Wahrscheinlichkeitsrechnung zur Bestimmung der Wähleranzahl in automatischen Fernsprechämtern”. Στο: *Elektrotechn. Zeitschrift* 31 (1918), σσ. 304–306.
- [34] A.K. Erlang. “Sandsynlighedsregning og Telefonsamtaler (Probability calculation and telephone conversations)”. Στο: *Nyt Tidsskrift for Matematik* 20.B (1909), σσ. 33–39.
- [35] A.K. Erlang. “Løsning af nogle Problemer fra Sandsynlighedsregningen af Betydning for de automatiske Telefoncentraler (Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges)”. Στο: *Elektroteknikerne* 13 (1917), σσ. 5–13.
- [36] A.K. Erlang. “Telefon-Ventetider. Et Stykke Sandsynlighedsregning (Telephone Waiting Times. A Bit of Probability Calculation)”. Στο: *Matematisk Tidsskrift B* 31 (1920), σσ. 25–42.
- [37] H.R. Gail, S.L. Hantler και B.A. Taylor. “Spectral analysis of M/G/1 and G/M/1 type Markov chains”. Στο: *Advances in Applied Probability* 28 (1996), σσ. 114–165.
- [38] H.R. Gail, S.L. Hantler και B.A. Taylor. “Use of characteristics roots for solving infinite state Markov chains”. Στο: *Grassmann, W. (ed) Computational Probability*. Springer, 2000. Κεφ. 7, σσ. 205–255. ISBN: 978-0-7923-8617-9. DOI: https://doi.org/10.1007/978-1-4757-4828-4_7.
- [39] E. Gelenbe και G. Pujolle. *Introduction to Queueing Networks, 2nd Edition*. Chichester: Wiley, 1998. ISBN: 978-0471962945.

- [40] S.C. Graves. “The applications of queueing theory to continuous perishable inventory systems”. Στο: *Management Science* 28 (1982), σσ. 400–406.
- [41] L. Green. “Queueing analysis in healthcare”. Στο: *Hall, R.W. (ed.) Patient flow: Reducing Delay in Healthcare Delivery*. Springer, 2006. Κεφ. 10, σσ. 281–307. ISBN: 978-1461495116.
- [42] G. Grimmett και D. Stirzaker. *Probability and Random Processes, 4th Edition*. Oxford, 2020. ISBN: 978-0198847595.
- [43] P. Guo και P. Zipkin. “The effects of the availability of waiting-time information on a balking queue”. Στο: *European Journal of Operational Research* 198 (2009), σσ. 199–209.
- [44] M. Harchol-Balter. *Performance Modeling and Design of Computer Systems: Queueing Theory in Action*. Cambridge, 2013. ISBN: 978-1107027503.
- [45] R. Hassin και M. Haviv. *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*. Kluwer, 2003. ISBN: 978-1402072031.
- [46] R. Hassin. “Consumer information in markets with random products quality: The case of queues and balking”. Στο: *Econometrica* 54 (1986), σσ. 1185–1195.
- [47] R. Hassin. *Rational Queueing*. CRC Press, 2016. ISBN: 978-1498745277.
- [48] R. Hassin και A. Koshman. “Profit maximization in the M/M/1 queue”. Στο: *Operations Research Letters* 45 (2017), σσ. 436–441.
- [49] R. Hassin και R. Roet-Green. “On queue-length information when customers travel to a queue”. Στο: *Manufacturing & Service Operations Management* 23 (2020), σσ. 989–1004.
- [50] M. Haviv. *Queues: A Course in Queueing Theory*. Springer, 2013. ISBN: 978-1461467649.
- [51] M. Haviv. “Regulating an M/G/1 queue when customers know their demand”. Στο: *Performance Evaluation* 77 (2014), σσ. 57–71.
- [52] M. Haviv και B. Oz. “Self-regulation of an unobservable queue”. Στο: *Management Science* 64 (2018), σσ. 2380–2389.
- [53] W.J. Hopp και M.L. Spearman. *Factory Physics, 3rd Edition*. Waveland Press Inc., 2011. ISBN: 978-1577667391.
- [54] M. Hu, Y. Li και J. Wang. “Efficient ignorance: Information heterogeneity in a queue”. Στο: *Management Science* 64 (2018), σσ. 2650–2671.
- [55] R. Ibrahim. “Sharing delay information in service systems: a literature survey”. Στο: *Queueing Systems* 89 (2018), σσ. 49–79. DOI: 10.1007/s11134-018-9577-y.
- [56] J.R. Jackson. “Networks of waiting lines”. Στο: *Operations Research* 5 (1957), σσ. 518–521.
- [57] J.R. Jackson. “Jobshop-like queueing systems”. Στο: *Management Science* 10 (1963), σσ. 131–142.
- [58] M.A. Johnson. “An empirical study of queueing approximations based on phase-type approximations”. Στο: *Stochastic Models* 9 (1993), σσ. 531–561.
- [59] E.P.C. Kao. *An Introduction to Stochastic Processes*. New York: Duxbury Press, 1997. ISBN: 978-0534255183.
- [60] W. Karush. “A queueing model for an inventory problem”. Στο: *Operations Research* 5 (1957), σσ. 693–703.
- [61] F.P. Kelly. *Reversibility and Stochastic Networks*. Chichester: Wiley, 1979. ISBN: 978-0471276012.
- [62] D.G. Kendall. “Some problems in the theory of queues”. Στο: *Journal of the Royal Statistical Society, Series B* 13 (1951), 151–173, discussion 173–185.

- [63] D.G. Kendall. “Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain”. Στο: *The Annals of Mathematical Statistics* 24 (1953), σσ. 338–354.
- [64] J.F.C. Kingman. “The first Erlang century - and the next”. Στο: *Queueing Systems* 63 (2009), σσ. 3–12.
- [65] L. Kleinrock. *Queueing Systems. Volume 1: Theory*. Wiley, 1975. ISBN: 978-0471491101.
- [66] L. Kleinrock. *Queueing Systems. Volume 2: Computer Applications*. Wiley, 1976. ISBN: 978-0471491118.
- [67] G. Koole. “Monotonicity in Markov Reward and Decision Chains: Theory and Applications”. Στο: *Foundations and Trends in Stochastic Systems* 1.1 (2007), σσ. 1–76.
- [68] G. Koole. *Call Center Optimization*. MG Books, 2013. ISBN: 978-9082017908.
- [69] V.G. Kulkarni. *Modeling and Analysis of Stochastic Systems, 2nd Edition*. CRC Press, 2010. ISBN: 978-1439808757.
- [70] G. Latouche και V. Ramaswami. *Introduction to Matrix Analytic Models in Stochastic Modeling*. ASA-SIAM Series on Statistics και Applied Probability, 1999. ISBN: 978-0898714258.
- [71] J.D.C. Little. “A proof of the queueing formula $L = \lambda W$ ”. Στο: *Operations Research* 9 (1961), σσ. 383–387.
- [72] P.M. Morse. *Queues, Inventories and Maintenance; the Analysis of Operational Systems with Variable Demand and Supply*. Wiley, 1958.
- [73] S. Nahmias και T. Lennon Olsen. *Production and Operations Analysis, 7th Edition*. Long Groves, Illinois: Waveland Press, Inc., 2015. ISBN: 978-1478623069.
- [74] P. Naor. “The regulation of queue size by levying tolls”. Στο: *Econometrica* 37 (1969), σσ. 15–24.
- [75] M.F. Neuts. *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*. The Johns Hopkins University Press, 1981. ISBN: 978-0801825606.
- [76] M.F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. CRC Press, 1989. ISBN: 978-0824782832.
- [77] F. Pollaczek. “Ueber eine Aufgabe der Wahrscheinlichkeitstheorie”. Στο: *Math. Z.* 32 (1930), σσ. 64–100.
- [78] S. Ross. *Stochastic Processes, 2nd Edition*. Wiley, 1995. ISBN: 978-0471120629.
- [79] S. Ross. *Introduction to Probability Models, 12th Edition*. Academic Publishers, 2019. ISBN: 978-0128143469.
- [80] L.I. Sennott. *Stochastic Dynamic Programming and the Control of Queueing Systems*. Wiley-Interscience, 1998. ISBN: 978-0471161202.
- [81] R. Serfozo. *Introduction to Stochastic Networks*. Springer, 1999. ISBN: 978-0387987736.
- [82] K. Sigman. *Stationary Marked Point Processes: An Intuitive Approach*. Chapman και Hall/CRC, 1995. ISBN: 978-0412984310.
- [83] J.F. Shortle, J.M. Thompson, D. Gross και C.M. Harris. *Fundamentals of Queueing Theory, 5th Edition*. Wiley, 2018. ISBN: 978-1118943526.
- [84] S.Jr Stidham. “ $L = \lambda W$: A discounted analogue and a new proof”. Στο: *Operations Research* 20 (1972), σσ. 1115–1120.
- [85] S.Jr Stidham. “A last word on $L = \lambda W$ ”. Στο: *Operations Research* 22 (1974), σσ. 417–421.

- [86] S.Jr Stidham. “Optimal control of admission to a queueing system”. Στο: *IEEE Transactions on Automatic Control* 30.8 (1985), σσ. 705–713.
- [87] S.Jr Stidham. “Analysis, design and control of queueing systems”. Στο: *Operations Research* 50 (2002), σσ. 197–216.
- [88] S.Jr. Stidham. *Optimal Design of Queueing Systems*. CRC Press, 2009. ISBN: 978-1584880769.
- [89] S.Jr. Stidham και R. Weber. “A survey of Markov decision models for control of networks of queues”. Στο: *Queueing Systems* 13 (1993), σσ. 291–314.
- [90] H.C. Tijms. *A First Course in Stochastic Models, 2nd Edition*. Chichester: Wiley, 2008. ISBN: 978-0471498803.
- [91] N.M. van Dijk. *Queueing Networks and Product Forms: A Systems Approach*. Chichester: Wiley, 1993. ISBN: 978-0471928488.
- [92] W. Whitt. “Approximating a point process by a renewal process I: two basic methods”. Στο: *Operations Research* 30 (1986), σσ. 125–147.
- [93] P. Whittle. *Systems in Stochastic Equilibrium*. Chichester: Wiley, 1986. ISBN: 978-0471908876.
- [94] H.S. Wilf. *Generatingfunctionology, 3rd Edition*. AK Peters/CRC, 2005. ISBN: 978-1568812793.
- [95] E.M.M. Winands, I.J.B.F. Adan και G.J.J.A.N. van Houtum. “Mean value analysis for polling systems”. Στο: *Queueing Systems: Theory and Applications* 54.1 (2006), σσ. 35–44.
- [96] R.W. Wolff. “Poisson arrivals see time averages”. Στο: *Operations Research* 30 (1982), σσ. 223–231.
- [97] R.W. Wolff. *Stochastic Modeling and the Theory of Queues*. Englewood Cliffs, New Jersey: Prentice Hall, 1989. ISBN: 978-0138466923.