

Ανάλυση Μέσης Τιμής

Ανάλυση Μέσης Τιμής - Μεθοδολογία

- Ανάλυση Μέσης Τιμής (AMT) $\rightarrow$ Γρήγορος υπολογισμός των $E[Q]$ και $E[S]$ ενός συστήματος.
- Βασική ιδέα: Επίλυση συστήματος δυο εξισώσεων με δυο αγνώστους.
- 1η εξίσωση: Ο N. Little:

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

- 2η εξίσωση: Από τον υπολογισμό του χρόνου παραμονής πελάτη, S , δεσμεύοντας στον αριθμό πελατών που βρίσκει ο πελάτης + PASTA:

$$E[S] = \sum_j \Pr[Q^- = j] E[S|Q^- = j]$$

$$\rightarrow E[S] = f(E[Q^-]) \stackrel{PASTA}{=} f(E[Q]).$$

Ανάλυση Μέσης Τιμής - Επεκτάσεις

- Εφαρμόζοντας τον N. Little σε διάφορα υποσυστήματα ενός συστήματος εξυπηρέτησης, μπορούμε να βρούμε όχι μόνο τις μέσες τιμές $E[Q]$, $E[S]$, αλλά και ολόκληρη την κατανομή (p_j) .

AMT - M/G/1/1 ουρά

- N. Little:

$$E[Q] = \lambda E[S] \Rightarrow p_1 = \lambda E[S].$$

- Δέσμευση για $E[S]$ στην Q^- :

$$\begin{aligned} E[S] &= \Pr[Q^- = 0]E[S|Q^- = 0] + \Pr[Q^- = 1]E[S|Q^- = 1] \\ &= a_0 \cdot b + a_1 \cdot 0 \stackrel{PASTA}{=} bp_0. \end{aligned}$$

- Άρα $p_1 = \lambda bp_0 = \rho p_0$. Επίσης $p_0 + p_1 = 1$, οπότε:

$$p_0 = \frac{1}{1 + \rho}, \quad p_1 = \frac{\rho}{1 + \rho}, \quad E[Q] = \frac{\rho}{1 + \rho}, \quad E[S] = \frac{1}{\mu(1 + \rho)}.$$

AMT - M/M/1 ουρά - Μέσες τιμές

- N. Little:

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

- Δέσμευση για $E[S]$ στην Q^- :

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

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.

AMT - M/M/1 ουρά - Μέσες τιμές (συνέχεια)

- N. Little:

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

- Δέσμευση για $E[S]$ στην Q^- + Αμνήμονη ιδιότητα εκθετικής + PASTA:

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

- Λύνοντας το σύστημα:

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

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

AMT - M/M/1 ουρά - Κατανομή ισορροπίας

- Εφαρμόζοντας N. Little στη θέση j του συστήματος:
Μέσο πλήθος πελατών στην θέση j
= Ρυθμός αφίξεων στη θέση j
× Μέσος χρόνος παραμονής στη θέση j .
- Μέσο πλήθος πελατών στην θέση j
= Πιθ. υπάρχει πελάτης στη θέση j $= \sum_{k=j}^{\infty} p_k$.
- Ρυθμός αφίξεων στη θέση j
= Ρυθμός αφίξεων × ποσοστό πελατών που περνάνε από τη θέση j
 $= \lambda \sum_{k=j-1}^{\infty} a_k \stackrel{\text{PASTA}}{=} \lambda \sum_{k=j-1}^{\infty} p_k$.
- Μέσος χρόνος παραμονής στη θέση $j = \frac{1}{\mu}$.
- Άρα

$$\sum_{k=j}^{\infty} p_k = \rho \sum_{k=j-1}^{\infty} p_k.$$

AMT - M/M/1 ουρά - Κατανομή ισορροπίας (συνέχεια)

- Αφαιρούμε την εξίσωση για την κατάσταση $j + 1$ από την εξίσωση για την κατάσταση j :

$$\begin{aligned}\sum_{k=j}^{\infty} p_k &= \rho \sum_{k=j-1}^{\infty} p_k \quad (\text{εξίσωση για } j) \\ - \sum_{k=j+1}^{\infty} p_k &= \rho \sum_{k=j}^{\infty} p_k \quad (\text{εξίσωση για } j+1)\end{aligned}$$

- Άρα:

$$p_j = \rho p_{j-1}, \quad j \geq 1.$$

- Επίσης, για $j = 1$ έχουμε $1 - p_0 = \rho$, οπότε

$$p_j = (1 - \rho)\rho^j, \quad j \geq 0.$$

AMT - M/G/1 ουρά - Μέσες τιμές

- N. Little:

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

- Δέσμευση για $E[S]$ στην Q^- : $S = \sum_{i=1}^{Q^-+1} B_i$, αλλά ο B_1 εξαρτάται από το Q^- .
- Οπότε δεσμεύουμε στο κατά πόσον $Q^- = 0$ ή $Q^- > 0$:

$$\begin{aligned} E[S] &= \Pr[Q^- = 0] E[B_1 | Q^- = 0] \\ &\quad + \Pr[Q^- > 0] E[B_1 | Q^- > 0] \\ &\quad + \Pr[Q^- > 0] E \left[\sum_{i=2}^{Q^-+1} B_i | Q^- > 0 \right]. \end{aligned}$$

AMT - M/G/1 ουρά - Μέσες τιμές (συνέχεια)

- Έχουμε:

$$\begin{aligned} E[S] &= \Pr[Q^- = 0] E[B_1 | Q^- = 0] \\ &\quad + \Pr[Q^- > 0] E[B_1 | Q^- > 0] \\ &\quad + \Pr[Q^- > 0] E \left[\sum_{i=2}^{Q^-+1} B_i | Q^- > 0 \right]. \end{aligned}$$

- $\Pr[Q^- = 0] = 1 - \rho$ (Πόρισμα N. Little - PASTA),
 $\Pr[Q^- > 0] = \rho$.
- $E[B_1 | Q^- = 0] = b$.
- $E \left[\sum_{i=2}^{Q^-+1} B_i | Q^- > 0 \right] = E[Q^- | Q^- > 0] b$.
- $E[B_1 | Q^- > 0] = E[R_B] = \frac{E[B^2]}{2E[B]} = \frac{b_2}{2b} = \frac{b^2 + \sigma_B^2}{2b}$ (μέσος υπολειπόμενος ανανεωτ. διαδικασίας με ενδιαμέσους χρόνους τους χρόνους εξυπηρέτησης).

AMT - M/G/1 ουρά - Μέσες τιμές (συνέχεια)

- Τελικά:

$$\begin{aligned}
 E[S] &= \Pr[Q^- = 0]E[B_1|Q^- = 0] \\
 &\quad + \Pr[Q^- > 0]E[B_1|Q^- > 0] \\
 &\quad + \Pr[Q^- > 0]E\left[\sum_{i=2}^{Q^-+1} B_i|Q^- > 0\right] \\
 &= (1 - \rho)b + \rho E[R_B] + \Pr[Q^- > 0]E[Q^-|Q^- > 0]b \\
 &= (1 - \rho)b + \rho E[R_B] + E[Q^- 1_{\{Q^- > 0\}}]b \\
 &= (1 - \rho)b + \rho E[R_B] + E[Q^-]b \\
 &= (1 - \rho)b + \rho E[R_B] + E[Q]b.
 \end{aligned}$$

- Άρα:

$$E[Q] = (1 - \rho)\lambda b + \lambda \rho E[R_B] + E[Q]\lambda b.$$

AMT - M/G/1 ουρά - Μέσες τιμές (συνέχεια)

- Επομένως:

$$E[Q] = \rho + \frac{\rho}{1 - \rho} \lambda E[R_B],$$

και διαιρώντας με λ παίρνουμε και τον μέσο χρόνο παραμονής πελάτη στο σύστημα

$$E[S] = b + \frac{\rho}{1 - \rho} E[R_B].$$

- Εύκολος μνημονικός τρόπος:

Μέσος χρόνος αναμονής στην M/G/1 ουρά = Μέσος αριθμός πελατών στην M/M/1 ουρά $\times$ μέση τιμή κατανομής ισορροπίας των χρόνων εξυπηρέτησης:

$$E[W] = \frac{\rho}{1 - \rho} E[R_B].$$

Σύγκριση $M/M/c$ και αντίστοιχων $M/M/c$ συστημάτων

- Κοινή ουρά για όλους τους υπηρέτες ή μια ουρά για κάθε υπηρέτη;
- Πολλοί αργοί ή λίγοι γρήγοροι υπηρέτες;

Κοινή ουρά για όλους τους υπηρέτες ή μια ουρά για κάθε υπηρέτη;

- Πρόκειται να σχεδιάσουμε σύστημα εξυπηρέτησης με c όμοιους υπηρέτες.
- Είναι προτιμότερο να έχουμε ξεχωριστή ουρά για κάθε υπηρέτη ή μια κοινή ουρά για όλους (pooling);
- Συγκεκριμένο πλαίσιο:
Poisson διαδικασία αφίξεων.
2 διαθέσιμοι υπηρέτες με $\text{Exp}(\mu)$ χρόνους εξυπηρέτησης.
Απεριόριστη χωρητικότητα.
- Σύστημα 1: Δυο παράλληλες M/M/1 ουρές με ρυθμό αφίξεων $\frac{\lambda}{2}$ και ρυθμό μ εξυπηρέτησης έκαστη.
- Σύστημα 2: Μια M/M/2 ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ .

Κοινή ουρά για όλους τους υπηρέτες ή μια ουρά για κάθε υπηρέτη; (συνέχεια)

- Θέτουμε $\rho = \frac{\lambda}{\mu}$.
- Σύστημα 1: Δυο παράλληλες M/M/1 ουρές με ρυθμό αφίξεων $\frac{\lambda}{2}$ και ρυθμό μ έκαστη. Μέσος χρόνος παραμονής:

$$E[S_1] = \frac{1}{\mu - \frac{\lambda}{2}} = \frac{2}{\mu(2 - \rho)}.$$

- Σύστημα 2: Μια M/M/2 ουρά με ρυθμό αφίξεων λ και ρυθμό μ . Μέσος χρόνος παραμονής:

$$E[S_2] = \frac{4}{\mu(2 + \rho)(2 - \rho)}.$$

Κοινή ουρά για όλους τους υπηρέτες ή μια ουρά για κάθε υπηρέτη; (συνέχεια)

- Έχουμε

$$\begin{aligned} E[S_1] - E[S_2] &= \frac{2}{\mu(2-\rho)} - \frac{4}{\mu(2+\rho)(2-\rho)} \\ &= \frac{2\rho}{\mu(2+\rho)(2-\rho)} > 0. \end{aligned}$$

- Συμπέρασμα: Προτιμότερο να υπάρχει μια κοινή ουρά για όλους τους υπηρέτες, αν ενδιαφερόμαστε για τον χρόνο παραμονής ενός πελάτη στο σύστημα.
- Το ίδιο ισχύει και για τον χρόνο αναμονής στην ουρά, αφού έχουμε $E[W_i] = E[S_i] - \frac{1}{\mu}$, $i = 1, 2$, οπότε

$$E[W_1] - E[W_2] = E[S_1] - E[S_2] > 0.$$

Κοινή ουρά για όλους τους υπηρέτες ή μια ουρά για κάθε υπηρέτη; (συνέχεια)

- Έχουμε:

$$\lim_{\rho \rightarrow 2^-} (E[S_1] - E[S_2]) = \lim_{\rho \rightarrow 2^-} \frac{2\rho}{\mu(2+\rho)(2-\rho)} = \infty,$$

δηλαδή η κοινή ουρά μειώνει σημαντικά τον μέσο χρόνο παραμονής σε συστήματα με ρυθμό συνωστισμού κοντά στην κρίσιμη τιμή για αστάθεια.

- Επίσης:

$$\frac{E[S_2]}{E[S_1]} = \frac{2}{2+\rho} \in \left(\frac{1}{2}, 1\right), \quad \rho \in (0, 2),$$

δηλαδή η κοινή ουρά μειώνει στο μισό τον χρόνο παραμονής για ρυθμό συνωστισμού κοντά στην κρίσιμη τιμή για αστάθεια.

Κοινή ουρά για όλους τους υπηρέτες ή μια ουρά για κάθε υπηρέτη; (συνέχεια)

- Η χρήση κοινής ουράς (pooling) βελτιώνει την απόδοση του συστήματος σε πολύ γενικότερες καταστάσεις.
- Η χρήση κοινής ουράς έχει υιοθετηθεί σε πολλές πρακτικές εφαρμογές των ουρών σε πραγματικά συστήματα τα τελευταία χρόνια.
- Αν οι πελάτες είναι ετερογενείς μπορεί να είναι προτιμότερο να έχουμε διαφορετικές ουρές για διαφορετικές κατηγορίες πελατών.
- Αν οι υπηρέτες δεν είναι όμοιοι, το ερώτημα θα πρέπει να εξεταστεί εξ αρχής με προσοχή.

Πολλοί αργοί ή λίγοι γρήγοροι υπηρέτες;

- Πρόκειται να σχεδιάσουμε σύστημα εξυπηρέτησης με συγκεκριμένη δυναμικότητα (συνολικό ρυθμό) εξυπηρέτησης.
- Είναι προτιμότερο να έχουμε έναν υπηρέτη με όλη τη δυναμικότητα ή να τη μοιράσουμε σε περισσότερους;
- Συγκεκριμένο πλαίσιο:
Poisson διαδικασία αφίξεων.
Δυνατότητα για έναν υπηρέτη με $\text{Exp}(2\mu)$ χρόνους εξυπηρέτησης ή δυο υπηρετών με $\text{Exp}(\mu)$ χρόνους εξυπηρέτησης.
- Σύστημα 1: M/M/1 ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης 2μ .
- Σύστημα 2: M/M/2 ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ ανά υπηρέτη.

Πολλοί αργοί ή λίγοι γρήγοροι υπηρέτες; (συνέχεια)

- Θέτουμε $\rho = \frac{\lambda}{\mu}$.
- Σύστημα 1: M/M/1 ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης 2μ . Μέσος χρόνος παραμονής:

$$E[S_1] = \frac{1}{2\mu - \lambda} = \frac{1}{\mu(2 - \rho)}.$$

- Σύστημα 2: M/M/2 ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ ανά υπηρέτη. Μέσος χρόνος παραμονής:

$$E[S_2] = \frac{4}{\mu(2 + \rho)(2 - \rho)}.$$

Πολλοί αργοί ή λίγοι γρήγοροι υπηρέτες; (συνέχεια)

- Έχουμε

$$\begin{aligned} E[S_1] - E[S_2] &= \frac{1}{\mu(2-\rho)} - \frac{4}{\mu(2+\rho)(2-\rho)} \\ &= \frac{\rho-2}{\mu(2+\rho)(2-\rho)} = -\frac{1}{\mu(2+\rho)} < 0. \end{aligned}$$

- Συμπέρασμα: Προτιμότερο να ανατεθεί όλη η δυναμικότητα σε έναν υπηρέτη, αν ενδιαφερόμαστε για τον χρόνο παραμονής ενός πελάτη στο σύστημα.

Πολλοί αργοί ή λίγοι γρήγοροι υπηρέτες; (συνέχεια)

- Εδώ:

$$\frac{E[S_1]}{E[S_2]} = \frac{\rho + 2}{4} \in \left(\frac{1}{2}, 1\right), \quad \rho \in (0, 2),$$

δηλαδή η ανάθεση όλης της δυναμικότητας εξυπηρέτησης σε έναν υπηρέτη βελτιώνει σημαντικά την λειτουργία του συστήματος όσο μικρότερος είναι ο συντελεστής συνωστισμού.

- Συγκεκριμένα οδηγεί σε μείωση του μέσου χρόνου παραμονής των πελατών με ένα συντελεστή που φθάνει κοντά στο $\frac{1}{2}$ για μικρές τιμές του ρυθμού συνωστισμού.

Πολλοί αργοί ή λίγοι γρήγοροι υπηρέτες; (συνέχεια)

- Για τους χρόνους αναμονής έχουμε:
Για την M/M/1 ουρά:

$$E[W_1] = E[S_1] - \frac{1}{2\mu} = \frac{\rho}{2\mu(2 - \rho)},$$

ενώ για την M/M/2 ουρά:

$$E[W_2] = E[S_2] - \frac{1}{\mu} = \frac{4}{\mu(2 + \rho)(2 - \rho)} - \frac{1}{\mu} = \frac{\rho^2}{\mu(2 + \rho)(2 - \rho)}.$$

Πολλοί αργοί ή λίγοι γρήγοροι υπηρέτες; (συνέχεια)

- Επομένως:

$$\begin{aligned}E[W_1] - E[W_2] &= \frac{\rho}{2\mu(2-\rho)} - \frac{\rho^2}{\mu(2+\rho)(2-\rho)} \\ &= \frac{\rho}{2\mu(2+\rho)} > 0.\end{aligned}$$

- Προτιμότερο να μοιραστεί η δυναμικότητα εξυπηρέτησης σε δυο υπηρέτες, αν ενδιαφερόμαστε για τον χρόνο αναμονής ενός πελάτη στην ουρά.
- Ισχύει το αντίθετο από αυτό που είδαμε για τον χρόνο παραμονής.