ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΣΗΜΕΙΩΣΕΙΣ ΠΙΘΑΝΟΤΗΤΩΝ

ΠΟΛΛΕΣ ΜΕΤΑΒΛΗΤΕΣ

Γιάννης Κοντογιάννης, Σταύρος Τουμπής

Ο.Π.Α. 2011

Σημείωση. Αυτές οι σημειώσεις είναι τρία κεφάλαια από τις σημειώσεις πιθανοτήτων των κ. Γ. Κοντογιάννη και Σ. Τουμπή. Έγιναν κάποιες μικρές αλλαγές στον συμβολισμό και στην ορολογία. Οι πλήρεις σημειώσεις υπάρχουν στις ιστοσελίδες των συγγραφέων. Τους ευχαριστώ για την άδεια να χρησιμοποιήσω αυτή τους την δουλειά.

Δημήτρης Χελιώτης

Οδηγίες Χρήσης

Το παρόν ΔΕΝ είναι διδακτικό βιβλίο. Είναι οι σημειώσεις των διαλέξεων και των φροντιστηρίων του μαθήματος «Πιθανότητες» όπως αυτό διδάσκεται στο Τμήμα Πληροφορικής του Οικονομικού Πανεπιστημίου Αθηνών. Το μάθημα διδάσκεται σε ένα εξάμηνο (το δεύτερο), και ό,τι περιέχεται εδώ καλύπτεται σε περίπου 40 δίωρες διαλέξεις (τα φροντιστήρια είναι ενσωματωμένα στις διαλέξεις). Οι σημειώσεις δίνονται ηλεκτρονικά στους φοιτητές.

Η διαμόρφωση της ύλης του μαθήματος, και συνεπώς των σημειώσεων, έχει γίνει λαμβάνοντας υπ' όψιν τις υφιστάμενες γνώσεις και το γεγονός ότι το μάθημα διδάσκεται στο δεύτερο εξάμηνο ενός τμήματος Πληροφορικής.

Δεν υπάρχει υποκατάστατο της σε βάθος μελέτης ενός καλού βιβλίου για την κατανόηση του αντικειμένου, όπως και δεν υπάρχει υποκατάστατο της φυσικής παρουσίας στο αμφιθέατρο. Συνεπώς, δεν συνιστάται στους φοιτητές, ούτε να χρησιμοποιήσουν τις σημειώσεις σαν βιβλίο, ούτε ως υποκατάστατο της παρακολούθησης. Αντίθετα, ο σκοπός τους είναι:

1. Να διευκολύνουν την παρακολούθηση των φοιτητών, μια που δεν θα χρειάζεται να αντιγράφουν ό,τι γράφεται στον πίνακα, συνήθως υπό δυσμενείς συνθήκες.

2. Να διευκολύνουν την μελέτη της ύλης που διδάχτηκε σε συνδυασμό με το βιβλίο.

3. Να βοηθούν τους φοιτητές που δεν παρακολούθησαν κάποιες διαλέξεις να μείνουν σε επαφή με το μάθημα.

Επίσης οι σημειώσεις ίσως βοηθήσουν ώστε όσοι φοιτητές βρίσκονται στο αμφιθέατρο να μην είναι εκεί απλώς για να «πάρουν τις σημειώσεις».

Το παρόν κείμενο είναι υπό διαρκή εξέλιξη. Θα εκτιμήσουμε ιδιαιτέρως οποιαδήποτε πρόταση για τη βελτίωσή του όπως και τυχόν παρατηρήσεις σχετικά με ορθογραφικά/τυπογραφικά σφάλματα, λάθη στις ασκήσεις και κάθε φύσεως πρόβλημα. Θα είναι χαρά μας επίσης αν το χρησιμοποιήσουν και άλλοι διδάσκοντες.

Γιάννης Κοντογιάννης, yiannis@aueb.gr
Σταύρος Τουμπής, toumpis@aueb.gr

Περιεχόμενα

Κεφάλαιο 1

Ζεύγη Διακριτών Τυχαίων Μεταβλητών

Η έννοια της συνάρτησης πιθανότητας μας επιτρέπει την μεμονωμένη μελέτη Τ.Μ. Σε προβλήματα όμως όπου εμφανίζονται πολλές Τ.Μ, η μεμονωμένη μελέτη κάθε μιας από αυτές δεν αποκαλύπτει τα πάντα. Φανταστείτε για παράδειγμα πως κάποιος ρίχνει δύο ζάρια με τέτοιο τρόπο ώστε

$$P((1,1)) = P((2,2)) = P((3,3)) = P((4,4)) = P((5,5)) = P((6,6)) = \frac{1}{6},$$

ενώ όλα τα άλλα ενδεχόμενα έχουν μηδενική πιθανότητα. Εύκολα μπορούμε να υπολογίσουμε ότι, μεμονωμένα, κάθε ένα από τα ζάρια είναι «δίκαιο», αφού η πιθανότητα να προκύψει οποιοδήποτε νούμερο είναι $\frac{1}{6}$. Ο συνδυασμός των ρίψεων, όμως, ΔΕΝ είναι «δίκαιος», και αυτό είναι κάτι που δεν μας αποκαλύπτουν οι συναρτήσεις πιθανοτήτων καθενός από τα δύο ζάρια! Το κενό αυτό καλύπτεται από την από κοινού συνάρτηση πιθανότητας, την συνδιακύμανση, και άλλες έννοιες που θα δούμε σε αυτό το κεφάλαιο.

1.1 Από Κοινού Συνάρτηση Πιθανότητας

Ορισμός 1.1. (Από κοινού και περιθώριες συναρτήσεις πιθανότητας) *Έστω δύο διακριτές Τ.Μ. X, Y που ορίζονται στον ίδιο χώρο πιθανότητας $(\Omega, \mathcal{A}, P)$.*

1. *Ορίζουμε την <u>από κοινού συνάρτηση πιθανότητας</u> ως την συνάρτηση $f_{X,Y}(x,y) : \mathbb{R} \times \mathbb{R} \to [0,1]$ που ορίζεται ως:*

$$f_{X,Y}(x,y) = P(X = x, Y = y), \quad \forall\, x, y \in \mathbb{R}.$$

2. *Οι συναρτήσεις πιθανότητας των X, Y καλούνται <u>περιθώριες συναρτήσεις πιθανότητας</u>.*

Συμβάσεις: 1) Σε αυτές τις σημειώσεις, όποτε σε ένα θεώρημα, παράδειγμα, ή άσκηση εμφανίζονται τουλάχιστον δύο τυχαίες μεταβλητές, θα υποθέτουμε ότι ορίζονται στον ίδιο χώρο πιθανότητας, και δεν θα το ζητάμε ρητά. Αυτό είναι απαραίτητο για να ορίζονται ποσότητες όπως $P(X = x, Y = y) = P(\{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\})$, $E(X + Y + Z), E(XY)$.

2) Για μία συνάρτηση $X : \Omega \to \mathbb{R}^n$ θέτουμε

$$S_X := X(\Omega) = \{X(\omega) : \omega \in \Omega\} \subset \mathbb{R}^n$$

το σύνολο τιμών της. Μία τέτοια X την λέμε n-διάστατη τυχαία μεταβλητή αν

$$X^{-1}(I) \in \mathcal{A} \text{ για καθε υποδιάστημα } I \text{ του } \mathbb{R}.$$

Παρατήρηση: Η (X, Y) του Ορισμού 1.1 είναι μια διδιάστατη τυχαία μεταβλητή. Τα σύνολα τιμών S_X, S_Y, των X, Y είναι αριθμήσιμα αφού είναι διακριτές τυχαίες μεταβλητές. Άρα και το σύνολο τιμών, $S_{(X,Y)}$, της (X, Y) είναι αριθμήσιμο ως υποσύνολο του $S_X \times S_Y$. Προφανώς τα μόνα (x, y) για τα οποία ενδεχομένως η $f_{X,Y}$ είναι διαφορετική από το μηδέν είναι τα στοιχεία του $S_{(X,Y)}$, γιατί για τα υπόλοιπα (x, y), το σύνολο $\{X = x, Y = y\} = \{\omega \in \Omega : (X(\omega), Y(\omega)) = (x, y)\} = \emptyset$, και άρα έχει πιθανότητα 0.

Παράδειγμα 1.1. *(Δύο προγράμματα)* Δύο διαφορετικά προγράμματα κατανέμονται τυχαία σε τρεις υπολογιστές, χωρίς κάποιο από αυτά να δείχνει προτίμηση σε κάποιον υπολογιστή, και ανεξάρτητα το ένα από το άλλο. Έστω X, Y, Z τα πλήθη των προγραμμάτων που κατέληξαν στον κάθε υπολογιστή, άρα $X + Y + Z = 2$. Ας εξετάσουμε τις δύο μεταβλητές X, Y. Η πιθανότητα να έχουμε $X = 0$ και $Y = 0$ είναι η πιθανότητα του ενδεχόμενου και τα δύο προγράμματα να πήγαν στον τρίτο υπολογιστή:

$$f_{X,Y}(0, 0) = P(X = 0, Y = 0) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}.$$

Παρομοίως, η πιθανότητα να έχουμε $X = 0$ και $Y = 1$ είναι η πιθανότητα το πρώτο πρόγραμμα να πήγε στον υπολογιστή 2 και το δεύτερο στον 3, ή αντίστροφα, δηλαδή,

$$f_{X,Y}(0, 1) = P(X = 0, Y = 1) = \frac{1}{3} \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{3} = \frac{2}{9}.$$

Με τον ίδιο τρόπο μπορούν να υπολογιστούν όλες οι τιμές της $f_{X,Y}$, που συνοψίζονται στον ακόλουθο πίνακα:

x / y	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

Έστω, τώρα, πως θέλουμε να υπολογίσουμε από τις άνω τιμές τις τιμές της περιθώριας συνάρτησης πιθανότητας της X. Χρησιμοποιώντας τον κανόνα της ολικής πιθανότητας έχουμε

$$\begin{aligned} f_X(1) &= P(X=1) = P(X=1, Y=0) + P(X=1, Y=1) + P(X=1, Y=2) \\ &= f_{X,Y}(1,0) + f_{X,Y}(1,1) + f_{X,Y}(1,2) = \frac{2}{9} + \frac{2}{9} + 0 = \frac{4}{9}. \end{aligned}$$

Άρα, για να βρούμε την $f_X(1)$, αθροίσαμε όλες τις τιμές του πίνακα που αντιστοιχούσαν στην στήλη $X = 1$. Με παρόμοιο τρόπο, αθροίζοντας τις τιμές στις στήλες $X = 0$ και $X = 2$ βρίσκουμε τις $f_X(0)$ και $f_X(2)$. Αντίστοιχα, αθροίζοντας τις τιμές που βρίσκονται στην κάθε γραμμή $Y = y$ του πίνακα, βρίσκουμε τις πιθανότητες $f_Y(y)$. Συμπληρώνεται έτσι ο ακόλουθος πίνακας:

x / y	0	1	2	$f_Y(y)$
0	1/9	2/9	1/9	4/9
1	2/9	2/9	0	4/9
2	1/9	0	0	1/9
$f_X(x)$	4/9	4/9	1/9	

Λόγω του ότι οι τιμές των μαζών $f_X(x)$ και $f_Y(y)$ που αφορούν μόνο μία απ᾽ τις δύο Τ.Μ. μπορούν να γραφτούν, όπως πιο πάνω, στο «περιθώριο» του πίνακα της από κοινού συνάρτησης πιθανότητας, οι επί μέρους συναρτήσεις πιθανότητάς τους ορίστηκαν άνω σαν *περιθώριες*.

Λήμμα 1.1. (Ιδιότητες από κοινού συνάρτησης πιθανότητας) *Έστω δύο διακριτές Τ.Μ. X, Y με συναρτήσεις πιθανότητας $f_X(x)$ και $f_Y(y)$, αντίστοιχα. Η από κοινού συνάρτηση πιθανότητάς τους $f_{X,Y}(x, y)$ έχει τις εξής ιδιότητες:*

1. $\displaystyle\sum_{(x,y)\in\mathbb{R}^2} f_{X,Y}(x, y) = 1.$

2. *Για κάθε $A \subset \mathbb{R}^2$,*
$$P((X, Y) \in A) = \sum_{(x,y)\in A} f_{X,Y}(x, y). \tag{1.1}$$

3. $$\begin{aligned} f_X(x) &= \sum_{y\in\mathbb{R}} f_{X,Y}(x, y) \quad \text{για κάθε } x \in \mathbb{R}, \text{ και} \\ f_Y(y) &= \sum_{x\in\mathbb{R}} f_{X,Y}(x, y) \quad \text{για κάθε } y \in \mathbb{R}. \end{aligned}$$

Απόδειξη. 1. Για διαφορετικά ζεύγη τιμών $(x, y) \in S_{(X,Y)}$ τα ενδεχόμενα $\{X = x, Y = y\}$ είναι ξένα μεταξύ τους, και προφανώς η ένωση τους καλύπτει όλο το δειγματικό χώρο Ω, άρα από την αριθμήσιμη προσθετικότητα της πιθανότητας (αφού $S_{(X,Y)}$ αριθμήσιμο) έχουμε

$$\begin{aligned} 1 &= P(\Omega) = P\left(\bigcup_{(x,y)\in S_{(X,Y)}} \{X = x, Y = y\}\right) = \sum_{(x,y)\in S_{(X,Y)}} P(X = x, Y = y) \\ &= \sum_{(x,y)\in S_{(X,Y)}} f_{X,Y}(x, y) = \sum_{(x,y)\in \mathbb{R}^2} f_{X,Y}(x, y) \end{aligned}$$

Η τελευταία ισότητα ισχύει γιατί στο δεύτερο άθροισμα οι όροι με $(x, y) \notin S_{(X,Y)}$ ισούνται με 0.

2. Με παρόμοιο τρόπο έχουμε

$$\begin{aligned} P((X, Y) \in A) &= P\left(\bigcup_{(x,y)\in A\cap S_{(X,Y)}} \{X = x, Y = y\}\right) \\ &= \sum_{(x,y)\in A\cap S_{(X,Y)}} P(X = x, Y = y) = \sum_{(x,y)\in A} f_{X,Y}(x, y). \end{aligned}$$

3. Τέλος, για δεδομένο x και για όλα τα διαφορετικά y, τα ενδεχόμενα $\{X = x, Y = y\}$ είναι ξένα μεταξύ τους και η ένωσή τους ισούται με το ενδεχόμενο $\{X = x\}$. Άρα, πάλι από την αριθμήσιμη προσθετικότητα της πιθανότητας, έχουμε την τρίτη ιδιότητα:

$$\begin{aligned} f_X(x) &= P(X = x) = P\left(\bigcup_{y\in S_Y} \{X = x, Y = y\}\right) \\ &= \sum_{y\in S_Y} P(X = x, Y = y) = \sum_{y\in S_Y} f_{X,Y}(x, y) = \sum_{y\in \mathbb{R}} f_{X,Y}(x, y). \end{aligned}$$

Η αντίστοιχη σχέση για την $f_Y(y)$ αποδεικνύεται με ακριβώς τον ίδιο τρόπο. (Παρατηρήστε ότι αυτό το σκέλος θα μπορούσε να προκύψει και ως ειδική περίπτωση του προηγούμενου.)

□

Παράδειγμα 1.2. *(Δύο ζάρια — συνέχεια)* Ρίχνουμε ένα ζάρι δύο φορές. Έστω πως οι ρίψεις είναι ανεξάρτητες και το ζάρι δίκαιο. Οι τιμές της από κοινού συνάρτησης πιθανότητας $f_{X,Y}(x,y)$ των X,Y εύκολα υπολογίζονται ως εξής: για κάθε ζεύγος (x,y) όπου $x,y=1,2,3,4,5,6$ έχουμε

$$f_{X,Y}(x,y) = P(X=x, Y=y) = P(X=x)P(Y=y) = \frac{1}{6}\times\frac{1}{6} = \frac{1}{36}.$$

Η δεύτερη ισότητα προέκυψε λόγω ανεξαρτησίας, και η τρίτη λόγω του ότι το ζάρι είναι δίκαιο. Προκύπτει τελικά ο ακόλουθος πίνακας:

x	1	2	3	4	5	6	
y							$f_Y(y)$
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/36
$f_X(x)$	1/6	1/6	1/6	1/6	1/6	1/6	

Παρατηρήστε ότι οι περιθώριες που προκύπτουν είναι οι αναμενόμενες.

Έστω τώρα ότι έρχονται πάντα διπλές, και όλες οι διπλές έχουν την ίδια πιθανότητα να εμφανιστούν. Η νέα από κοινού συνάρτηση πιθανότητας είναι η ακόλουθη:

x	1	2	3	4	5	6	
y							$f_Y(y)$
1	1/6	0	0	0	0	0	1/6
2	0	1/6	0	0	0	0	1/6
3	0	0	1/6	0	0	0	1/6
4	0	0	0	1/6	0	0	1/6
5	0	0	0	0	1/6	0	1/6
6	0	0	0	0	0	1/6	1/6
$f_X(x)$	1/6	1/6	1/6	1/6	1/6	1/6	

Παρατηρήστε ότι οι περιθώριες συναρτήσεις πιθανότητας των δύο περιπτώσεων ταυτίζονται, παρότι τα πειράματα από τα οποία προέρχονται είναι πολύ διαφορετικά.

Παράδειγμα 1.3. *(Δύο ζάρια — συνέχεια)* Ρίχνουμε ένα ζάρι δύο φορές, και υποθέτουμε ότι οι ρίψεις είναι ανεξάρτητες και ότι το ζάρι είναι δίκαιο.

1. Ποια είναι η από κοινού συνάρτηση πιθανότητας των $Z=\min(X,Y)$ και $W=\max(X,Y)$;

2. Ποιες είναι οι περιθώριες συναρτήσεις πιθανότητας των Z, W;

3. Με δεδομένο ότι $W = 5$, ποια η πιθανότητα του ενδεχόμενου $Z = 1$;

Για κάθε ένα από τα ερωτήματα έχουμε:

1. Έχουμε ένα πείραμα με $6 \times 6 = 36$ διαφορετικά αποτελέσματα. Επιπλέον, η από κοινού συνάρτηση πιθανότητας $f_{Z,W}(z, w)$ των Z, W πρέπει να υπολογιστεί για $6 \times 6 = 36$ ζεύγη τιμών (z, w). Αρχικά παρατηρούμε πως κάποια απ' αυτά είναι αδύνατον να εμφανιστούν, π.χ., δεν μπορεί το ελάχιστο απ' τις δύο ζαριές να είναι 5 και το μέγιστο 2. Συνεπώς, $f_{Z,W}(5, 2) = 0$, και γενικά $f_{Z,W}(z, w) = 0$ αν $z > w$. Για να υπολογίσουμε κάθε μια από τις υπόλοιπες τιμές πρέπει να βρούμε τα αποτελέσματα που της αντιστοιχούν.

$$\begin{aligned} f_{Z,W}(1, 1) &= P(X = 1, Y = 1) = \frac{1}{36}, \\ f_{Z,W}(1, 2) &= P(X = 1, Y = 2) + P(X = 2, Y = 1) = \frac{2}{36}. \end{aligned}$$

Συνεχίζοντας με τον ίδιο τρόπο, συμπληρώνουμε τον ακόλουθο πίνακα:

w / z	1	2	3	4	5	6	$f_Z(z)$
1	1/36	2/36	2/36	2/36	2/36	2/36	11/36
2	0	1/36	2/36	2/36	2/36	2/36	9/36
3	0	0	1/36	2/36	2/36	2/36	7/36
4	0	0	0	1/36	2/36	2/36	5/36
5	0	0	0	0	1/36	2/36	3/36
6	0	0	0	0	0	1/36	1/36
$f_W(w)$	1/36	3/36	5/36	7/36	9/36	11/36	

2. Οι συναρτήσεις πιθανότητας $f_W(w)$ και $f_Z(z)$ υπολογίζονται στο περιθώριο του πιο πάνω πίνακα, αθροίζοντας τις τιμές της από κοινού συνάρτησης πιθανότητας στην αντίστοιχη στήλη ή γραμμή.

3. Με χρήση των ορισμών της δεσμευμένης πιθανότητας και της από κοινού συνάρτησης πιθανότητας:

$$P(Z = 1|W = 5) = \frac{P(Z = 1, W = 5)}{P(W = 5)} = \frac{f_{Z,W}(1, 5)}{f_W(5)} = \frac{2/36}{9/36}.$$

Παράδειγμα 1.4. *(Ασανσέρ)* Σε ένα κτίριο με 3 ορόφους και ισόγειο, το ασανσέρ βρίσκεται στον όροφο X και ένας τυχαίος χρήστης στον όροφο Y όπου τα X, Y δίνονται από την ακόλουθη από κοινού συνάρτηση πιθανότητας:

x / y	0	1	2	3	$f_Y(y)$
0	1/6	1/12	1/12	1/12	15/36
1	1/12	1/18	1/36	1/36	7/36
2	1/12	1/36	1/18	1/36	7/36
3	1/12	1/36	1/36	1/18	7/36
$f_X(x)$	15/36	7/36	7/36	7/36	

Εννοείται πως ο όροφος 0 είναι το ισόγειο.

Παρατηρήστε καταρχήν πως οι πιθανότητες του άνω πίνακα ανήκουν όλες στο $[0,1]$ και αθροίζονται στη μονάδα. Παρατηρήστε επίσης πως το άνω μοντέλο λαμβάνει υπ' όψιν ότι οι μετακινήσεις από ή προς το ισόγειο είναι πολύ συχνότερες από ότι σε άλλους ορόφους, και το ότι συχνά το ίδιο άτομο χρησιμοποιεί το ασανσέρ δύο φορές στη σειρά, στην οποία περίπτωση το ασανσέρ παραμένει στον όροφο που είναι το άτομο. Έτσι, για παράδειγμα, το ενδεχόμενο $X=2,\ Y=3$ είναι λιγότερο πιθανό από το $X=2,\ Y=2$, το οποίο με τη σειρά του είναι λιγότερο πιθανό από το $X=2, Y=0$.

Ακολούθως, έστω ότι μας δίνεται ότι η καθυστέρηση στη χρήση του ασανσέρ είναι $Z=|X-Y|$, και καλούμαστε να υπολογίσουμε την συνάρτηση πιθανότητας της Z και τη μέση τιμή της. Χρησιμοποιώντας την από κοινού κατανομή, βρίσκουμε:

$$\begin{aligned}
f_Z(0) &= P(Z=0)=P(X=Y)=\frac{1}{6}+\frac{1}{18}+\frac{1}{18}+\frac{1}{18}=\frac{1}{3},\\
f_Z(1) &= P(Z=1)=P(X=Y+1)+P(X=Y-1)\\
&= \frac{1}{12}+\frac{1}{36}+\frac{1}{36}+\frac{1}{12}+\frac{1}{36}+\frac{1}{36}=\frac{5}{18},\\
f_Z(2) &= P(Z=2)=P(X=Y+2)+P(X=Y-2)\\
&= \frac{1}{12}+\frac{1}{36}+\frac{1}{12}+\frac{1}{36}=\frac{2}{9},\\
f_Z(3) &= P(Z=3)=P(X=3,Y=0)+P(X=0,Y=3)\\
&= \frac{1}{12}+\frac{1}{12}=\frac{1}{6},
\end{aligned}$$

και ακολούθως

$$E(|X-Y|)=E(Z)=0\times\frac{1}{3}+1\times\frac{5}{18}+2\times\frac{2}{9}+3\times\frac{1}{6}=\frac{11}{9}.$$

Παράδειγμα 1.5. *(Τελική βαθμολογία)* Βάσει εμπειρικών δεδομένων, ένα καλό μοντέλο για την από κοινού συνάρτηση πιθανότητας του βαθμού $X\in\{0,1,2,\ldots,8\}$ ενός φοιτητή στην τελική εξέταση ενός μαθήματος και του βαθμού $Y\in\{0,1,2\}$ στην πρόοδο αυτού του μαθήματος είναι το ακόλουθο:

y \ x	0	1	2	3	4	5	6	7	8	$f_Y(y)$
0	1/18	1/18	1/18	1/27	1/27	1/27	0	0	0	5/18
1	1/18	1/18	1/18	1/27	1/27	1/27	1/18	1/18	1/18	8/18
2	0	0	0	1/27	1/27	1/27	1/18	1/18	1/18	5/18
$f_X(x)$	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	

Οι περιθώριες προέκυψαν αθροίζοντας τις αντίστοιχες γραμμές/στήλες. Παρατηρήστε ότι το μοντέλο προβλέπει ότι φοιτητές που λαμβάνουν μεγάλο βαθμό στη μια εξέταση είναι αρκετά πιθανό να λάβουν μεγάλο βαθμό και στην άλλη.

Θα υπολογίσουμε τη μέση τιμή της τελικής βαθμολογίας $Z = X + Y$. Θα υπολογίσουμε καταρχήν την συνάρτηση πιθανότητας του Z:

$$
\begin{aligned}
f_Z(0) &= f_{X,Y}(0,0) = \frac{1}{18}, \\
f_Z(1) &= f_{X,Y}(1,0) + f_{X,Y}(0,1) = \frac{1}{9}, \\
f_Z(2) &= f_{X,Y}(2,0) + f_{X,Y}(1,1) + f_{X,Y}(0,2) = \frac{1}{9}, \\
f_Z(3) &= f_{X,Y}(3,0) + f_{X,Y}(2,1) + f_{X,Y}(1,2) = \frac{5}{54}, \\
f_Z(4) &= f_{X,Y}(4,0) + f_{X,Y}(3,1) + f_{X,Y}(2,2) = \frac{2}{27}, \\
f_Z(5) &= f_{X,Y}(5,0) + f_{X,Y}(4,1) + f_{X,Y}(3,2) = \frac{1}{9}, \\
f_Z(6) &= f_{X,Y}(6,0) + f_{X,Y}(5,1) + f_{X,Y}(4,2) = \frac{2}{27}, \\
f_Z(7) &= f_{X,Y}(7,0) + f_{X,Y}(6,1) + f_{X,Y}(5,2) = \frac{5}{54}, \\
f_Z(8) &= f_{X,Y}(8,0) + f_{X,Y}(7,1) + f_{X,Y}(6,2) = \frac{1}{9}, \\
f_Z(9) &= f_{X,Y}(8,1) + f_{X,Y}(7,2) = \frac{1}{9}, \\
f_Z(10) &= f_{X,Y}(8,2) = \frac{1}{18}.
\end{aligned}
$$

Η μέση τιμή $E(Z)$ προκύπτει

$$
\begin{aligned}
E(Z) &= 0 \times \frac{1}{18} + 1 \times \frac{1}{9} + 2 \times \frac{1}{9} + 3 \times \frac{5}{54} + 4 \times \frac{2}{27} \\
&\quad + 5 \times \frac{1}{9} + 6 \times \frac{2}{27} + 7 \times \frac{5}{54} + 8 \times \frac{1}{9} + 9 \times \frac{1}{9} + 10 \times \frac{1}{18} = 5.
\end{aligned}
$$

Παράδειγμα 1.6. *(Συμπλήρωση πίνακα)* Οι διακριτές Τ.Μ. X και Y έχουν από κοινού συνάρτηση πιθανότητας που δίνεται από τον παρακάτω πίνακα:

x / y	-5	-2	2	5	$f_Y(y)$
0	0.01	0.01	0.01		
1	0.05		0.05	0	0.19
2	0.07	0.07		0.07	
3	0.15		0	0.15	
$f_X(x)$		0.26	0.15		

Ζητούνται τα ακόλουθα:

1. Να συμπληρωθούν οι τιμές που λείπουν.
2. Να υπολογιστεί η $P(Y = 1|X = 2)$.

Για το κάθε ερώτημα έχουμε:

1. Πρέπει η τιμή της περιθώριας να ισούται με το άθροισμα της αντίστοιχης γραμμής ή στήλης. Με βάση αυτή την παρατήρηση, μπορεί να συμπληρωθεί ο ακόλουθος πίνακας:

x / y	-5	-2	2	5	$f_Y(y)$
0	0.01	0.01	0.01	**0.09**	**0.12**
1	0.05	**0.09**	0.05	0	0.19
2	0.07	0.07	**0.09**	0.07	**0.30**
3	0.15	**0.09**	0	0.15	**0.39**
$f_X(x)$	**0.28**	0.26	0.15	**0.31**	

2. Από τον ορισμό της δεσμευμένης πιθανότητας και τις τιμές του πιο πάνω πίνακα έχουμε

$$P(Y = 1|X = 2) = \frac{P(Y = 1, X = 2)}{P(X = 2)} = \frac{f_{X,Y}(2,1)}{f_X(2)} = \frac{0.05}{0.15} = \frac{1}{3}.$$

Παράδειγμα 1.7. Μερικές φορές το πλήθος των τιμών που παίρνει μια τυχαία μεταβλητή Y είναι απαγορευτικά μεγάλο για να αναπαραστήσουμε σε πίνακα την από κοινού συνάρτηση πιθανότητας της με μια άλλη Τ.Μ. X.

Έστω, για παράδειγμα, το ακόλουθο πείραμα. Επιλέγουμε στην τύχη έναν από τρεις καλαθοσφαιριστές, και του ζητάμε να εκτελεί ελεύθερες βολές μέχρι να βάλει το πρώτο καλάθι. Οι καλαθοσφαιριστές επιλέγονται χωρίς κάποια προτίμηση ο ένας από τον άλλο, και έχουν ποσοστά επιτυχίας 50%, 33%, και 25%.

Ορίζουμε την τυχαία μεταβλητή X ως

$$X = \begin{cases} 1/2, & \text{με πιθανότητα } 1/3, \\ 1/3, & \text{με πιθανότητα } 1/3, \\ 1/4, & \text{με πιθανότητα } 1/3. \end{cases}$$

Συνεπώς, η X εκφράζει την ευστοχία του τυχαία επιλεγμένου παίκτη. Ορίζουμε επίσης την Y ως μια Γεωμετρική τυχαία μεταβλητή με μέση τιμή X. Δηλαδή, η παράμετρος της τυχαίας μεταβλητής Y είναι επίσης τυχαία!

Η από κοινού συνάρτηση πιθανότητας των X, Y μπορεί εύκολα να υπολογιστεί από τον ορισμό της δεσμευμένης πιθανότητας: Για $x \in S_X$ και $y \in S_Y$,

$$f_{X,Y}(x,y) = P(X=x, Y=y) = P(X=x)P(Y=y|X=x) = \frac{1}{3}(1-x)^{y-1}x,$$

όπου χρησιμοποιήσαμε τον τύπο της συνάρτησης πιθανότητας της Γεωμ(x) κατανομής.

Η περιθώρια συνάρτηση πιθανότητας της X μας είναι εξ ορισμού γνωστή, ενώ εύκολα μπορούμε να υπολογίσουμε και τις τιμές της περιθώριας συνάρτησης πιθανότητας της Y. Για παράδειγμα,

$$\begin{aligned} f_Y(1) &= f_{X,Y}(1/2,1) + f_{X,Y}(1/3,1) + f_{X,Y}(1/4,1) \\ &= \frac{1}{3}\left(1-\frac{1}{2}\right)^{1-1}\frac{1}{2} + \frac{1}{3}\left(1-\frac{1}{3}\right)^{1-1}\frac{1}{3} + \frac{1}{3}\left(1-\frac{1}{4}\right)^{1-1}\frac{1}{4} = \frac{13}{36}. \end{aligned}$$

Από τον ορισμό της X γνωρίζουμε πως η πιθανότητα $P(X = 1/4) = 1/3$. Αλλά δεδομένου ότι το $Y = 1$, ποια θα ήταν η πιθανότητα να έχουμε $X = 1/4$; Το $X = 1/4$ αντιστοιχεί στην περίπτωση που το Y έχει γεωμετρική κατανομή με παράμετρο 1/4, δηλαδή με πιθανότητα «επιτυχίας» που είναι η μικρότερη από τις τρεις δυνατές περιπτώσεις (1/2, 1/3 ή 1/4). Άρα θεωρούμε ότι υπάρχει μικρή πιθανότητα να έχουμε $X = 1/4$ δεδομένου ότι είχαμε $Y = 1$, δηλαδή «επιτυχία» από το πρώτο κιόλας πείραμα. Με άλλα λόγια, περιμένουμε πως η δεσμευμένη πιθανότητα $P(X = 1/4|Y = 1)$ θα είναι μικρότερη από την αρχική $P(X = 1/4) = 1/3$. Πράγματι, από τον ορισμό της δεσμευμένης πιθανότητας βρίσκουμε

$$P(X=1/4|Y=1) = \frac{P(X=1/4, Y=1)}{P(Y=1)} = \frac{f_{X,Y}(1/4,1)}{f_Y(1)} = \frac{\frac{1}{3}\left(1-\frac{1}{4}\right)^{1-1}\frac{1}{4}}{13/36} = \frac{3}{13}.$$

1.2 Μέση Τιμή Συνάρτησης Δύο Τυχαίων Μεταβλητών

Λήμμα 1.2. (Μέση τιμή συνάρτησης δύο Τ.Μ.) *Έστω δύο διακριτές τυχαίες μεταβλητές X, Y, με από κοινού συνάρτηση πιθανότητας $f_{X,Y}(x,y)$, και τυχαία μεταβλητή $Z = g(X,Y)$. Η μέση τιμή της ισούται με*

$$E(Z) = E(g(X,Y)) = \sum_{(x,y)\in\mathbb{R}^2} g(x,y)f_{X,Y}(x,y). \tag{1.2}$$

Απόδειξη. Έστω $f_Z(z)$ η συνάρτηση πιθανότητας της Z. Παρατηρούμε πως

$$\begin{aligned} E(Z) &= \sum_{z\in\mathbb{R}} z f_Z(z) = \sum_{z\in\mathbb{R}} \left(z \sum_{(x,y)\in\mathbb{R}^2 : g(x,y)=z} f_{X,Y}(x,y) \right) \\ &= \sum_{z\in\mathbb{R}} \left(\sum_{(x,y)\in\mathbb{R}^2 : g(x,y)=z} g(x,y) f_{X,Y}(x,y) \right) \\ &= \sum_{(x,y)\in\mathbb{R}^2} g(x,y)f_{X,Y}(x,y). \end{aligned}$$

Η πρώτη ισότητα ισχύει εξ ορισμού. Στην δεύτερη ισότητα χρησιμοποιήσαμε τον νόμο της ολικής πιθανότητας, όπως αυτός εμφανίζεται στο δεύτερο σκέλος του Λήμματος 1.1. □

Παρατήρηση: Όπως είδαμε και σε προηγούμενα παραδείγματα, πολύ συχνά χρειάζεται να υπολογίσουμε την μέση τιμή Τ.Μ. Z που είναι συναρτήσεις δύο άλλων Τ.Μ. Σε πολλές περιπτώσεις, το άνω λήμμα απλουστεύει σημαντικά τους υπολογισμούς που χρειάζονται, γιατί μας επιτρέπει να αποφύγουμε τον υπολογισμό της συνάρτησης πιθανότητας της Z. Δείτε τα παραδείγματα που ακολουθούν.

Παράδειγμα 1.8. *(Ασανσέρ — συνέχεια)* Σε συνέχεια του Παραδείγματος 1.4, μπορούμε να υπολογίσουμε το μέσο χρόνο αναμονής $E(Z) = E(|X-Y|)$ και ως εξής:

$$\begin{aligned} E(|X-Y|) &= |0-0| \times \frac{1}{6} + (|1-1| + |2-2| + |3-3|) \times \frac{1}{18} \\ &\quad +(|1-0| + |2-0| + |3-0| + |0-1| + |0-2| + |0-3|) \times \frac{1}{12} \\ &\quad +(|2-1| + |3-1| + |1-2| + |3-2| + |1-3| + |2-3|) \times \frac{1}{36} \\ &= \frac{11}{9}. \end{aligned}$$

Παράδειγμα 1.9. *(Τελική Βαθμολογία — συνέχεια)* Σε συνέχεια του Παραδείγματος 1.5,

$$\begin{aligned} E(X) &= (0+1+2+\ldots+8)\times\frac{1}{9}=4,\\ E(Y) &= 0\times\frac{5}{18}+1\times\frac{8}{18}+2\times\frac{5}{18}=1,\\ E(X+Y) &= E(X)+E(Y)=5. \end{aligned}$$

Παράδειγμα 1.10. *(Δύο ζάρια — συνέχεια)* Έστω το ακόλουθο παιχνίδι: ρίχνουμε δύο δίκαια και ανεξάρτητα ζάρια, και κερδίζουμε $2\max\{X,Y\}+3\min\{X,Y\}$ ευρώ. Για να παίξουμε όμως, πρέπει αρχικά να καταβάλουμε 15 ευρώ. Έστω $P=2\max\{X,Y\}+3\min\{X,Y\}-4$ το καθαρό κέρδος. Μας συμφέρει να παίξουμε; Για να απαντήσουμε το ερώτημα, πρέπει να υπολογίσουμε την μέση τιμή

$$E(P)=E(2\max\{X,Y\}+3\min\{X,Y\}-15).$$

Με δεδομένο ότι γνωρίζουμε τις μέσες τιμές $E(\max\{X,Y\})$ και $E(\min\{X,Y\})$ από προηγούμενα παραδείγματα, μας συμφέρει να χρησιμοποιήσουμε το τρίτο σκέλος του άνω λήμματος ως εξής:

$$\begin{aligned} E(2\max\{X,Y\}+3\min\{X,Y\}-15) &= 2E(\max\{X,Y\})+3E(\min\{X,Y\})-15\\ &= 2\times\frac{161}{36}+3\times\frac{91}{36}-15=\frac{55}{36}. \end{aligned}$$

Επειδή η μέση τιμή προκύπτει θετική, μας συμφέρει να παίξουμε.

Εναλλακτικά, χωρίς χρήση του λήμματος, θα έπρεπε να υπολογίσουμε την συνάρτηση πιθανότητας του κέρδους P, και από αυτή την μέση τιμή της P με χρήση του ορισμού της μέσης τιμής.

Ορισμός 1.2. *Η συνδιακύμανση $\mathrm{Cov}(X,Y)$ μεταξύ δύο διακριτών τυχαίων μεταβλητών X,Y που έχουν πεπερασμένη δεύτερη ροπή (δηλαδή $E(X^2),E(Y^2)<\infty$), ορίζεται ως:*

$$\mathrm{Cov}(X,Y)\triangleq E\big[\big(X-E(X)\big)\big(Y-E(Y)\big)\big].$$

Παρατήρηση: Μια πρώτη διαισθητική ερμηνεία της συνδιακύμανσης είναι πως, όταν $\mathrm{Cov}(X,Y)>0$, τότε οι δύο Τ.Μ. X,Y τείνουν να παίρνουν τις «μεγάλες» και τις «μικρές» τιμές τους *ταυτόχρονα*. Αντίστοιχα, όταν $\mathrm{Cov}(X,Y)<0$, τότε όταν η μία Τ.Μ. παίρνει μεγάλες τιμές η άλλη τείνει να παίρνει μικρές τιμές. Άρα η συνδιακύμανση $\mathrm{Cov}(X,Y)$ παρέχει μια πρώτη ένδειξη για τη σχέση ανάμεσα στις X,Y. Δείτε τις συνδιακυμάνσεις των επόμενων παραδειγμάτων.

Παράδειγμα 1.11. *(Δύο προγράμματα — συνέχεια)* Για τις Τ.Μ. X, Y του Παραδείγματος 1.1 έχουμε

$$E(X) = E(Y) = 0 \times \frac{4}{9} + 1 \times \frac{4}{9} + 2 \times \frac{1}{9} = \frac{2}{3}.$$

Από τον ορισμό της συνδιακύμανσης και την από κοινού συνάρτηση πιθανότητας των X, Y, εύκολα υπολογίζουμε

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \frac{1}{9}\Big(0-\frac{2}{3}\Big)\Big(0-\frac{2}{3}\Big) + \frac{2}{9}\Big(1-\frac{2}{3}\Big)\Big(0-\frac{2}{3}\Big) + \frac{1}{9}\Big(2-\frac{2}{3}\Big)\Big(0-\frac{2}{3}\Big) \\ &\quad + \frac{2}{9}\Big(0-\frac{2}{3}\Big)\Big(1-\frac{2}{3}\Big) + \frac{2}{9}\Big(1-\frac{2}{3}\Big)\Big(1-\frac{2}{3}\Big) + \frac{1}{9}\Big(0-\frac{2}{3}\Big)\Big(2-\frac{2}{3}\Big) \\ &= -\frac{2}{9}, \end{aligned}$$

όπου στον υπολογισμό παραλείψαμε τα τρία ζεύγη τιμών (x, y) με μηδενική πιθανότητα.

Παρατήρηση: Στις περισσότερες περιπτώσεις, για τον υπολογισμό της συνδιακύμανσης δεν χρησιμοποιούμε τον ορισμό αλλά τον τύπο της ακόλουθης πρότασης.

Πρόταση 1.1. *Αν οι τυχαίες μεταβλητές X, Y έχουν πεπερασμένη δεύτερη ροπή, τότε*

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y). \tag{1.3}$$

Απόδειξη. Από τον ορισμό της συνδιακύμανσης, άμεσα έχουμε:

$$\begin{aligned} \mathrm{Cov}(X,Y) &= E\big[\big(X-E(X)\big)\big(Y-E(Y)\big)\big] \\ &= E\big[XY - E(X)\,Y - XE(Y) + E(X)E(Y)\big] \\ &= E[XY] - E[E(X)\,Y] - E[XE(Y)] + E[E(X)E(Y)]] \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

□

Πρόταση 1.2. (Ιδιότητες της συνδιακύμανσης) *Έστω X, Y, Z, $\{X_i : 1 \le i \le m\}$, $\{Y_j : 1 \le j \le n\}$ διακριτές τυχαίες μεταβλητές καθεμία με πεπερασμένη δεύτερη ροπή, και $a \in \mathbb{R}$. Τότε*

1. $\mathrm{Cov}(X,X) = \mathrm{V}(X)$.

2. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$.

3. $\mathrm{Cov}(a,X) = 0$.

4. $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$.

5. $\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$.

6. $\mathrm{Cov}\left(\sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{Cov}(X_i, Y_j)$.

Απόδειξη. 1. Από τον ορισμό της συνδιακύμανσης, άμεσα προκύπτει πως

$$\mathrm{Cov}(X, X) = E\big[\big(X - E(X)\big)\big(X - E(X)\big)\big] = E\big[(X - E(X))^2\big] = \mathrm{V}(X).$$

2. Προφανής.

3. $\mathrm{Cov}(a, X) = E((X - E(X))(a - E(a))) = E(0) = 0$, αφού $E(a) = a$.

4. $\mathrm{Cov}(aX, Y) = E\{(aX-E(aX))(Y-E(Y))\} = E\{a(X-E(X))(Y-E(Y))\} = aE\{(X - E(X))(Y - E(Y))\} = a\mathrm{Cov}(X, Y)$.

5.

$$\begin{aligned}
\mathrm{Cov}(X + Y, Z) &= E\{(X + Y - E(X + Y))(Z - E(Z))\} \\
&= E\{(X - E(X) + Y - E(Y))(Z - E(Z))\} \\
&= E\{(X - E(X))(Z - E(Z))\} + E\{(Y - E(Y))(Z - E(Z))\} \\
&= \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)
\end{aligned}$$

6. Προκύπτει από τα 2 και 5 με επαγωγή.

□

Πρόταση 1.3. *Έστω $\{X_i : 1 \leq i \leq n\}$ διακριτές τυχαίες μεταβλητές καθεμία με πεπερασμένη δεύτερη ροπή. Τότε*

$$\mathrm{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{V}(X_i) + 2\sum_{1\leq i<j\leq n} \mathrm{Cov}(X_i, X_j). \tag{1.4}$$

Απόδειξη. Από τις ιδιότητες 1 και 6 της πιό πάνω πρότασης έχουμε

$$\begin{aligned}
\mathrm{V}\left(\sum_{i=1}^{n} X_i\right) &= \mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \mathrm{Cov}(X_i, X_i) + 2\sum_{1\leq <j\leq n} \mathrm{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \mathrm{V}(X_i) + 2\sum_{1\leq i<j\leq n} \mathrm{Cov}(X_i, X_j)
\end{aligned}$$

□

Παράδειγμα 1.12. *(Δύο ζάρια — συνέχεια)* Έστω πως ρίχνουμε δύο ζάρια X και Y έτσι ώστε η από κοινού συνάρτηση πιθανότητας τους να είναι η πρώτη του Παραδείγματος 1.2. Δηλαδή είναι δύο συνήθη ζάρια. Κατά τα γνωστά, έχουμε:

$$\begin{aligned} E(X) &= E(Y) = \sum_{i=1}^{6} iP(X=i) = \sum_{i=1}^{6} i\frac{1}{6} = \frac{7}{2} \\ E(XY) &= \sum_{i,j=1}^{6} ijP(X=i, Y=j) \\ &= (1+2+3+4+5+6) \times (1+2+3+4+5+6)\,\frac{1}{36} = \frac{49}{4}, \\ \mathrm{Cov}(X,Y) &= E(XY) - E(X)E(Y) = \frac{441}{36} - \frac{7}{2} \times \frac{7}{2} = 0. \end{aligned}$$

Αν όμως τα ζάρια έχουν από κοινού συνάρτηση πιθανότητας την δεύτερη του Παραδείγματος 1.2, τότε, αφού και πάλι $E(X) = E(Y) = \frac{7}{2}$, έχουμε:

$$\begin{aligned} E(XY) &= (1^2 + 2^2 + 3^3 + 4^2 + 5^2 + 6^2) \times \frac{1}{6} = \frac{91}{6}, \\ \mathrm{Cov}(X,Y) &= E(XY) - E(X)E(Y) = \frac{91}{6} - \frac{7}{2} \times \frac{7}{2} = \frac{35}{12}. \end{aligned}$$

Παράδειγμα 1.13. *(Τελική βαθμολογία — συνέχεια)* Σε συνέχεια των Παραδειγμάτων 1.5 και 1.9, η συνδιακύμανση των βαθμών X, Y υπολογίζεται ως εξής:

$$\begin{aligned} E(XY) &= [1 \times (1+2+6+7+8) + 2 \times (6+7+8)] \times \frac{1}{18} \\ &\quad + (1 \times (3+4+5) + 2 \times (3+4+5)) \times \frac{1}{27} = 5, \\ \mathrm{Cov}(X,Y) &= E(XY) - E(X)E(Y) = 5 - 4 \times 1 = 1. \end{aligned}$$

Παράδειγμα 1.14. *(Ασανσέρ — συνέχεια)* Σε συνέχεια του Παραδείγματος 1.4,

$$\begin{aligned} E(X) = E(Y) &= 0 \times \frac{15}{36} + (1+2+3) \times \frac{7}{36} = \frac{7}{6}, \\ E(XY) &= (1^2 + 2^2 + 3^3) \times \frac{1}{18} \\ &\quad + (1+2) \times (3+4+5) \times \frac{1}{36} = \frac{25}{18}, \\ \mathrm{Cov}(X,Y) &= E(XY) - E(X)E(Y) = \frac{25}{18} - \frac{7}{6} \times \frac{7}{6} = \frac{1}{36}. \end{aligned}$$

Ορισμός 1.3. *Ο συντελεστής συσχέτισης $\rho(X,Y)$ μεταξύ δύο διακριτών τυχαίων μεταβλητών X,Y που έχουν πεπερασμένη και θετική διασπορά ορίζεται ως:*

$$\rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{V}(X)}\sqrt{\mathrm{V}(Y)}}.$$

Πρόταση 1.4. *Έστω X,Y διακριτές τυχαίες μεταβλητές με $V(X),V(Y)\in(0,\infty)$, και $a,b,c,d\in\mathbb{R}$ με $ac\neq 0$. Τότε*

$$\rho(aX+b,cY+d) = \begin{cases} \rho(X,Y) & \text{αν } ac>0, \\ -\rho(X,Y) & \text{αν } ac<0. \end{cases}$$

Απόδειξη. Χρησιμοποιώντας την Πρόταση (1.2), έχουμε

$$\begin{aligned}\mathrm{Cov}(aX+b,cY+d) &= \mathrm{Cov}(aX,cY)+\mathrm{Cov}(aX,d)+\mathrm{Cov}(b,cY)+\mathrm{Cov}(b,d)\\ &= \mathrm{Cov}(aX,cY)+0+0+0 = ac\mathrm{Cov}(X,Y).\end{aligned}$$

Επίσης $\mathrm{V}(aX+b)=\mathrm{V}(aX)=a^2\mathrm{V}(X)$ και $\mathrm{V}(cY+d)=c^2\mathrm{V}(Y)$. Άρα

$$\rho(aX+b,cY+d) = \frac{ac\mathrm{Cov}(X,Y)}{\sqrt{a^2\mathrm{V}(X)}\sqrt{c^2\mathrm{V}(Y)}} = \frac{ac}{|ac|}\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{V}(X)}\sqrt{\mathrm{V}(Y)}} = \frac{ac}{|ac|}\rho(X,Y).$$

Που είναι το ζητούμενο. □

Πρόταση 1.5. *Έστω X,Y διακριτές τυχαίες μεταβλητές με $V(X),V(Y)\in(0,\infty)$. Τότε*

1. $|\rho(X,Y)|\leq 1$.
2. $\rho(X,Y)=1 \Leftrightarrow$ *Υπάρχουν* $a>0, b\in\mathbb{R}$ *ώστε* $Y=aX+b$.
3. $\rho(X,Y)=-1 \Leftrightarrow$ *Υπάρχουν* $a<0, b\in\mathbb{R}$ *ώστε* $Y=aX+b$.

Δεν θα αποδείξουμε αυτή την πρόταση, απλώς θα κάνουμε κάποια σχόλια. Το 1 προκύπτει από την ανισότητα Cauchy-Schwarz. Για το 2, το αντίστροφο είναι εύκολο. Πράγματι, αν $Y=aX+b$ με $a>0, b\in\mathbb{R}$, τότε

$$\rho(X,Y)=\rho(X,aX+b)=\rho(X,X)=\frac{\mathrm{Cov}(X,X)}{\sqrt{\mathrm{V}(X)}\sqrt{\mathrm{V}(X)}}=\frac{\mathrm{V}(X)}{\mathrm{V}(X)}=1.$$

Το αντίστροφο στην 3 έπεται όμοια.

1.3 Ανεξάρτητες Τυχαίες Μεταβλητές

Ορισμός 1.4. (Ζεύγη ανεξάρτητων διακριτών Τ.Μ.) *Δύο διακριτές Τ.Μ. X, Y καλούνται ανεξάρτητες αν για οποιαδήποτε υποσύνολα $A, B \subset \mathbb{R}$, ισχύει*

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B). \tag{1.5}$$

Παρατήρηση: Συνεπώς, δύο Τ.Μ. είναι ανεξάρτητες αν δύο οποιαδήποτε ενδεχόμενα που αφορούν το καθένα αποκλειστικά τη μία από τις δύο Τ.Μ. είναι ανεξάρτητα. Η ανεξαρτησία λοιπόν των Τ.Μ. είναι άμεσα σχετισμένη με την έννοια της ανεξαρτησίας ενδεχόμενων.

Λήμμα 1.3. (Κριτήριο ανεξαρτησίας διακριτών Τ.Μ.) *Δύο διακριτές τυχαίες μεταβλητές X, Y, με από κοινού συνάρτηση πιθανότητας $f_{X,Y}(x,y)$, και περιθώριες συναρτήσεις πιθανότητας $f_X(x)$, $f_Y(y)$, είναι ανεξάρτητες αν και μόνο αν*

$$f_{X,Y}(x,y) = f_X(x)f_Y(y), \quad \forall\, x,y \in \mathbb{R}. \tag{1.6}$$

Απόδειξη. Έστω πως οι X, Y είναι ανεξάρτητες. Τότε η εξίσωση (1.6) προκύπτει, για κάθε x, y, με εφαρμογή του ορισμού της ανεξαρτησίας για τα σύνολα $A = \{x\}$, $B = \{y\}$. Αντιστρόφως, έστω πως ισχύει η (1.6). Έστω δύο υποσύνολα $A, B \subset \mathbb{R}$. Θα δείξουμε ότι ισχύει η (1.5). Πράγματι:

$$\begin{aligned} P(X \in A, Y \in B) &= \sum_{x\in A,\ y\in B} f_{X,Y}(x,y) = \sum_{x\in A}\sum_{y\in B} f_X(x)f_Y(y) \\ &= \sum_{x\in A} f_X(x) \sum_{y\in B} f_Y(y) = P(X \in A)P(Y \in B), \end{aligned}$$

και η απόδειξη ολοκληρώθηκε. (Στην πρώτη ισότητα χρησιμοποιήσαμε το Λήμμα 1.1 ενώ στην δεύτερη την υπόθεση.) □

Παρατηρήσεις

1. Παρατηρήστε πως όταν οι X, Y είναι ανεξάρτητες, τότε αν $P(X = x) = f_X(x) > 0$ θα έχουμε

$$\begin{aligned} P(Y = y|X = x) &= \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f_{X,Y}(x,y)}{f_X(x)} \\ &= \frac{f_X(x)f_Y(y)}{f_X(x)} = f_Y(y) = P(Y = y). \end{aligned}$$

Μια ανάλογη απόδειξη θα ισχύει αν αλλάξουμε τις θέσεις των X, Y. Άρα τελικά, αν τα X, Y είναι ανεξάρτητα, θα ισχύει ότι:

$$\begin{aligned} \forall x \in \mathbb{R}, \quad P(X=x) > 0 \ &\Rightarrow \ P(Y=y|X=x) = P(Y=y), \\ \forall y \in \mathbb{R}, \quad P(Y=y) > 0 \ &\Rightarrow \ P(X=x|Y=y) = P(X=x). \end{aligned}$$

Μπορούμε να αποδείξουμε, επιπλέον, ότι αν ισχύουν οι άνω, τότε οι X, Y είναι ανεξάρτητες.

2. Το αναγκαίο και ικανό κριτήριο του Λήμματος 1.3 είναι το βασικό εργαλείο που χρησιμοποιούμε για να διαπιστώνουμε την ανεξαρτησία δύο διακριτών Τ.Μ. Δείτε το επόμενο παράδειγμα.

Παράδειγμα 1.15. *(Δύο ζάρια — συνέχεια)* Παρατηρήστε πως όταν η από κοινού συνάρτηση πιθανότητας των αποτελεσμάτων X, Y δύο ζαριών είναι η πρώτη από τις δύο του Παραδείγματος 1.2, τότε σύμφωνα με το άνω κριτήριο οι X, Y είναι ανεξάρτητες, γιατί για κάθε $x, y \in \{1, \ldots, 6\}$ θα έχουμε

$$f_{X,Y}(x,y) = \frac{1}{36} = \frac{1}{6} \times \frac{1}{6} = f_X(x) f_Y(y).$$

Αν όμως η από κοινού συνάρτηση πιθανότητας των X, Y είναι η δεύτερη του Παραδείγματος 1.2, τότε οι X, Y δεν είναι ανεξάρτητες, γιατί, για παράδειγμα

$$f_{X,Y}(1,2) = 0 \neq \frac{1}{6} \times \frac{1}{6} = f_X(1) f_Y(2).$$

Παρατηρήσεις

1. Σε ποια άλλα από τα προηγούμενα παραδείγματα είχαμε ανεξάρτητες Τ.Μ.; Σε όλες τις περιπτώσεις, βεβαιωθείτε ότι συμβαδίζει η διαίσθησή σας με το κριτήριο του Λήμματος 1.3.

2. Συχνά η ανεξαρτησία δύο Τ.Μ. δεν είναι το ζητούμενο, αλλά δίνεται ως υπόθεση. Δείτε το ακόλουθο παράδειγμα.

Παράδειγμα 1.16. *(Ελάχιστο δύο ανεξάρτητων γεωμετρικών Τ.Μ.)* Έστω X και Y δύο ανεξάρτητες γεωμετρικές Τ.Μ., με παραμέτρους p_1 και p_2, αντίστοιχα. Θα υπολογίσουμε τη συνάρτηση πιθανότητας της νέας Τ.Μ. $Z = \min(X, Y)$.

Παρατηρούμε ότι το ελάχιστο Z είναι τουλάχιστον k, αν και μόνο αν και οι δύο Τ.Μ. X, Y παίρνουν τιμές τουλάχιστον k. Οπότε, για κάθε $k \geq 1$,

$$\begin{aligned} P(Z \geq k) &= P(\min(X,Y) \geq k) = P(X \geq k, \ Y \geq k) = P(X \geq k) P(Y \geq k) \\ &= (1-p_1)^{k-1}(1-p_2)^{k-1} = [(1-p_1)(1-p_2)]^{k-1}, \end{aligned}$$

όπου η τρίτη ισότητα ισχύει λόγω ανεξαρτησίας και η τέταρτη είναι γνωστή ιδιότητα της γεωμετρικής κατανομής.

Αν τώρα ορίσουμε $q = 1 - (1-p_1)(1-p_2)$, εφόσον

$$P(Z \geq k) = P(Z = k) + P(Z \geq k+1),$$

έχουμε

$$P(Z = k) = P(Z \geq k) - P(Z \geq k+1) = (1-q)^{k-1} - (1-q)^k = q(1-q)^{k-1}.$$

Συνεπώς, η Z ακολουθεί επίσης γεωμετρική κατανομή, αλλά με παράμετρο $q = 1-(1-p_1)(1-p_2)$.

Αυτό εξηγείται εύκολα: Αν διεξάγουμε ταυτόχρονα δύο ακολουθίες ανεξάρτητων πειραμάτων, μπορούμε να ορίσουμε ένα ολικό πείραμα στο οποίο θα έχουμε επιτυχία την πρώτη φορά που θα έχει επιτυχία ένα από τα δύο πειράματα. Σ' αυτή την περίπτωση, το πλήθος των επαναλήψεων που απαιτούνται για την πρώτη επιτυχία του ολικού πειράματος είναι $Z = \min(X, Y)$ όπου X, Y είναι το πλήθος επαναλήψεων μέχρι την πρώτη επιτυχία στο πρώτο και το δεύτερο πείραμα, αντίστοιχα. Το ολικό πείραμα αποτυγχάνει όταν αποτύχουν και τα δύο επί μέρους, δηλαδή με πιθανότητα $(1-p_1)(1-p_2)$, άρα η πιθανότητα επιτυχίας είναι $1-(1-p_1)(1-p_2) = q$. Συνεπώς, το πλήθος των προσπαθειών του ολικού πειράματος μέχρι την πρώτη επιτυχία ακολουθεί γεωμετρική κατανομή με παράμετρο q.

Λήμμα 1.4. (Ιδιότητες ανεξάρτητων διακριτών Τ.Μ.) *Έστω X, Y ανεξάρτητες διακριτές Τ.Μ.*

1. *Έστω συναρτήσεις $g : S_X \to \mathbb{R}$, και $h : S_Y \to \mathbb{R}$. Θα ισχύει*

$$E(g(X)h(Y)) = E(g(X))E(h(Y))).$$

   Ειδική περίπτωση της άνω είναι η

$$E(XY) = E(X)E(Y).$$

2. $\mathrm{Cov}(X, Y) = 0$.

3. $\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y)$.

Απόδειξη. 1. Εφαρμόζοντας το Λήμμα 1.2 έχουμε:

$$\begin{aligned} E(g(X)h(Y)) &= \sum_{(x,y)\in\mathbb{R}^2} h(x)g(y)f_{X,Y}(x,y) \\ &= \sum_{x\in\mathbb{R}}\sum_{y\in\mathbb{R}} g(x)h(y)f_X(x)f_Y(y) \\ &= \sum_{x\in\mathbb{R}} g(x)f_X(x)\sum_{y\in\mathbb{R}} h(y)f_Y(y) = E(g(X))E(h(Y)). \end{aligned}$$

2. Προκύπτει άμεσα από το πρώτο σκέλος και την Πρόταση 1.1.
3. Προκύπτει άμεσα από το δεύτερο σκέλος και την Πρόταση 1.3.

□

Παράδειγμα 1.17. *(Δύο ζάρια — συνέχεια)* Έστω πως η από κοινού συνάρτηση πιθανότητας των αποτελεσμάτων X, Y δύο ζαριών είναι η πρώτη από τις δύο του Παραδείγματος 1.2, οπότε οι X, Y είναι ανεξάρτητες. Θα έχουμε, για το άθροισμά τους $X+Y$:

$$\mathrm{V}(X+Y)=\mathrm{V}(X)+\mathrm{V}(Y)=\frac{35}{12}+\frac{35}{12}=\frac{35}{6}.$$

Παρατηρήστε επίσης ότι

$$E(X+Y)=E(X)+E(Y)=\frac{7}{2}+\frac{7}{2}=7.$$

Πρέπει να τονιστεί ότι η τελευταία ισότητα θα ίσχυε ακόμα και αν τα X,Y ΔΕΝ ήταν ανεξάρτητα, καθώς το Λήμμα 1.2 (πάνω στο οποίο βασιστήκαμε) δεν απαιτεί ανεξαρτησία!

Παράδειγμα 1.18. Θα δείξουμε ένα παράδειγμα δύο διακριτών Τ.Μ. X,Y με διασπορές $\mathrm{V}(X)=\mathrm{V}(Y)>0$, αλλά τέτοιες ώστε να έχουμε $\mathrm{V}(X+Y)=0$.

Έστω $X\sim \mathrm{Bern}(1/2)$ οπότε $\mathrm{V}(X)=(1/2)(1/2)=1/4>0$, και έστω $Y=-X$ οπότε $\mathrm{V}(Y)=\mathrm{V}((-1)X)=(-1)^2\mathrm{V}(X)=1/4>0$. Αλλά εφόσον εξ ορισμού η $X+Y=0$ είναι απλώς μια σταθερά, έχουμε $\mathrm{V}(X+Y)=0$, το οποίο φυσικά δεν ισούται με τον άθροισμα των διασπορών $\mathrm{V}(X)+\mathrm{V}(Y)=1/2$, όπως προβλέπει το άνω λήμμα. Αυτό συμβαίνει γιατί προφανώς οι X, Y δεν είναι ανεξάρτητες.

Παράδειγμα 1.19. *(Μη ανεξάρτητες Τ.Μ. με μηδενική συνδιακύμανση)* Είναι δυνατόν δύο Τ.Μ. να μην είναι ανεξάρτητες, να έχουν όμως μηδενική συνδιακύμανση. Για παράδειγμα, έστω δύο διακριτές Τ.Μ. X,Y με από κοινού συνάρτηση πιθανότητας $f_{X,Y}$ όπως στον πιο κάτω πίνακα:

x / y	-1	0	1	$f_Y(y)$
-1	0	1/4	0	1/4
0	1/4	0	1/4	1/2
1	0	1/4	0	1/4
$f_X(x)$	1/4	1/2	1/4	

Οι περιθώριες συναρτήσεις πιθανότητας $f_X(x), f_Y(y)$ έχουν επίσης υπολογιστεί στον πίνακα. Οι X,Y δεν είναι ανεξάρτητες, αφού

$$f_{X,Y}(0,0)=0\neq\frac{1}{2}\times\frac{1}{2}=f_X(0)f_Y(0).$$

Για τη συνδιακύμανσή τους, πρώτα υπολογίζουμε

$$E(X) = (-1) \times \frac{1}{4} + 0 \times \frac{1}{2} + 1 \times \frac{1}{4} = 0,$$

και παρομοίως $E(Y) = 0$ αφού οι X και Y έχουν την ίδια συνάρτηση πιθανότητας. Επιπλέον παρατηρούμε πως

$$E(XY) = \sum_{x=-1,0,1} \sum_{y=-1,0,1} xy f_{X,Y}(x,y) = 0,$$

διότι *όλοι* οι όροι του πιο πάνω αθροίσματος είναι μηδενικοί! (Ή $x = 0$ ή $y = 0$ ή $f_{X,Y}(x,y) = 0$, για όλα τα δυνατά ζεύγη τιμών (x,y).) Άρα,

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = 0 - 0 \times 0 = 0.$$

Ορισμός 1.5. (Ασυσχέτιστες Τ.Μ.) *Δύο Τ.Μ. X, Y, καλούνται ασυσχέτιστες αν η συνδιακύμανσή τους* $\mathrm{Cov}(X,Y)$ *είναι μηδενική, δηλαδή* $\mathrm{Cov}(X,Y) = 0$. *Από τα προηγούμενα είδαμε ότι δύο Τ.Μ. που είναι ανεξάρτητες είναι και ασυσχέτιστες, αλλά το αντίστροφο μπορεί να μην ισχύει.*

1.4 Άθροισμα Ανεξάρτητων Τυχαίων Μεταβλητών

Παράδειγμα 1.20. *(Συνέλιξη)* Έστω δύο ανεξάρτητες διακριτές Τ.Μ. X, Y με $S_X = S_Y = \mathbb{Z}$ και συναρτήσεις πιθανότητας $f_X(x)$, $f_Y(y)$ αντίστοιχα. Θα δείξουμε ότι η νέα Τ.Μ. $X+Y$ έχει συνάρτηση πιθανότητας:

$$f_{X+Y}(m) = \sum_{k=-\infty}^{\infty} f_X(k) f_Y(m-k), \quad m \in \mathbb{Z}.$$

Η πιο πάνω έκφραση είναι γνωστή ως η *συνέλιξη* των δύο μαζών $f_X(x)$ και $f_Y(y)$, και εμφανίζεται συχνά σε εφαρμογές, ακόμα και εκτός θεωρίας πιθανοτήτων. Αργότερα θα δούμε και την συνεχή εκδοχή της, που έχει τη μορφή ολοκληρώματος αντί για αθροίσματος, και της οποίας η χρήση είναι ακόμα πιο διαδεδομένη.

Πράγματι, παρατηρούμε πως, για οποιοδήποτε τιμή $m \in \mathbb{Z}$, η $X+Y$ ισούται με m αν και μόνο αν η $X = k$ και η $Y = m-k$ για κάποιο $k \in \mathbb{Z}$. Συνεπώς:

$$\begin{aligned} f_{X+Y}(m) &= P(X+Y=m) = P\left(\bigcup_{k=-\infty}^{\infty} (\{X=k\} \cap \{Y=m-k\})\right) \\ &= \sum_{k=-\infty}^{\infty} P(\{X=k\} \cap \{Y=m-k\}) = \sum_{k=-\infty}^{k=\infty} P(X=k)P(Y=m-k) \\ &= \sum_{k=-\infty}^{\infty} f_X(k) f_Y(m-k). \end{aligned}$$

Η πρώτη και η τελευταία ισότητα προκύπτουν από τον ορισμό της συνάρτησης πιθανότητας, η δεύτερη προκύπτει γράφοντας το ενδεχόμενο $\{X+Y=m\}$ σαν ένωση ξένων ενδεχόμενων, η τρίτη προκύπτει ακριβώς επειδή τα ενδεχόμενα είναι ξένα, ενώ η τέταρτη λόγω της ανεξαρτησίας των X, Y.

Παρατήρηση: Παρατηρήστε ότι αν μια Τ.Μ. X έχει σύνολο τιμών $S_X \subset \mathbb{Z}$ μπορούμε να θέσουμε ως νέο σύνολο τιμών της το $\mathbb{Z}$, θέτοντας μηδενικές τις πιθανότητες των ακεραίων εκτός του αρχικού S_X. Αντιστρόφως, μπορούμε να αφαιρέσουμε από το σύνολο τιμών S_X οποιαδήποτε τιμή x για την οποία $f_X(x) = 0$. Και οι δύο αυτές διαδικασίες δεν αλλάζουν ουσιωδώς την X, και γενικεύουν την εφαρμογή του άνω παραδείγματος. Δείτε τα παραδείγματα που ακολουθούν.

Παράδειγμα 1.21. *(Άθροισμα ανεξάρτητων Poisson Τ.Μ.)* Αν οι Τ.Μ. X και Y είναι ανεξάρτητες, με κατανομές $\text{Poisson}(\lambda)$ και $\text{Poisson}(\mu)$ αντίστοιχα, θα δείξουμε ότι η Τ.Μ. $X+Y$ έχει κατανομή $\text{Poisson}(\lambda+\mu)$.

Πράγματι, έστω $f_X(x)$ και $f_Y(y)$ οι συναρτήσεις πιθανότητας των X και Y, αντίστοιχα, και έστω $f_{X+Y}(m)$ η συνάρτηση πιθανότητας της $X+Y$. Από το προηγούμενο παράδειγμα, έχουμε:

$$\begin{aligned} f_{X+Y}(m) &= \sum_{k=-\infty}^{\infty} f_X(k)\, f_Y(m-k) = \sum_{k=0}^{m} \Big(\frac{e^{-\lambda}\lambda^k}{k!}\Big)\Big(\frac{e^{-\mu}\mu^{m-k}}{(m-k)!}\Big) \\ &= \frac{e^{-(\lambda+\mu)}}{m!}\sum_{k=0}^{m}\frac{m!}{k!(m-k)!}\lambda^k\mu^{m-k} = \frac{e^{-(\lambda+\mu)}}{m!}(\lambda+\mu)^m. \end{aligned}$$

Παρατηρήστε ότι σε αυτή την περίπτωση και οι δύο Τ.Μ. που απαρτίζουν το άθροισμα είναι μη αρνητικές, και συνεπώς δεν προκύπτει άπειρο άθροισμα όπως στην γενική περίπτωση του προηγούμενου παραδείγματος. Επίσης, η τελευταία ισότητα προκύπτει από το διωνυμικό θεώρημα:

$$(\alpha+b)^n = \sum_{k=0}^{n}\binom{n}{k}\alpha^k b^{n-k}.$$

Συνεπώς, $X+Y \sim \text{Poisson}(\lambda+\mu)$.

Παράδειγμα 1.22. *(Άθροισμα ανεξάρτητων διωνυμικών Τ.Μ.)* Σε αυτό το παράδειγμα θα δείξουμε ότι το άθροισμα δύο διωνυμικών Τ.Μ. X, Y με παραμέτρους n_1, p και n_2, p αντίστοιχα, ανεξάρτητων μεταξύ τους, έχει Διων(n_1+n_2, p) κατανομή.

Έστω $f_X(x)$ και $f_Y(y)$ οι συναρτήσεις πιθανότητας των X και Y, αντίστοιχα. Παρατηρήστε πως

$$\begin{aligned} f_{X+Y}(m) &= \sum_{k=-\infty}^{\infty} f_X(k) f_Y(m-k) \\ &= \sum_{k=\max\{0,m-n_2\}}^{\min\{n_1,m\}}\binom{n_1}{k}p^k(1-p)^{n_1-k}\binom{n_2}{m-k}p^{m-k}(1-p)^{n_2-m+k} \\ &= (p^m(1-p)^{n_1+n_2-m})\sum_{k=\max\{0,m-n_2\}}^{\min\{n_1,m\}}\binom{n_1}{k}\binom{n_2}{m-k}. \qquad (1.7) \end{aligned}$$

Τα όρια στο άθροισμα προέκυψαν κρατώντας μόνο τις τιμές του k για τις οποίες είναι θετικές και οι δύο συναρτήσεις πιθανότητας. Παρατηρήστε ότι όταν $m > n_1 + n_2$ ή $m < 0$, το άνω άθροισμα είναι κενό και προκύπτει πως $f_{X+Y}(m) = 0$, όπως και με την Διων(n_1+n_2, p) κατανομή. Αρκεί λοιπόν να επικεντρωθούμε στην περίπτωση $0 \le m \le n_1 + n_2$.

Σε αυτή την περίπτωση, παρατηρήστε πως

$$\sum_{k=\max\{0,m-n_2\}}^{\min\{n_1,m\}} \binom{n_1}{k}\binom{n_2}{m-k} = \binom{n_1+n_2}{m}. \tag{1.8}$$

Πράγματι, έστω το εξής πρόβλημα: Έχουμε $n_1 > 0$ άντρες και $n_2 > 0$ γυναίκες, και πρέπει να φτιάξουμε μια επιτροπή από m άτομα, με $0 \leq m \leq n_1 + n_2$. Το δεξί μέλος της (1.8) μας δίνει τους δυνατούς συνδυασμούς. Επιπλέον, κάθε ένας από τους όρους του αριστερού σκέλους της (1.8) μας δίνει τους τρόπους με τους οποίους μπορούμε να φτιάξουμε μια επιτροπή m ατόμων με k άντρες. Σχετικά με τα όρια του αθροίσματος, παρατηρήστε ότι σε κάθε όρο του αθροίσματος ο αριθμός των αντρών k θα πρέπει να είναι:

1. το πολύ ίσος με τον ολικό αριθμό αντρών, δηλαδή $k \leq n_1$,
2. το πολύ ίσος με το μέγεθος της επιτροπής, δηλαδή $k \leq m$,
3. μεγαλύτερος ή ίσος του μηδενός, δηλαδή $k \geq 0$, και τέλος,
4. αν $m > n_2$ (δηλαδή τα μέλη της επιτροπής είναι περισσότερα από τις γυναίκες) μεγαλύτερος ή ίσος από το $m - n_2$, δηλαδή $k \geq m - n_2$.

Συνδυάζοντας όλους αυτούς τους περιορισμούς, προκύπτουν τελικά τα όρια του αθροίσματος. Η ταυτότητα (1.8) είναι μια από τις απλούστερες μορφές της ταυτότητα του Vandermonde (και όχι Voldemort).

Εφαρμόζοντας τελικά την (1.8) στην (1.7), προκύπτει τελικά πως

$$f_{X+Y}(m) = \binom{n_1+n_2}{m} p^m (1-p)^{n_1+n_2-m},$$

δηλαδή το ζητούμενο.

Το αποτέλεσμα είναι αναμενόμενο. Πράγματι, το X εκφράζει το πλήθος των επιτυχιών σε n_1 πειράματα, ανεξάρτητα μεταξύ τους και με πιθανότητα επιτυχίας p, και το Y παρομοίως εκφράζει το πλήθος των επιτυχιών σε n_2 πειράματα, ανεξάρτητα μεταξύ τους, και ανεξάρτητα από τα πρώτα, επίσης με πιθανότητα επιτυχίας p. Άρα, λογικά το $X+Y$ εκφράζει τον πλήθος των επιτυχιών στο σύνολο των n_1+n_2 όμοιων πειραμάτων, και συνεπώς πρέπει να ακολουθεί την κατανομή Bernoulli.

1.5 Πολλές Τυχαίες Μεταβλητές

Στις προηγούμενες παραγράφους αυτού του κεφαλαίου περιοριστήκαμε στην περίπτωση δύο από κοινού κατανεμημένων διακριτών Τ.Μ. Όμως όλη η θεωρία που αναπτύξαμε μπορεί να επεκταθεί εύκολα στην περίπτωση πολλών διακριτών Τ.Μ. Σε αυτή την παράγραφο παρουσιάζουμε συνοπτικά την επέκταση της θεωρίας, και τα πιο χρήσιμα αποτελέσματα της. Παραλείπουμε όλες τις αποδείξεις, καθώς είναι όμοιες με αυτές της περίπτωσης των δύο Τ.Μ.

Ορισμός 1.6. (Από κοινού και περιθώριες συναρτήσεις πιθανότητας) *Έστω n διακριτές τυχαίες μεταβλητές $X_1, \ldots, X_n$.*

1. *Ορίζουμε την <u>από κοινού συνάρτηση πιθανότητας</u> ως την συνάρτηση*

$$f_{X_1,\ldots,X_n} : \mathbb{R}^n \to [0,1]$$

   με

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P(X_1 = x_1, \ldots, X_n = x_n), \quad \forall\,(x_1,\ldots,x_n) \in \mathbb{R}^n.$$

2. *Οι συναρτήσεις πιθανότητας των $X_1, \ldots, X_n$ καλούνται <u>περιθώριες συναρτήσεις πιθανότητας</u>.*

Λήμμα 1.5. (Ιδιότητες από κοινού συνάρτησης πιθανότητας) *Έστω n διακριτές Τ.Μ. $X_1, \ldots, X_n$, με συναρτήσεις πιθανότητας $f_{X_1}(x_1), \ldots, f_{X_n}(x_n)$ αντιστοίχως. Η από κοινού συνάρτηση πιθανότητας τους $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$ έχει τις εξής ιδιότητες:*

1. $\displaystyle\sum_{(x_1,\ldots,x_n)\in\mathbb{R}^n} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = 1.$

2. *Έστω $A \subset \mathbb{R}^n$. Τότε*

$$P((X_1,\ldots,X_n) \in A) = \sum_{(x_1,\ldots,x_n)\in A} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n). \tag{1.9}$$

3. $\displaystyle f_{X_i}(x) = \sum_{x_1\in\mathbb{R},\ldots,x_{i-1}\in\mathbb{R},x_{i+1}\in\mathbb{R},\ldots,x_n\in\mathbb{R}} f_{X_1,\ldots,X_n}(x_1,\ldots,x_{i-1},x,x_{i+1},\ldots,x_n)$

   για κάθε $x \in \mathbb{R}$.

Λήμμα 1.6. (Ιδιότητες μέσης τιμής συνάρτησης πολλών Τ.Μ.) *Έστω n διακριτές τυχαίες μεταβλητές $X_1, \ldots, X_n$, με από κοινού συνάρτηση πιθανότητας $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$. Έστω Τ.Μ. $Z = g(X_1,\ldots,X_n)$. Η μέση τιμή της Z ισούται με*

$$E(Z) = E(g(X_1,\ldots,X_n)) = \sum_{(x_1,\ldots,x_n)\in\mathbb{R}^n} g(x_1,\ldots,x_n) f_{X_1,\ldots,X_n}(x_1,\ldots,x_n).$$

Ορισμός 1.7. (n ανεξάρτητες Τ.Μ.) *n διακριτές Τ.Μ. $X_1, \dots, X_n$ καλούνται ανεξάρτητες αν για οποιαδήποτε υποσύνολα $A_i \subset \mathbb{R}$, $i = 1, \dots, n$ ισχύει*

$$P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n).$$

Λήμμα 1.7. (Κριτήριο ανεξαρτησίας Τ.Μ.) *Έστω n διακριτές Τ.Μ. $X_1, \dots, X_n$, με από κοινού συνάρτηση πιθανότητας $f_{X_1,\dots,X_n}(x_1, \dots, x_n)$, και περιθώριες συναρτήσεις πιθανότητας $f_{X_i}(x_i)$, $i = 1, \dots, n$. Οι $X_1, \dots, X_n$ είναι ανεξάρτητες αν και μόνο αν για οποιαδήποτε $x_i \in \mathbb{R}$ έχουμε*

$$f_{X_1,\dots,X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n).$$

Λήμμα 1.8. (Ιδιότητες ανεξάρτητων μεταβλητών) *Έστω n διακριτές ανεξάρτητες Τ.Μ. $X_1, \dots, X_n$.*

1. *Οποιοδήποτε υποσύνολο από τις $X_1, \dots, X_n$ είναι επίσης ανεξάρτητες Τ.Μ.*

2. *Έστω συναρτήσεις $g_i : S_{X_i} \to \mathbb{R}$. Θα ισχύει*

$$E(g_1(X_1) \cdots g_n(X_n)) = E(g_1(X_1)) \cdots E(g_n(X_n)).$$

   Ειδική περίπτωση της άνω είναι η

$$E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n).$$

3. $\mathrm{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{V}(X_i)$.

Παράδειγμα 1.23. Έστω τρεις ανεξάρτητες Τ.Μ. X_1, X_2, X_3 που έχουν όλες την ίδια συνάρτηση πιθανότητας, και, κατά συνέπεια, την ίδια μέση τιμή $\mu = E(X_i)$. Ορίζουμε δύο νέες Τ.Μ., τις $Y = X_1 + X_2$ και $Z = X_1 + X_3$. Εφόσον και οι δύο εξαρτώνται απ' την X_1 περιμένουμε πως δεν θα είναι ανεξάρτητες, και πως πιθανότατα η συνδιακύμανσή τους δεν θα είναι μηδενική. Έχουμε:

$$\begin{aligned}
\mathrm{Cov}(Y, Z) &= \mathrm{Cov}(X_1 + X_2, X_1 + X_3) \\
&= \mathrm{Cov}(X_1, X_1) + \mathrm{Cov}(X_1, X_3) + \mathrm{Cov}(X_2, X_1) + \mathrm{Cov}(X_2, X_3) \\
&= \mathrm{V}(X_1) + 0 + 0 + 0 = \mathrm{V}(X_1).
\end{aligned}$$

Άρα, η συνδιακύμανση μεταξύ Y και Z όχι μόνο είναι μη μηδενική (όπως περιμέναμε), αλλά επιπλέον ισούται ακριβώς με τη διασπορά της Τ.Μ. X_1 η οποία συνδέει τις Y και Z.

Παράδειγμα 1.24. *(Μέση τιμή και διασπορά της διωνυμικής κατανομής)* Θα χρησιμοποιήσουμε την θεωρία που έχουμε αναπτύξει μέχρι τώρα για να υπολογίσουμε, με πολύ εύκολο τρόπο, την μέση τιμή και τη διασπορά της διωνυμικής κατανομής. Έστω Τ.Μ. X που ακολουθεί την διωνυμική κατανομή με παραμέτρους N και p. Κατά τα γνωστά από τη θεωρία, η X εκφράζει τον αριθμό των επιτυχιών σε N ανεξάρτητα πειράματα, καθένα με πιθανότητα επιτυχίας p. Συνεπώς, η X μπορεί να γραφτεί ως $X = \sum_{i=1}^{N} X_i$ όπου οι X_i, $i = 1, \dots, N$ είναι ανεξάρτητες Τ.Μ. Bernoulli, όλες με παράμετρο p, και συνεπώς με μέση τιμή $E(X_i) = p$ και διασπορά $\mathrm{V}(X_i) = p(1-p)$. Έχουμε:

$$E(X) = E(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} E(X_i) = \sum_{i=1}^{N} p = Np.$$

Για να υπολογίσουμε τη διασπορά, παρατηρούμε απλώς πως

$$\mathrm{V}(X) = \mathrm{V}(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} \mathrm{V}(X_i) = Np(1-p).$$

Παράδειγμα 1.25. *(Το πρόβλημα των συζύγων)* Σε κάποιο πάρτι, N άτομα αφήνουν τα καπέλα τους σε ένα τραπέζι. Τα καπέλα ανακατεύονται και, φεύγοντας, κάθε άτομο παίρνει στην τύχη ένα καπέλο. Έστω X το πλήθος των ατόμων που παίρνουν το δικό τους καπέλο. Θα υπολογίσουμε τη μέση τιμή και τη διασπορά της X.

Θα χρησιμοποιήσουμε τις βοηθητικές Τ.Μ. X_i, $i = 1, 2, \dots, N$, όπου $X_i = 1$ αν το άτομο i πάρει το δικό του καπέλο, και $X_i = 0$ στην αντίθετη περίπτωση. Παρατηρούμε ότι $X = X_1 + X_2 + \ldots + X_N$. Επιπλέον, λόγω της συμμετρίας του προβλήματος, η πιθανότητα να πάρει ένα άτομο το δικό του καπέλο είναι $\frac{1}{N}$. Συνεπώς η μέση τιμή της κάθε X_i είναι

$$E(X_i) = 1 \times \frac{1}{N} + 0 \times (1 - \frac{1}{N}) = \frac{1}{N},$$

και συνεπώς

$$E(X) = E(X_1 + X_2 + \ldots + X_N) = E(X_1) + E(X_2) + \ldots + E(X_N) = N \times \frac{1}{N} = 1.$$

Άρα, κατά μέσο όρο, ακριβώς ένα άτομο θα φύγει με το δικό του καπέλο. Παρατηρήστε ότι το αποτέλεσμα είναι ανεξάρτητο του N!

Για να υπολογίσουμε τη διασπορά, μπορούμε να χρησιμοποιήσουμε την Πρόταση 1.3. Εναλλακτικά (και με περίπου τον ίδιο όγκο πράξεων) παρατηρούμε καταρχάς πως

$$\begin{aligned} E(X^2) &= E\left(\left(\sum_{i=1}^{N} X_i\right)^2\right) = E\left(\left(\sum_{i=1}^{N} X_i\right)\left(\sum_{j=1}^{N} X_j\right)\right) \\ &= E\left(\sum_{i=1}^{N}\sum_{j=1}^{N} X_i X_j\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} E(X_i X_j). \end{aligned}$$

Στο τελευταίο πιο πάνω διπλό άθροισμα, υπάρχουν N όροι της μορφής $E(X_i^2)$, και $N(N-1)$ όροι της μορφής $E(X_iX_j)$ για $i \neq j$. Εφόσον οι X_i είναι Τ.Μ. Bernoulli, πάντοτε έχουμε $X_i^2 = X_i$ και συνεπώς,

$$E(X_i^2) = E(X_i) = \frac{1}{N},$$

ενώ για $i \neq j$, βρίσκουμε, από τον ορισμό της δεσμευμένης πιθανότητας,

$$\begin{aligned} E(X_iX_j) &= P(X_i = X_j = 1) \times 1 + (1 - P(X_i = X_j = 1)) \times 0 \\ &= P(X_i = 1 \text{ και } X_j = 1) \\ &= P(X_i = 1)P(X_j = 1 | X_i = 1) = \frac{1}{N} \times \frac{1}{N-1}. \end{aligned}$$

Άρα, τελικά,

$$E(X^2) = E\left(\left(\sum_{i=1}^{N} X_i\right)^2\right) = N\frac{1}{N} + N(N-1)\frac{1}{N(N-1)} = 2,$$

και για τη διασπορά έχουμε

$$\mathrm{V}(X) = E(X^2) - (E(X))^2 = 1.$$

Αν και δεν το αποδεικνύουμε εδώ, αναφέρουμε ότι η κατανομή του X, καθώς το $N \to \infty$, τείνει στην κατανομή Poisson με παράμετρο $\lambda = 1$. Μπορείτε να εξηγήσετε διαισθητικά γιατί;

Κεφάλαιο 2

Ζεύγη Συνεχών Τυχαίων Μεταβλητών

Ανάλογα με ό,τι συμβαίνει στην περίπτωση των διακριτών Τ.Μ., η σχέση που έχουν μεταξύ τους δύο συνεχείς Τ.Μ. δεν μπορεί να αποτυπωθεί με χρήση των πυκνοτήτων τους, και είναι απαραίτητο να εισάγουμε την έννοια της *από κοινού* πυκνότητας. Μια βασική διαφορά της θεωρίας που προκύπτει σε σχέση με την περίπτωση της μελέτης μεμονωμένων συνεχών Τ.Μ., είναι ότι τα ολοκληρώματα που εμφανίζονται στον υπολογισμό πιθανοτήτων δεν είναι πια απλά, αλλά διπλά, και συνεπώς είναι αρκετά πιο σύνθετα στον υπολογισμό. Ξεκινάμε με μια σύντομη αναφορά στο βασικό θεώρημα που μας επιτρέπει τον υπολογισμό διπλών ολοκληρωμάτων, το Θεώρημα του Fubini. Στο υπόλοιπο κεφάλαιο, η ανάπτυξη της θεωρίας είναι ανάλογη με αυτή του αντίστοιχου κεφαλαίου για τα ζεύγη διακριτών Τ.Μ.

2.1 Θεώρημα του Fubini

Θεώρημα 2.1. (Fubini)

1. *Έστω χωρίο $R \subseteq \mathbb{R} \times \mathbb{R}$ της μορφής*

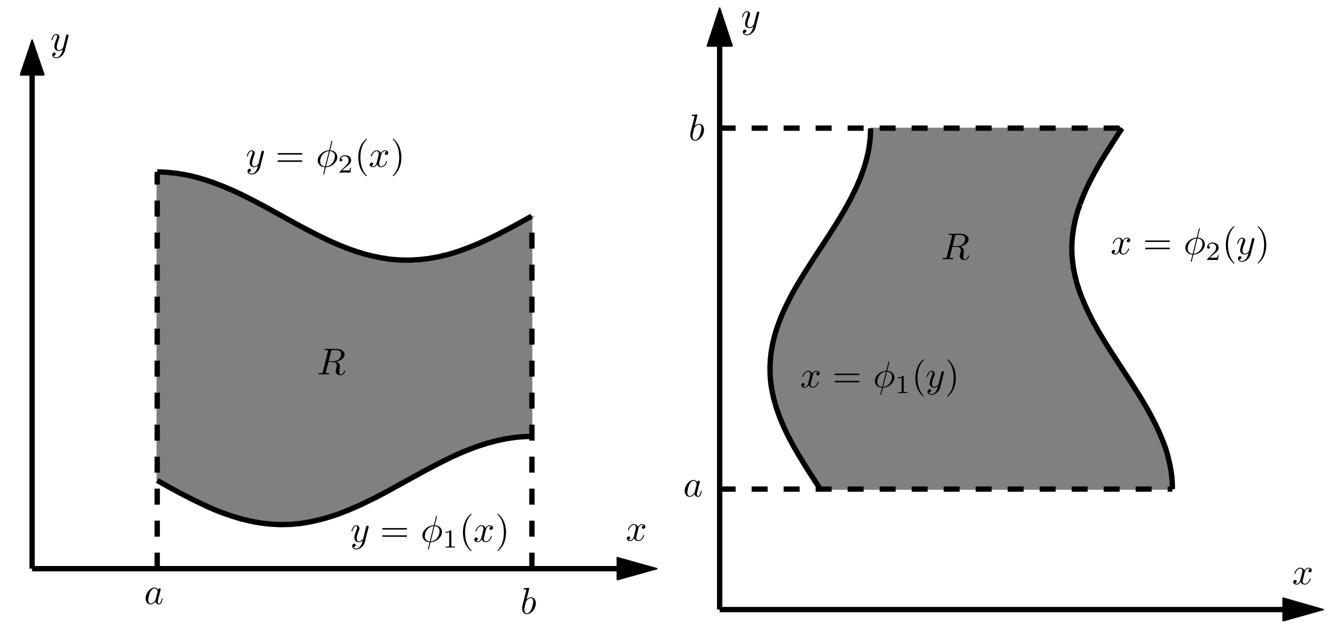
$$R = \{(x, y) : a \leq x \leq b,\ \phi_1(x) \leq y \leq \phi_2(x)\},$$

όπου $a, b \in \mathbb{R}$ με $a < b$, και οι συναρτήσεις $\phi_1(x), \phi_2(x) : [a, b] \to \mathbb{R}$ είναι συνεχείς με $\phi_1(x) \leq \phi_2(x)$ για κάθε $x \in [a, b]$. Έστω επίσης $f(x, y)$ συνεχής συνάρτηση στο R. Το διπλό ολοκλήρωμα της f στο R υπάρχει και μπορεί να υπολογιστεί μέσω μιας επαλληλίας απλών ολοκληρωμάτων ως εξής:

$$\iint\limits_R f(x, y)\, dA = \int_a^b \left(\int_{\phi_1(x)}^{\phi_2(x)} f(x, y)\, dy \right) dx. \tag{2.1}$$

2. *Έστω χωρίο $R \subseteq \mathbb{R} \times \mathbb{R}$ της μορφής*

$$R = \{(x, y) : a \leq y \leq b,\ \phi_1(y) \leq x \leq \phi_2(y)\},$$



Σχήμα 2.1: Τα χωρία των δύο περιπτώσεων του Θεωρήματος Fubini.

όπου $a, b \in \mathbb{R}$ με $a < b$, και οι συναρτήσεις $\phi_1(y), \phi_2(y) : [a, b] \to \mathbb{R}$ είναι συνεχείς με $\phi_1(y) \leq \phi_2(y)$ για κάθε $y \in [a, b]$. Έστω επίσης $f(x, y)$ συνεχής συνάρτηση στο R. Το διπλό ολοκλήρωμα της f στο R υπάρχει και μπορεί να υπολογιστεί μέσω μιας επαλληλίας απλών ολοκληρωμάτων ως εξής:

$$\iint_R f(x, y)\, dA = \int_a^b \left(\int_{\phi_1(y)}^{\phi_2(y)} f(x, y)\, dx \right) dy. \tag{2.2}$$

Παρατηρήσεις

1. Οι δύο μορφές του Θεωρήματος είναι εντελώς ανάλογες, με τις θέσεις των x, y ανεστραμμένες.

2. Η απόδειξη είναι αρκετά σύνθετη και ξεφεύγει από τους στόχους του μαθήματος.

3. Ενώ τα απλά ολοκληρώματα εκφράζουν *εμβαδόν*, τα διπλά εκφράζουν *όγκο*. Συγκεκριμένα, το διπλό ολοκλήρωμα της $f(x)$ στο R ισούται με τον όγκο του στερεού που περικλείεται μεταξύ του R και της $f(x)$, όπου αυτή είναι θετική, μείον τον όγκο του στερεού που περικλείεται μεταξύ του R και της $f(x)$, όπου αυτή είναι αρνητική.

4. Διαισθητικά, το θεώρημα μας λέει πως για να υπολογίσουμε τον όγκο ενός στερεού με αυθαίρετο σχήμα (το διπλό ολοκλήρωμα) μπορούμε να το κόψουμε σε πολύ λεπτές φέτες, να υπολογίσουμε το εμβαδόν κάθε μιας από αυτές (το εσωτερικό

ολοκλήρωμα), και μετά να προσθέσουμε τα εμβαδά αφού τα πολλαπλασιάσουμε με το (πολύ μικρό) πάχος τους (εκτελώντας έτσι την εξωτερική ολοκλήρωση).

5. Συχνά ένα διπλό ολοκλήρωμα μπορεί να υπολογιστεί με χρήση και των δύο μορφών του Θεωρήματος Fubini. Σε πολλές περιπτώσεις, όμως, η μια μορφή οδηγεί σε πολύ απλούστερους υπολογισμούς από την άλλη.

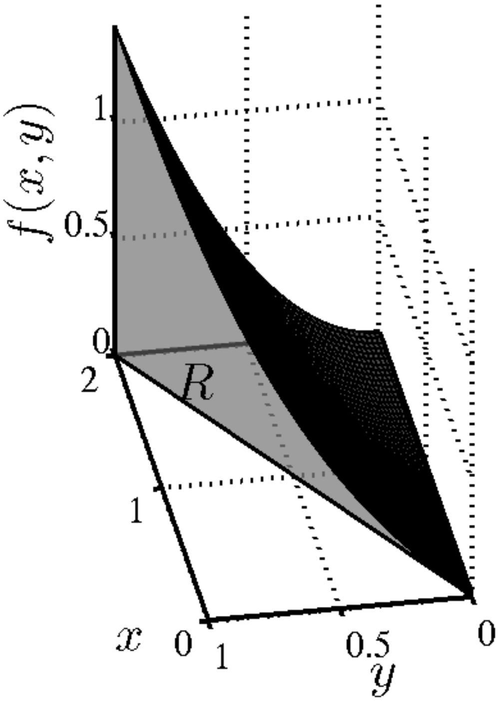
6. Όταν το R είναι ορθογώνιο, δηλαδή $R = [a,b] \times [c,d]$, έχουμε
$$\iint_{[a,b]\times[c,d]} f(x,y)\,dA = \begin{cases} \int_a^b \left(\int_c^d f(x,y)\,dy \right) dx \\ \int_c^d \left(\int_a^b f(x,y)\,dx \right) dy \end{cases}$$

7. Σε εξισώσεις όπου εμφανίζεται επαλληλία απλών ολοκληρωμάτων, όπως στις (2.1), (2.2), μερικές φορές θα παραλείπουμε να γράφουμε τις παρενθέσεις που χωρίζουν την εσωτερική από την εξωτερική ολοκλήρωση. Θα πρέπει πάντα να είμαστε προσεκτικοί στην αντιστοίχιση της κάθε μεταβλητής ολοκλήρωσης με τα σωστά όρια.

8. Συνηθίζεται να γράφουμε και $\iint_R f(x,y)\,dxdy$ αντί για $\iint_R f(x,y)\,dA$. Δεν θα χρησιμοποιήσουμε αυτή τη σύμβαση, για να αποφύγουμε τις παρανοήσεις.

Παράδειγμα 2.1. Σαν ένα απλό παράδειγμα, θα υπολογίσουμε το ολοκλήρωμα της $f(x,y) = y^2\sqrt{x}$ στο τρίγωνο R που ορίζεται από τα σημεία $(0,0), (2,0), (2,1)$. Το τρίγωνο, και η συνάρτηση σε αυτό, εμφανίζονται στο Σχήμα 2.2. Το ολοκλήρωμα που καλούμαστε να υπολογίσουμε είναι ο όγκος του σκιασμένου στερεού του σχήματος. Θα υπολογίσουμε το ολοκλήρωμα και με τις δύο μεθόδους που μας δίνει το Θεώρημα του Fubini.

Καταρχήν παρατηρούμε πως
$$R = \left\{ 0 \le x \le 2, \quad 0 \le y \le \frac{x}{2} \right\}.$$
Άρα,
$$\begin{aligned}
\iint_R f(x,y)\,dA &= \int_0^2 \left(\int_0^{\frac{x}{2}} y^2\sqrt{x}\,dy \right) dx = \int_0^2 \left(\int_0^{\frac{x}{2}} \left(\frac{y^3}{3}\sqrt{x} \right)' dy \right) dx \\
&= \int_0^2 \left[\frac{y^3}{3}\sqrt{x} \right]_0^{\frac{x}{2}} dx = \frac{1}{24}\int_0^2 x^{\frac{7}{2}}\,dx \\
&= \frac{1}{108}\int_0^2 \left(x^{\frac{9}{2}} \right)' dx = \frac{1}{108}\left[x^{\frac{9}{2}} \right]_0^2 = \frac{4\sqrt{2}}{27}.
\end{aligned}$$
Εναλλακτικά, παρατηρούμε πως το τρίγωνο μπορεί να περιγραφεί και ως εξής:
$$R = \{0 \le y \le 1, \quad 2y \le x \le 2\}.$$



Σχήμα 2.2: Το στερεό του Παραδείγματος 2.1.

Άρα,

$$\begin{aligned}
\iint_R f(x,y)\,dA &= \int_0^1 \left(\int_{2y}^2 y^2\sqrt{x}\,dx \right) dy = \int_0^1 \left(\int_{2y}^2 \left(\frac{2}{3} x^{\frac{3}{2}} y^2 \right)' \right) dy \\
&= \int_0^1 \left[\frac{2}{3} x^{\frac{3}{2}} y^2 \right]_{2y}^2 dy = \frac{4\sqrt{2}}{3} \int_0^1 \left(y^2 - y^{\frac{7}{2}} \right) dy \\
&= \frac{4\sqrt{2}}{3} \int_0^1 \left(\frac{1}{3} y^3 - \frac{2}{9} y^{\frac{9}{2}} \right)' dy = \frac{4\sqrt{2}}{3} \left[\frac{1}{3} y^3 - \frac{2}{9} y^{\frac{9}{2}} \right]_0^1 = \frac{4\sqrt{2}}{27}.
\end{aligned}$$

2.2 Ζεύγη Συνεχών Τυχαίων Μεταβλητών

Ορισμός 2.1. *Δύο Τ.Μ. X, Y καλούνται <u>από κοινού συνεχείς</u> όταν υπάρχει μια συνάρτηση $f_{X,Y}(x,y) : \mathbb{R}\times\mathbb{R} \to [0,\infty)$, η <u>από κοινού συνάρτηση πυκνότητας πιθανότητας</u>, ή <u>από κοινού πυκνότητα</u>, ή απλώς <u>πυκνότητα</u>, τέτοια ώστε για οποιοδήποτε χωρίο $R \subseteq \mathbb{R}\times\mathbb{R}$, να ισχύει ότι*

$$P\left[(X,Y)\in R\right] = \iint\limits_{R} f_{X,Y}(x,y)\,dA. \tag{2.3}$$

Παρατηρήσεις

1. Στην περίπτωση που έχουμε μια Τ.Μ., οι πιθανότητες που την αφορούν υπολογίζονται μέσω απλών ολοκληρωμάτων της πυκνότητας και ισοδυναμούν με *εμβαδά*. Στην περίπτωση δύο Τ.Μ., οι πιθανότητες υπολογίζονται μέσω διπλών ολοκληρωμάτων και επομένως ισοδυναμούν με *όγκους*! Έστω, για παράδειγμα, πως μας δίνεται η συνάρτηση $f_{X,Y}(x,y)$ του Σχήματος 2.3(a). Η πιθανότητα το ζεύγος (X,Y) να βρίσκεται σε κάποιο χωρίο R ισούται με τον όγκο του στερεού που οριοθετείται από το χωρίο R και το γράφημα της συνάρτησης. Δείτε για παράδειγμα το Σχήμα 2.3(b).

2. Στην απλούστερη περίπτωση, το χωρίο R είναι κλειστό και φραγμένο, και η συνάρτηση $f_{X,Y}$ είναι ολοκληρώσιμη σε αυτό με την αυστηρή έννοια του όρου.

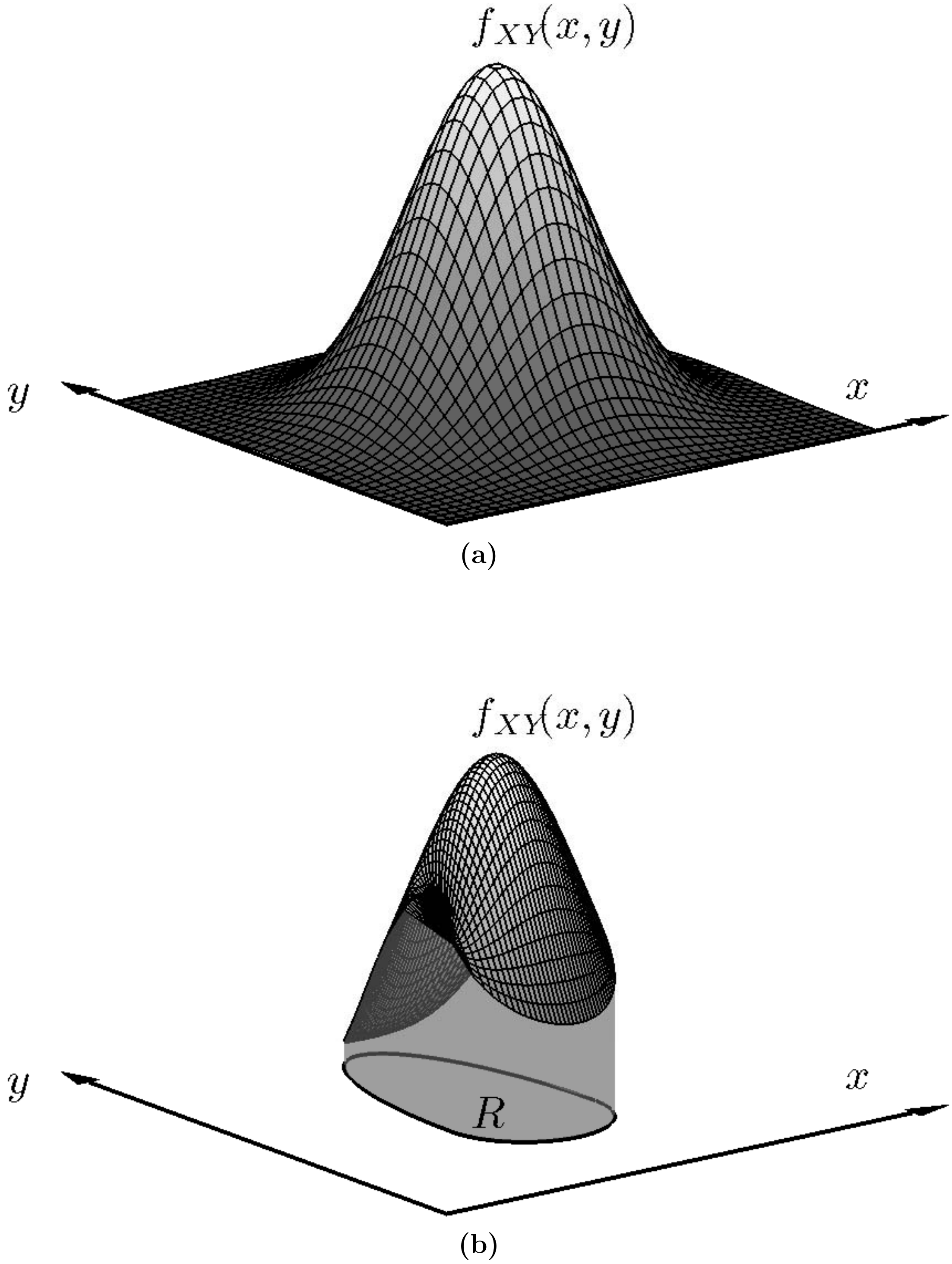
3. Για παράδειγμα, όταν το R είναι καρτεσιανό γινόμενο της μορφής $R = [a,b]\times[c,d]$, η (2.3) γίνεται

$$P\left[(X,Y)\in R\right] = \iint\limits_{[a,b]\times[c,d]} f_{X,Y}(x,y)\,dA = \begin{cases} \int_a^b \left(\int_c^d f_{X,Y}(x,y)dy\right)dx, \\ \int_c^d \left(\int_a^b f_{X,Y}(x,y)dx\right)dy. \end{cases} \tag{2.4}$$

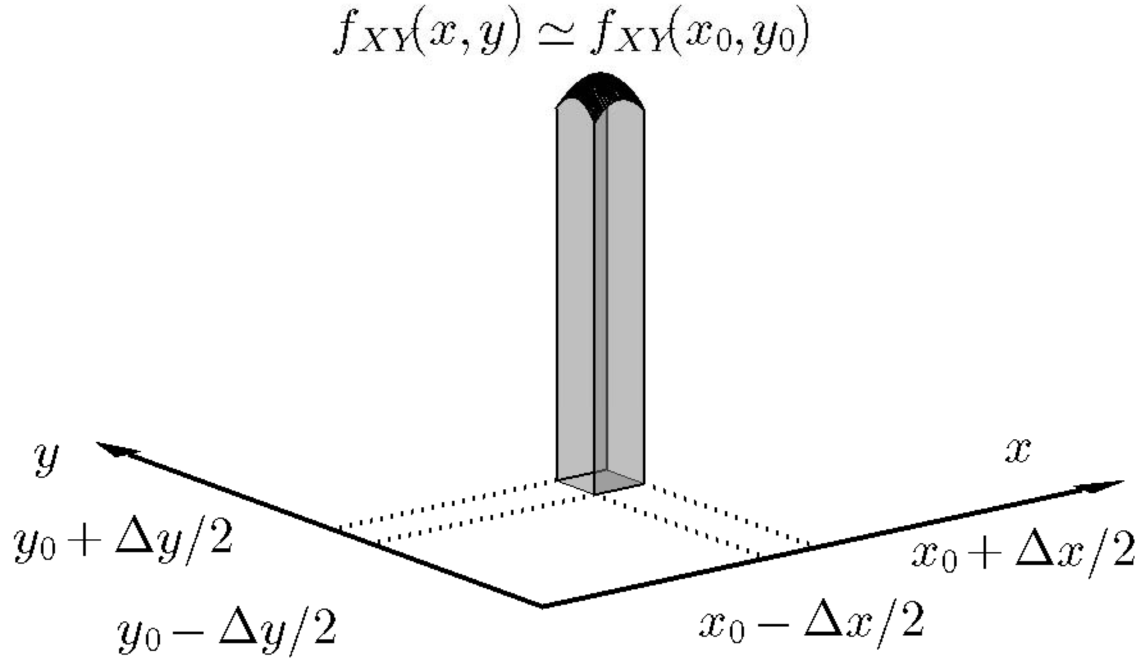
   Παρατηρήστε ότι το διπλό ολοκλήρωμα έχει γραφεί, με δύο διαφορετικούς τρόπους, ως δύο διαδοχικά απλά, με χρήση του Θεωρήματος του Fubini.

4. Στην γενικότερη περίπτωση όμως, το R μπορεί να μην είναι κλειστό (οπότε λαμβάνουμε το ολοκλήρωμα στην κλειστότητά του), ή να είναι ένωση διακριτών χωρίων (οπότε προσθέτουμε τα αντίστοιχα ολοκληρώματα), ή μπορεί το ολοκλήρωμα να υπάρχει μόνο ως καταχρηστικό. (Τα καταχρηστικά διπλά ολοκληρώματα ορίζονται ανάλογα με τα απλά καταχρηστικά ολοκληρώματα.)

5. Σε κάθε περίπτωση, πρέπει να υπολογίσουμε ένα ή περισσότερα διπλά ολοκληρώματα, πάντα με χρήση του Θεωρήματος του Fubini.



Σχήμα 2.3: **(a)** Μια από κοινού πυκνότητα πιθανότητας $f_{X,Y}(x,y)$. **(b)** Η πιθανότητα $P[(X,Y) \in R]$ το ζεύγος (X,Y) να είναι στο χωρίο R δίνεται από τον όγκο του σκιασμένου στερεού που βρίσκεται ανάμεσα στο γράφημα της $f_{X,Y}(x,y)$ και το R.



Σχήμα 2.4: Καθώς τα $\Delta x, \Delta y \to 0$, το στερεό που δημιουργείται ανάμεσα στο γράφημα της $f_{X,Y}(x,y)$ και το ορθογώνιο $\left[x_0 - \frac{\Delta x}{2}, x_0 + \frac{\Delta x}{2}\right] \times \left[y_0 - \frac{\Delta y}{2}, y_0 + \frac{\Delta y}{2}\right]$ τείνει σε ένα ορθογώνιο παραλληλεπίπεδο, και ο όγκος του ισούται, προσεγγιστικά, με $f_{X,Y}(x_0,y_0)\Delta x \Delta y$.

6. Αρκετά συχνά, ο υπολογισμός της πιθανότητας με την μία μέθοδο του Θεωρήματος του Fubini είναι αρκετά απλούστερος από τον υπολογισμό με την άλλη μέθοδο.

7. Διαισθητικά, η από κοινού πυκνότητα στη θέση (x_0,y_0) εκφράζει την πιθανότητα το ζεύγος (X,Y) να έχει τιμές «κοντά» στο (x_0,y_0). Πράγματι, έστω πως η $f(x,y)$ είναι συνεχής στο (x_0,y_0). Τότε

$$
\begin{aligned}
&P\left(x_0 - \frac{\Delta x}{2} \le X \le x_0 + \frac{\Delta x}{2}, y_0 - \frac{\Delta y}{2} \le Y \le y_0 + \frac{\Delta y}{2}\right) \\
&= P\left((X,Y) \in \left[x_0 - \frac{\Delta x}{2}, x_0 + \frac{\Delta x}{2}\right] \times \left[y_0 - \frac{\Delta y}{2}, y_0 + \frac{\Delta y}{2}\right]\right) \\
&= \iint\limits_{\left[x_0 - \frac{\Delta x}{2}, x_0 + \frac{\Delta x}{2}\right] \times \left[y_0 - \frac{\Delta y}{2}, y_0 + \frac{\Delta y}{2}\right]} f_{X,Y}(x_0,y_0)\,dy\,dx \simeq f_{X,Y}(x_0,y_0)\Delta x \Delta y,
\end{aligned}
$$

όπου το $\simeq$ σημαίνει ότι όταν τα Δx, Δy τείνουν στο 0, τότε το πηλίκο των δύο σκελών αριστερά και δεξιά του $\simeq$ τείνει στη μονάδα. Αυτό προκύπτει από μια γενίκευση του Θεμελιώδους Θεωρήματος του Ολοκληρωτικού Λογισμού στις 2 διαστάσεις, και εξηγείται διαισθητικά στο Σχήμα 2.4. Συνεπώς, όσο πιο μεγάλη είναι η τιμή της πυκνότητας, τόσο πιο πιθανό είναι το ζεύγος (X, Y) να βρεθεί «κοντά» στο (x_0, y_0), αν και βέβαια η πιθανότητα να πάρει ακριβώς την τιμή (x_0, y_0) είναι μηδέν.

8. Όπως και στην περίπτωση της μιας Τ.Μ.:

   (α′) Αν αλλάξουμε την πυκνότητα σε πεπερασμένο πλήθος σημείων, δεν αλλάζει το ολοκλήρωμά της σε οποιοδήποτε χωρίο. Άρα, ένα ζεύγος Τ.Μ. μπορεί να περιγράφεται επακριβώς, όσον αφορά την πιθανότητα να βρίσκεται σε κάποιο χωρίο, από περισσότερες από μια από κοινού πυκνότητες.

   (β′) Για ορισμένα περίπλοκα χωρία R, το ολοκλήρωμα (2.3) ενδεχομένως να μην ορίζεται, ακόμα και αν η $f_{X,Y}$ είναι πολύ απλή, π.χ., σταθερή. Αγνοούμε το πρόβλημα.

   (γ′) Οποιαδήποτε συνάρτηση $f_{X,Y}(x, y) \geq 0$ έχει ολοκλήρωμα

$$\iint_{\mathbb{R}\times\mathbb{R}} f_{X,Y}(x, y)\, dA = 1$$

   μπορεί να θεωρηθεί πυκνότητα ενός ζεύγους Τ.Μ.

Παράδειγμα 2.2. (Απλή πυκνότητα πιθανότητας) Έστω πως το ζεύγος (X, Y) έχει την από κοινού πυκνότητα πιθανότητας

$$f(x, y) = \begin{cases} \frac{1}{3}(2x + 4y), & 0 \leq x \leq 1,\ 0 \leq y \leq 1, \\ 0, & \text{αλλού}, \end{cases}$$

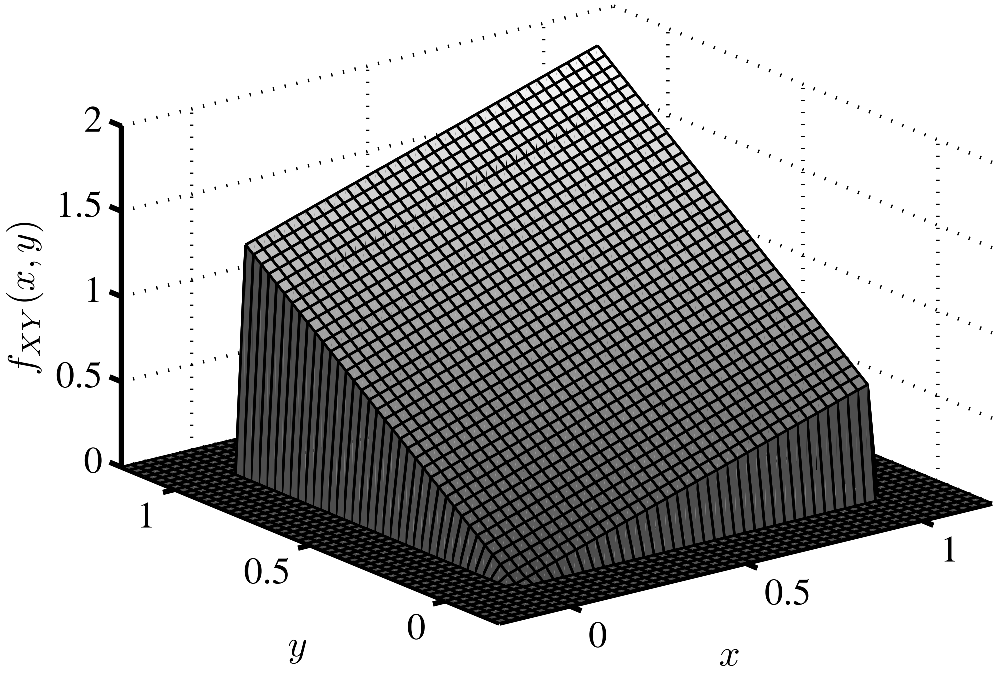
η οποία έχει σχεδιαστεί στο Σχήμα 2.5.

1. Θα υπολογίσουμε την πιθανότητα $P\left(0 \leq X \leq \frac{1}{2},\ 0 \leq Y \leq \frac{1}{2}\right)$.

2. Θα υπολογίσουμε την πιθανότητα $P(X < Y)$.

Καταρχάς, παρατηρούμε πως:

$$\begin{aligned}
\iint_{\mathbb{R}\times\mathbb{R}} f_{X,Y}(x, y)\, dA &= \iint_{[0,1]\times[0,1]} \frac{2x}{3}\, dA + \iint_{[0,1]\times[0,1]} \frac{4y}{3}\, dA \\
&= \int_0^1 \frac{2x}{3} \left(\int_0^1 dy\right) dx + \int_0^1 \frac{4y}{3} \left(\int_0^1 dx\right) dy \\
&= \int_0^1 \frac{2x}{3}\, dx + \int_0^1 \frac{4y}{3}\, dy = \int_0^1 \left(\frac{x^2}{3}\right)' dx + \int_0^1 \left(\frac{2y^2}{3}\right)' dy \\
&= \frac{1}{3} + \frac{2}{3} = 1.
\end{aligned}$$



Σχήμα 2.5: Η πυκνότητα πιθανότητας του Παραδείγματος 2.2. Εκτός των ορίων $0 \le x \le 1$, $0 \le y \le 1$, η πυκνότητα είναι 0, ενώ εντός είναι γραμμική ως προς τα x, y.

Στην πρώτη ισότητα χρησιμοποιήσαμε την γραμμικότητα του διπλού ολοκληρώματος, δηλαδή ότι το ολοκλήρωμα ενός αθροίσματος συναρτήσεων ισούται με το άθροισμα των αντίστοιχων ολοκληρωμάτων. Στην δεύτερη, εφαρμόσαμε το Θεώρημα του Fubini για κάθε ολοκλήρωμα, βγάζοντας έξω από κάθε εσωτερικό ολοκλήρωμα συντελεστές που δεν εξαρτώνται από την εσωτερική μεταβλητή ολοκλήρωσης. Το τελικό αποτέλεσμα αναμενόταν, καθώς η πιθανότητα να πάρει το ζεύγος (X, Y) οποιαδήποτε τιμή *πρέπει* να είναι μονάδα.

Για να υπολογίσουμε τις ζητούμενες πιθανότητες, θα χρησιμοποιήσουμε τον ορισμό της από κοινού πυκνότητας (2.3), ολοκληρώνοντας την στα αντίστοιχα χωρία. Στην πρώτη περίπτωση, έχουμε:

$$\begin{aligned}
&P\left(0 \le X \le \frac{1}{2},\ 0 \le Y \le \frac{1}{2}\right) \\
&= \iint_{[0,\frac{1}{2}]\times[0,\frac{1}{2}]} \left(\frac{2x}{3} + \frac{4y}{3}\right) dA = \int_0^{\frac{1}{2}} \left(\int_0^{\frac{1}{2}} \left(\frac{2x}{3} + \frac{4y}{3}\right) dy\right) dx \\
&= \int_0^{\frac{1}{2}} \left[\frac{2xy}{3} + \frac{2y^2}{3}\right]_0^{\frac{1}{2}} dx = \int_0^{\frac{1}{2}} \left(\frac{x}{3} + \frac{1}{6}\right) dx = \int_0^{\frac{1}{2}} \left(\frac{x^2}{6} + \frac{x}{6}\right)' dx = \frac{1}{8}.
\end{aligned}$$

Παρατηρήστε ότι εδώ επιλέξαμε να μην χρησιμοποιήσουμε την γραμμικότητα του ολοκληρώματος, και έτσι εφαρμόσαμε το Θεώρημα του Fubini μια φορά. Το στερεό του οποίου τον όγκο υπολογίσαμε έχει σχεδιαστεί στο Σχήμα 2.6.

Στην δεύτερη περίπτωση, το χωρίο έχει τριγωνικό σχήμα. Κατά τα γνωστά από το Θεώρημα Fubini, έχουμε:

$$\begin{aligned} P(X<Y) &= \int_0^1 \left(\int_0^y \left(\frac{2x}{3} + \frac{4y}{3} \right) dx \right) dy = \int_0^1 \left(\int_0^y \left(\frac{x^2}{3} + \frac{4yx}{3} \right)' dx \right) dy \\ &= \int_0^1 \left(\frac{y^2}{3} + \frac{4y^2}{3} \right) dy = \frac{5}{3} \int_0^1 y^2\, dy = \frac{5}{9} \int_0^1 (y^3)'\, dy = \frac{5}{9}. \end{aligned}$$

Το στερεό του οποίου τον όγκο υπολογίσαμε φαίνεται σκιασμένο στο Σχήμα 2.6.

Λήμμα 2.1. (Βασικές ιδιότητες της από κοινού πυκνότητας) *Έστω από κοινού συνεχείς Τ.Μ. με από κοινού πυκνότητα $f_{X,Y}(x,y)$. Ισχύουν τα ακόλουθα:*

1. $\displaystyle\iint_{\mathbb{R}\times\mathbb{R}} f(x)\, dA = 1.$

2. *Το ορθογώνιο χωρίο R με πλευρές $x=a$, $x=b$, $y=c$, $y=d$, όπου $a \le b$, $c \le d$, έχει πιθανότητα*

$$P((X,Y)\in R) = \iint_{[a,b]\times[c,d]} f_{X,Y}(x,y)\, dA = \begin{cases} \int_a^b \left(\int_c^d f_{X,Y}(x,y) dy \right) dx, \\ \int_c^d \left(\int_a^b f_{X,Y}(x,y) dx \right) dy, \end{cases}$$

ανεξάρτητα από το αν και πόσα από τα οριακά του σημεία ανήκουν σε αυτό ή όχι.

3. *$P(X=a, Y=b) = 0$ για οποιοδήποτε ζεύγος $a,b\in\mathbb{R}$.*

4. *Η X είναι συνεχής με πυκνότητα*

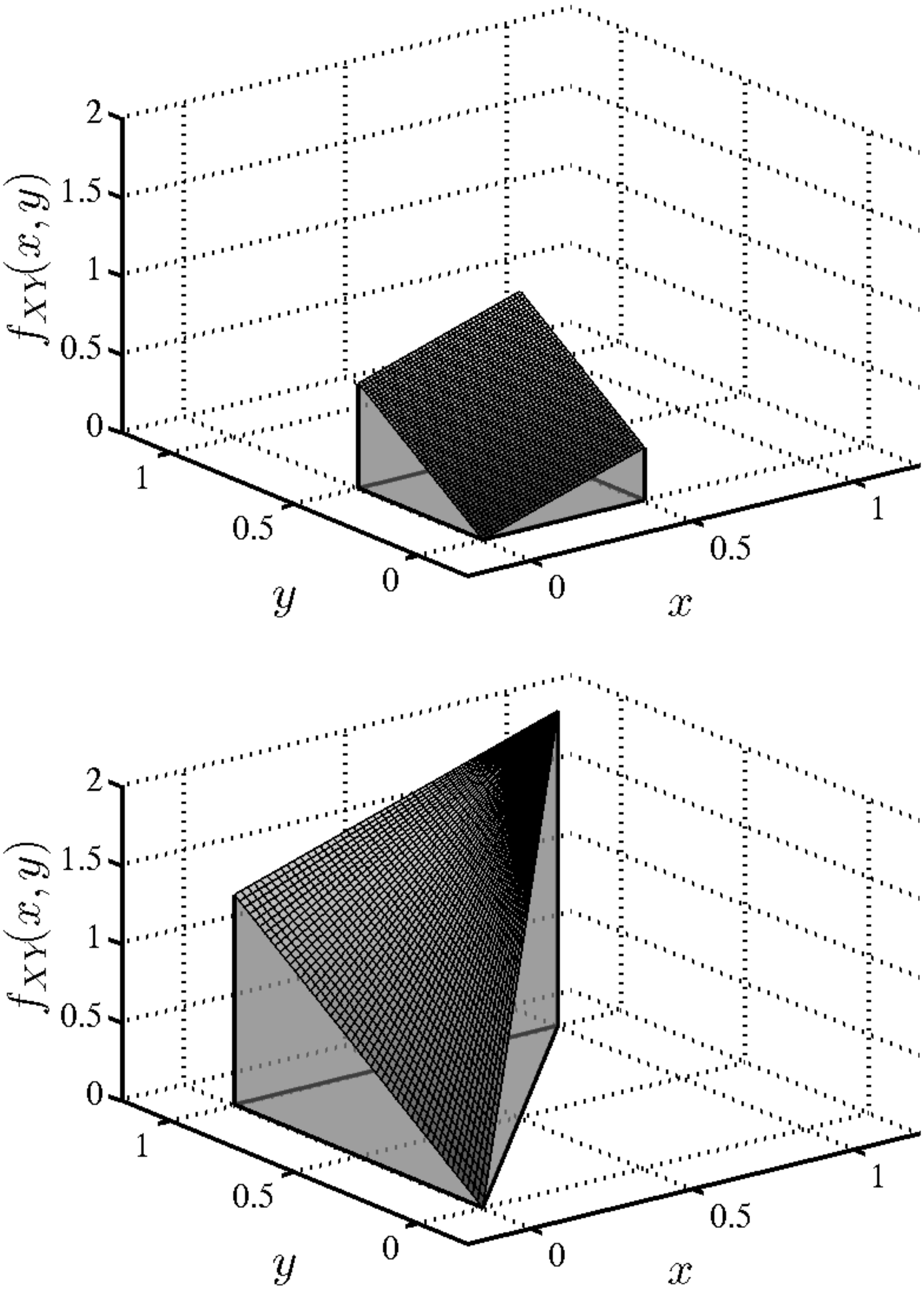
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy \quad \text{για κάθε } x\in\mathbb{R}. \tag{2.5}$$

5. *Η Y είναι συνεχής με πυκνότητα*

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx \quad \text{για κάθε } y\in\mathbb{R}. \tag{2.6}$$

Απόδειξη. 1. Προκύπτει από το γεγονός ότι η πιθανότητα να πάρει το ζεύγος (X,Y) οποιαδήποτε τιμή, που ισούται με το δοσμένο ολοκλήρωμα, πρέπει να είναι μονάδα.

2. Προκύπτει από την σχέση (2.3).



Σχήμα 2.6: Οι όγκοι των στερεών ισούνται με τις πιθανότητες $P\left[0 \leq X \leq \frac{1}{2},\ 0 \leq Y \leq \frac{1}{2}\right]$ και $P[X < Y]$ του Παραδείγματος 2.2.

3. Προκύπτει από την σχέση (2.3).

4. Παρατηρήστε πως, για οποιοδήποτε $A \subseteq \mathbb{R}$,

$$P(X \in A) = P(X \in A, Y \in \mathbb{R}) = \iint\limits_{A\times\mathbb{R}} f_{X,Y}(x,y)\,dA$$

$$= \int_A \left(\int_{\mathbb{R}} f_{X,Y}(x,y)\,dy \right) dx \Rightarrow P(X \in A) = \int_A \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \right) dx.$$

Άρα, βρήκαμε μια συνάρτηση $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$ η οποία λαμβάνει μη αρνητικές τιμές και η οποία επιπλέον έχει την ιδιότητα, για οποιοδήποτε $A \subseteq \mathbb{R}$,

$$P(X \in A) = \int_A f_X(x)\,dx.$$

Άρα εξ ορισμού των συνεχών Τ.Μ., η X είναι συνεχής, με πυκνότητα που δίνεται από την (2.5).

5. Προκύπτει εντελώς ανάλογα με την προηγούμενη περίπτωση.

□

Ορισμός 2.2. *Αν οι X, Y είναι από κοινού συνεχείς, οι πυκνότητές τους $f_X(x)$, $f_Y(y)$ καλούνται* <u>*περιθώριες πυκνότητες πιθανότητας*</u>.

Παράδειγμα 2.3. Θα υπολογίσουμε τις περιθώριες πυκνότητες του Παραδείγματος 2.2. Για να υπολογίσουμε την $f_X(x)$, παρατηρούμε πως αν $x < 0$ ή $x > 1$, τότε προκύπτει ότι $f_X(x) = 0$. Πράγματι, σε αυτή την περίπτωση η $f_X(x)$ ισούται με το ολοκλήρωμα μιας συνάρτησης που είναι παντού 0.

Αν $0 \leq x \leq 1$, τότε

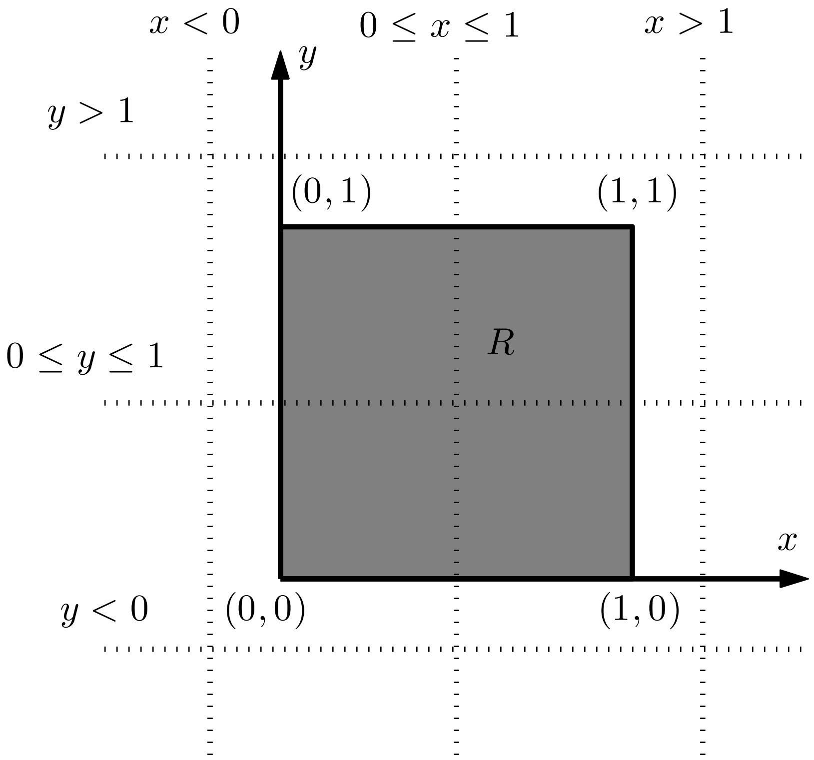
$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^1 \left(\frac{2x}{3} + \frac{4y}{3} \right) dy \\ &= \frac{2x}{3}\int_0^1 dy + \frac{4}{6}\int_0^1 \left(y^2\right)'\,dy = \frac{2}{3}(x+1). \end{aligned}$$

Συγκεντρωτικά:

$$f_X(x) = \begin{cases} \frac{2}{3}(x+1), & x \in [0,1], \\ 0, & x \notin [0,1]. \end{cases}$$

Παρομοίως,

$$f_Y(y) = \begin{cases} \frac{1}{3}(4y+1), & y \in [0,1], \\ 0, & y \notin [0,1]. \end{cases}$$



Σχήμα 2.7: Παράδειγμα 2.3: Η από κοινού πυκνότητα είναι μη μηδενική στο χωρίο $R = [0, 1] \times [0, 1]$. Για να υπολογίσουμε τις περιθώριες πυκνότητες $f_X(x)$, $f_Y(y)$, πρέπει να ολοκληρώσουμε την από κοινού κατά μήκος ευθειών καθέτων ή οριζοντίων (αντίστοιχα), έχοντας την μεταβλητή που δεν ολοκληρώνεται, x ή y αντίστοιχα, ως μια παράμετρο που θα επηρεάσει την τιμή του ολοκληρώματος μέσω της θέσης της ευθείας.

Στο Σχήμα 2.7 έχουμε σχεδιάσει το χωρίο R στο οποίο η $f_{X,Y}(x, y)$ είναι μη μηδενική, καθώς και τις διάφορες ευθείες πάνω στις οποίες υπολογίσαμε το (απλό) ολοκλήρωμα της από κοινού πυκνότητας, προκειμένου να υπολογίσουμε τις περιθώριες πυκνότητες.

2.3 Μέση Τιμή και Συνδιακύμανση

Πρόταση 2.1. (Μέση τιμή συνάρτησης δύο Τ.Μ.) *Έστω δύο από κοινού συνεχείς Τ.Μ. X, Y, με από κοινού πυκνότητα $f_{X,Y}(x,y)$. Έστω Τ.Μ. $Z = g(X,Y)$. Η μέση τιμή της ισούται με*

$$E(Z) = \iint\limits_{\mathbb{R}\times\mathbb{R}} g(x,y) f_{X,Y}(x,y)\, dA = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\, dxdy. \tag{2.7}$$

Παρατήρηση: Η πρόταση μας επιτρέπει να υπολογίσουμε μέσες τιμές τυχαίων μεταβλητών με πολύ λιγότερο κόπο σε σχέση με την περίπτωση που χρησιμοποιούσαμε τον ορισμό της μέσης τιμής. Δείτε το ακόλουθο παράδειγμα.

Παράδειγμα 2.4. Για τις Τ.Μ. X, Y του Παραδείγματος 2.2 έχουμε

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\, dxdy = \frac{1}{3}\int_0^1 \left(\int_0^1 xy(2x+4y)\, dy\right) dx \\ &= \frac{1}{3}\int_0^1 \left(\int_0^1 \left(x^2y^2 + \frac{4}{3}xy^3\right)' dy\right) dx = \frac{1}{3}\int_0^1 \left(x^2 + \frac{4x}{3}\right) dx \\ &= \frac{1}{3}\int_0^1 \left(\frac{x^3}{3} + \frac{2x^2}{3}\right)' dx = \frac{1}{3}\left(\frac{1}{3}+\frac{2}{3}\right) = \frac{1}{3}. \end{aligned}$$

Χωρίς χρήση της πιο πάνω πρότασης, για να υπολογίσουμε την μέση τιμή $E(Z)$ της Τ.Μ. $Z \triangleq XY$, θα έπρεπε

1. να υπολογίσουμε την κατανομή $F_Z(z) = P(Z \le z) = P(XY \le z)$, μέσω του διπλού ολοκληρώματος
$$P(XY \le z) = \iint\limits_{xy \le z} f_{X,Y}(x,y)\, dA = \iint\limits_{xy\le z,\ 0\le x,y\le 1} \left(\frac{2x}{3} + \frac{4y}{3}\right) dA.$$
2. να υπολογίσουμε την πυκνότητα της Z, με χρήση της σχέσης $f_Z(z) = F'_Z(z)$, και
3. κατόπιν να εφαρμόσουμε τον ορισμό της μέσης τιμής $E(Z) = \int_{-\infty}^{\infty} z f_Z(z)\, dz$.

Η διαδικασία θα ήταν πολύ πιο μακροσκελής (ιδιαίτερα το πρώτο βήμα).

Η συνδιακύμανση $\mathrm{Cov}(X,Y)$ και ο συντελεστής συσχέτησης $\rho(X,Y)$ δύο συνεχών Τ.Μ. X, Y ορίζονται ακριβώς όπως και στην διακριτή περίπτωση ως

$$\mathrm{Cov}(X,Y) \triangleq E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big],$$
$$\rho(X,Y) \triangleq \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{V}(X)}\sqrt{\mathrm{V}(Y)}}.$$

Όπου για να ορίσουμε την συνδιακύμανση υποθέτουμε ότι καθεμία από τις X, Y έχει πεπερασμένη δεύτερη ροπή, ενώ για να ορίσουμε τον συντελεστή συσχέτισης υποθέτουμε ότι καθεμία έχει πεπερασμένη και θετική διασπορά.

Παρατήρηση: Οι προτάσεις 1.1, 1.2, 1.3, 1.4, 1.5 ισχύουν και στην περίπτωση συνεχών τυχαίων μεταβλητών, με την ίδια απόδειξη. Δεν τις διατυπώνουμε ξανά.

Παράδειγμα 2.5. Για τις Τ.Μ. X, Y του Παραδείγματος 2.2 έχουμε, χρησιμοποιώντας τις περιθώριες πυκνότητες που βρήκαμε στο Παράδειγμα 2.3:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_0^1 x \frac{2}{3}(x+1)\, dx \\ &= \int_0^1 \left(\frac{2x^3}{9} + \frac{x^2}{3} \right)' dx = \frac{2}{9} + \frac{1}{3} = \frac{5}{9}. \end{aligned}$$

Παρομοίως,

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} y f_Y(y)\, dy = \int_0^1 y \frac{1}{3}(4y+1)\, dy \\ &= \int_0^1 \frac{1}{3} \left(\frac{4y^3}{3} + \frac{y^2}{2} \right)' dy = \frac{1}{3} \left(\frac{4}{3} + \frac{1}{2} \right) = \frac{11}{18}. \end{aligned}$$

Τέλος, στο Παράδειγμα 2.4 βρήκαμε πως $E(XY) = \frac{1}{3}$, άρα τελικά

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{3} - \frac{5}{9} \times \frac{11}{18} = -\frac{1}{162}.$$

Παράδειγμα 2.6. Έστω X, Y συνεχείς τυχαίες μεταβλητές με από κοινού πυκνότητα

$$f(x,y) = \begin{cases} 2 & \text{αν } 0 < x < y < 1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Τότε

$$E(XY) = \iint_{\mathbb{R}^2} xy f(x,y)\, dxdy = 2 \int_0^1 \int_0^y xy\, dxdy = 2 \int_0^1 \frac{y^3}{2}\, dy = \frac{1}{4},$$

και όμοια βρίσκουμε ότι για $r > 0$ ισχύει

$$\begin{aligned} E(X^r) &= \int_0^1 \int_0^y x^r 2\, dxdy = \ldots = \frac{2}{(r+1)(r+2)}, \\ E(Y^r) &= \int_0^1 \int_0^y y^r 2\, dxdy = \ldots = \frac{2}{r+2}. \end{aligned}$$

Άρα χρησιμοποιώντας αυτές τις σχέσεις για $r = 1, 2$, βρίσκουμε

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= E(XY) - EXEY = \frac{1}{4} - \frac{1}{3}\frac{2}{3} = \frac{1}{36}, \\
V(X) &= E(X^2) - (EX)^2 = \frac{2}{3 \times 4} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}, \\
V(Y) &= \frac{2}{4} - \frac{4}{9} = \frac{1}{18}, \\
\rho(X,Y) &= \frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{1/36}{1/18} = \frac{1}{2}.
\end{aligned}$$

2.4 Ανεξάρτητες Συνεχείς Τυχαίες Μεταβλητές

Ορισμός 2.3. (Ζεύγη ανεξάρτητων συνεχών Τ.Μ.) *Δύο συνεχείς Τ.Μ. X, Y καλούνται ανεξάρτητες αν για οποιαδήποτε υποσύνολα $A, B \subseteq \mathbb{R}$ ισχύει*

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B). \tag{2.8}$$

Παρατήρηση: Όπως και στην περίπτωση των διακριτών Τ.Μ., δύο Τ.Μ. είναι ανεξάρτητες αν δύο οποιαδήποτε ενδεχόμενα που αφορούν το καθένα αποκλειστικά τη μία από τις δύο Τ.Μ. είναι ανεξάρτητα. Η ανεξαρτησία λοιπόν και των συνεχών Τ.Μ. είναι άμεσα σχετισμένη με την ανεξαρτησία των ενδεχόμενων.

Λήμμα 2.2. (Ικανό κριτήριο ανεξαρτησίας Τ.Μ.) *Δύο από κοινού συνεχείς Τ. Μ. X, Y, με από κοινού πυκνότητα $f_{X,Y}(x,y)$ και περιθώριες πυκνότητες $f_X(x)$, $f_Y(y)$ είναι ανεξάρτητες αν*

$$f_{X,Y}(x,y) = f_X(x)f_Y(y), \quad \forall\, x, y \in \mathbb{R}. \tag{2.9}$$

Απόδειξη. Έστω πως ισχύει η (2.9) και έστω δύο ενδεχόμενα $A, B \subseteq \mathbb{R}$. Παρατηρούμε πως

$$\begin{aligned} P(X \in A, Y \in B) &= \iint\limits_{A\times B} f_{X,Y}(x,y)\,dA = \int_A \left(\int_B f_X(x)f_Y(y)\,dy \right) dx \\ &= \int_A f_X(x) \left(\int_B f_Y(y)\,dy \right) dx \\ &= \left(\int_B f_Y(y)\,dy \right) \left(\int_A f_X(x)\,dx \right) = P(X \in A)P(Y \in B). \end{aligned}$$

Η δεύτερη ισότητα προέκυψε από εφαρμογή του Θεωρήματος του Fubini (σε μια πιο γενική μορφή από αυτή που έχουμε δει) και την υπόθεση, ενώ η τρίτη και η τέταρτη βγάζοντας σταθερές εκτός ολοκληρωμάτων. □

Παρατηρήσεις

1. Σε αντίθεση με το αντίστοιχο κριτήριο της διακριτής περίπτωσης, το κριτήριο είναι απλώς ικανό, και όχι αναγκαίο. Για να καταλάβετε γιατί, σκεφτείτε ως εξής: Έστω δύο Τ.Μ. X, Y για τις οποίες η (2.9) ικανοποιείται, οπότε οι Τ.Μ. είναι ανεξάρτητες. Αν αλλάξουμε την από κοινού πυκνότητα σε ένα μόνο σημείο, τότε το κριτήριο δεν ικανοποιείται πλέον, όμως οι X, Y συνεχίζουν να είναι ανεξάρτητες, αφού η αλλαγή της από κοινού πυκνότητας σε ένα σημείο δεν μπορεί να επηρεάσει τον υπολογισμό του ολοκληρώματος της σε κανένα σύνολο, άρα και τις τιμές των πιθανοτήτων που εμφανίζονται στα δύο μέλη της (2.8).

2. Αναφέρουμε όμως, χωρίς απόδειξη, ότι αν δύο Τ.Μ. X, Y, με πυκνότητες, αντιστοίχως, $f_X(x)$, $f_Y(y)$, είναι ανεξάρτητες, τότε μπορούμε να θέσουμε ως από κοινού πυκνότητά τους την $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

Λήμμα 2.3. (Ιδιότητες ανεξάρτητων συνεχών Τ.Μ.) *Έστω X, Y ανεξάρτητες συνεχείς Τ.Μ.*

1. *Έστω συναρτήσεις $g : \mathbb{R} \to \mathbb{R}$, και $h : \mathbb{R} \to \mathbb{R}$. Θα ισχύει*

$$E(g(X)h(Y)) = E(g(X))E(h(Y))).$$

Ειδική περίπτωση της άνω είναι η

$$E(XY) = E(X)E(Y).$$

2. $\mathrm{Cov}(X,Y) = 0$.

3. $\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y)$.

Απόδειξη. 1. Εφαρμόζοντας το Λήμμα 2.1 έχουμε:

$$\begin{aligned} E(g(X)h(Y)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{X,Y}(x,y)\,dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)\,dxdy \\ &= \left(\int_{-\infty}^{\infty} g(x)f_X(x)\,dx\right)\left(\int_{-\infty}^{\infty} h(y)f_Y(y)\,dy\right). \end{aligned}$$

Η δεύτερη ισότητα προκύπτει από την υπόθεση της ανεξαρτησίας. Η τρίτη βγάζοντας σταθερές έξω από ολοκληρώματα.

2. Προκύπτει άμεσα από το προηγούμενα σκέλος και την εφαρμογή της (1.3), που ισχύει και για συνεχείς τυχαίες μεταβλητές.

3. Προκύπτει άμεσα από το προηγούμενο σκέλος και την εφαρμογή της (1.4)

□

Ορισμός 2.4. (Ασυσχέτιστες Τ.Μ.) *Δύο συνεχείς Τ.Μ. X, Y, είναι ασυσχέτιστες αν η συνδιακύμανσή τους $\mathrm{Cov}(X,Y)$ είναι μηδενική, δηλαδή $\mathrm{Cov}(X,Y) = 0$.*

Παρατήρηση: Όπως και στη διακριτή περίπτωση, δύο ασυσχέτιστες Τ.Μ. μπορεί να μην είναι ανεξάρτητες. Για παράδειγμα, αν η Z ακολουθεί την ομοιόμορφη κατανομή στο $(-1,1)$, και πάρουμε $X = Z, Y = Z^2$. Τότε

$$\mathrm{Cov}(X,Y) = E(Z \cdot Z^2) - E(Z)E(Z^2) = 0,$$

αφού $E(Z) = E(Z^3) = 0$. Όμως

$$P(X \in (0,1/2), X^2 \in (1/2,1)) = P(\emptyset) = 0 < P(X \in (0,1/2))P(X^2 \in (1/2,1)).$$

Παράδειγμα 2.7. Θα εξετάσουμε κατά πόσον οι X, Y του Παραδείγματος 2.2 είναι ανεξάρτητες. Παρατηρούμε καταρχήν, πως υπάρχουν ζεύγη τιμών των x, y για τα οποία δεν εφαρμόζεται το Λήμμα 2.2 (μπορείτε να βρείτε μερικά;). Παρατηρούμε επίσης πως η συνδιακύμανση, που έχουμε υπολογίσει στο Παράδειγμα 2.5, δεν είναι μηδενική, άρα από το άνω λήμμα προκύπτει πως οι X, Y δεν μπορεί να είναι ανεξάρτητες.

2.5 Παραδείγματα

Παράδειγμα 2.8. Ο Σταύρος και ο Γιάννης έχουν δώσει ραντεβού σε ένα μπαρ, για τις 03:00 π.μ. Έστω X, Y οι χρόνοι καθυστέρησής τους, σε ώρες. Υποθέτουμε ότι οι Τ.Μ. X, Y είναι ανεξάρτητες η μια από την άλλη, και επιπλέον είναι ομοιόμορφα κατανεμημένες από το 0 ως το 1. Θα απαντήσουμε τα ακόλουθα ερωτήματα:

1. Ποια είναι η πιθανότητα η διαφορά στους χρόνους αφίξεων σε απόλυτη τιμή, δηλαδή το $|X - Y|$, να είναι κάτω από 0.5 ώρες;

2. Για να τους αποτρέπει να αργούν, ο μπάρμαν τους χρεώνει στο λογαριασμό, εκτός από τα ποτά, και ένα κόστος $20X + 10Y$ Ευρώ. Κατά μέσο όρο, πόσο πληρώνουν κάθε φορά στον μπάρμαν λόγω της καθυστέρησής τους;

3. Ποια είναι η πιθανότητα να πληρώσουν πάνω από 20 Ευρώ κόστος καθυστέρησης;

4. Πόση είναι η συνδιακύμανση $\mathrm{Cov}(X, Y)$;

Για να απαντήσουμε τα ερωτήματα, παρατηρούμε καταρχήν πως οι περιθώριες κατανομές είναι οι

$$f_X(x) = \begin{cases} 1, & x \in [0,1], \\ 0, & x \notin [0,1], \end{cases} \quad f_Y(y) = \begin{cases} 1, & y \in [0,1], \\ 0, & y \notin [0,1], \end{cases}$$

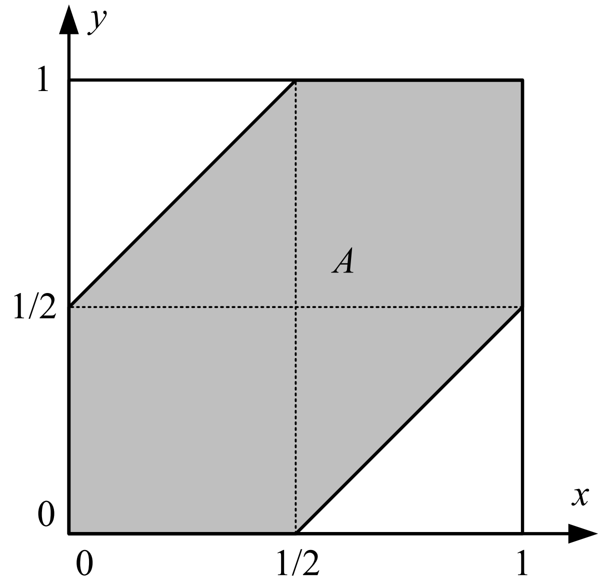
και χρησιμοποιώντας την υπόθεση της ανεξαρτησίας μπορούμε να βρούμε την από κοινού πυκνότητα:

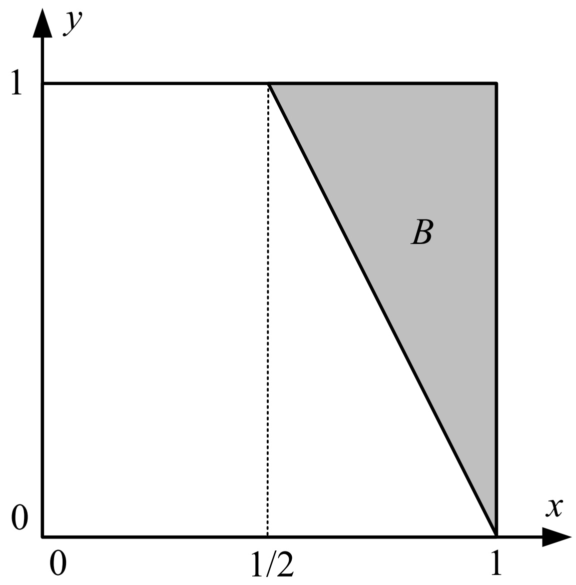
$$f_{X,Y}(x,y) = f_X(x) f_Y(y) = \begin{cases} 1, & (x,y) \in [0,1] \times [0,1], \\ 0, & (x,y) \notin [0,1] \times [0,1]. \end{cases}$$

Έχοντας την από κοινού κατανομή, μπορούμε να απαντήσουμε τα άνω ερωτήματα:

1. $$P\left(|X - Y| < \frac{1}{2}\right) = \iint\limits_{|x-y|<\frac{1}{2}} f_{X,Y}(x,y)\, dA = \iint\limits_{|x-y|<\frac{1}{2}, 0 \le x,y \le 1} 1\, dA.$$

   Στη δεύτερη ισότητα λάβαμε υπ΄ όψιν ότι η ολοκληρωτέα συνάρτηση είναι μηδέν εκτός του τετραγώνου $[0,1] \times [0,1]$, και ίση με τη μονάδα εντός. Το χωρίο $A = \{(x,y) : |x - y| < \frac{1}{2}, 0 \le x, y \le 1\}$ έχει σχεδιαστεί στο Σχήμα 2.8. Επειδή η ολοκληρωτέα συνάρτηση ισούται με τη μονάδα, το ολοκλήρωμα ισούται με το εμβαδόν του A, που είναι $\frac{3}{4}$. Άρα, με πιθανότητα $\frac{3}{4}$ η αναμονή του πρώτου για την άφιξη του δεύτερου δεν θα ξεπεράσει την μισή ώρα.





Σχήμα 2.8: Τα χωρία A, B του Παραδείγματος 2.8. Η από κοινού πυκνότητα είναι μονάδα εντός του τετραγώνου $\{(x, y):\ 0 \leq x, y \leq 1\}$, και μηδενική εκτός.

2. Η μέση τιμή του κόστους είναι εύκολο να υπολογιστεί:
$$E(Z) = E(20X + 10Y) = 20E(X) + 10E(Y) = 20 \times \frac{1}{2} + 10 \times \frac{1}{2} = 15.$$

3. Υπολογίζουμε την πιθανότητα $P(Z > 20)$ ως εξής:
$$\begin{aligned} P(Z > 20) &= P(20X + 10Y > 20) = P(2X + Y > 2) \\ &= \iint_{2x+y>2} f_{X,Y}(x, y)\, dA = \iint_{2x+y>2,\ 0\leq x,y\leq 1} 1\, dA. \end{aligned}$$

   Και το χωρίο $B = \{(x, y) : 2x + y > 2,\ 0 \leq x, y \leq 1\}$ εμφανίζεται σκιασμένο στο Σχήμα 2.8. Το εμβαδόν του είναι $\frac{1}{4}$, άρα, επειδή η πυκνότητα είναι σταθερή και ίση με τη μονάδα, έχουμε $P(Z > 20) = \frac{1}{4}$.

4. Αφού οι X, Y είναι ανεξάρτητες, η συνδιακύμανση $\mathrm{Cov}(X, Y) = 0$.

Παράδειγμα 2.9. Έστω τώρα πως ο Σταύρος και ο Γιάννης φτάνουν σε τυχαίους χρόνους X και Y αντίστοιχα, για τους οποίους γνωρίζουμε ότι η από κοινού πυκνότητα πιθανότητας είναι η
$$f_{X,Y}(x, y) = \begin{cases} cx(1 - y), & (x, y) \in [0, 1] \times [0, 1], \\ 0, & (x, y) \notin [0, 1] \times [0, 1]. \end{cases}$$

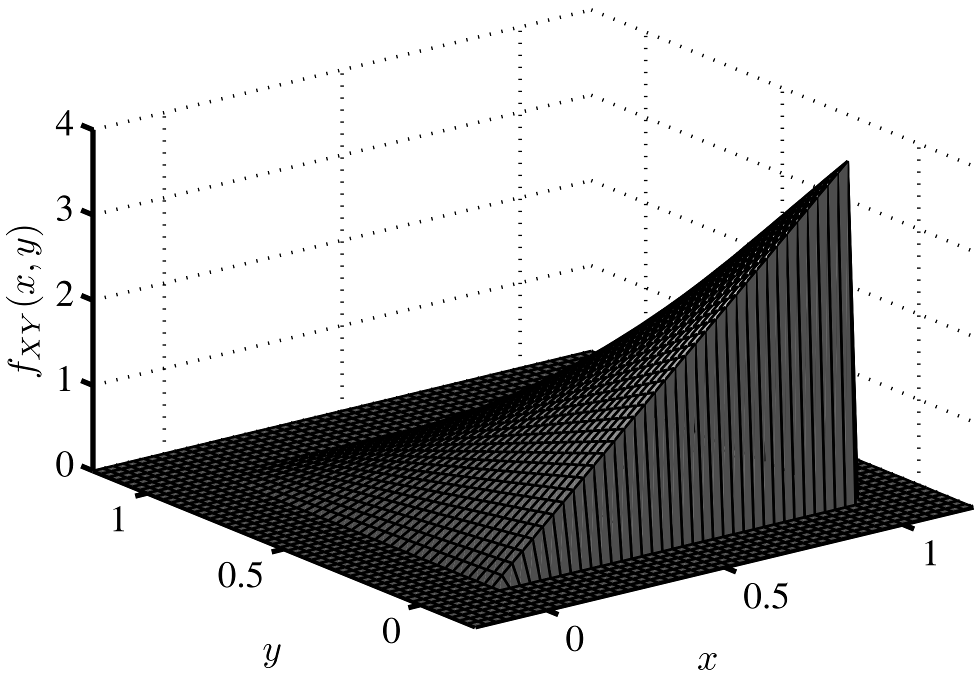
Θα απαντήσουμε τα ακόλουθα ερωτήματα:

1. Ποια είναι η τιμή της σταθεράς c;
2. Ποιες είναι οι περιθώριες πυκνότητες των X, Y;
3. Είναι οι Τ.Μ. X, Y ανεξάρτητες;
4. Πόση είναι η συνδιακύμανση $\mathrm{Cov}(X,Y)$;
5. Πόση είναι η πιθανότητα $P(X \leq Y)$;

Έχουμε:

1. Θα χρησιμοποιήσουμε το γεγονός ότι το ολοκλήρωμα της πυκνότητας σε όλο το επίπεδο πρέπει να ισούται με τη μονάδα. Έχουμε
$$\begin{aligned} \iint\limits_{\mathbb{R}\times\mathbb{R}} f_{X,Y}(x,y)\,dA &= c\int_0^1 \left(\int_0^1 x(1-y)\,dx\right)\,dy \\ &= c\int_0^1 (1-y)\left(\int_0^1 x\,dx\right)\,dy \\ &= c\left(\int_0^1 x\,dx\right)\left(\int_0^1 (1-y)\,dy\right) = c\times\frac{1}{2}\times\frac{1}{2}, \end{aligned}$$
συνεπώς πρέπει $c = 4$. Η πυκνότητα έχει σχεδιαστεί στο Σχήμα 2.9.

2. Κατά τα γνωστά από τη θεωρία,
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy.$$
Προφανώς όταν $x > 1$ ή $x < 0$, το άνω ολοκλήρωμα είναι μηδέν, γιατί τότε η ολοκληρωτέα συνάρτηση είναι παντού μηδενική.
Στην περίπτωση που $0 \leq x \leq 1$, έχουμε:
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^1 4x(1-y)\,dy = 4x\int_0^1 (1-y)\,dy = 2x.$$
Άρα, συγκεντρωτικά
$$f_X(x) = \begin{cases} 2x, & x \in [0,1], \\ 0, & x \notin [0,1]. \end{cases}$$
Παρομοίως, η περιθώρια $f_Y(y) = 0$ όταν $y \notin [0,1]$, ενώ όταν $y \in [0,1]$ έχουμε:
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^1 4x(1-y)\,dx = (1-y)\int_0^1 4x\,dx = 2(1-y),$$



Σχήμα 2.9: Η πυκνότητα πιθανότητας $f_{X,Y}(x,y)$ του Παραδείγματος 2.9.

και συγκεντρωτικά $f_Y(y) = \begin{cases} 2(1-y), & y \in [0,1], \\ 0, & y \notin [0,1]. \end{cases}$

Για να κατανοήσετε καλύτερα τον υπολογισμό των περιθωρίων πυκνοτήτων, μπορείτε να δείτε το Σχήμα 2.7, που εφαρμόζεται ως έχει και σε αυτό το παράδειγμα.

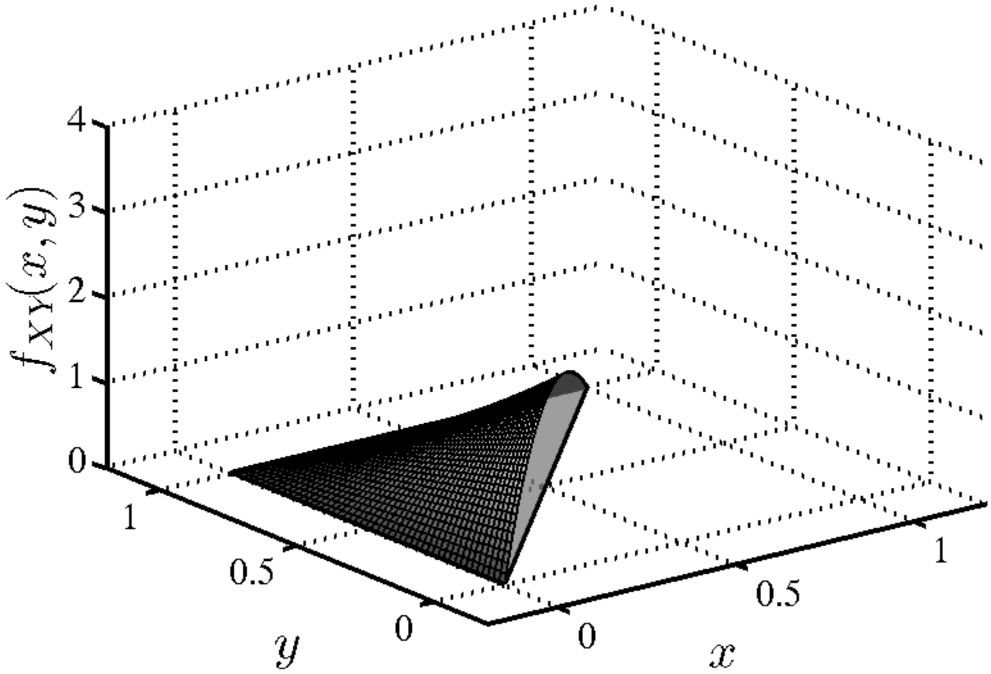
3. Για κάθε ζεύγος $(x,y) \in \mathbb{R} \times \mathbb{R}$, ισχύει ότι $f_{X,Y}(x,y) = f_X(x) f_Y(y)$. (Εξετάστε περιπτώσεις για να σιγουρευτείτε.) Άρα οι X, Y είναι ανεξάρτητες.

4. Αφού οι X, Y είναι ανεξάρτητες, η συνδιακύμανση $\mathrm{Cov}(X,Y) = 0$.

5.

$$\begin{aligned} P(X \leq Y) &= \iint\limits_{(x,y):x\leq y} f_{X,Y}(x,y)\, dA = \int_0^1 \left(\int_0^y 4x(1-y)\, dx \right) dy \\ &= 4 \int_0^1 \left((1-y) \int_0^y x\, dx \right) dy = 4 \int_0^1 (1-y) \frac{y^2}{2}\, dy \\ &= 4 \int_0^1 \left(\frac{y^3}{6} - \frac{y^4}{8} \right)' dy = 4 \left(\frac{1}{6} - \frac{1}{8} \right) = \frac{1}{6}. \end{aligned}$$

Στο Σχήμα 2.10 έχουμε σχεδιάσει το στερεό του οποίου ο όγκος αντιστοιχεί με την άνω πιθανότητα.



Σχήμα 2.10: Ο όγκος του στερεού ισούται με την $P[X \le Y]$ του Παραδείγματος 2.9.

Παράδειγμα 2.10. Έστω, τέλος, πως ο Σταύρος και ο Γιάννης φτάνουν σε τυχαίους χρόνους X και Y αντίστοιχα, για τους οποίους γνωρίζουμε ότι

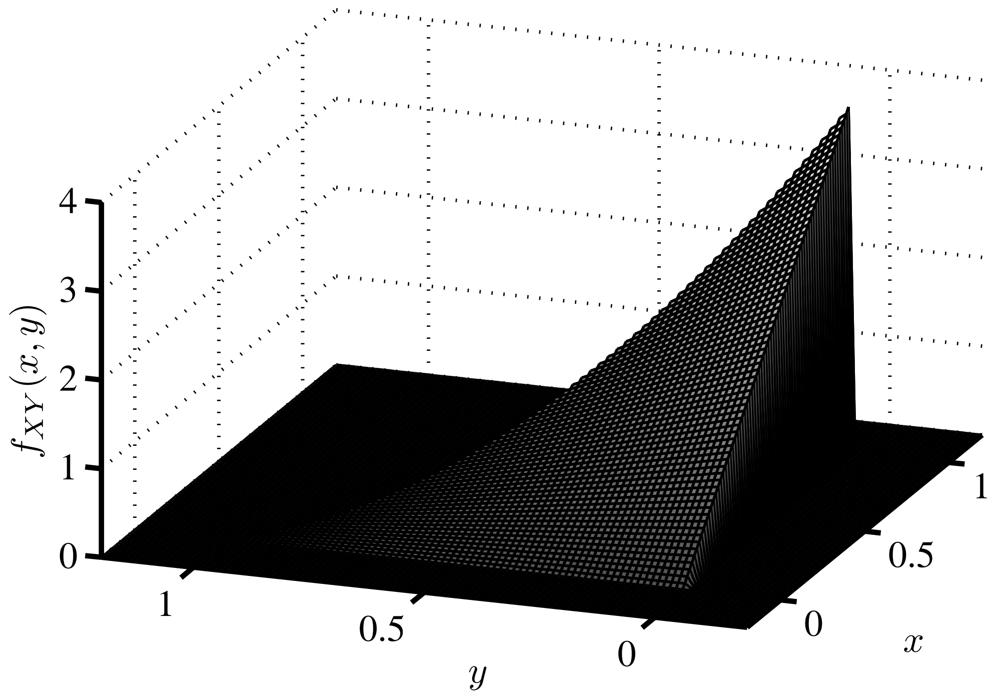
$$f_{X,Y}(x,y) = \begin{cases} cx(1-y), & 0 \le x, y \le 1,\ x+y \le 1, \\ 0, & \text{αλλού.} \end{cases}$$

1. Ποια είναι η τιμή της σταθεράς c;
2. Ποιες είναι οι περιθώριες πυκνότητες των X, Y;
3. Είναι οι Τ.Μ. X, Y ανεξάρτητες;
4. Πόση είναι η συνδιακύμανση $\mathrm{Cov}(X,Y)$;

Έχουμε, κατά περίπτωση:

1. Και πάλι, πρέπει το ολοκλήρωμα της $f_{X,Y}(x,y)$ στο $\mathbb{R}\times\mathbb{R}$ να είναι μονάδα. Όμως,

$$\begin{aligned}
&\iint\limits_{\mathbb{R}\times\mathbb{R}} f_{X,Y}(x,y)\,dA \\
&= \int_0^1 \left(\int_0^{1-y} cx(1-y)\,dx\right) dy = \int_0^1 \left(c(1-y)\int_0^{1-y} x\,dx\right) dy \\
&= \int_0^1 c(1-y)\frac{(1-y)^2}{2}\,dy = c\int_0^1 \left(-\frac{(1-y)^4}{8}\right)' dy = \frac{c}{8}.
\end{aligned}$$



Σχήμα 2.11: Η πυκνότητα πιθανότητας του Παραδείγματος 2.10.

Άρα τελικά $c = 8$. Η πυκνότητα έχει σχεδιαστεί στο Σχήμα 2.11.

2. Όταν $x \notin [0,1]$, έχουμε $f_X(x) = 0$. Αν $x \in [0,1]$,

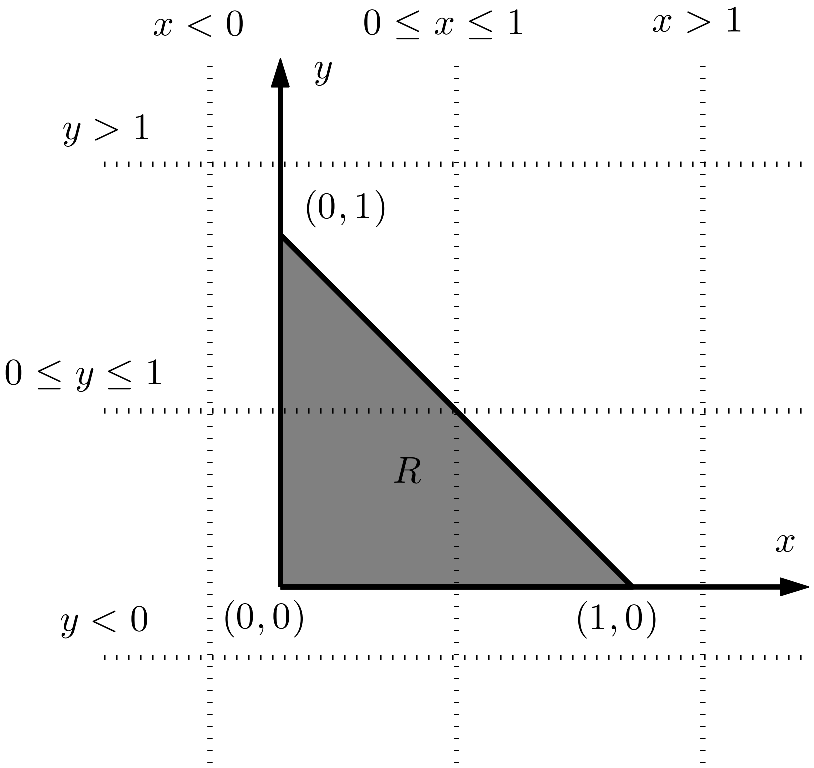
$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^{1-x} 8x(1-y)\,dy \\ &= 8x\int_0^{1-x}(1-y)\,dy = -8x\int_0^{1-x}\left(\frac{(1-y)^2}{2}\right)'\,dy = 4x(1-x^2), \end{aligned}$$

και συγκεντρωτικά

$$f_X(x) = \begin{cases} 4x(1-x^2), & x \in [0,1], \\ 0, & x \notin [0,1]. \end{cases}$$

Παρόμοια, αν $y \notin [0,1]$, τότε $f_Y(y) = 0$. Αν $y \in [0,1]$, τότε

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^{1-y} 8x(1-y)\,dx \\ &= 4(1-y)\int_0^{1-y}(x^2)'\,dx = 4(1-y)^3. \end{aligned}$$



Σχήμα 2.12: Παράδειγμα 2.10: Η από κοινού πυκνότητα είναι μη μηδενική στο χωρίο R. Για να υπολογίσουμε τις περιθώριες πυκνότητες $f_X(x)$, $f_Y(y)$, πρέπει να ολοκληρώσουμε την από κοινού κατά μήκος ευθειών καθέτων ή οριζοντίων (αντίστοιχα), έχοντας την μεταβλητή που δεν ολοκληρώνεται, x ή y αντίστοιχα, ως μια παράμετρο που θα επηρεάσει την τιμή του ολοκληρώματος μέσω της θέσης της ευθείας.

Τελικά,

$$f_Y(y) = \begin{cases} 4(1-y)^3, & y \in [0,1], \\ 0, & y \notin [0,1]. \end{cases}$$

Στο Σχήμα 2.12 έχουμε σχεδιάσει το χωρίο R στο οποίο η $f_{X,Y}(x, y)$ είναι μη μηδενική, καθώς και τις διάφορες ευθείες πάνω στις οποίες υπολογίσαμε το (απλό) ολοκλήρωμα της από κοινού πυκνότητας, προκειμένου να υπολογίσουμε τις περιθώριες πυκνότητες.

3. Οι X, Y δεν είναι ανεξάρτητες γιατί

$$P\left[\frac{1}{2} \le X \le 1\right] > 0, \qquad P\left[\frac{1}{2} \le Y \le 1\right] > 0,$$

ενώ

$$P\left[\frac{1}{2} \le X \le 1, \quad \frac{1}{2} \le Y \le 1\right] = \iint_{\frac{1}{2} \le x, y \le 1} 0\, dA = 0,$$

άρα η συνθήκη (2.8) δεν ικανοποιείται για $A = B = \left[\frac{1}{2}, 1\right]$.

4. Αφού οι X, Y, δεν είναι ανεξάρτητες, η συνδιακύμανση ενδεχομένως να μην είναι μηδενική, και πρέπει να υπολογιστεί. Έχουμε:

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^1 4x^2(1-x^2)\,dx = 4\int_0^1 \left(\frac{1}{3}x^3 - \frac{1}{5}x^5\right)' dx \\
&= 4\left[\frac{1}{3}x^3 - \frac{1}{5}x^5\right]_0^1 = \frac{8}{15}, \\
E(Y) &= \int_0^1 y f_Y(y)\,dy = 4\int_0^1 y(1-y)^3\,dy \\
&= 4\int_0^1 (y - 3y^2 + 3y^3 - y^4)\,dy = 4\int_0^1 \left(\frac{1}{2}y^2 - y^3 + \frac{3}{4}y^4 - \frac{1}{5}y^5\right)' dy \\
&= 4\left[\frac{1}{2}y^2 - y^3 + \frac{3}{4}y^4 - \frac{1}{5}y^5\right]_0^1 = \frac{1}{5}, \\
E(XY) &= \iint\limits_{\mathbb{R}\times\mathbb{R}} xy f_{X,Y}(x,y)\,dA = \iint\limits_{0\le x,y\le 1,\ x+y\le 1} 8x^2(1-y)y\,dA \\
&= \int_0^1 \left(\int_0^{1-x} 8x^2(1-y)y\,dy\right) dx = \int_0^1 8x^2 \left(\int_0^{1-x} (y-y^2)\,dy\right) dx \\
&= \int_0^1 8x^2 \left(\int_0^{1-x} \left(\frac{y^2}{2} - \frac{y^3}{3}\right)' dy\right) dx = \int_0^1 8x^2 \left[\frac{y^2}{2} - \frac{y^3}{3}\right]_0^{1-x} dx \\
&= \int_0^1 8x^2 \left(\frac{1}{2}(1-x)^2 - \frac{1}{3}(1-x)^3\right) dx \\
&= \int_0^1 \frac{4}{3}x^2\left(2x^3 - 3x^2 + 1\right) dx = \int_0^1 \left(\frac{8}{3}x^5 - 4x^4 + \frac{4}{3}x^2\right) dx \\
&= \int_0^1 \left(\frac{4}{9}x^6 - \frac{4}{5}x^5 + \frac{4}{9}x^3\right)' dx = \left[\frac{4}{9}x^6 - \frac{4}{5}x^5 + \frac{4}{9}x^3\right]_0^1 = \frac{4}{45},
\end{aligned}$$

και τελικά

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{4}{45} - \frac{8}{15} \times \frac{1}{5} = -\frac{4}{225}.$$

Παράδειγμα 2.11. *(Συνέλιξη)* Έστω δύο ανεξάρτητες συνεχείς Τ.Μ. X, Y με πυκνότητες $f_X(x)$, $f_Y(y)$ και με από κοινού πυκνότητα (δεδομένης της ανεξαρτησίας τους) την $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. Η πυκνότητα του αθροίσματός τους, έστω $Z = X + Y$, ισούται με

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(t) f_Y(z-t)\,dt.$$

(Η πιο πάνω έκφραση είναι η συνεχής μορφή της συνέλιξης δύο συναρτήσεων. Την συναντάμε πολύ συχνά στην επεξεργασία σήματος. Για παράδειγμα, ο ήχος που φτάνει στο αυτί σας από το στερεοφωνικό σας είναι η συνέλιξη της κυματομορφής που είναι αποθηκευμένη στο όποιο μέσο αποθήκευσης (CD, κτλ.) και μιας συνάρτησης που ενσωματώνει τις ρυθμίσεις του equalizer σας, τα χαρακτηριστικά του δωματίου που βρίσκεστε, τη συμπεριφορά των αυτιών σας, κ.ο.κ.)

Για να αποδείξουμε την άνω σχέση, θα υπολογίσουμε πρώτα την κατανομή της Z. Έχουμε:

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(X+Y \leq z) \\ &= \iint_{(x,y):x+y\leq z} f_{X,Y}(x,y)\,dA = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z-x} f_X(x)f_Y(y)\,dy\right)dx \\ &= \int_{-\infty}^{\infty} f_X(x)\left(\int_{-\infty}^{z-x} f_Y(y)\,dy\right)dx = \int_{-\infty}^{\infty} f_X(x)F_Y(z-x)\,dx. \end{aligned}$$

Παραγωγίζοντας, προκύπτει τελικά

$$\begin{aligned} f_Z(z) &= F_Z'(z) = \frac{d}{dz}\left(\int_{-\infty}^{\infty} f_X(x)F_Y(z-x)\,dx\right) \\ &= \int_{-\infty}^{\infty} f_X(x)F_Y'(z-x)\,dx = \int_{-\infty}^{\infty} f_X(t)f_Y(z-t)\,dt. \end{aligned}$$

Στην τρίτη ισότητα, αλλάξαμε την σειρά της ολοκλήρωσης και της παραγώγισης (στην συγκεκριμένη περίπτωση ικανοποιούνται οι συνθήκες που επιτρέπουν αυτή την εναλλαγή.) Στην τέταρτη ισότητα αλλάξαμε την μεταβλητή ολοκλήρωσης από x σε t.

Παράδειγμα 2.12. *(Ελάχιστο και μέγιστο)* Έστω οι ανεξάρτητες συνεχείς Τ.Μ. X, Y με κατανομές $F_X(x)$, $F_Y(y)$ και πυκνότητες $f_X(x)$, $f_Y(y)$. Έστω επίσης οι συνεχείς Τ.Μ. $W = \max\{X, Y\}$ και $V = \min\{X, Y\}$. Θα υπολογίσουμε τις κατανομές $F_W(w)$, $F_V(v)$ και τις πυκνότητες $f_W(w)$, $f_V(v)$, των W και V, συναρτήσει των $F_X(x)$, $F_Y(y)$ και $f_X(x)$, $f_Y(y)$.

Για να βρούμε την κατανομή της W, χρησιμοποιώντας το γεγονός ότι οι X, Y είναι ανεξάρτητες, έχουμε

$$\begin{aligned} F_W(w) &= P(W \leq w) = P\big(\{X \leq w\} \cap \{Y \leq w\}\big) \\ &= P(X \leq w)P(Y \leq w) = F_X(w)F_Y(w). \end{aligned}$$

Παραγωγίζοντας,

$$f_W(w) = f_X(w)F_Y(w) + F_X(w)f_Y(w).$$

Παρομοίως, για να βρούμε την συνάρτηση κατανομής της V έχουμε:

$$\begin{aligned} F_V(v) &= P(V \leq v) = 1 - P(V > v) = 1 - P\big(\{X > v\} \cap \{Y > v\}\big) \\ &= 1 - P(X > v)P(Y > v) = 1 - (1 - F_X(v))(1 - F_Y(v)). \end{aligned}$$

Παραγωγίζοντας:

$$f_V(v) = f_X(v)(1 - F_Y(v)) + f_Y(v)(1 - F_X(v)).$$

Όλα τα αποτελέσματα έχουν διαισθητική εξήγηση. Μπορείτε να τη βρείτε;

2.6 Πολλές Συνεχείς Τυχαίες Μεταβλητές

Όπως και στην περίπτωση των διακριτών Τ.Μ., η θεωρία που έχουμε αναπτύξει μέχρι τώρα επεκτείνεται και στην περίπτωση περισσότερων από δύο συνεχών Τ.Μ. Σε αυτή την παράγραφο παρουσιάζουμε συνοπτικά την επέκταση της θεωρίας, και μερικά χρήσιμα αποτελέσματα της. Παραλείπουμε όλες τις αποδείξεις, καθώς είναι ανάλογες με αυτές της περίπτωσης των δύο Τ.Μ.

Ορισμός 2.5. *Ένα πλήθος n Τ.Μ. $X_1, \ldots, X_n$ καλούνται <u>από κοινού συνεχείς</u> όταν υπάρχει μια συνάρτηση $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) : \mathbb{R}^n \to [0, \infty)$, την οποία καλούμε <u>από κοινού συνάρτηση πυκνότητας πιθανότητας</u>, ή <u>από κοινού πυκνότητα</u>, τέτοια ώστε για οποιοδήποτε χωρίο $R \subseteq \mathbb{R}^n$, να ισχύει ότι*

$$P[(X_1, \ldots, X_n) \in R] = \int \cdots \int_R f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n. \tag{2.10}$$

Λήμμα 2.4. (Ιδιότητες από κοινού πυκνότητας) *Έστω n από κοινού συνεχείς Τ.Μ. $X_1, \ldots, X_n$. Η από κοινού πυκνότητά τους $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ έχει τις εξής ιδιότητες:*

1. $\displaystyle\int \cdots \int_{\mathbb{R}^n} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = 1.$

2. *Η Τ.Μ. X_i είναι συνεχής και έχει <u>περιθώρια πυκνότητα</u> την*

$$f_{X_i}(x_i) = \int \cdots \int_{(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)\in\mathbb{R}^{n-1}} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n.$$

Λήμμα 2.5. (Ιδιότητες μέσης τιμής συνάρτησης πολλών Τ.Μ.) *Έστω n συνεχείς Τ.Μ. $X_1, \ldots, X_n$ με από κοινού πυκνότητα $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$. Έστω Τ.Μ. $Z = g(X_1, \ldots, X_n)$. Η μέση τιμή της ισούται με*

$$E(Z) = \int \cdots \int_{\mathbb{R}^n} g(x_1, \ldots, x_n) f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n.$$

Ορισμός 2.6. *Ένα πλήθος n από κοινού συνεχών Τ.Μ. $X_1, \ldots, X_n$ είναι <u>ανεξάρτητες</u> αν για οποιαδήποτε υποσύνολα $A_i \subset \mathbb{R}$, $i = 1, \ldots, n$ ισχύει*

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n).$$

Λήμμα 2.6. (Κριτήριο ανεξαρτησίας από κοινού συνεχών Τ.Μ.) *Έστω n από κοινού συνεχείς Τ.Μ. $X_1, \ldots, X_n$, με από κοινού πυκνότητα $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$, και περιθώριες πυκνότητες $f_{X_i}(x_i)$, $i = 1, \ldots, n$. Οι $X_1, \ldots, X_n$ είναι ανεξάρτητες αν*

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \ldots f_{X_n}(x_n) \quad \forall\, (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Λήμμα 2.7. (Ιδιότητες ανεξάρτητων συνεχών Τ.Μ.) *Έστω n συνεχείς ανεξάρτητες Τ.Μ. $X_1, \ldots, X_n$.*

1. *Οποιοδήποτε υποσύνολο από τις $X_1, \ldots, X_n$ είναι επίσης ανεξάρτητες Τ.Μ.*

2. *Έστω συναρτήσεις $g_i : S_{X_i} \to \mathbb{R}$. Θα ισχύει*

   $$E(g_1(X_1) \cdots g_n(X_n)) = E(g_1(X_1)) \cdots E(g_n(X_n)).$$

   Ειδική περίπτωση της άνω είναι η

   $$E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n).$$

3. $\mathrm{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{V}(X_i).$

Παράδειγμα 2.13. *(Μηνύματα)* Η διάρκεια αποστολής ενός SMS σε δευτερόλεπτα έχει ομοιόμορφη κατανομή στο διάστημα $[1, 3]$, και η διάρκεια αποστολής ενός MMS έχει εκθετική κατανομή με μέση τιμή τα 8 δευτερόλεπτα. Στέλνουμε, σε τρεις ανεξάρτητες αποστολές, 2 SMS και ένα MMS. Θα απαντήσουμε στα ακόλουθα:

1. Ποια είναι η μέση τιμή της συνολικής διάρκειας αποστολής;

2. Ποια είναι η πιθανότητα και τα 2 SMS και το MMS να έχουν όλα διάρκεια πάνω από 2 δευτερόλεπτα το καθένα;

3. Ποια είναι η πιθανότητα το MMS να έχει διάρκεια μεγαλύτερη από τη μέση τιμή της συνολικής διάρκειας των 2 SMS;

Για να απαντήσουμε τα άνω, ορίζουμε τις ακόλουθες Τ.Μ.:

X_1	η διάρκεια αποστολής του πρώτου SMS,
X_2	η διάρκεια αποστολής του δεύτερου SMS,
Y	η διάρκεια αποστολής του MMS, και
Z	η συνολική διάρκεια αποστολής των δύο SMS και του MMS,

οπότε $Z = X_1 + X_2 + Y$.

1. Για τη μέση τιμή της συνολικής διάρκειας αποστολής έχουμε:

   $$E(Z) = E(X_1 + X_2 + Y) = E(X_1) + E(X_2) + E(Y) = \frac{3+1}{2} + \frac{3+1}{2} + 8 = 12,$$

   αφού οι X_1, X_2 έχουν ομοιόμορφη κατανομή στο διάστημα $[1, 3]$ και η Y έχει εκθετική κατανομή με μέση τιμή 8.

2. Εφόσον οι τρεις αποστολές είναι ανεξάρτητες μεταξύ τους έχουμε ότι η πιθανότητα τα 2 SMS και το MMS να έχουν διάρκεια πάνω από 2 δευτερόλεπτα το καθένα είναι:

$$\begin{aligned} & P(X_1 > 2,\ X_2 > 2,\ Z > 2) \\ = \ & P(X_1 > 2)P(X_2 > 2)P(Z > 2) = \left[\int_2^3 \frac{1}{3-1}\,dx\right]^2 \times e^{-2/8} = \frac{e^{-1/4}}{4}. \end{aligned}$$

3. Η πιθανότητα το MMS να έχει διάρκεια μεγαλύτερη από τη μέση τιμή της συνολικής διάρκειας των 2 SMS είναι

$$P(Y > E(X_1 + X_2)) = P(Y > 2 + 2) = 1 - P(Y \leq 4) = e^{-4/8} \simeq 0.6065.$$

Παράδειγμα 2.14. *(Ελάχιστο n εκθετικών Τ.Μ.)* Έστω n Τ.Μ. $X_1, \ldots, X_n$, όλες εκθετικά κατανεμημένες, και με μέσες τιμές $\theta_1, \ldots, \theta_n$. Θα υπολογίσουμε την κατανομή του ελαχίστου τους, $Z = \min\{X_1, \ldots, X_n\}$, υποθέτοντας πως είναι ανεξάρτητες.

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = 1 - P(Z > z) = 1 - P\left(X_1 > z, \ldots, X_n > z\right) \\ &= 1 - P(X_1 > z) \cdots P(X_n > z) = 1 - \exp\left[-\frac{z}{\theta_1}\right] \cdots \exp\left[-\frac{z}{\theta_n}\right] \\ &= 1 - \exp\left[-z\left(\sum_{i=1}^n \frac{1}{\theta_i}\right)\right]. \end{aligned}$$

Άρα, το ελάχιστο Z είναι επίσης κατανεμημένο εκθετικά, με μέση τιμή $\left[\sum_{i=1}^n \theta_i^{-1}\right]^{-1}$.

Κεφάλαιο 3

Οριακά Θεωρήματα

3.1 Ανισότητα του Markov

Περιληπτικά, η ανισότητα του Markov λέει πως αν μια Τ.Μ. έχει «μικρή» μέση τιμή τότε δεν μπορεί να παίρνει μεγάλες τιμές με μεγάλη πιθανότητα. Για να κατανοήσουμε αυτό τον αναγκαίο περιορισμό, έστω πως ακούμε ένα βιολόγο να ισχυρίζεται πως «το μέσο βάρος του αφρικανικού χελιδονιού είναι 100 γραμμάρια, ενώ το 60% των αφρικανικών χελιδονιών έχει βάρος άνω των 200 γραμμαρίων». Προφανώς ο ισχυρισμός δεν μπορεί να ευσταθεί, γιατί ακόμα και αν το 40% των χελιδονιών δεν είχε καθόλου βάρος, και το υπόλοιπο 60% είχε το ελάχιστο επιτρεπτό, δηλαδή 200, τότε πάλι αυτό το υπόλοιπο 60% θα αρκούσε ώστε το μέσο βάρος να είναι τουλάχιστον $200 \times 0.6 = 120$ γραμμάρια, και θα είχαμε άτοπο. Η γενίκευση ακριβώς αυτού του συλλογισμού οδηγεί στην ανισότητα του Markov και την απόδειξή της.

Λήμμα 3.1. (Ανισότητα του Markov) *Έστω μια Τ.Μ. X που παίρνει πάντα τιμές $X \geq 0$. Τότε:*

$$P(X \geq c) \leq \frac{E(X)}{c}, \quad \forall\, c > 0.$$

Απόδειξη. Η ανισότητα ισχύει για όλα τα είδη Τ.Μ. Θα δούμε την απόδειξή της σε δύο ειδικές περιπτώσεις: όταν η X είναι διακριτή, και όταν είναι συνεχής.

Έστω καταρχήν πως η X είναι συνεχής με πυκνότητα $f(x)$. Ξεκινώντας από τον ορισμό της μέσης τιμής,

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^{\infty} x\, f(x)\, dx = \int_0^{c} x\, f(x)\, dx + \int_c^{\infty} x\, f(x)\, dx.$$

Στην δεύτερη ισότητα λάβαμε υπόψη ότι $f(x) = 0$ όταν $x < 0$. Εφόσον όλες οι τιμές x στο πρώτο από τα δύο ολοκληρώματα του δεξιού σκέλους είναι μεγαλύτερες ή ίσες του μηδενός, το ολοκλήρωμα είναι κι αυτό μεγαλύτερο ή ίσο του μηδενός, συνεπώς

$$E(X) \geq \int_c^{\infty} x\, f(x)\, dx \geq \int_c^{\infty} c\, f(x)\, dx = c \int_c^{\infty} x f(x)\, dx,$$

όπου, στην δεύτερη ανισότητα, χρησιμοποιήσαμε το γεγονός ότι στο ολοκλήρωμα που απομένει $x \geq c \Rightarrow xf(x) \geq cf(x)$. Παρατηρώντας, τέλος, πως από τον ορισμό της πυκνότητας το τελευταίο ολοκλήρωμα άνω ισούται με $P(X \geq c)$, έχουμε

$$E(X) \geq c\,P(X \geq c),$$

που είναι η ζητούμενη ανισότητα.

Έστω τώρα η περίπτωση που η X είναι διακριτή, με σύνολο τιμών $S \subset [0, \infty)$ και συνάρτηση πιθανότητας $f(x)$. Ξεκινώντας από τον ορισμό της μέσης τιμής,

$$E(X) = \sum_{x \in S} x\,f(x) = \sum_{x \in S : x < c} x\,f(x) + \sum_{x \in S : x \geq c} x\,f(x),$$

όπου χωρίσαμε το άθροισμα σε δύο μέρη, αυτό που αντιστοιχεί σε τιμές $x \in S$ μικρότερες του c, και τις τιμές $x \geq c$. Εφόσον όλες οι τιμές $x \in S$ της X είναι μεγαλύτερες ή ίσες του μηδενός, το πρώτο άθροισμα άνω είναι κι αυτό μεγαλύτερο ή ίσο του μηδενός, συνεπώς

$$E(X) \geq \sum_{x \in S : x \geq c} x\,f(x) \geq \sum_{x \in S : x \geq c} c\,f(x) = c \sum_{x \in S : x \geq c} f(x),$$

όπου, για την δεύτερη ανισότητα, χρησιμοποιήσαμε το γεγονός ότι στο άθροισμα που απομένει όλα τα x είναι μεγαλύτερα ή ίσα του c. Παρατηρώντας, τέλος, πως το τελευταίο άθροισμα άνω ισούται με $P(X \geq c)$ έχουμε

$$E(X) \geq c\,P(X \geq c),$$

που είναι η ζητούμενη ανισότητα. □

Παράδειγμα 3.1. *(Αφρικάνικο χελιδόνι.)* Ένας βιολόγος ισχυρίζεται το εξής: «Το μέσο βάρος του αφρικανικού χελιδονιού είναι 100 γραμμάρια, ενώ το 60% των αφρικανικών χελιδονιών έχει βάρος άνω των 200 γραμμαρίων». Βάσει της ανισότητας του Markov, ο ισχυρισμός του δεν ευσταθεί.

Πράγματι, αν X είναι το βάρος ενός τυχαίου χελιδονιού, θα πρέπει να έχουμε,

$$P(X \geq 200) \leq \frac{E(X)}{200} = \frac{100}{200} = \frac{1}{2},$$

που δεν συμφωνεί με τον ισχυρισμό του βιολόγου ότι αυτή η πιθανότητα είναι ίση με 60%.

Παράδειγμα 3.2. *(Γεωμετρική κατανομή)* Έστω μια Τ.Μ. $X \sim$ Γεωμ$(1/5)$, οπότε η X έχει μέση τιμή $E(X) = \frac{1}{1/5} = 5$. Από την ανισότητα του Markov έχουμε πως

$$P(X \geq 15) \leq 5/15 = \frac{1}{3}.$$

Στην πραγματικότητα όμως, η πιο πάνω πιθανότητα είναι (χρησιμοποιώντας γνωστό τύπο για την γεωμετρική κατανομή)

$$P(X \geq 15) = P(X > 14) = \Big(1 - \frac{1}{5}\Big)^{14} \simeq 0.044,$$

δηλαδή σημαντικά μικρότερη.

Παρατήρηση: Το φράγμα που δίνει η ανισότητα του Markov είναι εξαιρετικά χρήσιμο. Για παράδειγμα, όπως θα δούμε σύντομα, χρησιμεύει στην απόδειξη της ανισότητας του Chebychev και μέσω αυτής στην απόδειξη του Νόμου των Μεγάλων Αριθμών. Αλλά σε πολλές περιπτώσεις, όπως στο άνω παράδειγμα, είναι σχετικά ασθενές. Η ισχύς του έγκειται κυρίως στο γεγονός πως *δεν* χρησιμοποιεί κάποια πληροφορία για την κατανομή της X πέραν της μέσης τιμής της $E(X)$, και συνεπώς ισχύει για *οποιαδήποτε* Τ.Μ. X με μέση τιμή $E(X)$.

Παράδειγμα 3.3. *(Χρόνος εκτέλεσης αλγορίθμου)* Ένας αλγόριθμος έχει ως δεδομένα εισόδου έναν ακέραιο αριθμό Y και μία σειρά από n bits, $X_1, X_2, \ldots, X_n$. Ο χρόνος εκτέλεσης του αλγορίθμου είναι

$$T = Y + 2\sum_{i=1}^{n} X_i + \left(\sum_{i=1}^{n} X_i\right)^2 \qquad \text{δευτερόλεπτα.}$$

Αν υποθέσουμε ότι τα $Y, X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες Τ.Μ. όπου $X_i \sim \text{Bern}(1/4)$ και το Y παίρνει τις τιμές $0, 2, 5$ και 9 με πιθανότητα $1/4$ την κάθε μία:

1. Θα υπολογίσουμε τον μέσο χρόνο εκτέλεσης.
2. Θα δείξουμε ότι η πιθανότητα ο χρόνος εκτέλεσης να ξεπερνά τα n^2 δευτερόλεπτα είναι το πολύ
$$\frac{1}{16} + \frac{11}{16n} + \frac{4}{n^2}.$$
(Συνεπώς, για μεγάλα n αυτή η πιθανότητα είναι το πολύ $\simeq \frac{1}{16} = 6.25\%$.)

Πράγματι:

1. Καταρχάς, αν ορίσουμε
$$Z = \sum_{i=1}^{n} X_i,$$
τότε η $Z \sim \text{Διων}(n, 1/4)$ και συνεπώς
$$E(Z) = \frac{n}{4} \quad \text{και} \quad \mathrm{V}(Z) = n \times \frac{1}{4} \times \Big(1 - \frac{1}{4}\Big) = \frac{3n}{16}.$$

Επιπλέον,

$$E(Z^2) = \mathrm{V}(Z) + [E(Z)]^2 = \frac{3n}{16} + \frac{n^2}{16}.$$

Παρατηρώντας πως ο χρόνος εκτέλεσης T μπορεί να εκφραστεί ως $T = Y + 2Z + Z^2$, βρίσκουμε

$$\begin{aligned} E(T) &= E(Y + 2Z + Z^2) = E(Y) + 2E(Z) + E(Z^2) \\ &= \frac{1}{4} \times 0 + \frac{1}{4} \times 2 + \frac{1}{4} \times 5 + \frac{1}{4} \times 9 + 2 \times \frac{n}{4} + \frac{3n}{16} + \frac{n^2}{16} \\ &= 4 + \frac{11n}{16} + \frac{n^2}{16}. \end{aligned}$$

2. Από το προηγούμενο σκέλος και την ανισότητα του Markov έχουμε

$$P(T \geq n^2) \leq \frac{E(T)}{n^2} = \frac{1}{16} + \frac{11}{16n} + \frac{4}{n^2}.$$

Παράδειγμα 3.4. *(Ανισότητα του Chernoff)* Θα δείξουμε πως για οποιαδήποτε Τ.Μ. X και για κάθε $c, \lambda \in \mathbb{R}$, $\lambda > 0$, ισχύει

$$P(X \geq c) \leq \frac{E(e^{\lambda X})}{e^{\lambda c}}.$$

Η ανισότητα είναι μια από τις πιο απλές μορφές της *ανισότητας του Chernoff*. Παρατηρήστε ότι το φράγμα που δίνει μειώνεται εκθετικά με το c, και για αυτό το λόγο η ανισότητα αυτή είναι πολύ χρήσιμη στη Θεωρία Πιθανοτήτων.

Για να την αποδείξουμε, παρατηρήστε πως η Τ.Μ. $e^{\lambda X}$ προφανώς παίρνει πάντα τιμές μεγαλύτερες ή ίσες του μηδενός, και έχει μέση τιμή $E(e^{\lambda X})$. Άρα μπορούμε να εφαρμόσουμε την ανισότητα του Markov για αυτή την Τ.Μ., έτσι ώστε

$$P(X \geq c) = P(\lambda X \geq \lambda c) = P\Big(e^{\lambda X} \geq e^{\lambda c}\Big) \leq \frac{E(e^{\lambda X})}{e^{\lambda c}}$$

και έτσι προκύπτει η ζητούμενη ανισότητα.

Ας εφαρμόσουμε την ανισότητα του Chernoff στην περίπτωση που $X \sim \text{Εκθ}(1/\theta)$. Παρατηρήστε πως

$$E(e^{\lambda X}) = \int_{-\infty}^{\infty} e^{\lambda x}\, f(x)\, dx = \int_0^{\infty} e^{\lambda x}\, \frac{1}{\theta}\, e^{-x/\theta}\, dx = \frac{1}{\theta} \int_0^{\infty} e^{-x(1/\theta - \lambda)}\, dx.$$

Έστω πως $\lambda = \frac{1}{\theta}$. Τότε προφανώς $E(e^{\lambda X}) = \infty$. Έστω πως $\lambda \neq \frac{1}{\theta}$. Τότε έχουμε

$$\begin{aligned} E(e^{\lambda X}) &= \frac{1}{\theta} \int_0^{\infty} e^{-x(1/\theta - \lambda)}\, dx = \frac{1}{\theta(1/\theta - \lambda)} \Big[-e^{-x(1/\theta - \lambda)}\Big]_0^{\infty} \\ &= \frac{1}{\theta(1/\theta - \lambda)} \Big[1 - \lim_{x \to \infty} e^{-x(1/\theta - \lambda)}\Big]. \end{aligned}$$

Το όριο που εμφανίζεται είναι μηδέν για $\lambda < 1/\theta$, και ∞ για $\lambda > 1/\theta$. Άρα, συγκεντρωτικά,

$$E(e^{\lambda X}) = \begin{cases} \infty, & \lambda \geq 1/\theta, \\ \frac{1}{1-\lambda\theta}, & \lambda < \frac{1}{\theta}. \end{cases}$$

Με αντικατάσταση της μέσης τιμής που βρήκαμε στην ανισότητα του Chernoff, λαμβάνουμε τελικά

$$P(X \geq c) \leq \frac{e^{-\lambda c}}{1-\lambda\theta},$$

εφόσον βέβαια $\lambda < 1/\theta$.

Για παράδειγμα, για $\lambda = 1/2 < 1/\theta = 1$ και $c = 10$, το πιο πάνω φράγμα είναι περίπου ίσο με 0.0135, ενώ το φράγμα που μας δίνει η ανισότητα Markov είναι $1/c = 0.1$. Συνεπώς το φράγμα που αποδείξαμε εδώ είναι σημαντικά μικρότερο, και κατά συνέπεια πολύ ισχυρότερο από αυτό που θα μας έδινε η απευθείας χρήση της ανισότητας του Markov.

3.2 Ανισότητα του Chebychev

Το επόμενο αποτέλεσμα (το οποίο προκύπτει από μια σχετικά απλή εφαρμογή της ανισότητας του Markov) λέει πως αν μια Τ.Μ. έχει «μικρή» διασπορά, τότε δεν μπορεί να παίρνει τιμές μακρυά απ' την μέση τιμή της με μεγάλη πιθανότητα.

Λήμμα 3.2. (Ανισότητα του Chebychev) *Έστω μια Τ.Μ. X με πεπερασμένη μέση τιμή $E(X)$ και διασπορά $\mathrm{V}(X)$. Τότε:*

$$P(|X - E(X)| \geq c) \leq \frac{\mathrm{V}(X)}{c^2}, \quad \forall\, c > 0.$$

Απόδειξη. Έστω μια νέα Τ.Μ. $Y = (X - \mu)^2$ για την οποία, εξ ορισμού, πάντα έχουμε $Y \geq 0$. Από τον ορισμό της διασποράς, παρατηρούμε πως το Y έχει μέση τιμή $E(Y) = E[(X - \mu)^2] = \mathrm{V}(X)$. Εφαρμόζοντας την ανισότητα του Markov για την Τ.Μ. Y βρίσκουμε

$$P(|X - \mu| \geq c) = P\left((X - \mu)^2 \geq c^2\right) = P\left(Y \geq c^2\right) \leq \frac{E(Y)}{c^2} = \frac{\mathrm{V}(X)}{c^2}.$$

□

Παράδειγμα 3.5. Έστω μια Τ.Μ. X με κατανομή Υπερ$(500, 150, 15)$. Από τις γνωστές ιδιότητες της υπεργεωμετρικής κατανομής, η X έχει μέση τιμή

$$E(X) = \frac{150 \times 15}{500} = 4.5$$

και διασπορά

$$\mathrm{V}(X) = \frac{150 \times 15 \times (500 - 150) \times (500 - 15)}{500^2 \times (500 - 1)} \simeq 3.06162.$$

Έστω τώρα πως θέλουμε να υπολογίσουμε την πιθανότητα $P(|X - 4.5| \geq 3)$, δηλαδή την πιθανότητα το X να ισούται με 0 ή με 1 ή με κάποιον ακέραιο από το 8 ως και το 15. Εφόσον γνωρίζουμε την συνάρτηση πιθανότητας $f(x)$ της X, μπορούμε να υπολογίσουμε αυτή την πιθανότητα ως $f(0) + f(1) + f(8) + f(9) + \cdots + f(15)$. Αλλά λόγω της πολυπλοκότητας του τύπου της συνάρτησης πιθανότητας $f(x)$, και επιπλέον επειδή απαιτεί τον υπολογισμό παραγοντικών $k!$ για μεγάλα k – πράγμα το οποίο συχνά οδηγεί σε σημαντικά αριθμητικά σφάλματα – είναι πολύ πιο εύκολο να χρησιμοποιήσουμε την ανισότητα Chebychev για να βρούμε ένα *φράγμα* για την ζητούμενη πιθανότητα:

$$P(|X - 4.5| \geq 3) = P(|X - E(X)| \geq 3) \leq \frac{\mathrm{V}(X)}{3^2} \simeq 0.34018.$$

Με βοήθεια υπολογιστή μπορούμε να βρούμε πως

$$P(|X - 4.5| \geq 3) \simeq 0.0807.$$

Παράδειγμα 3.6. Έστω ένας εξυπηρετητής που δέχεται αιτήματα από χρήστες και αποκρίνεται, μετά από κάποια επεξεργασία του κάθε αιτήματος. Από εμπειρικές μετρήσεις γνωρίζουμε πως ο χρόνος απόκρισης X (σε δευτερόλεπτα) έχει μέση τιμή $E(X)$ και τυπική απόκλιση $\sigma = 6$ δευτερόλεπτα.

Θέλουμε να βρούμε ένα «φράγμα» θ τέτοιο ώστε να μπορούμε να εγγυηθούμε ότι, με πιθανότητα τουλάχιστον 99%, η χρονική απόκριση *δεν* θα απέχει από την μέση τιμή πάνω από $\pm\theta$. Δηλαδή, θέλουμε $P(|X - E(X)| < \theta) \geq 0.99$, ή, ισοδύναμα,

$$P(|X - E(X)| \geq \theta) \leq 0.01. \tag{3.1}$$

Αλλά από την ανισότητα του Chebychev έχουμε πως

$$P(|X - E(X)| \geq \theta) \leq \frac{\sigma^2}{\theta^2} = \frac{6^2}{\theta^2}.$$

Συνεπώς, για να ισχύει το ζητούμενο (3.1), αρκεί να διαλέξουμε το θ έτσι ώστε να έχουμε $36/\theta^2 \leq 0.01$, για το οποίο αρκεί να διαλέξουμε $\theta = 60$ δευτερόλεπτα. Με άλλα λόγια, μπορούμε να εγγυηθούμε πως, με πιθανότητα τουλάχιστον 99%, ο χρόνος απόκρισης δεν θα απέχει από τη μέση τιμή για πάνω από ένα λεπτό.

Παράδειγμα 3.7. *(Απόσταση Τ.Μ. από τη μέση τιμή της)* Για μια Τ.Μ. X με άγνωστη κατανομή και με μέση τιμή μ, κάποιος στατιστικολόγος ισχυρίζεται ότι η πιθανότητα η τιμή της X να απέχει από τη μέση τιμή της κατά παραπάνω από 3μ είναι μικρή. Υποστηρίζει αυτό το συμπέρασμα διότι γνωρίζει πως, στη συγκεκριμένη περίπτωση, η διασπορά της X, σ^2, είναι αρκετά μικρότερη από τη μέση τιμή στο τετράγωνο (δηλαδή από το μ^2). Έχει δίκιο στον ισχυρισμό του ή όχι;

Για να απαντήσουμε, παρατηρούμε πως, αφού έχουμε $\sigma^2 \ll \mu^2$ (δηλαδή η διασπορά της X είναι «πολύ μικρότερη» από τη μέση τιμή στο τετράγωνο), από την ανισότητα του Chebychev προκύπτει πως:

$$P(|X - \mu| \geq 3\mu) \leq \frac{\sigma^2}{(3\mu)^2} \ll \frac{\mu^2}{9\mu^2} = \frac{1}{9}.$$

Επομένως, ο στατιστικολόγος έχει δίκιο.

3.3 Ο Νόμος των Μεγάλων Αριθμών

Σε αυτό το σημείο έχουμε πλέον αναπτύξει αρκετά μαθηματικά εργαλεία ώστε να είμαστε σε θέση να διατυπώσουμε και να αποδείξουμε (σε μια απλή μορφή του) ίσως το πιο θεμελιώδες αποτέλεσμα της Θεωρίας Πιθανοτήτων, τον *Νόμο των Μεγάλων Αριθμών*. Χωρίς να μπούμε ακόμα σε μαθηματικές λεπτομέρειες, ο Νόμος των Μεγάλων Αριθμών λέει το εξής:

> Έστω ένα «μεγάλο» πλήθος N από ανεξάρτητες Τ.Μ. $X_1, \ldots, X_N$, οι οποίες έχουν όλες την ίδια κατανομή, και κατά συνέπεια την ίδια μέση τιμή $E(X_i) = \mu$. Τότε
> $$\frac{1}{N}\sum_{i=1}^{N} X_i \;\simeq\; \mu, \quad \text{με πιθανότητα} \simeq 1.$$

Η αυστηρή μαθηματική διατύπωση είναι η ακόλουθη:

Θεώρημα 3.1. (Νόμος των Μεγάλων Αριθμών (Ν.Μ.Α.)) *Έστω μια ακολουθία από ανεξάρτητες Τ.Μ. $X_1, X_2, \ldots$ που έχουν όλες την ίδια κατανομή με πεπερασμένη μέση τιμή $\mu = E(X_i)$ και πεπερασμένη διασπορά $\sigma^2 = \mathrm{V}(X_i)$. Έστω ο* <u>*εμπειρικός μέσος όρος*</u>

$$\bar{X}_N \triangleq \frac{1}{N}\sum_{i=1}^{N} X_i.$$

Για κάθε $\epsilon > 0$, έχουμε

$$P\left(|\bar{X}_N - \mu| < \epsilon\right) \to 1, \quad \textit{καθώς το } N \to \infty.$$

Απόδειξη. Η απόδειξη είναι μια απλή εφαρμογή της Ανισότητας του Chebychev. Καταρχάς, παρατηρούμε πως η μέση τιμή της Τ.Μ. $\bar{X}_N$ είναι

$$E(\bar{X}_N) = E\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{1}{N}\sum_{i=1}^{N} E(X_i) = \frac{1}{N} \times N \times \mu = \mu.$$

Παρομοίως, χρησιμοποιώντας το γεγονός ότι τα X_i είναι ανεξάρτητα, η διασπορά της Τ.Μ. $\bar{X}_N$ είναι

$$\begin{aligned} \mathrm{V}(\bar{X}_N) &= \mathrm{V}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{1}{N^2}\mathrm{V}\left(\sum_{i=1}^{N} X_i\right) = \frac{1}{N^2}\sum_{i=1}^{N} \mathrm{V}(X_i) \\ &= \frac{1}{N^2} \times N \times \sigma^2 = \frac{\sigma^2}{N}. \end{aligned}$$

Τέλος, δεδομένου του ότι $\epsilon > 0$, από την ανισότητα του Chebychev έχουμε

$$P\left(|\bar{X}_N - \mu| \geq \epsilon\right) = P\left(|\bar{X}_N - E(\bar{X}_N)| \geq \epsilon\right) \leq \frac{\mathrm{V}(\bar{X}_N)}{\epsilon^2} = \frac{\sigma^2}{N\epsilon^2}.$$

Άρα,

$$P\left(|\bar{X}_N - \mu| < \epsilon\right) = 1 - P\left(|\bar{X}_N - \mu| \geq \epsilon\right) \geq 1 - \frac{\sigma^2}{N\epsilon^2},$$

το οποίο προφανώς τείνει στο 1, καθώς το $N \to \infty$. □

Παρατηρήσεις

1. Το αποτέλεσμα ισχύει όποιο και αν είναι το είδος των X_i, αρκεί να υπάρχουν η μέση τιμή και η διασπορά τους.

2. Στην παρούσα του μορφή, ο Ν.Μ.Α. μας λέει ότι η ακολουθία των Τ.Μ. (και όχι αριθμών!) $\bar{X}_N$ τείνει στην σταθερά μ καθώς το $N \to \infty$ με την έννοια ότι η πιθανότητα να απέχει έστω και απειροελάχιστα το $\bar{X}_N$ απ' τη μ τείνει στο μηδέν. Δηλαδή, για οποιαδήποτε απόκλιση $\epsilon > 0$, οσοδήποτε μικρή, έχουμε

$$P\left(|\bar{X}_N - \mu| \geq \epsilon\right) \to 0, \quad \text{καθώς το } N \to \infty.$$

   Στη γλώσσα των πιθανοτήτων, το είδος αυτής της σύγκλισης ονομάζεται «σύγκλιση κατά πιθανότητα».

3. Η συγκεκριμένη διατύπωση του Ν.Μ.Α. καλείται *«Ασθενής Νόμος των Μεγάλων Αριθμών»*. Υπάρχει και ένας *«Ισχυρός Νόμος των Μεγάλων Αριθμών»*, σύμφωνα με τον οποίο

$$P\left(\lim_{N\to\infty} \bar{X}_N = \mu\right) = 1.$$

   Στην γλώσσα των πιθανοτήτων, η ακολουθία $\bar{X}_N$ συγκλίνει στο μ «σχεδόν σίγουρα». Οι δύο νόμοι μοιάζουν αρκετά στην διατύπωση, όμως ο Ισχυρός Ν.Μ.Α. αποδεικνύεται πολύ πιο δύσκολα, δεν απαιτεί την ύπαρξη της διασποράς σ^2, και είναι πολύ πιο χρήσιμος σε θεωρητικά προβλήματα πιθανοτήτων.

4. Όπως φαίνεται και από τα επόμενα παραδείγματα, ο Ν.Μ.Α. μας εξασφαλίζει ότι ισχύουν πολλές ιδιότητες των Πιθανοτήτων που διαισθητικά είναι αναμενόμενες, αλλά δεν τις είχαμε αποδείξει μέχρι τώρα. Κατά μία έννοια, ο Ν.Μ.Α. είναι «απόδειξη» ότι η Θεωρία Πιθανοτήτων «δουλεύει»!

Παράδειγμα 3.8. *(Τι θα πει «πιθανότητα»;)* Έστω πως μας ενδιαφέρει η πιθανότητα $p = P(A)$ του να συμβεί ένα συγκεκριμένο ενδεχόμενο A. Λόγου χάρη, το A θα μπορούσε να αντιστοιχεί στα «η θεραπεία του ασθενή ήταν επιτυχής» ή «το δίκτυο παρουσίασε σφάλμα» ή «η εκτέλεση του αλγορίθμου ολοκληρώθηκε κανονικά» κλπ. Ένας τρόπος για να δώσουμε ένα διαισθητικό νόημα στην πιθανότητα p είναι να φανταστούμε πως επαναλαμβάνεται το ίδιο ακριβώς πείραμα πάρα πολλές, ανεξάρτητες φορές, έτσι ώστε, τουλάχιστον στο διαισθητικό επίπεδο: «η πιθανότητα $p = P(A)$ είναι το ποσοστό των φορών, μακροπρόθεσμα, που συμβαίνει το A». Ο Ν.Μ.Α. είναι ακριβώς το αποτέλεσμα εκείνο που τεκμηριώνει αυτόν το συλλογισμό μαθηματικά, γιατί δείχνει ότι ο τρόπος που ορίσαμε την πιθανότητα είναι συμβατός με την άνω ερμηνεία της πιθανότητας $P(A)$.

Συγκεκριμένα, το πιο πάνω φανταστικό σενάριο μπορεί να περιγραφεί ως εξής. Έστω πως έχουμε ένα μεγάλο πλήθος, N, ανεξάρτητων επαναλήψεων του ίδιο πειράματος, και ας ορίσουμε, για την κάθε επανάληψη i, μια Τ.Μ. X_i που να ισούται με 1 αν το A συνέβη τη φορά i, αλλιώς να ισούται με 0. Τότε,

$$\text{«ποσοστό των φορών που συνέβη το } A\text{»} = \frac{1}{N}\sum_{i=1}^{N} X_i,$$

και ο Ν.Μ.Α. μας λέει πως, με πιθανότητα που τείνει στη μονάδα, αυτό θα απέχει από την μέση τιμή των X_i λιγότερο από ϵ. Εφόσον τα X_i είναι Bernoulli Τ.Μ., η μέση τιμή $E(X_i)$ είναι απλά η πιθανότητα $p = P(A)$ του να έχουμε $X_i = 1$, δηλαδή του να συμβεί το A. Με άλλα λόγια:

> *Η πιθανότητα p του να συμβεί ένα οποιοδήποτε ενδεχόμενο A, ισούται με τη συχνότητα με την οποία θα συμβεί, μακροπρόθεσμα, το A, σε πολλές ανεξάρτητες επαναλήψεις του ίδιου πειράματος.*

Παράδειγμα 3.9. *(Γιατί γίνονται δημοσκοπήσεις;)* Έστω πως σε έναν πληθυσμό M ατόμων οι $M/4$ είναι ψηφοφόροι κάποιου κόμματος, δηλαδή το κόμμα αυτό στις εκλογές θα πάρει ποσοστό 25%. Μια δημοσκόπηση επιλέγει στην τύχη ένα «μεγάλο» πλήθος ατόμων N (με επανατοποθέτηση) και τα ρωτά αν θα ψηφίσουν το κόμμα αυτό. (Σε τυπικές δημοσκοπήσεις έχουμε N μεταξύ 1000 και 2000 ερωτηθέντων σε έναν πληθυσμό M μεταξύ 500 χιλιάδων και 8 εκατομμυρίων ατόμων.)

Έστω X_i μια Τ.Μ. που ισούται με 1 αν το άτομο i που ρωτάται είναι ψηφοφόρος του κόμματος, και 0 αν όχι, οπότε οι X_i είναι ανεξάρτητες Τ.Μ. με κατανομή Bern(0.25). Τότε,

$$\text{«ποσοστό ερωτηθέντων που είναι ψηφοφόροι του κόμματος»} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

και ο Ν.Μ.Α. μας λέει πως, με πιθανότητα που τείνει στη μονάδα, αυτό θα απέχει από την μέση τιμή των X_i, δηλαδή το 0.25, λιγότερο από ϵ. Άρα, από ένα αρκετά μεγάλο (τυχαίο) δείγμα, μπορούμε να εκτιμήσουμε το πραγματικό ποσοστό ψήφων που αυτό το κόμμα θα πάρει στις εκλογές.

Παράδειγμα 3.10. *(Δειγματοληψία)* Έστω πως θέλουμε να υπολογίσουμε το μέσο ύψος $\bar{y}$ (σε εκατοστά) που έχουν M κορίτσια σε ένα σχολείο. Αν y_k είναι το ύψος του κοριτσιού $k = 1, 2, \ldots, M$, θέλουμε να εκτιμήσουμε το

$$\text{«μέσο ύψος»} = \bar{y} = \frac{1}{M}\sum_{i=1}^{M} y_k.$$

Αντί να εξετάσουμε ολόκληρο τον πληθυσμό, επιλέγουμε τυχαία, με επανατοποθέτηση, N μέλη του πληθυσμού, όπου το μέγεθος N του δείγματός μας είναι μεν σχετικά μεγάλο, αλλά είναι σημαντικά μικρότερου του συνολικού μεγέθους M του πληθυσμού (όπως και στο προηγούμενο παράδειγμα).

Έστω $X_j =$«ύψος του κοριτσιού j» που επιλέξαμε. Τότε τα X_j είναι ανεξάρτητα, και το κάθε X_j έχει σύνολο τιμών $S = \{y_1, y_2, \ldots, y_M\}$ και συνάρτηση πιθανότητας

$$p(y_k) = P(X_j = y_k) = P(\text{«επιλέξαμε το κορίτσι } k\text{»}) = \frac{1}{M}.$$

(Για απλότητα, υποθέσαμε ότι όλα τα ύψη είναι διαφορετικά, ας πούμε γιατί τα έχουμε μετρήσει με άπειρη ακρίβεια.) Άρα τα X_j έχουν μέση τιμή

$$E(X_j) = \sum_{k=1}^{M} y_k \times p(y_k) = \sum_{k=1}^{M} y_k \times \frac{1}{M} = \frac{1}{M}\sum_{k=1}^{M} y_k = \bar{y}.$$

Αν εκτιμήσουμε λοιπόν το $\bar{y}$ ως τον μέσο όρο απ΄ τα ύψη των κοριτσιών που επιλέξαμε, ο Ν.Μ.Α. είναι αυτός που μας διαβεβαιώνει πως, αν το μέγεθος N του δείγματός μας είναι αρκετά μεγάλο, τότε η εκτίμησή μας θα είναι ακριβής με πιθανότητα κοντά στο 100%:

$$\text{«μέσο ύψος δειγμάτων»} = \frac{1}{N}\sum_{i=1}^{N} X_i \simeq E(X_i) = \bar{y}, \quad \text{με πιθανότητα} \simeq 1.$$

3.4 Το Κεντρικό Οριακό Θεώρημα

Έστω X_i, $i = 1, 2, \ldots$ μια ακολουθία από ανεξάρτητες Τ.Μ. με την ίδια κατανομή, πεπερασμένη μέση τιμή $E(X) = \mu$ και πεπερασμένη διασπορά $\mathrm{V}(X) = \sigma^2$.

Το πρώτο θεμελιώδες αποτέλεσμα των πιθανοτήτων, ο Ν.Μ.Α., μας εξασφαλίζει ότι η πιθανότητα ο εμπειρικός μέσος όρος

$$\bar{X}_N \triangleq \frac{1}{N}\sum_{i=1}^{N} X_i$$

να παρουσιάζει απόκλιση από το μ μεγαλύτερη από ϵ τείνει στο 0, δηλαδή

$$\lim_{N\to\infty} P(|\bar{X}_N - \mu| > \epsilon) = 0.$$

Δύο βασικά και μάλλον προφανή ερωτήματα που γεννιούνται από τον Ν.Μ.Α. είναι τα εξής:

1. Πόσο μεγάλο πρέπει να είναι το πλήθος N των δειγμάτων, ώστε να έχουμε κάποια σχετική βεβαιότητα πως ο εμπειρικός μέσος όρος $\bar{X}_N$ θα είναι αρκετά κοντά στη μέση τιμή μ;
2. Επιπλέον, δεδομένου του πλήθους N, πόσο μικρή είναι η πιθανότητα το $\bar{X}_N$ να απέχει κατά πολύ από το μ;

Αν εξετάσουμε προσεκτικά την απόδειξη του Ν.Μ.Α. που δώσαμε θα δούμε πως, πέρα απ᾽ το ασυμπτωτικό αποτέλεσμα του θεωρήματος (το οποίο ισχύει μόνο καθώς το $N \to \infty$), περιέχει και κάποιες πρώτες απαντήσεις στα πιο πάνω ερωτήματα. Συγκεκριμένα, δείξαμε πως η μέση τιμή και η διασπορά του $\bar{X}_N$ είναι, αντίστοιχα,

$$E(\bar{X}_N) = \mu \quad \text{και} \quad \mathrm{V}(\bar{X}_N) = \frac{\sigma^2}{N}, \quad \text{για } \textit{κάθε } N.$$

Επιπλέον βρήκαμε ένα ακριβές ποσοτικό φράγμα για την πιθανότητα ο εμπειρικός μέσος όρος $\bar{X}_N$ να απέχει από την μέση τιμή μ κατά τουλάχιστον ϵ:

$$P\left(|\bar{X}_N - \mu| \ge \epsilon\right) \le \frac{\sigma^2}{N\epsilon^2}.$$

Αν και μαθηματικά σωστό, το πιο πάνω φράγμα στις περισσότερες περιπτώσεις δεν είναι αρκετά ακριβές ώστε να είναι χρήσιμο στην πράξη. Με άλλα λόγια, η ζητούμενη πιθανότητα στη σχέση (3.4) είναι συνήθως σημαντικά μικρότερη από το φράγμα $\sigma^2/(N\epsilon^2)$.

Το δεύτερο θεμελιώδες αποτέλεσμα των πιθανοτήτων, το *Κεντρικό Οριακό Θεώρημα* καλύπτει αυτή την αδυναμία του Ν.Μ.Α., δίνοντας μια ακριβή προσέγγιση της πιθανότητας $P\left(|\bar{X}_N - \mu| \ge \epsilon\right)$.

Συγκεκριμένα, το Κεντρικό Οριακό Θεώρημα λέει πως, κάτω από ορισμένες συνθήκες, η κατανομή του εμπειρικού μέσου όρου $\bar{X}_N$ μπορεί να προσεγγιστεί με μεγάλη ακρίβεια από την *κανονική κατανομή* με παραμέτρους μ και $\frac{\sigma^2}{N}$, όπου σ^2 είναι η διασπορά των X_i.

Ισοδύναμα, επειδή για τις κανονικές Τ.Μ. γνωρίζουμε ότι αν η $X \sim N(\mu, \sigma^2)$ τότε η $Y = \frac{X-\mu}{\sigma} \sim N(0,1)$, θα ισχύει επίσης ότι το άθροισμα

$$\frac{\bar{X}_N - \mu}{\sqrt{\sigma^2/N}} = \frac{\sqrt{N}}{\sigma}\left[\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) - \mu\right] = \frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}(X_i - \mu)$$

ακολουθεί κατά προσέγγιση την τυπική κανονική κατανομή $N(0,1)$. Συνεπώς, οι πιθανότητες ενδεχόμενων που αφορούν τον $\bar{X}_N$ μπορούν εύκολα να υπολογιστούν.

Θεώρημα 3.2. (Κεντρικό Οριακό Θεώρημα (Κ.Ο.Θ.)) *Έστω μια ακολουθία από ανεξάρτητες Τ.Μ. $X_1, X_2, \ldots$ που έχουν όλες την ίδια κατανομή με πεπερασμένη μέση τιμή $\mu = E(X_i)$ και πεπερασμένη διασπορά $\sigma^2 = \mathrm{V}(X_i)$. Θεωρούμε το* <u>*κανονικοποιημένο άθροισμα*</u>

$$\bar{S}_N \triangleq \frac{\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) - \mu}{\sigma\sqrt{1/N}} = \frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}(X_i - \mu) = \frac{\left(\sum_{i=1}^{N} X_i\right) - N\mu}{\sigma\sqrt{N}}.$$

Έστω επίσης και Z μιά τυχαία μεταβλητή με κατανομή την τυπική κανονική $N(0,1)$. Τότε για κάθε διάστημα $I \subset \mathbb{R}$ ισχύει

$$P\left(\bar{S}_N \in I\right) \to P(Z \in I)$$

καθώς $N \to \infty$.

Ειδικές περιπτώσεις. Για $a, b \in \mathbb{R}$ με $a < b$ ισχύει

$$P\left(a \le \bar{S}_N \le b\right) \;\to\; \Phi(b) - \Phi(a) \tag{3.2}$$

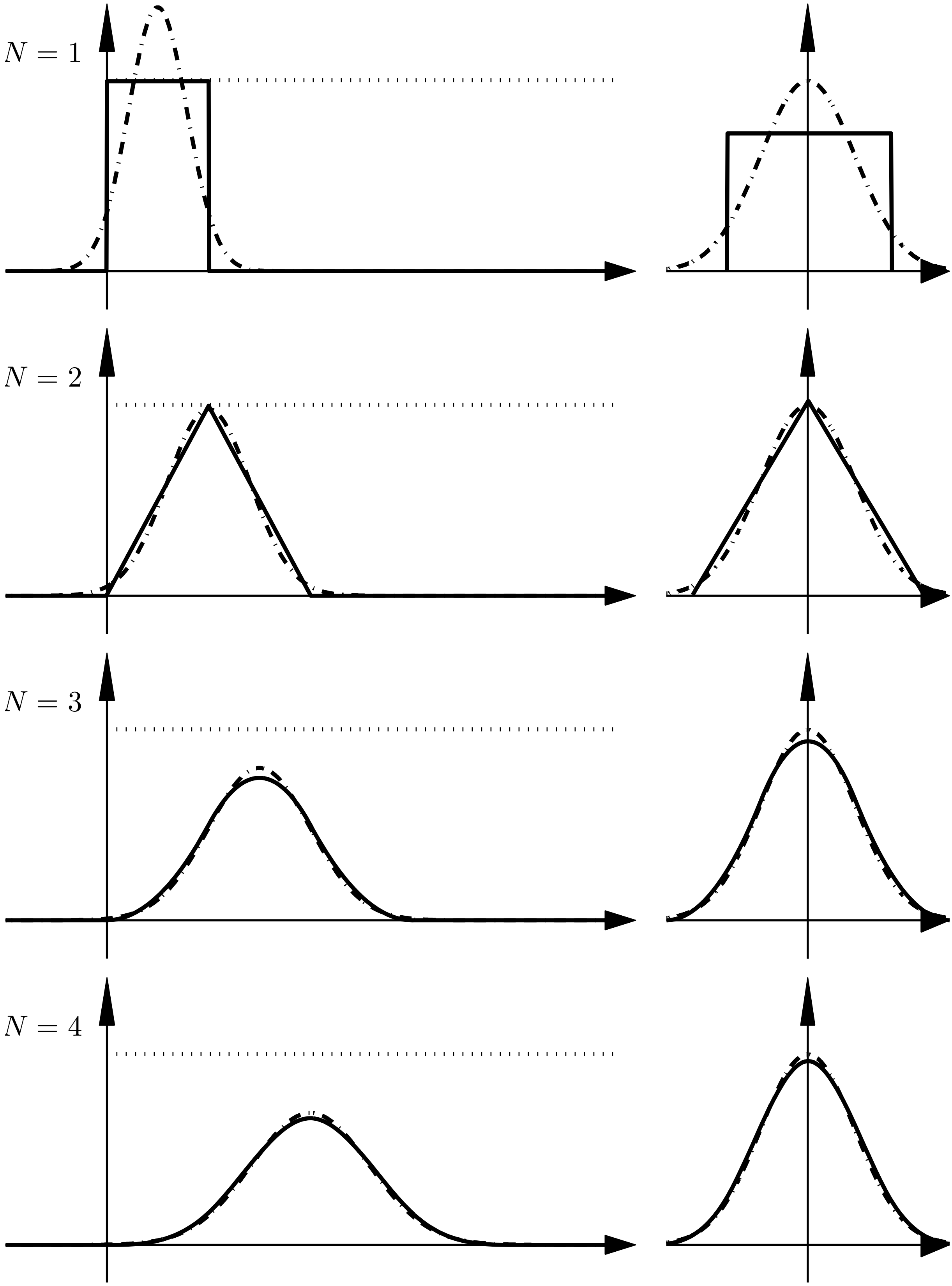
$$P\left(\bar{S}_N \le b\right) \;\to\; \Phi(b) \tag{3.3}$$

$$P\left(a \le \bar{S}_N\right) \;\to\; 1 - \Phi(a) \tag{3.4}$$

Φ είναι η συνάρτηση κατανομής της τυπικής κανονικής $N(0,1)$.

Παρατηρήσεις

1. Η απόδειξη του Κ.Ο.Θ. είναι πολύ τεχνική, και ξεφεύγει κατά πολύ από τους στόχους του παρόντος μαθήματος. Συνεπώς την παραλείπουμε.



Σχήμα 3.1: Σύγκλιση της κατανομής των αθροισμάτων $\sum_{i=1}^{N} X_i$ (αριστερά) και των κανονικοποιημένων αθροισμάτων $\bar{S}_N = \frac{\left(\sum_{i=1}^{N} X_i\right) - N\mu}{\sigma\sqrt{N}}$ (δεξιά) στην κανονική κατανομή, στην περίπτωση $X_i \sim U[0,1]$.

2. Παρατηρήστε ότι έχουμε ένα είδος σύγκλισης που δεν έχουμε ξαναδεί. Συγκεκριμένα, η κατανομή μιας ακολουθίας Τ.Μ. $\bar{S}_N$ συγκλίνει σε μια άλλη κατανομή, στην συγκεκριμένη περίπτωση την τυπική κανονική. Αυτού του είδους η σύγκλιση καλείται «σύγκλιση κατά κατανομή».

3. Ένα παράδειγμα της σύγκλισης που προβλέπει το Κ.Ο.Θ. φαίνεται στο Σχήμα 3.1. Αριστερά φαίνονται με συνεχή γραμμή οι πυκνότητες, όπως έχουν υπολογιστεί με χρήση Η/Υ, των αθροισμάτων $\sum_{i=1}^{N} X_i$, για $N = 1, 2, 3, 4$, και όταν τα X_i είναι κατανεμημένα ομοιόμορφα στο $[0, 1]$. Φαίνονται επίσης, με διακεκομμένη γραμμή, οι κανονικές κατανομές στις οποίες συγκλίνουν οι πυκνότητες. Στο δεξί μέρος έχουμε σχεδιάσει τις πυκνότητες των κανονικοποιημένων αθροισμάτων $\bar{S}_N = \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^{N}(X_i - \mu)$ και τις αντίστοιχές τους κανονικές. Παρατηρήστε ότι η σύγκλιση είναι πολύ καλή, ακόμα και για την μικρή τιμή $N = 4$.

4. Η πρακτική χρησιμότητα του Κ.Ο.Θ. είναι ότι μας επιτρέπει να υπολογίζουμε προσεγγιστικά πιθανότητες που αφορούν μέσους όρους και/ή αθροίσματα ενός πλήθους N ανεξάρτητων Τ.Μ. $X_1, X_2, \ldots, X_N$, όλων με την ίδια κατανομή. Η μεθοδολογία είναι πάντα η ίδια:

   (α′) Υπολογίζουμε (αν δεν γνωρίζουμε ήδη) τη μέση τιμή και τη διασπορά των X_i.

   (β′) Εκφράζουμε το ενδεχόμενο του οποίου την πιθανότητα θέλουμε να υπολογίσουμε ως μια ανισότητα ή συνδυασμό ανισοτήτων που αφορούν το κανονικοποιημένο άθροισμα $\bar{S}_N$, σε οποιαδήποτε από τις μορφές του μας βολεύει.

   (γ′) Επικαλούμαστε το Κ.Ο.Θ. για να προσεγγίσουμε την πιθανότητα του ενδεχόμενου που βρήκαμε στο προηγούμενο βήμα μέσω της κανονικής κατανομής.

5. Το σφάλμα που κάνουμε στην προσέγγιση (3.2) εξαρτάται από πολλούς παράγοντες:

   (α′) Την απόσταση μεταξύ των a, b.

   (β′) Την σχέση ανάμεσα στα a, b, μ.

   (γ′) Την κατανομή των X_i.

   (δ′) Την τιμή του N.

   Γενικώς, όταν το N είναι αρκετά μεγάλο, πχ μεγαλύτερο του 50, και όταν τα a, b απέχουν αρκετά μεταξύ τους και είναι αρκετά κοντά στο μ, ώστε η πιθανότητα που βρίσκουμε να μην είναι πολύ μικρή, π.χ., μεγαλύτερη του 0.01, τότε η προσέγγιση είναι ικανοποιητική.

6. Δείτε τα ακόλουθα παραδείγματα για εφαρμογές αυτής της μεθοδολογίας.

Παράδειγμα 3.11. Έστω πως X_i είναι η θερμοκρασία ενός επεξεργαστή την ημέρα i, όπου θεωρούμε πως τα X_i είναι ανεξάρτητες Τ.Μ. με κατανομή $X_i \sim U[10, 50]$. Ποια η πιθανότητα η μέση θερμοκρασία του επεξεργαστή σε μία περίοδο 3 μηνών (δηλαδή 90 ημερών) να ξεπερνά τους 31 βαθμούς;

Για να εφαρμόσουμε το Κ.Ο.Θ. αρχικά υπολογίζουμε την μέση τιμή και την διασπορά των X_i οι οποίες, από τις αντίστοιχες ιδιότητες της ομοιόμορφης κατανομής είναι

$$\mu = E(X_i) = \frac{10+50}{2} = 30, \quad \sigma^2 = \mathrm{V}(X_i) = \frac{(50-10)^2}{12} = \frac{400}{3}$$

αντίστοιχα. Άρα, η ζητούμενη πιθανότητα μπορεί να εκφραστεί ως προς το κανονικοποιημένο άθροισμα του Κ.Ο.Θ. ως εξής:

$$P\Big(\frac{1}{N}\sum_{i=1}^{N} X_i > 31\Big) = P\Big(\frac{1}{N}\sum_{i=1}^{N}(X_i-\mu) > 31-30\Big) = P\Big(\frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}(X_i-\mu) > \frac{\sqrt{N}}{\sigma}\Big),$$

όπου η τιμή $\sqrt{N}/\sigma$ είναι $\sqrt{90}/\sqrt{400/3} \simeq 0.8216$. Άρα, από το Κ.Ο.Θ., η ζητούμενη πιθανότητα μπορεί να προσεγγιστεί ως

$$P\Big(\frac{1}{N}\sum_{i=1}^{N} X_i > 31\Big) \simeq P(Z > 0.82) = 1 - P(Z \leq 0.82) = 1 - \Phi(0.82),$$

όπου η Z έχει κατανομή $N(0, 1)$. Αντικαθιστώντας την τιμή $\Phi(0.82) \simeq 0.7939$, τελικά βρίσκουμε πως η ζητούμενη πιθανότητα είναι $\simeq 1 - 0.7939 = 0.2061$.

Παράδειγμα 3.12. Από εμπειρικές μετρήσεις παρατηρούμε πως η διάρκεια εκτέλεσης ενός αλγορίθμου διαρκεί κατά μέσο όρο 17.5 δευτερόλεπτα, με τυπική απόκλιση ± 4 δευτερόλεπτα. Ποια είναι η πιθανότητα, σε 400 διαδοχικές, ανεξάρτητες χρήσεις του αλγορίθμου, ο συνολικός χρόνος εκτέλεσης να μην ξεπερνά τις δύο ώρες;

Έστω X_i ο χρόνος εκτέλεσης του αλγορίθμου την φορά i για $i = 1, \ldots, N = 400$. Από τα δεδομένα του προβλήματος έχουμε πως τα X_i είναι ανεξάρτητες Τ.Μ. με μέση τιμή $\mu = 17.5$ δευτερόλεπτα και τυπική απόκλιση $\sigma = 4$ δευτερόλεπτα. Συνεπώς, η ζητούμενη πιθανότητα μπορεί να εκφραστεί χρησιμοποιώντας το κανονικοποιημένο άθροισμα του Κ.Ο.Θ. ως εξής:

$$\begin{aligned}
P\Big(\sum_{i=1}^{N} X_i \leq 2 \times 60 \times 60\Big) &= P\Big(\sum_{i=1}^{N}(X_i - \mu) \leq 7200 - 400 \times 17.5\Big) \\
&= P\Big(\frac{1}{\sigma\sqrt{N}}\sum_{i=1}^{N}(X_i - \mu) \leq \frac{200}{4 \times \sqrt{400}}\Big) \\
&\simeq P(Z \leq 2.5),
\end{aligned}$$

όπου $Z \sim N(0,1)$ και πάλι εφαρμόσαμε το Κ.Ο.Θ. Άρα η ζητούμενη πιθανότητα είναι

$$P\Big(\sum_{i=1}^{N} X_i \leq 7200\Big) \simeq P(Z \leq 2.5) = \Phi(2.5) \simeq 0.9938,$$

όπου αντικαταστήσαμε την τιμή $\Phi(2.5) \simeq 0.9938$.

Παρατήρηση: Η άνω μεθοδολογία μπορεί να εφαρμοστεί για οποιοδήποτε είδος κατανομής των X_i. Στην ειδική περίπτωση που οι X_i είναι ακέραιες, μπορούμε να κάνουμε μια μικρή τροποποίηση, που βελτιώνει κάπως την ακρίβεια της προσέγγισης, ιδιαιτέρως όταν η ζητούμενη πιθανότητα είναι πολύ μικρή. Συγκεκριμένα, έστω $m < M$ και έστω πως πρέπει να υπολογίσουμε την πιθανότητα

$$P(m \leq \sum_{i=1}^{N} X_i \leq M).$$

Τότε, προσεγγίσουμε αντί αυτής την πιθανότητα

$$P\left(m - \frac{1}{2} \leq \sum_{i=1}^{N} X_i \leq M + \frac{1}{2}\right),$$

Παρατηρήστε ότι αφού οι X_i είναι ακέραιες, θα είναι και το άθροισμά τους ακέραιο, και άρα οι δύο πιθανότητες είναι ακριβώς ίσες. Οι προσεγγίσεις τους όμως θα διαφέρουν μεταξύ τους, και η προσέγγιση που θα βασίζεται στο δεύτερο ενδεχόμενο θα είναι καλύτερη. Διαισθητικά, είναι λογικό να αντιστοιχίσουμε στο ενδεχόμενο $\sum_{i=1}^{N} X_i = m$ το διάστημα $(m-\frac{1}{2}, m+\frac{1}{2})$, στο ενδεχόμενο $\sum_{i=1}^{N} X_i = m+1$ το διάστημα $(m+\frac{1}{2}, m+\frac{3}{2})$, κ.ο.κ., και τελικά στο ενδεχόμενο $\sum_{i=1}^{N} X_i = M$ το διάστημα $(M-\frac{1}{2}, M+\frac{1}{2})$. Παρατηρήστε μάλιστα πως όταν $M = m$ η πρώτη προσέγγιση θα δώσει πιθανότητα ακριβώς 0. Δείτε τα ακόλουθα παραδείγματα.

Παράδειγμα 3.13. Για $N = 150$ ρίψεις ενός δίκαιου νομίσματος ($p = 1/2$), ποια είναι η πιθανότητα οι συνολικές κορώνες να είναι μεταξύ των 90 και των 105;

Για να απαντήσουμε, έστω $X_1, X_2, \ldots, X_N$ Τ.Μ. όπου $X_i = 1$ αν η i-οστή ρίψη είναι κορώνα, και $X_i = 0$ αλλιώς. Η μέση τιμή των X_i είναι $\frac{1}{2}$ και η διασπορά $p(1-p) = \frac{1}{4}$, κατά τα γνωστά για την κατανομή Bernoulli. Έστω επίσης $Y = \sum_{i=1}^{N} X_i$ το άθροισμα από κορώνες. Παρατηρούμε πως

$$\begin{aligned}
&P(90 \le Y \le 105)\\
&= P\left(90 - \frac{1}{2} \le Y \le 105 + \frac{1}{2}\right)\\
&= P\left(90 - \frac{1}{2} - 150 \times 0.5 \le Y - 150 \times 0.5 \le 105 + \frac{1}{2} - 150 \times 0.5\right)\\
&= P\left(\frac{90 - \frac{1}{2} - 150 \times 0.5}{\sqrt{150\frac{1}{4}}} \le \frac{Y - 150 \times 0.5}{\sqrt{150\frac{1}{4}}} \le \frac{105 + \frac{1}{2} - 150 \times 0.5}{\sqrt{150\frac{1}{4}}}\right)\\
&\simeq \Phi(4.98) - \Phi(2.37) \simeq 1 - 0.9911 = 0.0089,
\end{aligned}$$

Αν δεν είχαμε τροποποιήσει αρχικά το εύρος τιμών, το αποτέλεσμα θα ήταν

$$\Phi(4.90) - \Phi(2.45) \simeq 1 - 0.9928 = 0.0072,$$

Η ακριβής τιμή, όπως προκύπτει με χρήση υπολογιστή, είναι 0.0088. Η τροποποιημένη μεθοδολογία δίνει καλύτερο αποτέλεσμα, αλλά πάλι υπάρχει μια μικρή διαφορά, που οφείλεται στο ότι η ζητούμενη πιθανότητα είναι πολύ μικρή.

Παράδειγμα 3.14. Έστω πως η διάρκεια κάθε τηλεφωνικής κλήσης σε ένα δίκτυο έχει εκθετική κατανομή με μέση διάρκεια 85 δευτερόλεπτα, και έστω πως οι διάρκειες των διαδοχικών κλήσεων είναι ανεξάρτητες. Μια κλήση θεωρείται «σύντομη» αν διαρκέσει λιγότερο από ένα λεπτό. Ποια είναι η πιθανότητα, το πλήθος των σύντομων κλήσεων ανάμεσα στις 250 που έγιναν σε μία μέρα να είναι μικρότερο 120;

Έστω μια Τ.Μ. $Y \sim$ Εκθ$(1/85)$. Η πιθανότητα μια κλήση να είναι σύντομη είναι

$$p = P(Y \le 60) = 1 - e^{-60/85} \simeq 0.5063.$$

Έστω, τώρα, ανεξάρτητες Bern(p) Τ.Μ. X_i, όπου η κάθε $X_i = 1$ αν η κλήση i είναι σύντομη. Η μέση τιμή των X_i είναι $E(X_i) = p$ και η διασπορά τους $\sigma^2 = p(1-p)$. Η ζητούμενη πιθανότητα είναι

$$\begin{aligned}
&P\Big(\sum_{i=1}^{250} X_i < 120\Big)\\
&= P\Big(\sum_{i=1}^{250} X_i \le 119.5\Big) = P\Big(\sum_{i=1}^{250}(X_i - p) \le 119.5 - 250 \times p\Big)\\
&= P\left(\frac{1}{\sqrt{p(1-p) \times 250}} \sum_{i=1}^{250}(X_i - \mu) \le \frac{119.5 - 250 \times p}{\sqrt{p(1-p) \times 250}}\right)\\
&\simeq P(Z \le -0.8959) = \Phi(-0.8959) \simeq 0.1851,
\end{aligned}$$

Αν δεν είχαμε τροποποιήσει αρχικά το εύρος τιμών, το αποτέλεσμα θα ήταν (επιβεβαιώστε το)

$$\Phi(-0.8326) \simeq 0.2025,$$

Η ακριβής τιμή, όπως προκύπτει με χρήση υπολογιστή, είναι επίσης $\simeq 0.1851$! Για την ακρίβεια, το σφάλμα του πρώτου μας υπολογισμού είναι περίπου ίσο με 10^{-6}.

Παρατήρηση: Παρατηρήστε ότι στα άνω δύο παραδείγματα είχαμε την προσέγγιση ενός αθροίσματος όμοια κατανεμημένων και ανεξάρτητων Τ.Μ. Bernoulli μέσω της κανονικής κατανομής. Παρατηρήστε όμως ότι ένα άθροισμα όμοια κατανεμημένων και ανεξάρτητων Τ.Μ. Bernoulli ακολουθεί την διωνυμική κατανομή. Άρα, μπορούμε να προσεγγίσουμε την διωνυμική κατανομή μέσω της κανονικής! Αυστηρά, έχουμε το ακόλουθο λήμμα, που δίνουμε χωρίς απόδειξη:

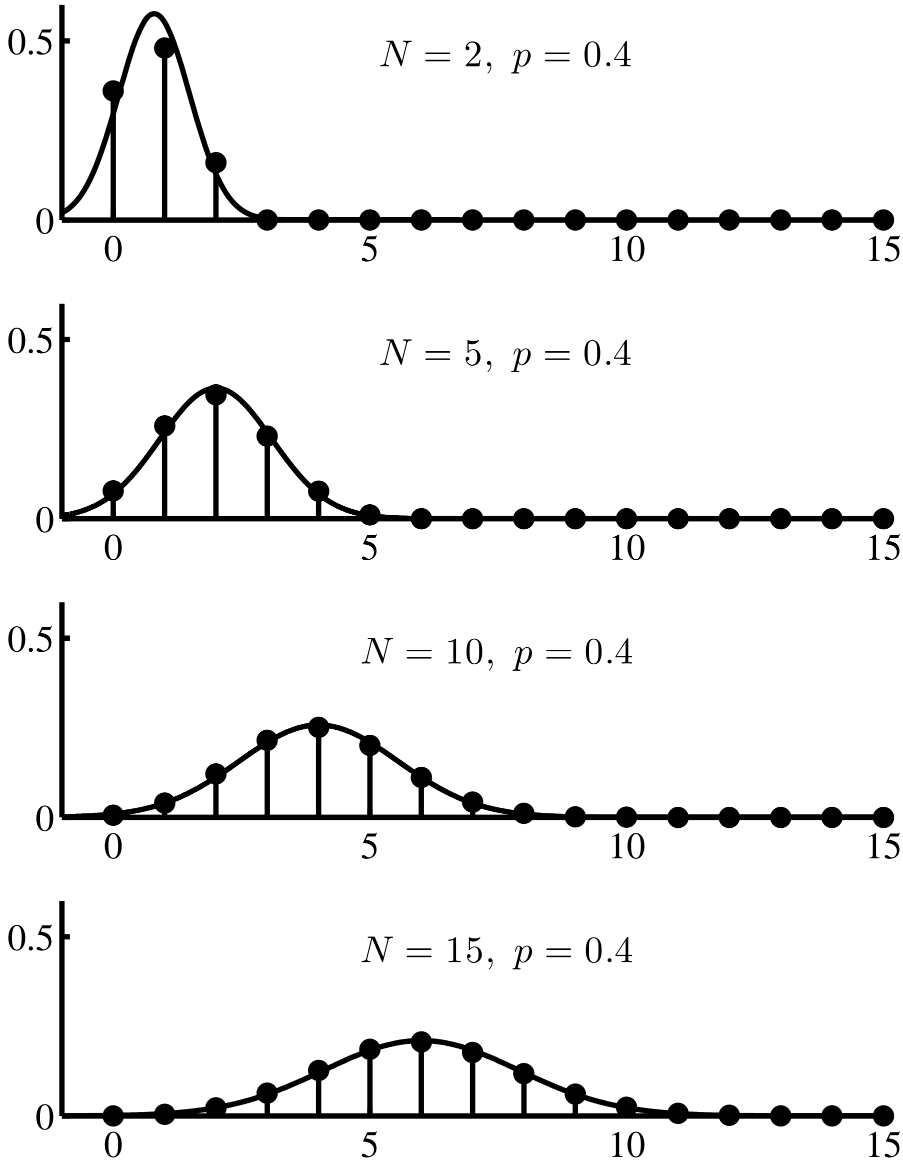
Λήμμα 3.3. (Προσέγγιση της Διωνυμικής Κατανομής από την Κανονική) *Έστω ακολουθία από διωνυμικές Τ.Μ. X_N με παραμέτρους N, p. Τότε, καθώς $N \to \infty$:*

$$\begin{aligned}
P\left(a \leq \frac{X_N - Np}{\sqrt{Np(1-p)}} \leq b\right) &\to \Phi(b) - \Phi(a), \\
P\left(\frac{X_N - Np}{\sqrt{Np(1-p)}} \leq b\right) &\to \Phi(b), \\
P\left(\frac{X_N - Np}{\sqrt{Np(1-p)}} \geq a\right) &\to 1 - \Phi(a).
\end{aligned}$$

Παρατηρήσεις

1. Στο Σχήμα 3.2 έχουμε σχεδιάσει τις συναρτήσεις πιθανότητας μιάς ακολουθίας από διωνυμικές Τ.Μ. με $p = 0.4$ και $N = 2, 5, 10, 15$. Παρατηρήστε πως, όπως προβλέπει το Κ.Ο.Θ., καθώς μεγαλώνει το N, οι μάζες τείνουν στην κανονική κατανομή

2. Πρακτικά, η άνω προσέγγιση είναι ικανοποιητική όταν το N είναι αρκετά μεγάλο, π.χ., μεγαλύτερο του 50, και ταυτόχρονα το Np είναι αρκετά μεγαλύτερο της μονάδας, π.χ. μεγαλύτερο του 10.



Σχήμα 3.2: Σύγκλιση της διωνυμικής κατανομής στην κανονική, για $p = 0.4$ και για αυξανόμενες τιμές του N ($N = 2, 5, 10, 15$.)