



ΚΕΦΑΛΑΙΟ 5

ΚΑΤΑΝΟΜΕΣ ΠΙΘΑΝΟΤΗΤΑΣ

Σκοπός

Οι δειγματικοί χώροι, ανάλογα με τη φύση και τον τρόπο έκφρασης των ενδεχομένων τους, κατατάσσονται σε ποσοτικούς και ποιοτικούς. Προφανώς ο υπολογισμός πιθανοτήτων ενδεχομένων είναι ευκολότερος στην περίπτωση ποσοτικών δεδομένων, των οποίων τα στοιχεία εκφράζονται ως αριθμοί. Επειδή, όμως, σε πολλές πρακτικές εφαρμογές οι δειγματικοί χώροι είναι ποιοτικοί, είναι επιθυμητό να οριστεί μία συνάρτηση δια της οποίας να αντιστοιχείται ένας πραγματικός αριθμός σε κάθε στοιχείο του ποιοτικού δειγματικού χώρου. Η συνάρτηση αυτή ονομάζεται **τυχαία μεταβλητή**. Σκοπός του κεφαλαίου αυτού είναι να εισαγάγει την έννοια της τυχαίας μεταβλητής και της κατανομής πιθανότητάς της.

Έννοιες Κλειδιά

- | | | |
|--|--|------------------------------------|
| • Αθροιστική συνάρτηση κατανομής πιθανότητας | • Διάμεσος | • Συνάρτηση πυκνότητας πιθανότητας |
| • Αναμενόμενη τιμή | • Επικρατούσα τιμή | • Συνδιακύμανση |
| • Ανεξάρτητες τυχαίες μεταβλητές | • Κατανομή πιθανότητας | • Συνεχής τυχαία μεταβλητή |
| • Βαθμωτή συνάρτηση | • Μαθηματική ελπίδα | • Συντελεστής συσχέτισης |
| • Διακριτή ή απαριθμητή τυχαία μεταβλητή | • Μέση απόκλιση τετραγώνου ή τυπική απόκλιση | • Τυχαία μεταβλητή |
| • Διακύμανση | • Μέση τιμή | |
| | • Ροπή | |
| | • Συνάρτηση πιθανότητας | |

5.1 ΕΙΣΑΓΩΓΗ

Η εφαρμογή των κανόνων υπολογισμού πιθανοτήτων ενδεχόμενων ή συνδυασμού ενδεχομένων, οι οποίοι παρουσιάστηκαν στο προηγούμενο κεφάλαιο, είναι ευκολότερη στην περίπτωση ποσοτικών δεδομένων. Επειδή, όμως, σε πολλές πρακτικές εφαρμογές, τα δεδομένα είναι ποιοτικά, θα ήταν χρήσιμο να οριστεί μία συνάρτηση που να αντιστοιχεί έναν πραγματικό αριθμό σε κάθε στοιχείο ενός ποιοτικού δειγματικού χώρου. Η συνάρτηση αυτή ονομάζεται **τυχαία μεταβλητή**.

Σκοπός του πέμπτου κεφαλαίου είναι να εισαγάγει την έννοια της τυχαίας μεταβλητής και της κατανομής πιθανότητάς της. Στην **πρώτη ενότητα** ορίζεται η έννοια της τυχαίας μεταβλητής, γίνεται διάκριση μεταξύ διακριτών και συνεχών τυχαίων μεταβλητών και στη συνέχεια ορίζεται, αφενός, η συνάρτηση πιθανότητας με το διάγραμμά της για τις διακριτές μεταβλητές και, αφετέρου, η συνάρτηση πυκνότητας πιθανότητας με το διάγραμμά της για τις συνεχείς τυχαίες μεταβλητές. Στη **δεύτερη ενότητα** ορίζεται η αθροιστική συνάρτηση κατανομής για διακριτές και συνεχείς τυχαίες μεταβλητές. Στην **τρίτη**, στην **τέταρτη**, στην **πέμπτη και στην έκτη ενότητα** παρουσιάζονται αντίστοιχα τα μέτρα κεντρικής τάσης, τα μέτρα διασποράς και οι ροπές τόσο για διακριτές όσο και για συνεχείς τυχαίες μεταβλητές. Τέλος, στην **έβδομη ενότητα** γίνεται μια σύντομη αναφορά στις διμεταβλητές κατανομές πιθανότητας.

5.2 ΤΥΧΑΙΕΣ ΜΕΤΑΒΛΗΤΕΣ

Ο τρόπος με τον οποίο εκφράζονται τα στοιχειώδη ενδεχόμενα ενός τυχαίου πειράματος εξαρτάται από τη φύση των απλών αυτών ενδεχομένων. Με βάση, λοιπόν, τη φύση και κατά συνέπεια τον τρόπο έκφρασης των ενδεχομένων τους, οι δειγματικοί χώροι που συνδέονται με τα συγκεκριμένα τυχαία πειράματα κατατάσσονται σε δύο κατηγορίες:

α. Ποσοτικοί ή αριθμητικοί

β. Ποιοτικοί ή περιγραφικοί

Παράδειγμα ποσοτικού δειγματικού χώρου είναι ο χώρος που συνδέεται με το τυχαίο πείραμα της ρίψης ενός ζαριού και έχει τα στοιχειώδη ενδεχόμενα 1, 2, 3, 4, 5, 6.

Παράδειγμα ποιοτικού δειγματικού χώρου είναι ο χώρος που συνδέεται με το τυχαίο πείραμα της ρίψης ενός νομίσματος και έχει τα στοιχειώδη ενδεχόμενα «Πρόσωπο» (Π) και «Γράμματα» (Γ).

Είναι προφανές ότι ο υπολογισμός των πιθανοτήτων ενδεχομένων ή συνδυασμού ενδεχομένων είναι ευκολότερος στην περίπτωση των ποσοτικών δειγματικών χώρων όπου χρησιμοποιούνται αριθμοί. Είναι επιθυμητό, λοιπόν, για την ενιαία αντιμετώπιση των δειγματικών χώρων, να οριστεί ένας κανόνας, δηλαδή μία συνάρτηση, που να αντιστοιχεί έναν πραγματικό αριθμό σε κάθε στοιχείο του δειγματικού χώρου S . Η συνάρτηση αυτή ονομάζεται **τυχαία μεταβλητή** (*random variable*) και ορίζεται ως εξής:

Έστω ο δειγματικός χώρος S . Ορίζουμε ως τυχαία μεταβλητή μια πραγματική συνάρτηση με πεδίο ορισμού τον S και τιμές στην ευθεία των πραγματικών αριθμών $R: X: S \rightarrow R$ τέτοια ώστε το σύνολο $X^{-1}(I)$ να είναι ένα ενδεχόμενο για κάθε διάστημα $I \subseteq R$.

Σημειώνεται ότι οι τυχαίες μεταβλητές συμβολίζονται συνήθως με τα κεφαλαία γράμματα X, Y, Z, W και οι τιμές τους με τα αντίστοιχα μικρά γράμματα x, y, z, w . Θα πρέπει να υπογραμμιστεί ότι σε κάθε στοιχειώδες ενδεχόμενο, s , του δειγματικού χώρου S ($s \in S$) αντιστοιχεί ακριβώς μία τιμή της τυχαίας μεταβλητής, η τιμή $x = X(s)$. Η τιμή αυτή μπορεί να αντιστοιχεί σε περισσότερα από ένα στοιχειώδη ενδεχόμενα του S . Τέλος, για κάθε τιμή x της X θα πρέπει να καθορίζεται η πιθανότητα $P(X = x)$, η οποία θα πρέπει να ικανοποιεί τα βασικά αξιώματα του Kolmogorov.

Οι τυχαίες μεταβλητές διακρίνονται σε δύο βασικές κατηγορίες:

α. Διακριτές

β. Συνεχείς

Καθεμιά από τις κατηγορίες αυτές εξετάζεται αναλυτικά στη συνέχεια.

5.2.1 Διακριτές τυχαίες μεταβλητές

Έστω η τυχαία μεταβλητή $X: S \rightarrow R$. Αν το πεδίο τιμών $X(S)$ της X , δηλαδή το σύνολο των τιμών της, είναι πεπερασμένο ή απείρως αριθμήσιμο, τότε η μεταβλητή αυτή

καλείται **διακριτή τυχαία μεταβλητή** (*discrete random variable*). Αυτό πρακτικά σημαίνει ότι στην περίπτωση μιας διακριτής τυχαίας μεταβλητής X μπορούμε να αριθμήσουμε τις δυνατές τιμές της και, αν μεν το $X(S)$ είναι πεπερασμένο, η αρίθμηση περατώνεται, ενώ, αν το $X(S)$ είναι απείρως αριθμήσιμο, η αρίθμηση συνεχίζεται επ' άπειρο.

Έστω X μια διακριτή τυχαία μεταβλητή και $X(S)$ το πεδίο τιμών της. Μία συνάρτηση: $p_x \equiv P_X(x) \equiv P(x) \equiv P(X=x)$ με πεδίο ορισμού τις τιμές της X και πεδίο τιμών τις πιθανότητες των τιμών αυτών, η οποία ικανοποιεί τις συνθήκες:

$$P_X(x) \geq 0, \quad \forall x \in X(S) \subseteq R$$

$$\sum_{x \in X(S)} P_X(x) = 1$$

καλείται **συνάρτηση πιθανότητας** (*probability function*) της τυχαίας μεταβλητής X .

Το σύνολο των ζευγών $\{(x, P_X(x)), x \in X(S)\}$ καλείται **κατανομή πιθανότητας** (*probability distribution*) της διακριτής τυχαίας μεταβλητής X . Η γραφική απεικόνιση της συνάρτησης πιθανότητας ονομάζεται **διάγραμμα πιθανότητας** (*probability plot*) της τυχαίας μεταβλητής X .

Το παράδειγμα που ακολουθεί θα βοηθήσει στην καλύτερη κατανόηση των εννοιών που παρουσιάστηκαν παραπάνω.

Παράδειγμα 5.1

Έστω μία οικογένεια με 3 παιδιά. Αν συμβολίσουμε με X τον αριθμό των αγοριών, να καθοριστεί η κατανομή πιθανότητας της τυχαίας μεταβλητής X και να παρασταθεί γραφικά.

Λύση

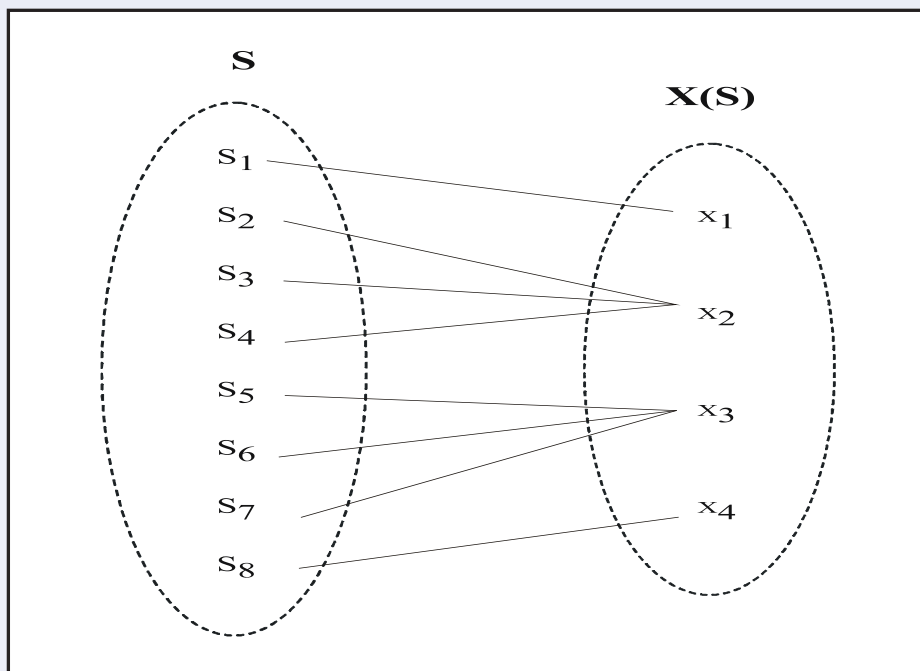
Έστω A το ενδεχόμενο γέννησης αγοριού και K το ενδεχόμενο γέννησης κοριτσιού. Ο αριθμός των αγοριών στην οικογένεια εξαρτάται από το αποτέλεσμα καθενός από τους τρεις τοκετούς. Είναι, λοιπόν, δυνατό να εμφανιστούν κανένα, ένα, δύο ή τρία αγόρια. Κατά συνέπεια, η μεταβλητή X , η οποία εκφράζει τον αριθμό των αγοριών είναι μία διακριτή τυχαία μεταβλητή η οποία μπορεί να πάρει τις τιμές $x_1 = 0, x_2 = 1, x_3 = 2, x_4 = 3$.

Πίνακας 5.1 Οικογένεια με τρία παιδιά: Τα στοιχειώδη ενδεχόμενα του δειγματικού χώρου και οι αντίστοιχες πιθανότητές τους

Στοιχειώδη ενδεχόμενα δειγμ. χώρου S	Αριθμός αγοριών X	Πιθανότητα
$s_1 : KKK$	0	1/8
$s_2 : AKK$	1	1/8
$s_3 : KAK$	1	1/8
$s_4 : KKA$	1	1/8
$s_5 : AAK$	2	1/8
$s_6 : AKA$	2	1/8
$s_7 : KAA$	2	1/8
$s_8 : AAA$	3	1/8

Ο Πίνακας 5.1 περιέχει όλα τα στοιχειώδη ενδεχόμενα του δειγματικού χώρου (δηλαδή όλα τα δυνατά αποτελέσματα των τριών τοκετών) καθώς και τις αντίστοιχες πιθανότητές τους. Όπως φαίνεται, σε κάθε στοιχειώδες ενδεχόμενο s_i του δειγματικού χώρου αντιστοιχεί ακριβώς μια τιμή της τυχαίας μεταβλητής X και επιπλέον η ίδια τιμή μπορεί να αντιστοιχεί σε διαφορετικά στοιχειώδη ενδεχόμενα του S .

Σχήμα 5.1 Οικογένεια με τρία παιδιά: Αντιστοιχία των στοιχειωδών ενδεχομένων του δειγματικού χώρου και των τιμών της τυχαίας μεταβλητής



Η αντιστοιχία μεταξύ των στοιχειωδών ενδεχομένων και των τιμών της τυχαίας μεταβλητής X απεικονίζεται γραφικά στο Σχήμα 5.1.

Αφού προσδιοριστούν οι δυνατές τιμές της τυχαίας μεταβλητής X , στη συνέχεια μπορεί να οριστεί η συνάρτηση πιθανότητάς της $p_x \equiv P_x(x) \equiv P(x) \equiv P(X = x)$. Με τη συνάρτηση αυτή, σε κάθε στοιχείο x του πεδίου τιμών $X(S)$ της μεταβλητής X αντιστοιχούμε έναν πραγματικό αριθμό $P(X = x)$, ο οποίος εκφράζει την πιθανότητα εμφάνισης της τιμής αυτής.

Συγκεκριμένα,

$$p_0 \equiv P_x(0) \equiv P(0) \equiv P(X = 0) = \frac{1}{8}$$

$$p_1 \equiv P_X(1) \equiv P(1) \equiv P(X=1) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$

$$p_2 \equiv P_X(2) \equiv P(2) \equiv P(X=2) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$

$$p_3 \equiv P_X(3) \equiv P(3) \equiv P(X=3) = \frac{1}{8}$$

Από τα παραπάνω προκύπτει:

$$\begin{aligned} \sum_{i=1}^4 P(X=x_i) &= P(X=x_1) + P(X=x_2) + P(X=x_3) + P(X=x_4) \\ &= \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1 \end{aligned}$$

Είναι προφανές ότι για τη συνάρτηση πιθανότητας ικανοποιούνται οι βασικές συνθήκες που απορρέουν από τα αξιώματα του Kolmogorov. Συγκεκριμένα, όλες οι τιμές της είναι μεγαλύτερες του μηδενός και, επιπλέον, το άθροισμα όλων των τιμών της είναι ίσο με τη μονάδα.

Οι δυνατές τιμές της τυχαίας μεταβλητής X και οι αντίστοιχες πιθανότητές τους συνοψίζονται στον Πίνακα 5.2. Ο Πίνακας αυτός ή, με άλλα λόγια, το σύνολο των ζευγών

$$\{(0,1/8), (1,3/8), (2,3/8), (3,1/8)\}$$

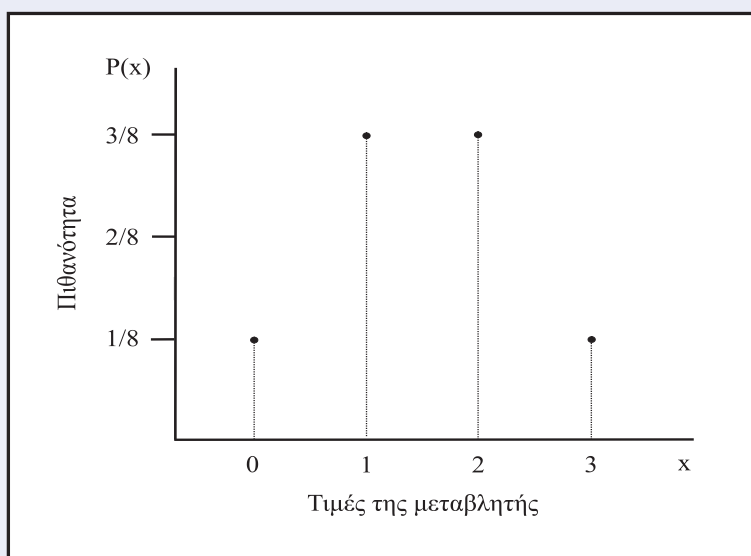
ονομάζεται κατανομή πιθανότητας της τυχαίας μεταβλητής X .

Πίνακας 5.2 Οικογένεια με τρία παιδιά. Κατανομή πιθανότητας του αριθμού των αγοριών

Δυνατές τιμές της τυχαίας μεταβλητής X x_i	Πιθανότητες εμφάνισής τους $P(X = x_i)$
0	1/8
1	3/8
2	3/8
3	1/8

Η γραφική απεικόνιση του συνόλου αυτών των ζευγών, δηλαδή το αντίστοιχο διάγραμμα πιθανότητας, παρουσιάζεται στο Σχήμα 5.2.

Σχήμα 5.2 Οικογένεια με τρία παιδιά: Διάγραμμα πιθανότητας



5.2.2 Συνεχείς τυχαίες μεταβλητές

Η τυχαία μεταβλητή X , με τιμές x , ονομάζεται **συνεχής τυχαία μεταβλητή** (*continuous random variable*) αν:

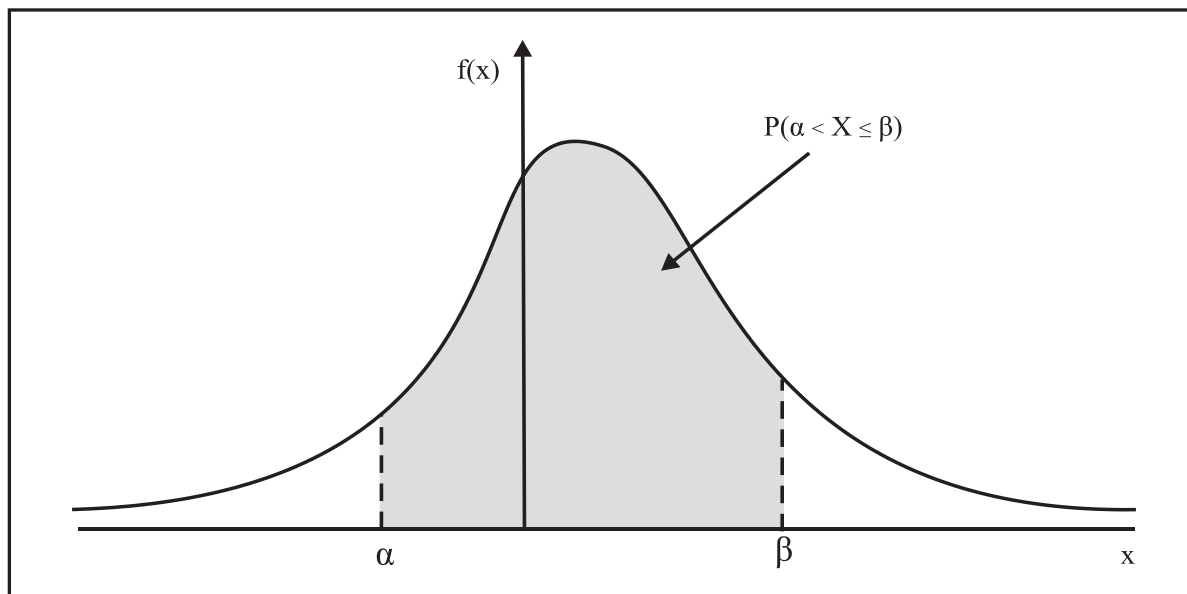
- i. το πεδίο τιμών της $X(S)$ είναι οι πραγματικοί αριθμοί
- ii. υπάρχει συνεχής συνάρτηση $f(x)$ τέτοια ώστε $f(x) \geq 0$ για όλες τις τιμές x στο διάστημα $-\infty < x < \infty$
- iii. για δύο οποιουδήποτε πραγματικούς αριθμούς a και β ($a < \beta$) ισχύει:

$$P(a < X \leq \beta) = \int_a^\beta f(x)dx \quad (5.1)$$

Η συνάρτηση $f(x)$ ονομάζεται **συνάρτηση πυκνότητας πιθανότητας** (*probability density function*) της συνεχούς τυχαίας μεταβλητής X ή, απλώς, **πυκνότητα πιθανότητας**. Από τον ορισμό προκύπτει ότι η συνάρτηση πυκνότητας πιθανότητας $f(x)$ ικανοποιεί τις συνθήκες

i. $f(x) \geq 0, \forall x \in X(S)$ όπου $-\infty < x < +\infty$

ii. $\int_{-\infty}^{+\infty} f(x)dx = 1$

Σχήμα 5.3 Διάγραμμα Συνάρτησης Πυκνότητας Πιθανότητας

Η συνάρτηση $f(x)$ ορίζει μία συνεχή καμπύλη η οποία βρίσκεται εξ ολοκλήρου πάνω από τον άξονα των τετμημένων (Σχήμα 5.3). Η καμπύλη αυτή είναι το **διάγραμμα της συνάρτησης πυκνότητας πιθανότητας** (*probability plot*) της τυχαίας μεταβλητής X .

Όπως φαίνεται στο Σχήμα 5.3, η πιθανότητα $P(\alpha < X \leq \beta)$ απεικονίζεται γραφικά ως το εμβαδόν της επιφάνειας που περιέχεται μεταξύ της καμπύλης που ορίζει η συνάρτηση πυκνότητας πιθανότητας $f(x)$ και του άξονα των x , περιορίζεται δε από τις κατακόρυφες στα α και β . Να σημειωθεί ότι λόγω της σχέσης (5.1) ισχύει ότι $P(X = \alpha) = P(X = \beta) = 0$. Κατά συνέπεια, ισχύει επίσης ότι $P(\alpha < X \leq \beta) = P(\alpha \leq X \leq \beta) = P(\alpha \leq X < \beta) = P(\alpha < X < \beta)$. Επομένως, με το γραμμοσκεπασμένο εμβαδόν του Σχήματος 5.3 απεικονίζεται οποιαδήποτε εκ των πιθανοτήτων: $P(\alpha < X \leq \beta)$, $P(\alpha \leq X \leq \beta)$, $P(\alpha \leq X < \beta)$, $P(\alpha < X < \beta)$.

Το παράδειγμα που ακολουθεί θα βοηθήσει στην καλύτερη κατανόηση των εννοιών που παρουσιάστηκαν παραπάνω.

Παράδειγμα 5.2

Έστω X μια συνεχής τυχαία μεταβλητή της οποίας η συνάρτηση πυκνότητας πιθανότητας είναι

$$f(x) = \begin{cases} 4x^3 & \forall x \in [0, 1] \\ 0 & \forall x \notin [0, 1]. \end{cases}$$

Να υπολογιστεί η πιθανότητα να λάβει η τυχαία μεταβλητή X τιμή

(α) μικρότερη ή ίση του $1/3$.

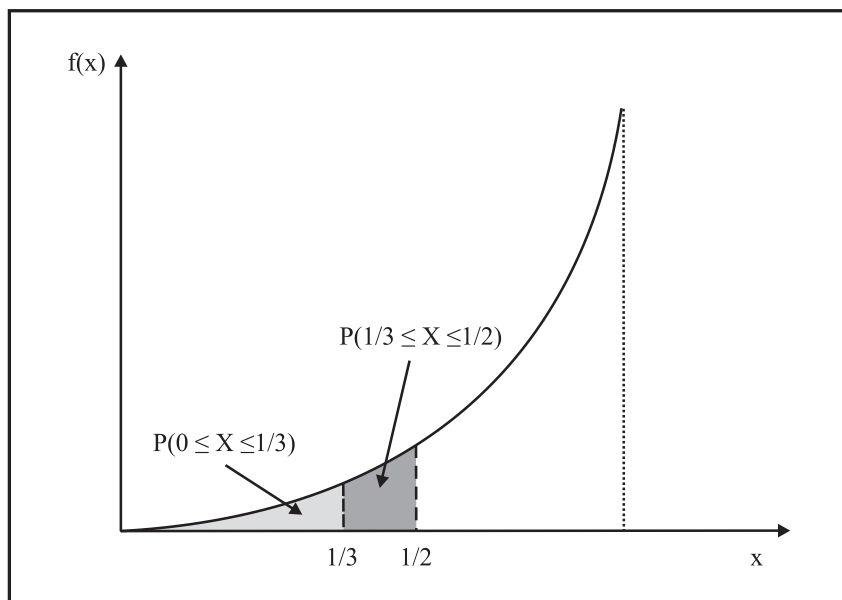
(β) μεταξύ του $1/3$ και του $1/2$.

Λύση

$$\begin{aligned} \alpha. P(0 \leq X \leq 1/3) &= \int_0^{1/3} 4x^3 dx = [x^4]_0^{1/3} \\ &= [(1/3)^4] - [0] = 1/81 \end{aligned}$$

$$\begin{aligned} \beta. P(1/3 \leq X \leq 1/2) &= \int_{1/3}^{1/2} 4x^3 dx = [x^4]_{1/3}^{1/2} \\ &= [(1/2)^4] - [(1/3)^4] = 65/1296 \end{aligned}$$

Σχήμα 5.4 Συνάρτηση πυκνότητας πιθανότητας της $f(x) = 4x^3$



Στο Σχήμα 5.4 απεικονίζονται γραφικά η συνάρτηση πυκνότητας πιθανότητας $f(x)$ του Παραδείγματος 5.2 και οι δύο πιθανότητες που υπολογίστηκαν.

5.3 ΑΘΡΟΙΣΤΙΚΗ ΣΥΝΑΡΤΗΣΗ ΚΑΤΑΝΟΜΗΣ ΠΙΘΑΝΟΤΗΤΑΣ ΤΥΧΑΙΑΣ ΜΕΤΑΒΛΗΤΗΣ

Έστω X τυχαία μεταβλητή (διακριτή ή συνεχή). Η συνάρτηση $F_X : R \rightarrow R$ που ορίζεται ως $F_X(\alpha) \equiv F(\alpha) = P(X \leq \alpha)$, $\forall \alpha \in R$ και η οποία εκφράζει την πιθανότητα να λάβει η τυχαία μεταβλητή X οποιαδήποτε τιμή μικρότερη ή ίση του α , καλείται **αθροιστική συνάρτηση κατανομής πιθανότητας** (*cumulative probability distribution function*) ή, απλώς, **συνάρτηση κατανομής** της τυχαίας μεταβλητής X .

Η συνάρτηση $F(\alpha)$ που ορίζεται με τον τρόπο αυτόν έχει τις παρακάτω πολύ σημαντικές ιδιότητες:

- i. Αν $\alpha < \beta \Rightarrow F(\alpha) \leq F(\beta)$, είναι δηλαδή μη φθίνουσα συνάρτηση.

- ii. $\lim_{a \rightarrow +\infty} F(\alpha) = 1$ και $\lim_{a \rightarrow -\infty} F(\alpha) = 0$. Με άλλα λόγια, η μέγιστη τιμή της είναι η μονάδα και η ελάχιστη το μηδέν.
- iii. Η συνάρτηση $F(\alpha)$ είναι συνεχής τουλάχιστον από δεξιά $\forall \alpha \in R$.

5.3.1 Αθροιστική συνάρτηση κατανομής πιθανότητας διακριτής τυχαίας μεταβλητής

Έστω η διακριτή τυχαία μεταβλητή X με τιμές $x_i, i = 0, 1, 2, \dots$ Τότε,

$$F(\alpha) = P(X \leq \alpha) = \sum_{x_i \leq \alpha} P(x_i), \quad \forall \alpha \in R \quad (5.2)$$

όπου η άθροιση εκτείνεται σε όλα τα $x_i, i = 0, 1, 2, \dots$ τα οποία είναι μικρότερα ή ίσα του α . Σημειώνεται ότι η συνάρτηση κατανομής F μιας διακριτής τυχαίας μεταβλητής X είναι σταθερή κατά διαστήματα και αυξάνει μόνο με άλματα στα σημεία x_i . Με άλλα λόγια, είναι μια **βαθμωτή** (ή **βαθμιδωτή**) **συνάρτηση** (*step function*) με ύψος άλματος σε κάθε x_i , ίσο με την πιθανότητα να πάρει η τυχαία μεταβλητή X την τιμή x_i . Δηλαδή

$$F(x_i) - F(x_{i-1}) = P(X = x_i) \quad (5.3)$$

Αποδεικνύεται ότι η πιθανότητα να λάβει η διακριτή τυχαία μεταβλητή X τιμή μεταξύ δύο οποιωνδήποτε πραγματικών αριθμών α και β , με $\alpha < \beta$, δίνεται από τις σχέσεις:

i. $P(\alpha < X \leq \beta) = F(\beta) - F(\alpha)$

- ii. $P(\alpha < X < \beta) = F(\beta - 1) - F(\alpha)$
- iii. $P(\alpha \leq X \leq \beta) = F(\beta) - F(\alpha - 1)$
- iv. $P(\alpha \leq X < \beta) = F(\beta - 1) - F(\alpha - 1)$

Το παράδειγμα που ακολουθεί αναδεικνύει τη διαδικασία που πρέπει να ακολουθήσουμε προκειμένου να σχηματίσουμε την αθροιστική συνάρτηση κατανομής μιας διακριτής τυχαίας μεταβλητής.

Παράδειγμα 5.3

Να υπολογιστεί η αθροιστική συνάρτηση κατανομής της τυχαίας μεταβλητής X που εκφράζει τον αριθμό των αγοριών σε μία οικογένεια με τρία παιδιά έτσι όπως αυτή ορίστηκε στο Παράδειγμα 5.1 και να παρασταθεί γραφικά.

Λύση

Όπως έχει ήδη βρεθεί, η τυχαία μεταβλητή X μπορεί να πάρει τις τιμές 0, 1, 2, 3 με αντίστοιχες πιθανότητες $1/8, 3/8, 3/8, 1/8$. Σύμφωνα με τον ορισμό που ήδη δόθηκε, οι τιμές της αθροιστικής συνάρτησης κατανομής θα είναι:

$$\text{Για } \alpha < 0, \quad F(a) = \sum_{x_i \leq a} P(x_i) = 0$$

$$\text{Για } 0 \leq \alpha < 1, \quad F(a) = \sum_{x_i \leq a} P(x_i) = P(0) = \frac{1}{8}$$

$$\text{Για } 1 \leq \alpha < 2, \quad F(a) = \sum_{x_i \leq a} P(x_i) = P(0) + P(1) = \frac{1}{8} + \frac{3}{8} = \frac{4}{8}$$

$$\text{Για } 2 \leq \alpha < 3, \quad F(a) = \sum_{x_i \leq a} P(x_i) = P(0) + P(1) + P(2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$$

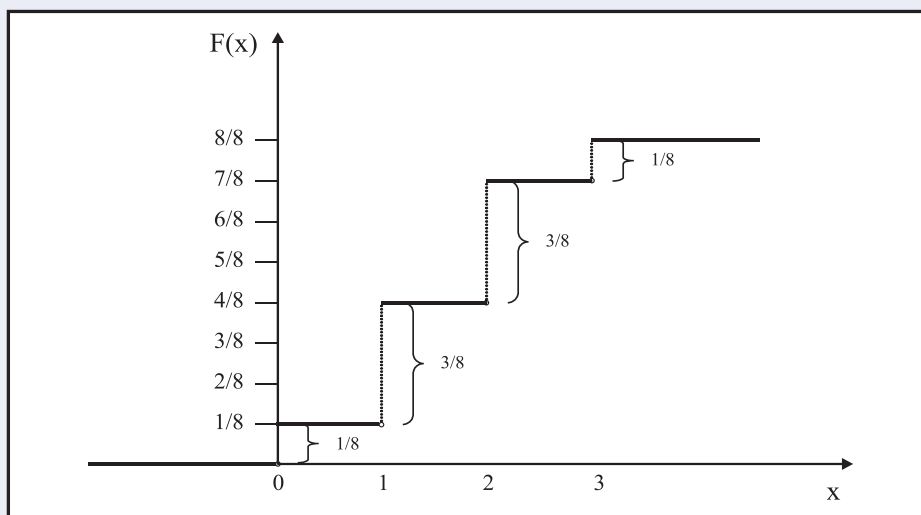
$$\text{Για } \alpha \geq 3, \quad F(a) = \sum_{x_i \leq a} P(x_i) = P(0) + P(1) + P(2) + P(3) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$$

Συνοπτικά, η αθροιστική συνάρτηση κατανομής παρουσιάζεται ως ακολούθως:

$$F(a) = \begin{cases} 0 & a < 0 \\ 1/8 & 0 \leq a < 1 \\ 4/8 & 1 \leq a < 2 \\ 7/8 & 2 \leq a < 3 \\ 1 & a \geq 3 \end{cases}$$

Η γραφική απεικόνιση της αθροιστικής συνάρτησης κατανομής δίνεται στο Σχήμα 5.5.

Σχήμα 5.5 Οικογένεια με τρία παιδιά: Αθροιστική συνάρτηση κατανομής



Από το Σχήμα 5.5 βλέπουμε ότι η $F(x)$ είναι πράγματι μία βαθμωτή συνάρτηση στην οποία το ύψος της βαθμίδας (του άλματος) που αντιστοιχεί στην τιμή x_i είναι κάθε φορά ίσο με την πιθανότητα $P(x_i)$.

5.3.2 Αθροιστική συνάρτηση κατανομής πιθανότητας συνεχούς τυχαίας μεταβλητής

Στην περίπτωση που η τυχαία μεταβλητή X είναι συνεχής, έχουμε

$$P(X \leq x) = \int_{-\infty}^x f(t)dt.$$

Άρα,

$$F(x) = \int_{-\infty}^x f(t)dt. \quad (5.4)$$

Προφανώς η μορφή της καμπύλης της αθροιστικής συνάρτησης κατανομής $F(x)$ εξαρτάται από τη μορφή της καμπύλης της συνάρτησης πυκνότητας πιθανότητας $f(x)$.

Για κάθε συνεχή τυχαία μεταβλητή X ισχύει ότι

$$F'(x) = \frac{dF(x)}{dx} = f(x) \text{ σχεδόν παντού.} \quad (5.5)$$

Σημειώνεται ότι οι σχέσεις (5.3) και (5.4) είναι αντίστοιχες των σχέσεων (5.1) και (5.2) για την περίπτωση συνεχών τυχαίων μεταβλητών. Η συνάρτηση πυκνότητας πιθανότητας δεν εκφράζει την πιθανότητα εμφάνισης κάποιου ενδεχομένου, δηλαδή μιας συγκεκριμένης τιμής x της συνεχούς τυχαίας μεταβλητής X . Μόνο όταν η συνάρτηση ολοκληρώνεται μεταξύ δύο σημείων, εκφράζει πιθανότητα ενδεχομένου (Παράδειγμα 5.2). Σημειώνεται, πάντως, ότι για πολύ μικρό $\Delta x > 0$ έχουμε

$$P(x < X < x + \Delta x) \cong f(x)\Delta x$$

Η ποσότητα $f(x)\Delta x$ ονομάζεται **στοιχείο πιθανότητας** (*propability element*) ή **διαφορικό πιθανότητας** ή **στοιχειώδης πιθανότητα**.

Αποδεικνύεται ότι η πιθανότητα να λάβει η συνεχής τυχαία μεταβλητή X τιμή μεταξύ δύο οποιωνδήποτε πραγματικών αριθμών α και β , με $\alpha < \beta$, δίνεται από τις σχέσεις

$$\begin{aligned} P(\alpha < X \leq \beta) &= P(\alpha < X < \beta) = P(\alpha \leq X \leq \beta) = P(\alpha \leq X < \beta) \\ &= F(\beta) - F(\alpha) \end{aligned} \quad (5.6)$$

δεδομένου ότι $P(X = a) = P(X = \beta) = 0, \forall a \in R$

Στην περίπτωση που γνωρίζουμε την $F(x)$, μπορούμε από τη σχέση (5.5) να βρούμε την τιμή της $f(x)$ σε ένα σημείο x της X , υπολογίζοντας την τιμή $\frac{dF(x)}{dx}$ της παραγώγου της $F(x)$, εφόσον βέβαια αυτή υπάρχει στο σημείο αυτό. Η διαδικασία αυτή αναδεικνύεται μέσω της εναλλακτικής θεώρησης του Παραδείγματος 5.2, έτσι όπως διατυπώνεται στο Παράδειγμα 5.4.

Παράδειγμα 5.4

Έστω X μια συνεχής τυχαία μεταβλητή της οποίας η συνάρτηση πυκνότητας πιθανότητας είναι

$$f(x) = \begin{cases} 4x^3 & \forall x \in [0,1] \\ 0 & \forall x \notin [0,1] \end{cases}$$

Να προσδιοριστεί η αθροιστική συνάρτηση κατανομής της τυχαίας μεταβλητής X και από αυτήν να υπολογιστούν οι πιθανότητες $P(0 \leq X \leq 1/3)$ και $P(1/3 \leq X \leq 1/2)$.

Λύση

Αρχικά, θα υπολογίσουμε την αθροιστική συνάρτηση κατανομής της X , $\forall x \in \mathbb{R}$.

$$\text{Για } x < 0, \quad F(x) = 0$$

$$\text{Για } 0 \leq x \leq 1,$$

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(t)dt = \int_{-\infty}^0 f(t)dt + \int_0^x f(t)dt \\ &= \int_0^x 4t^3 dt = \left[t^4 + c \right]_0^x = x^4 \end{aligned}$$

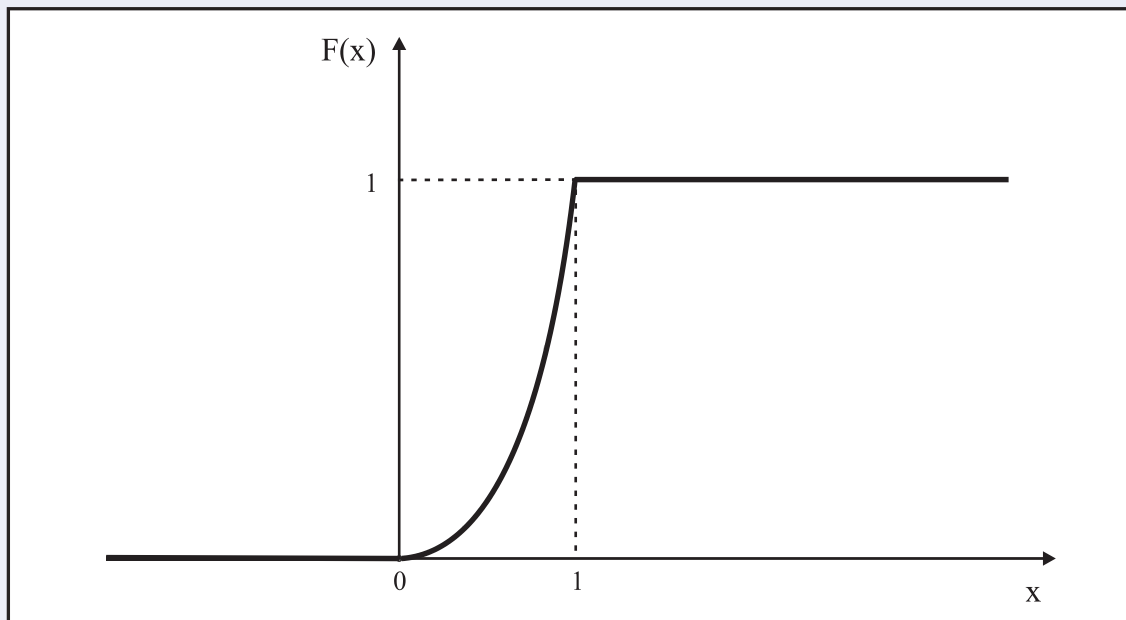
$$\text{Τέλος, για } x > 1, \quad F(x) = 1$$

Άρα, καταλήξαμε ότι

$$F(x) = \begin{cases} 0 & x < 0 \\ x^4 & 0 \leq x \leq 1 \\ 1 & x \geq 1 \end{cases}$$

Η γραφική απεικόνιση της αθροιστικής συνάρτησης της τυχαίας μεταβλητής X δίνεται στο Σχήμα 5.6.

Σχήμα 5.6 Αθροιστική συνάρτηση κατανομής της $f(x) = 4x^3$



Χρησιμοποιώντας τη μορφή της αθροιστικής συνάρτησης κατανομής και τη σχέση (5.5), προκύπτει ότι οι ζητούμενες πιθανότητες είναι

$$\begin{aligned} P(0 \leq X \leq 1/3) &= F(1/3) - F(0) \\ &= (1/3)^4 - 0 = 1/81 \end{aligned}$$

και

$$\begin{aligned} P(1/3 \leq X \leq 1/2) &= F(1/2) - F(1/3) \\ &= (1/2)^4 - (1/3)^4 = 65/1296. \end{aligned}$$

Καταλήγουμε, δηλαδή, στο ίδιο αποτέλεσμα όπως και στο Παράδειγμα 5.2.

5.4 ΜΕΤΡΑ ΚΕΝΤΡΙΚΗΣ ΤΑΣΗΣ

Η κατανομή πιθανότητας μιας τυχαίας μεταβλητής δίνει πληροφορίες για την πιθανότητα εμφάνισης των δυνατών τιμών της. Σε πολλές περιπτώσεις, όμως, απαιτείται μια συνοπτική περιγραφή της πιθανοθεωρητικής συμπεριφοράς της τυχαίας μεταβλητής. Στην περίπτωση αυτή απαιτείται η θεώρηση ορισμένων χαρακτηριστικών μέτρων της κατανομής πιθανότητας, τα οποία να την περιγράφουν με ικανοποιητικό τρόπο. Στην ενότητα αυτή θα παρουσιαστούν τα κυριότερα **μέτρα κεντρικής τάσης** (*measures of central tendency*).

5.4.1 Μέση τιμή

Το βασικότερο μέτρο κεντρικής τάσης της κατανομής μιας τυχαίας μεταβλητής είναι η μέση τιμή της. Η **μέση τιμή** (*mean value*), η οποία καλείται επίσης και **αναμενόμενη τιμή** (*expected value*) ή **μαθηματική ελπίδα** (*mathematical expectation*), ορίζεται ως εξής:

α. Μέση τιμή διακριτής τυχαίας μεταβλητής

Έστω X διακριτή τυχαία μεταβλητή με συνάρτηση πιθανότητας $P(X = x_i)$. Η μέση τιμή της δίνεται από τον τύπο

$$E(X) = \sum_{i=1}^{\infty} x_i P(x_i) \quad (5.7)$$

με την προϋπόθεση ότι η σειρά συγκλίνει απολύτως.

Αν η X παίρνει πεπερασμένο πλήθος τιμών, τότε έχουμε

$$E(X) = \sum_{i=1}^n x_i P(x_i). \quad (5.8)$$

Αξίζει να τονιστεί ότι η τιμή της $E(X)$ δεν είναι υποχρεωτικά κάποια από τις τιμές που παίρνει η τυχαία μεταβλητή X .

Με το παράδειγμα που ακολουθεί παρουσιάζουμε τη διαδικασία υπολογισμού της αναμενόμενης τιμής της διακριτής τυχαίας μεταβλητής X .

Παράδειγμα 5.5

Επιστρέφουμε στο Παράδειγμα 5.1, όπου η τυχαία μεταβλητή εκφράζει τον αριθμό των αγοριών σε μια οικογένεια με τρία παιδιά, και θέλουμε να υπολογίσουμε τη μέση τιμή της.

Λύση

Χρησιμοποιώντας τα αποτελέσματα του Παραδείγματος 5.1 και εφαρμόζοντας τον τύπο (5.8) έχουμε

$$\begin{aligned} E(X) &= \sum_{i=1}^4 x_i P(X = x_i) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) = \\ &= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 0 + \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = \frac{12}{8} = \frac{3}{2} \Rightarrow \end{aligned}$$

$$E(X) = 1,5$$

β. Μέση τιμή συνεχούς τυχαίας μεταβλητής

Έστω X συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας $f(x)$. Η μέση τιμή της δίνεται από τον τύπο

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx \quad (5.9)$$

με την προϋπόθεση ότι το ολοκλήρωμα συγκλίνει απολύτως.

Αν η πιθανοθεωρητική συμπεριφορά της τυχαίας μεταβλητής X είναι τέτοια ώστε η αντίστοιχη $f(x)$ να παίρνει μη μηδενικές τιμές μόνο στο διάστημα $[\alpha, \beta]$, τότε μπορούμε να γράψουμε

$$E(X) = \int_{\alpha}^{\beta} xf(x)dx \quad (5.10)$$

Με το παράδειγμα που ακολουθεί παρουσιάζουμε τη διαδικασία υπολογισμού της αναμενόμενης τιμής μιας συνεχούς τυχαίας μεταβλητής X .

Παράδειγμα 5.6

Επιστρέφουμε στο Παράδειγμα 5.2, όπου ήταν γνωστή η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής X , δηλαδή

$$f(x) = \begin{cases} 4x^3 & \forall x \in [0, 1] \\ 0 & \forall x \notin [0, 1] \end{cases}$$

και ζητάμε να υπολογίσουμε τη μέση τιμή της X .

Λύση

Γνωρίζουμε ότι στην περίπτωση αυτή ισχύει

$$\begin{aligned} E(X) &= \int_a^b xf(x)dx = \int_0^1 x4x^3dx \\ &= \left[\frac{4x^5}{5} \right]_0^1 = [(4/5)] - [0] = 4/5. \end{aligned}$$

Στη συνέχεια θα αναφέρουμε ένα θεώρημα το οποίο είναι ιδιαίτερα χρήσιμο, δεδομένου ότι δίνει τη δυνατότητα υπολογισμού της μέσης τιμής μιας πραγματικής συνάρτησης, έστω $g(X)$, της τυχαίας μεταβλητής X , χωρίς να γνωρίζουμε την κατανομή πιθανότητας της $g(X)$.

Θεώρημα 5.1

Έστω X μία τυχαία μεταβλητή και $g(\cdot)$ μια πραγματική συνάρτηση. Τότε,

- αν η τυχαία μεταβλητή X είναι διακριτή, η μέση τιμή της τυχαίας μεταβλητής $g(X)$ δίνεται από τη σχέση

$$E[g(X)] = \sum_{i=1}^{\infty} g(x_i) P(X = x_i) \quad (5.11)$$

- αν η τυχαία μεταβλητή X είναι συνεχής και η πραγματική συνάρτηση είναι τέτοια ώστε η τυχαία μεταβλητή $g(X)$ να είναι συνεχής, η μέση της τιμή δίνεται από τη σχέση

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x) dx \quad (5.12)$$

με την προϋπόθεση, βέβαια, ότι τόσο η σειρά όσο και το ολοκλήρωμα, κατά περίπτωση, συγκλίνουν απολύτως.

Εφαρμόζοντας το συμπέρασμα του θεωρήματος 5.1 μπορούμε να διατυπώσουμε το παρακάτω χρήσιμο θεώρημα, το οποίο, μαζί με τα πορίσματα που θα ακολουθήσουν, δίνει ευελιξία κατά τον υπολογισμό των αναμενόμενων τιμών συναρτήσεων της τυχαίας μεταβλητής X .

Θεώρημα 5.2

Έστω X μία τυχαία μεταβλητή και $g(X)$ αλγεβρική συνάρτηση της X , της οποίας η μέση τιμή υπάρχει. Έστω επίσης α, β σταθερές. Τότε αποδεικνύεται ότι

$$E[ag(X) + \beta] = aE[g(X)] + \beta \quad (5.13)$$

Με βάση το Θεώρημα 5.2 είναι πλέον εύκολο να κατανοηθούν τα συμπεράσματα που περιέχονται στο Πόρισμα 5.1 που ακολουθεί.

Πόρισμα 5.1

Έστω X μία τυχαία μεταβλητή της οποίας η μέση τιμή υπάρχει. Έστω επίσης a, β σταθερές. Τότε αποδεικνύεται ότι:

$$\text{i. } E(aX + \beta) = aE(X) + \beta, \text{ όπου } a, \beta \text{ είναι σταθερές.} \quad (5.14)$$

Ειδικά $E(\beta) = \beta$, όπου β είναι σταθερά

$$E(aX) = aE(X), \text{ όπου } a \text{ είναι σταθερά.}$$

$$\text{ii. } E(X - E(X)) = 0 \quad (5.15)$$

Θεώρημα 5.3

Έστω X μία τυχαία μεταβλητή και $g_1(X), g_2(X)$ αλγεβρικές συναρτήσεις της X των οποίων οι μέσες τιμές υπάρχουν. Έστω επίσης a_1, a_2 σταθερές. Τότε αποδεικνύεται ότι

$$E[a_1 g_1(X) + a_2 g_2(X)] = a_1 E[g_1(X)] + a_2 E[g_2(X)] \quad (5.16)$$

Το συμπέρασμα αυτό γενικεύεται για $g_i(X)$, $i = 1, 2, \dots, n$ συναρτήσεις μιας τυχαίας μεταβλητής X και a_i , $i = 1, 2, \dots, n$ σταθερές, οπότε

$$E\left[\sum_{i=1}^n a_i g_i(X)\right] = \sum_{i=1}^n a_i E[g_i(X)] \quad (5.17)$$

5.4.2 Διάμεσος

Διάμεσος (*median*) μιας τυχαίας μεταβλητής X είναι εκείνη η τιμή που χωρίζει την κατανομή σε δύο ίσα μέρη. Η διάμεσος μιας τυχαίας μεταβλητής X με συνάρτηση κατανομής $F(x)$, $\forall x \in R$ είναι η τιμή M για την οποία ισχύει

$$P(X < M) \leq 1/2 \quad \text{και} \quad P(X \leq M) \geq 1/2 \quad (5.18)$$

Για την περίπτωση διακριτών τυχαίων μεταβλητών, η διάμεσος είναι η μικρότερη τιμή που δίνει $F(x) \geq 1/2$. Να σημειώσουμε ότι, αν η $F(x)$ είναι συνεχής, η διάμεσος της κατανομής είναι κάθε λύση M της εξίσωσης

$$F(M) = 1/2 \quad (5.19)$$

Αν επιπλέον η $F(x)$ είναι γνησίως αύξουσα, τότε η εξίσωση (5.19) έχει μοναδική λύση. Τα παραδείγματα που ακολουθούν αποσαφηνίζουν τα όσα ήδη αναφέρθηκαν.

Παράδειγμα 5.7

Επιστρέφουμε στο Παράδειγμα 5.1, όπου η τυχαία μεταβλητή εκφράζει τον αριθμό των αγοριών σε μια οικογένεια με τρία παιδιά. Θέλουμε να υπολογίσουμε τη διάμεσό της.

Λύση

Χρησιμοποιώντας τα αποτελέσματα του Παραδείγματος 5.3, παρατηρούμε ότι η μικρότερη τιμή για την οποία $F(x) \geq \frac{1}{2}$, είναι η τιμή $x = 1$. Κατά συνέπεια, η τιμή της διαμέσου είναι $M = 1$.

Παράδειγμα 5.8

Επιστρέφουμε στο Παράδειγμα 5.2, όπου ήταν γνωστή η συνάρτηση πυκνότητας της τυχαίας μεταβλητής X . Δηλαδή

$$f(x) = \begin{cases} 4x^3, & \forall x \in [0, 1] \\ 0, & \forall x \notin [0, 1]. \end{cases}$$

Ζητάμε να υπολογίσουμε τη διάμεσο τιμή της X .

Λύση

Γνωρίζουμε ότι η διάμεσος τιμή M ικανοποιεί την εξίσωση (5.19). Κατά συνέπεια,

$$\int_{-\infty}^M f(x)dx = 1/2, \quad \forall M \in [0,1]$$

Επομένως

$$\int_0^M 4x^3 dx = 1/2 \Rightarrow \left[x^4 \right]_0^M = 1/2 \Rightarrow M^4 = 1/2 \Rightarrow M = \sqrt[4]{1/2}$$

Η άλλη πραγματική λύση της εξίσωσης, δηλαδή η $M = -\sqrt[4]{1/2}$, απορρίπτεται, δεδομένου ότι δεν ανήκει στο διάστημα $[0, 1]$.

5.4.3 Ποσοστιαία σημεία

Όπως έχει ήδη αναφερθεί, η διάμεσος μιας κατανομής είναι το 50ό εκατοστιαίο σημείο της. Γενικά **p-ποσοστιαίο σημείο** (*quantile*) μιας τυχαίας μεταβλητής X με συνάρτηση κατανομής $F(x)$, $\forall x \in R$ είναι η τιμή x_p για την οποία ισχύει:

$$P(X < x_p) \leq p \quad \text{και} \quad P(X \leq x_p) \geq p$$

Να σημειώσουμε ότι, αν η $F(x)$ είναι συνεχής, το p -ποσοστιαίο σημείο της κατανομής είναι κάθε λύση, x_p , της εξίσωσης

$$F(x_p) = p.$$

Αν επιπλέον η $F(x)$ είναι γνησίως αύξουσα, τότε η εξίσωση (5.19) έχει μοναδική λύση. Προφανώς:

Αν $p = 0.5$, το x_p είναι η διάμεσος (M)

Αν $p = 0.25$, το x_p είναι το 1° τεταρτημόριο (Q_1)

Αν $p = 0.75$, το x_p είναι το 3° τεταρτημόριο (Q_3)

5.4.4 Επικρατούσα τιμή

Επικρατούσα τιμή (*mode*) μιας τυχαίας μεταβλητής X είναι εκείνη η τιμή που έχει τη μεγαλύτερη πιθανότητα εμφάνισης από κάθε άλλη τιμή. Η επικρατούσα τιμή ορίζεται ως εξής:

α. Επικρατούσα τιμή διακριτής τυχαίας μεταβλητής

Έστω X διακριτή τυχαία μεταβλητή με συνάρτηση πιθανότητας $P(x)$. Η επικρατούσα τιμή είναι εκείνη η τιμή T_0 για την οποία ισχύει

$$P(X = T_0) \geq P(X = x), \quad \forall x$$

Παράδειγμα 5.9

Επιστρέφουμε στο Παράδειγμα 5.1, όπου η τυχαία μεταβλητή εκφράζει τον αριθμό των αγοριών σε μια οικογένεια με τρία παιδιά. Θέλουμε να υπολογίσουμε την επικρατούσα τιμή της.

Λύση

Χρησιμοποιώντας τα αποτελέσματα του Παραδείγματος 5.1 διαπιστώνουμε ότι υπάρχουν δύο τιμές με τη μεγαλύτερη πιθανότητα εμφάνισης, οι τιμές $x = 1$ και $x = 2$. Κατά συνέπεια, η υπό μελέτη κατανομή είναι δικόρυφη με $T_0 = 1$ και $T_0 = 2$

β. Επικρατούσα τιμή συνεχούς τυχαίας μεταβλητής

Έστω X συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας $f(x)$. Η επικρατούσα τιμή είναι εκείνη η τιμή T_0 για την οποία ισχύει

$$f(T_0) \geq f(x), \quad \forall x.$$

Σημειώνεται ότι, αν η συνάρτηση πυκνότητας πιθανότητας έχει παράγωγο δεύτερης τάξης, τότε ένα σημείο T_0 με $f'(T_0) = 0$ και $f''(T_0) < 0$ είναι επικρατούσα τιμή της κατανομής.

Παράδειγμα 5.10

Έστω X συνεχής τυχαία μεταβλητή με συνάρτηση κατανομής την

$$F(x) = \begin{cases} 0, & -\infty < x < 0 \\ 1 - e^{-\lambda x^2}, & 0 \leq x < \infty, \quad \lambda > 0. \end{cases}$$

Να υπολογιστεί η επικρατούσα τιμή της X .

Λύση

Η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής X υπολογίζεται από τη συνάρτηση κατανομής με βάση τον τύπο (5.5). Επομένως,

$$f(x) = \frac{dF(x)}{dx} = 2\lambda x e^{-\lambda x^2}, \quad 0 \leq x < \infty$$

Παραγωγίζοντας την $f(x)$ έχουμε διαδοχικά:

$$f'(x) = 2\lambda(1 - 2\lambda x^2)e^{-\lambda x^2} \quad \text{και} \quad f''(x) = -4\lambda^2 x(3 - 2\lambda x^2)e^{-\lambda x^2}$$

Παρατηρούμε ότι

$$f'(x) = 0 \Rightarrow 2\lambda(1 - 2\lambda x^2)e^{-\lambda x^2} = 0 \Rightarrow 1 - 2\lambda x^2 = 0 \Rightarrow x = \pm 1/\sqrt{2\lambda}$$

Επίσης,

$$f''(1/\sqrt{2\lambda}) = -4\lambda\sqrt{2\lambda}e^{-1/2} < 0, \quad \text{δεδομένου ότι } \lambda > 0$$

Η τιμή $x = -1/\sqrt{2\lambda}$ δεν εξετάζεται, δεδομένου ότι βρίσκεται έξω από το διάστημα $[0, \infty)$ στο οποίο η συνάρτηση πυκνότητας πιθανότητας είναι θετική. Άρα, $T_0 = 1/\sqrt{2\lambda}$.

5.5 ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

Η μέση τιμή και τα άλλα μέτρα κεντρικής τάσης μιας τυχαίας μεταβλητής X , τα οποία παρουσιάστηκαν στην προηγούμενη ενότητα, εκφράζουν μια αντιπροσωπευτική τιμή της X , αλλά δεν είναι αρκετά για να μας επιτρέψουν να περιγράψουμε ικανοποιητικά την κατανομή πιθανότητάς της και να συγκρίνουμε διαφορετικές κατανομές. Για το σκοπό αυτόν, χρειαζόμαστε κάποια άλλα μέτρα που να μας δίνουν τη διασπορά των τιμών της μεταβλητής γύρω από την αντιπροσωπευτική τιμή της. Τα μέτρα αυτά ονομάζονται μέτρα διασποράς και τα κυριότερα από αυτά παρουσιάζονται στη συνέχεια.

5.5.1 Διακύμανση και τυπική απόκλιση

Το βασικότερο μέτρο διασποράς μιας τυχαίας μεταβλητής είναι η διακύμανσή της, η οποία ορίζεται ως εξής:

Έστω X τυχαία μεταβλητή (διακριτή ή συνεχής) με μέση τιμή $\mu = E(X)$. Η **διακύμανση** (*variance*) της X συμβολίζεται με $V(X)$ ή σ^2 και δίνεται από τον τύπο

$$V(X) = \sigma^2 = E[X - E(X)]^2 = E(X - \mu)^2 \quad (5.20)$$

Πολλές φορές για διευκόλυνση, αντί του τύπου (5.17), χρησιμοποιούμε τον ισοδύναμο τύπο του

$$V(X) = \sigma^2 = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2 \quad (5.21)$$

Όπως φαίνεται από τον παραπάνω ορισμό, η διακύμανση δεν εκφράζεται στις μονάδες της X αλλά στα τετράγωνα των μονάδων της. Για να ξεπεράσουμε το πρόβλημα αυτό, χρησιμοποιούμε συνήθως, αντί της διακύμανσης, τη θετική τετραγωνική της ρίζα

$$\sigma = +\sqrt{V(X)}$$

η οποία καλείται **μέση απόκλιση τετραγώνου** ή **τυπική απόκλιση** (*mean square deviation or standard deviation*) της X .

Για τη διακύμανση μιας τυχαιάς μεταβλητής αποδεικνύεται ότι ισχύουν ορισμένες ιδιότητες, οι σπουδαιότερες από τις οποίες παρουσιάζονται στη συνέχεια:

$$1. \quad V(a) = 0, \text{ όπου } a \text{ σταθερά} \quad (5.22)$$

$$2. \quad V(X + \beta) = V(X), \text{ όπου } \beta \text{ σταθερά} \quad (5.23)$$

$$3. \quad V(aX) = a^2 V(X), \text{ όπου } a \text{ σταθερά} \quad (5.24)$$

$$4. \quad V(aX + \beta) = a^2 V(X), \text{ όπου } a, \beta \text{ σταθερές} \quad (5.25)$$

Με το παράδειγμα που ακολουθεί παρουσιάζουμε τη διαδικασία υπολογισμού της διακύμανσης και της τυπικής απόκλισης για την περίπτωση διακριτής τυχαιάς μεταβλητής X .

Παράδειγμα 5.11

Επιστρέφουμε στο Παράδειγμα 5.1, όπου η τυχαία μεταβλητή εκφράζει τον αριθμό των αγοριών σε μια οικογένεια με τρία παιδιά και θέλουμε να υπολογίσουμε τη διακύμανση και την τυπική απόκλιση της.

Λύση

Γνωρίζουμε ότι η διακύμανση υπολογίζεται με βάση τον τύπο

$$V(X) = \sigma^2 = E(X^2) - [E(X)]^2.$$

Έχει ήδη υπολογιστεί (Παράδειγμα 5.5) ότι $E(X) = \frac{3}{2} = 1.5$

Άρα, αρκεί να υπολογίσουμε το $[E(X^2)]$ και αυτό μπορεί να γίνει με βάση το συμπέρασμα του Θεωρήματος 5.1. Δηλαδή

$$\begin{aligned} E(X^2) &= \sum_{i=0}^3 x_i^2 P(X = x_i) = 0^2 \cdot P(X = 0) + 1^2 \cdot P(X = 1) + 2^2 \cdot P(X = 2) + 3^2 \cdot P(X = 3) \\ &= 1 \cdot \frac{3}{8} + 4 \cdot \frac{3}{8} + 9 \cdot \frac{1}{8} = \frac{3}{8} + \frac{12}{8} + \frac{9}{8} = \frac{24}{8} \Rightarrow \\ E(X^2) &= 3 \end{aligned}$$

Άρα,

$$V(X) = 3 - \left(\frac{3}{2}\right)^2 = 3 - \frac{9}{4} = \frac{12-9}{4} = \frac{3}{4} \text{ και } \sigma = +\sqrt{V(X)} = +\sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2}.$$

Με το παράδειγμα που ακολουθεί παρουσιάζουμε τη διαδικασία υπολογισμού της διακύμανσης και της τυπικής απόκλισης για την περίπτωση συνεχούς τυχαίας μεταβλητής X .

Παράδειγμα 5.12

Επιστρέφουμε στο Παράδειγμα 5.2, όπου ήταν γνωστή η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής X , δηλαδή

$$f(x) = \begin{cases} 4x^3 & \forall x \in [0,1] \\ 0 & \forall x \notin [0,1] \end{cases}$$

και ζητάμε να υπολογίσουμε τη διακύμανση και τη μέση απόκλιση τετραγώνου της X .

Λύση

Γνωρίζουμε ότι η διακύμανση υπολογίζεται με βάση τον τύπο

$$V(X) = \sigma^2 = E(X^2) - [E(X)]^2$$

Έχει ήδη υπολογιστεί (Παράδειγμα 5.6) ότι $E(X) = 4/5$. Αρκεί, λοιπόν, να υπολογίσουμε την $E(X^2)$. Ο υπολογισμός θα γίνει με βάση το συμπέρασμα του θεωρήματος 5.1. Κατά συνέπεια,

$$\begin{aligned}
 E(X^2) &= \int_a^b x^2 f(x) dx = \int_0^1 x^2 4x^3 dx \\
 &= \left[\frac{4x^6}{6} + c \right]_0^1 = \left[\frac{4}{6} \right] - [0] = 2/3.
 \end{aligned}$$

Άρα,

$$V(X) = 2/3 - [4/5]^2 = 2/75$$

Επίσης, $\sigma = \sqrt{\frac{2}{75}}$

5.5.2 Ενδοτεταρτημοριακό εύρος και τεταρτημοριακή απόκλιση

Το **ενδοτεταρτημοριακό εύρος** (*interquartile range*) ορίζεται ως $Q_3 - Q_1$ και η

τεταρτημοριακή απόκλιση (*quartile deviation*) ως $Q = \frac{Q_3 - Q_1}{2}$

Το Q_1 (1° τεταρτημόριο ή 25ό εκατοστιαίο σημείο) και το Q_3 (3° τεταρτημόριο ή 75ό εκατοστιαίο σημείο) υπολογίζονται με βάση τη γενική μεθοδολογία που δόθηκε σε προηγούμενη ενότητα για το p -ποσοστιαίο σημείο.

5.6 ΡΟΠΕΣ

Οι περισσότερες κατανομές πιθανότητας περιγράφονται πλήρως όταν γνωρίζουμε τη μέση τιμή και τη διακύμανσή τους. Υπάρχουν, όμως, και ορισμένες περιπτώσεις όπου κατανομές πιθανότητας με την ίδια μέση τιμή και διακύμανση έχουν τελείως διαφορετική μορφή. Δύο τέτοιες κατανομές πιθανότητας απεικονίζονται στο Σχήμα 5.7.

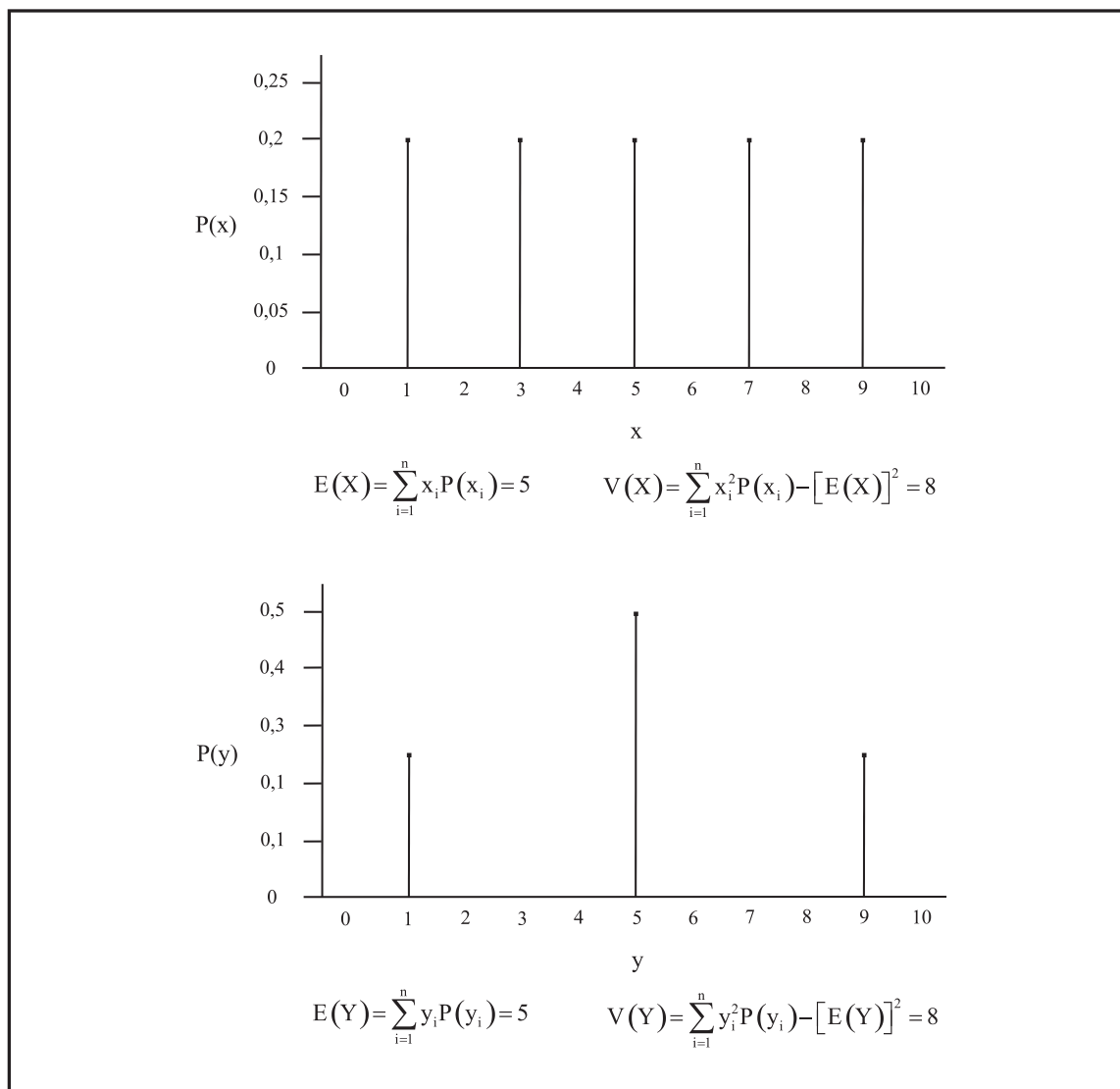
Η **ροπή** (*moment*) k **τάξης** ($k = 1, 2, 3, \dots$) **γύρω από την αρχή** μιας τυχαίας μεταβλητής X ορίζεται ως $E(X^k)$ και συμβολίζεται με μ'_k . Αν η μ'_k υπάρχει, τότε

$$\mu'_k = E(X^k) = \begin{cases} \sum_x x^k P(X=x), & \text{αν } X \text{ διακριτή} \\ \int_{-\infty}^{+\infty} x^k f(x)dx, & \text{αν } X \text{ συνεχής.} \end{cases} \quad (5.26)$$

Προφανώς, η ροπή πρώτης τάξης γύρω από την αρχή, δηλαδή

$$\mu'_1 = E(X) = \begin{cases} \sum_x xP(X=x), & \text{αν } X \text{ διακριτή} \\ \int_{-\infty}^{+\infty} xf(x)dx, & \text{αν } X \text{ συνεχής} \end{cases} \quad (5.27)$$

εκφράζει τη μέση τιμή της κατανομής πιθανότητας της X .

Σχήμα 5.7 Κατανομές πιθανότητας με την ίδια μέση τιμή και την ίδια διακύμανση

Η ροπή k τάξης ($k = 1, 2, 3, \dots$) γύρω από τον μέσο (ή κεντρική ροπή k τάξης) μιας τυχαίας μεταβλητής X ορίζεται ως η $E(X - \mu)^k$ και συμβολίζεται με μ_k . Αν η μ_k υπάρχει, τότε

$$\mu_k = E((X - \mu)^k) = \begin{cases} \sum_x (x - \mu)^k P(X = x), & \text{αν } X \text{ διακριτή} \\ \int_{-\infty}^{+\infty} (x - \mu)^k f(x) dx, & \text{αν } X \text{ συνεχής.} \end{cases} \quad (5.28)$$

Προφανώς, η ροπή δεύτερης τάξης γύρω από τον μέσο, δηλαδή

$$\mu_2 = E\left((X - \mu)^2\right) = \begin{cases} \sum_x (x - \mu)^2 P(X = x), & \text{αν } X \text{ διακριτή} \\ \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx, & \text{αν } X \text{ συνεχής} \end{cases} \quad (5.29)$$

εκφράζει τη διακύμανση της κατανομής πιθανότητας της X .

Οι κεντρικές ροπές έχουν ιδιαίτερη σημασία γιατί δίνουν μία ένδειξη για το πώς μια τυχαία μεταβλητή κατανέμεται γύρω από τη μέση τιμή της.

Η τρίτη κεντρική ροπή μάς δίνει πληροφορίες για το βαθμό ασυμμετρίας της κατανομής. Αν η κατανομή είναι συμμετρική, τότε $\mu_3 = 0$.

Τέλος, η τέταρτη κεντρική ροπή μάς δίνει πληροφορίες για το βαθμό κύρτωσης (αιχμηρότητας) της κατανομής.

5.7 ΔΙΜΕΤΑΒΛΗΤΕΣ ΚΑΤΑΝΟΜΕΣ ΠΙΘΑΝΟΤΗΤΑΣ

Στις ενότητες που προηγήθηκαν, ασχοληθήκαμε με τη μελέτη ενός φαινομένου όπως αυτό εκφράζεται μέσω συγκεκριμένης τυχαίας μεταβλητής. Στην πράξη, όμως, συμβαίνει συχνά να ενδιαφερόμαστε για τη μελέτη φαινομένων τα οποία περιγράφονται από παραπάνω από μία συνιστώσες, εκφράζονται δηλαδή με περισσότερες από μία τυχαίες μεταβλητές, οι οποίες είναι πιθανό να αλληλοεξαρτώνται. Στις περιπτώσεις αυτές, σκοπός μας είναι η από κοινού μελέτη των τυχαίων αυτών μεταβλητών.

Όπως και στην περίπτωση της μελέτης μιας τυχαίας μεταβλητής, μπορούμε να περιγράψουμε τη μεταβλητότητα πολλών τυχαίων μεταβλητών με τη χρήση πολυμεταβλητών κατανομών πιθανότητας και αντίστοιχων μέτρων κεντρικής τάσης, διασποράς κ.ο.κ.

Η απλούστερη περίπτωση, η οποία και θα παρουσιαστεί στην ενότητα αυτή, είναι η περίπτωση ζεύγους τυχαίων μεταβλητών X, Y όπου και οι δύο τυχαίες μεταβλητές είτε είναι διακριτές είτε συνεχείς.

5.7.1 Διμεταβλητές κατανομές πιθανότητας

Έστω X, Y ζεύγος τυχαίων μεταβλητών. Για τις τυχαίες αυτές μεταβλητές θα ορίσουμε στη συνέχεια ορισμένες συναρτήσεις, οι οποίες περιγράφουν την από κοινού συμπεριφορά του ζεύγους των τυχαίων αυτών μεταβλητών.

Ως **από κοινού συνάρτηση πιθανότητας** (*joint probability function*) (αν X, Y διακριτές τυχαίες μεταβλητές) ή **από κοινού συνάρτηση πυκνότητας πιθανότητας** (*joint probability density function*) (αν X, Y συνεχείς τυχαίες μεταβλητές) ορίζεται η συνάρτηση $P_{X,Y}(x, y)$ ή $f_{X,Y}(x, y)$ με πεδίο ορισμού το $R \times R$ και πεδίο τιμών το R ως εξής:

- Αν X, Y **διακριτές** τυχαίες μεταβλητές, τότε η από κοινού συνάρτηση πιθανότητας αυτών ορίζεται ως

$$P_{X,Y}(x, y) = P(X = x, Y = y)$$

και θα πρέπει να ικανοποιεί τις συνθήκες:

$$\text{i. } P_{X,Y}(x, y) = P(X = x, Y = y) \geq 0,$$

$$\forall x \in X(S) \subseteq R, y \in Y(S) \subseteq R$$

$$\text{ii. } \sum_{x \in X(S)} \sum_{y \in Y(S)} P_{X,Y}(x, y) = 1$$

- Αν X, Y **συνεχείς** τυχαίες μεταβλητές, τότε η από κοινού συνάρτηση πυκνότητας πιθανότητας αυτών ορίζεται ως

$$f_{X,Y}(x,y) = \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \frac{P(x \leq X < x + \Delta x, y \leq Y < y + \Delta y)}{\Delta x \Delta y}$$

και θα πρέπει να ικανοποιεί τις συνθήκες:

$$\text{i. } f_{X,Y}(x,y) \geq 0 \quad \forall x \in X(S), y \in Y(S)$$

$$\text{ii. } \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dx dy = 1.$$

Η συνάρτηση $F_{X,Y} : R \times R \rightarrow R$, που ορίζεται ως $F_{X,Y}(a,b) = P(X \leq a, Y \leq b)$, $\forall a, b \in R$ και η οποία εκφράζει την πιθανότητα να λάβει η τυχαία μεταβλητή X οποιαδήποτε τιμή μικρότερη ή ίση του a και η τυχαία μεταβλητή Y οποιαδήποτε τιμή μικρότερη ή ίση του b , καλείται **από κοινού αθροιστική συνάρτηση κατανομής πιθανότητας** (*joint cumulative probability distribution function*) και ορίζεται ως εξής

$$F_{X,Y}(a,b) = P(X \leq a, Y \leq b) = \begin{cases} \sum_{x \leq a} \sum_{y \leq b} P_{X,Y}(x,y) & \text{αν } X, Y \text{ διακριτές} \\ \int_{-\infty}^a \int_{-\infty}^b f_{X,Y}(x,y) dx dy & \text{αν } X, Y \text{ συνεχείς.} \end{cases}$$

Αν X και Y τυχαίες μεταβλητές των οποίων η από κοινού κατανομή είναι καλά ορισμένη, τότε η κατανομή πιθανότητας της X (αντίστοιχα της Y) ονομάζεται **περιθώρια κατανομή πιθανότητας** (*marginal probability distribution*) της X (αντίστοιχα της Y) και υπολογίζεται ως εξής:

- Αν X, Y **διακριτές** (*discrete*) τυχαίες μεταβλητές, η περιθώρια κατανομή της X συμβολίζεται με $P_X(x)$, ορίζεται ως

$$P_X(x) = \sum_y P_{X,Y}(x, y)$$

και θα πρέπει να ικανοποιεί τις συνθήκες:

i. $P_X(x) \geq 0 \quad \forall x$

ii. $\sum_x P_X(x) = 1$

Ομοίως, ορίζεται η περιθώρια κατανομή της Y , $P_Y(y) = \sum_x P_{X,Y}(x, y)$

- Αν X, Y **συνεχείς** (*continuous*) τυχαίες μεταβλητές, η περιθώρια κατανομή της X συμβολίζεται με $f_X(x)$, υπολογίζεται ως

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy$$

και θα πρέπει να ικανοποιεί τις συνθήκες:

i. $f_X(x) \geq 0 \quad \forall x$

$$\text{ii. } \int_{-\infty}^{+\infty} f_X(x) dx = 1$$

Ομοίως, ορίζεται η περιθώρια κατανομή της Y , $f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dx$

Με βάση την περιθώρια κατανομή πιθανότητας μιας τυχαίας μεταβλητής X είναι δυνατό να υπολογιστεί και η αντίστοιχη **περιθώρια αθροιστική συνάρτηση κατανομής** (*marginal cumulative distribution function*) ως ακολούθως:

$$F_X(a) = P(X \leq a) = \begin{cases} \sum_{x \leq a} P_X(x) = \sum_{x \leq a} \sum_y P_{X,Y}(x,y) & \text{αν } X, Y \text{ διακριτές} \\ \int_{-\infty}^a f_X(x) dx = \int_{-\infty}^a \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dx dy & \text{αν } X, Y \text{ συνεχείς.} \end{cases}$$

Με βάση, λοιπόν, τα παραπάνω προκύπτει ότι, όταν γνωρίζουμε την από κοινού κατανομή πιθανοτήτων δύο τυχαίων μεταβλητών X και Y , μπορούμε να προσδιορίσουμε την περιθώρια κατανομή για καθεμιά από τις X και Y . Το αντίστροφο δεν ισχύει (εκτός από κάποιες ειδικές περιπτώσεις). Σημειώνουμε επίσης ότι οι περιθώριες κατανομές, οι οποίες υπολογίζονται από την από κοινού κατανομή πιθανοτήτων, είναι ίδιες με τις κατανομές που προκύπτουν όταν κάθε τυχαία μεταβλητή εξετάζεται χωριστά.

Οι **δεσμευμένες κατανομές πιθανότητας** (*conditional probability distribution*) περιγράφουν τον τρόπο με τον οποίο μεταβάλλεται η μία, έστω η X , από τις τυχαίες μεταβλητές για δεδομένη τιμή της άλλης τυχαίας μεταβλητής, έστω Y . Πιο συγκεκριμένα:

- Για X, Y **διακριτές** τυχαίες μεταβλητές

$$P_{X|Y}(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{P_{X,Y}(x, y)}{P_Y(y)}$$

- Για X, Y **συνεχείς** τυχαίες μεταβλητές

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Με ανάλογο τρόπο προκύπτει η δεσμευμένη κατανομή της τυχαίας μεταβλητής Y , δεδομένης της X .

5.7.2 Ανεξαρτησία τυχαίων μεταβλητών

Δύο τυχαίες μεταβλητές X και Y θα λέγονται **ανεξάρτητες** (*independent*) αν για κάθε ζεύγος συνόλων A και B πραγματικών αριθμών ισχύει:

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Ο παραπάνω ορισμός ανεξαρτησίας αποδεικνύεται (στηριζόμενοι στα τρία αξιώματα των πιθανοτήτων) ότι είναι ισοδύναμος με τις εξής σχέσεις:

- Για X, Y **διακριτές** τυχαίες μεταβλητές

$$P_{X,Y}(x,y) = P_X(x)P_Y(y), \quad \forall (x,y)$$

- Για X, Y **συνεχείς** τυχαίες μεταβλητές

$$f_{X,Y}(x,y) = f_X(x)f_Y(y), \quad \forall (x,y)$$

5.7.3 Περιγραφικά μέτρα της από κοινού κατανομής δύο τυχαίων μεταβλητών

Όπως και στην περίπτωση μια τυχαίας μεταβλητής, έτσι και η από κοινού κατανομή δύο τυχαίων μεταβλητών είναι δυνατό να συνοψιστεί με ορισμένα βασικά περιγραφικά μέτρα.

Στην περίπτωση του διμεταβλητού προβλήματος εισάγεται ένα νέο μέτρο το οποίο χρειάζεται για τη μελέτη της συμμεταβλητότητας του ζεύγους των τυχαίων μεταβλητών (X, Y) . Το μέτρο αυτό ονομάζεται **συνδιακύμανση** (*covariance*) αυτών και ορίζεται ως εξής:

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(X \cdot Y) - E(X) \cdot E(Y)$$

Η συνδιακύμανση έχει τις παρακάτω ιδιότητες:

- Αν X, Y ανεξάρτητες, $\text{Cov}(X, Y) = 0$
- $\text{Cov}(X + Z, Y) = \text{Cov}(X, Y) + \text{Cov}(Z, Y)$
- $\text{Cov}(a + bX, c + dY) = bd \text{Cov}(X, Y)$

Παρατηρούμε ότι $Cov(X, X) \equiv V(X)$

Όπως προκύπτει από τα όσα αναφέρθηκαν παραπάνω, η συνδιακύμανση δύο τυχαίων μεταβλητών εκφράζει ποσοτικά την ένταση της συμμεταβολής των τυχαίων μεταβλητών και έχει μονάδες μέτρησης το γινόμενο των μονάδων μέτρησης των δύο τυχαίων μεταβλητών. Αυτό σημαίνει, για παράδειγμα, ότι, αν μεταβληθούν οι μονάδες μέτρησης των τυχαίων μεταβλητών, θα μεταβληθεί και η τιμή της συνδιακύμανσής τους. Δημιουργείται, λοιπόν, η ανάγκη ορισμού ενός μέτρου συμμεταβολής, το οποίο δεν θα επηρεάζεται από τις μονάδες μέτρησης των τυχαίων μεταβλητών. Το πρόβλημα αυτό αντιμετωπίζεται στη συνέχεια.

Έστω δύο τυχαίες μεταβλητές X, Y που κατανέμονται με μέσες τιμές μ_X, μ_Y και τυπικές αποκλίσεις σ_X, σ_Y αντίστοιχα. Θεωρούμε επίσης τις τυχαίες μεταβλητές

$$Z_X = \frac{X - \mu_X}{\sigma_X} \text{ και } Z_Y = \frac{Y - \mu_Y}{\sigma_Y}, \text{ δηλαδή τις τυποποιημένες μορφές των αρχικών τυχαίων}$$

μεταβλητών. Ορίζουμε ως **συντελεστή συσχέτισης** (*coefficient of correlation*) των X, Y και συμβολίζουμε με $\rho(X, Y)$ τη συνδιακύμανση των τυποποιημένων μορφών τους. Δηλαδή

$$\rho(X, Y) = Cov(Z_X, Z_Y) = Cov\left(\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y}\right) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Να σημειωθεί ότι αν είτε $\sigma_X = 0$ είτε $\sigma_Y = 0$ ορίζουμε ότι $\rho(X, Y) = 0$. Οι τυχαίες μεταβλητές X, Y για τις οποίες $\rho(X, Y) = 0$ ονομάζονται **ασυσχέτιστες** (*uncorrelated*).

Ο συντελεστής συσχέτισης των τυχαίων μεταβλητών X, Y εκφράζει την ένταση της γραμμικής εξάρτησης των X, Y , έχει δε τις παρακάτω ιδιότητες:

$$i. \quad \rho(a + bX, c + dY) = \begin{cases} \rho(X, Y) & \text{αν } bd > 0 \\ -\rho(X, Y) & \text{αν } bd < 0 \end{cases}$$

$$ii. \quad |\rho(X, Y)| \leq 1$$

Όταν $|\rho(X, Y)| = 1$, έχουμε πλήρη γραμμική εξάρτηση. Τιμές του συντελεστή κοντά στο 1 αποτελούν ένδειξη για θετική γραμμική συσχέτιση μεταξύ των X και Y (δηλαδή, όσο αυξάνουν οι τιμές της μιας τυχαίας μεταβλητής, αυξάνουν και οι τιμές της άλλης), ενώ τιμές του συντελεστή κοντά στο -1 αποτελούν ένδειξη για αρνητική γραμμική συσχέτιση μεταξύ των X και Y (δηλαδή, όσο αυξάνουν οι τιμές της μιας τυχαίας μεταβλητής, μειώνονται οι τιμές της άλλης).

iii. Έστω δύο τυχαίες μεταβλητές X, Y για τις οποίες ισχύει ότι $Y = a + bX$. Τότε

$$\rho(X, Y) = \begin{cases} 1, & a > 0 \\ -1, & a < 0 \end{cases}$$

iv. Έστω δύο τυχαίες μεταβλητές X, Y για τις οποίες ισχύει ότι $\rho(X, Y) = \pm 1$. Τότε υπάρχουν $a, b \in R$ με $b \neq 0$ τέτοια ώστε $Y = a + bX$.

Ολοκληρώνοντας την παρουσίαση του συντελεστή συσχέτισης $\rho(X, Y)$, να τονίσουμε ότι:

- Ο συντελεστής συσχέτισης θα πρέπει να χρησιμοποιείται ως ένδειξη γραμμικής μόνο εξάρτησης και όχι ως ένδειξη οποιασδήποτε μορφής εξάρτησης.
- Αν δύο τυχαίες μεταβλητές X, Y είναι ανεξάρτητες, τότε θα είναι οπωσδήποτε και ασυσχέτιστες. Το αντίστροφο, όμως, δεν ισχύει. Δηλαδή, δύο τυχαίες μεταβλητές X, Y μπορεί να είναι ασυσχέτιστες χωρίς απαραίτητα να είναι και ανεξάρτητες.

5.7.4 Συναρτήσεις τυχαίων μεταβλητών

Τα μέτρα που ήδη ορίστηκαν είναι ιδιαίτερα χρήσιμα για τη συστηματική μελέτη της συμπεριφοράς (μέσω βασικών περιγραφικών μέτρων) των συναρτήσεων τυχαίων μεταβλητών. Ορισμένα όμως βασικά θεωρήματα, που βοηθούν στον υπολογισμό της αναμενόμενης τιμής και της διακύμανσης συναρτήσεων των τυχαίων μεταβλητών δίνονται στη συνέχεια.

Θεώρημα 5.4

Έστω $g(X, Y)$ μία συνάρτηση των τυχαίων μεταβλητών X και Y . Τότε αποδεικνύεται ότι

$$E(g(X, Y)) = \begin{cases} \sum_x \sum_y g(x, y) P_{X,Y}(x, y), & \text{αν } X, Y \text{ διακριτές} \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{X,Y}(x, y) dx dy, & \text{αν } X, Y \text{ συνεχείς} \end{cases}$$

Θεώρημα 5.5

Έστω $X_1, \dots, X_n$ τυχαίες μεταβλητές. Τότε η αναμενόμενη τιμή του αθροίσματος των τυχαίων μεταβλητών ισούται με το άθροισμα των αναμενόμενων τιμών τους. Δηλαδή

$$E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i)$$

Θεώρημα 5.6

Έστω $X_1, \dots, X_n$ **ανεξάρτητες** τυχαίες μεταβλητές. Τότε η αναμενόμενη τιμή του γινομένου των τυχαίων μεταβλητών ισούται με το γινόμενο των αναμενόμενων τιμών τους. Δηλαδή

$$E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i)$$

Θεώρημα 5.7

Έστω $X_1, \dots, X_n$ τυχαίες μεταβλητές. Τότε η διακύμανση του αθροίσματος των τυχαίων μεταβλητών δίνεται από τη σχέση

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) + 2 \sum_{\substack{i,j=1 \\ i < j}}^n \text{Cov}(X_i, X_j)$$

Πόρισμα 5.2

Έστω $X_1, \dots, X_n$ **ανεξάρτητες** τυχαίες μεταβλητές. Τότε η διακύμανση του αθροίσματος των τυχαίων μεταβλητών ισούται με το άθροισμα των διακυμάνσεών τους. Δηλαδή

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

ΣΥΝΟΨΗ

Με βάση τη φύση και, κατά συνέπεια, τον τρόπο έκφρασης των ενδεχομένων τους, οι δειγματικοί χώροι που συνδέονται με συγκεκριμένα τυχαία πειράματα κατατάσσονται σε δύο κατηγορίες, σε ποιοτικούς και ποσοτικούς. Η εφαρμογή των κανόνων του λογισμού των πιθανοτήτων, οι οποίοι παρουσιάστηκαν στο προηγούμενο κεφάλαιο, είναι ευκολότερη στην περίπτωση ποσοτικών δειγματικών χώρων. Σε πολλές πρακτικές εφαρμογές, όμως, τα δεδομένα είναι ποιοτικά και για το λόγο αυτόν θα ήταν χρήσιμο να οριστεί μία συνάρτηση που να αντιστοιχεί έναν πραγματικό αριθμό σε κάθε στοιχείο ενός ποιοτικού δειγματικού χώρου. Η συνάρτηση αυτή ονομάζεται τυχαία μεταβλητή και αποτέλεσε αντικείμενο της μελέτης του κεφαλαίου αυτού.

Στην πρώτη ενότητα ορίστηκε η έννοια της τυχαίας μεταβλητής, έγινε διάκριση μεταξύ διακριτών και συνεχών τυχαίων μεταβλητών και στη συνέχεια ορίστηκε, αφενός, η συνάρτηση πιθανότητας με το διάγραμμά της για τις διακριτές μεταβλητές και, αφετέρου, η συνάρτηση πυκνότητας πιθανότητας με το διάγραμμά της για τις συνεχείς τυχαίες μεταβλητές. Στη δεύτερη ενότητα ορίστηκε η αθροιστική συνάρτηση κατανομής για διακριτές και συνεχείς τυχαίες μεταβλητές. Στην τρίτη, στην τέταρτη και στην πέμπτη ενότητα παρουσιάστηκαν αντίστοιχα τα μέτρα κεντρικής τάσης, τα μέτρα διασποράς και οι ροπές τόσο για διακριτές όσο και για συνεχείς τυχαίες μεταβλητές. Τέλος, στην έκτη ενότητα έγινε μια σύντομη αναφορά στις διμεταβλητές κατανομές πιθανότητας.

ΑΣΚΗΣΕΙΣ

Άσκηση 5.1

Στρίβουμε ένα αμερόληπτο νόμισμα τέσσερις φορές και συμβολίζουμε με X την τυχαία μεταβλητή που εκφράζει το συνολικό αριθμό εμφανίσεων της ένδειξης «Κεφάλι» στις τέσσερις αυτές ρίψεις.

- Να καταγραφούν, υπό μορφή πίνακα, τα στοιχειώδη ενδεχόμενα του πειράματος αυτού και η τιμή της μεταβλητής X που αντιστοιχεί σε καθένα από αυτά.
- Να κατασκευαστούν ο πίνακας της κατανομής πιθανότητας της μεταβλητής X και το αντίστοιχο διάγραμμα πιθανότητας.

- iii. Να υπολογιστούν η μέση τιμή και η τυπική απόκλιση της μεταβλητής X .
- iv. Να κατασκευαστούν ο πίνακας της αθροιστικής κατανομής της μεταβλητής X και το αντίστοιχο διάγραμμα.

Άσκηση 5.2

Έστω X μια συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \begin{cases} ke^{-3x}, & \text{για } x > 0 \\ 0, & \text{αλλού.} \end{cases}$$

- i. Να υπολογιστεί η τιμή του k .
- ii. Να υπολογιστεί η πιθανότητα η τυχαία μεταβλητή X να λαμβάνει τιμές στο διάστημα $[0.5, 1]$.
- iii. Να υπολογιστεί η αθροιστική συνάρτηση κατανομής της X .

Άσκηση 5.3

Έστω X μια συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \begin{cases} \frac{1}{8}(x+1), & \text{για } 2 < x < 4 \\ 0, & \text{αλλού.} \end{cases}$$

Να υπολογιστούν η μέση τιμή, η διακύμανση και η τυπική απόκλιση της τυχαίας μεταβλητής X .

Άσκηση 5.4

Έστω X διακριτή τυχαία μεταβλητή που παίρνει τις τιμές 0, 1, 2 και 3 με πιθανότητες $\frac{1}{125}, \frac{12}{125}, \frac{48}{125}$ και $\frac{64}{125}$ αντίστοιχα. Να υπολογιστεί η μέση τιμή και η διακύμανσή της.

Άσκηση 5.5

Δίνεται η συνάρτηση $P(x) = \frac{x+2}{25}$ για $x = 1, 2, 3, 4, 5$. Να εξεταστεί αν η συνάρτηση αυτή είναι συνάρτηση πιθανότητας μιας διακριτής τυχαίας μεταβλητής.

Άσκηση 5.6

Έστω X μια συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \begin{cases} e^{-x}, & \text{για } x > 0 \\ 0, & \text{αλλού.} \end{cases}$$

Να υπολογιστεί η αναμενόμενη τιμή της τυχαίας μεταβλητής $g(X) = e^{3X/4}$

Άσκηση 5.7

Να εξετασθεί ποιοι από τους παρακάτω πίνακες αποτελούν πίνακες κατανομής πιθανότητας της διακριτής τυχαίας μεταβλητής X .

(α)

x	-2	-1	0	1	2
$P(x)$	0.1	0.2	0.4	0.2	0.1

(β)

x	0	1	2	3
$P(x)$	0.4	0.3	0.1	0.1

(γ)

x	-4	0	4	8
$P(x)$	0.2	1.2	-0.4	0

(δ)

x	-2	-1	0	1	2
$P(x)$	0.4	0.2	0.1	0.2	0.1

Άσκηση 5.8

Ένας ασφαλιστής σε κανονικές συνθήκες επισκέπτεται οκτώ υποψήφιους πελάτες την ημέρα. Έστω X η τυχαία μεταβλητή που εκφράζει τον αριθμό των ασφαλιστηρίων συμβολαίων που συνάπτει σε μία μέρα και έστω ότι η συνάρτηση πιθανότητας της X είναι:

$$P(x) = \frac{8!}{(8-x)!x!} (0.1)^x (0.9)^{8-x}; \quad x = 0, 1, \dots, 8$$

- (α) Να υπολογισθεί η πιθανότητα $P(x)$ για κάθε τιμή της μεταβλητής X και να απεικονιστεί γραφικά η κατανομή πιθανότητας.
- (β) Να υπολογισθεί η πιθανότητα ένας ασφαλιστής να συνάψει τουλάχιστον ένα ασφαλιστήριο συμβόλαιο σε μία τυχαία επιλεγμένη μέρα.
- (γ) Να υπολογισθεί η πιθανότητα ένας ασφαλιστής να συνάψει το πολύ ένα ασφαλιστήριο συμβόλαιο σε μία τυχαία επιλεγμένη μέρα.
- (δ) Να υπολογισθεί η πιθανότητα ένας ασφαλιστής να συνάψει ακριβώς ένα ασφαλιστήριο συμβόλαιο κατά τη διάρκεια μίας τυχαία επιλεγμένης μέρας.
- (ε) Να υπολογισθεί η αναμενόμενη τιμή του αριθμού των ασφαλιστηρίων συμβολαίων που συνάπτονται από τον ασφαλιστή σε μια μέρα.
- (στ) Να εκτιμηθεί η διακύμανση και η τυπική απόκλιση του αριθμού των ασφαλιστηρίων συμβολαίων που συνάπτονται από τον ασφαλιστή σε μια μέρα.

Άσκηση 5.9

Με βάση ιστορικά στοιχεία, η κατανομή της τυχαίας μεταβλητής X που εκφράζει τον αριθμό των πελατών που φτάνουν σε ένα κατάστημα μέσα σε ένα χρονικό διάστημα 5 λεπτών είναι η:

x	0	1	2	3	4	5	6	7	8	9	10
$P(x)$	0.01	0.04	0.10	0.12	0.16	0.20	0.17	0.08	0.07	0.04	0.01

- (α) Να απεικονιστεί γραφικά η κατανομή πιθανότητας της τυχαίας μεταβλητής X .
- (β) Να υπολογισθεί η αθροιστική κατανομή της τυχαίας μεταβλητής X .
- (γ) Να υπολογισθεί η πιθανότητα τουλάχιστον τρεις πελάτες να φτάσουν στο κατάστημα μέσα σε χρονικό διάστημα 5 λεπτών.

Άσκηση 5.10

Με βάση ιστορικά στοιχεία, η κατανομή της τυχαίας μεταβλητής X που εκφράζει τον αριθμό των υπεράριθμων κρατήσεων μίας αεροπορικής εταιρείας είναι η:

x	0	1	2	3	4	5	6
$P(x)$	0.05	0.10	0.20	0.25	0.20	0.15	0.05

- (α) Να απεικονιστεί γραφικά η κατανομή πιθανότητας της τυχαίας μεταβλητής X .
- (β) Να υπολογισθεί η αθροιστική κατανομή της τυχαίας μεταβλητής X .
- (γ) Να υπολογισθεί η πιθανότητα να γίνει κράτηση τουλάχιστον τεσσάρων υπεράριθμων θέσεων.
- (δ) Να υπολογισθεί η πιθανότητα να μη γίνει κράτηση υπεράριθμων θέσεων.
- (ε) Να υπολογισθεί η αναμενόμενη τιμή του αριθμού των υπεράριθμων κρατήσεων.
- (στ) Να υπολογισθεί η διακύμανση και η τυπική απόκλιση του αριθμού των υπεράριθμων κρατήσεων.

Άσκηση 5.11

Έστω ότι η συνάρτηση πυκνότητας πιθανότητας μιας συνεχούς τυχαίας μεταβλητής X είναι:

$$f(x) = \frac{x}{2} \quad \text{όπου} \quad 0 \leq x \leq 2$$

Να υπολογιστούν οι ακόλουθες πιθανότητες:

- (α) Η X να πάρει τιμές μεγαλύτερες από 1.

(β) Η X να πάρει τιμές μικρότερες από $\frac{1}{2}$

(γ) Η X να πάρει τιμές στο διάστημα $\left(\frac{1}{2}, 1\right)$

(δ) Η X να πάρει την τιμή 1

(ε) Η X να πάρει τιμές μεγαλύτερες από 3

Άσκηση 5.12

Έστω ότι η αθροιστική συνάρτηση κατανομής της τυχαίας μεταβλητής X είναι:

$$F(x) = \frac{x^2}{4} \quad \text{όπου} \quad 0 \leq x \leq 2$$

Να χρησιμοποιηθούν οι ιδιότητες της αθροιστικής συνάρτησης κατανομής για να απαντηθούν τα ερωτήματα της άσκησης 5.11. Τι παρατηρείτε; Αιτιολογείστε τα ευρήματά σας.

Άσκηση 5.13

Έστω ότι η συνάρτηση πυκνότητας πιθανότητας μιας συνεχούς τυχαίας μεταβλητής X είναι:

$$f(x) = \frac{1}{8} \quad \text{όπου} \quad -2 \leq x \leq 6$$

Να υπολογιστούν οι ακόλουθες πιθανότητες:

(α) Η X να πάρει τιμές μικρότερες από 0.

(β) Η X να πάρει τιμές μεγαλύτερες από $\frac{3}{2}$.

(γ) Η X να πάρει τιμές μεγαλύτερες από 4.

Άσκηση 5.14

Έστω ότι η αθροιστική συνάρτηση κατανομής της τυχαίας μεταβλητής X είναι:

$$F(x) = \frac{x+2}{8} \quad \text{όπου} \quad -2 \leq x \leq 6$$

Να χρησιμοποιηθούν οι ιδιότητες της αθροιστικής συνάρτησης κατανομής για να απαντηθούν τα ερωτήματα της άσκησης 5.12. Τι παρατηρείτε; Αιτιολογείστε τα ευρήματά σας.

Άσκηση 5.15

Έστω ότι η συνάρτηση πυκνότητας πιθανότητας μιας συνεχούς τυχαίας μεταβλητής X είναι:

$$f(x) = \frac{1}{2} \quad \text{όπου} \quad 0 \leq x \leq 2$$

Να υπολογιστούν οι ακόλουθες πιθανότητες:

(α) Η X να πάρει τιμές μεγαλύτερες από $\frac{3}{2}$

(β) Η X να πάρει τιμές μικρότερες από $\frac{1}{2}$

(γ) Η X να πάρει τιμές στο διάστημα $\left(\frac{1}{2}, \frac{3}{2}\right)$