# *Κεφάλαιο 2* Εισαγωγικές έννοιες θεωρίας πιθανοτήτων

## 2.1 Αξιωματική θεμελίωση της θεωρίας πιθανοτήτων

Η σύγχρονη προσέγγιση της θεωρίας πιθανοτήτων (ή πιθανοθεωρίας) βασίζεται στη θεωρία συνόλων και είναι πολύ απλή (βλ. Papoulis, 1965, 1990· Benjamin & Cornell, 1970· Taylor and Karlin, 1984· Cooper, 1989). Δομείται από τρεις θεμελιώδεις έννοιες και τρία αξιώματα. Οι θεμελιώδεις έννοιες είναι οι ακόλουθες:

1. Ο *δειγματικός χώρος* ορίζεται ως το σύνολο $\Omega$, του οποίου τα στοιχεία $\omega$ αντιστοιχούν στις δυνατές εκβάσεις ενός πειράματος.
2. Η *οικογένεια γεγονότων* ορίζεται ως μια συλλογή $\mathcal{F}$ υποσυνόλων (γεγονότων) $A$ του $\Omega$: Λέμε ότι το γεγονός $A$ *συμβαίνει* (ή *πραγματοποιείται*) όταν η έκβαση $\omega$ του πειράματος είναι στοιχείο του $A$.
3. Το *μέτρο πιθανότητας αποτελεί μια* συνάρτηση $P$ επί του $\mathcal{F}$. Μέσω αυτής, σε κάθε γεγονός $A$ αντιστοιχίζουμε έναν αριθμό $P(A)$ που λέγεται *πιθανότητα του γεγονότος A*.

Η τριάδα $(\Omega, \mathcal{F}, P)$ ονομάζεται *χώρος πιθανότητας*. Τα *αξιώματα της θεωρίας πιθανοτήτων* αποτελούν συνθήκες τις οποίες οφείλει να ικανοποιεί η συνάρτηση *P,* και είναι τα ακόλουθα:

$$
\begin{aligned}
&\text{I.} && P(A) \geq 0 \\
&\text{II.} && P(\Omega) = 1 \\
&\text{IIIα.} && P(A \cup B) = P(A) + P(B), \text{ με την προϋπόθεση ότι } A \cap B = \varnothing \\
&\text{IIIβ.} && P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i), \text{ με την προϋπόθεση ότι } A_i \cap A_j = \varnothing, i \neq j
\end{aligned}
\tag{2.1}
$$

Το αξίωμα IIIβ αποτελεί επέκταση του IIIα για άπειρο αριθμό γεγονότων, αλλά δεν είναι συνέπεια του IIIα , γι' αυτό και εισάγεται ως ανεξάρτητο αξίωμα.

Υπενθυμίζεται ότι η ένωση δύο γεγονότων (συμβολικά $A \cup B$, ή $A + B$, διαβάζεται "*A* ή *B*") είναι επίσης ένα γεγονός που πραγματοποιείται αν πραγματοποιηθεί ένα από τα δύο γεγονότα. Αντίστοιχα, η τομή δύο γεγονότων (συμβολικά $A \cap B$, ή $A\ B$, διαβάζεται "*A* και *B*") είναι ένα γεγονός που πραγματοποιείται αν πραγματοποιηθούν ταυτόχρονα και τα δύο γεγονότα. Ο δειγματικός χώρος, που σύμφωνα με το αξίωμα II έχει πιθανότητα ίση με ένα, λέγεται και *βέβαιο γεγονός*. Το κενό σύνολο ($\varnothing$) συμβολίζει το *αδύνατο γεγονός* και, όπως εύκολα προκύπτει από τα αξιώματα II και IIIα, έχει μηδενική πιθανότητα. Τέλος δύο γεγονότα που η τομή τους είναι το αδύνατο γεγονός (όπως τα *A* και *B* στο αξίωμα IIIα ) λέμε ότι είναι *αμοιβαία αποκλειόμενα*.

Στην περίπτωση που ο δειγματικός χώρος είναι αριθμήσιμο σύνολο, η οικογένεια γεγονότων $\mathcal{F}$ περιλαμβάνει όλα τα υποσύνολα του δειγματικού χώρου. Αν όμως ο δειγματικός χώρος είναι άπειρος μη αριθμήσιμος δεν είναι δυνατό να οριστεί μέτρο πιθανότητας σε κάθε υποσύνολο του Ω και γι' αυτό δεν είναι εφικτό να οριστεί η $\mathcal{F}$ ως το σύνολο όλων των υποσυνόλων του Ω. Το πιο χαρακτηριστικό παράδειγμα μη αριθμήσιμου απειροσυνόλου αποτελεί το σύνολο των πραγματικών αριθμών. Αποδεικνύεται ότι μπορούν να κατασκευαστούν υποσύνολά του τέτοια ώστε να μη μπορεί να οριστεί πιθανότητα σε αυτά και κατ' αυτή την έννοια δεν αποτελούν γεγονότα (πάντως η κατασκευή τους είναι πολύπλοκη). Βέβαια αυτού του είδους τα "παθολογικά" υποσύνολα δεν έχουν καμιά εφαρμογή και μπορούμε να τα αγνοήσουμε. Σε κάθε περίπτωση η οικογένεια $\mathcal{F}$ θα πρέπει να έχει τις ακόλουθες ιδιότητες:

$$
\begin{aligned}
&\text{α.} && \varnothing \in \mathcal{F} \\
&\text{β.} && \Omega \in \mathcal{F} \\
&\text{γ.} && (\forall A \in \mathcal{F})\ \left[\overline{A} \in \mathcal{F}\right], \text{ όπου } \overline{A} \equiv \{\omega \in \Omega, \omega \notin A\} \\
&\text{δ.} && (\forall A_i \in \mathcal{F}, i = 1,2,\ldots)\ \left[\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}\right]
\end{aligned}
\tag{2.2}
$$

Υπενθυμίζεται ότι το γεγονός $\overline{A}$ ονομάζεται *συμπλήρωμα* του $A$. Η συλλογή $\mathcal{F}$ που έχει τις παραπάνω ιδιότητες ονομάζεται σ-άλγεβρα.

## 2.2 Τυχαία μεταβλητή και συνάρτηση κατανομής

*Τυχαία μεταβλητή* είναι απλώς μια συνάρτηση ορισμένη επί του δειγματικού χώρου. Μέσω αυτής, σε κάθε δυνατή έκβαση $\omega$ αντιστοιχίζουμε, βάσει ενός προκαθορισμένου κανόνα, έναν αριθμό $X(\omega)$. Κάθε συνάρτηση μιας τυχαίας μεταβλητής είναι επίσης μια τυχαία μεταβλητή. Για λόγους απλοποίησης στο συμβολισμό μιας τυχαίας μεταβλητής παραλείπουμε το στοιχείο $\omega$ στο οποίο αντιστοιχεί και γράφουμε απλώς $X$. Επιπλέον χρησιμοποιούμε κεφαλαία ψηφία για την παράσταση της τυχαίας μεταβλητής και μικρά για την παράσταση αριθμητικών τιμών της. Για παράδειγμα γράφουμε $\{X \le x\}$ εννοώντας το γεγονός εκείνο (ήτοι υποσύνολο του δειγματικού χώρου) που αποτελείται από όλα τα στοιχεία $\omega$, τέτοια ώστε οι τιμές τους $X(\omega)$ να είναι μικρότερες του αριθμού $x$. Την πιθανότητα του γεγονότος αυτού τη συμβολίζουμε με $P(\{X \le x\})$ ή απλούστερα $P(X \le x)$.

*Συνάρτηση κατανομής* είναι η συνάρτηση της πραγματικής μεταβλητής $x$, που δίνεται από την εξίσωση

$$F_X(x) := P(X \le x) \tag{2.3}$$

και ορίζεται για κάθε πραγματικό αριθμό $x$, από $-\infty$ μέχρι $+\infty$. Σημειώνεται ότι η συνάρτηση κατανομής *δεν* είναι συνάρτηση της τυχαίας μεταβλητής (με την αυστηρώς μαθηματική έννοια) αλλά συνάρτηση συνδεδεμένη με την τυχαία μεταβλητή. Γι’ αυτό και η τυχαία μεταβλητή δεν είναι όρισμα της συνάρτησης αλλά δείκτης του συμβόλου της (ο δείκτης μπορεί πάντως να παραλείπεται αν δεν υπάρχει κίνδυνος παρανόησης). Επίσης, το πεδίο ορισμού της συνάρτησης δεν ταυτίζεται με το πεδίο τιμών της τυχαίας μεταβλητής αλλά είναι πάντα όλο το σύνολο των πραγματικών αριθμών. Είναι πάντα αύξουσα συνάρτηση, συνεχής από δεξιά, και υπακούει στη σχέση

$$0 = F_X(-\infty) \leq F_X(x) \leq F_X(+\infty) = 1 \tag{2.4}$$

Η συνάρτηση κατανομής συχνά αποκαλείται και *αθροιστική συνάρτηση κατανομής* ή και *πιθανότητα μη υπέρβασης*. Αντίστοιχα, λέγεται *πιθανότητα υπέρβασης* η παράσταση

$$F_{1_X}(x) := P(X > x) = 1 - F_X(x) \tag{2.5}$$

που είναι φθίνουσα συνάρτηση και υπακούει στη σχέση

$$1 = F_{1_X}(-\infty) \geq F_{1X}(x) \geq F_{1X}(+\infty) = 0 \tag{2.6}$$

Η τυχαία μεταβλητή $X$ λέγεται *συνεχής* αν η συνάρτηση κατανομής $F_X(x)$ είναι συνεχής συνάρτηση για κάθε $x$. Στην περίπτωση αυτή ο δειγματικός χώρος είναι άπειρο μη αριθμήσιμο σύνολο. Η $X$ λέγεται διακριτή μεταβλητή αν η συνάρτηση κατανομής της είναι κλιμακωτή. Στην περίπτωση αυτή ο δειγματικός χώρος είναι πεπερασμένο σύνολο ή άπειρο αλλά αριθμήσιμο. Τέλος, η $X$ λέγεται μικτή μεταβλητή αν η συνάρτηση κατανομής της είναι ασυνεχής χωρίς όμως να είναι κλιμακωτή.

Η παράγωγος της συνάρτησης κατανομής, ήτοι η

$$f_X(x) := \frac{dF(x)}{dx} \tag{2.7}$$

λέγεται *συνάρτηση πυκνότητας πιθανότητας*. Για συνεχείς τυχαίες μεταβλητές η συνάρτηση αυτή ορίζεται παντού, δε συμβαίνει όμως το ίδιο στην περίπτωση των διακριτών μεταβλητών (πάντως και στην τελευταία περίπτωση μπορεί να οριστεί η συνάρτηση πυκνότητας πιθανότητας με χρήση συναρτήσεων $\delta$ του Dirac). Οι βασικές της ιδιότητες που προκύπτουν άμεσα από τον ορισμό της είναι

$$f_X(x) \geq 0, \qquad \int_{-\infty}^{\infty} f_X(x)dx = 1 \tag{2.8}$$

Είναι προφανές ότι η συνάρτηση πυκνότητας πιθανότητας δεν παριστάνει πιθανότητα και γι' αυτό μπορεί να πάρει τιμές μεγαλύτερες της μονάδας. Η σχέση της με την πιθανότητα προσδιορίζεται από την εξίσωση

$$f_X(x) = \lim_{\Delta x \to 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x} \tag{2.9}$$

Η συνάρτηση κατανομής υπολογίζεται από τη συνάρτηση πυκνότητας πιθανότητας σύμφωνα με την ακόλουθη σχέση, αντίστροφη της (2.7)

$$F_X(x) = \int_{-\infty}^{x} f_X(\xi)d\xi \tag{2.10}$$

Για συνεχείς τυχαίες μεταβλητές υπάρχει η αντίστροφη συνάρτηση $F_X^{-1}(u)$ της $F_X(x)$ . Κατά συνέπεια η εξίσωση $u = F_X(x)$ έχει μία μοναδική λύση ως προς $x$, την $x_u = F_X^{-1}(u)$. Η τιμή $x_u$, η οποία αντιστοιχεί σε δεδομένη τιμή $u$ της συνάρτησης κατανομής, λέγεται *u-ποσοστημόριο* της μεταβλητής $X$.

**Εφαρμογή 2.2**

Για καλύτερη κατανόηση των βασικών εννοιών της θεωρίας πιθανοτήτων δίνουμε το ακόλουθο παράδειγμα από το χώρο της υδρολογίας. Σε ένα συγκεκριμένο τόπο και μια συγκεκριμένη περίοδο του έτους μάς ενδιαφέρει η μαθηματική περιγραφή του ενδεχομένου αν βρέχει κατά τη διάρκεια μιας μέρας ή όχι. Θεωρούμε το φυσικό αυτό φαινόμενο ως ισοδύναμο με ένα πείραμα τύχης με δύο δυνατές εκβάσεις: βροχερή μέρα (συμβολικά Β) και στεγνή μέρα (συμβολικά Σ).

Ο δειγματικός χώρος είναι βέβαια πεπερασμένος, αφού περιλαμβάνει δύο μόνο στοιχεία. Συγκεκριμένα

$$\Omega = \{\mathrm{B}, \Sigma\}$$

Αντίστοιχα, η οικογένεια γεγονότων περιλαμβάνει όλα τα υποσύνολα του δειγματικού χώρου, άρα

$$F = \{\varnothing, \{\mathrm{B}\}, \{\Sigma\}, \Omega\}$$

Για να ορίσουμε την πιθανότητα επί του $\mathcal{F}$ φτάνει να ορίσουμε την πιθανότητα ενός από τα δύο ενδεχόμενα, έστω την $P$(Β). Για να γίνει αυτό με αξιόπιστο τρόπο χρειάζεται να έχουμε στη διάθεσή μας ιστορικά δεδομένα και να χρησιμοποιήσουμε τις αρχές της στατιστικής που θα συζητηθούν παρακάτω. Για την ώρα, ας θεωρήσουμε αυθαίρετα ότι $P$(Β) = 0.2. Οι υπόλοιπες πιθανότητες προκύπτουν με βάση τα αξιώματα. Έτσι έχουμε $P(\varnothing)$ = 0 και $P$(Ω) = 1. Δεδομένου ότι τα Β και Σ είναι αμοιβαία αποκλειόμενα ισχύει $P$(Β) + $P$(Σ) = $P$(Β€+€Σ) = $P$(Ω) = 1, άρα $P$(Σ) = 0.8.

Ας ορίσουμε την τυχαία μεταβλητή $X$ με τον ακόλουθο τρόπο

$$X(\Sigma) = 0, \qquad X(\mathrm{B}) = 1$$

Μπορούμε τώρα να υπολογίσουμε εύκολα τη συνάρτηση κατανομής της $X$. Για $x < 0$ έχουμε $F_X(x) = P(X \le x) = 0$ (δεδομένου ότι η $X$, όπως την ορίσαμε, δεν μπορεί να πάρει αρνητικές τιμές). Για $0 \le x < 1$ έχουμε

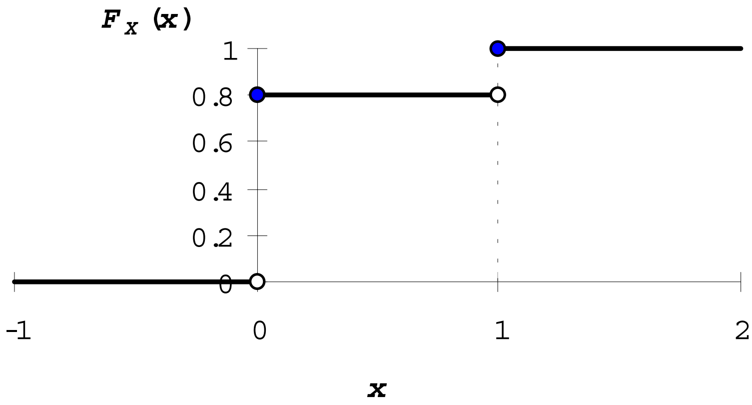
$$F_X(x) = P(X \le x) = P(X = 0) = 0.8$$

Τέλος για $1 \le x$ έχουμε

$$F_X(x) = P(X \le x) = P(X = 0) + P(X = 1) = 1$$

Η γραφική παράσταση της συνάρτησης κατανομής φαίνεται στο Σχ. 2.1. Το κλιμακωτό σχήμα χαρακτηρίζει το γεγονός ότι η τυχαία μεταβλητή είναι διακριτή.

Κλείνουμε την εφαρμογή αυτή με δύο παρατηρήσεις. Πρώτον, η πιθανότητα βροχερής ημέρας και κατά συνέπεια και η συνάρτηση κατανομής εξαρτάται από τη συγκεκριμένη περίοδο του έτους (π.χ. μήνας), καθώς επίσης και από το συγκεκριμένο τόπο, στον οποίο αναφερόμαστε. Δεύτερον, στο συγκεκριμένο πιθανοθεωρητικό μοντέλο που κατασκευάσαμε δε γίνεται καθόλου αναφορά στην αλληλουχία βροχερών-στεγνών ημερών. Αυτό βέβαια δεν είναι λάθος, απλώς περιορίζει την προγνωστική αξία του μοντέλου. Ένα πληρέστερο μοντέλο θα περιέγραφε π.χ. ξεχωριστά την πιθανότητα μιας στεγνής μέρας που ακολουθεί αμέσως μετά από επίσης στεγνή μέρα, η οποία, όπως εμπειρικά γνωρίζουμε, είναι αυξημένη σε σχέση με την πιθανότητα στεγνής μέρας ύστερα από βροχερή μέρα. Τέτοιου είδους μοντέλα, που λαμβάνουν υπόψη τη χρονική αλληλουχία γεγονότων, στηρίζονται στη θεωρία των *στοχαστικών ανελίξεων* και ανήκουν στον ιδιαίτερο κλάδο της *στοχαστικής υδρολογίας*, η ανάπτυξη του οποίου (πέρα από μια εισαγωγή που δίνεται στην ενότητα 2.7 και στο κεφάλαιο 4) ξεφεύγει από το σκοπό αυτού του κειμένου.




**Σχ. 2.1** Συνάρτηση κατανομής της εφαρμογής 2.2.

## 2.3 Ανεξάρτητα και εξαρτημένα γεγονότα, δεσμευμένη πιθανότητα

Δύο γεγονότα $A$ και $B$ λέγονται *ανεξάρτητα* (ή *στοχαστικώς ανεξάρτητα*), αν

$$P(A \cap B) = P(A)P(B) \tag{2.11}$$

Σε αντίθετη περίπτωση τα γεγονότα λέγονται στοχαστικώς εξαρτημένα. Ο ορισμός της στοχαστικής ανεξαρτησίας επεκτείνεται και για περισσότερα γεγονότα $A_1$, $A_2$, …, τα οποία λέγονται ανεξάρτητα αν ισχύει

$$P\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_n}\right) = P\left(A_{i_1}\right)P\left(A_{i_2}\right)\cdots P\left(A_{i_n}\right) \tag{2.12}$$

για κάθε πεπερασμένο σύνολο διακεκριμένων δεικτών $i_1$, $i_2$, …, $i_n$.

Ως *δεσμευμένη πιθανότητα* γεγονότος $A$ δεδομένου ότι έχει εμφανιστεί το γεγονός $B$, συμβολικά $P(A|B)$, ορίζεται ο λόγος

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \tag{2.13}$$

Προφανώς, στην περίπτωση που $P(B) = 0$ δεν ορίζεται η δεσμευμένη πιθανότητα $P(A|B)$. Συνέπεια της (2.13) είναι η

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) \tag{2.14}$$

Άμεσα προκύπτει από τις εξισώσεις (2.11) και (2.13) ότι για ανεξάρτητα γεγονότα $A$ και $B$ ισχύει $P(A|B) = P(A)$, οπότε και η (2.14) μεταπίπτει στην (2.11).

**Εφαρμογή 2.3**

*α. Να υπολογιστεί η πιθανότητα να εμφανιστεί επί δύο διαδοχικά χρόνια πλημμύρα μεγαλύτερη ή ίση αυτής που αντιστοιχεί σε πιθανότητα υπέρβασης $F_1$ = 20%.*

Έστω $X$ η τυχαία μεταβλητή που περιγράφει τη μέγιστη ετήσια πλημμυρική παροχή και $x_{0.8}$ το 0.8-ποσοστημόριό της, δηλαδή η τιμή της που αντιστοιχεί στην τιμή 0.8 της συνάρτησης κατανομής, ή, ισοδύναμα, σε πιθανότητα υπέρβασης 0.2. Συμβολίζουμε με $A$ το γεγονός να έχουμε σε ένα έτος πλημμύρα μεγαλύτερη ή ίση της τιμής $x_{0.8}$, ήτοι $A = \{X \geq x_{0.8}\}$ και με $\overline{A}$ το συμπληρωματικό του, δηλαδή το να έρθει πλημμύρα μικρότερη της $x_{0.8}$. Οι πιθανότητες των δύο αυτών συμπληρωματικών γεγονότων είναι

$$p \equiv P(A) = P(X \geq x_{0.8}) = F_{1_X}(x_{0.8}) = 0.2$$

και

$$q \equiv P(\overline{A}) = 1 - p = 0.8$$

Αν αναφερθούμε στο πρόβλημα των δύο διαδοχικών ετών, ο δειγματικός χώρος είναι

$$\Omega = \left\{A_1 A_2, \overline{A}_1 A_2, A_1 \overline{A}_2, \overline{A}_1 \overline{A}_2\right\}$$

όπου οι δείκτες 1 και 2 αντιστοιχούν στο πρώτο και δεύτερο έτος, αντίστοιχα. Με την παραδοχή ότι η εμφάνιση ή όχι της συγκεκριμένης πλημμύρας στο πρώτο έτος είναι στοχαστικά ανεξάρτητη από την εμφάνιση ή όχι στο δεύτερο έτος (η παραδοχή αυτή είναι απολύτως λογική, οι μέγιστες πλημμύρες των δύο ετών προκύπτουν ως συνέπειες τελείως διαφορετικών υδρομετεωρολογικών φαινομένων) η ζητούμενη πιθανότητα θα είναι

$$P_1 = (A_1 A_2) = P(A_1)P(A_2) = p^2 = 0.04$$

Για λόγους πληρότητας υπολογίζουμε ομοίως και τις πιθανότητες των υπόλοιπων ενδεχομένων, που είναι:

$$P\left(\overline{A}_1 A_2\right) = P\left(A_1 \overline{A}_2\right) = pq = 0.16, \qquad P\left(\overline{A}_1 \overline{A}_2\right) = q^2 = 0.64$$

Όπως αναμενόταν, το άθροισμα των πιθανοτήτων των τεσσάρων ενδεχομένων είναι 1.

*β. Να υπολογιστεί η πιθανότητα να εμφανιστεί επί δύο διαδοχικά χρόνια πλημμύρα μεγαλύτερη ή ίση αυτής που αντιστοιχεί σε $F_1$ = 20%, αν είναι γνωστό ότι σε ένα από τα δύο χρόνια έχει εμφανιστεί τέτοια πλημμύρα.*

Το να ξέρουμε ότι σε ένα από τα δύο χρόνια έχει εμφανιστεί πλημμύρα μεγέθους $x_{0.8}$ ή μεγαλύτερου ισοδυναμεί με το να ξέρουμε ότι έχει πραγματοποιηθεί το σύνθετο γεγονός $A_1A_2 \cup \overline{A}_1 A_2 \cup A_1 \overline{A}_2$ (έχει αποκλειστεί μόνο το $\overline{A}_1\overline{A}_2$). Άρα η πιθανότητα που ζητούμε είναι η $P_2 = P(A_1A_2 | A_1A_2 \cup \overline{A}_1 A_2 \cup A_1 \overline{A}_2)$, η οποία σύμφωνα με τον ορισμό της δεσμευμένης πιθανότητας είναι

$$P_2 = \frac{P\left(A_1A_2 \cap \left(A_1A_2 \cup \overline{A}_1 A_2 \cup A_1 \overline{A}_2\right)\right)}{P\left(A_1A_2 \cup \overline{A}_1 A_2 \cup A_1 \overline{A}_2\right)}$$

και, παίρνοντας υπόψη ότι όλοι οι συνδυασμοί γεγονότων είναι αμοιβαία αποκλειόμενοι, γίνεται

$$P_2 = \frac{P\left(A_1A_2\right)}{P\left(A_1A_2\right) + P\left(\overline{A}_1 A_2\right) + P\left(A_1 \overline{A}_2\right)} = \frac{p^2}{p^2 + 2pq} = \frac{p}{p + 2q} = 0.111\ldots$$

*γ. Να υπολογιστεί η πιθανότητα να εμφανιστεί επί δύο διαδοχικά χρόνια πλημμύρα μεγαλύτερη ή ίση αυτής που αντιστοιχεί σε $F_1$ = 20%, αν είναι γνωστό ότι στο πρώτο από τα δύο χρόνια έχει εμφανιστεί τέτοια πλημμύρα.*

Αν και φαίνεται ότι το ερώτημα αυτό είναι ταυτόσημο με το προηγούμενο, στην πραγματικότητα δεν είναι.* Στο προηγούμενο ερώτημα είχαμε μια δεδομένη πληροφορία που αφορούσε κάποιο από τα δύο έτη, χωρίς να ξέρουμε ποιο, ενώ τώρα η πληροφορία αυτή συγκε-

* Το πρόβλημα αυτό είναι ισοδύναμο με το λεγόμενο “παράδοξο των παπαγάλων”, το οποίο ο ενδιαφερόμενος αναγνώστης μπορεί να βρει μαζί με άλλα ενδιαφέροντα μαθηματικά παράδοξα στον Gardner (1982).

κριμενοποιείται στο πρώτο έτος (το ίδιο θα ήταν αν η πληροφορία αναφερόταν στο δεύτερο). Η συγκεκριμενοποίηση της δεδομένης πληροφορίας μεταβάλλει τις πιθανότητες, όπως θα επιβεβαιώσουμε αμέσως.

Αυτό που τώρα γνωρίζουμε είναι ότι έχει πραγματοποιηθεί το σύνθετο γεγονός $A_1A_2 \cup A_1\overline{A}_2$ (έχουν αποκλειστεί τα $\overline{A}_1A_2$ και $\overline{A}_1\overline{A}_2$). Κατά συνέπεια η ζητούμενη πιθανότητα τώρα είναι η $P_3 = P(A_1A_2 | A_1A_2 \cup A_1\overline{A}_2)$, που σύμφωνα με τον ορισμό της δεσμευμένης πιθανότητας είναι

$$P_3 = \frac{P\left(A_1A_2 \cap \left(A_1A_2 \cup A_1\overline{A}_2\right)\right)}{P\left(A_1A_2 \cup A_1\overline{A}_2\right)}$$

και γίνεται

$$P_3 = \frac{P(A_1A_2)}{P(A_1A_2) + P\left(A_1\overline{A}_2\right)} = \frac{p^2}{p^2 + pq} = \frac{p}{p+q} = p = 0.2$$

## 2.4 Αναμενόμενες τιμές και παράμετροι κατανομών

Αν $X$ είναι συνεχής τυχαία μεταβλητή και $g(X)$ είναι συνάρτηση της $X$, τότε ορίζεται ως *αναμενόμενη τιμή* ή *προσδοκία της* $g(X)$ το μέγεθος

$$E[g(X)] := \int_{-\infty}^{\infty} g(x) f_X(x) dx \tag{2.15}$$

Η αντίστοιχη σχέση για διακριτή τυχαία μεταβλητή $X$, που παίρνει τις τιμές $x_1$, $x_2$, …, είναι

$$E[g(X)] := \sum_{i=1}^{\infty} g(x_i) P(X = x_i) \tag{2.16}$$

Ειδικότερα:

1. Για $g(X) = X^r$, όπου $r = 0, 1, 2, \ldots$, το μέγεθος

$$m_X^{(r)} := E\left[X^r\right] \tag{2.17}$$

ονομάζεται *ροπή περί την αρχή* (ή απλώς *ροπή*)*τάξης r* της $X$.

2. Για $g(X) = X$, το μέγεθος

$$m_X := E[X] \tag{2.18}$$

(δηλαδή η ροπή τάξης 1) ονομάζεται *αναμενόμενη τιμή* ή *μέση τιμή* της $X$. Εναλλακτικά για τη μέση τιμή χρησιμοποιείται και το σύμβολο $\mu_X$.

3. Για $g(X) = (X - m_X)^r$, το μέγεθος

$$\mu_X^{(r)} := E\left[(X - m_X)^r\right] \tag{2.19}$$

ονομάζεται *κεντρική ροπή τάξης r* της $X$.

4. Για $g(X) = (X - m_X)^2$, το μέγεθος

$$\sigma_X^2 := \mu_X^{(2)} = E\left[(X - m_X)^2\right] \tag{2.20}$$

(δηλαδή η κεντρική ροπή τάξης 2) ονομάζεται *διασπορά* της $X$. Η διασπορά συμβολίζεται ακόμη και με $\mathrm{Var}[X]$.

Σημειώνεται ότι η μηδενική ροπή, περί την αρχή ή κεντρική, είναι πάντα ίση με ένα ενώ η πρώτη κεντρική ροπή είναι μηδέν. Αν δοθούν οι ροπές περί την αρχή μιας μεταβλητής μπορούν να υπολογιστούν οι κεντρικές ροπές από τις εξισώσεις που δίνονται στο Παράρτημα 2.Α (σ. 45). Στα παραπάνω μεγέθη μπορεί να παραλείπεται ο δείκτης $X$ αν δεν υπάρχει κίνδυνος παρανόησης.

Οι ροπές περί την αρχή και οι κεντρικές ροπές των τεσσάρων πρώτων τάξεων χρησιμοποιούνται πολύ συχνά στην τεχνική υδρολογία (και σε άλλες εφαρμογές) δεδομένου ότι έχουν συγκεκριμένο νόημα. Συγκεκριμένα περιγράφουν, κατά σειρά, τη *θέση*, τη *διασπορά*, την *ασυμμετρία* και την *κύρτωση* της κατανομής, όπως εξηγείται παρακάτω. Εναλλακτικά χρησιμοποιούνται και άλλες παράμετροι με αντίστοιχο νόημα, που επίσης εξηγούνται παρακάτω.

### 2.4.1 Παράμετροι θέσης

Η μέση τιμή ουσιαστικά περιγράφει τη θέση του κέντρου βάρους του σχήματος που ορίζει η συνάρτηση πυκνότητας πιθανότητας με τον οριζόντιο άξονα (Σχ. 2.2α). Είναι επίσης ισοδύναμη με τη στατική ροπή περί τον κατακόρυφο άξονα του εν λόγω γεωμετρικού σχήματος (δεδομένου ότι το εμβαδό του σχήματος είναι ίσο με 1). Συχνά χρησιμοποιούνται εναλλακτικά και οι ακόλουθοι τύποι παραμέτρων θέσης:

1. Η *πιθανότερη τιμή* ή *κορυφή*, συμβολικά $x_p$, είναι η τιμή της μεταβλητής $x$ για την οποία η $f_X(x)$ γίνεται μέγιστη, προκειμένου για συνεχείς μεταβλητές, ή που έχει τη μεγαλύτερη πιθανότητα να συμβεί, αν η μεταβλητή είναι διακριτή. Αν η $f_X(x)$ παρουσιάζει ένα, δύο, κτλ., μέγιστα, λέμε ότι η κατανομή είναι *μονοκόρυφη*, *δικόρυφη*, κτλ., αντίστοιχα.

2. Η *διάμεσος*, συμβολικά $x_{0.5}$, είναι εκείνη η τιμή της μεταβλητής για την οποία ισχύει $P(X \leq x_{0.5}) = P(X \geq x_{0.5}) = 1/2$.* Έτσι, η διάμεσος αντιστοιχεί στο σημείο που χωρίζει την καμπύλη πυκνότητας πιθανότητας σε δύο ισοδύναμα τμήματα με εμβαδό 1/2.

Γενικά η μέση τιμή, η κορυφή και η διάμεσος δεν ταυτίζονται, εκτός αν η κατανομή είναι συμμετρική και μονοκόρυφη.

### 2.4.2 Παράμετροι διασποράς

Η διασπορά μιας τυχαίας μεταβλητής δείχνει το μέγεθος της συγκέντρωσης της πυκνότητας πιθανότητας γύρω από τη μέση τιμή. Έτσι, μικρή διασπορά δείχνει συγκεντρωμένη κατανομή (Σχ. 2.2β). Η οριακή τιμή μηδενικής διασποράς αντιστοιχεί σε μεταβλητή που παίρνει μία μόνο τιμή με πλήρη βεβαιότητα (αρνητική τιμή της διασποράς είναι αδύνατη). Το γεωμετρικό αντίστοιχο της διασποράς είναι η ροπή αδρανείας περί τον κατακόρυφο κεντροβαρικό άξονα του σχήματος που ορίζει η συνάρτηση πυκνότητας πιθανότητας με τον οριζόντιο άξονα.

Η τετραγωνική ρίζα της διασποράς,

---

* Αυστηρά η σχέση αυτή ισχύει μόνο για συνεχείς μεταβλητές. Κατ’ αναλογία ορίζεται η διάμεσος σε διακριτές μεταβλητές.

$$\sigma_X = \sqrt{\mathrm{Var}[X]} \tag{2.21}$$

που έχει διαστάσεις ίδιες με την τυχαία μεταβλητή, λέγεται *τυπική απόκλιση*. Η αδιάστατη παράμετρος

$$C_{v_X} := \frac{\sigma_X}{m_X} \tag{2.22}$$

λέγεται *συντελεστής μεταβλητότητας*.

Εναλλακτική παράμετρος διασποράς, που πάντως δεν χρησιμοποιείται πολύ συχνά, είναι το *κεντρικό ή διατεταρτημοριακό πλάτος* που ορίζεται ως η διαφορά $x_{0.75} - x_{0.25}$, δηλαδή η διαφορά των τιμών της μεταβλητής που αντιστοιχούν στα ποσοστημόρια 0.75 και 0.25. Οι τιμές αυτές ($x_{0.75}$ και $x_{0.25}$) ονομάζονται *άνω* και *κάτω τεταρτημόριο*, αντίστοιχα. Το εμβαδό της συνάρτησης πυκνότητας πιθανότητας ανάμεσα σε αυτές τις δύο θέσεις είναι ίσο με 0.5.

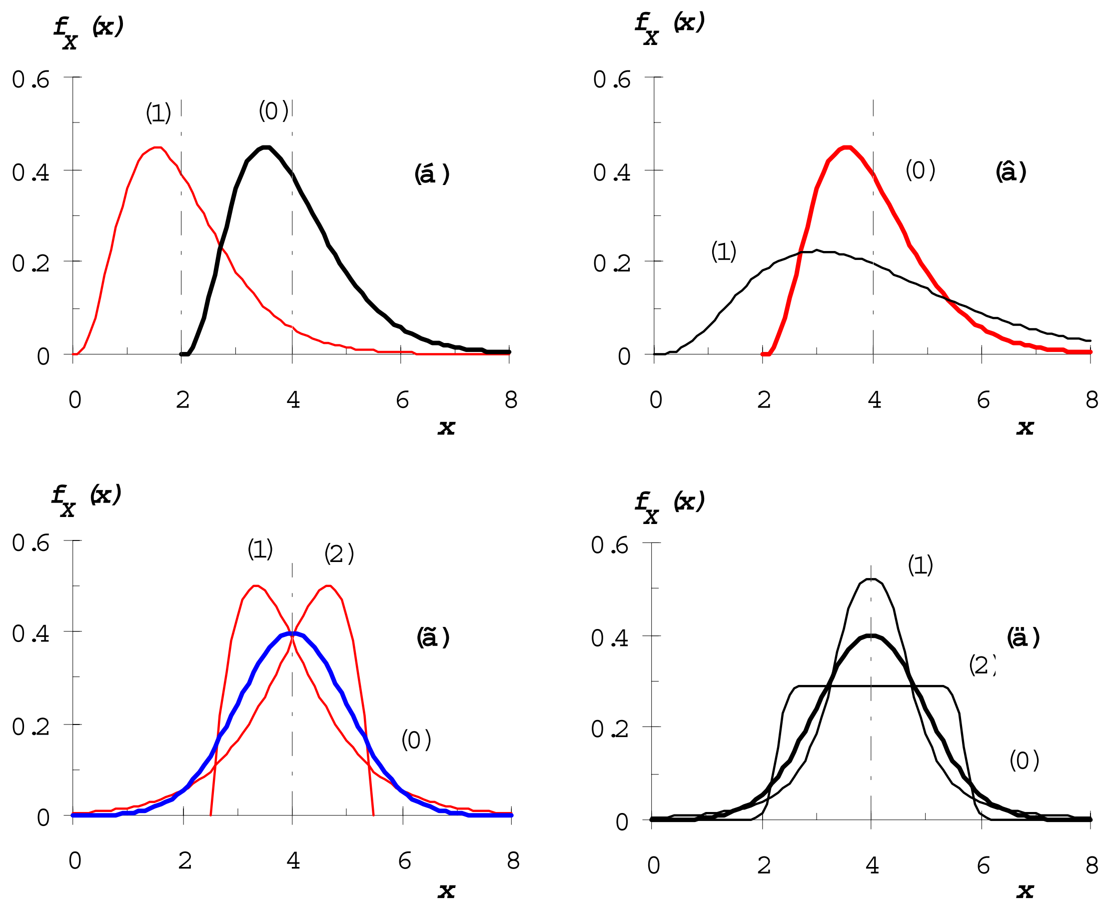
### 2.4.3 Παράμετροι ασυμμετρίας

Η τρίτη κεντρική ροπή χρησιμοποιείται για την περιγραφή της ασυμμετρίας της κατανομής. Μηδενική τιμή της τρίτης ροπής δείχνει συμμετρική κατανομή. Αν η τρίτη ροπή είναι μεγαλύτερη ή μικρότερη από το μηδέν λέμε ότι η κατανομή είναι θετικά ασύμμετρη ή αρνητικά ασύμμετρη, αντίστοιχα (Σχ. 2.2γ). Στην περίπτωση θετικά ασύμμετρης μονοκόρυφης κατανομής ισχύει $x_{\mathrm{p}} \le x_{0.5} \le m_X$ ενώ η ανάστροφη ανισότητα ισχύει για αρνητικά ασύμμετρη κατανομή. Αποτελεσματικότερο μέτρο της ασυμμετρίας της κατανομής είναι ο αδιάστατος *συντελεστής ασυμμετρίας* που ορίζεται από τη σχέση

$$C_{s_X} := \frac{\mu_X^{(3)}}{\sigma_X^3} \tag{2.23}$$

### 2.4.4 Παράμετροι κύρτωσης

Ο όρος *κύρτωση* περιγράφει το πόσο "αιχμηρή" ή όχι είναι η συνάρτηση πυκνότητας πιθανότητας γύρω από την κορυφή της. Μέτρο αυτής της ιδιότητας είναι η τέταρτη κεντρική ροπή, ή καλύτερα ο αδιάστατος *συντελεστής κύρτωσης* που ορίζεται μέσω της τελευταίας από τη σχέση




**Σχ. 2.2** Σχέση χαρακτηριστικών παραμέτρων κατανομής και σχήματος συνάρτησης πυκνότητας πιθανότητας: *(α) Επίδραση της μέσης τιμής*. Οι καμπύλες (0) και (1) έχουν μέσες τιμές 4 και 2, αντίστοιχα, ενώ και οι δύο έχουν τυπική απόκλιση 1, συντελεστή ασυμμετρίας 1 και συντελεστή κύρτωσης 4.5. *(β) Επίδραση της τυπικής απόκλισης*. Οι καμπύλες (0) και (1) έχουν τυπική απόκλιση 1 και 2 αντίστοιχα, ενώ και οι δύο έχουν μέση τιμή 4, συντελεστή ασυμμετρίας 1 και συντελεστή κύρτωσης 4.5. *(γ) Επίδραση του συντελεστή ασυμμετρίας*. Οι καμπύλες (0), (1) και (2) έχουν συντελεστή ασυμμετρίας 0, +1.33 και -1.33 αντίστοιχα, ενώ και οι τρεις έχουν μέση τιμή 4 και τυπική απόκλιση 1 (ο συντελεστής κύρτωσης είναι 3 για την καμπύλη (0) και 5.67 για τις καμπύλες (1) και (2)). *(δ) Επίδραση του συντελεστή κύρτωσης*. Οι καμπύλες (0), (1) και (2) έχουν συντελεστή κύρτωσης 3, 5 και 2, αντίστοιχα, ενώ και οι τρεις έχουν μέση τιμή 4, τυπική απόκλιση 1 και συντελεστή ασυμμετρίας 0.

$$C_{k_X} := \frac{\mu_X^{(4)}}{\sigma_X^4} \qquad (2.24)$$

Ο συντελεστής αυτός συνήθως συγκρίνεται με την τιμή 3, η οποία αντιστοιχεί στην κανονική κατανομή (βλ. εδάφιο 2.8.2). Τιμή του συντελεστή μεγαλύτερη του 3 αντιστοιχεί σε λεπτόκυρτη (αιχμηρή) κατανομή ενώ τιμή μικρότερη του 3 αντιστοιχεί σε πλατύκυρτη (επίπεδη) κατανομή (Σχ. 2.2δ).

**Εφαρμογή 2.4**

*Κατά τις μέρες που βρέχει το ημερήσιο ύψος βροχής, X, μετρημένο σε συγκεκριμένο σταθμό και εκφρασμένο σε mm, βρέθηκε ότι ακολουθεί την εκθετική κατανομή με συνάρτηση κατανομής*

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

*όπου λ= 0.05 mm$^{-1}$.* Ζητείται ο υπολογισμός των παραμέτρων θέσης, διασποράς, ασυμμετρίας και κύρτωσης της κατανομής.*

Παραγωγίζοντας τη συνάρτηση κατανομής υπολογίζουμε τη συνάρτηση πυκνότητας πιθανότητας

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

Προκειμένου να υπολογίσουμε τη μέση τιμή, εφαρμόζουμε την εξίσωση (2.15) για $g(X) = X$ και έχουμε

$$m_X = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} \lambda x e^{-\lambda x} dx$$

και μετά τις πράξεις

$$m_X = 1/\lambda = 20 \text{ mm}$$

Με παρόμοιο τρόπο βρίσκουμε για τυχόν $r \geq 0$

$$m_X^{(r)} = E\left[X^r\right] = r!/\lambda^r$$

και τελικά, εφαρμόζοντας τις εξισώσεις (2.68)

$$\sigma_X^2 = 1/\lambda^2 = 400 \text{ mm}^2, \quad \mu_X^{(3)} = 2/\lambda^3 = 16\,000 \text{ mm}^3$$

$$\mu_X^{(4)} = 9/\lambda^4 = 1\,440\,000 \text{ mm}^4$$

---

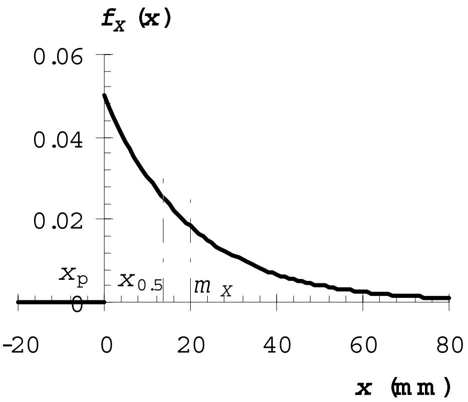
* Επειδή η τυχαία μεταβλητή στο παράδειγμα ορίζεται μόνο για τις μέρες που βρέχει, η κατανομή της είναι συνεχής παντού. Αν οριζόταν για όλες τις μέρες, και τις μη βροχερές, τότε η τιμή $X = 0$ θα είχε συγκεκριμένη πιθανότητα διαφορετική από το μηδέν, και έτσι η συνάρτηση κατανομής θα είχε άλμα στη θέση $x = 0$.

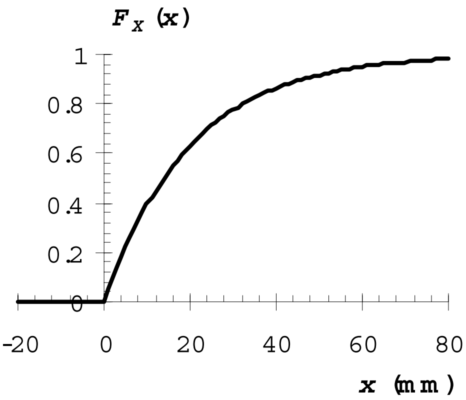
Η πιθανότερη τιμή είναι προφανώς μηδέν (βλ. Σχ. 2.3). Η διάμεσος τιμή προσδιορίζεται από τη σχέση

$$F_X(x_{0.5}) = 1/2 \Leftrightarrow 1 - e^{-\lambda x_{0.5}} = 1/2 \Leftrightarrow x_{0.5} = \ln 2 / \lambda = 13.9 \text{ mm}$$

Παρατηρούμε ότι ισχύει η ανισοτική σχέση $x_p \le x_{0.5} \le m_X$ που χαρακτηρίζει τις θετικά ασύμμετρες κατανομές.

Η τυπική απόκλιση είναι $\sigma_X$ = 20 mm και ο συντελεστής μεταβλητότητας $C_{v_X} = 1$. Αυτή η τιμή του συντελεστή μεταβλητότητας δείχνει πολύ μεγάλη διασπορά της κατανομής.




**Σχ. 2.3** Συνάρτηση πυκνότητας πιθανότητας και συνάρτηση κατανομής του ημερήσιου ύψους βροχής της εφαρμογής 2.4.

Ο συντελεστής ασυμμετρίας υπολογίζεται από τη σχέση (2.23) και είναι $C_{s_X} = (2/\lambda^3)/(1/\lambda^3) = 2$. Αυτό επιβεβαιώνει τη θετική ασυμμετρία της κατανομής, πράγμα που φαίνεται επίσης και στο Σχ. 2.3. Ειδικότερα, παρατηρούμε ότι η συνάρτηση πυκνότητας πιθανότητας του παραδείγματος έχει *σχήμα ανεστραμμένου J*, σε αντίθεση με τις συνηθέστερες κατανομές (π.χ. στο Σχ. 2.2) που έχουν *κωδωνοειδές σχήμα.*

Ο συντελεστής κύρτωσης υπολογίζεται από τη σχέση (2.24) και είναι $C_{k_X} = (9/\lambda^4)/(1/\lambda^4) = 9$. Αυτό δείχνει ότι η κατανομή είναι λεπτόκυρτη, όπως επιβεβαιώνεται και από το Σχ. 2.3.

## Χρονική κλίμακα και σχήμα κατανομής

Στο παραπάνω παράδειγμα είδαμε ότι η κατανομή ενός φυσικού μεγέθους έντονα τυχαίου, όπως η βροχόπτωση, σε μια μικρή χρονική κλίμακα, όπως η ημερήσια, εμφανίζει μεγάλη διασπορά, έντονη θετική ασυμμετρία και σχήμα πυκνότητας πιθανότητας τύπου ανεστραμμένου J. Αυτό συμβαίνει επειδή το κύριο σώμα των τιμών του μεγέθους εμφανίζεται κοντά στο μηδέν, ενώ παράλληλα εμφανίζονται και εξαιρετικά μεγάλες (θετικές) τιμές του μεγέθους, με μικρή βέβαια πιθανότητα. Το γεγονός ότι οι αρνητικές τιμές του φυσικού μεγέθους αποκλείονται οδηγεί σε θετικά ασύμμετρες κατανομές, με μεγάλη "ουρά" προς τα δεξιά, που είναι και οι συχνότερες στην υδρολογία. Ωστόσο, καθώς προχωρούμε σε μεγαλύτερες χρονικές κλίμακες, π.χ. από την ημερήσια στην ωριαία, η μέση τιμή του μεγέθους αυξάνει, χωρίς αντίστοιχου βαθμού αύξηση των μεγαλύτερων ροπών, πράγμα που οδηγεί σε μικρότερους συντελεστές διασποράς, ασυμμετρίας και κύρτωσης. Έτσι οι κατανομές έχουν πια κωδωνοειδές σχήμα και τείνουν να είναι συμμετρικές. Όπως θα δούμε παρακάτω, υπάρχουν σοβαροί θεωρητικοί λόγοι που οδηγούν σε αυτή τη συμπεριφορά για τις μεγάλες χρονικές κλίμακες (βλ. εδάφιο 2.8.1, το κεντρικό οριακό θεώρημα).

## 2.5 Αλλαγή μεταβλητής

Πολλές φορές στην υδρολογία προτιμούμε, αντί να χρησιμοποιούμε απευθείας τη μεταβλητή $X$ που περιγράφει ένα φυσικό μέγεθος, να μελετούμε μια άλλη παράγωγη μεταβλητή που προκύπτει από αμφιμονοσήμαντο μετασχηματισμό της $X$, έστω την $Y = g(X)$. Αν η $X$ είναι τυχαία μεταβλητή τότε και η $Y$ είναι επίσης τυχαία μεταβλητή και το γεγονός $\{Y \le y\}$ είναι ταυτόσημο με το γεγονός $\{X \le g^{-1}(y)\}$ όπου $g^{-1}$ είναι η αντίστροφη συνάρτηση της $g$. Κατά συνέπεια οι συναρτήσεις κατανομής των $X$ και $Y$ συνδέονται με την

$$F_Y(y) = P\{Y \le y\} = P\{X \le g^{-1}(y)\} = F_X\left(g^{-1}(y)\right) \tag{2.25}$$

Στην περίπτωση που οι μεταβλητές είναι συνεχείς και η συνάρτηση $g$ παραγωγίσιμη, αποδεικνύεται ότι η συνάρτηση πυκνότητας πιθανότητας της $Y$ δίνεται συναρτήσει της πυκνότητας της $X$ από την

$$f_Y(y) = \frac{f_X\left(g^{-1}(y)\right)}{\left|g'\left(g^{-1}(y)\right)\right|} \tag{2.26}$$

όπου $g'$ είναι η παράγωγος της $g$. Η εφαρμογή της (2.26) διασαφηνίζεται στα παραδείγματα που ακολουθούν. Ειδικότερα, ο μετασχηματισμός $Z = (X - m_X) / \sigma_X$, που μελετάται στην Εφαρμογή 2.5.α, χρησιμοποιείται πολύ συχνά στη στατιστική και κατ' επέκταση στην υδρολογία. Η μεταβλητή $Z$, που αναφέρεται ως *τυποποιημένη μεταβλητή*, είναι αδιάστατη και, όπως αποδεικνύεται παρακάτω, έχει (α) μηδενική μέση τιμή, (β) μοναδιαία τυπική απόκλιση και (γ) τρίτη και τέταρτη ροπή ίση με το συντελεστή ασυμμετρίας και κύρτωσης της $X$ αντίστοιχα.

**Εφαρμογή 2.5.α**

*Να βρεθούν η συνάρτηση κατανομής, η συνάρτηση πυκνότητας πιθανότητας και οι ροπές της μεταβλητής $Z = g(x) = (X - m_X) / \sigma_X$ συναρτήσει των αντίστοιχων μεγεθών της X.*

Δεδομένου ότι $X = g^{-1}(Z) = \sigma_X Z + m_X$, από την (2.25) προκύπτει άμεσα ότι

$$F_Z(z) = F_X\left(g^{-1}(z)\right) = F_X(\sigma_X z + m_X)$$

Αντίστοιχα, δεδομένου ότι $g'(x) = 1 / \sigma_X$, από την (2.26) προκύπτει ότι

$$f_Z(z) = \frac{f_X\left(g^{-1}(z)\right)}{\left|g'\left(g^{-1}(z)\right)\right|} = \sigma_X\, f_X(\sigma_X z + m_X)$$

Εξ άλλου, από την (2.15) έχουμε

$$E[Z] = E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx = \int_{-\infty}^{\infty} \frac{x - m_X}{\sigma_X} f_X(x) dx =$$

$$= \frac{1}{\sigma_X} \int_{-\infty}^{\infty} x f_X(x) dx - \frac{m_X}{\sigma_X} \int_{-\infty}^{\infty} f_X(x) dx = \frac{1}{\sigma_X} m_X - \frac{m_X}{\sigma_X} 1$$

και τελικά

$$m_Z = E[Z] = 0$$

πράγμα που σημαίνει ότι οι κεντρικές και οι περί την αρχή ροπές της $W$ ταυτίζονται. Έτσι η ροπή τάξης $r$ είναι

$$E\left[Z^r\right] = E\left[\left(g(X)\right)^r\right] = \int_{-\infty}^{\infty} \left(\frac{x - m_X}{\sigma_X}\right)^r f_X(x) dx =$$

$$= \frac{1}{\sigma_X^r} \int_{-\infty}^{\infty} (x - m_X)^r f_X(x) dx = \frac{1}{\sigma_X^r} \mu_X^{(r)}$$

και τελικά

$$\mu_Z^{(r)} = m_Z^{(r)} = \frac{\mu_X^{(r)}}{\sigma_X^r}$$

**Εφαρμογή 2.5.β**

*Να βρεθούν η συνάρτηση κατανομής και η συνάρτηση πυκνότητας πιθανότητας της μεταβλητής $Y = e^X$, όπου η $X$ έχει οριστεί στην Εφαρμογή 2.4.*

Είναι

$$f_X(x) = \lambda e^{-\lambda x}, \quad F_X(x) = 1 - e^{-\lambda x}$$

Εξ άλλου

$$Y = g(X) = e^X, \quad g^{-1}(Y) = \ln Y, \quad g'(X) = e^X$$

όπου $X \geq 0$ και $Y \geq 1$. Από την (2.25) προκύπτει

$$F_Y(y) = F_Y\left(g^{-1}(y)\right) = F_X(\ln y) = 1 - e^{-\lambda \ln y} = 1 - y^{-\lambda}$$

Αντίστοιχα, από την (2.26) προκύπτει

$$f_Y(y) = \frac{f_X\left(g^{-1}(y)\right)}{\left|g'\left(g^{-1}(y)\right)\right|} = \frac{\lambda e^{-\lambda \ln y}}{e^{\ln y}} = \frac{\lambda y^{-\lambda}}{y} = \lambda y^{-(\lambda+1)}$$

Η τελευταία προκύπτει και ευκολότερα με παραγώγιση της $F_Y(y)$. Η συγκεκριμένη κατανομή της $Y$ είναι γνωστή στη βιβλιογραφία ως κατανομή Pareto (βλ. εδάφιο 6.4.2).

## 2.6 Από κοινού και περιθώριες συναρτήσεις κατανομής

Στις παραπάνω ενότητες αναπτύχθηκαν οι έννοιες της θεωρίας πιθανοτήτων που αναφέρονται σε μια μεμονωμένη τυχαία μεταβλητή. Συχνά όμως ενδιαφέρει η ταυτόχρονη μελέτη δύο μεταβλητών. Έστω το ζεύγος τυχαίων μεταβλητών ($X$, $Y$) που, σύμφωνα με τα παραπάνω, είναι συναρτήσεις των δειγματικών χώρων ($\Omega_X$, $\Omega_Y$) αντίστοιχα. Η ένωση γεγονότων

$\{X \le x\} \cup \{Y \le y\} = \{X \le x, Y \le y\}$ είναι επίσης γεγονός του δειγματικού χώρου $\Omega_{XY} = \Omega_X \times \Omega_Y$. Βάσει αυτού του γεγονότος ορίζεται η *από κοινού συνάρτηση κατανομής* του ζεύγους μεταβλητών ($X$, $Y$) ως η συνάρτηση των πραγματικών μεταβλητών ($x$, $y$):

$$F_{XY}(x,y) := P(X \le x, Y \le y) \tag{2.27}$$

Οι δείκτες $X$, $Y$ συχνά παραλείπονται, εφόσον δεν υπάρχει κίνδυνος σύγχυσης. Με την προϋπόθεση ότι η $F_{XY}$ είναι παραγωγίσιμη, η συνάρτηση

$$f_{XY}(x,y) := \frac{\partial^2 F_{XY}(x,y)}{\partial x\, \partial y} \tag{2.28}$$

είναι η από κοινού συνάρτηση πυκνότητας πιθανότητας των μεταβλητών. Προφανώς ισχύει

$$F_{XY}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(\xi,\omega)\, d\omega\, d\xi \tag{2.29}$$

Οι συναρτήσεις

$$\begin{aligned} F_X(x) &:= P(X \le x) = \lim_{y\to\infty} F_{XY}(x,y) \\ F_Y(y) &:= P(Y \le y) = \lim_{x\to\infty} F_{XY}(x,y) \end{aligned} \tag{2.30}$$

λέγονται *περιθώριες συναρτήσεις κατανομής* των $X$ και $Y$, αντίστοιχα. Ορίζονται επίσης *περιθώριες συναρτήσεις πυκνότητας πιθανότητας* που δίνονται από τις σχέσεις

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx \tag{2.31}$$

Ακόμη, ενδιαφέρον παρουσιάζει η *δεσμευμένη συνάρτηση κατανομής* και η *δεσμευμένη συνάρτηση πυκνότητας πιθανότητας* της $X$ για δεδομένη τιμή της $Y$ που δίνονται από τις εξισώσεις

$$F_X(x|Y=y)=\frac{\int_{-\infty}^{x} f_{XY}(\xi,y)\,d\xi}{f_Y(y)}, \quad f_X(x|Y=y)=\frac{f_{XY}(x,y)}{f_Y(y)} \tag{2.32}$$

αντίστοιχα. Εναλλάσσοντας τα $X$ και $Y$ παίρνουμε τις δεσμευμένες συναρτήσεις της $Y$.

### 2.6.1 Αναμενόμενες τιμές - ροπές

Η αναμενόμενη τιμή ή προσδοκία της συνάρτησης $g(X, Y)$ ορίζεται από την εξίσωση

$$E\left[g(X,Y)\right]:=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)\,dy\,dx \tag{2.33}$$

Το μέγεθος $E[X^p Y^q]$ ονομάζεται *από κοινού ροπή τάξης $p + q$* των $X$ και $Y$. Αντίστοιχα, το μέγεθος $E[(X-m_X)^p (Y-m_Y)^q]$ ονομάζεται *από κοινού κεντρική ροπή τάξης $p + q$* των $X$ και $Y$. Από τις κεντρικές ροπές συχνότερα χρησιμοποιείται η

$$\sigma_{XY} := E\left[(X-m_X)(Y-m_Y)\right] = E[XY]-m_X m_Y \tag{2.34}$$

που ονομάζεται και *συνδιασπορά* των $X$ και $Y$. Εναλλακτικά η συνδιασπορά συμβολίζεται ως $\mathrm{Cov}[X,Y]$. Διαιρώντας τη συνδιασπορά με τις τυπικές αποκλίσεις $\sigma_X$ και $\sigma_Y$ παίρνουμε το *συντελεστή συσχέτισης*

$$\rho_{XY} := \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}} \equiv \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{2.35}$$

ο οποίος είναι αδιάστατος και με τιμές $-1 \le \rho_{XY} \le 1$. Όπως θα δούμε παρακάτω, ο συντελεστής αυτός αποτελεί σημαντική παράμετρο για τη μελέτη της συσχέτισης μεταξύ δύο μεταβλητών.

Η *δεσμευμένη αναμενόμενη τιμή* δεδομένης συνάρτησης μιας μεταβλητής, π.χ. της $g(X)$ ορίζεται από την εξίσωση

$$E\left[g(X)|Y=y\right]:=\int_{-\infty}^{\infty} g(x) f_X(x|Y=y)\,dx \tag{2.36}$$

Σπουδαιότερο μέγεθος αυτού του τύπου είναι η δεσμευμένη μέση τιμή της $X$:

$$E\left[X|Y=y\right]:=\int_{-\infty}^{\infty} x f_X(x|Y=y)\,dx \tag{2.37}$$

Αντίστοιχα ορίζεται και η δεσμευμένη μέση τιμή της $Y$.

### 2.6.2 Ανεξαρτησία μεταβλητών

Οι μεταβλητές ($X$, $Y$) λέγονται *ανεξάρτητες* αν για κάθε ζεύγος τιμών ($x$, $y$) ισχύει

$$F_{XY}(x,y)=F_X(x)F_Y(y) \tag{2.38}$$

Σε αυτή την περίπτωση ισχύει επίσης η σχέση

$$f_{XY}(x,y)=f_X(x)f_Y(y) \tag{2.39}$$

η οποία είναι ισοδύναμη με την (2.38), καθώς και οι ακόλουθες

$$\sigma_{XY}=0 \Leftrightarrow \rho_{XY}=0 \Leftrightarrow E[XY]=E[X]E[Y] \tag{2.40}$$

$$E[X|Y=x]=E[X], \quad E[Y|X=x]=E[Y] \tag{2.41}$$

οι οποίες αποτελούν απλές συνέπειες της (2.38) και όχι ικανές συνθήκες για να είναι ανεξάρτητες οι μεταβλητές ($X$, $Y$). Δύο μεταβλητές ($X$, $Y$) για τις οποίες ισχύει η (2.40) λέγονται *ασυσχέτιστες*.

### 2.6.3 Αθροίσματα μεταβλητών

Συνέπεια του ορισμού της αναμενόμενης τιμής (εξίσωση (2.33))είναι η σχέση

$$E\left[c_1 g_1(X,Y)+c_2 g_2(X,Y)\right]=c_1 E\left[g_1(X,Y)\right]+c_2 E\left[g_2(X,Y)\right] \tag{2.42}$$

όπου $c_1$ και $c_2$ είναι τυχούσες σταθερές ενώ $g_1$ και $g_2$ είναι τυχούσες συναρτήσεις. Προφανώς η ιδιότητα αυτή επεκτείνεται και για περισσότερες συναρτήσεις $g_i$. Εφαρμόζοντας τη (2.42) για το άθροισμα δύο μεταβλητών έχουμε

$$E[X+Y] = E[X] + E[Y] \tag{2.43}$$

Κατά παρόμοιο τρόπο

$$\begin{aligned} E\left[(X - m_X + Y - m_Y)^2\right] = E\left[(X - m_X)^2\right] + E\left[(Y - m_Y)^2\right] + \\ +2E\left[(X - m_X)(Y - m_Y)\right] \end{aligned} \tag{2.44}$$

Κατά συνέπεια

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X, Y] \tag{2.45}$$

Η συνάρτηση πυκνότητας πιθανότητας του αθροίσματος $Z = X + Y$ είναι γενικά δύσκολο να υπολογιστεί. Στην περίπτωση που οι $X$ και $Y$ είναι ανεξάρτητες, αποδεικνύεται ότι

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - w)\, f_Y(w)\, dw \tag{2.46}$$

Το ολοκλήρωμα αυτό είναι γνωστό ως *ολοκλήρωμα συνέλιξης* των συναρτήσεων $f_X(x)$ και $f_Y(y)$.

**Εφαρμογή 2.6.3**

*Σε στεγανή λίμνη έκτασης 10.0 km$^2$ η εισροή του Απριλίου, η οποία συνίσταται από τη βροχόπτωση επί της επιφάνειας της λίμνης και την απορροή από την επιφάνεια της υδρολογικής της λεκάνης, είναι τυχαία μεταβλητή με μέση τιμή $4.0 \times 10^6$ m$^3$ και τυπική απόκλιση $1.5 \times 10^6$ m$^3$. Η εξάτμιση του ίδιου μήνα, που αποτελεί τη μοναδική εκροή από τη λίμνη, είναι επίσης τυχαία μεταβλητή με μέση τιμή 90.0 mm και τυπική απόκλιση 20.0 mm. Με την προϋπόθεση ότι η εισροή και η εκροή είναι στοχαστικά ανεξάρτητες, ζητείται η μέση τιμή και η τυπική απόκλιση της μεταβολής της στάθμης της λίμνης τον Απρίλιο, καθώς και ο συντελεστής συσχέτισης της τελευταίας με την εισροή και με την εκροή.*

Αρχικά εκφράζουμε την εισροή σε ίδιες μονάδες όπως στην εκροή. Για το σκοπό αυτό διαιρούμε τον όγκο εισροής με την έκταση της λίμνης, οπότε παίρνουμε τη μεταβολή της στάθμης της λίμνης στην οποία αντιστοιχεί η εισροή. Έτσι έχουμε για τη μέση τιμή $4.0 \times 10^6$ / $10.0 \times 10^6 = 0.4$ m = 400.0 mm και για την τυπική απόκλιση $1.5 \times 10^6$ / $10.0 \times 10^6 = 0.15$ m = 150.0 mm.

Συμβολίζουμε με $X$ και $Y$ την εισροή και την εκροή του Απριλίου, αντίστοιχα, και με $Z$ τη μεταβολή της στάθμης της λίμνης τον ίδιο μήνα. Όλα τα μεγέθη τα εκφράζουμε σε mm. Προφανώς ισχύει

$$Z = X - Y \tag{α}$$

Έχουμε δεδομένα τα μεγέθη

$$m_X = E[X] = 400.0 \text{ mm}, \quad \sigma_X = \sqrt{Var[X]} = 150.0 \text{ mm}$$

$$m_Y = E[Y] = 90.0 \text{ mm}, \quad \sigma_Y = \sqrt{Var[Y]} = 20.0 \text{ mm}$$

Συνδυάζοντας την (α) και τη (2.42) παίρνουμε

$$E[Z] = E[X - Y] = E[X] - E[Y] \Rightarrow m_Z = m_X - m_Y \tag{β}$$

απ' όπου τελικά προκύπτει $m_Z$ = 310.0 mm. Αφαιρώντας κατά μέλη τις (α) και (β) παίρνουμε

$$Z - m_z = (X - m_X) - (Y - m_Y) \tag{γ}$$

Υψώνοντας την παραπάνω στο τετράγωνο βρίσκουμε

$$(Z - m_z)^2 = (X - m_X)^2 + (Y - m_Y)^2 - 2(X - m_X)(Y - m_Y)$$

Παίρνοντας αναμενόμενες τιμές στην παραπάνω, καταλήγουμε στην ακόλουθη σχέση, αντίστοιχη της (2.45)

$$\mathrm{Var}[Z] = \mathrm{Var}[X - Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] - 2\mathrm{Cov}[X, Y] \tag{δ}$$

Δεδομένου, όμως, ότι οι μεταβλητές $X$ και $Y$ είναι ανεξάρτητες θα είναι Cov[X, Y] = 0 (εξίσωση 2.40). Κατά συνέπεια από τη (δ) βρίσκουμε

$$\sigma_Z^2 = 150.0^2 + 20.0^2 = 22\,900.0 \text{ mm}^2$$

και $\sigma_Z$ = 151.3 mm.

Πολλαπλασιάζοντας τα δύο μέλη της (γ) επί $(X - m_X)$ και παίρνοντας μετά αναμενόμενες τιμές βρίσκουμε

$$E\left[(Z - m_z)(X - m_X)\right] = E\left[(X - m_X)^2\right] - E\left[(X - m_X)(Y - m_Y)\right]$$

ή απλούστερα

$$\mathrm{Cov}[Z, X] = \mathrm{Var}[X] - \mathrm{Cov}[X, Y] \qquad (\varepsilon)$$

Ο τελευταίος όρος του δεξιού μέλους της (ε) είναι μηδέν, οπότε

$$\sigma_{ZY} = \sigma_X^2 = 150.0^2 = 22\,500.0 \text{ mm}^2$$

και κατά συνέπεια

$$\rho_{ZX} = \sigma_{ZX} / \left(\sigma_Z \sigma_X\right) = 22500.0 / \left(151.3 * 150.0\right) = 0.991$$

Ομοίως έχουμε

$$\mathrm{Cov}[Z, Y] = \mathrm{Cov}[X, Y] - \mathrm{Var}[Y] \qquad (\zeta)$$

Δεδομένου ότι ο πρώτος όρος του δεξιού μέλους της (ζ) είναι μηδέν θα έχουμε

$$\sigma_{ZY} = -\sigma_Y^2 = -20.0^2 = -400.0 \text{ mm}^2$$

και κατά συνέπεια

$$\rho_{ZY} = \sigma_{ZY} / \left(\sigma_Z \sigma_Y\right) = -400.0 / \left(151.3 * 20.0\right) = -0.132$$

Η θετική τιμή του $\rho_{ZX}$ περιγράφει το γεγονός ότι η στάθμη αυξάνεται με την αύξηση της εισροής (θετική συσχέτιση των $X$ και $Z$). Αντίστοιχα, η αρνητική τιμή του $\rho_{ZY}$ περιγράφει το γεγονός ότι η στάθμη μειώνεται με την αύξηση της εκροής (αρνητική συσχέτιση των $Y$ και $Z$). Η πολύ μεγάλη , κοντά στο 1, τιμή του $\rho_{ZX}$ σε σχέση με την κατ' απόλυτο πολύ μικρότερη τιμή του $\rho_{ZY}$ περιγράφει το γεγονός ότι η μεταβολή της στάθμης εξαρτάται πρωτίστως από την εισροή, η οποία έχει και μεγαλύτερο μέγεθος και μεγαλύτερες διακυμάνσεις (τυπική απόκλιση) και δευτερευόντως από την εκροή.

### 2.6.4 Πολλές μεταβλητές

Όσα αναφέρονται σε δύο μεταβλητές μπορούν εύκολα να επεκταθούν σε περισσότερες από δύο μεταβλητές. Για παράδειγμα, η συνάρτηση κατανομής των $n$ μεταβλητών $X_1, X_2, \ldots, X_n$ είναι

$$F_{X_1,\cdots,X_n}(x_1,\ldots,x_n) := P(X_1 \le x_1,\ldots,X_n \le x_n) \tag{2.47}$$

και συνδέεται με τη συνάρτηση πυκνότητας πιθανότητας με την

$$F_{X_1,\cdots,X_n}(x_1,\ldots,x_n) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f_{X_1,\cdots,X_n}(\xi_1,\ldots,\xi_n)d\xi_n\cdots d\xi_1 \tag{2.48}$$

Οι μεταβλητές $X_1$, $X_2$, …, $X_n$ λέγονται ανεξάρτητες αν για όλες τις τιμές $x_1$, $x_2$, …, $x_n$ ισχύει

$$F_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = F_{X_1}(x_1)\ldots F_{X_n}(x_n) \tag{2.49}$$

Οι αναμενόμενες τιμές και ροπές ορίζονται κατά παρόμοιο τρόπο όπως στις δύο μεταβλητές, ενώ η ιδιότητα (2.42) γενικεύεται για συναρτήσεις $g_i$ πολλών μεταβλητών.

## 2.7 Η έννοια της στοχαστικής ανέλιξης

Η θεωρία των στοχαστικών ανελίξεων αποτελεί ιδιαίτερο κλάδο της θεωρίας πιθανοτήτων, ο οποίος μάλιστα είναι από τους πιο προχωρημένους. Ο σκοπός μας εδώ δεν είναι να δώσουμε πλήρη εικόνα αυτού του κλάδου, αλλά μόνο ορισμένες εισαγωγικές έννοιες που είναι χρήσιμες για την ορθή κατανόηση της μαθηματικής περιγραφής των υδρολογικών διεργασιών και των προϋποθέσεων στις οποίες στηρίζεται αυτή η περιγραφή. Οι έννοιες αυτές και η σύνδεση τους με την τεχνική υδρολογία θα διασαφηνιστούν στο κεφάλαιο 4.

Οι στοχαστικές ανελίξεις αποτελούν οικογένειες τυχαίων μεταβλητών, όπως αυτές της προηγούμενης ενότητας, που, όμως, μπορεί να είναι και απειροπληθείς. Έτσι, *στοχαστική ανέλιξη* ονομάζεται μια οικογένεια μεταβλητών $X_t$ όπου $t$ είναι παράμετρος που παίρνει τιμές από ένα κατάλληλο σύνολο $T$ (Κάκουλος, 1978· Papoulis, 1965· Taylor and

Karlin, 1984). Εναλλακτικά χρησιμοποιείται και ο συμβολισμός $X(t)$ αντί του $X_t$. Αν και γενικά το *δεικτοσύνολο* $T$ μπορεί να είναι οποιοδήποτε σύνολο, συχνότατα παριστάνει χρόνο. Σε περίπτωση που το δεικτοσύνολο αντιστοιχεί σε διακριτές μονάδες χρόνου, $T = \{0, 1, 2, \ldots\}$, έχουμε μια *ανέλιξη σε διακριτό χρόνο*. Αντίστοιχα, αν το δεικτοσύνολο αντιστοιχεί σε συνεχή χρόνο, δηλαδή $T = [0, \infty)$, μιλούμε για *ανέλιξη σε συνεχή χρόνο*.

Το σύνολο τιμών της στοχαστικής ανέλιξης, δηλαδή το σύνολο των δυνατών τιμών των τυχαίων μεταβλητών $X_t$ ονομάζεται *φασικός χώρος*.

Μια υλοποίηση της στοχαστικής ανέλιξης, δηλαδή ένα σύνολο παρατηρήσεων $x(t)$ της $X(t)$, για μεταβαλλόμενο χρόνο $t$, ονομάζεται *δειγματοσυνάρτηση* ή *χρονοσειρά* της ανέλιξης.

### 2.7.1 Συνάρτηση κατανομής

Η συνάρτηση κατανομής της τυχαίας μεταβλητής $X_t$, δηλαδή η

$$F(x;t) := P\big(X(t) \le x\big) \tag{2.50}$$

ονομάζεται *συνάρτηση κατανομής πρώτης τάξης* της ανέλιξης. Αντίστοιχα, έχουμε τη συνάρτηση δεύτερης τάξης

$$F(x_1, x_2; t_1, t_2) := P\big(X(t_1) \le x_1, X(t_2) \le x_2\big) \tag{2.51}$$

και κατ' επέκταση τη συνάρτηση $n$ τάξης

$$F(x_1, \ldots, x_n; t_1, \ldots, t_n) := P\big(X(t_1) \le x_1, \ldots, X(t_n) \le x_n\big) \tag{2.52}$$

Με παραγώγιση των συναρτήσεων κατανομής παίρνουμε κατά τα γνωστά τις αντίστοιχες συναρτήσεις πυκνότητας πιθανότητας.

Μια στοχαστική ανέλιξη είναι στατιστικά ορισμένη αν γνωρίζουμε τη συνάρτηση κατανομής $n$ τάξης για κάθε τιμή του $n$.

### 2.7.2 Ροπές

Οι ροπές ορίζονται και εδώ με τον ίδιο τρόπο όπως στην ενότητα 2.4. Ιδιαίτερο ενδιαφέρον παρουσιάζουν:

1. Η *μέση τιμή της ανέλιξης*, δηλαδή η αναμενόμενη τιμή της μεταβλητής $X(t)$:

$$\mu(t) := E\left[X(t)\right] = \int_{-\infty}^{\infty} x\, f(x;t)\,dt \tag{2.53}$$

2. Η *αυτοσυνδιασπορά* της ανέλιξης, δηλαδή η συνδιασπορά των τυχαίων μεταβλητών $X(t_1)$ και $X(t_2)$:

$$C(t_1, t_2) := \mathrm{Cov}\left[X(t_1), X(t_2)\right] = E\left[\left(X(t_1) - \mu(t_1)\right)\left(X(t_2) - \mu(t_2)\right)\right] \tag{2.54}$$

Σημειώνεται ότι η διασπορά της ανέλιξης, δηλαδή η διασπορά της μεταβλητής $X(t)$, είναι $\mathrm{Var}[X(t)] = C(t, t)$. Κατά συνέπεια ο *συντελεστής αυτοσυσχέτισης*, δηλαδή ο συντελεστής συσχέτισης των τυχαίων μεταβλητών $X(t_1)$ και $X(t_2)$ ορίζεται από τη σχέση

$$\rho(t_1, t_2) := \frac{\mathrm{Cov}\left[X(t_1), X(t_2)\right]}{\sqrt{Var\left[X(t_1)\right]Var\left[X(t_2)\right]}} = \frac{C(t_1, t_2)}{\sqrt{C(t_1, t_1)C(t_2, t_2)}} \tag{2.55}$$

### 2.7.3 Στασιμότητα

Όπως υπονοείται από τον παραπάνω συμβολισμό, στη γενική περίπτωση οι στατιστικές παράμετροι μιας στοχαστικής ανέλιξης, π.χ. η μέση τιμή και η διασπορά της, είναι συναρτήσεις του χρόνου και κατά συνέπεια μεταβάλλονται με τη μεταβολή του χρόνου. Μια ειδική κατηγορία ανελίξεων, σαφώς απλούστερων στη μελέτη, είναι οι στάσιμες ανελίξεις στις οποίες δεν υπάρχει μεταβολή των στατιστικών χαρακτηριστικών με την πάροδο του χρόνο. Έτσι, μια στοχαστική ανέλιξη λέγεται *στάσιμη με την αυστηρή έννοια*, ή απλώς *στάσιμη*, όταν η συνάρτηση κατανομής της δεν επηρεάζεται από τη μετατόπιση του χρόνου, δηλαδή αν, για τυχούσα χρονική μετατόπιση $\tau$, η συνάρτηση κατανομής οποιασδήποτε τάξης της $X(t + \tau)$ ταυτίζεται με τη συνάρτηση κατανομής της ίδιας τάξης της $X(t)$. Λέγεται δε *στάσιμη με την ευρεία* (ή *ελαστική*) *έννοια* αν η μέση τιμή της είναι σταθερή και η αυτοσυνδιασπορά της εξαρτάται μόνο από τη διαφορά του χρόνου, δηλαδή αν

$$E[X(t)] = \mu = \text{σταθερά}, \quad E[(X(\tau) - \mu)\,(X(t + \tau) - \mu)] = C(\,\tau) \tag{2.56}$$

### 2.7.4 Εργοδικότητα

Η έννοια της εργοδικότητας μιας στοχαστικής ανέλιξης έχει σχέση με το πρόβλημα του προσδιορισμού της κατανομής της από μια απλή σειρά παρατηρήσεών της. Έτσι μια στάσιμη στοχαστική ανέλιξη είναι *εργοδική* αν κάθε παράμετρος της κατανομής μπορεί να προσδιοριστεί από μια απλή δειγματοσυνάρτηση της ανέλιξης. Δεδομένου ότι οι παράμετροι υπολογίζονται ως χρονικές μέσες τιμές, ο παραπάνω ορισμός εκφράζεται και με τον εξής τρόπο: Μια ανέλιξη είναι εργοδική αν οι χρονικοί μέσοι είναι ίσοι με τους συνολικούς μέσους (δηλαδή τις αναμενόμενες τιμές). Για παράδειγμα, μια ανέλιξη είναι εργοδική ως προς τη μέση τιμή αν

$$E\left[X(t)\right] = \lim_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N} X(t) \qquad \text{(για ανέλιξη διακριτού χρόνου)}$$

$$E\left[X(t)\right] = \lim_{T\to\infty} \frac{1}{T} \int_0^T X(t)\,dt \qquad \text{(για ανέλιξη συνεχούς χρόνου)} \tag{2.57}$$

Το αριστερό μέλος των παραπάνω παριστάνει το συνολικό μέσο, δηλαδή την αναμενόμενη τιμή, ενώ το δεξιό μέλος παριστάνει το χρονικό μέσο, στην οριακή περίπτωση άπειρου χρόνου. Ενώ το αριστερό μέλος είναι παράμετρος και όχι τυχαία μεταβλητή, το δεξιό μέλος είναι τυχαία μεταβλητή (ως άθροισμα ή ολοκλήρωμα τυχαίων μεταβλητών). Η εξίσωση λοιπόν μιας παραμέτρου με μια τυχαία μεταβλητή υπονοεί ότι η τυχαία μεταβλητή έχει μηδενική διασπορά. Αυτή ακριβώς είναι και η προϋπόθεση που κάνει την ανέλιξη εργοδική, προϋπόθεση που, ας σημειωθεί, δεν ισχύει απαραίτητα για κάθε στοχαστική ανέλιξη.

## 2.8 Το κεντρικό οριακό θεώρημα και μερικές συναρτήσεις κατανομής

Το *κεντρικό οριακό θεώρημα* είναι ένα από τα πιο σημαντικά της θεωρίας πιθανοτήτων. Αφορά στην οριακή κατανομή αθροίσματος τυχαίων μεταβλητών-συνιστωσών, η οποία, ανεξάρτητα από τις κατανομές των συνιστωσών του αθροίσματος, είναι πάντα η ίδια, η γνωστή *κανονική κατανομή*. Η κατανομή αυτή είναι η πιο γενικευμένη σε χρήση, τόσο στη θεωρία πιθανοτήτων, όσο και σε άλλες επιστήμες, και προκύπτει όχι μόνο ως συνέπεια του κεντρικού οριακού θεωρήματος, αλλά και από άλλες θεωρήσεις, π.χ. από την αρχή της μέγιστης εντροπίας (Papoulis, 1990, σ. 422-430).

Στην ενότητα αυτή περιγράφουμε το κεντρικό οριακό θεώρημα, την κανονική κατανομή, καθώς και ορισμένες άλλες κατανομές που συνδέονται άμεσα με την κανονική ($\chi^2$ και Student). Όλες αυτές οι κατανομές είναι θεμελιώδεις για τη στατιστική και, όπως θα δούμε στο κεφάλαιο 3, χρησιμοποιούνται άμεσα για τη στατιστική εκτίμηση και πρόγνωση. Πέρα από αυτό, η κανονική κατανομή έχει και άμεσες εφαρμογές στην τεχνική υδρολογία. Αυτές θα εξεταστούν ειδικότερα στο κεφάλαιο 6 μαζί με τους άλλους τύπους κατανομών που χρησιμοποιούνται στην τεχνική υδρολογία.

### 2.8.1 Το κεντρικό οριακό θεώρημα και η σημασία του

Έστω οι τυχαίες μεταβλητές $X_i$ ($i = 1, \ldots, n$) και το άθροισμά τους $Z = X_1 + X_2 + \cdots + X_n$, με $E[Z] = m_Z$ και $\mathrm{Var}[Z] = \sigma_Z^2$. Το κεντρικό οριακό θεώρημα λέει ότι η κατανομή της τυχαίας μεταβλητής $Z$, υπό ορισμένες γενικές προϋποθέσεις (συνοψίζονται πιο κάτω), έχει συγκεκριμένο όριο όταν το $n$ τείνει στο άπειρο, που δίνεται από τη σχέση

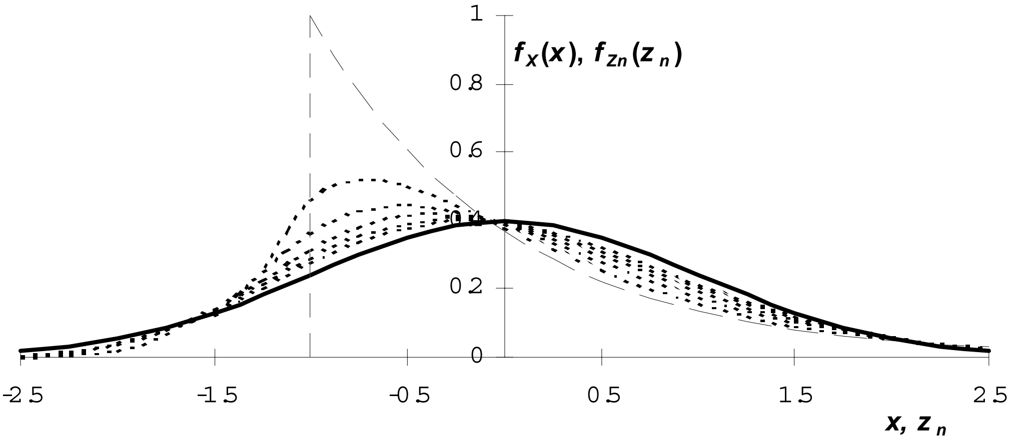
$$F_Z(z) \xrightarrow[n\to\infty]{} \int_{-\infty}^{z} \frac{1}{\sigma_Z\sqrt{2\pi}}\, e^{-\frac{(\zeta - m_Z)^2}{2\sigma_Z^2}}\, d\zeta \tag{2.58}$$

και επιπλέον, αν οι $X_i$ είναι συνεχείς, η πυκνότητα πιθανότητας της $Z$ έχει αντίστοιχο όριο:

$$f_Z(z) \xrightarrow[n\to\infty]{} \frac{1}{\sigma_Z\sqrt{2\pi}}\, e^{-\frac{(z - m_Z)^2}{2\sigma_Z^2}} \tag{2.59}$$

Η συνάρτηση κατανομής του δεξιού μέλους της (2.58) είναι η γνωστή *κανονική κατανομή* ή *κατανομή Gauss* και, αντίστοιχα, η συνάρτηση του δεξιού μέλους της (2.59) είναι η κανονική συνάρτηση πυκνότητας πιθανότητας.

Πρακτικά, η σύγκλιση για $n \to \infty$ μπορεί να θεωρηθεί ως προσέγγιση όταν το $n$ είναι αρκετά μεγάλο. Το πόσο μεγάλο πρέπει να είναι το $n$ ώστε η προσέγγιση να είναι ικανοποιητική εξαρτάται από τις συναρτήσεις κατανομής των επιμέρους $X_i$. Πάντως, στις περισσότερες εφαρμογές θεωρείται ικανοποιητική η τιμή $n = 30$ (με την προϋπόθεση ότι οι $X_i$ έχουν την ίδια συνάρτηση κατανομής). Στο Σχ. 2.4 δίνεται μια γραφική ερμηνεία του κεντρικού οριακού θεωρήματος βασισμένη σε ένα παράδειγμα. Ξεκινώντας από μεταβλητές $X_i$ με ασύμμετρη εκθετική κατανομή, υπολογίζουμε την κατανομή του αθροίσματος 2, 4, 8, 16 και 32 μεταβλητών. Είναι εμφανής η σύγκλιση της κατανομής του αθροίσματος προς την (συμμετρική) κανονική κατανομή όσο αυξάνεται το $n$, καθώς και ο ικανοποιητικός βαθμός προσέγγισης της κατανομής του αθροίσματος 32 μεταβλητών προς την κανονική κατανομή. Αν οι αρχικές μεταβλητές είχαν συμμετρική κατανομή, η σύγκλιση θα ήταν πολύ ταχύτερη.




**Σχ. 2.4** Σύγκλιση της συνάρτησης πυκνότητας πιθανότητας αθροίσματος μεταβλητών εκθετικής κατανομής προς την κανονική πυκνότητα πιθανότητας (έντονη γραμμή). Η διακεκομμένη γραμμή με μέγιστο στη θέση −1 παριστάνει την πυκνότητα πιθανότητας των αρχικών μεταβλητών $X_i$ που είναι $f_X(x) = e^{-(x-1)}$ (μέση τιμή 0 και τυπική απόκλιση 1). Οι εστιγμένες γραμμές (ξεκινώντας από τις περισσότερο προς τις λιγότερο αιχμηρές) παριστάνουν τις πυκνότητες πιθανότητας των αθροισμάτων $Z_n = (X_1 + \dots + X_n) / n$ για $n = 2, 4, 8, 16$ και 32. Η διαίρεση του αθροίσματος με $n$ έγινε για να έχουν όλες οι $Z_i$ την ίδια μέση τιμή και διασπορά (0 και 1, αντίστοιχα), ώστε να είναι άμεσα συγκρίσιμες οι κατανομές τους, και δεν επηρεάζει την ουσία του κεντρικού οριακού θεωρήματος. Οι πυκνότητες πιθανότητας των $Z_i$ υπολογίστηκαν θεωρητικά (με ολοκληρώματα συνέλιξης όπως αυτό της ((2.46)- βλ. και κεφάλαιο 6 κατανομή γάμα).

Οι προϋποθέσεις για να ισχύει το κεντρικό οριακό θεώρημα είναι αρκετά γενικές ώστε να ισχύουν για πάρα πολλές μεταβλητές με φυσικό νόημα. Διάφορα σύνολα προϋποθέσεων (βλ. και Papoulis, 1990, σ. 215, Benjamin & Cornell, 1970, σ. 251) με ιδιαίτερο ενδιαφέρον είναι π.χ. (α) οι μεταβλητές $X_i$ είναι ανεξάρτητες με κοινή συνάρτηση κατανομής και πεπερασμένη τρίτη ροπή, (β) οι μεταβλητές $X_i$ είναι ανεξάρτητες, πεπερασμένες και με διασπορά μεγαλύτερη από 0, (γ) οι μεταβλητές $X_i$ είναι ανεξάρτητες και με πεπερασμένη τρίτη ροπή, ενώ η διασπορά της $Z$ τείνει στο άπειρο όταν το $n$ τείνει στο άπειρο (δ) οι μεταβλητές δεν είναι ανεξάρτητες αλλά είναι πεπερασμένες και ο συντελεστής συσχέτισης είναι ουσιαστικώς μηδέν μεταξύ κάθε μεταβλητής και όλων των άλλων εκτός από ένα περιορισμένο αριθμό.

Στις (2.58) και (2.59) παρατηρούμε ότι τα όρια των συναρτήσεων $F_Z(z)$ και $f_Z(z)$ δεν εξαρτώνται από τις συναρτήσεις κατανομής των $X_i$, δηλαδή το αποτέλεσμα είναι ίδιο για οποιοδήποτε τύπο συνάρτησης κατανομής των $X_i$. Έτσι: (α) μπορούμε να ξέρουμε την κατανομή του αθροίσματος χωρίς να ξέρουμε τις συναρτήσεις κατανομής των συνιστωσών του, και (β) η ίδια ακριβώς συνάρτηση κατανομής περιγράφει οποιαδήποτε μεταβλητή προκύπτει ως άθροισμα πολλών επιμέρους συνιστωσών. Εκεί ακριβώς βρίσκεται και η σπουδαιότητα του κεντρικού οριακού θεωρήματος και η μεγάλη διάδοση της κανονικής κατανομής σε όλες σχεδόν τις επιστήμες (φυσικές, κοινωνικές, οικονομικές κτλ.). Ειδικότερα στη στατιστική, όπως θα δούμε στο κεφάλαιο 3, το κεντρικό οριακό θεώρημα συνεπάγεται ότι η μέση τιμή δείγματος οποιασδήποτε μεταβλητής ακολουθεί κανονική κατανομή (για μέγεθος δείγματος αρκετά μεγάλο).

Στην τεχνική υδρολογία, όπως θα δούμε και αλλού αναλυτικότερα, η κανονική κατανομή περιγράφει με ικανοποιητική προσέγγιση μεταβλητές που αναφέρονται σε μεγάλη χρονική κλίμακα π.χ. ετήσια. Έτσι η ετήσια βροχόπτωση σε μια όχι άνυδρη περιοχή μπορεί να θεωρηθεί ως άθροισμα πολλών (π.χ. περισσότερων από 30) επεισοδίων βροχής κατά τη διάρκεια του έτους, των οποίων τα ύψη βροχής είναι πρακτικώς ανεξάρτητα. Αντίστοιχα η (χρονικά) μέση ετήσια παροχή μπορεί να θεωρηθεί ως μέσος όρος π.χ. 365 ημερήσιων παροχών, οι οποίες είναι μεν εξαρτημένες αλλά η στοχαστική εξάρτηση κατά κανόνα περιορίζεται σε ένα μικρό σχετικά διάστημα (πριν και μετά) ώστε και πάλι να ισχύει το κεντρικό οριακό θεώρημα. Όμως, θα πρέπει να σημειωθεί, τα συμπεράσματα αυτά δεν εφαρμόζονται για άνυδρες περιοχές, όπου π.χ. τα επεισόδια βροχής κάθε χρόνο είναι λίγα.

### 2.8.2 Η κανονική κατανομή

Λέμε ότι η μεταβλητή X είναι *κανονική* (ή ότι ακολουθεί κανονική κατανομή ή κατανομή Gauss) με παραμέτρους $\mu$ και $\sigma$ (συμβολικά $N(\mu, \sigma)$) αν η πυκνότητα πιθανότητάς της είναι η συνάρτηση

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.60}$$

Η αντίστοιχη συνάρτηση κατανομής είναι η

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(\xi-\mu)^2}{2\sigma^2}} d\xi \tag{2.61}$$

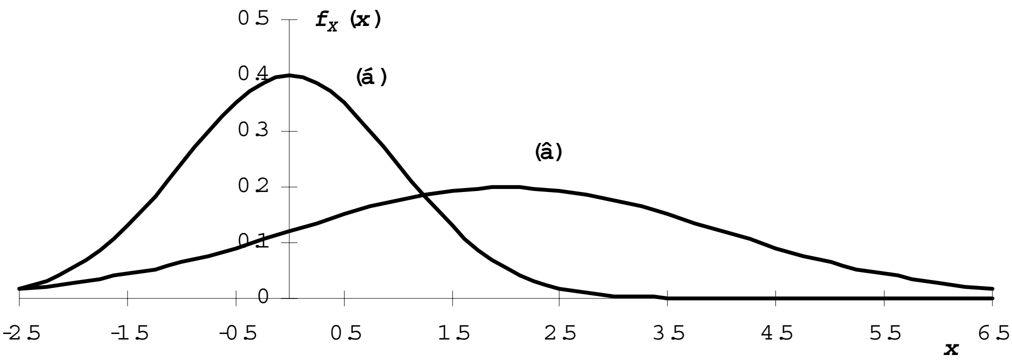
Η μέση τιμή και η τυπική απόκλιση της κατανομής είναι $\mu$ και $\sigma$, αντίστοιχα. Η κατανομή είναι συμμετρική (βλ. Σχ. 2.5) και επομένως η τρίτη κεντρική ροπή της είναι μηδέν (άρα $C_s = 0$). Η τέταρτη κεντρική ροπή της είναι $3\sigma^4$ (άρα $C_k = 3$).

Το ολοκλήρωμα στο δεξιό μέλος της (2.61) δεν υπολογίζεται αναλυτικά. Έτσι για τους τυπικούς υπολογισμούς ($x \to F_X(x)$ και $F_X(x) \to x$) κατά κανόνα χρησιμοποιούνται πίνακες (βλ. Παράρτημα στατιστικών πινάκων στο τέλος του βιβλίου), οι οποίοι δίνουν την κατανομή της *τυποποιημένης κανονικής μεταβλητής* $Z$ που προκύπτει από τη $X$ με το μετασχηματισμό

$$Z = \frac{X-\mu}{\sigma} \Leftrightarrow X = \mu + \sigma Z \tag{2.62}$$

και έχει κατανομή $N(0,1)$. Εύκολα αποδεικνύεται (βλ. Εφαρμογή 2.5.α) ότι

$$F_X(x) = F_Z(z) = F_Z\left(\frac{x-\mu}{\sigma}\right) \tag{2.63}$$




**Σχ. 2.5** Δύο παραδείγματα κανονικής πυκνότητας πιθανότητας (α) $N(0,1)$ και (β) $N(2, 2)$.

### Εφαρμογή 2.8.2

*Το ετήσιο ύψος βροχής σταθμού της βορειοδυτικής Ελλάδας ακολουθεί κανονική κατανομή με $\mu$ = 1750 mm και $\sigma$ = 410 mm. Ζητούνται:*

*(α) η πιθανότητα υπέρβασης της τιμής 2500 mm*
*(β) η βροχόπτωση πιθανότητας υπέρβασης 1/50.*

(α) Για $x = 2500$ βρίσκουμε $z = (2500 - 1750) / 410 = 1.83$. Από τον πίνακα της κανονικής κατανομής (Παράρτημα Π1) προκύπτει $F_Z(z) = 0.9664$ ($= F_X(x)$). Άρα $F_{1X}(x) = 1 - 0.9664 = 0.0336$.

(β) Έχουμε $F_X(x) = F_Z(z) = 1 - 1/50 = 1 - 0.02 = 0.98$ και από τον ίδιο πίνακα $z = 2.05$ οπότε $x = 1750 + 410 \times 2.05 = 2590.5$ mm.

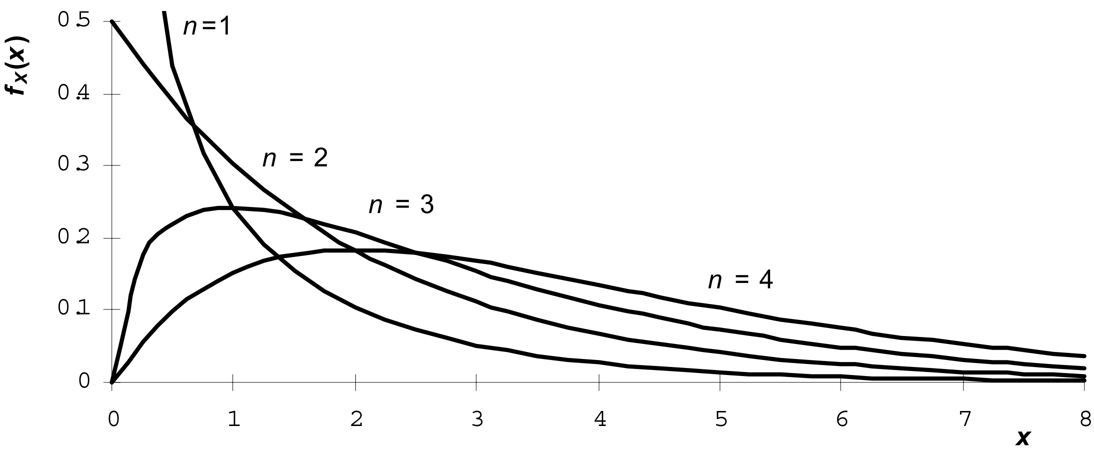
## 2.8.3 Η κατανομή $\chi^2$

Λέμε ότι η μεταβλητή $X$ ακολουθεί την κατανομή $\chi^2$ με $n$ βαθμούς ελευθερίας (συμβολικά $\chi^2(n)$) αν η πυκνότητα πιθανότητάς της είναι η συνάρτηση

$$f_X(x) = \frac{1}{2^{n/2}\,\Gamma(n/2)}\, x^{n/2-1}\, e^{-x/2} \qquad x \geq 0,\, n = 1, 2, \ldots \tag{2.64}$$

όπου Γ() η συνάρτηση γάμα (βλ. Παράρτημα 2.Β, σ. 45).

Πρόκειται για θετικά ασύμμετρη κατανομή (βλ. Σχ. 2.7) μιας παραμέτρου ($n$), η οποία αποτελεί ειδικότερη περίπτωση της *κατανομής γάμα* (βλ. κεφάλαιο 6). Η μέση τιμή και η διασπορά της κατανομής είναι $n$ και $2n$, αντίστοιχα. Οι συντελεστές ασυμμετρίας και κύρτωσης είναι $C_s = 2\sqrt{2/n}$ και $C_k = 3 + 12/n$.




**Σχ. 2.6** Παραδείγματα πυκνότητας πιθανότητας $\chi^2(n)$ για διάφορες τιμές του $n$.

Το ολοκλήρωμα της (2.64) δεν υπολογίζεται αναλυτικά. Έτσι για τους τυπικούς υπολογισμούς κατά κανόνα χρησιμοποιούνται πίνακες της κατανομής $\chi^2$ ή της κατανομής γάμα (βλ. Παράρτημα Π2 και Π4).

Η κατανομή $\chi^2$ δεν χρησιμοποιείται άμεσα για την περιγραφή υδρολογικών μεταβλητών. Αντί αυτής χρησιμοποιείται η γενικότερη κατανομή γάμα. Ωστόσο, η κατανομή $\chi^2$ έχει μεγάλη σημασία στη στατιστική. Η σημασία της προκύπτει από το ακόλουθο θεώρημα της στατιστικής: Αν οι μεταβλητές $X_i$ ($i$ = 1, …, $n$) ακολουθούν κανονική κατανομή $N(0, 1)$ τότε το άθροισμα των τετραγώνων τους

$$Q = \sum_{i=1}^{n} X_i^2 \tag{2.65}$$

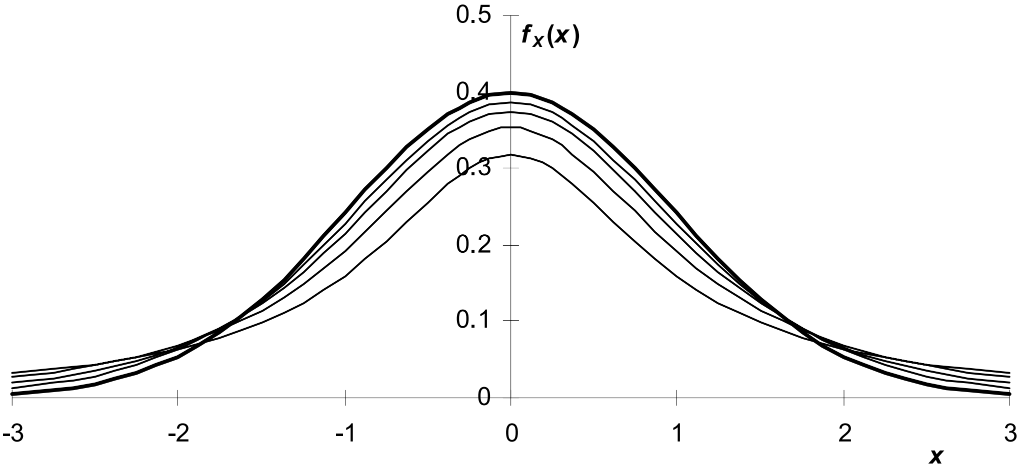
ακολουθεί κατανομή $\chi^2(n)$. Με συνδυασμό αυτού του θεωρήματος με το κεντρικό οριακό θεώρημα προκύπτει ότι για μεγάλο $n$ η $\chi^2(n)$ τείνει στην κανονική κατανομή.

### 2.8.4 Η κατανομή Student (*t*)

Λέμε ότι η μεταβλητή $X$ ακολουθεί την κατανομή Student ή $t$ με $n$ βαθμούς ελευθερίας (συμβολικά $t(n)$) αν η πυκνότητα πιθανότητάς της είναι η συνάρτηση

$$f_X(x) = \frac{\Gamma\left[(n+1)/2\right]}{\sqrt{\pi n}\,\Gamma(n/2)} \frac{1}{\sqrt{\left(1 + x^2/n\right)^{n+1}}} \qquad n = 1,2,\ldots \tag{2.66}$$

Πρόκειται για συμμετρική κατανομή (βλ. Σχ. 2.7) μιας παραμέτρου ($n$), με μέση τιμή 0 και διασπορά $n/(n-2)$. Για μεγάλες τιμές του $n$ ($\geq 30$) πρακτικώς ταυτίζεται με την κανονική κατανομή.




**Σχ. 2.7** Παραδείγματα πυκνότητας πιθανότητας $t(n)$ για $n$ = 1, 2, 4 και 8 (συνεχείς λεπτές γραμμές από κάτω προς τα πάνω), σε σύγκριση και με την τυποποιημένη κανονική κατανομή $N(0, 1)$ (χοντρή γραμμή).

Το ολοκλήρωμα της (2.66) δεν υπολογίζεται αναλυτικά. Έτσι για τους τυπικούς υπολογισμούς κατά κανόνα χρησιμοποιούνται πίνακες της κατανομής $t$ (βλ. Παράρτημα Π3).

Η κατανομή $t$ δεν χρησιμοποιείται άμεσα για την περιγραφή υδρολογικών μεταβλητών αλλά έχει μεγάλη σημασία στη στατιστική. Η σημασία της προκύπτει από το ακόλουθο θεώρημα της στατιστικής: Αν οι τυχαίες μεταβλητές $Z$ και $W$ είναι ανεξάρτητες και ακολουθούν κατανομές $N(0, 1)$ και $\chi^2(n)$, αντίστοιχα, τότε ο λόγος

$$T = \frac{Z}{\sqrt{W/n}} \tag{2.67}$$

ακολουθεί κατανομή $t(n)$.

**Παράρτημα 2.A: Σχέση κεντρικών ροπών και ροπών περί την αρχή**
Οι κεντρικές ροπές συνδέονται με τις αντίστοιχες ροπές περί την αρχή με τις σχέσεις

$$\sigma_X^2 = m_X^{(2)} - m_X^2$$

$$\mu_X^{(3)} = m_X^{(3)} - 3m_X^{(2)} m_X + 2m_X^3$$

$$\mu_X^{(4)} = m_X^{(4)} - 4m_X^{(3)} m_X + 6m_X^{(2)} m_X^2 - 3m_X^4 \qquad (2.68)$$

$$\vdots$$

$$\mu_X^{(r)} = m_X^{(r)} - \binom{r}{1} m_X^{(r-1)} m_X + \cdots + (-1)^j \binom{r}{j} m_X^{(r-j)} m_X^j + \cdots (-1)^r m_X^{(0)} m_X^r$$

Ισοδύναμες με τις παραπάνω είναι οι ακόλουθες σχέσεις, που διευκολύνουν τον υπολογισμό των ροπών περί την αρχή τάξης δύο έως τέσσερα, αν είναι γνωστές οι αντίστοιχες κεντρικές ροπές και η μέση τιμή:

$$m_X^{(2)} = \sigma_X^2 + m_X^2$$

$$m_X^{(3)} = \mu_X^{(3)} + 3\sigma_X^2 m_X + m_X^3 \qquad (2.69)$$

$$m_X^{(4)} = \mu_X^{(4)} + 4\mu_X^{(3)} m_X + 6\sigma_X^2 m_X^2 + m_X^4$$

**Παράρτημα 2.B: Ιδιότητες της συνάρτησης Γ( )**
Η συνάρτηση Γ( ) που υπεισέρχεται στον τύπο της κατανομής $\chi^2$ καθώς και άλλων κατανομών (γάμα†, *t* κ.ά.) ορίζεται με βάση το ακόλουθο ολοκλήρωμα:

---

† Δεν πρέπει να συγχέουμε τη συνάρτηση γάμα με τη συνάρτηση κατανομής γάμα.

$$\Gamma(a) = \int_0^\infty y^{a-1} e^{-y} dy \tag{2.70}$$

Το ολοκλήρωμα συγκλίνει για $a > 0$. Παρακάτω δίνονται μερικές χαρακτηριστικές ιδιότητες της συνάρτησης.

$$\begin{array}{llll} \Gamma(1) = 1 & \Gamma(1/2) = \sqrt{\pi} & \Gamma(a+1) = a\Gamma(a) & \\ \Gamma(n+1) = n! & \Gamma\left(n+\frac{1}{2}\right) = \left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\cdots\frac{3}{2}\sqrt{\pi} & n = 1,2,\ldots & \end{array} \tag{2.71}$$