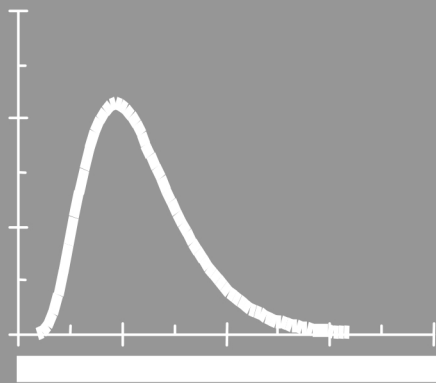


Δημήτρης Κουτσογιάννης
Εθνικό Μετσόβιο Πολυτεχνείο
Τομέας Υδατικών Πόρων

ΣΤΑΤΙΣΤΙΚΗ ΥΔΡΟΛΟΓΙΑ



Ελληνικά Ακαδημαϊκά Ηλεκτρονικά
Συγγράμματα και Βοηθήματα
www.kallipos.gr

HEALLINK
Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών

ΣΤΑΤΙΣΤΙΚΗ ΥΔΡΟΛΟΓΙΑ

Δημήτρης Κουτσογιάννης
Επίκουρος Καθηγητής
Τομέας Υδατικών Πόρων
Εθνικό Μετσόβιο Πολυτεχνείο

Έκδοση 5
Αθήνα 2016



**Ελληνικά Ακαδημαϊκά Ηλεκτρονικά
Συγγράμματα και Βοηθήματα**
www.kallipos.gr

HEALLINK
Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών

ΣΤΑΤΙΣΤΙΚΗ ΥΔΡΟΛΟΓΙΑ

Έκδοση 5

Copyright © Δ. Κουτσογιάννης, 1995, 1996, 1997, 1998, 2016.

Δεν επιτρέπεται η αναπαραγωγή του βιβλίου, τμηματική ή ολική, σε οποιαδήποτε μορφή, χωρίς την έγγραφη άδεια του συγγραφέα.

ISBN: 978-960-603-507-4

Πληκτρολόγηση και επεξεργασία κειμένου: Δ. Κουτσογιάννης

Επιμέλεια κειμένου: Δ. Κουτσογιάννης

Σχέδια: Δ. Κουτσογιάννης

Μακέτα εξωφύλλου: Α. Πατρικίου

Εκτύπωση Offset και βιβλιοδεσία: Τυπογραφείο Εθνικού Μετσοβίου
Πολυτεχνείου

Τεχνική επιμέλεια εκτύπωσης και βιβλιοδεσίας: Γ. Καραγκιοζόπουλος,
Ν. Γκάνης, Α. Μιχαλόπουλος

Αφιερώνεται στην Αννούσκα

Πρόλογος

Οι σημειώσεις αυτές απευθύνονται, ως ένα συμπληρωματικό βοήθημα, στους φοιτητές που παρακολουθούν το εισαγωγικό μάθημα της τεχνικής υδρολογίας. Καλύπτουν μόνο ένα μέρος της ύλης του μαθήματος και συγκεκριμένα τις πιθανοθεωρητικές και στατιστικές μεθόδους που χρησιμοποιούνται στην τεχνική υδρολογία.

Θα ήθελα να κάνω απ' την αρχή μια προειδοποίηση προς τους φοιτητές: το αντικείμενο αυτών των σημειώσεων δεν είναι εύκολο. Η θεωρία πιθανοτήτων και η στατιστική είναι από τους πιο δύσκολους κλάδους των μαθηματικών. Επιπλέον, η φυσική υδρολογική πραγματικότητα με την μεγάλη ποικιλία φαινομένων, κατά κανόνα χαοτικού χαρακτήρα, δύσκολα τυποποιείται και προσαρμόζεται στα αφαιρετικά μαθηματικά μοντέλα της θεωρίας πιθανοτήτων. Ωστόσο, η εφαρμογή των πιθανοθεωρητικών και στατιστικών μεθόδων στην τεχνική υδρολογία αποτελεί το μόνο επιστημονικά άρτιο τρόπο ποσοτικοποίησης της αβεβαιότητας που διέπει τα υδρολογικά φαινόμενα και των κινδύνων για τα έργα του μηχανικού που αυτή συνεπάγεται.

Στην παρουσίαση της ύλης απέφυγα την εύκολη μέθοδο της προσανατολισμένης προς την εφαρμογή τεχνητής απλούστευσης του αντικειμένου. Αντίθετα, επιδίωξα να παρουσιάσω τα θέματα σε βάθος και συνεκτικά. Έτσι, το κείμενο αυτό αντανακλά την πραγματική δυσκολία του επιστημονικού αντικειμένου που διαπραγματεύεται. Για το λόγο αυτό είναι πολύ πιθανό (ίσως και στατιστικώς βέβαιο) ότι η πρώτη ανάγνωσή του δεν θα ενθουσιάσει το φοιτητή. Ελπίζω όμως, ότι η βαθύτερη μελέτη

του θα οδηγήσει σε εξοικείωση με τις πολλαπλές αρετές της τεχνικής υδρολογίας.

Για λόγους πληρότητας και αυτοτέλειας έχω συμπεριλάβει (κεφάλαια 2 και 3) μια εισαγωγή στις απαραίτητες έννοιες της πιθανοθεωρίας και της στατιστικής. Η ανάπτυξη των εννοιών αυτών είναι περιληπτική και σαφώς προσανατολισμένη στα προβλήματα της υδρολογίας. Για πληρέστερη και αυστηρότερη μαθηματική ανάπτυξη αυτού του θεωρητικού υποβάθρου, οι ενδιαφερόμενοι φοιτητές παραπέμπονται στα βιβλία θεωρίας πιθανοτήτων και στατιστικής (βλ. βιβλιογραφικές αναφορές στο τέλος του βιβλίου). Οι εξοικειωμένοι με αυτές τις έννοιες φοιτητές μπορούν να παραλείψουν τη μελέτη αυτών των κεφαλαίων, ή να περιοριστούν στην ανάγνωση των παραδειγμάτων, τα οποία βοηθούν στη σύνδεση της πιθανοθεωρίας και στατιστικής με την τεχνική υδρολογία.

Δεν θα αποκρύψω από τους φοιτητές ότι η σημερινή τάση των υδρολογικών επιστημών στο διεθνή χώρο, είναι προς την κατεύθυνση της πληρέστερης κατανόησης των φυσικών υδρολογικών διεργασιών, εις βάρος ίσως της αυστηρότητας στη μαθηματική αναπαράστασή τους. Αυτή η τάση αντανακλάται και στα πιο σύγχρονα καθιερωμένα συγγράμματα υδρολογίας, στα οποία η αναφορά στις στατιστικές μεθοδολογίες είναι σχετικά περιορισμένη και συχνά έχει εμπειρικό χαρακτήρα. Αυτό το κείμενο δεν ακολουθεί αυτή την τάση, αλλά αντίθετα προβάλλει τα μαθηματικό χαρακτήρα του αντικειμένου που διαπραγματεύεται προσπαθώντας να διατηρεί ένα βαθμό μαθηματικής αυστηρότητας. Δεν νομίζω ότι αυτό είναι μια άσκοπη θεωρητικοποίηση που μειώνει τον τεχνολογικό προσανατολισμό του κειμένου. Αντίθετα, η ακολουθούμενη προσέγγιση νομίζω ότι είναι βαθειά τεχνολογική, δεδομένου ότι καλύπτει τις πραγματικές ανάγκες των εφαρμογών, οι οποίες δεν απαιτούν κατ' ανάγκη την πλήρη κατανόηση των πολύπλοκων φυσικών μηχανισμών που κρύβονται πίσω από τις υδρολογικές διεργασίες. Άλλωστε, “η υψηλή τεχνολογία είναι η μαθηματική τεχνολογία”.[†]

[†] Διαπίστωση του Εθνικού Συμβουλίου Έρευνας των ΗΠΑ (βλ. και άρθρο του Γ. Ν. Παντελίδη, “Τα Πολυτεχνεία πρέπει να εκπαιδεύουν μηχανικούς για την τεχνολογία του μέλλοντος”, Πρακτικά του διημέρου *Το γενικό επιστημονικό υπόβαθρο στις σπουδές του μηχανικού του ΕΜΠ και ο ρόλος του Γενικού Τμήματος*, ΕΜΠ, Αθήνα, 1994).

Τελειώνοντας αυτό τον πρόλογο θα ήθελα να γράψω και μερικά λόγια για τη δική μου πρώτη (αντιφατική) εμπειρία με την τεχνική υδρολογία, που απέκτησα το 1976, ως τριτοετής φοιτητής της Σχολής Πολιτικών Μηχανικών του ΕΜΠ. Ομολογώ ότι η πρώτη μου εντύπωση από το μάθημα της υδρολογίας ήταν τελείως αρνητική. Είχα κάνει πολλές συζητήσεις με τους συμφοιτητές μου στις οποίες είχα εκφράσει την άποψη ότι το μάθημα αυτό ήταν εντελώς άχρηστο για έναν πολιτικό μηχανικό. Οι περισσότεροι συμφωνούσαν μαζί μου· θυμάμαι μόνο τον Π. Παπανικολάου να έχει τεκμηριωμένη αντίθετη άποψη: ότι η υδρολογία αποτελεί ένα απαραίτητο κρίκο στην αλυσίδα των μαθημάτων γύρω από το νερό. Το βασικό έναυσμα για να αλλάξω γνώμη ήταν το διάβασμα του βιβλίου του Θ. Ξανθόπουλου “Μαθήματα στατιστικής υδρολογίας” (έκδοση Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, 1975), το οποίο μας μοιράστηκε περί το τέλος του εξαμήνου. Διαβάζοντας το βιβλίο αυτό άρχισα να καταλαβαίνω το επιστημονικό βάθος και τη γοητεία της τεχνικής υδρολογίας, τη χρησιμότητα των μεθόδων που χρησιμοποιεί σε εφαρμογές και άλλων επιστημονικών και τεχνολογικών κλάδων, αλλά και τη συμβολή της στην κατανόηση και μαθηματική περιγραφή των πιο πολύπλοκων φυσικών διεργασιών. Στην πορεία μου η υδρολογία έγινε η βασική δίοδος επικοινωνίας μου με την επιστημονική και τεχνική κοινότητα και τους φοιτητές. Θα ήθελα, λοιπόν, να ευχαριστήσω τους καθηγητές μου, τωρινούς συνεργάτες, που με μύησαν σε αυτή την κατεύθυνση: τον Θ. Ξανθόπουλο, τον Μ. Αφτιά, τη Μ. Μιμίκου και τον Η. Βασιλόπουλο. Ακόμη, θα ήθελα να αναφέρω στο σημείο αυτό και τον αείμνηστο Ι. Παπαδάκη, τότε επιμελητή της Έδρας Υδρολογίας και Υδραυλικών Έργων, που ακόμη θυμάμαι τον ιδιότυπο ορισμό του για την τεχνική υδρολογία: “είναι η επιστήμη που σου δίνει το Q ”.

Τέλος, ευχαριστώ θερμά τους συνεργάτες Ηλία Βασιλόπουλο, Αντώνη Χριστοφίδη και Νίκο Μαμάση για τις παρατηρήσεις τους στο κείμενο και τις υποδείξεις τους.

Αθήνα, Φεβρουάριος 1996
Δημήτρης Κουτσογιάννης

ΠΕΡΙΕΧΟΜΕΝΑ

| | |
|---|----------|
| Κεφάλαιο 1 Σύνδεση τεχνικής υδρολογίας και πιθανοθεωρίας | 1 |
| Μη προβλεψιμότητα και τύχη..... | 3 |
| Κεφάλαιο 2 Εισαγωγικές έννοιες θεωρίας πιθανοτήτων | 9 |
| 2.1 Αξιοματική θεμελίωση της θεωρίας πιθανοτήτων..... | 9 |
| 2.2 Τυχαία μεταβλητή και συνάρτηση κατανομής..... | 11 |
| Εφαρμογή 2.2..... | 13 |
| 2.3 Ανεξάρτητα και εξαρτημένα γεγονότα, δεσμευμένη πιθανότητα..... | 15 |
| Εφαρμογή 2.3..... | 16 |
| 2.4 Αναμενόμενες τιμές και παράμετροι κατανομών..... | 18 |
| 2.4.1 Παράμετροι θέσης..... | 20 |
| 2.4.2 Παράμετροι διασποράς | 20 |
| 2.4.3 Παράμετροι ασυμμετρίας..... | 21 |
| 2.4.4 Παράμετροι κύρτωσης | 21 |
| Εφαρμογή 2.4..... | 23 |
| Χρονική κλίμακα και σχήμα κατανομής..... | 24 |
| 2.5 Αλλαγή μεταβλητής | 25 |
| Εφαρμογή 2.5.α..... | 26 |
| Εφαρμογή 2.5.β..... | 27 |

| | |
|--|-----------|
| 2.6 Από κοινού και περιθώριες συναρτήσεις κατανομής..... | 27 |
| 2.6.1 Αναμενόμενες τιμές - ροπές..... | 29 |
| 2.6.2 Ανεξαρτησία μεταβλητών | 30 |
| 2.6.3 Αθροίσματα μεταβλητών | 30 |
| Εφαρμογή 2.6.3..... | 31 |
| 2.6.4 Πολλές μεταβλητές | 33 |
| 2.7 Η έννοια της στοχαστικής ανέλιξης | 34 |
| 2.7.1 Συνάρτηση κατανομής | 35 |
| 2.7.2 Ροπές | 35 |
| 2.7.3 Στασιμότητα | 36 |
| 2.7.4 Εργοδικότητα | 36 |
| 2.8 Το κεντρικό οριακό θεώρημα και μερικές συναρτήσεις κατανομής..... | 37 |
| 2.8.1 Το κεντρικό οριακό θεώρημα και η σημασία του..... | 38 |
| 2.8.2 Η κανονική κατανομή | 41 |
| Εφαρμογή 2.8.2..... | 42 |
| 2.8.3 Η κατανομή χ^2 | 42 |
| 2.8.4 Η κατανομή Student (t) | 43 |
| Παράρτημα 2.A: Σχέση κεντρικών ροπών και ροπών περί την αρχή | 45 |
| Παράρτημα 2.B: Ιδιότητες της συνάρτησης $\Gamma(\cdot)$ | 45 |
| Κεφάλαιο 3 Εισαγωγικές έννοιες στατιστικής..... | 47 |
| 3.1 Εννοιολογία και ορισμοί | 48 |
| 3.1.1 Δείγμα..... | 48 |
| 3.1.2 Στατιστική συνάρτηση | 49 |
| 3.1.3 Εκτιμήτριες και εκτιμήσεις | 50 |
| 3.1.4 Εκτίμηση διαστήματος και όρια εμπιστοσύνης | 51 |
| 3.2 Τυπικές σημειακές εκτιμήτριες | 53 |
| 3.2.1 Γενικά για τις εκτιμήτριες ροπών..... | 53 |
| 3.2.2 Δειγματική μέση τιμή..... | 55 |
| 3.2.3 Διασπορά και τυπική απόκλιση..... | 55 |
| 3.2.4 Τρίτη κεντρική ροπή και συντελεστής ασυμμετρίας | 58 |
| 3.2.5 Συνδιασπορά και συσχέτιση..... | 60 |
| 3.3 Τυπικά όρια εμπιστοσύνης..... | 62 |
| 3.3.1 Μέση τιμή - γνωστή διασπορά πληθυσμού..... | 62 |

| | |
|--|----|
| 3.3.2 Μέση τιμή - άγνωστη διασπορά πληθυσμού..... | 64 |
| Εφαρμογή 3.3.2..... | 65 |
| 3.3.3 Διασπορά και τυπική απόκλιση..... | 67 |
| Εφαρμογή 3.3.3..... | 67 |
| 3.3.4 Ποσοστημόριο κανονικής κατανομής – Τυπικό σφάλμα | 68 |
| Εφαρμογή 3.3.4..... | 70 |
| 3.3.5 Συντελεστής συσχέτισης | 71 |
| 3.4 Εκτίμηση παραμέτρων συναρτήσεων κατανομής..... | 72 |
| 3.4.1 Η μέθοδος των ροπών | 73 |
| Υπολογισμός παραμέτρων της κανονικής κατανομής με τη μέθοδο των ροπών | 74 |
| 3.4.2 Η μέθοδος της μέγιστης πιθανοφάνειας..... | 75 |
| Υπολογισμός παραμέτρων της κανονικής κατανομής με τη μέθοδο της μέγιστης πιθανοφάνειας..... | 76 |
| 3.5 Έλεγχος υποθέσεων..... | 77 |
| 3.5.1 Ορολογία | 78 |
| 3.5.2 Διαδικασία ελέγχου..... | 80 |
| 3.5.3 Έλεγχος σημαντικότητας του συντελεστή συσχέτισης..... | 81 |
| Εφαρμογή 3.5.3..... | 83 |

Κεφάλαιο 4 Ειδικές έννοιες θεωρίας πιθανοτήτων στην υδρολογία

85

| | |
|---|-----|
| 4.1 Πιθανοθεωρητική περιγραφή υδρολογικών διεργασιών..... | 85 |
| 4.1.1 Ανεξίξεις σε διακριτό χρόνο | 87 |
| 4.1.2 Ανεξίξεις ακροτάτων..... | 90 |
| 4.1.3 Βασικές απλουστευτικές παραδοχές | 92 |
| 4.1.4 Τελικό συμπέρασμα | 94 |
| Εφαρμογή 4.1..... | 94 |
| 4.2 Η έννοια της περιόδου επαναφοράς | 96 |
| Εφαρμογή 4.2..... | 98 |
| 4.3 Η έννοια της διακινδύνευσης | 98 |
| Εφαρμογή 4.3..... | 101 |

| | |
|--|------------|
| Κεφάλαιο 5 Τυπική στατιστική ανάλυση μιας υδρολογικής μεταβλητής..... | 103 |
| 5.1 Στατιστικά χαρακτηριστικά δείγματος..... | 104 |
| Εφαρμογή 5.1..... | 107 |
| 5.2 Ιστογράμματα..... | 112 |
| Εφαρμογή 5.2..... | 112 |
| 5.3 Εμπειρική συνάρτηση κατανομής..... | 113 |
| 5.3.1 Κλασική εμπειρική συνάρτηση κατανομής..... | 114 |
| 5.3.2 Θέσεις σχεδίασης..... | 114 |
| 5.3.3 Χαρτιά πιθανότητας..... | 117 |
| Εφαρμογή 5.3..... | 119 |
| 5.4 Επιλογή και προσαρμογή θεωρητικής συνάρτησης κατανομής..... | 120 |
| Ενδείξεις καταλληλότητας της κανονικής κατανομής για υδρολογικές μεταβλητές..... | 122 |
| Εφαρμογή 5.4..... | 124 |
| 5.5 Έλεγχος προσαρμογής συνάρτησης κατανομής..... | 124 |
| 5.5.1 Η δοκιμή χ^2 | 124 |
| Εφαρμογή 5.5..... | 127 |
| 5.6 Στατιστική πρόγνωση..... | 129 |
| Εφαρμογή 5.6..... | 129 |
| Κεφάλαιο 6 Τυπικές συναρτήσεις κατανομής στην τεχνική υδρολογία..... | 131 |
| 6.1 Κανονική κατανομή και μετασχηματισμοί της..... | 131 |
| 6.1.1 Κανονική κατανομή..... | 131 |
| Τυπικοί υπολογισμοί..... | 133 |
| Εκτίμηση παραμέτρων..... | 134 |
| Τυπικό σφάλμα ποσοστημορίου και όρια εμπιστοσύνης..... | 134 |
| Χαρτί κανονικής κατανομής..... | 134 |
| 6.1.2 Λογαριθμοκανονική κατανομή δύο παραμέτρων..... | 135 |
| Τυπικοί υπολογισμοί..... | 136 |
| Εκτίμηση παραμέτρων..... | 136 |
| Τυπικό σφάλμα ποσοστημορίου και όρια εμπιστοσύνης..... | 137 |

| | |
|---|-----|
| Χαρτί λογαριθμοκανονικής κατανομής | 138 |
| Εφαρμογή 6.1.2..... | 138 |
| 6.1.3 Λογαριθμοκανονική κατανομή τριών παραμέτρων (Galton) | 141 |
| Τυπικοί υπολογισμοί..... | 141 |
| Εκτίμηση παραμέτρων | 142 |
| 6.2 Η ομάδα των κατανομών γάμα..... | 143 |
| 6.2.1 Εκθετική κατανομή | 143 |
| 6.2.2 Κατανομή γάμα δύο παραμέτρων | 144 |
| Τυπικοί υπολογισμοί..... | 146 |
| Εκτίμηση παραμέτρων..... | 147 |
| Τυπικό σφάλμα ποσοστημορίου και όρια εμπιστοσύνης | 147 |
| Χαρτί κατανομής γάμα | 148 |
| Εφαρμογή 6.2.2..... | 148 |
| 6.2.3 Κατανομή γάμα τριών παραμέτρων (Pearson III)..... | 150 |
| 6.2.4 Κατανομή Log-Pearson III..... | 151 |
| Τυπικοί υπολογισμοί..... | 152 |
| Εκτίμηση παραμέτρων..... | 153 |
| Χαρτί κατανομής Log-Pearson III | 153 |
| 6.3 Ασυμπτωτικές κατανομές ακροτάτων | 153 |
| 6.3.1 Κατανομή μεγίστων τύπου I (Gumbel)..... | 155 |
| Τυπικοί υπολογισμοί..... | 155 |
| Εκτίμηση παραμέτρων..... | 156 |
| Τυπικό σφάλμα ποσοστημορίου και όρια εμπιστοσύνης | 157 |
| Χαρτί κατανομής Gumbel..... | 158 |
| Εφαρμογή 6.3.1 | 158 |
| 6.3.2 Κατανομή ελαχίστων τύπου I (Gumbel) | 160 |
| 6.3.3 Κατανομή ελαχίστων τύπου III (Weibull) | 162 |
| Τυπικοί υπολογισμοί..... | 163 |
| Εκτίμηση παραμέτρων..... | 163 |
| Χαρτί κατανομής Weibull..... | 164 |
| Εφαρμογή 6.3.3..... | 164 |
| 6.4 Άλλες χρήσιμες κατανομές..... | 167 |
| 6.4.1 Κατανομή βήτα δύο παραμέτρων..... | 167 |

| | |
|--|------------|
| 6.4.2 Κατανομή Pareto | 168 |
| Παράρτημα 6.A: Αριθμητικές προσεγγίσεις της κανονικής κατανομής..... | 170 |
| Παράρτημα 6.B: Αριθμητικές προσεγγίσεις της κατανομής γάμα | 171 |
| Κεφάλαιο 7 Ανάλυση δύο τυχαίων μεταβλητών - Εκτιμήσεις ελάχιστων τετραγώνων | 177 |
| 7.1 Τυπική γραμμική παλινδρόμηση..... | 179 |
| 7.1.1 Προσδιοριστική αντιμετώπιση..... | 179 |
| Γεωμετρική ερμηνεία..... | 179 |
| Αναλυτική προσέγγιση | 180 |
| Συντελεστής προσδιορισμού..... | 182 |
| Αντιστροφή του γραμμικού νόμου | 185 |
| 7.1.2 Πιθανοτική αντιμετώπιση | 187 |
| Εξισώσεις εκτίμησης παραμέτρων..... | 187 |
| Η έννοια της παλινδρόμησης | 189 |
| Ροπές του σφάλματος εκτίμησης..... | 190 |
| Ροπές της εκτιμήτριας..... | 191 |
| Συντελεστής προσδιορισμού..... | 191 |
| Εφαρμογή 7.1..... | 192 |
| 7.2 Άλλες γραμμικές εκτιμήσεις δύο μεταβλητών..... | 200 |
| 7.2.1 Ομογενής ευθεία..... | 200 |
| 7.2.2 Οργανική συσχέτιση..... | 202 |
| Εφαρμογή 7.2..... | 205 |
| 7.3 Γενίκευση της γραμμικής παλινδρόμησης | 207 |
| 7.3.1 Γραμμική παλινδρόμηση πολλών μεταβλητών | 207 |
| 7.3.2 Μη γραμμική παλινδρόμηση με γραμμικούς συντελεστές..... | 209 |
| 7.3.3 Γραμμικοποίηση με μετασχηματισμούς μεταβλητών | 209 |
| Εφαρμογή 7.3.3..... | 210 |
| 7.4 Γενική παλινδρόμηση..... | 213 |
| 7.4.1 Μεταβλητές με τυχούσα συνάρτηση κατανομής | 213 |
| 7.4.2 Μεταβλητές με κανονική κατανομή..... | 215 |
| Εισαγωγικές έννοιες για τη διδιάστατη κανονική κατανομή..... | 215 |

| | |
|---|-----|
| Η γραμμική παλινδρόμηση ως συνέπεια της γενικής παλινδρόμησης..... | 215 |
| Ομοσκεδαστικότητα..... | 219 |
| Εφαρμογή 7.4.2..... | 219 |
| 7.5 Όρια εμπιστοσύνης και δοκιμές σημαντικότητας | 222 |
| Εφαρμογή 7.5..... | 225 |
| 7.6 Συμπλήρωση και επέκταση δειγμάτων..... | 227 |
| 7.6.1 Χρήση της γραμμικής παλινδρόμησης χωρίς όρο σφάλματος..... | 230 |
| 7.6.2 Χρήση της οργανικής συσχέτισης..... | 233 |
| 7.6.3 Χρήση της ομογενούς ευθείας..... | 234 |
| 7.6.4 Χρήση της γραμμικής παλινδρόμησης με όρο σφάλματος..... | 234 |
| Γεννήτριες τυχαίων αριθμών ομοιόμορφης και κανονικής κατανομής..... | 236 |
| Εφαρμογή 7.6..... | 238 |
| Παράρτημα 7Α: Παραγωγή των εκτιμήσεων του διευρυμένου δείγματος και σχετικές αποδείξεις | 243 |

Κεφάλαιο 8 Ανάλυση τυχαίας μεταβλητής εξαρτημένης από παράμετρο - Όμβριες καμπύλες249

| | |
|--|-----|
| 8.1 Έννοια και τρόπος κατασκευής των μεταβλητών | 251 |
| Εφαρμογή 8.1..... | 254 |
| 8.2 Συμβατική στατιστική μέθοδος εξαγωγής όμβριων καμπυλών | 256 |
| 8.2.1 Μεμονωμένες καμπύλες για κάθε διάρκεια | 257 |
| Εφαρμογή 8.2.1..... | 258 |
| 8.2.2 Ενιαία ημιεμπειρική έκφραση όμβριων καμπυλών..... | 265 |
| Εφαρμογή 8.2.2..... | 267 |
| 8.3 Συνεπείς μέθοδοι εξαγωγής όμβριων καμπυλών..... | 270 |
| 8.3.1 Εναλλακτικοί τύποι συνάρτησης κατανομής | 271 |
| Κατανομή Gumbel..... | 271 |
| Κατανομή γάμα..... | 272 |
| Κατανομή Log Pearson III..... | 273 |
| Κατανομή Pareto..... | 273 |
| Εκθετική κατανομή..... | 274 |

| | |
|---|------------|
| 8.3.2 Εκτίμηση παραμέτρων | 275 |
| Εκτίμηση με ενοποίηση διαρκειών | 276 |
| Καθολική εκτίμηση..... | 278 |
| Εφαρμογή 8.3.2..... | 279 |
| 8.4 Συμπληρωματικά στοιχεία | 284 |
| 8.4.1 Ενοποίηση δειγμάτων διαφορετικών σταθμών | 284 |
| 8.4.2 Επίδραση της χρονικής ευκρίνειας..... | 286 |
| 8.4.3 Σχέση σειρών ετήσιων μεγίστων και σειρών υπεράνω κατωφλίου | 287 |
| 8.4.4 Όρια εμπιστοσύνης όμβριων καμπυλών | 288 |
| 8.5 Εξαγωγή όμβριων καμπυλών με χρήση δεδομένων από βροχόμετρα..... | 289 |
| 8.6 Γεωγραφική μεταβλητότητα των όμβριων καμπυλών | 293 |
| Παράρτημα Στατιστικοί πίνακες..... | 299 |
| Αναφορές..... | 307 |

Κεφάλαιο 1 Σύνδεση τεχνικής υδρολογίας και πιθανοθεωρίας

Υπάρχουν τρεις τουλάχιστον λόγοι για τους οποίους η πιθανοθεωρία και η στατιστική αποτελούν το βασικό μαθηματικό εργαλείο της τεχνικής υδρολογίας, σε βαθμό που να θεωρείται η τεχνική υδρολογία ως ιδεώδες πεδίο εφαρμογής αυτών των μαθηματικών κλάδων:

1. Η *τύχη* ενυπάρχει σε όλα τα φαινόμενα της υδρολογίας, ξεκινώντας από την καταιγίδα ή την ανομβρία και καταλήγοντας στην πλημμύρα και την ξηρασία. Έτσι οι νόμοι των πιθανοτήτων κυβερνούν σε μεγαλύτερο ή μικρότερο βαθμό αυτά τα φαινόμενα. Ο όρος *τύχη* χρησιμοποιείται εδώ με την έννοια της μη (αιτιοκρατικής) προβλεψιμότητας. Το θέμα αυτό αναλύεται περισσότερο αμέσως πιο κάτω, στο ένθετο εδάφιο *Μη προβλεψιμότητα και τύχη*.
2. Η υδρολογική πληροφορία απαρτίζεται από μετρήσεις υδρολογικών διεργασιών, η επεξεργασία των οποίων προϋποθέτει τη χρήση στατιστικών μεθόδων. Ο έλεγχος των σφαλμάτων των μετρήσεων και η συμπλήρωση των ελλείψεων ιστορικών δειγμάτων είναι δύο τυπικές περιπτώσεις που αντιμετωπίζονται με χρήση στατιστικών μεθόδων.
3. Στην τεχνική υδρολογία η λήψη τεχνικο-οικονομικών αποφάσεων, που αναφέρονται είτε στο σχεδιασμό έργων είτε στη λειτουργία τους, γίνεται πάντα υπό καθεστώς αβεβαιότητας. Η ορθολογικότερη προσέγγιση και η ποσοτικοποίηση της αβεβαιότητας επιτυγχάνεται με τη θεωρία πιθανοτήτων.

Οι εκτιμήσεις και προβλέψεις της τεχνικής υδρολογίας κατά κανόνα συνάγονται μετά από επεξεργασία της διαθέσιμης ιστορικής υδρολογικής πληροφορίας με πιθανοθεωρητικές και στατιστικές μεθόδους. Η ιστορική υδρολογική πληροφορία είναι ένα σύνολο μετρήσεων της φυσικής διεργασίας που ενδιαφέρει σε συγκεκριμένη περιοχή (π.χ. υδρολογική λεκάνη) ή θέση (π.χ. διατομή ποταμού). Δεν βασίζονται, λοιπόν, οι υδρολογικές εκτιμήσεις και προβλέψεις σε γενικούς νόμους παγκόσμιας ισχύος (όπως είναι π.χ. οι νόμοι της φυσικής). Αυτό, βεβαίως, δεν σημαίνει ότι η τεχνική υδρολογία αδιαφορεί για τους φυσικούς νόμους που διέπουν τις υδρολογικές διεργασίες. Αντίθετα, χρησιμοποιεί τη φυσική περιγραφή των διεργασιών αυτών για την κατανόησή τους, αλλά και για την ποσοτική περιγραφή τους όπου (και σε όποιο βαθμό) αυτό είναι δυνατό. Ωστόσο, θα πρέπει να τονίσουμε, η φυσική περιγραφή των διεργασιών χωρίς ιστορική υδρολογική πληροφορία αποτελεί ανεπαρκή βάση για τη συναγωγή εκτιμήσεων και προβλέψεων.

Είναι βεβαίως αυτονόητο ότι η λήψη των κατάλληλων μετρήσεων και η κατάρτιση ενός επαρκούς υδρολογικού δείγματος προηγείται και αποτελεί τη βάση της στατιστικής επεξεργασίας και συναγωγής συμπερασμάτων. Η θεωρία πιθανοτήτων και η στατιστική δεν μπορούν να αντιμετωπίσουν την έλλειψη υδρολογικής πληροφορίας. Επίσης, είναι προφανές ότι, όσο πιο μεγάλο είναι το μέγεθος του υδρολογικού δείγματος και όσο μεγαλύτερη είναι η αξιοπιστία των μετρήσεων* τόσο πιο αξιόπιστες είναι και οι εκτιμήσεις και οι προβλέψεις. Ένας μεγάλος κλάδος της τεχνικής υδρολογίας βασισμένος στη στατιστική, ασχολείται με τον έλεγχο, αλλά και τη βελτίωση, της αξιοπιστίας της υδρολογικής πληροφορίας.

Η εφαρμογή των μεθόδων της πιθανοθεωρίας και στατιστικής στην τεχνική υδρολογία δεν είναι μια τυποποιημένη και αυτόματη διαδικασία. Σε κάθε πρόβλημα που αντιμετωπίζουμε χρειάζεται μια αρχική εξερεύνηση και κατανόηση των δεδομένων και των ιδιοτήτων τους. Αυτή χρησιμοποιεί διάφορες τεχνικές, συχνά εμπειρικές, όπως κατάλληλη πινακοποίηση, συμπύκνωση, γεωμετρική (γραφική) απεικόνισή τους, κ.ά.

* Η αξιοπιστία των μετρήσεων εξαρτάται από πολλούς παράγοντες, όπως από την καταλληλότητα της θέσης μέτρησης, την καταλληλότητα και συντήρηση των οργάνων, την εμπειρία του προσωπικού που εκτελεί μετρήσεις, τον τρόπο οργάνωσης της αρμόδιας υδρολογικής υπηρεσίας, κ.ά.

Η διαδικασία αυτή, που σήμερα διευκολύνεται από τη χρήση ηλεκτρονικών υπολογιστών, είναι γνωστή με τον όρο *εξερευνητική ανάλυση δεδομένων* (Hirsch et al., 1993). Αλλά και στο στάδιο της σύνθεσης, δηλαδή της συναγωγής συμπερασμάτων (εκτιμήσεων ή προβλέψεων) κάθε άλλο παρά τυποποιημένη είναι η διαδικασία. Έτσι λοιπόν και στην τεχνική υδρολογία, όπως και σε άλλες επιστήμες του μηχανικού, η επιτυχής αντιμετώπιση των προβλημάτων προϋποθέτει, πέρα από τη γνώση των μαθηματικών εργαλείων, ευρύτητα πνεύματος, εμπειρία, αλλά και φαντασία και διαίσθηση.

Μη προβλεψιμότητα και τύχη

Ένα φυσικό φαινόμενο που διέπεται από προσδιοριστικούς (ντετερμινιστικούς) νόμους είναι πάντα προβλέψιμο; Η φυσική μας διαίσθηση απαντά θετικά στο ερώτημα αυτό, με την προϋπόθεση ότι είναι γνωστοί οι νόμοι που διέπουν το φαινόμενο. Ίσως γιατί η διαίσθησή μας βασίζεται σε συστήματα που διέπονται από γραμμικούς νόμους. Ωστόσο η πραγματικότητα για την πλειονότητα των φυσικών συστημάτων είναι διαφορετική. Τις τελευταίες δεκαετίες έχει διερευνηθεί μια μεγάλη ποικιλία μη γραμμικών συστημάτων, φυσικών ή μαθηματικών, τα οποία εμφανίζουν μη περιοδική, ακανόνιστη, απρόβλεπτη, τυχαία συμπεριφορά, αν και διέπονται από προσδιοριστικούς νόμους. Η συμπεριφορά αυτή έχει γίνει γνωστή με τον όρο *προσδιοριστικό χάος*. Τα υδρολογικά και μετεωρολογικά συστήματα είναι τυπικοί αντιπρόσωποι αυτής της συμπεριφοράς, από τη μικροσκοπική κλίμακα (π.χ. τύρβη) μέχρι τη μακροσκοπική (πλημμύρες, ξηρασίες κτλ.).

Ο πρώτος που αναγνώρισε αυτή τη συμπεριφορά των μη γραμμικών συστημάτων ήταν ο διάσημος φυσικομαθηματικός Henri Poincaré που πριν από 100 περίπου χρόνια έγραφε στο *Επιστήμη και μέθοδος* (από τον Stewart, 1990):

“Μια πολύ ανεπαίσθητη αιτία, η οποία μας διαφεύγει, καθορίζει ένα αξιοσημείωτο φαινόμενο που δεν μπορούμε να μη δούμε, και τότε λέμε ότι το φαινόμενο αυτό οφείλεται στην τύχη. Αν μπορούσαμε να ξέρουμε επακριβώς τους νόμους της φύσης και την κατάσταση του σύμπαντος στην αρχική του στιγμή, θα μπορούσαμε να προβλέψουμε επακριβώς την κατάσταση αυτού του ίδιου του σύμπαντος σε μια μεταγενέστερη χρονική στιγμή. Αλλά, ακόμα και αν οι φυσικοί νόμοι δεν είχαν άλλα μυστικά από εμάς, θα μπορούσαμε να ξέρουμε την αρχική κατάσταση μόνο κατά προσέγγιση. Αν αυτό μας επιτρέπει να προβλέψουμε τη μεταγενέστερη κατάσταση με τον ίδιο βαθμό προσέγγισης, αυτό αρκεί

για να πούμε ότι το φαινόμενο είχε προβλεφθεί, ότι υπόκειται σε νόμους. Όμως, το ζήτημα δεν είναι πάντοτε έτσι: υπάρχει περίπτωση οι πολύ λεπτές διαφορές στις αρχικές συνθήκες να παράγουν πολύ μεγαλύτερες διαφορές στα τελικά φαινόμενα, ένα ελάχιστο σφάλμα στην αρχή να προκαλεί ένα τεράστιο σφάλμα στο τέλος. Η πρόβλεψη τότε γίνεται αδύνατη κι έτσι έχουμε το φαινόμενο της τύχης.

”Γιατί άραγε οι μετεωρολόγοι συναντούν τόσο μεγάλες δυσκολίες στην πρόβλεψη του καιρού; Γιατί ακόμη οι βροχές και οι καταιγίδες μάς φαίνονται ότι έρχονται στην τύχη, με αποτέλεσμα πολλοί άνθρωποι να θεωρούν εντελώς φυσικό να προσεύχονται για να μην έρθει βροχή ή για να σταματήσει, ενώ τους φαίνεται γελοίο να προσευχηθούν για μια έκλειψη; Παρατηρούμε ότι οι μεγάλες διαταραχές συμβαίνουν γενικά σε περιοχές όπου η ατμόσφαιρα βρίσκεται σε ασταθή ισορροπία. Οι μετεωρολόγοι γνωρίζουν ότι αυτή η ισορροπία είναι ασταθής, ότι εμφανίζεται κάπου ένας κυκλώνας — όμως δεν μπορούν να ξέρουν πού. Ένα δέκατο του βαθμού περισσότερο ή λιγότερο σε κάποιο σημείο και ο κυκλώνας ξεσπά εδώ και όχι εκεί, σπέρνοντας τον όλεθρο σε χώρες που θα τον είχαν αποφύγει. Αυτό θα μπορούσαμε να το προβλέψουμε αν ξέραμε αυτό το δέκατο του βαθμού, οι παρατηρήσεις όμως δεν είναι αρκετά πυκνές ούτε αρκετά ακριβείς — και για το λόγο αυτό όλα φαίνονται να οφείλονται στον παράγοντα τύχη.”

Αξίζει να σημειωθεί ότι ο Poincaré ασχολήθηκε κυρίως με κινήσεις αστρικών σωμάτων, όπου και κει, αντίθετα με την κρατούσα αντίληψη που θέλει να επικρατεί πλήρης τάξη και περιοδικότητα, ανακάλυψε χαοτικά φαινόμενα (βλ. π.χ. Schroeder, 1990). Οι ανακαλύψεις του Poincaré, ωστόσο, έμειναν ανεξερεύνητες μέχρι τη δεκαετία του 1960, οπότε η εισαγωγή των ηλεκτρονικών υπολογιστών επέτρεψε την πληρέστερη μελέτη των μη γραμμικών συστημάτων. Στις αρχές της δεκαετίας του 1960 ο μετεωρολόγος Edward Lorenz τελείως τυχαία, κάνοντας ένα πείραμα σε ηλεκτρονικό υπολογιστή, επιβεβαίωσε τη χαοτική συμπεριφορά της εξέλιξης του καιρού. Ο Lorenz κατάρτισε ένα απλοποιημένο μοντέλο που προσέγγιζε τα φαινόμενα μεταφοράς στην ατμόσφαιρα, το οποίο αποτελείται από τρεις συνήθεις μη γραμμικές διαφορικές εξισώσεις (βλ. Cooper, 1989). Τις εξισώσεις αυτές τις επέλυε αριθμητικά με έναν από τους πρώτους ηλεκτρονικούς υπολογιστές. Ο ίδιος περιγράφει την ανακάλυψή του με τον ακόλουθο τρόπο:

“Κατά τη διάρκεια των υπολογισμών μας αποφασίσαμε να εξετάσουμε μια από τις λύσεις σε μεγαλύτερη λεπτομέρεια, και

διαλέξαμε κάποιες ενδιάμεσες συνθήκες που είχαν εκτυπωθεί από τον υπολογιστή, τις οποίες εισαγάγαμε ως νέες αρχικές συνθήκες. Όταν γυρίσαμε στον υπολογιστή, μια ώρα αργότερα, και αυτός είχε προσομοιώσει περίπου δύο μήνες “καιρού”, ανακαλύψαμε ότι υπήρχε πλήρης διαφωνία με την προηγούμενη λύση. Στην αρχή αυτό το αποδώσαμε σε πρόβλημα της μηχανής, πράγμα που δεν ήταν ασυνήθιστο, αλλά γρήγορα αντιληφθήκαμε ότι οι δύο λύσεις προέρχονταν από τις ίδιες συνθήκες. Οι υπολογισμοί έγιναν με υπολογιστική ακρίβεια περίπου έξι δεκαδικών ψηφίων, αλλά η εκτύπωση περιείχε μόνο τρία, έτσι ώστε οι νέες αρχικές συνθήκες αποτελούνταν από τις παλιές συν ορισμένες μικρές αποκλίσεις (διαταραχές). Αυτές οι αποκλίσεις μεγεθύνονταν σχεδόν εκθετικά, και διπλασιάζονταν σε περίπου τέσσερις μέρες, έτσι που σε δύο μήνες οι δύο λύσεις έπαιρναν ξεχωριστούς δρόμους.”

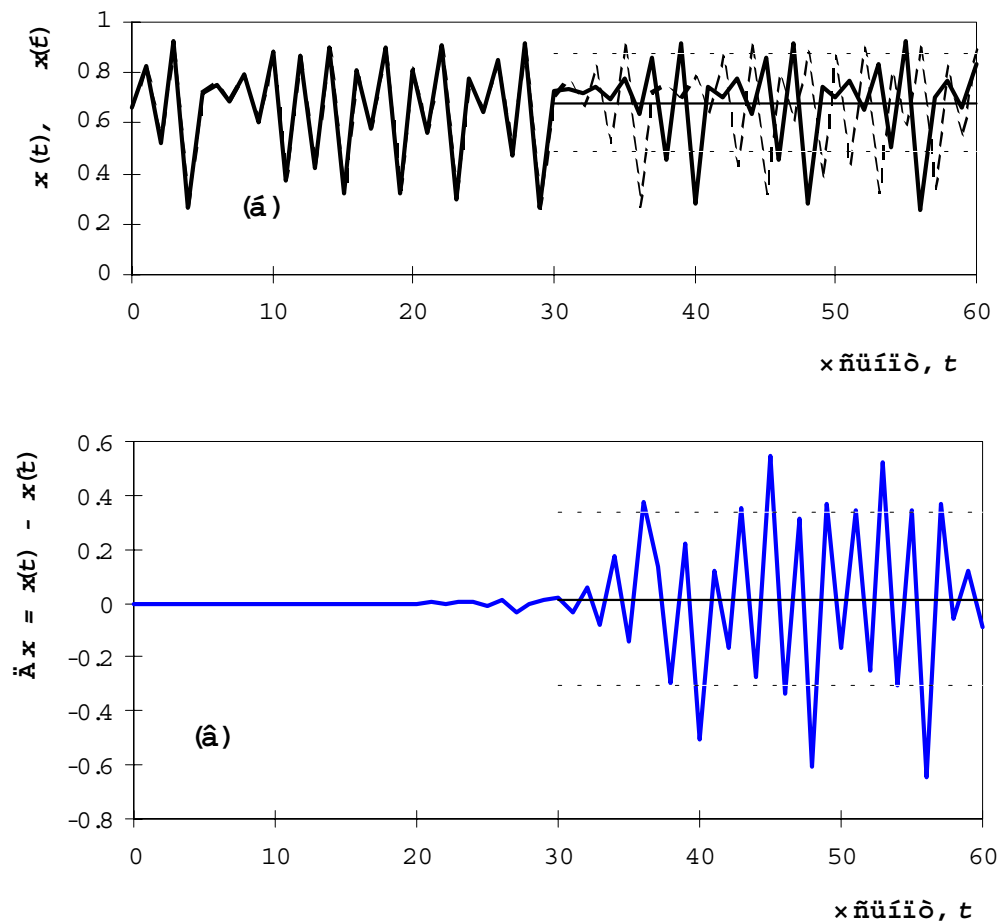
Ένα πολύ απλό παράδειγμα για να δούμε τη μεγέθυνση των διαταραχών στην εξέλιξη ενός μη γραμμικού φαινομένου δίνεται στο Σχ. 1.1. Ας θεωρήσουμε ένα σύστημα το οποίο περιγράφεται από μια μοναδική μεταβλητή $x(t)$, όπου t είναι ο χρόνος που στο παράδειγμα θεωρείται διακριτός. Θεωρούμε ότι η κατάσταση του συστήματος στην τρέχουσα χρονική στιγμή t εξαρτάται από την κατάστασή του στην αμέσως προηγούμενη χρονική στιγμή $t-1$, και ότι η μετάβαση από τη μια στιγμή στην άλλη περιγράφεται από την εξίσωση.

$$x(t) = k x(t-1) [1 - x(t-1)] \quad (1.1)$$

όπου k είναι αριθμητική σταθερά. Η εξίσωση αυτή είναι γνωστή στη βιβλιογραφία ως *λογιστική απεικόνιση* (βλ. π.χ. Cooper, 1989· Schroeder, 1990· Stewart, 1990, όπου υπάρχει αναλυτικότερη μελέτη της εξίσωσης) και βέβαια είναι μια από τις απλούστερες μη γραμμικές εξισώσεις.*

Τονίζεται ότι το σύστημα που περιγράφει η εξίσωση είναι πλήρως κλειστό και προσδιοριστικό, δηλαδή δεν μεσολαβεί στην εξέλιξή του κανένας εξωτερικός παράγοντας, ούτε προστίθεται καμιά τυχαία συνιστώσα.

* Η λογιστική απεικόνιση δεν χρησιμοποιείται σε υδρολογικές εφαρμογές, αφού δεν χαρακτηρίζει κανένα τυπικό υδρολογικό σύστημα. Απαντά σε άλλες επιστημονικές περιοχές, όπως π.χ. στην οικολογία. Ωστόσο, έχει επιλεγεί για το παράδειγμά μας λόγω της μαθηματικής απλότητάς της.



Σχ. 1.1 Εξέλιξη της λογιστικής απεικόνισης (εξίσωση 1.1) για σταθερά $k = 3.7$. Στο διάγραμμα (α) εμφανίζεται με χοντρή συνεχή γραμμή η πραγματική εξέλιξη της $x(t)$ για αρχική τιμή $x(0) = 0.660001$ και με διακεκομμένη η εξέλιξη $x'(t)$ αν η αρχική τιμή ληφθεί (π.χ. για λόγους στρογγύλευσης) ίση με 0.66 (η διαφορά από την πραγματική τιμή είναι μόνο 10^{-6}). Στο διάγραμμα (β) δίνεται η διαφορά των $x(t)$ και $x'(t)$ συναρτήσει του t . Και στα δύο διαγράμματα έχει απεικονιστεί επίσης η μέση τιμή της περιόδου $30 \leq t \leq 60$ (συνεχείς ευθείες) και το μέσο σφάλμα γύρω από τη μέση τιμή (διακεκομμένες ευθείες). Το μέσο σφάλμα είναι 0.19 για το διάγραμμα (α) και 0.31 για το διάγραμμα (β).

Παρατηρούμε ότι μετά από κάποιο αρχικό χρόνο (30 χρονικές μονάδες στο παράδειγμά μας) δύο “λύσεις” που αρχικά έχουν ασήμαντη απόκλιση, μόνο 10^{-6} για $t = 0$, αποκλίνουν τελείως και κάθε μια ακολουθεί το δικό της ανεξάρτητο δρόμο. Εύκολα συμπεραίνουμε ότι η προσδιοριστική πρόγνωση είναι πολύ αποτελεσματική για μικρά χρονικά διαστήματα (οι αποκλίσεις των δύο “λύσεων” στο Σχ. 1.1 είναι δυσδιάκριτες) αλλά οδηγεί σε πολύ σημαντικά σφάλματα για μεγαλύτερα χρονικά διαστήματα. Για τα μεγάλα διαστήματα η στατιστική πρόγνωση είναι πιο αποτελεσματική, αφού και μόνο η

στατιστική μέση τιμή δίνει μικρότερο σφάλμα στην εκτίμηση της πραγματικής εξέλιξης της $x(t)$ από το σφάλμα που δίνει η $x'(t)$. Επιπλέον, όπως θα δούμε παρακάτω, η στατιστική προσέγγιση έχει τη δυνατότητα να προσδιορίζει όρια μέσα στα οποία κυμαίνεται μια μεταβλητή, για καθορισμένη πιθανότητα.

Βεβαίως τα φυσικά υδρομετεωρολογικά συστήματα είναι απείρως πολυπλοκότερα από το σύστημα του παραπάνω παραδείγματος. Εύκολα λοιπόν μπορεί να συναγάγει κανείς την αδυναμία πλήρους προσδιοριστικής προσέγγισής τους και την αναγκαιότητα της πιθανοθεωρητικής προσέγγισης.

Κεφάλαιο 2 Εισαγωγικές έννοιες θεωρίας πιθανοτήτων

2.1 Αξιοματική 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 Ω , του οποίου τα στοιχεία ω αντιστοιχούν στις δυνατές εκβάσεις ενός πειράματος.
2. Η *οικογένεια γεγονότων* ορίζεται ως μια συλλογή \mathcal{F} υποσυνόλων (γεγονότων) A του Ω : Λέμε ότι το γεγονός A *συμβαίνει* (ή *πραγματοποιείται*) όταν η έκβαση ω 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 (Ω, \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:

$$\begin{aligned}
 \text{I.} & & P(A) & \geq 0 \\
 \text{II.} & & P(\Omega) & = 1 \\
 & & & \text{(2.1)}
 \end{aligned}$$

IIIα. $P(A \cup B) = P(A) + P(B)$, με την προϋπόθεση ότι $A \cap B = \emptyset$

IIIβ. $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$, με την προϋπόθεση ότι $A_i \cap A_j = \emptyset, i \neq j$

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 (\emptyset) 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:

$$\begin{aligned}
\alpha. & \quad \emptyset \in \mathcal{F} \\
\beta. & \quad \Omega \in \mathcal{F} \\
\gamma. & \quad (\forall A \in \mathcal{F}) [\bar{A} \in \mathcal{F}], \text{ όπου } \bar{A} \equiv \{\omega \in \Omega, \omega \notin A\} \\
\delta. & \quad (\forall A_i \in \mathcal{F}, i=1,2,\dots) \left[\bigcup_{i=1}^{\infty} A_i \in \mathcal{F} \right]
\end{aligned} \tag{2.2}$$

Υπενθυμίζεται ότι το γεγονός \bar{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 ω 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 ω 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 $\{X \leq x\}$ εννοώντας το γεγονός εκείνο (ήτοι υποσύνολο του δειγματικού χώρου) που αποτελείται από όλα τα στοιχεία ω , τέτοια ώστε οι τιμές τους $X(\omega)$ να είναι μικρότερες του αριθμού x . Την πιθανότητα του γεγονότος αυτού τη συμβολίζουμε με $P(\{X \leq x\})$ ή απλούστερα $P(X \leq x)$.

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

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

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

$$0 = F_X(-\infty) \leq F_X(x) \leq F_X(+\infty) = 1 \quad (2.4)$$

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

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

που είναι φθίνουσα συνάρτηση και υπακούει στη σχέση

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

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

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

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 δ του Dirac). Οι βασικές της ιδιότητες που προκύπτουν άμεσα από τον ορισμό της είναι

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

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

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

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

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

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

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

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

$$F = \{\emptyset, \{B\}, \{\Sigma\}, \Omega\}$$

Για να ορίσουμε την πιθανότητα επί του \mathcal{F} φτάνει να ορίσουμε την πιθανότητα ενός από τα δύο ενδεχόμενα, έστω την $P(B)$. 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 $P(B) = 0.2$. Οι υπόλοιπες πιθανότητες προκύπτουν με βάση τα αξιώματα. Έτσι έχουμε $P(\emptyset) = 0$ και $P(\Omega) = 1$. Δεδομένου ότι τα B και Σ είναι αμοιβαία αποκλειόμενα ισχύει $P(B) + P(\Sigma) = P(B \cup \Sigma) = P(\Omega) = 1$, άρα $P(\Sigma) = 0.8$.

Ας ορίσουμε την τυχαία μεταβλητή X με τον ακόλουθο τρόπο

$$X(\Sigma) = 0, \quad X(B) = 1$$

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

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

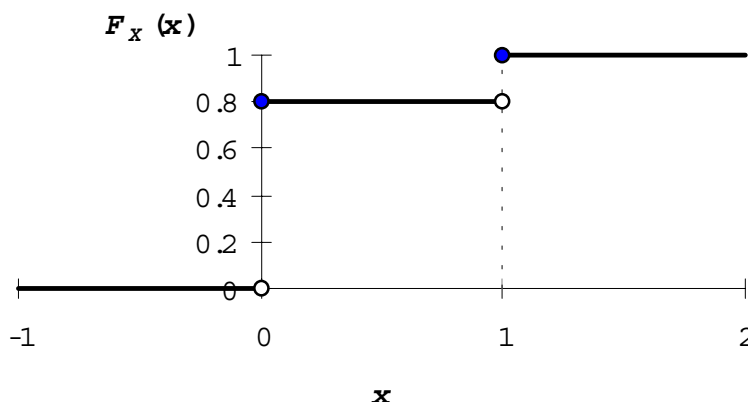
Τέλος για $1 \leq x$ έχουμε

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

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 ανεξίτητων* 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

$$P(A \cap B) = P(A)P(B) \quad (2.11)$$

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

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_n}) \quad (2.12)$$

για κάθε πεπερασμένο σύνολο διακεκριμένων δεικτών i_1, i_2, \dots, i_n .

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

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

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

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) \quad (2.14)$$

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 \bar{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

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

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

$$\Omega = \{A_1 A_2, \bar{A}_1 A_2, A_1 \bar{A}_2, \bar{A}_1 \bar{A}_2\}$$

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:

$$P(\bar{A}_1 A_2) = P(A_1 \bar{A}_2) = pq = 0.16, \quad P(\bar{A}_1 \bar{A}_2) = q^2 = 0.64$$

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 $A_1 A_2 \cup \bar{A}_1 A_2 \cup A_1 \bar{A}_2$ (έχει αποκλειστεί μόνο το $\bar{A}_1 \bar{A}_2$). Άρα η πιθανότητα που ζητούμε είναι η $P_2 = P(A_1 A_2 | A_1 A_2 \cup \bar{A}_1 A_2 \cup A_1 \bar{A}_2)$, η οποία σύμφωνα με τον ορισμό της δεσμευμένης πιθανότητας είναι

$$P_2 = \frac{P(A_1 A_2 \cap (A_1 A_2 \cup \bar{A}_1 A_2 \cup A_1 \bar{A}_2))}{P(A_1 A_2 \cup \bar{A}_1 A_2 \cup A_1 \bar{A}_2)}$$

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

$$P_2 = \frac{P(A_1 A_2)}{P(A_1 A_2) + P(\bar{A}_1 A_2) + P(A_1 \bar{A}_2)} = \frac{p^2}{p^2 + 2pq} = \frac{p}{p + 2q} = 0.111\dots$$

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 $A_1 A_2 \cup A_1 \bar{A}_2$ (έχουν αποκλειστεί τα $\bar{A}_1 A_2$ και $\bar{A}_1 \bar{A}_2$). Κατά συνέπεια η ζητούμενη πιθανότητα τώρα είναι η $P_3 = P(A_1 A_2 | A_1 A_2 \cup A_1 \bar{A}_2)$, που σύμφωνα με τον ορισμό της δεσμευμένης πιθανότητας είναι

$$P_3 = \frac{P(A_1 A_2 \cap (A_1 A_2 \cup A_1 \bar{A}_2))}{P(A_1 A_2 \cup A_1 \bar{A}_2)}$$

και γίνεται

$$P_3 = \frac{P(A_1 A_2)}{P(A_1 A_2) + P(A_1 \bar{A}_2)} = \frac{p^2}{p^2 + pq} = \frac{p}{p + q} = p = 0.2$$

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

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

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

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

Ειδικότερα:

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

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

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

2. Για $g(X) = X$, το μέγεθος

$$m_X := E[X] \quad (2.18)$$

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

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

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

ονομάζεται *κεντρική ροπή τάξης r* της X .

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

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

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

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.A (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.

$$\sigma_X = \sqrt{\text{Var}[X]} \quad (2.21)$$

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

$$C_{v_X} := \frac{\sigma_X}{m_X} \quad (2.22)$$

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.

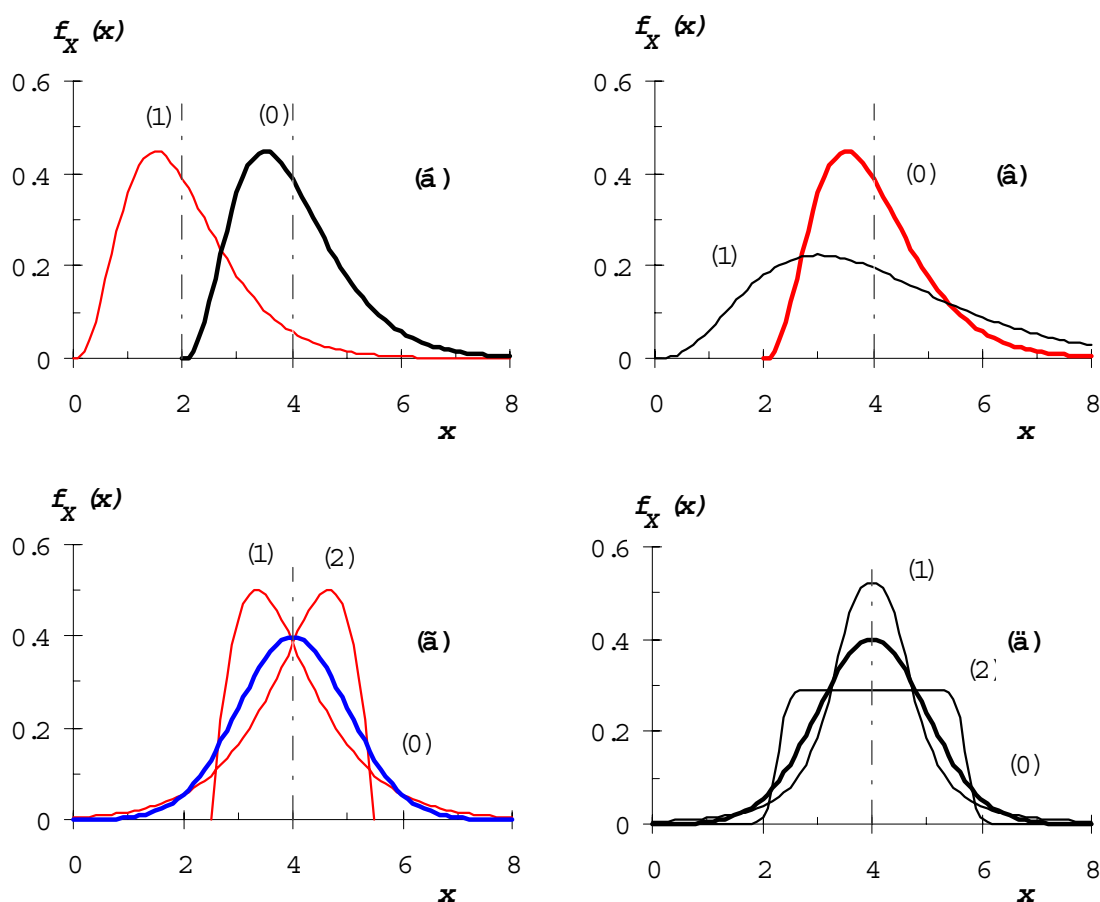
2.4.3 Παράμετροι ασυμμετρίας

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 $x_p \leq x_{0.5} \leq m_X$ 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

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

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.

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

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 $\lambda = 0.05 \text{ mm}^{-1}$. * 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 $r \geq 0$

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

και τελικά, εφαρμόζοντας τις εξισώσεις (2.68)

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

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

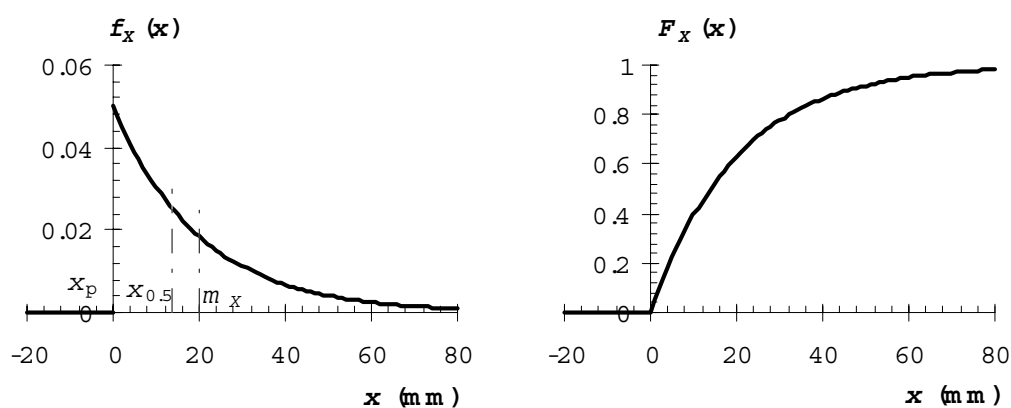
* Επειδή η τυχαία μεταβλητή στο παράδειγμα ορίζεται μόνο για τις μέρες που βρέχει, η κατανομή της είναι συνεχής παντού. Αν οριζόταν για όλες τις μέρες, και τις μη βροχερές, τότε η τιμή $X = 0$ θα είχε συγκεκριμένη πιθανότητα διαφορετική από το μηδέν, και έτσι η συνάρτηση κατανομής θα είχε άλμα στη θέση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

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 $x_p \leq x_{0.5} \leq m_X$ που χαρακτηρίζει τις θετικά ασύμμετρες κατανομές.

Η τυπική απόκλιση είναι $\sigma_X = 20 \text{ mm}$ 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 $\{Y \leq y\}$ είναι ταυτόσημο με το γεγονός $\{X \leq g^{-1}(y)\}$ όπου g^{-1} είναι η αντίστροφη συνάρτηση της g . Κατά συνέπεια οι συναρτήσεις κατανομής των X και Y συνδέονται με την

$$F_Y(y) = P\{Y \leq y\} = P\{X \leq g^{-1}(y)\} = F_X(g^{-1}(y)) \quad (2.25)$$

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

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

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, χρησιμοποιεί-

ται 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

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

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

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

Εξ άλλου, από την (2.15) έχουμε

$$\begin{aligned} E[Z] &= E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx = \int_{-\infty}^{\infty} \frac{x - m_X}{\sigma_X} f_X(x) dx = \\ &= \frac{1}{\sigma_X} \int_{-\infty}^{\infty} x f_X(x) dx - \frac{m_X}{\sigma_X} \int_{-\infty}^{\infty} f_X(x) dx = \frac{1}{\sigma_X} m_X - \frac{m_X}{\sigma_X} 1 \end{aligned}$$

και τελικά

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

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

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

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

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

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

Αντίστοιχα, από την (2.26) προκύπτει

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

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 (Ω_X, Ω_Y) αντίστοιχα. Η ένωση γεγονότων

$\{X \leq x\} \cup \{Y \leq y\} = \{X \leq x, Y \leq y\}$ 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) :

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

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

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

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

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

Οι συναρτήσεις

$$F_X(x) := P(X \leq x) = \lim_{y \rightarrow \infty} F_{XY}(x, y) \quad (2.30)$$

$$F_Y(y) := P(Y \leq y) = \lim_{x \rightarrow \infty} F_{XY}(x, y)$$

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

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

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

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

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

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

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

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

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

$$\rho_{XY} := \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]\text{Var}[Y]}} \equiv \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad (2.35)$$

ο οποίος είναι αδιάστατος και με τιμές $-1 \leq \rho_{XY} \leq 1$. 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

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

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

$$E[X|Y=y] := \int_{-\infty}^{\infty} xf_X(x|Y=y)dx \quad (2.37)$$

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

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

Σε αυτή την περίπτωση ισχύει επίσης η σχέση

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

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

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

$$E[X|Y=x] = E[X], \quad E[Y|X=x] = E[Y] \quad (2.41)$$

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

$$E[c_1g_1(X, Y) + c_2g_2(X, Y)] = c_1E[g_1(X, Y)] + c_2E[g_2(X, Y)] \quad (2.42)$$

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

$$E[X + Y] = E[X] + E[Y] \quad (2.43)$$

Κατά παρόμοιο τρόπο

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

Κατά συνέπεια

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y] \quad (2.45)$$

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

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

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 $4.0 \times 10^6 \text{ m}^3$ και τυπική απόκλιση $1.5 \times 10^6 \text{ m}^3$. 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 $4.0 \times 10^6 / 10.0 \times 10^6 = 0.4 \text{ m} = 400.0 \text{ mm}$ και για την τυπική απόκλιση $1.5 \times 10^6 / 10.0 \times 10^6 = 0.15 \text{ m} = 150.0 \text{ mm}$.

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

$$Z = X - Y \quad (\alpha)$$

Έχουμε δεδομένα τα μεγέθη

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

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

Συνδυάζοντας την (α) και τη (2.42) παίρνουμε

$$E[Z] = E[X - Y] = E[X] - E[Y] \Rightarrow m_Z = m_X - m_Y \quad (\beta)$$

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

$$Z - m_Z = (X - m_X) - (Y - m_Y) \quad (\gamma)$$

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)

$$\text{Var}[Z] = \text{Var}[X - Y] = \text{Var}[X] + \text{Var}[Y] - 2\text{Cov}[X, Y] \quad (\delta)$$

Δεδομένου, όμως, ότι οι μεταβλητές X και Y είναι ανεξάρτητες θα είναι $\text{Cov}[X, Y] = 0$ (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 $\sigma_Z = 151.3 \text{ mm}$.

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

$$E[(Z - m_Z)(X - m_X)] = E[(X - m_X)^2] - E[(X - m_X)(Y - m_Y)]$$

ή απλούστερα

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

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

$$\sigma_{ZY} = \sigma_X^2 = 150.0^2 = 22500.0 \text{ mm}^2$$

και κατά συνέπεια

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

Ομοίως έχουμε

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

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

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

Η θετική τιμή του ρ_{ZX} περιγράφει το γεγονός ότι η στάθμη αυξάνεται με την αύξηση της εισροής (θετική συσχέτιση των X και Z). Αντίστοιχα, η αρνητική τιμή του ρ_{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 ρ_{ZX} σε σχέση με την κατ' απόλυτο πολύ μικρότερη τιμή του ρ_{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, \dots, X_n είναι

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) := P(X_1 \leq x_1, \dots, X_n \leq x_n) \quad (2.47)$$

και συνδέεται με τη συνάρτηση πυκνότητας πιθανότητας με την

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

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

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n) \quad (2.49)$$

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, $T = \{0, 1, 2, \dots\}$, 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

$$F(x;t) := P(X(t) \leq x) \quad (2.50)$$

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

$$F(x_1, x_2; t_1, t_2) := P(X(t_1) \leq x_1, X(t_2) \leq x_2) \quad (2.51)$$

και κατ' επέκταση τη συνάρτηση n τάξης

$$F(x_1, \dots, x_n; t_1, \dots, t_n) := P(X(t_1) \leq x_1, \dots, X(t_n) \leq x_n) \quad (2.52)$$

Με παραγωγή 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 $X(t)$:

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

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

$$C(t_1, t_2) := \text{Cov}[X(t_1), X(t_2)] = E[(X(t_1) - \mu(t_1))(X(t_2) - \mu(t_2))] \quad (2.54)$$

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

$$\rho(t_1, t_2) := \frac{\text{Cov}[X(t_1), X(t_2)]}{\sqrt{\text{Var}[X(t_1)]\text{Var}[X(t_2)]}} = \frac{C(t_1, t_2)}{\sqrt{C(t_1, t_1)C(t_2, t_2)}} \quad (2.55)$$

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 χρόνου. 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 τ , 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

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

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

$$E[X(t)] = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=0}^N X(t) \quad (\text{για ανέλιξη διακριτού χρόνου}) \quad (2.57)$$

$$E[X(t)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T X(t) dt \quad (\text{για ανέλιξη συνεχούς χρόνου})$$

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 (χ^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 ($i = 1, \dots, n$) και το άθροισμά τους $Z = X_1 + X_2 + \dots + X_n$, με $E[Z] = m_Z$ και $\text{Var}[Z] = \sigma_Z^2$. 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

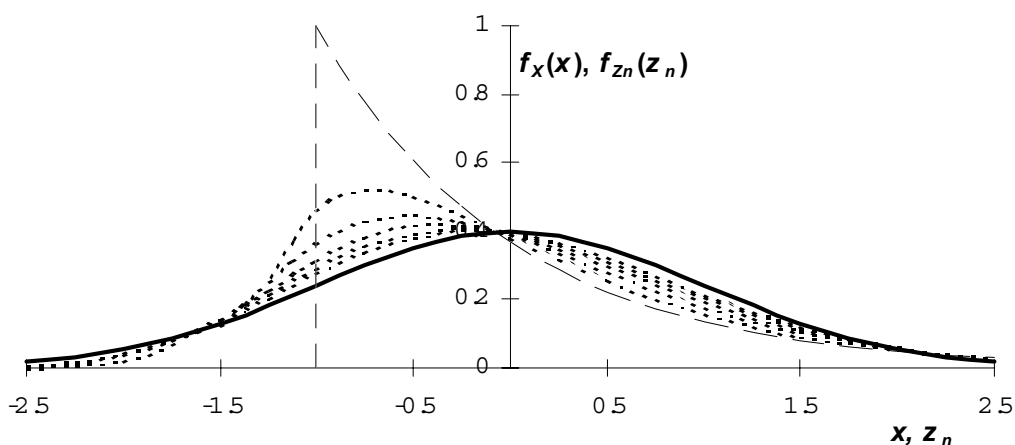
$$F_Z(z) \xrightarrow{n \rightarrow \infty} \int_{-\infty}^z \frac{1}{\sigma_Z \sqrt{2\pi}} e^{-\frac{(\zeta - m_Z)^2}{2\sigma_Z^2}} d\zeta \quad (2.58)$$

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

$$f_Z(z) \xrightarrow{n \rightarrow \infty} \frac{1}{\sigma_Z \sqrt{2\pi}} e^{-\frac{(z - m_Z)^2}{2\sigma_Z^2}} \quad (2.59)$$

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 $n \rightarrow \infty$ 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 μ και σ (συμβολικά $N(\mu, \sigma)$) αν η πυκνότητα πιθανότητάς της είναι η συνάρτηση

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

Η αντίστοιχη συνάρτηση κατανομής είναι η

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

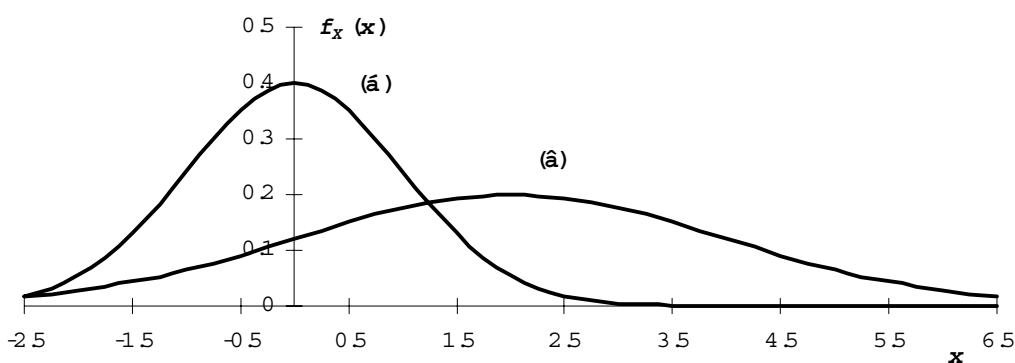
Η μέση τιμή και η τυπική απόκλιση της κατανομής είναι μ και σ , αντίστοιχα. Η κατανομή είναι συμμετρική (βλ. Σχ. 2.5) και επομένως η τρίτη κεντρική ροπή της είναι μηδέν (άρα $C_3 = 0$). Η τέταρτη κεντρική ροπή της είναι $3\sigma^4$ (άρα $C_4 = 3$).

Το ολοκλήρωμα στο δεξιό μέλος της (2.61) δεν υπολογίζεται αναλυτικά. Έτσι για τους τυπικούς υπολογισμούς ($x \rightarrow F_X(x)$ και $F_X(x) \rightarrow x$) 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

$$Z = \frac{X - \mu}{\sigma} \Leftrightarrow X = \mu + \sigma Z \quad (2.62)$$

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

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



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.

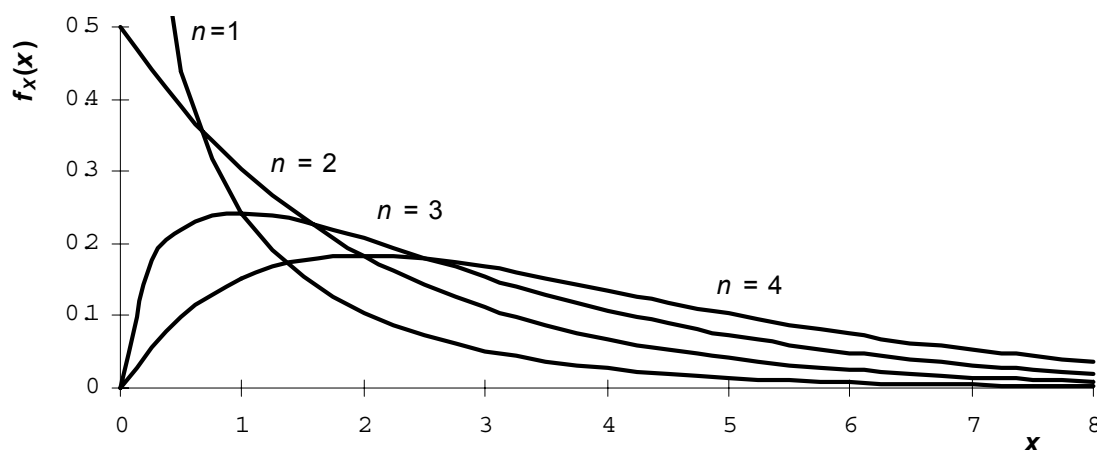
2.8.3 Η κατανομή χ^2

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

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

όπου $\Gamma()$ η συνάρτηση γάμα (βλ. Παράρτημα 2.B, 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 χ^2 ή της κατανομής γάμα (βλ. Παράρτημα Π2 και Π4).

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

$$Q = \sum_{i=1}^n X_i^2 \quad (2.65)$$

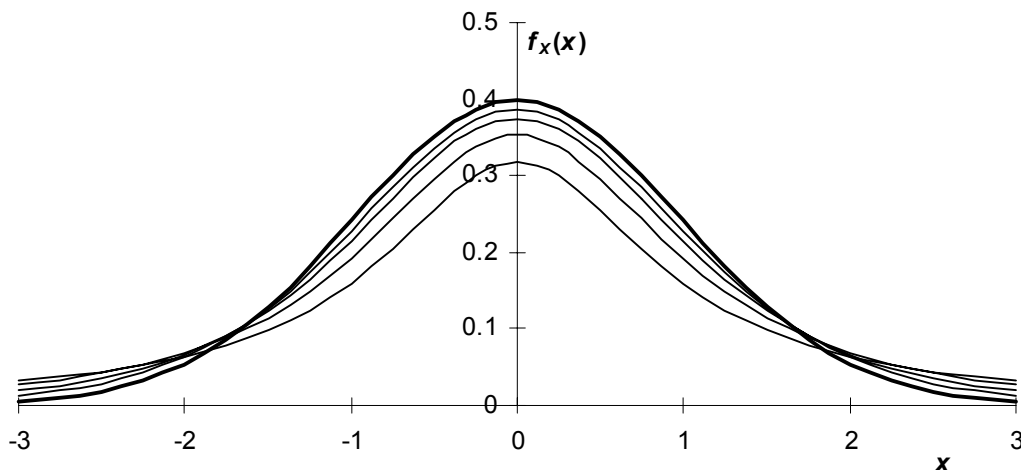
ακολουθεί κατανομή $\chi^2(n)$. Με συνδυασμό αυτού του θεωρήματος με το κεντρικό οριακό θεώρημα προκύπτει ότι για μεγάλο n η $\chi^2(n)$ τείνει στην κανονική κατανομή.

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

$$f_X(x) = \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n} \Gamma(n/2)} \frac{1}{\sqrt{(1+x^2/n)^{n+1}}} \quad n = 1, 2, \dots \quad (2.66)$$

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 (≥ 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

$$T = \frac{Z}{\sqrt{W/n}} \quad (2.67)$$

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

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

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

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

⋮

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

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:

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

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

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

Παράρτημα 2.B: Ιδιότητες της συνάρτησης $\Gamma(\cdot)$

Η συνάρτηση $\Gamma(\cdot)$ που υπεισέρχεται στον τύπο της κατανομής χ^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.

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

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

$$\begin{aligned} \Gamma(1) &= 1 & \Gamma(1/2) &= \sqrt{\pi} & \Gamma(a+1) &= a\Gamma(a) \\ \Gamma(n+1) &= n! & \Gamma(n + \frac{1}{2}) &= (n - \frac{1}{2})(n - \frac{3}{2}) \cdots \frac{3}{2} \sqrt{\pi} & n &= 1, 2, \dots \end{aligned} \quad (2.71)$$

Κεφάλαιο 3 Εισαγωγικές έννοιες στατιστικής

Η στατιστική είναι εφαρμοσμένος κλάδος της πιθανοθεωρίας ο οποίος ασχολείται με πραγματικά προβλήματα, επιδιώκοντας την εξαγωγή συμπερασμάτων βασισμένων σε παρατηρήσεις. Τα συμπεράσματα αφορούν ένα πληθυσμό, ενώ εξάγονται με βάση ένα δείγμα. Αν και το περιεχόμενο του όρου *πληθυσμός* δεν ορίζεται με ενιαίο και αυστηρό τρόπο στη στατιστική βιβλιογραφία, γενικά θεωρούμε ότι ο όρος περιγράφει οποιαδήποτε συλλογή αντικειμένων των οποίων ενδιαφέρουν οι μετρήσιμες ιδιότητες. Ο πληθυσμός μπορεί να είναι συγκεκριμένος και πεπερασμένος (π.χ. ο πληθυσμός των ετήσιων απορροών του υδρολογικού έτους 1990-91 για όλες τις υδρολογικές λεκάνες της γης με μέγεθος άνω των 100 km^2) ή άπειρος και αφαιρετικά ορισμένος (π.χ. ο πληθυσμός των δυνατών ετήσιων απορροών μιας υδρολογικής λεκάνης). Ο όρος *δείγμα* αναφέρεται σε ένα σύνολο μετρήσεων για το συγκεκριμένο πληθυσμό (βλ. και ορισμό στο εδάφιο 3.1.1).

Το πιο σημαντικό αντικείμενο της στατιστικής είναι η *εκτίμηση*. Διακρίνεται σε *εκτίμηση παραμέτρων* και *πρόγνωση*. Για να διευκρινίσουμε τις έννοιες αυτές, θεωρούμε κάποιο πληθυσμό που αντιπροσωπεύεται από την τυχαία μεταβλητή X με συνάρτηση κατανομής $F(X, \eta)$, όπου η παράμετρος. Στην περίπτωση που η παράμετρος είναι άγνωστη και επιδιώκουμε την εκτίμηση της, έχουμε ένα πρόβλημα εκτίμησης παραμέτρων. Στην περίπτωση που είναι γνωστή η παράμετρος και επιδιώκουμε την εκτίμηση της μεταβλητής X ή μιας συνάρτησής της, έχουμε ένα πρόβλημα πρόγνωσης. Όπως θα δούμε παρακάτω, και τα δύο

προβλήματα αντιμετωπίζονται με παρόμοιες μεθόδους της στατιστικής, γι' αυτό άλλωστε αποδίδονται και τα δύο με τον όρο *εκτίμηση*.

Η εκτίμηση μπορεί να είναι *σημειακή* ή *διαστήματος*. Στην πρώτη περίπτωση το ζητούμενο μέγεθος περιγράφεται με μια αριθμητική τιμή. Στη δεύτερη περιγράφεται από ένα διάστημα μέσα στο οποίο περιέχεται το ζητούμενο μέγεθος με δεδομένο βαθμό ακρίβειας ή αξιοπιστίας. Αντίστροφα, για ένα δεδομένο διάστημα διακύμανσης ενός μεγέθους, η στατιστική μπορεί να υπολογίσει τον αντίστοιχο βαθμό ακρίβειας ή, αντίθετα, το βαθμό αβεβαιότητας που υπάρχει γύρω από το ζητούμενο μέγεθος.

Μια άλλη σπουδαία περιοχή της στατιστικής είναι ο *έλεγχος υποθέσεων*, που αποτελεί τη βάση της *θεωρίας αποφάσεων*. Η διαδικασία του ελέγχου προϋποθέτει τη διατύπωση δύο υποθέσεων: μιας βασικής H_0 , που αναφέρεται με τον όρο *μηδενική υπόθεση*, και μιας *εναλλακτικής υπόθεσης* H_1 . Ξεκινάμε τη διαδικασία ελέγχου θεωρώντας ότι η μηδενική υπόθεση είναι σωστή και χρησιμοποιούμε τις παρατηρήσεις, προκειμένου να αποφασίσουμε αν η υπόθεση αυτή πρέπει να απορριφθεί. Αυτό γίνεται με τη χρήση στατιστικών θεωρήσεων. Παρόλο που και ο έλεγχος υποθέσεων βασίζεται στο ίδιο θεωρητικό υπόβαθρο όπως η εκτίμηση, η διαφορά εδώ είναι ότι εξετάζουμε δύο εναλλακτικά μοντέλα, ενώ στην εκτίμηση χρησιμοποιούμε ένα μόνο μοντέλο.

Το βασικό υπόβαθρο για όλα τα παραπάνω δίνεται στις ενότητες αυτού του κεφαλαίου, αναλυτικότερα όμως εφαρμόζεται στα κεφάλαια που ακολουθούν, όπου δίνονται και συγκεκριμένα αριθμητικά παραδείγματα. Βεβαίως, υπάρχουν και άλλες περιοχές της στατιστικής, όπως για παράδειγμα η ανάλυση Bayes, που δεν καλύπτονται σε αυτό το κείμενο.

3.1 Εννοιολογία και ορισμοί

3.1.1 Δείγμα

Θεωρούμε μια τυχαία μεταβλητή X με πυκνότητα πιθανότητας $f(X)$, η οποία ορίζεται βάσει ενός δειγματικού χώρου Ω . Θα θεωρούμε ότι η μεταβλητή αυτή ταυτίζεται εννοιολογικά με κάποιο πληθυσμό. *Δείγμα* μεγέθους (ή μήκους) n της μεταβλητής είναι μια ακολουθία n ανεξάρτητων τυχαίων μεταβλητών X_1, X_2, \dots, X_n με κοινή πυκνότητα πιθανότητας $f(x)$

που ορίζεται στο δειγματικό χώρο $\Omega^n = \Omega \times \cdots \times \Omega$ (Papoulis, 1990, σ. 238). Καθεμιά από τις μεταβλητές X_i αντιστοιχεί στις δυνατές εκβάσεις μιας μέτρησης ή παρατήρησης της μεταβλητής X . Αφού εκτελεστούν οι μετρήσεις, για καθεμιά μεταβλητή θα έχουμε μία μέτρηση, άρα συνολικά θα έχουμε την αριθμητική ακολουθία x_1, x_2, \dots, x_n , την οποία λέμε *παρατηρημένο δείγμα* ή απλώς *παρατηρήσεις*.

Η έννοια του δείγματος είναι, λοιπόν, συνυφασμένη με δύο ειδών ακολουθίες: μια αφαιρετική ακολουθία τυχαίων μεταβλητών και την αντίστοιχη συγκεκριμένη ακολουθία των αριθμητικών τιμών τους. Πολλές φορές στην τεχνική υδρολογία ο όρος *δείγμα* χρησιμοποιείται αδιακρίτως και για τη δεύτερη ακολουθία, με παράλειψη του όρου *παρατηρημένο*. Για το λόγο αυτό πρέπει να είμαστε προσεκτικοί ώστε να διακρίνουμε κάθε φορά αν ο όρος αναφέρεται με την αφαιρετική ή τη συγκεκριμένη έννοιά του.

Η λήψη ενός δείγματος μεγέθους n ή *δειγματοληψία* στη στατιστική γίνεται κατά κανόνα με την εκτέλεση n επαναλήψεων ενός πειράματος τύχης, δηλαδή είναι μια πειραματική διαδικασία. Οι επαναλήψεις πρέπει να είναι ανεξάρτητες μεταξύ τους και να εκτελούνται κάτω από ουσιαστικά ισοδύναμες συνθήκες. Αντίθετα, στην τεχνική υδρολογία δεν υπάρχει η δυνατότητα του πειράματος, και έτσι η δειγματοληψία είναι διαδικασία πολλαπλών μετρήσεων ενός φυσικού φαινομένου σε διάφορες χρονικές στιγμές. Κατά συνέπεια δεν είναι δυνατό να εξασφαλιστούν πάντα οι προϋποθέσεις της ανεξαρτησίας και των ισοδύναμων συνθηκών. Ωστόσο, για ορισμένες υδρολογικές μεταβλητές εξασφαλίζονται κατά προσέγγιση οι προϋποθέσεις αυτές (οι οποίες ισοδυναμούν με αυτές της ανεξαρτησίας, στασιμότητας και εργοδικότητας, βλ. κεφάλαιο 4) και έτσι μπορούμε να χρησιμοποιούμε για τη μελέτη τους τις τρέχουσες μεθόδους της στατιστικής.

3.1.2 Στατιστική συνάρτηση

Με τον όρο *στατιστική συνάρτηση* εννοούμε κάθε συνάρτηση των τυχαίων μεταβλητών του δείγματος, της μορφής $\Theta = g(X_1, \dots, X_n)$. Από τις παρατηρήσεις του δείγματος μπορούμε να υπολογίσουμε άμεσα την αριθμητική τιμή $\theta = g(x_1, \dots, x_n)$ της στατιστικής συνάρτησης. Προφανώς η στατιστική συνάρτηση δεν ταυτίζεται με την αριθμητική της τιμή, αφού η πρώτη, ως συνάρτηση τυχαίων μεταβλητών, είναι και η ίδια τυχαία

μεταβλητή, με συγκεκριμένη συνάρτηση κατανομής. Ενώ η αριθμητική τιμή της στατιστικής συνάρτησης υπολογίζεται με απλό αριθμητικό τρόπο από τις παρατηρήσεις του δείγματος, η συνάρτηση κατανομή της συνάγεται βάσει θεωρημάτων της στατιστικής, όπως θα δούμε σε επόμενες ενότητες. Παραδείγματα τυπικών στατιστικών συναρτήσεων δίνονται στα επόμενα εδάφια.

3.1.3 Εκτιμήτριες και εκτιμήσεις

Οι στατιστικές συναρτήσεις χρησιμοποιούνται για την εκτίμηση παραμέτρων του πληθυσμού. Για κάθε παράμετρο η του πληθυσμού μπορούν να βρεθούν μία ή περισσότερες στατιστικές συναρτήσεις της μορφής $\Theta = g(X_1, \dots, X_n)$ κατάλληλες για την εκτίμηση αυτής της παραμέτρου. Σε αυτή την περίπτωση λέμε ότι η $\Theta = g(X_1, \dots, X_n)$ είναι *εκτιμήτρια* της παραμέτρου η και ότι η αριθμητική τιμή της $\theta = g(x_1, \dots, x_n)$ αποτελεί *εκτίμηση* της η .

Δεν υπάρχει αυστηρό κριτήριο για το πότε μια στατιστική συνάρτηση μπορεί να χρησιμοποιηθεί για την εκτίμηση μιας παραμέτρου ενός πληθυσμού. Πολύ συχνά ο σχηματισμός τέτοιων συναρτήσεων είναι εμπειρικός και συνηθέστατα ο τύπος της στατιστικής συνάρτησης Θ είναι ο ίδιος με αυτόν που δίνει την παράμετρο η για μια τυχαία μεταβλητή με διακριτό και πεπερασμένο σύνολο τιμών. Για παράδειγμα, έστω ότι μας ενδιαφέρει να βρούμε μια εκτιμήτρια της μέσης τιμής $\eta = m_X$ μιας μεταβλητής X , με βάση το δείγμα (X_1, \dots, X_n) με παρατηρήσεις (x_1, \dots, x_n) . Θεωρούμε προς στιγμήν ότι η X , είναι διακριτή μεταβλητή με δυνατές τιμές (x_1, \dots, x_n) , σε καθεμιά από τις οποίες αντιστοιχεί η ίδια πιθανότητα $P(X = x_i) = 1/n$. Από τον τύπο της μέσης τιμής διακριτής μεταβλητής (εξισώσεις (2.16) και (2.18)) βρίσκουμε $\eta = (x_1 + \dots + x_n)/n$. Στην τελευταία εξίσωση αντικαθιστούμε τις αριθμητικές τιμές με τις αντίστοιχες μεταβλητές και παίρνουμε τη στατιστική συνάρτηση $\Theta = (X_1 + \dots + X_n)/n$. Όπως θα δούμε και παρακάτω, αυτή η στατιστική συνάρτηση είναι πράγματι εκτιμήτρια της μέσης τιμής οποιασδήποτε μεταβλητής, λέγεται *δειγματική μέση τιμή* και συμβολίζεται με \bar{X} . Πάντως, αυτή η εμπειρική μέθοδος δεν δίνει πάντα την καλύτερη δυνατή εκτιμήτρια.

Παρόλο που, όπως είδαμε παραπάνω, η εκτιμήτρια στη γενική περίπτωση δεν ορίζεται με αυστηρό μαθηματικό τρόπο, διάφορες κατηγορίες εκτιμητριών έχουν αυστηρούς ορισμούς. Έτσι:

1. Μια στατιστική συνάρτηση Θ είναι *αμερόληπτη εκτιμήτρια* της παραμέτρου η αν $E[\Theta] = \eta$. Διαφορετικά είναι *μεροληπτική εκτιμήτρια* και η διαφορά $E[\Theta] - \eta$ λέγεται *μεροληψία*.
2. Μια στατιστική συνάρτηση Θ είναι *συνεπής εκτιμήτρια* της παραμέτρου η αν το σφάλμα εκτίμησης $\Theta - \eta$ τείνει στο μηδέν με πιθανότητα 1 όταν $n \rightarrow \infty$. Διαφορετικά είναι *ασυνεπής εκτιμήτρια*.
3. Μια στατιστική συνάρτηση Θ είναι *βέλτιστη εκτιμήτρια* της παραμέτρου η αν το μέσο τετραγωνικό σφάλμα εκτίμησης $(\Theta - \eta)^2$ είναι ελάχιστο.
4. Μια στατιστική συνάρτηση Θ είναι *η πιο αποτελεσματική εκτιμήτρια* της παραμέτρου η αν είναι αμερόληπτη και έχει την ελάχιστη διασπορά.

Ας σημειωθεί ότι η εκτιμήτρια \bar{X} του πιο πάνω παραδείγματος είναι αμερόληπτη και συνεπής, και μάλιστα για ορισμένες συναρτήσεις κατανομής είναι ταυτόχρονα και βέλτιστη και η πιο αποτελεσματική.

Στην πράξη γίνεται προσπάθεια να χρησιμοποιούνται αμερόληπτες και συνεπείς εκτιμήτριες, ενώ ο υπολογισμός της βέλτιστης και της πιο αποτελεσματικής εκτιμήτριας έχει περισσότερο θεωρητικό ενδιαφέρον. Για κάθε παράμετρο μπορεί να υπάρχουν περισσότερες από μία αμερόληπτες ή συνεπείς εκτιμήτριες. Συχνά ο προσδιορισμός αμερόληπτων εκτιμητριών είναι αρκετά δύσκολος, οπότε καταφεύγουμε στη χρήση μεροληπτικών.

3.1.4 Εκτίμηση διαστήματος και όρια εμπιστοσύνης

Εκτίμηση διαστήματος μιας παραμέτρου η είναι ένα διάστημα της μορφής (θ_1, θ_2) , όπου $\theta_1 = g_1(x_1, \dots, x_n)$ και $\theta_2 = g_2(x_1, \dots, x_n)$ είναι συναρτήσεις των παρατηρήσεων του δείγματος. Το διάστημα (Θ_1, Θ_2) που ορίζουν οι αντίστοιχες στατιστικές συναρτήσεις $\Theta_1 = g_1(X_1, \dots, X_n)$ και $\Theta_2 = g_2(X_1, \dots, X_n)$ λέγεται *εκτιμήτρια διαστήματος* της η .

Το διάστημα (Θ_1, Θ_2) ονομάζεται *διάστημα εμπιστοσύνης γ της παραμέτρου η αν*

$$P(\Theta_1 < \eta < \Theta_2) = \gamma \quad (3.1)$$

όπου γ είναι δεδομένη σταθερά ($0 < \gamma < 1$). Η σταθερά αυτή ονομάζεται *συντελεστής εμπιστοσύνης* και κατά κανόνα παίρνει τιμές κοντά στο 1 (π.χ. 0.9, 0.95, 0.99, έτσι ώστε η πιθανότητα στην (3.1) να γίνεται “σχεδόν βεβαιότητα”). Η διαφορά $\alpha = 1 - \gamma$ ονομάζεται *επίπεδο εμπιστοσύνης* και τα όρια Θ_1 και Θ_2 ονομάζονται *όρια εμπιστοσύνης*. Καταχρηστικώς, ο όρος *όρια εμπιστοσύνης* χρησιμοποιείται στην πράξη και για τις αριθμητικές τιμές θ_1 και θ_2 των στατιστικών συναρτήσεων και το ίδιο συμβαίνει και για τον όρο *διάστημα εμπιστοσύνης*.

Προκειμένου να δώσουμε ένα γενικό τρόπο υπολογισμού του διαστήματος εμπιστοσύνης, ας θεωρήσουμε ότι η στατιστική συνάρτηση $\Theta = g(X_1, \dots, X_n)$ είναι αμερόληπτη (σημειακή) εκτιμήτρια της παραμέτρου η και ότι η συνάρτηση κατανομής της είναι η $F_\Theta(\theta)$. Με βάση αυτή τη συνάρτηση κατανομής μπορούν να υπολογιστούν δύο θετικοί αριθμοί ξ_1 και ξ_2 , έτσι ώστε το σφάλμα εκτίμησης $\Theta - \eta$ να βρίσκεται στο διάστημα $(-\xi_1, \xi_2)$ με πιθανότητα γ , ήτοι

$$P(\eta - \xi_1 < \Theta < \eta + \xi_2) = \gamma \quad (3.2)$$

και παράλληλα το διάστημα $(-\xi_1, \xi_2)$ να είναι το ελάχιστο δυνατό.* Η τελευταία εξίσωση μετασχηματίζεται άμεσα στην

$$P(\Theta - \xi_1 < \eta < \Theta + \xi_2) = \gamma \quad (3.3)$$

Κατά συνέπεια τα ζητούμενα όρια εμπιστοσύνης είναι $\Theta_1 = \Theta - \xi_1$ και $\Theta_2 = \Theta + \xi_2$.

Παρόλο που οι εξισώσεις (3.2) και (3.3) είναι ισοδύναμες, η στατιστική ερμηνεία τους είναι διαφορετική. Η πρώτη είναι προγνωστική, με την έννοια ότι μας δίνει όρια εμπιστοσύνης[†] για την τυχαία μεταβλητή Θ . Η δεύτερη είναι εκτιμητική, αφού μας δίνει όρια εμπιστοσύνης για την άγνωστη παράμετρο η , η οποία προφανώς δεν είναι τυχαία μεταβλητή.

* Αν η κατανομή της Q είναι συμμετρική τότε η απαίτηση αυτή ικανοποιείται για $\xi_1 = \xi_2$. Διαφορετικά, επειδή είναι δύσκολος ο υπολογισμός του ελάχιστου διαστήματος για ασύμμετρες κατανομές, απλοποιούμε το πρόβλημα διασπώντας την (3.2) στις εξισώσεις $P(\Theta < \eta - \xi_1) = P(\Theta > \eta + \xi_2) = (1 - \gamma) / 2$.

[†] Οι όροι *όρια εμπιστοσύνης*, *διάστημα εμπιστοσύνης*, *συντελεστής εμπιστοσύνης* κτλ. χρησιμοποιούνται και για αυτή την προγνωστική μορφή της εξίσωσης.

3.2 Τυπικές σημειακές εκτιμήτριες

Στην ενότητα αυτή δίνουμε τις πιο τυπικές εκτιμήτριες που αναφέρονται σε στατιστικές ροπές του πληθυσμού και είναι ανεξάρτητες της συνάρτησης κατανομής του $F(x)$. Συγκεκριμένα, δίνουμε εκτιμήτριες για τη μέση τιμή, τη διασπορά και την τρίτη κεντρική ροπή μιας μεταβλητής. Δεν επεκτεινόμαστε σε ροπές μεγαλύτερης τάξης, γιατί και η εξαγωγή αμερόληπτων εκτιμητριών είναι αρκετά πολύπλοκη, αλλά και η αποτελεσματικότητά τους είναι περιορισμένη για μικρά δείγματα (δηλαδή οι εκτιμήτριες έχουν μεγάλη διασπορά) έτσι ώστε οι αντίστοιχες εκτιμήσεις να είναι ασαφείς. Γι' αυτό, άλλωστε, στην τεχνική υδρολογία δεν χρησιμοποιούνται ροπές τάξης μεγαλύτερης από τρία. Ακόμη και η εκτίμηση της τρίτης ροπής είναι αρκετά ασαφής για μικρά δείγματα, αλλά όμως αυτή χρησιμοποιείται συχνά στην τεχνική υδρολογία για το λόγο ότι περιγράφει την ασυμμετρία της κατανομής, που είναι βασικό χαρακτηριστικό για τις, κατά κανόνα θετικά ασύμμετρες, υδρολογικές μεταβλητές.

Πέρα από τις παραπάνω ροπές μιας μεταβλητής, δίνουμε και εκτιμήτριες της συνδιασποράς και του συντελεστή συσχέτισης δύο μεταβλητών, που χρησιμοποιούνται κατά την ταυτόχρονη στατιστική ανάλυση δύο μεταβλητών.

3.2.1 Γενικά για τις εκτιμήτριες ροπών

Οι εκτιμήτριες των ροπών περί την αρχή για μία ή δύο μεταβλητές, ήτοι των $m_X^{(r)}$ και $m_{XY}^{(rs)}$ (όπου τα r και s είναι τυχόντες ακέραιοι), κατασκευασμένες με την εμπειρική μέθοδο που περιγράφηκε στο εδάφιο 3.1.3, είναι οι εξής:

$$\tilde{M}_X^{(r)} = \frac{\sum_{i=1}^n X_i^r}{n} \quad \tilde{M}_{XY}^{(rs)} = \frac{\sum_{i=1}^n X_i^r Y_i^s}{n} \quad (3.4)$$

Αποδεικνύεται (Kendall and Stewart, 1968, σ. 229) ότι

$$E\left[\tilde{M}_X^{(r)}\right] = m_X^{(r)} \quad E\left[\tilde{M}_{XY}^{(rs)}\right] = m_{XY}^{(rs)} \quad (3.5)$$

Κατά συνέπεια οι εκτιμήτριες είναι αμερόληπτες. Οι διασπορές των εκτιμητριών αυτών είναι

$$\text{Var}[\tilde{M}_X^{(r)}] = \frac{1}{n} \left[m_X^{(2r)} - (m_X^{(r)})^2 \right] \quad E[\tilde{M}_{XY}^{(rs)}] = \frac{1}{n} \left[m_{XY}^{(2r,2s)} - (m_{XY}^{(rs)})^2 \right] \quad (3.6)$$

Οι τελευταίες εξισώσεις δείχνουν ότι η διασπορές τείνουν στο 0 όταν $n \rightarrow \infty$ και κατά συνέπεια ότι οι εκτιμήτριες είναι συνεπείς.

Για την κατασκευή εκτιμητριών των κεντρικών ροπών $\mu_X^{(r)}$ ή $\mu_{XY}^{(rs)}$ μίας ή δύο μεταβλητών, αντίστοιχα, χρησιμοποιούνται ως βάση οι στατιστικές συναρτήσεις $\hat{M}_X^{(r)}$ και $\hat{M}_{XY}^{(rs)}$ που ορίζονται από τις εξισώσεις

$$\hat{M}_X^{(r)} = \frac{\sum_{i=1}^n (X_i - \bar{X})^r}{n} \quad \hat{M}_{XY}^{(rs)} = \frac{\sum_{i=1}^n (X_i - \bar{X})^r (Y_i - \bar{Y})^s}{n} \quad (3.7)$$

Οι εκτιμήτριες αυτές έχουν κατασκευαστεί με την εμπειρική μέθοδο που περιγράφηκε στο εδάφιο 3.1.3. Ωστόσο, είναι μεροληπτικές εκτιμήτριες (για $r + s > 1$). Αν $\hat{M}_X^{*(r)}$ και $\hat{M}_{XY}^{*(rs)}$ είναι κάποιες αμερόληπτες εκτιμήτριες των ίδιων ροπών του πληθυσμού, τότε οι λόγοι

$$\phi(\hat{M}_X^{(r)}, n) = \frac{\hat{M}_X^{*(r)}}{\hat{M}_X^{(r)}} \quad \phi(\hat{M}_{XY}^{(rs)}, n) = \frac{\hat{M}_{XY}^{*(rs)}}{\hat{M}_{XY}^{(rs)}} \quad (3.8)$$

ονομάζονται *συντελεστές διόρθωσης μεροληψίας*. Αντιστρέφοντας και γενικεύοντας το συμβολισμό, αν δοθεί μια μεροληπτική εκτιμήτρια Θ και ο αντίστοιχος συντελεστής διόρθωσης μεροληψίας $\phi(\Theta, n)$ τότε προσδιορίζεται άμεσα και η αντίστοιχη αμερόληπτη εκτιμήτρια $\Theta^* = \phi(\Theta, n) \Theta$. Ο προσδιορισμός συντελεστών διόρθωσης μεροληψίας δεν είναι πάντα εύκολος και γι' αυτό σε ορισμένες περιπτώσεις δίνονται προσεγγιστικοί συντελεστές. Επίσης, για τυχούσα μη γραμμική συνάρτηση $g()$, η $g(\Theta^*)$ δεν παραμένει γενικά αμερόληπτη, γεγονός που δυσκολεύει τον προσδιορισμό συντελεστών διόρθωσης για μετασχηματισμούς μιας ή περισσότερων στατιστικών συναρτήσεων.

3.2.2 Δειγματική μέση τιμή

Η πιο κοινή στατιστική συνάρτηση είναι η δειγματική μέση τιμή, η οποία, όπως είδαμε και παραπάνω, αποτελεί εκτιμήτρια της μέσης τιμής και ορίζεται από τη σχέση

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (3.9)$$

που αποτελεί ειδική περίπτωση της (3.4) για $r = 1$. Η αριθμητική τιμή της

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.10)$$

ονομάζεται *παρατηρημένη ή αριθμητική μέση τιμή* ή απλώς *μέσος όρος* του δείγματος. Τα μεγέθη \bar{X} και \bar{x} δεν πρέπει να συγχέονται εννοιολογικά μεταξύ τους, ούτε με τη μέση τιμή της μεταβλητής X , δηλαδή $m_X = E[X]$ που ορίζεται με βάση τις εξισώσεις (2.15) ή (2.16) και (2.18). Ωστόσο τα τρία μεγέθη είναι στενά συνδεδεμένα. Εφαρμόζοντας τις (3.5) και (3.6) βρίσκουμε

$$E[\bar{X}] = E[X] \quad \text{Var}[\bar{X}] = \frac{\text{Var}[X]}{n} \quad (3.11)$$

και αυτό ανεξάρτητα από το ποια είναι η κατανομή της X . Οι παραπάνω εξισώσεις δείχνουν ότι η εκτιμήτρια είναι αμερόληπτη και συνεπής.

3.2.3 Διασπορά και τυπική απόκλιση

Η μεροληπτική εκτιμήτρια της διασποράς σ_X^2 του πληθυσμού είναι η ακόλουθη:

$$S_X^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \quad (3.12)$$

για την οποία αποδεικνύεται (Kendall and Stewart, 1968, σ. 277) ότι

$$E[S_X^2] = \frac{n-1}{n} \sigma_X^2 \quad (3.13)$$

$$\text{Var}[S_X^2] = \frac{\mu_X^{(4)} - \sigma_X^4}{n} - \frac{2(\mu_X^{(4)} - 2\sigma_X^4)}{n^2} + \frac{\mu_X^{(4)} - 3\sigma_X^4}{n^3}$$

όπου $\mu_X^{(4)}$ η τέταρτη κεντρική ροπή του πληθυσμού. Στην τελευταία εξίσωση μπορούν να παραλειφθούν οι δύο τελευταίοι όροι για μεγάλο n . Από την πρώτη από τις πιο πάνω εξισώσεις προκύπτει άμεσα ότι ο συντελεστής διόρθωσης μεροληψίας είναι

$$\phi(S_X^2, n) = \frac{n}{n-1} \quad (3.14)$$

και κατά συνέπεια η αμερόληπτη (και συνεπής) εκτιμήτρια της σ_X^2 είναι η ακόλουθη, γνωστή ως *δειγματική διασπορά*:

$$S_X^{*2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad (3.15)$$

Για μεγάλο μέγεθος δείγματος n οι δύο εκτιμήτριες πρακτικώς ταυτίζονται. Στην περίπτωση κανονικής κατανομής του πληθυσμού αποδεικνύεται ότι

$$\text{Var}[S_X^2] = \frac{2(n-1)\sigma_X^4}{n^2} \quad \text{Var}[S_X^{*2}] = \frac{2\sigma_X^4}{(n-1)} \quad (3.16)$$

Ως εκτιμήτριες της τυπικής απόκλισης του πληθυσμού χρησιμοποιούνται οι τετραγωνικές ρίζες των παραπάνω, δηλαδή οι S_X και S_X^* , οι οποίες πάντως δεν είναι αμερόληπτες. Έτσι (Yevjevich, 1972, σ. 193· Kendall and Stewart, 1968, σ. 233)

$$E[S_X] = \sigma_X + O\left(\frac{1}{n}\right) \quad \text{Var}[S_X] = \frac{\mu_X^{(4)} - \sigma_X^4}{4\sigma_X^2 n} + O\left(\frac{1}{n^2}\right) \quad (3.17)$$

όπου οι όροι $O(1/n)$ και $O(1/n^2)$ είναι ποσότητες ανάλογες των $1/n$ και $1/n^2$, αντίστοιχα, και μπορούν να παραλειφθούν για μέγεθος δείγματος σχετικά μεγάλο ($n \geq 20$). Στην περίπτωση κανονικής κατανομής του πληθυσμού, για την S_X χρησιμοποιούνται οι ακόλουθες προσεγγιστικές εξισώσεις

$$E[S_X] \approx \sigma_X \sqrt{\frac{n-1}{n}} \quad \text{Var}[S_X] \approx \frac{\sigma_X^2}{2n} \quad (3.18)$$

Τα σφάλματα αυτών των εξισώσεων είναι μικρότερα από 2.5% και 2.7%, αντίστοιχα, για $n \geq 10$ και πρακτικώς μηδενίζονται για $n \geq 100$. Οι αντίστοιχες εξισώσεις για την S_X^* είναι

$$E[S_X^*] \approx \sigma_X \quad \text{Var}[S_X^*] \approx \frac{\sigma_X^2}{2(n-1)} \quad (3.19)$$

ενώ ακριβέστερες προσεγγίσεις δίνουν οι

$$E[S_X^*] \approx \sigma_X \sqrt{\frac{n - \frac{5}{4}}{n - \frac{3}{4}}} \quad \text{Var}[S_X^*] \approx \frac{\sigma_X^2}{2(n - \frac{3}{4})} \quad (3.20)$$

των οποίων τα σφάλματα είναι μικρότερα από 0.005% και 0.2%, αντίστοιχα, για $n \geq 10$.*

Τέλος, για το συντελεστή μεταβλητότητας χρησιμοποιείται μια από τις δύο εκτιμήτριες

$$\hat{C}_{v_X} = \frac{S_X}{\bar{X}} \quad \hat{C}_{v_X}^* = \frac{S_X^*}{\bar{X}} \quad (3.21)$$

Αν η X είναι θετική, τότε αποδεικνύεται ότι οι εκτιμήτριες αυτές είναι άνω φραγμένες ($\hat{C}_{v_X} \leq \sqrt{n-1}$) χωρίς να ισχύει το ίδιο και για τις αντί-

* Παραθέτουμε εδώ και τις ακριβείς εξισώσεις, που είναι

$$E[S_X^*] = \sigma_X \frac{\Gamma(\frac{n}{2})}{\sqrt{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} \quad \text{Var}[S_X^*] = \sigma_X^2 \left[1 - \frac{\Gamma^2(\frac{n}{2})}{\frac{n-1}{2} \Gamma^2(\frac{n-1}{2})} \right]$$

στοιχεις παραμέτρους του πληθυσμού. Αυτό εισάγει ασφαλώς μεροληψία, η οποία πάντως για τις τρέχουσες εφαρμογές είναι αμελητέα.*

3.2.4 Τρίτη κεντρική ροπή και συντελεστής ασυμμετρίας

Η μεροληπτική εκτιμήτρια της τρίτης κεντρικής ροπής $\mu_X^{(3)}$ του πληθυσμού είναι η ακόλουθη

$$\hat{M}_X^{(3)} = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{n} \quad (3.22)$$

για την οποία αποδεικνύεται (Kendall and Stewart, σ. 278-281) ότι†

$$E[\hat{M}_X^{(3)}] = \frac{(n-1)(n-2)}{n^2} \mu_X^{(3)} \quad (3.23)$$

Από την παραπάνω εξίσωση προκύπτει άμεσα ότι ο συντελεστής διόρθωσης μεροληψίας είναι‡

* Η έκφραση για τη διασπορά της εκτιμήτριας είναι αρκετά πολύπλοκη και παραλείπεται (βλ. Kendall and Stewart, 1968, σ. 233). Στην περίπτωση κανονικής κατανομής της X ισχύει

$$\text{Var}[\hat{C}_{v_X}] \approx C_{v_X}^2 / 2n$$

† Η έκφραση για τη διασπορά της εκτιμήτριας είναι αρκετά πολύπλοκη:

$$\text{Var}[\hat{M}_X^{(3)}] \approx \frac{1}{n} \left[\mu_X^{(6)} - (\mu_X^{(3)})^2 - 6\mu_X^{(4)}\sigma_X^2 + 9\sigma_X^6 \right]$$

Για κανονική κατανομή της X ισχύει (Kendall and Stewart, 1968, σ. 243)

$$\text{Var}[\hat{M}_X^{(3)}] = 6\sigma_X^6 / n$$

‡ Όπως προαναφέρθηκε, για μεγαλύτερης τάξης ροπές, ο υπολογισμός αμερόληπτων εκτιμητριών καθίσταται ιδιαίτερα πολύπλοκος. Για παράδειγμα, ο συντελεστής διόρθωσης μεροληψίας της τέταρτης κεντρικής ροπής είναι

$$\phi(\hat{M}_X^{(4)}, n) = \frac{n^3}{(n-1)(n-2)(n-3)} \left[1 - \frac{2n-3}{n^2} \left(\frac{3}{\hat{C}_{k_X}} + 1 \right) \right]$$

όπου

$$\phi(\hat{M}_X^{(3)}, n) = \frac{n^2}{(n-1)(n-2)} \quad (3.24)$$

και κατά συνέπεια η αμερόληπτη (και συνεπής) εκτιμήτρια της $\mu_X^{(3)}$ είναι η

$$\hat{M}_X^{*(3)} = \frac{n \sum_{i=1}^n (X_i - \bar{X})^3}{(n-1)(n-2)} \quad (3.25)$$

Για μεγάλο μέγεθος δείγματος n οι δύο εκτιμήτριες πρακτικώς ταυτίζονται.

Για την εκτίμηση του συντελεστή ασυμμετρίας C_{s_X} του πληθυσμού χρησιμοποιείται κατά βάση η εκτιμήτρια

$$\hat{C}_{s_X} = \frac{\hat{M}_X^{(3)}}{S_X^3} \quad (3.26)$$

η οποία δεν είναι αμερόληπτη. Η μεροληψία δεν οφείλεται μόνο στο γεγονός ότι οι δύο εκτιμήτριες ροπών από τις οποίες υπολογίζεται (αριθμητής και παρονομαστής) δεν είναι αμερόληπτες, αλλά κυρίως στο γεγονός ότι η \hat{C}_{s_X} είναι άνω και κάτω φραγμένη σε αντίθεση με την C_{s_X} του πληθυσμού που δεν είναι φραγμένη. Αυτό οφείλεται στο πεπερασμένο μέγεθος του δείγματος, το οποίο καθορίζει και το άνω και κάτω όριο. Έτσι αποδεικνύεται (Kirby, 1974· Wallis et al., 1974) ότι $\hat{C}_{s_X} \leq (n-2) / \sqrt{n-1}$.

Έχουν προταθεί διάφοροι προσεγγιστικοί συντελεστές διόρθωσης μεροληψίας, χωρίς όμως κανένα απ' αυτούς να οδηγεί σε αυστηρώς αμε-

$$\hat{C}_{k_X} = \hat{M}_X^{(4)} / (\hat{M}_X^{(2)})^2$$

(Η εξίσωση αυτή προκύπτει από τις εξισώσεις (12.29), (12.16), (12.13) και (3.43) του Kendall and Stewart (1968)). Ας σημειωθεί ότι, σε πολλά βιβλία στατιστικής υδρολογίας, εσφαλμένα παραλείπεται ο όρος που βρίσκεται μέσα στην αγκύλη, πράγμα που οδηγεί σε υπερεκτίμηση της τέταρτης ροπής του πληθυσμού.

ρόληπτη εκτιμήτρια του συντελεστή ασυμμετρίας. Παραθέτουμε τους σημαντικότερους απ' αυτούς:

$$1. \quad \phi(\hat{C}_{s_x}, n) = \frac{\sqrt{n(n-1)}}{n-2} \quad (3.27)$$

Αυτός προκύπτει αν στην (3.26) αντικατασταθούν οι μεροληπτικές εκτιμήτριες των ροπών από τις αντίστοιχες αμερόληπτες.

$$2. \quad \phi(\hat{C}_{s_x}, n) = \frac{n^2}{(n-1)(n-2)} \quad (3.28)$$

Αυτός προκύπτει αν στην (3.26) αντικατασταθεί η μεροληπτική εκτιμήτρια της τρίτης κεντρικής ροπής από την αντίστοιχη αμερόληπτη (Yevjevich, 1978, σ. 110).

$$3. \quad \phi(\hat{C}_{s_x}, n) = \frac{\sqrt{n(n-1)}}{n-2} \left(1 + \frac{8.5}{n}\right) \quad (3.29)$$

Αυτός έχει προταθεί από τον Hazen (1924).

$$4. \quad \phi(\hat{C}_{s_x}, n) = 1 + \left(\frac{6.51}{n} + \frac{20.20}{n^2}\right) + \left(\frac{1.48}{n} + \frac{6.77}{n^2}\right) \hat{C}_{s_x}^2 \quad (3.30)$$

Αυτός έχει προταθεί από τους Bobie and Robitaille (1975), οι οποίοι στηρίχτηκαν σε αποτελέσματα των Wallis et al. (1974).

3.2.5 Συνδιασπορά και συσχέτιση

Η μεροληπτική εκτιμήτρια της συνδιασποράς σ_{XY} ενός πληθυσμού δύο μεταβλητών X και Y είναι η ακόλουθη:

$$S_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n} \quad (3.31)$$

για την οποία αποδεικνύεται (Paroulis, 1990, σ. 295) ότι

$$E[S_{XY}] = \frac{n-1}{n} \sigma_{XY} \quad (3.32)$$

και επομένως ο συντελεστής διόρθωσης μεροληψίας είναι

$$\phi(S_{XY}, n) = \frac{n}{n-1} \quad (3.33)$$

Έτσι, η αμερόληπτη (και συνεπής) εκτιμήτρια της σ_X^2 είναι η ακόλουθη, γνωστή ως *δειγματική συνδιασπορά*:*

$$S_{XY}^* = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \quad (3.34)$$

Για μεγάλο μέγεθος δείγματος n οι δύο εκτιμήτριες πρακτικώς ταυτίζονται.

Η εκτιμήτρια του συντελεστή συσχέτισης ρ_{XY} είναι η ακόλουθη, γνωστή ως *δειγματικός συντελεστής συσχέτισης*:

$$R_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{S_{XY}^*}{S_X^* S_Y^*} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (3.35)$$

Η ακριβής συνάρτηση κατανομής της και οι ροπές της είναι δύσκολο να προσδιοριστούν αναλυτικά, αλλά πάντως θεωρείται κατά προσέγγιση αμερόληπτη εκτιμήτρια.

* Σε πολλά βιβλία στον παρονομαστή της (3.34) υπάρχει ο όρος $n-2$, ο οποίος όμως δεν είναι σωστός.

3.3 Τυπικά όρια εμπιστοσύνης

3.3.1 Μέση τιμή - γνωστή διασπορά πληθυσμού

Έστω X τυχαία μεταβλητή που αντιπροσωπεύει κάποιο πληθυσμό, με μέση τιμή μ_X και τυπική απόκλιση σ_X . Σύμφωνα με το κεντρικό οριακό θεώρημα και την εξίσωση (3.11), η δειγματική μέση τιμή \bar{X} (ως μέσος όρος n τυχαίων μεταβλητών) ακολουθεί κανονική κατανομή $N(\mu_X, \sigma_X / \sqrt{n})$, υπό την προϋπόθεση ότι το n είναι αρκετά μεγάλο.

Επίσης ακολουθεί την ίδια κανονική κατανομή, ανεξαρτήτως του μεγέθους n , αν η X είναι κανονική.

Το πρόβλημα που μας απασχολεί εδώ είναι ο προσδιορισμός των ορίων εμπιστοσύνης της μ_X για συντελεστή εμπιστοσύνης γ . Συμβολίζουμε με $z_{(1+\gamma)/2}$ το $((1+\gamma)/2)$ -ποσοστημόριο της τυποποιημένης κανονικής κατανομής $N(0, 1)$ (δηλαδή το σημείο z που αντιστοιχεί σε πιθανότητα μη υπέρβασης $(1+\gamma)/2$). Μάλιστα, λόγω συμμετρίας ισχύει $z_{(1-\gamma)/2} = -z_{(1+\gamma)/2}$ (βλ. Σχ. 3.1). Θα έχουμε

$$P\left(\mu_X - \frac{z_{(1+\gamma)/2} \sigma_X}{\sqrt{n}} < \bar{X} < \mu_X + \frac{z_{(1+\gamma)/2} \sigma_X}{\sqrt{n}}\right) = \gamma \quad (3.36)$$

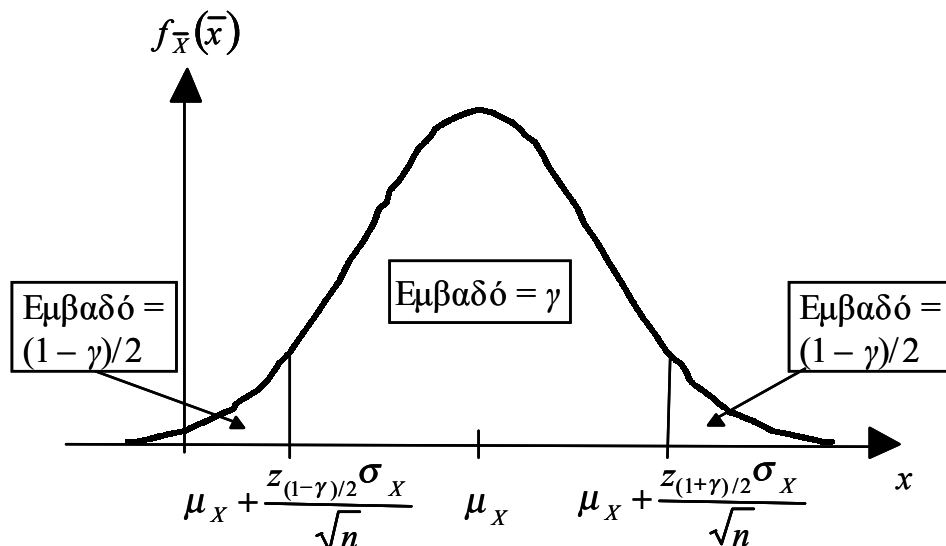
ή ισοδύναμα

$$P\left(\bar{X} - \frac{z_{(1+\gamma)/2} \sigma_X}{\sqrt{n}} < \mu_X < \bar{X} + \frac{z_{(1+\gamma)/2} \sigma_X}{\sqrt{n}}\right) = \gamma \quad (3.37)$$

Η (3.37) δίνει τα ζητούμενα όρια εμπιστοσύνης. Για τον αριθμητικό τους υπολογισμό αντικαθιστούμε στην (3.37) την εκτιμήτρια \bar{X} με την αντίστοιχη εκτίμηση \bar{x} .

Στον Πίν. 3.1 δίνονται για τους πιο συνηθισμένους συντελεστές εμπιστοσύνης οι αντίστοιχες τιμές των κανονικών ποσοστημορίων $z_{(1+\gamma)/2}$ (οι ίδιες τιμές προκύπτουν και από τον πίνακα της κανονικής κατανομής του Παρατήματος). Παρατηρούμε ότι όσο πλησιέστερος προς το 1 είναι ο συντελεστής εμπιστοσύνης, δηλαδή όσο αυξάνει η αξιοπιστία της εκτίμησης, τόσο μεγαλύτερο είναι διάστημα εμπιστοσύνης, δηλαδή τόσο ασαφέστερη είναι η εκτίμηση. Αντίθετα, αν μικρύνουμε το συντελεστή

εμπιστοσύνης, παίρνουμε πιο “συγκεντρωμένη εκτίμηση”, δηλαδή μικρό διάστημα εμπιστοσύνης, αλλά τότε μειώνεται η αξιοπιστία της εκτίμησης, ή, ισοδύναμα, αυξάνεται ο βαθμός αβεβαιότητας.



Σχ. 3.1 Επεξηγηματικό σκαρίφημα για τα όρια εμπιστοσύνης της μέσης τιμής.

Από την εξίσωση (3.37) διαπιστώνουμε ότι υπάρχει τρόπος να αυξηθεί η αξιοπιστία χωρίς να διευρυνθεί το διάστημα εμπιστοσύνης, ή να μειωθεί το διάστημα εμπιστοσύνης χωρίς να μειωθεί η αξιοπιστία της εκτίμησης. Ο τρόπος αυτός είναι να αυξηθεί το μέγεθος του δείγματος n με εκτέλεση και άλλων μετρήσεων (βλ. και την Εφαρμογή που ακολουθεί).

Πίν. 3.1 Τυπικές τιμές κανονικών ποσοστημορίων $z_{(1+\gamma)/2}$ για τον υπολογισμό ορίων εμπιστοσύνης.

| | | | | |
|--------------------|-------|-------|-------|--------|
| γ | 0.90 | 0.95 | 0.99 | 0.999 |
| $(1+\gamma)/2$ | 0.95 | 0.975 | 0.995 | 0.9995 |
| $z_{(1+\gamma)/2}$ | 1.645 | 1.960 | 2.576 | 3.291 |

Η προϋπόθεση της γνωστής διασποράς του πληθυσμού, στην οποία στηρίχτηκε η προηγούμενη ανάλυση, στην τεχνική υδρολογία δεν είναι γενικά ρεαλιστική. Αυτό επειδή η πληροφορία που έχουμε για μια υδρολογική μεταβλητή προέρχεται πάντα από ένα δείγμα. Ωστόσο τα αποτελέσματα έχουν πρακτικό ενδιαφέρον, δεδομένου ότι η (3.37) μπορεί να χρησιμοποιηθεί ως προσέγγιση αν το n είναι αρκετά μεγάλο (> 30), με αντικατάσταση της διασποράς του πληθυσμού από την εκτίμησή της από το δείγμα.

3.3.2 Μέση τιμή - άγνωστη διασπορά πληθυσμού

Η ανάλυση που παρουσιάζουμε εδώ μπορεί να χρησιμοποιηθεί για άγνωστη διασπορά πληθυσμού και για οποιοδήποτε μέγεθος δείγματος. Ωστόσο, και η ανάλυση αυτή έχει μια περιοριστική προϋπόθεση: ότι η τυχαία μεταβλητή X που αντιπροσωπεύει τον πληθυσμό είναι κανονική $N(\mu_X, \sigma_X)$. Σε αυτή την περίπτωση προκύπτουν τα ακόλουθα συμπεράσματα:

1. Η δειγματική μέση τιμή ακολουθεί κανονική κατανομή $N(\mu_X, \sigma_X / \sqrt{n})$. Το συμπέρασμα αυτό είναι συνέπεια της ιδιότητας της κανονικής κατανομής να διατηρείται με την άθροιση (ιδιότητα συμβατή με το κεντρικό οριακό θεώρημα).
2. Η συνάρτηση της δειγματικής διασποράς $(n-1)S_X^{*2} / \sigma_X^2$ ακολουθεί κατανομή $\chi^2(n-1)$. Το συμπέρασμα αυτό προκύπτει από το θεώρημα του εδαφίου 2.8.3, σύμφωνα με το οποίο το άθροισμα των τετραγώνων ενός αριθμού τυποποιημένων κανονικών μεταβλητών ακολουθεί κατανομή χ^2 .
3. Οι τυχαίες μεταβλητές \bar{X} και S_X^{*2} είναι ανεξάρτητες. Αυτό το συμπέρασμα είναι απόρροια θεωρήματος της στατιστικής (βλ. π.χ. Papoulis, 1990, σ. 222).
4. Ο λόγος $(\bar{X} - \mu_X) / (S_X^* / \sqrt{n})$ ακολουθεί κατανομή Student $t(n-1)$. Το συμπέρασμα αυτό βασίζεται στο θεώρημα του εδαφίου 2.8.4.

Συμβολίζουμε με $t_{(1+\gamma)/2}$ το $[(1+\gamma)/2]$ -ποσοστημόριο της κατανομής Student $t(n-1)$ (δηλαδή το σημείο t που αντιστοιχεί σε πιθανότητα μη υπέρβασης $(1+\gamma)/2$, για $n-1$ βαθμούς ελευθερίας). Λόγω της συμμετρίας ισχύει $t_{(1-\gamma)/2} = -t_{(1+\gamma)/2}$. Θα έχουμε

$$P\left(-t_{(1+\gamma)/2} < \frac{\bar{X} - \mu_X}{S_X^* / \sqrt{n}} < t_{(1+\gamma)/2}\right) = \gamma \quad (3.38)$$

ή ισοδύναμα

$$P\left(\bar{X} - \frac{t_{(1+\gamma)/2} S_X^*}{\sqrt{n}} < \mu_X < \bar{X} + \frac{t_{(1+\gamma)/2} S_X^*}{\sqrt{n}}\right) = \gamma \quad (3.39)$$

Η (3.39) δίνει τα ζητούμενα όρια εμπιστοσύνης. Για τον αριθμητικό τους υπολογισμό αντικαθιστούμε στην (3.39) τις εκτιμήτριες \bar{X} και S_X^* με τις αντίστοιχες εκτιμήσεις \bar{x} και s_X^* .

Παρόλο που οι (3.37) και (3.39) διαφέρουν σημαντικά ως προς τη θεωρητική τους υπόσταση και τις παραδοχές βάσει των οποίων προκύπτουν, υπολογιστικά είναι όμοιες. Μάλιστα, για μεγάλο n (>30) πρακτικά ταυτίζονται δεδομένου ότι ισχύει $t_{(1+\gamma)/2} \approx z_{(1+\gamma)/2}$ (ακριβέστερα $t_{(1+\gamma)/2} \approx z_{(1+\gamma)/2} \sqrt{(n-1)/(n-3)}$, για $n-1$ βαθμούς ελευθερίας).

Οι παραπάνω αναλύσεις αφήνουν ένα κενό, ήτοι την περίπτωση μικρού δείγματος, άγνωστης διασποράς του πληθυσμού και μη κανονικής κατανομής του πληθυσμού. Αυτή η περίπτωση δεν καλύπτεται με γενικό και αυστηρό τρόπο από τη στατιστική.* Προσεγγιστικά, πάντως, μπορούμε να χρησιμοποιήσουμε την πιο πάνω μεθοδολογία και σε αυτές τις περιπτώσεις, αρκεί η κατανομή του πληθυσμού να είναι κωδωνοειδής, δηλαδή όχι ιδιαίτερα ασύμμετρη. Άλλωστε ελάχιστες είναι οι περιπτώσεις που, όπως στις αναλύσεις που προηγήθηκαν, υπολογίζονται θεωρητικώς ακριβή όρια εμπιστοσύνης. Στα περισσότερα από τα προβλήματα που θα ακολουθήσουν θα αρκεστούμε σε προσεγγιστικά όρια εμπιστοσύνης.

Εφαρμογή 3.3.2

Από δείγμα ετήσιων παροχών ποταμού (χρονικά μέσων) μήκους 15 ετών βρέθηκε η δειγματική μέση τιμή ίση με $10.05 \text{ m}^3/\text{s}$ και η δειγματική τυπική απόκλιση $2.80 \text{ m}^3/\text{s}$. Ζητούνται (1) τα όρια εμπιστοσύνης 95% της μέσης υπερετήσιας παροχής και (2) το μέγεθος του δείγματος που, με βαθμό εμπιστοσύνης 95%, εξασφαλίζει ακρίβεια 10% στην εκτίμηση της μέσης υπερετήσιας παροχής.

* Υπάρχουν βέβαια ειδικές περιπτώσεις που υπολογίζονται θεωρητικώς ακριβή όρια εμπιστοσύνης. Μια τέτοια περίπτωση αναφέρεται στις συναρτήσεις κατανομής μιας παραμέτρου (π.χ. εκθετική). Το θέμα αυτό καλύπτεται από βιβλία μαθηματικής στατιστικής, π.χ., Papoulis (1990, σ. 280).

(1) Θεωρούμε ότι η μέση ετήσια παροχή ακολουθεί κανονική κατανομή (βλ. και ενότητα 2.8 καθώς και κεφάλαιο 5) και χρησιμοποιούμε την εξίσωση (3.39). Από τον πίνακα της κατανομής Student (Παράρτημα Π3) βρίσκουμε ότι για $n - 1 = 14$ βαθμούς ελευθερίας $t_{(1+\gamma)/2} = t_{0.975} = 2.14$. Κατά συνέπεια τα όρια εμπιστοσύνης 95% είναι*

$$10.05 - 2.14 \times 2.80 / \sqrt{15} < \mu_X < 10.05 + 2.14 \times 2.80 / \sqrt{15}$$

ήτοι

$$8.50 < \mu_X < 11.60$$

Για λόγους σύγκρισης θα υπολογίσουμε τα όρια εμπιστοσύνης και με τη (3.37), αν και δεν είναι σωστό. Από τον Πίν. 3.1 βρίσκουμε $z_{(1+\gamma)/2} = z_{0.975} = 1.96$. Κατά συνέπεια τα όρια εμπιστοσύνης 95% είναι

$$10.05 - 1.96 \times 2.80 / \sqrt{15} < \mu_X < 10.05 + 1.96 \times 2.80 / \sqrt{15}$$

ήτοι

$$8.63 < \mu_X < 11.47$$

Το διάστημα εμπιστοσύνης σε αυτή την περίπτωση είναι ελαφρώς στενότερο.

(2) Υποθέτουμε ότι $n \geq 30$ οπότε μπορούμε να χρησιμοποιήσουμε την (3.37). Θα πρέπει

$$1.96 \times 2.8 / \sqrt{n} = 10\% \times 10.05,$$

άρα $n = 30$. Διαπιστώνουμε ότι καλύπτεται οριακά η προϋπόθεση που τέθηκε ($n \geq 30$). (Αν δεν καλυπτόταν θα συνεχίζαμε με δοκιμές, χρησιμοποιώντας πλέον την (3.39)).

* Εφαρμόζοντας τη (3.39), δηλαδή αντικαθιστώντας τις εκτιμήτριες με τις εκτιμήσεις, δεν θα ήταν μαθηματικώς ορθό να γράψουμε

$$P(10.05 - 2.62 \times 2.80 / \sqrt{15} < \mu_X < 10.05 + 2.62 \times 2.80 / \sqrt{15}) = 0.95$$

γιατί η μ_X είναι (άγνωστη) παράμετρος και όχι τυχαία μεταβλητή, άρα δεν ορίζεται συνάρτηση πιθανότητας για αυτή. Επίσης, δεν είναι σωστό να πούμε π.χ. ότι “με πιθανότητα 95% η μέση τιμή βρίσκεται στο διάστημα (8.16, 11.94)”. Το σωστό είναι να χρησιμοποιήσουμε τον όρο “με εμπιστοσύνη 95%”.

3.3.3 Διασπορά και τυπική απόκλιση

Όπως και στο προηγούμενο εδάφιο, θα υποθέσουμε και εδώ ότι η μεταβλητή X ακολουθεί κανονική κατανομή $N(\mu_X, \sigma_X)$. Όπως ήδη αναφέρθηκε, σε αυτή την περίπτωση η συνάρτηση της δειγματικής διασποράς $(n-1)S_X^{*2} / \sigma_X^2$ ακολουθεί κατανομή $\chi^2(n-1)$.

Συμβολίζουμε με $\chi_{(1+\gamma)/2}^2$ το $[(1+\gamma)/2]$ -ποσοστημόριο της κατανομής $\chi^2(n-1)$ και με $\chi_{(1-\gamma)/2}^2$ το $[(1-\gamma)/2]$ -ποσοστημόριο της ίδιας κατανομής (εδώ τα δύο αυτά μεγέθη δεν συνδέονται με σχέση συμμετρίας, γιατί η κατανομή χ^2 δεν είναι συμμετρική). Θα έχουμε

$$P\left(\chi_{(1-\gamma)/2}^2 < \frac{(n-1)S_X^{*2}}{\sigma_X^2} < \chi_{(1+\gamma)/2}^2\right) = \gamma \quad (3.40)$$

ή ισοδύναμα

$$P\left(\frac{(n-1)S_X^{*2}}{\chi_{(1+\gamma)/2}^2} < \sigma_X^2 < \frac{(n-1)S_X^{*2}}{\chi_{(1-\gamma)/2}^2}\right) = \gamma \quad (3.41)$$

Η (3.41) δίνει τα ζητούμενα όρια εμπιστοσύνης. Εύκολα προκύπτει ότι τα όρια εμπιστοσύνης της τυπικής απόκλισης δίνονται από την εξίσωση

$$P\left(\frac{\sqrt{n-1} S_X^*}{\sqrt{\chi_{(1+\gamma)/2}^2}} < \sigma_X < \frac{\sqrt{n-1} S_X^*}{\sqrt{\chi_{(1-\gamma)/2}^2}}\right) = \gamma \quad (3.42)$$

Εφαρμογή 3.3.3

Να υπολογιστούν τα όρια εμπιστοσύνης 95% της τυπικής απόκλισης της μέσης ετήσιας παροχής στο πρόβλημα της Εφαρμογής 3.3.2.

Υπενθυμίζουμε ότι η δειγματική τυπική απόκλιση στο παραπάνω πρόβλημα ήταν $2.8 \text{ m}^3/\text{s}$. Θεωρούμε και πάλι ότι η μέση ετήσια παροχή ακολουθεί κανονική κατανομή και χρησιμοποιούμε την εξίσωση (3.42). Από τον πίνακα της κατανομής χ^2 (Παράρτημα Π2) βρίσκουμε ότι για $n-1 = 14$ βαθμούς ελευθερίας $\chi_{(1+\gamma)/2}^2 = \chi_{0.975}^2 = 26.12$ και $\chi_{(1-\gamma)/2}^2 = \chi_{0.025}^2 = 5.63$. Κατά συνέπεια τα όρια εμπιστοσύνης 95% είναι

$$\frac{\sqrt{14} * 2.80}{\sqrt{26.12}} < \sigma_X < \frac{\sqrt{14} * 2.80}{\sqrt{5.63}}$$

ήτοι

$$2.05 < \sigma_X < 4.41$$

3.3.4 Ποσοστημόριο κανονικής κατανομής – Τυπικό σφάλμα

Στην τεχνική υδρολογία το συχνότερο πρόβλημα ορίων εμπιστοσύνης που αντιμετωπίζουμε αφορά σε μεγέθη υδρολογικού σχεδιασμού, π.χ. στην παροχή σχεδιασμού ενός έργου. Έστω η υδρολογική μεταβλητή X , η οποία ακολουθεί δεδομένη κατανομή $F_X(x)$. Εδώ θα θεωρήσουμε ότι η $F_X(x)$ είναι κανονική $N(\mu_X, \sigma_X)$, της οποίας η αντιμετώπιση είναι σχετικά απλή, ενώ στο κεφάλαιο 6 θα δώσουμε αναλυτικούς τύπους για όλες τις κατανομές που χρησιμοποιούνται στην τεχνική υδρολογία. Για δεδομένη πιθανότητα μη υπέρβασης $u = F_X(x)$, η αντίστοιχη τιμή της μεταβλητής X (συμβολικά $x_u = u$ -ποσοστημόριο) θα είναι

$$x_u = \mu_X + z_u \sigma_X \quad (3.43)$$

όπου z_u το u -ποσοστημόριο της τυποποιημένης κανονικής κατανομής $N(0, 1)$. Ωστόσο, στην παραπάνω εξίσωση οι παράμετροι πληθυσμού μ_X και σ_X στην πράξη είναι άγνωστες. Χρησιμοποιώντας τις σημειακές εκτιμήσεις τους, παίρνουμε την αντίστοιχη εκτίμηση $\hat{x}_u = \bar{x} + z_u s_X$, η οποία μπορεί να θεωρηθεί ως τιμή της τυχαίας μεταβλητής

$$\hat{X}_u = \bar{X} + z_u S_X \quad (3.44)$$

Η εξίσωση αυτή μπορεί να χρησιμοποιηθεί για τον προσδιορισμό ορίων εμπιστοσύνης της x_u . Ωστόσο, ο ακριβής προσδιορισμός είναι πρακτικά αδύνατος, λόγω της πολυπλοκότητας της συνάρτησης κατανομής της \hat{X}_u . Θα αρκεστούμε λοιπόν (όπως και σε επόμενες ακόμη πιο πολύπλοκες περιπτώσεις) σε προσεγγιστικά όρια εμπιστοσύνης, βασισμένα στην παραδοχή ότι η \hat{X}_u ακολουθεί κανονική κατανομή.

Η μέση τιμή της \hat{X}_u δίνεται από την εξίσωση (2.43), η οποία αν συνδυαστεί και με τις (3.11) και (3.17) γράφεται

$$E[\hat{X}_u] = E[\bar{X}] + z_u E[S_X] \approx \mu_X + z_u \sigma_X = x_u \quad (3.45)$$

όπου θεωρήσαμε ότι το n είναι αρκετά μεγάλο και παραλείψαμε τον όρο $O(1/n)$ στην $E[S_X]$.^{*} Αντίστοιχα, η διασπορά της \hat{X}_u δίνεται από την εξίσωση (2.45), η οποία γράφεται

$$\text{Var}[\hat{X}_u] = \text{Var}[\bar{X}] + z_u^2 \text{Var}[S_X] + 2z_u \text{Cov}[\bar{X}, S_X] \quad (3.46)$$

Δεδομένου ότι η X ακολουθεί κανονική κατανομή, ο τρίτος όρος της (3.46) είναι μηδέν (όπως αναφέρθηκε πιο πάνω, οι μεταβλητές \bar{X} και S_X είναι ανεξάρτητες). Συνδυάζοντας και τις (3.11) και (3.18), η (3.46) γράφεται

$$\varepsilon_u^2 := \text{Var}[\hat{X}_u] \approx \frac{\sigma_X^2}{n} + z_u^2 \frac{\sigma_X^2}{2n} = \frac{\sigma_X^2}{n} \left(1 + \frac{z_u^2}{2}\right) \quad (3.47)$$

Η ποσότητα ε_u είναι γνωστή στη βιβλιογραφία με τον όρο *τυπικό σφάλμα ποσοστημορίου* ή απλώς *τυπικό σφάλμα*.

Χρησιμοποιώντας την παραδοχή ότι η \hat{X}_u ακολουθεί κανονική κατανομή $N(x_u, \varepsilon_u)$ μπορούμε να γράψουμε

$$P\left(-z_{(1+\gamma)/2} < \frac{\hat{X}_u - x_u}{\varepsilon_u} < z_{(1+\gamma)/2}\right) = \gamma \quad (3.48)$$

όπου γ ο συντελεστής εμπιστοσύνης. Ισοδύναμα

$$P\left(\hat{X}_u - z_{(1+\gamma)/2} \varepsilon_u < x_u < \hat{X}_u + z_{(1+\gamma)/2} \varepsilon_u\right) = \gamma \quad (3.49)$$

Αν αντικαταστήσουμε στην παραπάνω το ε_u από την (3.47), και στη συνέχεια αντικαταστήσουμε την τυπική απόκλιση σ_X με την αντίστοιχη εκτιμήτρια, παίρνουμε την ακόλουθη τελική σχέση

$$P\left(\hat{X}_u - z_{(1+\gamma)/2} \sqrt{1 + \frac{z_u^2}{2}} \frac{S_X}{\sqrt{n}} < x_u < \hat{X}_u + z_{(1+\gamma)/2} \sqrt{1 + \frac{z_u^2}{2}} \frac{S_X}{\sqrt{n}}\right) = \gamma \quad (3.50)$$

^{*} Λόγω του προσεγγιστικού χαρακτήρα της ανάλυσης, δεν κάνουμε διάκριση των εκτιμητριών S_X και S_X^* , θεωρώντας ότι το n είναι αρκετά μεγάλο, ώστε πρακτικά να ταυτίζονται.

Η τελευταία εξίσωση είναι προσεγγιστική, με βαθμό ακρίβειας που αυξάνεται με την αύξηση του n . Επίσης, ισχύει μόνο όταν η κατανομή της X είναι κανονική. Ωστόσο, η (3.49) χρησιμοποιείται και για άλλες κατανομές της X , με τη διαφορά ότι το τυπικό σφάλμα έχει διαφορετική έκφραση και διαφορετικό τρόπο υπολογισμού. Ο ενδιαφερόμενος αναγνώστης για μια γενική έκφραση του τυπικού σφάλματος παραπέμπεται στον Kite (1988, σ. 33-38).

Οι σημειακές εκτιμήσεις των ορίων εμπιστοσύνης

$$\hat{x}_{u1} = \hat{x}_u - z_{(1+\gamma)/2} \sqrt{1 + \frac{z_u^2}{2}} \frac{s_X}{\sqrt{n}} \quad \hat{x}_{u2} = \hat{x}_u + z_{(1+\gamma)/2} \sqrt{1 + \frac{z_u^2}{2}} \frac{s_X}{\sqrt{n}} \quad (3.51)$$

είναι προφανώς συναρτήσεις του u ή, ισοδύναμα, της πιθανότητας υπέρβασης, $1 - u$. Τα γραφήματα των συναρτήσεων αυτών βρίσκονται εκατέρωθεν της καμπύλης x_u και λέγονται *καμπύλες εμπιστοσύνης* (ή *καμπύλες ασφάλειας της x_u*).

Εφαρμογή 3.3.4

Να υπολογιστούν τα όρια εμπιστοσύνης 95% της παροχής που έχει πιθανότητα υπέρβασης (α) 1% και (β) 99% στο πρόβλημα της Εφαρμογής 3.3.2.

Κατ' αρχήν, επειδή το μέγεθος του δείγματος είναι πολύ μικρό, δεν περιμένουμε πολύ καλό βαθμό ακρίβειας στους υπολογισμούς μας.

Θα υπολογίσουμε πρώτα τις σημειακές εκτιμήσεις. Για την παροχή πιθανότητας υπέρβασης $F_1 = 0.01$ έχουμε $u = 1 - F_1 = 0.99$ και $z_u = 2.326$ (από τον πίνακα της κανονικής κατανομής του Παραρτήματος Π1). Άρα η σημειακή εκτίμηση είναι $\hat{x}_u = 10.05 + 2.326 \times 2.80 = 16.56$. Αντίστοιχα, για την παροχή πιθανότητας υπέρβασης $F_1 = 0.99$ έχουμε $u = 1 - F_1 = 0.01$ και $z_u = -2.326$, άρα $\hat{x}_u = 10.05 - 2.326 \times 2.80 = 3.54$.

Προχωρούμε τώρα στον υπολογισμό των ορίων εμπιστοσύνης. Για $\gamma = 95\%$ και $z_{(1+\gamma)/2} = 1.96$, τα όρια για την παροχή πιθανότητας υπέρβασης 1% θα είναι:

$$\hat{x}_{u1} = 16.56 - 1.96 \sqrt{1 + \frac{2.326^2}{2}} \frac{2.80}{\sqrt{15}} = 13.83$$

$$\hat{x}_{u2} = 16.56 + 1.96 \sqrt{1 + \frac{2.326^2}{2}} \frac{2.80}{\sqrt{15}} = 19.29$$

Αντίστοιχα, τα όρια για την παροχή πιθανότητας υπέρβασης 99% θα είναι:

$$\hat{x}_{u1} = 3.54 - 1.96 \sqrt{1 + \frac{2.326^2}{2} \frac{2.80}{\sqrt{15}}} = 0.81$$

$$\hat{x}_{u2} = 3.54 + 1.96 \sqrt{1 + \frac{2.326^2}{2} \frac{2.80}{\sqrt{15}}} = 6.27$$

3.3.5 Συντελεστής συσχέτισης

Για τον υπολογισμό των ορίων εμπιστοσύνης του συντελεστή συσχέτισης ρ ενός πληθυσμού δύο μεταβλητών X και Y , χρησιμοποιούμε τη βοηθητική μεταβλητή Z που ορίζεται από το λεγόμενο μετασχηματισμό Fisher:

$$Z = \frac{1}{2} \ln \frac{1+R}{1-R} \Leftrightarrow R = \frac{e^{2Z} - 1}{e^{2Z} + 1} = \tanh Z \quad (3.52)$$

όπου R ο δειγματικός συντελεστής συσχέτισης. Παρατηρούμε ότι για $-1 < R < 1$ το διάστημα μεταβολής της Z είναι $-\infty < Z < \infty$, ενώ για $R = 0$ έχουμε $Z = 0$. Αποδεικνύεται ότι αν οι X και Y ακολουθούν κανονικές κατανομές, τότε η Z ακολουθεί κατά προσέγγιση κανονική κατανομή $N(\mu_Z, \sigma_Z)$ όπου

$$\mu_Z = E[Z] \approx \frac{1}{2} \ln \frac{1+\rho}{1-\rho} \quad \sigma_Z^2 = \text{Var}[Z] \approx \frac{1}{n-3} \quad (3.53)^*$$

Κατά συνέπεια, αν $\zeta_{(1+\gamma)/2}$ είναι το $(1+\gamma)/2$ -ποσοστημόριο της τυποποιημένης κανονικής κατανομής,[†] θα έχουμε

* Ακριβέστερες προσεγγίσεις δίνονται από τις ακόλουθες εξισώσεις

$$\mu_Z = E[Z] \approx \frac{1}{2} \ln \frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)} \quad \sigma_Z^2 = \text{Var}[Z] \approx \frac{1}{n-1} + \frac{4-\rho^2}{2(n-1)^2}$$

πράγμα που σημαίνει ότι οι προσεγγίσεις της (3.53) ισχύουν για μικρές τιμές του ρ ή μεγάλες τιμές του n (βλ. Kendall and Stuart, 1963, σ. 391).

[†] Εδώ χρησιμοποιήσαμε το ελληνικό σύμβολο $\zeta_{(1+\gamma)/2}$ αντί του συνήθους λατινικού $z_{(1+\gamma)/2}$, προκειμένου να αποφευχθεί η σύγχυση με τη μεταβλητή του Fisher Z η οποία επίσης ακολουθεί κανονική κατανομή, αλλά όχι τυποποιημένη.

$$P\left(\mu_Z - \frac{\zeta_{(1+\gamma)/2}}{\sqrt{n-3}} < Z < \mu_Z + \frac{\zeta_{(1+\gamma)/2}}{\sqrt{n-3}}\right) = \gamma \quad (3.54)$$

ή ισοδύναμα

$$P\left(Z - \frac{\zeta_{(1+\gamma)/2}}{\sqrt{n-3}} < \mu_Z < Z + \frac{\zeta_{(1+\gamma)/2}}{\sqrt{n-3}}\right) = \gamma \quad (3.55)$$

Αντικαθιστώντας στην (3.55) το μ_Z από την (3.53) και επιλύοντας ως προς ρ , αφού λάβουμε υπόψη και τη μονοτονικότητα του μετασχηματισμού (3.52), παίρνουμε

$$P(R_1 < \rho < R_2) = \gamma \quad (3.56)$$

όπου

$$\begin{aligned} R_1 &= \frac{e^{2Z_1} - 1}{e^{2Z_1} + 1} & R_2 &= \frac{e^{2Z_2} - 1}{e^{2Z_2} + 1} \\ \left. \begin{array}{l} Z_2 \\ Z_1 \end{array} \right\} &= Z \pm \frac{\zeta_{(1+\gamma)/2}}{\sqrt{n-3}} & Z &= \frac{1}{2} \ln \frac{1+R}{1-R} \end{aligned} \quad (3.57)$$

Για τον αριθμητικό υπολογισμό των ορίων εμπιστοσύνης εφαρμόζουμε τις εξισώσεις (3.57) ξεκινώντας από το τέλος προς την αρχή και αντικαθιστώντας τις εκτιμήτριες με τις αντίστοιχες εκτιμήσεις (π.χ. $R = r$, κτλ.).

3.4 Εκτίμηση παραμέτρων συναρτήσεων κατανομής

Έστω τυχαία μεταβλητή X με γνωστό τύπο συνάρτησης κατανομής, ο οποίος περιέχει τις άγνωστες παραμέτρους $\theta_1, \theta_2, \dots, \theta_r$. Έτσι, η πυκνότητα πιθανότητας της X είναι μια συνάρτηση $f_X(x, \theta_1, \theta_2, \dots, \theta_r)$. Θα εξετάσουμε το πρόβλημα της εκτίμησης αυτών των παραμέτρων με βάση ένα δείγμα X_1, X_2, \dots, X_n . Συγκεκριμένα θα παραθέσουμε τις δύο πιο κλασικές μεθόδους της στατιστικής για την εκτίμηση των παραμέτρων: τη μέθοδο των ροπών και τη μέθοδο της μέγιστης πιθανοφάνειας.

Η στατιστική έχει αναπτύξει και άλλες γενικές μεθόδους για την εκτίμηση παραμέτρων κατανομών, όπως π.χ. τη μέθοδο της μέγιστης εντροπίας ή τη μέθοδο των L-ροπών. Τέτοιες μέθοδοι έχουν συχνά χρησιμοποιηθεί και στην τεχνική υδρολογία, αλλά δεν παρατίθενται στο κείμενο αυτό (ο ενδιαφερόμενος αναγνώστης παραπέμπεται στους Singh and Rajagopal (1986) και Stedinger et al. (1993), αντίστοιχα). Επιπλέον, στην τεχνική υδρολογία χρησιμοποιούνται κατά περίπτωση και ειδικότερες μέθοδοι, γραφικές ή υπολογιστικές και κατά κανόνα εμπειρικές ή ημιεμπειρικές. Δόκιμα παραδείγματα τέτοιων μεθόδων θα παραθέσουμε στο κεφάλαιο 6 για συγκεκριμένες κατανομές.

3.4.1 Η μέθοδος των ροπών

Η μέθοδος των ροπών βασίζεται στην εξίσωση των θεωρητικών ροπών της κατανομής της X με τις αντίστοιχες δειγματικές εκτιμήσεις των ροπών. Έτσι αν r είναι ο αριθμός των άγνωστων παραμέτρων της κατανομής, μπορούμε να γράψουμε r εξισώσεις της μορφής

$$m_X^{(k)} = \hat{m}_X^{(k)} \quad k = 1, 2, \dots, r \quad (3.58)$$

όπου τα $m_X^{(k)}$ είναι οι θεωρητικές ροπές περί την αρχή, οι οποίες είναι συναρτήσεις των άγνωστων παραμέτρων και δίνονται από τη σχέση

$$m_X^{(k)} = \int_{-\infty}^{\infty} x^k f_X(x, \theta_1, \dots, \theta_r) dx \quad (3.59)$$

ενώ τα $\hat{m}_X^{(k)}$ είναι αριθμητικές εκτιμήσεις που υπολογίζονται από το δείγμα σύμφωνα με τη σχέση

$$\hat{m}_X^{(k)} = \frac{1}{n} \sum_{i=1}^n x_i^k \quad (3.60)$$

Με την κατάστρωση και στη συνέχεια επίλυση των r εξισώσεων υπολογίζονται οι άγνωστες παράμετροι $\theta_1, \theta_2, \dots, \theta_r$. Σημειώνουμε ότι κατά κανόνα το σύστημα των εξισώσεων δεν είναι γραμμικό και έτσι μπορεί να μην επιλύεται αναλυτικά αλλά μόνο αριθμητικά.

Ισοδύναμα μπορούμε να χρησιμοποιήσουμε τις κεντρικές ροπές, αντί των ροπών περί την αρχή, για $k > 1$, οπότε το σύστημα των εξισώσεων γίνεται

$$\mu_X = \bar{x} \quad \mu_X^{(k)} = \hat{\mu}_X^{(k)} \quad k = 2, \dots, r \quad (3.61)$$

όπου $\mu_X = m_X^{(1)}$ η θεωρητική μέση τιμή του πληθυσμού, $\bar{x} = \hat{m}_X^{(1)}$ η δειγματική μέση τιμή, $\mu_X^{(k)}$ οι θεωρητικές κεντρικές ροπές που δίνονται από τη σχέση

$$\mu_X^{(k)} = \int_{-\infty}^{\infty} (x - \mu_X)^k f_X(x, \theta_1, \dots, \theta_r) dx \quad (3.62)$$

και $\hat{\mu}_X^{(k)}$ οι αντίστοιχες δειγματικές εκτιμήσεις που υπολογίζονται από τη σχέση

$$\hat{\mu}_X^{(k)} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k \quad (3.63)$$

Υπενθυμίζεται ότι οι ροπές περί την αρχή που υπολογίζονται από την (3.60) είναι αμερόληπτες εκτιμήσεις, ενώ οι κεντρικές ροπές της (3.63) είναι μεροληπτικές. Πάντως, συχνά χρησιμοποιούνται οι αμερόληπτες εκτιμήσεις των κεντρικών ροπών αντί των μεροληπτικών. Σημειώνεται ότι η μέθοδος των ροπών γενικά δεν οδηγεί σε αμερόληπτες εκτιμήσεις των παραμέτρων $\theta_1, \theta_2, \dots, \theta_r$. (εκτός από ορισμένες ειδικές περιπτώσεις), όποιες και αν είναι οι εκτιμήσεις των ροπών που χρησιμοποιούνται.

Υπολογισμός παραμέτρων της κανονικής κατανομής με τη μέθοδο των ροπών

Ως παράδειγμα εφαρμογής της μεθόδου των ροπών, θα υπολογίσουμε τις παραμέτρους της κανονικής κατανομής. Η συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής είναι:

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

και περιλαμβάνει δύο παραμέτρους, τις μ και σ . Κατά συνέπεια χρειάζονται δύο εξισώσεις. Με βάση την (3.61) οι εξισώσεις αυτές θα είναι

$$\mu_X = \bar{x} \quad \sigma_X^2 = s_X^2$$

όπου για τη δεύτερη εξίσωση έχουμε χρησιμοποιήσει τα πιο κοινά σύμβολα σ_X^2 και s_X^2 για τη θεωρητική και δειγματική διασπορά (δηλαδή δεύτερη κεντρική ροπή) της X , αντίστοιχα. Γνωρίζουμε όμως για τις θεωρητικές ροπές (βλ. εδάφιο 2.8.2) ότι

$$\mu_X = \mu \quad \sigma_X^2 = \sigma^2$$

Αντικαθιστώντας τις τιμές αυτές στην προηγούμενη εξίσωση βρίσκουμε άμεσα τις τελικές εκτιμήσεις

$$\mu = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \sigma = s_X = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Η παραπάνω εκτίμηση της σ είναι μεροληπτική, ενώ της μ είναι αμερόληπτη. Αν χρησιμοποιούσαμε την αμερόληπτη εκτίμηση της διασποράς, στην τελευταία εξίσωση θα είχαμε στον παρονομαστή $(n-1)$ στη θέση του n . Ακόμη και τότε όμως δεν θα παίρναμε αμερόληπτη εκτίμηση της σ για τους λόγους που εξηγήσαμε στην ενότητα 3.2.

Όπως είδαμε σε αυτό το συγκεκριμένο παράδειγμα η εφαρμογή της μεθόδου των ροπών είναι απλούστατη και αυτό επεκτείνεται και στους άλλους τύπους συναρτήσεων κατανομής (χωρίς βέβαια η λύση να είναι πάντα τόσο προφανής όσο στο παραπάνω παράδειγμα).

3.4.2 Η μέθοδος της μέγιστης πιθανοφάνειας

Έστω τυχαία μεταβλητή X με πυκνότητα πιθανότητας $f_X(x, \theta_1, \theta_2, \dots, \theta_r)$ όπου $\theta_1, \theta_2, \dots, \theta_r$ παράμετροι, και δείγμα X_1, X_2, \dots, X_n της μεταβλητής. Λόγω της ανεξαρτησίας των μεταβλητών X_1, X_2, \dots, X_n , η από κοινού συνάρτηση πυκνότητας πιθανότητάς τους είναι

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n, \theta_1, \dots, \theta_r) = \prod_{i=1}^n f_X(x_i, \theta_1, \dots, \theta_r) \quad (3.64)$$

Για δεδομένες τιμές των τυχαίων μεταβλητών ίσες με τις παρατηρήσεις x_1, \dots, x_n η $f_{X_1, \dots, X_n}(\cdot)$ είναι συνάρτηση των παραμέτρων $\theta_1, \theta_2, \dots, \theta_r$ και λέγεται *συνάρτηση πιθανοφάνειας* αυτών των παραμέτρων.

Η μέθοδος της μέγιστης πιθανοφάνειας εκτιμά τις παραμέτρους $\theta_1, \theta_2, \dots, \theta_r$ σε τρόπο ώστε η συνάρτηση πιθανοφάνειας να γίνει μέγιστη. Στην περίπτωση αυτή θα πρέπει

$$\frac{\partial f_{X_1, \dots, X_n}(x_1, \dots, x_n, \theta_1, \dots, \theta_r)}{\partial \theta_k} = 0 \quad k = 1, \dots, r \quad (3.65)$$

Από τις r εξισώσεις της παραπάνω μορφής προκύπτουν οι r άγνωστες παράμετροι. Επειδή όμως ο χειρισμός των εξισώσεων αυτών είναι πολύπλοκος, αντί της μεγιστοποίησης της συνάρτησης πιθανοφάνειας, επιδιώκουμε τη μεγιστοποίηση του λογαρίθμου της

$$\begin{aligned} L(x_1, \dots, x_n, \theta_1, \dots, \theta_r) &= \ln f_{X_1, \dots, X_n}(x_1, \dots, x_n, \theta_1, \dots, \theta_r) \\ &= \sum_{i=1}^n \ln f_X(x_i, \theta_1, \dots, \theta_r) \end{aligned} \quad (3.66)$$

Η συνάρτηση $L(\cdot)$ λέγεται *λογαριθμική συνάρτηση πιθανοφάνειας*. Για να είναι μέγιστη θα πρέπει

$$\frac{\partial L(x_1, \dots, x_n, \theta_1, \dots, \theta_r)}{\partial \theta_k} = \sum_{i=1}^n \frac{1}{f_X(x_i, \theta_1, \dots, \theta_r)} \frac{\partial f_X(x_i, \theta_1, \dots, \theta_r)}{\partial \theta_k} = 0 \quad (3.67)$$

για $k = 1, \dots, r$. Από τη λύση αυτών των r εξισώσεων προκύπτουν οι r άγνωστες παράμετροι.

Υπολογισμός παραμέτρων της κανονικής κατανομής με τη μέθοδο της μέγιστης πιθανοφάνειας

Ως παράδειγμα εφαρμογής της μεθόδου της μέγιστης πιθανοφάνειας, θα υπολογίσουμε τις παραμέτρους της κανονικής κατανομής με συνάρτηση πυκνότητας πιθανότητας

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

Με βάση την (3.64) σχηματίζουμε τη συνάρτηση πιθανοφάνειας

$$f(x_1, \dots, x_n, \mu, \sigma) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

Λογαριθμίζοντας παίρνουμε τη λογαριθμική συνάρτηση πιθανοφάνειας:

$$L(x_1, \dots, x_n, \mu, \sigma) = -\frac{n}{2} \ln(2\pi) - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Υπολογίζουμε τις παραγώγους ως προς τις άγνωστες παραμέτρους μ και σ και τις εξισώνουμε με 0, οπότε παίρνουμε

$$\frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \quad \frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 = 0$$

Λύνοντας το σύστημα παίρνουμε τις τελικές εκτιμήσεις των παραμέτρων:

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \quad \sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = s_X$$

Συγκρίνοντας με τα αποτελέσματα του προηγούμενου ένθετου εδαφίου, παρατηρούμε κατ' αρχήν ότι η εφαρμογή της μεθόδου της μέγιστης πιθανοφάνειας είναι πιο πολύπλοκη σε σχέση με τη μέθοδο των ροπών. Επίσης, παρατηρούμε ότι τα αποτελέσματα των δύο μεθόδων είναι τα ίδια. Αυτό όμως δεν είναι κανόνας για όλες τις συναρτήσεις κατανομής. Αντίθετα, στις περισσότερες περιπτώσεις οι δύο μέθοδοι δίνουν διαφορετικά αποτελέσματα.

3.5 Έλεγχος υποθέσεων

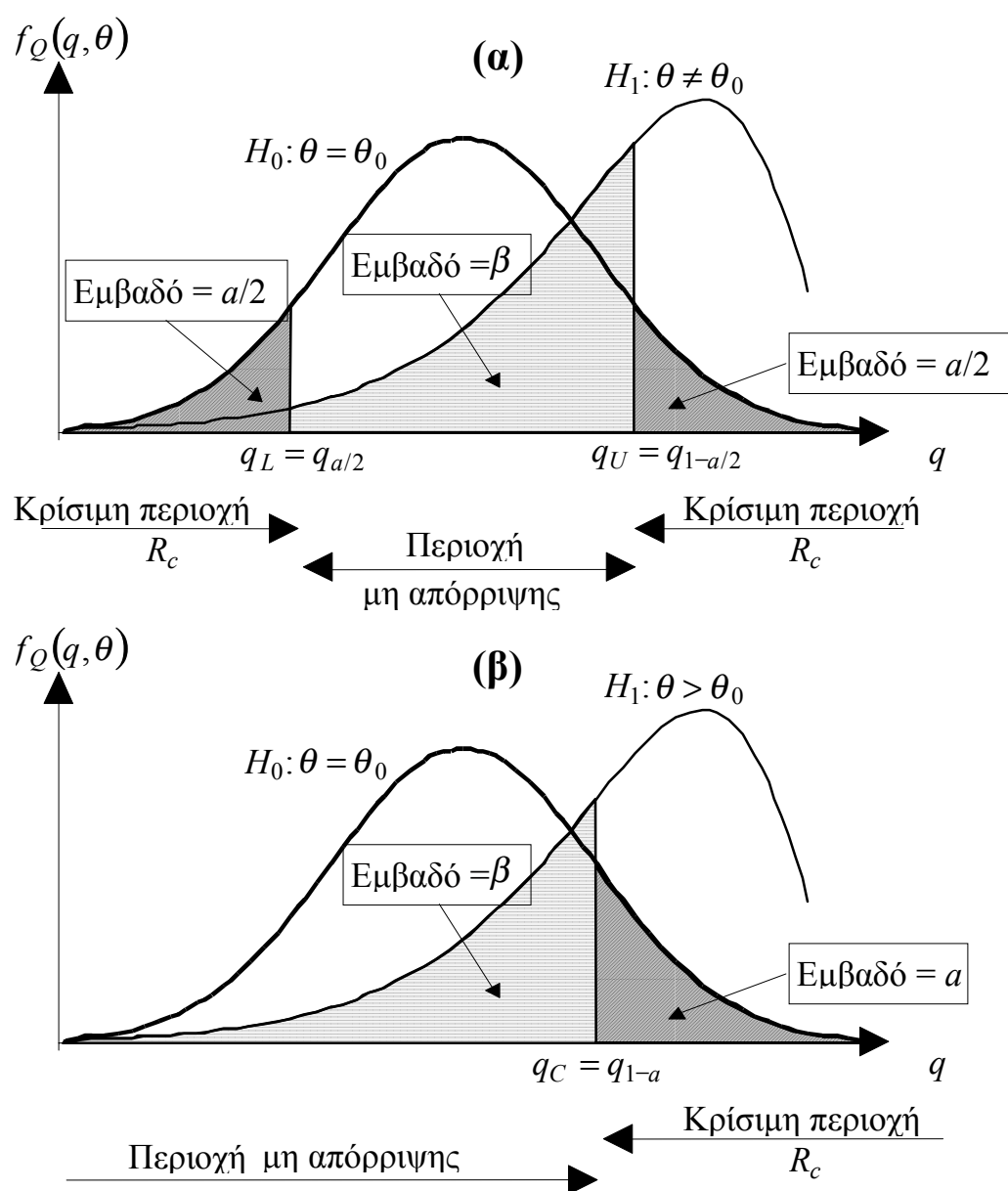
Μια στατιστική υπόθεση είναι μια υπόθεση σχετικά με τις τιμές μιας ή περισσότερων παραμέτρων ενός στατιστικού μοντέλου, το οποίο περιγράφεται από μια συνάρτηση κατανομής. Ο έλεγχος μιας υπόθεσης είναι μια δισήμαντη στατιστική διαδικασία που οδηγεί είτε στην απόρριψη είτε στη αποδοχή (ακριβέστερα: μη απόρριψη) της υπόθεσης.

Στην ενότητα αυτή δίνουμε πολύ περιληπτικά τη σχετική ορολογία και διαδικασία, ενώ σε επόμενα κεφάλαια θα δούμε ορισμένες εφαρμο-

γές. Πάντως, στην τεχνική υδρολογία υπάρχει πληθώρα περιπτώσεων όπου αξιοποιούμε τη θεωρία του ελέγχου υποθέσεων, οι οποίες δεν καλύπτονται στο εισαγωγικό αυτό κείμενο. Ο αναγνώστης που ενδιαφέρεται για λεπτομερέστερη παρουσίαση της θεωρίας παραπέμπεται στα βιβλία στατιστικής (όπως Papoulis, 1990, σ. 321-387, Freund et al., 1988, σ. 310-542), ενώ για μια περιεκτική επισκόπηση των υδρολογικών εφαρμογών παραπέμπεται στους Hirsch et al. (1993, σ. 17.11-29).

3.5.1 Ορολογία

- *Μηδενική υπόθεση* είναι η υπόθεση που ελέγχεται (συμβολικά H_0). Συνήθως πρόκειται για μια υπόθεση της μορφής $\theta = \theta_0$, όπου θ μια παράμετρος που συνδέεται με την κατανομή δεδομένης μεταβλητής και θ_0 μια συγκεκριμένη αριθμητική τιμή.
- *Εναλλακτική υπόθεση* είναι μια δεύτερη υπόθεση η οποία δεν πρέπει να συναληθεύει με τη μηδενική (συμβολικά H_1). Μπορεί να είναι είτε απλή, όπως $\theta = \theta_1$, είτε (συνηθέστερα) σύνθετη, όπως $\theta \neq \theta_0$, $\theta > \theta_0$ ή $\theta < \theta_0$.
- *Στατιστική συνάρτηση ελέγχου* είναι μια κατάλληλα επιλεγμένη στατιστική συνάρτηση του δείγματος, η οποία χρησιμοποιείται για τον έλεγχο (συμβολικά Q).
- *Κρίσιμη περιοχή* είναι ένα διάστημα πραγματικών αριθμών που, όταν βρίσκεται σε αυτό η τιμή της στατιστικής συνάρτησης ελέγχου, τότε απορρίπτουμε τη μηδενική υπόθεση (συμβολικά R_c , βλ. Σχ. 3.2).
- *Μονόπλευρος* λέγεται ένας έλεγχος όταν η εναλλακτική υπόθεση είναι της μορφής $\theta > \theta_0$ ή $\theta < \theta_0$. Στην περίπτωση αυτή η κρίσιμη περιοχή είναι μια ημιευθεία της μορφής ($q > q_c$) ή ($q < q_c$), αντίστοιχα.
- *Αμφίπλευρος* λέγεται ένας έλεγχος όταν η εναλλακτική υπόθεση είναι της μορφής $\theta \neq \theta_0$. Στην περίπτωση αυτή η κρίσιμη περιοχή αποτελείται από τις δύο ημιευθείες ($q < q_L$) και ($q > q_U$).
- *Παραμετρικός* λέγεται ένας έλεγχος όταν μπορεί να εφαρμοστεί μόνο για συγκεκριμένο τύπο συνάρτησης κατανομής του πληθυσμού.



Σχ. 3.2 Επεξηγηματικό σκαρίφημα για τις έννοιες που υπεισέρχονται στο στατιστικό έλεγχο: (α) αμφίπλευρος έλεγχος, (β) μονόπλευρος έλεγχος.

- *Μη παραμετρικός* λέγεται ένας έλεγχος όταν εφαρμόζεται για οποιοδήποτε τύπο συνάρτησης κατανομής του πληθυσμού.
- *Κανόνας απόφασης* είναι ο κανόνας βάσει του οποίου απορρίπτουμε ή όχι τη μηδενική υπόθεση. Κατά κανόνα εκφράζεται ως εξής:

απορρίπτουμε την H_0 αν $q \in R_c$

- Σφάλμα τύπου I είναι η απόρριψη (με βάση τον κανόνα απόφασης) μιας αληθούς μηδενικής υπόθεσης.
- Σφάλμα τύπου II είναι η μη απόρριψη (με βάση τον κανόνα απόφασης) μιας ψευδούς μηδενικής υπόθεσης.
- Επίπεδο σημαντικότητας του ελέγχου είναι η πιθανότητα σφάλματος τύπου I, δηλαδή η πιθανότητα απόρριψης αληθούς μηδενικής υπόθεσης. Συμβολικά

$$\alpha = P(Q \in R_c | H_0) \quad (3.68)$$

- Ισχύς του ελέγχου είναι η πιθανότητα απόρριψης ψευδούς μηδενικής υπόθεσης. Συμβολικά

$$P = 1 - \beta = P(Q \in R_c | H_1) \quad (3.69)$$

όπου β είναι η πιθανότητα σφάλματος τύπου II, δηλαδή*

$$\beta = P(Q \notin R_c | H_1) \quad (3.70)$$

3.5.2 Διαδικασία ελέγχου

Η διαδικασία ελέγχου περιλαμβάνει τα ακόλουθα βήματα:

1. Σχηματοποίηση της μηδενικής υπόθεσης H_0 και της εναλλακτικής H_1 .
2. Επιλογή της στατιστικής συνάρτησης ελέγχου $Q = g(X_1, \dots, X_n)$ και καθορισμός της συνάρτησης πυκνότητας πιθανότητάς της $f_Q(q, \theta)$.

* Διευκρινίζεται ότι τα μεγέθη P και β είναι συναρτήσεις της άγνωστης υπό έλεγχο παραμέτρου θ και όχι αριθμητικές σταθερές. Ο προσδιορισμός τους γίνεται με βάση τις εξισώσεις

$$\beta(\theta) = \int_{R-R_c} f_Q(q, \theta) dq \quad P(\theta) = 1 - \beta(\theta)$$

όπου $f_Q(q, \theta)$ η συνάρτηση πυκνότητας πιθανότητας της Q και $R-R_c$ το διάστημα μη απόρριψης της H_0 .

3. Επιλογή του επιπέδου σημαντικότητας α του ελέγχου και καθορισμός της κρίσιμης περιοχής R_c .
4. Υπολογισμός της τιμής $q = g(x_1, \dots, x_n)$ της Q από το δείγμα.
5. Εφαρμογή του κανόνα απόφασης και απόρριψη ή αποδοχή της H_0 .
6. Υπολογισμός της ισχύος P του ελέγχου.

Το τελευταίο βήμα, λόγω της πολυπλοκότητας του, στις πρακτικές εφαρμογές συνήθως παραλείπεται. Τα πιο πάνω βήματα καθώς και η ορολογία διασαφηνίζονται στο εδάφιο που ακολουθεί.

3.5.3 Έλεγχος σημαντικότητας του συντελεστή συσχέτισης

Ως παράδειγμα εφαρμογής της πιο πάνω διαδικασίας θα δώσουμε τον έλεγχο της σημαντικότητας του συντελεστή συσχέτισης δύο τυχαίων μεταβλητών X και Y , βάσει του οποίου μπορούμε να αποφαινόμαστε αν οι μεταβλητές είναι γραμμικά συσχετισμένες.

Αν οι μεταβλητές είναι γραμμικά συσχετισμένες τότε ο συντελεστής συσχέτισής τους $\rho = \rho_{XY}$ θα είναι διαφορετικός από 0, αλλιώς θα είναι 0. Με βάση αυτή την παρατήρηση, προχωρούμε στα ακόλουθα βήματα της διαδικασίας στατιστικού ελέγχου.

1. Η μηδενική υπόθεση θα είναι $\rho = 0$ και η εναλλακτική υπόθεση $\rho \neq 0$. Κατά συνέπεια θα προβούμε σε αμφίπλευρο έλεγχο. (Αν θέλαμε να αποφασίσουμε σχετικά με το αν οι δύο μεταβλητές είναι θετικά γραμμικά συσχετισμένες θα διατυπώναμε την εναλλακτική υπόθεση ως $\rho > 0$, δηλαδή θα προβαίναμε σε μονόπλευρο έλεγχο).
2. Επιλέγουμε ως στατιστική συνάρτηση ελέγχου την

$$Q = \frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) \sqrt{n-3} = Z \sqrt{n-3} \quad (3.71)$$

όπου R ο δειγματικός συντελεστής συσχέτισης και Z η βοηθητική μεταβλητή του Fisher (βλ.εδάφιο 3.3.5), η οποία, με την υπόθεση H_0 , έχει κατά προσέγγιση κανονική κατανομή με μέση τιμή 0 και τυπική

απόκλιση $1/\sqrt{n-3}$. Κατά συνέπεια η Q ακολουθεί την τυποποιημένη κανονική κατανομή $N(0, 1)$.[†]

3. Επιλέγουμε μια τιμή του επιπέδου σημαντικότητας, έστω $\alpha = 0.05$. Αν $z_{1-\alpha/2}$ είναι το $(1-\alpha/2)$ -ποσοστημόριο της κανονικής κατανομής, τότε η αντίστοιχη κρίσιμη περιοχή R_c είναι η $|q| > z_{1-\alpha/2}$ ή $|q| > z_{0.975}$, ή τελικά $|q| > 1.96$, δεδομένου ότι

$$\begin{aligned} P(|Q| > z_{1-\alpha/2}) &= P(Q < -z_{1-\alpha/2}) + P(Q > z_{1-\alpha/2}) \\ &= 2P(Q < z_{\alpha/2}) = 2\alpha/2 = \alpha \end{aligned}$$

(Υπενθυμίζουμε ότι λόγω συμμετρίας της κανονικής πυκνότητας πιθανότητας ισχύει $z_{1-u} = z_u$. Ακόμη, σημειώνουμε ότι στην περίπτωση του μονόπλευρου ελέγχου η κρίσιμη περιοχή θα ήταν $q > z_{1-\alpha}$).

4. Η αριθμητική τιμή q προκύπτει από το δείγμα ως εξής:

$$q = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) \sqrt{n-3} \quad r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.72)$$

5. Ο κανόνας απόφασης θα είναι ο εξής:

$$\text{απορρίπτουμε την } H_0 \text{ αν } |q| > z_{1-\alpha/2}$$

και για $\alpha = 0.05$

[†] Συχνά στα βιβλία στατιστικής χρησιμοποιείται για τον ίδιο σκοπό η μεταβλητή

$$T = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$$

η οποία έχει κατανομή $t(n-1)$. Πρακτικώς τα αποτελέσματα των δύο μεθόδων ταυτίζονται (δεν διαφέρουν περισσότερο από 1% για συνήθη επίπεδα σημαντικότητας). Εδώ έχουμε προτιμήσει τον έλεγχο βάσει της μεταβλητής του Fisher λόγω της γενικότερης εφαρμογής της για κάθε ρ , σε αντίθεση με την T που εφαρμόζεται μόνο για $\rho = 0$.

απορρίπτουμε την H_0 αν $|q| = \frac{1}{2} \left| \ln \left(\frac{1+r}{1-r} \right) \right| \sqrt{n-3} > 1.96$

Επιλύοντας την παραπάνω εξίσωση ως προς r βρίσκουμε ότι το κρίσιμο όριο r_c του δειγματικού συντελεστή συσχέτισης, το οποίο ορίζει την κρίσιμη περιοχή R_c για τη στατιστική συνάρτηση R , είναι

$$r_c = \frac{e^{3.92/\sqrt{n-3}} - 1}{e^{3.92/\sqrt{n-3}} + 1} \quad (3.73)$$

Ένας απλός και εύκολος στην απομνημόνευση τύπος που προσεγγίζει ικανοποιητικά την παραπάνω εξίσωση είναι ο ακόλουθος:

$$r_c = \frac{2}{\sqrt{n}} \quad (3.74)$$

Κατά συνέπεια, μπορούμε να κάνουμε τον έλεγχο της υπόθεσης με πιο άμεσο τρόπο, συγκρίνοντας την απόλυτη τιμή του r με την κρίσιμη τιμή r_c . Αν $|r| > r_c$ τότε θεωρούμε ότι υπάρχει στατιστικά σημαντική συσχέτιση ανάμεσα στις μεταβλητές

Εφαρμογή 3.5.3

Από ταυτόχρονα δείγματα ετήσιας βροχόπτωσης και απορροής μιας λεκάνης, μήκους 18 ετών, υπολογίστηκε συντελεστής συσχέτισης 0.58. Υπάρχει γραμμική συσχέτιση ανάμεσα στην ετήσια βροχόπτωση και στην απορροή της λεκάνης;

Υπολογίζουμε την κρίσιμη τιμή r_c χρησιμοποιώντας μία από τις εξισώσεις (3.73) ή (3.74). Εδώ για λόγους σύγκρισης χρησιμοποιούμε και τις δύο.

$$r_c = \frac{e^{3.92/\sqrt{15}} - 1}{e^{3.92/\sqrt{15}} + 1} = 0.470 \quad r_c = \frac{2}{\sqrt{18}} = 0.471$$

Παρατηρούμε ότι και οι δύο εξισώσεις δίνουν πρακτικώς το ίδιο αποτέλεσμα. Αφού $0.58 > 0.47$ συμπεραίνουμε ότι υπάρχει στατιστικά σημαντική γραμμική συσχέτιση ανάμεσα στην ετήσια βροχόπτωση και απορροή της λεκάνης.

Κεφάλαιο 4 Ειδικές έννοιες θεωρίας πιθανοτήτων στην υδρολογία

4.1 Πιθανοθεωρητική περιγραφή υδρολογικών διεργασιών

Από την οπτική γωνία της θεωρίας πιθανοτήτων οι υδρολογικές διεργασίες είναι στοχαστικές ανελίξεις. Ας πάρουμε για παράδειγμα την παροχή σε μια συγκεκριμένη θέση ενός ποταμού. Σε κάθε χρονική στιγμή t μέσα στο συνεχή χρόνο η παροχή $X(t)$ είναι μια τυχαία μεταβλητή, με την έννοια ότι δεν υπάρχει προσδιοριστική μέθοδος καθορισμού της τιμής της παροχής με πλήρη βεβαιότητα. Έτσι λοιπόν η $X(t)$, σύμφωνα με τον ορισμό που δόθηκε στην ενότητα 2.7, είναι στοχαστική ανέλιξη σε συνεχή χρόνο, ενώ μια σειρά μετρήσεων της σε τακτά χρονικά διαστήματα αποτελεί μια χρονοσειρά.

Στο σημείο αυτό θα πρέπει να δώσουμε ορισμένες διευκρινίσεις προκειμένου να αποφευχθούν παρανοήσεις στις οποίες είναι εύκολο να οδηγηθεί κανείς από την εισαγωγή της έννοιας της στοχαστικής ανελίξης για την περιγραφή μιας φυσικής διεργασίας.

Κατ' αρχήν, το γεγονός ότι μια φυσική διεργασία περιγράφεται από μια στοχαστική ανέλιξη δεν σημαίνει ότι η πρώτη δεν υπακούει σε κανέναν είδος αιτιοκρατία. Αντίθετα, είναι γνωστό ότι τα υδρολογικά μεγέθη εμφανίζουν περιοδικές διακυμάνσεις μέσα στη διάρκεια ενός έτους, οι οποίες προφανώς οφείλονται στην ετήσια κίνηση της γης και στα μετεω-

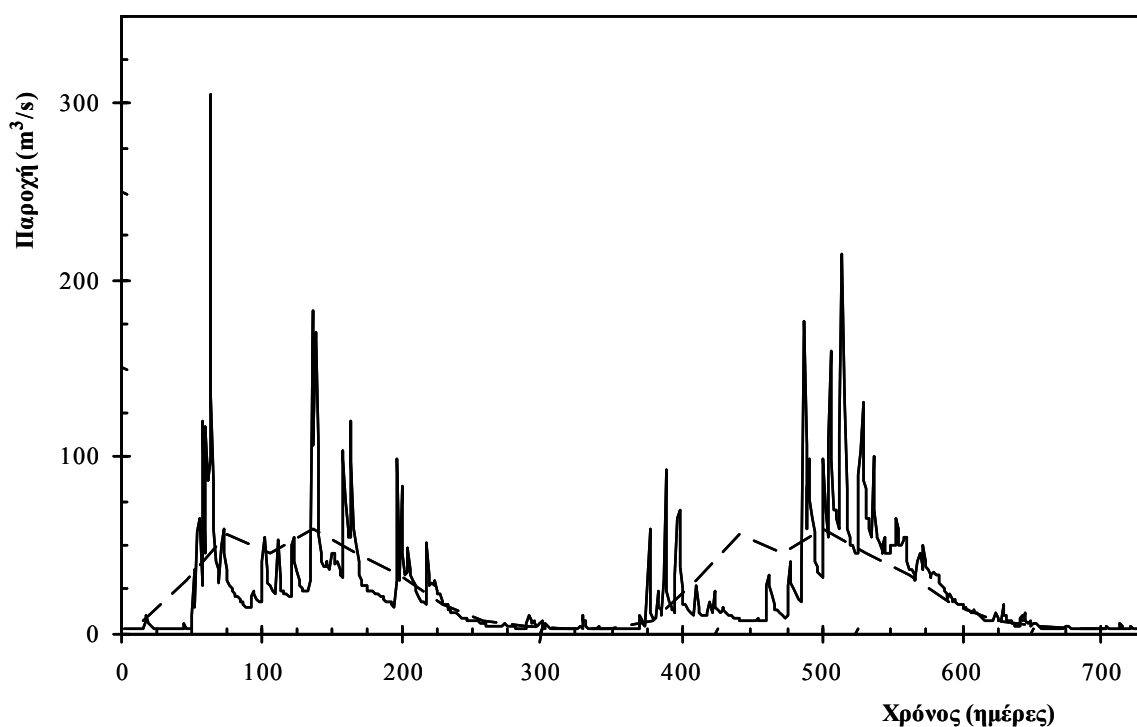
ρολογικά φαινόμενα που αυτή επισύρει. Αυτές οι περιοδικές διακυμάνσεις αποτελούν την ονομαζόμενη *προσδιοριστική συνιστώσα* των διεργασιών. Η στοχαστική ανέλιξη $X(t)$ δεν έχει δυσκολία να συμπεριλάβει και να περιγράψει μαθηματικά και αυτή τη συνιστώσα. Συχνά μάλιστα γίνεται διάκριση των δύο συνιστωσών θεωρώντας ότι

$$X(t) = D(t) + \Xi(t) \quad (4.1)$$

όπου $D(t)$ είναι η προσδιοριστική (εν προκειμένω περιοδική) συνιστώσα και $\Xi(t)$ είναι η στοχαστική συνιστώσα. Στο Σχ. 4.1 παρουσιάζεται η μεταβολή της παροχής ενός ποταμού στη διάρκεια δύο ετών, όπου είναι εμφανής η ετήσια περιοδικότητα.

Σε πολλά κείμενα υδρολογίας (π.χ. Haan, 1977· Kottegoda, 1980) στην προσδιοριστική συνιστώσα συγκαταλέγονται ακόμη οι *τάσεις*, δηλαδή οι μακροχρόνιες, πολυετείς, βαθμιαίες μεταβολές των μέσων χαρακτηριστικών της $X(t)$, καθώς και τα *άλματα*, δηλαδή οι απότομες αλλαγές στα μέσα χαρακτηριστικά της $X(t)$, οι οποίες εν συνεχεία διατηρούνται επίσης για μεγάλα χρονικά διαστήματα. Ωστόσο, εκτός από σπάνιες περιπτώσεις, οι μηχανισμοί που προκαλούν αυτές τις τάσεις και τα άλματα δεν είναι γνωστοί. Κατά συνέπεια, δεν είναι προβλέψιμος ο τρόπος με τον οποίο θα εξελιχθούν αυτά τα φαινόμενα, αντίθετα με τα φαινόμενα που οφείλονται στην περιοδικότητα της κίνησης της γης, τα οποία είναι βέβαια προβλέψιμα. Με αυτή την έννοια, στο κείμενο αυτό οι τάσεις και τα άλματα θεωρούνται κατά κανόνα ως τυχαίες διακυμάνσεις, παρά ως προσδιοριστικές, οι οποίες όμως συντελούνται σε πολύ μεγαλύτερη χρονική κλίμακα (π.χ. δεκαετιών, αιώνων κοκ.) από την κλίμακα των συνηθών τυχαίων διακυμάνσεων

Ειδικότερα, το στοχαστικό μέρος της ανέλιξης δεν είναι πλήρως τυχαίο, αλλά, όπως λέμε, έχει *στοχαστική δομή* ή *μνήμη*. Αυτό σημαίνει ότι υπάρχει στοχαστική εξάρτηση των τιμών της ανέλιξης σε γειτονικές χρονικές στιγμές, ή, με την ορολογία του εδαφίου 2.7.2, ισχυρός συντελεστής αυτοσυσχέτισης της ανέλιξης. Για παράδειγμα, αν η τιμή της παροχής ενός ποταμού στο χρόνο t_0 είναι $X(t_0) = 500 \text{ m}^3/\text{s}$, είναι απίθανο, μετά από ένα μικρό χρονικό διάστημα Δt π.χ. 1 ώρα, να γίνει $X(t_0 + \Delta t) = 0.5 \text{ m}^3/\text{s}$. Το πιθανότερο είναι ότι θα κυμανθεί γύρω από μια τιμή κοντά στα $500 \text{ m}^3/\text{s}$.



Σχ. 4.1 Διακύμανση της ημερήσιας παροχής του ποταμού Ευήνου στη θέση Πόρος Ρηγανίου (υδρολογικά έτη 1971-72 και 1972-73 – έναρξη χρόνου 1.10.1971). Με διακεκομμένη γραμμή φαίνονται οι υπερετιήσιες μέσες τιμές για κάθε μήνα υπολογισμένες από το ιστορικό δείγμα 1970-71 μέχρι 1989-90.

4.1.1 Ανελίξεις σε διακριτό χρόνο

Η μελέτη της πλήρους ανέλιξης σε συνεχή χρόνο μιας υδρολογικής μεταβλητής $X(t)$ είναι αρκετά δύσκολο πρόβλημα, αλλά και δεν είναι απαραίτητη για τα περισσότερα από τα πρακτικά προβλήματα του μηχανικού. Για το λόγο αυτό ξεφεύγει από τους στόχους αυτού του κειμένου. Θα εντοπίσουμε την προσοχή μας σε ορισμένες παράγωγες ανελίξεις, οι οποίες αναφέρονται σε διακριτό χρόνο και αναλύονται ευκολότερα. Πριν προχωρήσουμε στον ορισμό αυτών των ανελίξεων, ας ορίσουμε δύο χρονικά μεγέθη. Το πρώτο είναι το *υδρολογικό έτος*, μέσα στο οποίο πραγματοποιείται ένας πλήρης κύκλος περιόδων υδρολογικών διακυμάνσεων. Η διάρκειά του D ταυτίζεται με τη διάρκεια του ημερολογιακού έτους, συνήθως όμως ξεκινάει όχι στη 1 Ιανουαρίου, αλλά στην αρχή της βροχερής περιόδου του έτους. Έτσι, στην Ελλάδα, κατά σύμβαση, το υδρολογικό έτος ξεκινάει στη 1 Οκτωβρίου. Το δεύτερο είναι το *χρονικό βήμα* Δ , δηλαδή το χρονικό παράθυρο μέσα από το οποίο βλέπουμε την ανέλιξη. Σε αντίθεση με το υδρολογικό έτος, το χρονικό βήμα δεν είναι στα-

θερό, αλλά εξαρτάται από τη συγκεκριμένο πρόβλημα που αντιμετωπίζουμε. Συνήθως κυμαίνεται από μερικά λεπτά της ώρας μέχρι ένα υδρολογικό έτος. Βοηθητικό για τους ορισμούς που θα ακολουθήσουν είναι το Σχ. 4.2 (σ. 92), στο οποίο ο συνεχής χρόνος t (στον οριζόντιο άξονα) μετριέται σε υδρολογικά έτη, ενώ για σχεδιαστικούς λόγους έχει θεωρηθεί $\Delta = D/4$.

Η πρώτη απλοποίηση της πλήρους ανέλιξης $X(t)$ (Σχ. 4.2(1)) αφορά στη διακριτοποίηση του χρόνου (Σχ. 4.2(2)). Το διακριτό χρόνο k τον ορίζουμε χωρίζοντας το συνεχή χρόνο σε διαστήματα μεγέθους Δ . Έτσι οι τιμές $k = 1, 2, \dots$, αντιστοιχούν στα χρονικά διαστήματα $[0, \Delta)$, $[\Delta, 2\Delta)$, κ.ο.κ. Ως τιμή της νέας ανέλιξης $X_{\Delta}(k)$ στη χρονική θέση k ορίζουμε το χρονικό μέσο της $X(t)$ στο αντίστοιχο διάστημα, ήτοι

$$X_{\Delta}(k) := \frac{1}{\Delta} \int_{(k-1)\Delta}^{k\Delta} X(t) dt \quad (4.2)$$

Αν για παράδειγμα η $X(t)$ παριστάνει τη στιγμιαία παροχή ποταμού, και το Δ ληφθεί μία ημέρα ή ένας μήνας, τότε η $X_{\Delta}(k)$ παριστάνει την ημερήσια (ακριβέστερα: τη χρονικά μέση ημερήσια) ή τη μηνιαία (ακριβέστερα: τη χρονικά μέση μηνιαία) παροχή, αντίστοιχα. Μερικές φορές ως παράγωγο μέγεθος παίρνουμε όχι το χρονικό μέσο αλλά το χρονικό άθροισμα στο αντίστοιχο διάστημα Δ , δηλαδή το μέγεθος

$$X_{\Delta}^*(k) := \int_{(k-1)\Delta}^{k\Delta} X(t) dt \quad (4.3)$$

Στο παραπάνω παράδειγμα το μέγεθος $X_{\Delta}^*(k)$ παριστάνει τον ημερήσιο ή το μηνιαίο όγκο απορροής. Αντίστοιχα αν η $X(t)$ παριστάνει τη στιγμιαία ένταση βροχής σε δεδομένο σημείο μιας λεκάνης απορροής, και το Δ ληφθεί μία ημέρα ή ένας μήνας, τότε η $X_{\Delta}^*(k)$ παριστάνει το ημερήσιο ή το μηνιαίο ύψος βροχής, αντίστοιχα.

Αν και η διακριτοποίηση του χρόνου είναι ένα βήμα προς την απλοποίηση της μελέτης μιας υδρολογικής διεργασίας, η μαθηματική περιγραφή της $X_{\Delta}(k)$ ή της $X_{\Delta}^*(k)$ είναι ακόμη αρκετά πολύπλοκη, γιατί προϋποθέτει την ανάλυση της περιοδικότητας και της ισχυρής αυτοσυσχέτισης της ανέλιξης. Έτσι, η μελέτη και αυτού του τύπου των

ανελιξέων δεν καλύπτεται από αυτό το εισαγωγικό κείμενο τεχνικής υδρολογίας και ο ενδιαφερόμενος αναγνώστης παραπέμπεται στους Haan (1977), Kottegoda (1980) Bras and Rodriguez-Iturbe (1985) και Salas et al. (1988). Οι επόμενες απλοποιήσεις είναι αρκετά δραστικότερες και οι ανελιξεις που προκύπτουν είναι πολύ ευκολότερο να μελετηθούν αλλά και είναι πολύ χρησιμότερες στα πρακτικά προβλήματα του μηχανικού.

Αν στην ανέλιξη $X_\Delta(k)$ (ή την $X_\Delta^*(k)$) θεωρήσουμε χρονικό βήμα ίσο με ένα υδρολογικό έτος ($\Delta = D$) τότε παίρνουμε την ετήσια ανέλιξη, συμβολικά $X_D^*(\tau)$, όπου χρησιμοποιήσαμε ως σύμβολο του διακριτού χρόνου (που πλέον εκφράζεται σε ακέραιες μονάδες υδρολογικού έτους) το τ αντί του k (Σχ. 4.2(3)). Έτσι έχουμε

$$X_D(\tau) := \frac{1}{D} \int_{(\tau-1)D}^{\tau D} X(t) dt, \quad X_D^*(\tau) := \int_{(\tau-1)D}^{\tau D} X(t) dt \quad (4.4)$$

Σε αυτή την ανέλιξη πλέον έχει εξαλειφθεί τελείως η περιοδικότητα, δεδομένου ότι δεν είναι “ορατά” χρονικά διαστήματα μικρότερα του έτους, και έχει περιοριστεί η αυτοσυσχέτιση σε μεγάλο βαθμό, λόγω του μεγάλου χρονικού βήματος ολοκλήρωσης. Εξ άλλου η ανέλιξη αυτή, η οποία εκφράζει τη διαδοχή στα διάφορα υδρολογικά έτη μιας συνολικής ετήσιας υδρολογικής ποσότητας, είναι επαρκής για προβλήματα εκτίμησης υδατικού δυναμικού.

Ένας τρόπος για να πάμε σε μικρότερο από το ετήσιο χρονικό βήμα, εξαλείφοντας παράλληλα την περιοδικότητα και την αυτοσυσχέτιση, φαίνεται στο Σχ. 4.2(4). Σε κάθε υδρολογικό έτος $\tau = 1, 2, \dots$, παίρνουμε ένα μόνο διάστημα μεγέθους Δ , και συγκεκριμένα το $[(\tau-1)D + (k-1)\Delta, (\tau-1)D + k\Delta)$. Το k εδώ είναι ένας δεδομένος ακέραιος με δυνατές τιμές $k = 1, 2, \dots, D/\Delta$ (στο Σχ. 4.2(4) έχει ληφθεί $k = 1$). Η προκύπτουσα ανέλιξη είναι:

$$Y_\Delta(\tau) := \frac{1}{\Delta} \int_{(\tau-1)D + (k-1)\Delta}^{(\tau-1)D + k\Delta} X(t) dt, \quad Y_\Delta^*(\tau) := \int_{(\tau-1)D + (k-1)\Delta}^{(\tau-1)D + k\Delta} X(t) dt \quad (4.5)$$

Για παράδειγμα αν η $X(t)$ παριστάνει τη στιγμιαία παροχή ποταμού, το Δ ληφθεί ένας μήνας και το $k = 1$, τότε η $Y_\Delta(\tau)$ είναι η σειρά των (μέσων)

μηνιαίων παροχών του Οκτωβρίου κάθε έτους και η $Y_{\Delta}^*(\tau)$ είναι η σειρά των αντίστοιχων όγκων απορροής.

4.1.2 Ανελίξεις ακροτάτων

Σε πολλά προβλήματα, αντί των χρονικών μέσων ή των αθροιστικών μεγεθών δεδομένης διάρκειας, ενδιαφέρουν τα ακρότατα μεγέθη, δηλαδή τα μέγιστα (για τη μελέτη πλημμυρών) ή τα ελάχιστα (για τη μελέτη ξηρασιών). Για τη μελέτη αυτών των μεγεθών σχηματίζουμε αντίστοιχες ανελίξεις. Έτσι, στο Σχ. 4.2(5) έχει σχηματιστεί η ανέλιξη των *στιγμιαίων ετήσιων μεγίστων* $Z_0(\tau)$. Σε κάθε υδρολογικό έτος πήραμε μόνο μια τιμή, τη στιγμιαία μέγιστη τιμή που εμφανίζεται κατά τη διάρκεια όλου του υδρολογικού έτους, δηλαδή

$$Z_0(\tau) := \max_{\tau-1 \leq t < \tau} \{X(t)\} \quad (4.6)$$

Με αντίστοιχο τρόπο σχηματίζεται και η ανέλιξη των *στιγμιαίων ετήσιων ελαχίστων*. Αυτές οι ανελίξεις δεν εμφανίζουν περιοδικότητα (μια τιμή μόνο ανά έτος) ούτε αυτοσυσχέτιση (απομακρυσμένες χρονικά τιμές που προκύπτουν από τελείως διαφορετικά υδρομετεωρολογικά φαινόμενα) και γι' αυτό είναι εύκολο να μελετηθούν.

Αν ενδιαφέρει η μέγιστη τιμή του μεγέθους μέσα σε μια δεδομένη διάρκεια Δ , τότε ορίζουμε και μελετούμε αντίστοιχα τις *ανελίξεις ετήσιων μεγίστων δεδομένης διάρκειας*, ήτοι (Σχ. 4.2(6))

$$Z_{\Delta}(\tau) := \max_{\tau-1 \leq s < \tau-\Delta} \left\{ \frac{1}{\Delta} \int_s^{s+\Delta} X(t) dt \right\}, \quad Z_{\Delta}^*(\tau) := \max_{\tau-1 \leq s < \tau-\Delta} \left\{ \int_s^{s+\Delta} X(t) dt \right\} \quad (4.7)$$

Ο ορισμός των ανελίξεων αυτών βασίστηκε στη ανέλιξη συνεχούς χρόνου $X(t)$. Εναλλακτικά, αλλά με μικρότερη ακρίβεια, μπορούν να οριστούν από τις ανελίξεις διακριτού χρόνου (Σχ. 4.2(7)) με τις σχέσεις

$$Z'_{\Delta}(\tau) := \max_{k_1 \leq k \leq k_2} \{X_{\Delta}(k)\}, \quad Z'^*_{\Delta}(\tau) := \max_{k_1 \leq k \leq k_2} \{X^*_{\Delta}(k)\} \quad (4.8)$$

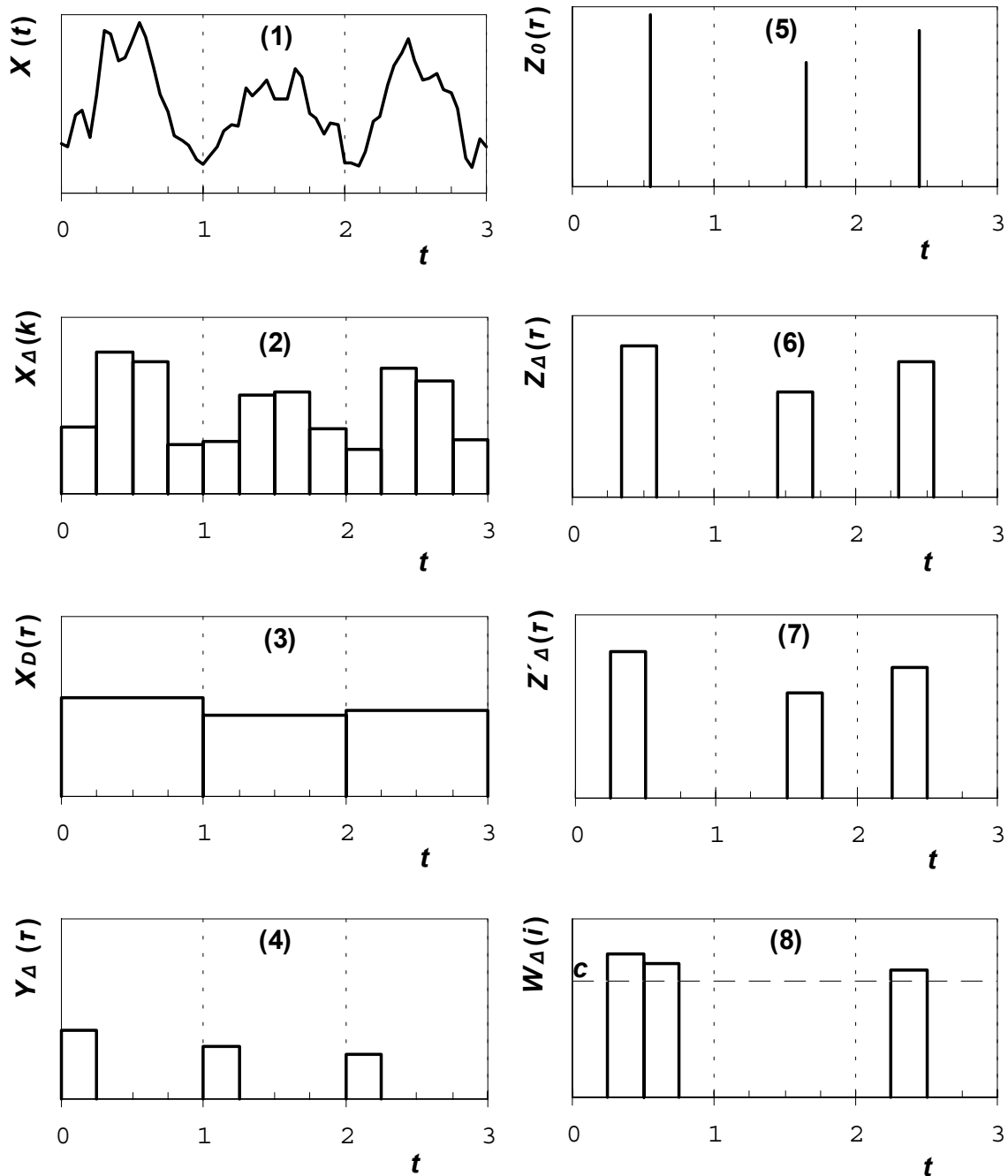
όπου $k_1 := (\tau - 1)D / \Delta + 1$ $\hat{=}$ $k_2 := \tau D / \Delta$. Στα Σχ. 4.2(6) και Σχ. 4.2(7) είναι εμφανές ότι οι μεταβλητές $Z_{\Delta}(\tau)$ και $Z'_{\Delta}(\tau)$ δεν ταυτίζονται ούτε ως προς το μέγεθος ούτε ως προς τη χρονική τοποθέτηση, χωρίς πά-

ντως να διαφέρουν πολύ. Με ανάλογο τρόπο σχηματίζονται και οι ανελίξεις των ετήσιων ελαχίστων δεδομένης διάρκειας. Στα προβλήματα πλημμυρών και ξηρασιών οι τυπικές τιμές του χρονικού βήματος Δ κυμαίνονται από μερικά λεπτά (π.χ. στις καταιγίδες σχεδιασμού δικτύων αποχέτευσης ομβρίων) μέχρι μερικές μέρες ή ακόμη και ένα μήνα (π.χ. στη μελέτη ποιότητας υδάτων ποταμού σε συνθήκες ξηρασίας).

Μια τελευταία σειρά μεγίστων που ονομάζεται *σειρά υπεράνω κατώφλιου* ή *σειρά μερικής διάρκειας* φαίνεται στο Σχ. 4.2(8). Η σειρά αυτή, όπως και αυτή του Σχ. 4.2(7), προκύπτει από την ανέλιξη διακριτού χρόνου $X_{\Delta}(k)$. Αντί όμως να πάρουμε τη μέγιστη τιμή κάθε υδρολογικού έτους, σχηματίζουμε τη σειρά όλων των τιμών που υπερβαίνουν ένα όριο c , ανεξάρτητα από τη θέση των τιμών αυτών στα διάφορα υδρολογικά έτη, δηλαδή

$$\{W_{\Delta}(i), i = 1, 2, \dots\} := \{X_{\Delta}(k) : X_{\Delta}(k) \geq c, k = 1, 2, \dots\} \quad (4.9)$$

(Αντίστοιχα ορίζεται και η ανέλιξη $W_{\Delta}^*(i)$ από τη $X_{\Delta}^*(k)$). Εδώ η μεταβλητή i που επέχει θέση χρόνου στην πραγματικότητα αντιπροσωπεύει απλώς τον αύξοντα αριθμό που έχει η κάθε τιμή στη σειρά των χρονικά διαδοχικών τιμών. Το κατώφλι c συνήθως επιλέγεται έτσι ώστε σε κάθε έτος να αντιστοιχεί κατά μέσο όρο μια τιμή μεγαλύτερη από το κατώφλι. Έτσι, στο Σχ. 4.2(8) το κατώφλι που αντιπροσωπεύεται από την οριζόντια διακεκομμένη γραμμή έχει επιλεγεί έτσι ώστε να υπάρχουν τρεις τιμές σε σύνολο τριών υδρολογικών ετών. Στο συγκεκριμένο παράδειγμα βλέπουμε ότι στο πρώτο υδρολογικό έτος προκύπτουν δύο τιμές πάνω από το κατώφλι, ενώ στο δεύτερο καμιά. Το γεγονός ότι στις σειρές μερικής διάρκειας μπορεί να εμφανίζονται τιμές που αντιστοιχούν σε γειτονικές θέσεις του πραγματικού χρόνου (όπως οι δύο τιμές του πρώτου υδρολογικού έτους του παραδείγματος) ενδέχεται να εισάγει μη αμελητέα στοχαστική εξάρτηση στις διαδοχικές τιμές της ανέλιξης. Αν είναι επιθυμητή η κατασκευή σειράς ανεξάρτητων τιμών θα πρέπει είτε να τεθεί και ένα όριο ελάχιστης χρονικής απόστασης διαδοχικών τιμών, είτε να χρησιμοποιηθούν άλλες εμπειρικές μέθοδοι (βλ. Kottegodā, 1980, σ. 247).



Σχ. 4.2 Βοηθητικό σκαρίφημα για τον ορισμό των διάφορων τύπων ανελίξεων (βλ. περιγραφή στο κείμενο).

4.1.3 Βασικές απλουστευτικές παραδοχές

Πιο πάνω διευκρινίσαμε ότι στο εισαγωγικό αυτό κείμενο δεν καλύπτεται ούτε η μελέτη των ανελίξεων συνεχούς χρόνου, ούτε των πλήρων ανελίξεων διακριτού χρόνου. Ορίσαμε, ωστόσο, άλλους έξι τύπους ανελίξεων με τις οποίες θα ασχοληθούμε και στα επόμενα κεφάλαια. Σε

όλους αυτούς τους τύπους ο “χρόνος” είναι διακριτός και δεν ταυτίζεται κατ' ανάγκη με τον πραγματικό χρόνο. Επίσης, καθεμιά από τις ανελιξεις περιλαμβάνει ακριβώς μία τιμή ανά έτος, εκτός από την ανέλιξη μερικής διάρκειας που περιλαμβάνει κατά μέσο όρο μία τιμή ανά έτος. Για τη μελέτη μας θα κάνουμε από τώρα τις ακόλουθες παραδοχές:

1. Οι ανελιξεις είναι στάσιμες: η κατανομή κάθε μεταβλητής παραμένει ίδια από έτος σε έτος.
2. Οι ανελιξεις είναι εργοδικές: οι αναμενόμενες τιμές είναι ίσες με τους χρονικούς μέσους.
3. Οι μεταβλητές που αντιστοιχούν σε διαφορετικούς χρόνους είναι στοχαστικά ανεξάρτητες.

Για να διαλευκάνουμε το νόημα αυτών των παραδοχών δίνουμε ένα παράδειγμα. Έστω ότι η $X(t)$ αντιπροσωπεύει τη στιγμιαία παροχή στο χρόνο t , οπότε σύμφωνα με τα παραπάνω η $X_D(\tau)$ είναι η μέση ετήσια παροχή του υδρολογικού έτους τ . Έστω ότι διαθέτουμε μετρήσεις 30 ετών, ξέρουμε δηλαδή τις τιμές $x_D(1), \dots, x_D(30)$, των τυχαίων μεταβλητών $X_D(1), \dots, X_D(30)$. Είναι βέβαια προφανές, αλλά ωστόσο τονίζουμε ότι για κάθε μεταβλητή μπορούμε να έχουμε μόνο μια τιμή (δεν μπορούμε να κάνουμε πολλαπλά πειράματα τύχης με διαφορετικές εκβάσεις ώστε να σχηματίσουμε ένα δείγμα πολλαπλών τιμών για την ίδια μεταβλητή, π.χ. την ετήσια παροχή του υδρολογικού έτους με αριθμό 25). Τίθεται τώρα το ερώτημα αν με αυτά τα δεδομένα μπορούμε να εκτιμήσουμε τη μέση τιμή της παροχής του επόμενου υδρολογικού έτους, του υπ' αριθμόν 31. Η απάντηση είναι αρνητική αν δεν ισχύει η πρώτη από τις παραπάνω παραδοχές. Αν η κατανομή της ετήσιας παροχής είναι διαφορετική κάθε έτος, τότε δεν μπορούμε να συνδέσουμε τα προηγούμενα χρόνια με το τρέχον.[†] Αλλά εξακολουθεί να είναι αρνητική και αν ακόμη ισχύει η πρώτη παραδοχή αλλά όχι η δεύτερη. Στην πραγματικότητα, αυτό που μπορούμε να υπολογίσουμε από τα διαθέσιμα δεδομένα είναι ο χρονικός μέσος όρος, δηλαδή η μέση υπερετήσια παροχή των προηγούμενων τριάντα ετών. Η εργοδικότητα μας εξασφαλίζει ότι αυτός ο χρονικός μέσος τείνει προς τη θεωρητική μέση τιμή της κάθε μεταβλητής.

[†] Υπάρχουν πάντως τρόποι να αντιμετωπιστεί η μη στασιμότητα, οι οποίοι ξεφεύγουν από το σκοπό του παρόντος κειμένου.

Έτσι οι δύο πρώτες παραδοχές είναι θεμελιώδεις. Η τρίτη παραδοχή, της ανεξαρτησίας, είναι λιγότερο σημαντική, αλλά βοηθά πολύ σε περιπτώσεις που μας ενδιαφέρει η πιθανότητα διαδοχικών γεγονότων, όπως στο παράδειγμα που ακολουθεί στο τέλος αυτής της ενότητας.

Η πρώτη από τις πιο πάνω παραδοχές αποτελεί απλώς μια προσέγγιση, δεδομένου ότι τα υδρομετεωρολογικά φαινόμενα δεν περιγράφονται με αυστηρώς στάσιμες ανελίξεις. Η ορθότητα της δεύτερης παραδοχής συναρτάται σε μεγάλο βαθμό από την ορθότητα της πρώτης. Για την τρίτη παραδοχή έχει ήδη γίνει συζήτηση: οι ανελίξεις ετήσιων τιμών και μέγιστων τιμών μερικής διάρκειας δεν είναι πάντα απαλλαγμένες από στοχαστική εξάρτηση, η οποία πάντως είναι αρκετά μικρή και μπορεί να αγνοηθεί, ιδίως όταν μας ενδιαφέρει η περιθώρια συνάρτηση κατανομής ενός μεμονωμένου μεγέθους. Είναι απαραίτητο πάντως να τονίσουμε ότι στην υδρολογία δεν υπάρχουν γενικά αποδεκτές *a priori* παραδοχές και γι' αυτό πάντα θα πρέπει να γίνεται έλεγχος των παραδοχών *a posteriori*.

4.1.4 Τελικό συμπέρασμα

Οι παραδοχές της στασιμότητας, της εργοδικότητας και της ανεξαρτησίας που έγιναν για τις ετήσιες ανελίξεις ετησίων τιμών, μηνιαίων τιμών συγκεκριμένου μήνα, ετήσιων μέγιστων ή ελαχίστων (στιγμιαίων ή δεδομένης διάρκειας) και μέγιστων μερικής διάρκειας, οδηγούν σε μια τελική απλοποίηση που διευκολύνει πολύ τη μελέτη: μπορούμε καθένα απ' αυτά τα μεγέθη να το θεωρήσουμε ως μια απλή τυχαία μεταβλητή και τις τιμές του στα διάφορα υδρολογικά έτη να τις θεωρήσουμε ως διαφορετικές εμφανίσεις αυτής της ίδιας τυχαίας μεταβλητής. Έτσι μπορούμε να αντικαταστήσουμε την έννοια της ανέλιξης με την έννοια της απλής τυχαίας μεταβλητής, και, πράγματι, αυτό θα κάνουμε στις επόμενα κεφάλαια. Για παράδειγμα, στο εξής δεν θα μιλούμε για την ανέλιξη των ετήσιων παροχών αλλά για την τυχαία μεταβλητή της ετήσιας παροχής. Αντίστοιχα, αντί να μιλούμε για τη χρονοσειρά των ετήσιων παροχών θα μιλούμε για το δείγμα της ετήσιας παροχής, αγνοώντας τη χρονική διάταξη των ιστορικών δεδομένων των διαδοχικών υδρολογικών ετών.

Εφαρμογή 4.1

Με βάση τα ιστορικά δεδομένα παροχών ενός ποταμού που διατίθενται για μια μακρά περίοδο, εκτιμήσαμε ότι το ενδεχόμενο να εμφανιστεί ετήσιος όγκος απορροής μικρότερος από 500 hm^3 † έχει πολύ μικρή πιθανότητα, ίση με 10^{-2} . (α) Να υπολογιστεί η πιθανότητα να εμφανιστεί επί πέντε διαδοχικά χρόνια όγκος απορροής μικρότερος από 500 hm^3 (βλ. και Εφαρμογή 2.3).

Με την προϋπόθεση ότι ισχύουν οι παραδοχές της στασιμότητας, της εργοδικότητας και της ανεξαρτησίας η ζητούμενη πιθανότητα είναι απλώς $(10^{-2})^5 = 10^{-10}$. Πρόκειται βέβαια για εξαιρετικά μικρή πιθανότητα. Ωστόσο έχει παρατηρηθεί ότι στη φύση τέτοια φαινόμενα διαδοχής εξαιρετικά ξηρών ετών δεν είναι τόσο απίθανα.

Αν δεν ισχύει η παραδοχή της ανεξαρτησίας, τότε κατά κανόνα η συσχέτιση των τιμών των διαφορετικών ετών (που περιγράφεται από το συντελεστή αυτοσυσχέτισης της ανέλιξης για χρονικό βήμα ένα) είναι θετική. Σε αυτή την περίπτωση η πιθανότητα του υπόψη ενδεχομένου θα είναι μεγαλύτερη από 10^{-10} , αλλά δεν μπορούμε να υπολογίσουμε την τιμή της χωρίς να χρησιμοποιήσουμε προχωρημένες τεχνικές στοχαστικών ανελίξεων.

(β) Αν τα προηγούμενα τέσσερα χρόνια έχει εμφανιστεί όγκος απορροής μικρότερος από 500 hm^3 , ποια είναι η πιθανότητα να εμφανιστεί και εφέτος το ίδιο ενδεχόμενο;

Προφανώς, αν ισχύει η παραδοχή της ανεξαρτησίας, η πιθανότητα που ζητείται είναι 10^{-2} . Το τι συνέβη τα προηγούμενα χρόνια δεν επηρεάζει το παρόν αν υπάρχει ανεξαρτησία των γεγονότων. Αν δεχτούμε ότι υπάρχει θετική συσχέτιση μεταξύ των διαδοχικών ετών, τότε η πιθανότητα θα είναι μεγαλύτερη, αφού είναι πιθανότερο ένα ξηρό υδρολογικό έτος να ακολουθείται από επίσης ξηρό.

Ενδεχομένως, η διαισθητική απάντηση σε αυτό το ερώτημα είναι αντίθετη: η διαίσθηση, παρασυρμένη από την εξαιρετικά μικρότερη πιθανότητα της διαδοχής πέντε διαδοχικών ξηρών ετών, που όπως υπολογίσαμε είναι 10^{-10} , έχει την τάση να αποδώσει πιθανότητα μικρότερη από 10^{-2} , ίσως και 10^{-10} . Πρόκειται όμως για καθαρή πλάνη.‡

† Υπενθυμίζεται ότι η μονάδα hm^3 σημαίνει κυβικά εκατόμετρα ($1 \text{ hm}^3 = (100 \text{ m})^3 = 1\,000\,000 \text{ m}^3$).

‡ Βέβαια δεν είναι η μοναδική πλάνη του κοινού νου γύρω από τις πιθανότητες. Σε τέτοιου είδους φαινομενικά “αθώες” πλάνες έχουν στηριχτεί πολλές απάτες, ιδίως σε τυχερά παιχνίδια, κληρώσεις κ.ο.κ.

4.2 Η έννοια της περιόδου επαναφοράς

Ο όρος *περίοδος επαναφοράς* είναι από αυτούς που χρησιμοποιούνται πολύ συχνά στην τεχνική υδρολογία, αλλά έχει τύχει κριτικής ως προς την ευστοχία και την σαφήνιά του. Πράγματι, ο όρος αυτός είναι υπεύθυνος για πολλές παρανοήσεις που θα εκτεθούν παρακάτω προκειμένου να διασαφηνιστεί όσο γίνεται καλύτερα.

Αρχικά, η περίοδος επαναφοράς, T , μιας δεδομένης τιμής x της τυχαίας μεταβλητής X (η οποία στην πραγματικότητα αντιπροσωπεύει μια στοχαστική ανέλιξη) ορίζεται ως ο μέσος αριθμός χρονικών διαστημάτων (εν προκειμένω υδρολογικών ετών) που μεσολαβεί μεταξύ δύο διαδοχικών εμφανίσεων της τυχαίας μεταβλητής με μέγεθος μεγαλύτερο ή ίσο της δεδομένης τιμής x . Για παράδειγμα, αν η τιμή $500 \text{ m}^3/\text{s}$ της μέγιστης ετήσιας παροχής έχει περίοδο επαναφοράς 50, αυτό σημαίνει ότι κατά μέσο όρο μεσολαβεί ένα διάστημα 50 ετών ανάμεσα σε δύο εμφανίσεις πλημμυρικής παροχής μεγαλύτερης ή ίσης των $500 \text{ m}^3/\text{s}$. Αποδεικνύεται (Kottegoda, 1980, σ. 213) ότι η περίοδος επαναφοράς της τιμής x είναι

$$T = \frac{1}{P(X > x)} = \frac{1}{F_{1_x}(x)} = \frac{1}{1 - F_x(x)} \quad (4.10)$$

δηλαδή η περίοδος επαναφοράς είναι το αντίστροφο της πιθανότητας υπέρβασης. Προϋποθέσεις για να ισχύει η παραπάνω σχέση είναι (α) να είναι συνεχής η τυχαία μεταβλητή και (β) να ισχύει η παραδοχή ανεξαρτησίας της προηγούμενης ενότητας, δηλαδή κάθε εμφάνιση να είναι στοχαστικά ανεξάρτητη από τις προηγούμενες και επόμενες της. Δεδομένου ότι οι προϋποθέσεις αυτές ισχύουν κατά κανόνα για τα μεγέθη που εξετάζουμε σε αυτό το κείμενο, για τις ανάγκες των προβλημάτων των επόμενων κεφαλαίων μπορούμε να θεωρούμε την εξίσωση (4.10) ως ισοδύναμο ορισμό της περιόδου επαναφοράς.

Τα παραπάνω ισχύουν κατά βάση για τις περιπτώσεις που ενδιαφέρουν μεγάλες τιμές υδρολογικών μεγεθών, π.χ. πλημμύρες. Κατ' αναλογία ορίζεται η περίοδος επαναφοράς για μικρά μεγέθη, π.χ. για παροχές ξηρασίας. Στην περίπτωση αυτή, στον αρχικό ορισμό θα πρέπει να αντικαταστήσουμε τη φράση “μεγαλύτερο ή ίσο” με τη φράση “μικρότερο ή ίσο” ή ισοδύναμα να θεωρήσουμε

$$T = \frac{1}{P(X \leq x)} = \frac{1}{F_X(x)} \quad (4.11)$$

Συμπερασματικά λοιπόν, στα πλαίσια αυτού του κειμένου θα θεωρούμε την περίοδο επαναφοράς είτε ως το αντίστροφο της πιθανότητας υπέρβασης, όταν ενδιαφέρει μέγιστο μέγεθος, είτε ως το αντίστροφο της πιθανότητας μη υπέρβασης (συνάρτησης κατανομής) όταν ενδιαφέρει ελάχιστο μέγεθος.

Ας έρθουμε τώρα στην κριτική της έννοιας της περιόδου επαναφοράς. Κατ' αρχήν η διττή σημασία της (διαφορετική έννοια για μέγιστα ή ελάχιστα μεγέθη) τείνει να δημιουργήσει σύγχυση. Επίσης ο όρος *περίοδος* δημιουργεί τον κίνδυνο να εκληφθεί το μέγεθος στο οποίο αναφέρεται ως κυμαινόμενο περιοδικά, πράγμα που είναι τελείως εσφαλμένο αφού αναφέρεται σε τυχαίες μεταβλητές. Για το λόγο αυτό έχει συχνά χρησιμοποιηθεί, χωρίς όμως να επικρατήσει, ο όρος *διάστημα επανόδου*. Ωστόσο, και εδώ υπάρχει σύγχυση, γιατί ο τελευταίος όρος έχει χρησιμοποιηθεί συχνά και με άλλο νόημα, ήτοι του πραγματικού (και όχι του μέσου) χρόνου ανάμεσα σε διαδοχικές υπερβάσεις μιας τιμής, οπότε έχει την έννοια τυχαίας μεταβλητής (π.χ. στους Chow et al., 1988). Ένα τελευταίο πρόβλημα αφορά στη διάσταση του μεγέθους T . Όπως έχει οριστεί παραπάνω, το μέγεθος είναι αδιάστατο και ως αδιάστατο χρησιμοποιείται σε όλες τις μαθηματικές σχέσεις. Ωστόσο, στην πράξη αποδίδονται στο μέγεθος διαστάσεις χρόνου με μονάδα ίση με το υδρολογικό έτος. Έτσι λέμε π.χ. ότι η (μέγιστη ετήσια) πλημμυρική παροχή των 500 m³/s έχει περίοδο επαναφοράς 50 ετών, ή αλλιώς ότι η πλημμύρα 50ετίας είναι 500 m³/s.

Παρόλα τα παραπάνω προβλήματα η περίοδος επαναφοράς έχει χρησιμοποιηθεί ευρέως στην υδρολογία, γιατί γίνεται αμεσότερα αντιληπτή από την έννοια της πιθανότητας και γιατί η αριθμητική της έκφραση είναι ανετότερη από αυτή της πιθανότητας. Για παράδειγμα είναι αμεσότερη η έκφραση “πλημμύρα 50ετίας” από την έκφραση “πλημμύρα πιθανότητας υπέρβασης 0.02”. Ωστόσο, συνδυάζοντας τους δύο εκφράσεις, θα μπορούσαμε να πούμε ακριβέστερα “πλημμύρα πιθανότητας υπέρβασης 1 : 50”.

Εφαρμογή 4.2

α. Σε ποια πιθανότητα υπέρβασης και ποια πιθανότητα μη υπέρβασης αντιστοιχούν οι πλημμύρες 2, 5, 10, 20, 50, 100, 500, 1000, 5 000 και 10 000 ετών;

Με βάση την εξίσωση (4.10) εύκολα βρίσκουμε τις τιμές του Πίν. 4.1.

Πίν. 4.1 Αντιστοιχία περιόδου επαναφοράς, πιθανότητας υπέρβασης και πιθανότητας μη υπέρβασης της Εφαρμογής 4.2.

| Περίοδος επαναφοράς T | Πιθανότητα υπέρβασης F_1 (%) | Πιθανότητα μη υπέρβασης F (%) |
|-------------------------|--------------------------------|---------------------------------|
| 2 | 50 | 50 |
| 5 | 20 | 80 |
| 10 | 10 | 90 |
| 20 | 5 | 95 |
| 50 | 2 | 98 |
| 100 | 1 | 99 |
| 500 | 0.2 | 99.8 |
| 1000 | 0.1 | 99.9 |
| 5 000 | 0.02 | 99.98 |
| 10 000 | 0.001 | 99.99 |

β. Σε ποια πιθανότητα υπέρβασης και ποια πιθανότητα μη υπέρβασης αντιστοιχούν οι ξηρασίες των ίδιων περιόδων επαναφοράς, όπως στο ερώτημα α;

Η απάντηση δίνεται από τον Πίν. 4.1, αν αντιμετωπισθούν αμοιβαία οι στήλες της πιθανότητας υπέρβασης και πιθανότητας μη υπέρβασης.

4.3 Η έννοια της διακινδύνευσης

Κατά το σχεδιασμό ενός έργου είναι συνηθέστατη πρακτική να υπολογίζεται με βάση τα διαθέσιμα δεδομένα η τιμή σχεδιασμού ή φόρτιση σχεδιασμού L , αλλά η ικανότητα του έργου C να υιοθετείται μεγαλύτερη, έτσι ώστε να υπάρχει δεδομένο περιθώριο ασφάλειας

$$SM := C - L \quad (4.12)$$

ή δεδομένος, μεγαλύτερος από 1, συντελεστής ασφάλειας

$$SF := \frac{C}{L} \quad (4.13)$$

Στο χώρο των κατασκευών πολιτικού μηχανικού το μέγεθος L μπορεί, για παράδειγμα, να είναι το δυσμενέστερο φορτίο λειτουργίας ενός υποστυλώματος, οπότε το C είναι το φορτίο που προκαλεί θραύση του υποστυλώματος. Στο χώρο της τεχνικής υδρολογίας, το L μπορεί να είναι, για παράδειγμα, η παροχή σχεδιασμού ενός υπερχειλιστή φράγματος, οπότε το μέγεθος C είναι η παροχетеυτικότητα του υπερχειλιστή, δηλαδή η παροχή που μπορεί να διοδευτεί με ασφάλεια από τον υπερχειλιστή χωρίς κίνδυνο υπερπήδησης του φράγματος.

Η υιοθέτηση συγκεκριμένης αριθμητικής τιμής για το συντελεστή ασφάλειας (π.χ. $SF = 3$) ή το περιθώριο ασφάλειας αποτελεί εμπειρική προσέγγιση στην αντιμετώπιση της αβεβαιότητας που υπάρχει στην εκτίμηση της φόρτισης ενός έργου ή της ικανότητας του. Ωστόσο, αυτή η πρακτική δεν δίνει πολύ σαφές μέτρο της ασφάλειας και μάλιστα δίνει την εσφαλμένη εντύπωση της πλήρους εξάλειψης του κινδύνου αστοχίας του συγκεκριμένου έργου. Αντίθετα, η προσέγγιση της αβεβαιότητας μέσω της θεωρίας πιθανοτήτων αποτελεί τη μόνη ορθολογική απάντηση στην ποσοτικοποίηση του κινδύνου αστοχίας. Σύμφωνα με αυτή την προσέγγιση, τα μεγέθη SM και SF θεωρούνται τυχαίες μεταβλητές και, ως μέτρο του κινδύνου αστοχίας, ορίζεται η *διακινδύνευση* (ή *επικινδυνότητα - risk*):

$$R := P(SF < 1) = P(SM < 0) \quad (4.14)$$

Στην τυπικότερη κατηγορία προβλημάτων της τεχνικής υδρολογίας, τα προβλήματα σχεδιασμού, η ικανότητα του έργου (π.χ. η παροχетеυτικότητα ή η αποθηκευτική ικανότητα) μπορεί να θεωρηθεί ότι έχει δεδομένη τιμή με πλήρη βεβαιότητα ($C = c$), οπότε δεν θεωρείται ως τυχαία μεταβλητή.[†] Σε αυτή την περίπτωση η διακινδύνευση είναι

$$R = P(L > c) = 1 - P(L \leq c) \quad (4.15)$$

[†] Ο αναγνώστης που ενδιαφέρεται για συνθετότερα προβλήματα, όπου και η ικανότητα είναι τυχαία μεταβλητή, παραπέμπεται στους Chow et al. (1988).

και οφείλεται μόνο στην ενυπάρχουσα αβεβαιότητα των υδρολογικών διεργασιών. Για το λόγο αυτό μιλούμε για φυσική ή ενυπάρχουσα διακινδύνευση.

Το μέγεθος L εξαρτάται από τη μεταβλητότητα της φυσικής διεργασίας που ενδιαφέρει (π.χ. της πλημμυρικής παροχής) καθώς και από τη διάρκεια που θα εκτεθεί το έργο στο φυσικό κίνδυνο, δηλαδή τη διάρκεια ζωής του έργου. Εάν X παριστάνει το μέγιστο μέγεθος της φυσικής διεργασίας σε ετήσια βάση (π.χ. τη μέγιστη ετήσια πλημμύρα) και n είναι η διάρκεια ζωής του έργου, τότε το γεγονός $\{L \leq c\}$ ισοδυναμεί με n διαδοχικές εμφανίσεις του γεγονότος $\{X \leq c\}$. Για να μην έχουμε, δηλαδή, υπέρβαση του μεγέθους c σε όλη τη διάρκεια ζωής του έργου θα πρέπει να μην έχουμε υπέρβαση σε όλα τα n χρόνια αυτής της διάρκειας. Θεωρώντας ότι οι πλημμύρες των διαδοχικών ετών είναι στοχαστικά ανεξάρτητες και εφαρμόζοντας τη σχέση (2.12) παίρνουμε

$$R = 1 - [P(X \leq c)]^n = 1 - [F_X(c)]^n \quad (4.16)$$

όπου F_X είναι η συνάρτηση κατανομής του φυσικού μεγέθους που μας ενδιαφέρει, σε ετήσια βάση. Αντικαθιστώντας τη με την αντίστοιχη περίοδο επαναφοράς παίρνουμε την ακόλουθη βασική σχέση, η οποία συνδέει τα τρία βασικά μεγέθη υδρολογικού σχεδιασμού ενός έργου, δηλαδή την περίοδο επαναφοράς, τη διάρκεια ζωής και τη διακινδύνευση:

$$R = 1 - \left(1 - \frac{1}{T}\right)^n \quad (4.17)$$

Γραφική απεικόνιση της (4.17) δίνεται στο Σχ. 4.3 για χαρακτηριστικές περιόδους επαναφοράς.

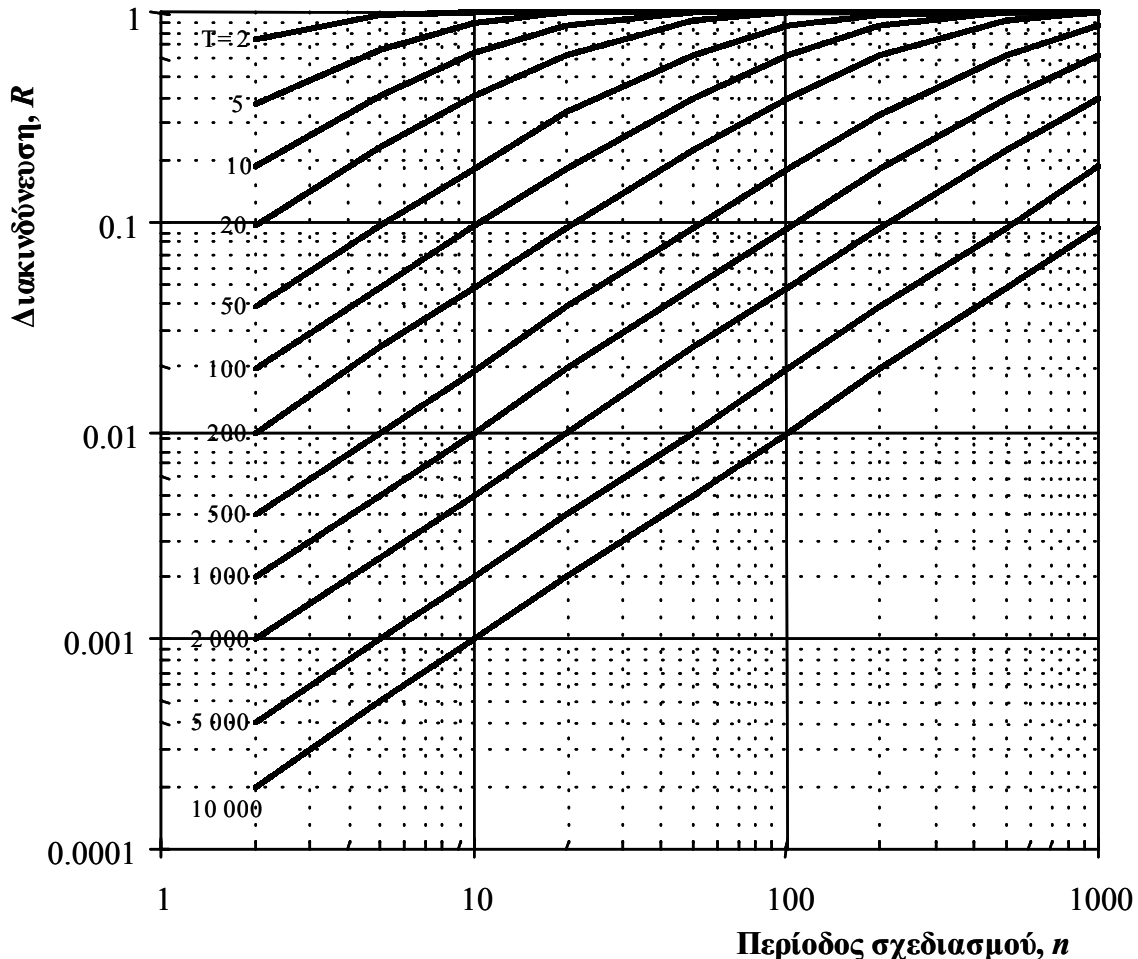
Ισοδύναμη έκφραση της (4.17) είναι η ακόλουθη, η οποία έχει επιλυθεί ως προς την περίοδο επαναφοράς:

$$T = \frac{1}{1 - (1 - R)^{1/n}} \quad (4.18)$$

Τέλος, δεδομένου ότι $\ln(1 - x)^n = n \ln(1 - x) = n(-x - x^2/2 - \dots) \approx -nx$, παίρνουμε και την ακόλουθη προσεγγιστική έκφραση του R :

$$R \approx 1 - e^{-n/T} \quad (4.19)$$

η οποία ισχύει με σφάλμα $< 1\%$ για $T \geq 50$.



Σχ. 4.3 Γραφική απεικόνιση της σχέσης των χαρακτηριστικών μεγεθών υδρολογικού σχεδιασμού (εξίσωση 4.17).

Εφαρμογή 4.3

α. Διώρυγα εκτροπής σχεδιάζεται να λειτουργήσει κατά την περίοδο κατασκευής φράγματος, η οποία εκτιμάται σε 5 χρόνια. Ποια πρέπει να είναι η περίοδος επαναφοράς της πλημμύρας σχεδιασμού, ώστε η διακινδύνευση να μην υπερβαίνει το 10%;

Από την (4.18) παίρνουμε

$$T = \frac{1}{1 - (1 - R)^{1/n}} = \frac{1}{1 - (1 - 0.1)^{1/5}} = 47.9$$

Στρογγυλεύουμε σε $T = 50$ χρόνια.

β. Πόση είναι η διακινδύνευση αν ένα έργο σχεδιαστεί με περίοδο επαναφοράς ίση με τη διάρκεια ζωής του;

Αν δεχτούμε ότι η διάρκεια ζωής του έργου είναι αρκετά μεγάλη (≥ 50 χρόνια), τότε από την (4.19) παίρνουμε $R = 1 - e^{-1} = 0.632 = 63.2\%$. Διαφορετικά η διακινδύνευση υπολογίζεται από την (4.17).

***Κεφάλαιο 5* Τυπική στατιστική ανάλυση μιας υδρολογικής μεταβλητής**

Στο κεφάλαιο αυτό θα εφαρμόσουμε τις αρχές και μεθόδους της στατιστικής, τις οποίες παρουσιάσαμε ήδη στο κεφάλαιο 3, σε ένα από τα πιο τυπικά προβλήματα της στατιστικής υδρολογίας: στην ανάλυση μιας υδρολογικής μεταβλητής. Η υπό μελέτη υδρολογική μεταβλητή μπορεί να είναι οποιαδήποτε: βροχόπτωση ή απορροή, μέση ή μέγιστη, ετήσια ή μηνιαία, κτλ. Η ανάλυση περιλαμβάνει συνήθως τα εξής τρία βήματα:

1. *Συμπύκνωση του δείγματος.* Αυτή γίνεται είτε με αριθμητικό τρόπο, χρησιμοποιώντας ορισμένους αντιπροσωπευτικούς στατιστικούς δείκτες, είτε με γραφικό τρόπο, χρησιμοποιώντας ιστογράμματα και εμπειρικές συναρτήσεις κατανομής.
2. *Προσαρμογή θεωρητικού μοντέλου.* Αυτή περιλαμβάνει την επιλογή κατάλληλης συνάρτησης κατανομής για τη μεταβλητή, εκτίμηση των παραμέτρων της και στατιστικό έλεγχο της προσαρμογής.
3. *Στατιστική πρόγνωση.* Ανάλογα με το σκοπό της μελέτης, η πρόγνωση αποσκοπεί στην εκτίμηση της τιμής της μεταβλητής για δεδομένη περίοδο επαναφοράς. Η εκτίμηση αυτή μπορεί να είναι είτε σημειακή είτε εκτίμηση διαστήματος.

Παρόλο που η τυπική αυτή στατιστική ανάλυση εφαρμόζεται για κάθε τύπο θεωρητικού μοντέλου, στα παραδείγματα αυτού του κεφαλαίου θα κάνουμε χρήση μόνο της κανονικής κατανομής, την οποία έχουμε ήδη μελετήσει στο εδάφιο 2.8.2. Στο αμέσως επόμενο κεφάλαιο θα δώσουμε

το θεωρητικό υλικό καθώς και αντίστοιχα παραδείγματα για όλες τις άλλες συναρτήσεις κατανομής που χρησιμοποιούνται στην τεχνική υδρολογία.

5.1 Στατιστικά χαρακτηριστικά δείγματος

Με τον όρο *στατιστικά χαρακτηριστικά δείγματος* εννοούμε διάφορους στατιστικούς δείκτες που συνοψίζουν σε λίγους αριθμούς τις πιο χαρακτηριστικές ιδιότητες του δείγματος. Διακρίνουμε δύο κατηγορίες στατιστικών χαρακτηριστικών. Η πρώτη περιλαμβάνει τις *δειγματικές ροπές και τα παράγωγά τους χαρακτηριστικά*. Συγκεκριμένα περιλαμβάνει: (α) τη δειγματική μέση τιμή, η οποία, όπως είδαμε, αποτελεί παράμετρο θέσης· (β) τη δειγματική διασπορά και τους παράγωγους δείκτες διασποράς (τυπική απόκλιση και συντελεστή μεταβλητότητας)· (γ) την τρίτη κεντρική ροπή και το συντελεστή ασυμμετρίας. Η δεύτερη κατηγορία περιλαμβάνει απλούστερους στατιστικούς δείκτες που ο υπολογισμός τους προϋποθέτει τη διάταξη του δείγματος σε φθίνουσα ή αύξουσα σειρά και γι' αυτό εδώ αναφέρονται ως *συνοπτικά χαρακτηριστικά διατεταγμένου δείγματος*. Πρόκειται για την ελάχιστη και μέγιστη τιμή του δείγματος, τη διάμεσο (παράμετρος θέσης), το άνω και κάτω τεταρτημόριο και το διατεταρτημοριακό πλάτος (παράμετρος διασποράς).

Οι δειγματικές ροπές και τα παράγωγά τους χαρακτηριστικά υπολογίζονται με εφαρμογή των αντίστοιχων εκτιμητριών που έχουν δοθεί στο κεφάλαιο 3. Οι αντίστοιχοι τύποι δίνονται επίσης συγκεντρωμένοι στον Πίν. 5.1, και μάλιστα σε μορφή απλούστερη για υπολογισμούς. Δίνονται ακόμη οι συντελεστές διόρθωσης μεροληψίας (στήλη 3), με τους οποίους πρέπει να πολλαπλασιαστούν οι απλές εκτιμήσεις (στήλη 2) προκειμένου να υπολογιστούν οι αντίστοιχες αμερόληπτες εκτιμήσεις. Στον Πίν. 5.1 δίνεται επίσης και ο τρόπος εύρεσης των συνοπτικών χαρακτηριστικών του διατεταγμένου δείγματος.

Πίν. 5.1 Τυπικά χαρακτηριστικά δείγματος και τρόπος υπολογισμού τους

| Στατιστικός δείκτης | Απλή εκτίμηση | Συντελεστής διόρθωσης μεροληψίας |
|--|--|----------------------------------|
| <i>1. Δειγματικές ροπές και παράγωγα χαρακτηριστικά</i> | | |
| Μέση τιμή | $\bar{x} = \frac{1}{n} \sum x_i$ | — |
| Διασπορά | $s_X^2 = \frac{1}{n} \sum x_i^2 - \frac{1}{n^2} (\sum x_i)^2$ $= \frac{1}{n} \sum x_i^2 - \bar{x}^2$ | $\frac{n}{n-1}$ |
| Τυπική απόκλιση | s_X | $\approx \sqrt{\frac{n}{n-1}}$ |
| Συντελεστής μεταβλητότητας | $\hat{C}_{v_X} = \frac{s_X}{\bar{x}}$ | $\approx \sqrt{\frac{n}{n-1}}$ |
| Τρίτη κεντρική ροπή | $\hat{\mu}_X^{(3)} = \frac{1}{n} \sum x_i^3 - \frac{3}{n^2} (\sum x_i)(\sum x_i^2) + \frac{2}{n^3} (\sum x_i)^3$ $= \frac{1}{n} \sum x_i^3 - 3\bar{x}s_X^2 - \bar{x}^3$ | $\frac{n^2}{(n-1)(n-2)}$ |
| Συντελεστής ασυμμετρίας | $\hat{C}_{s_X} = \frac{\hat{\mu}_X^{(3)}}{s_X^3}$ | βλ. εδάφιο 3.2.4 |
| <i>2. Συνοπτικά χαρακτηριστικά διατεταγμένου δείγματος</i> | | |
| Ελάχιστη τιμή | $\hat{x}_{\min} = \min(x_1, x_2, \dots, x_n)$ | — |
| Μέγιστη τιμή | $\hat{x}_{\max} = \max(x_1, x_2, \dots, x_n)$ | — |
| Διάμεσος | $\hat{x}_{0.5}$: Ο μεσαίος όρος του διατεταγμένου δείγματος, ή για άρτιο μέγεθος δείγματος, ο μέσος όρος των δύο μεσαίων όρων του. | — |
| Κάτω τεταρτημόριο | $\hat{x}_{0.25}$: Η διάμεσος του τμήματος του δείγματος που περιλαμβάνει τις τιμές $x_i \leq \hat{x}_{0.5}$. | — |
| Άνω τεταρτημόριο | $\hat{x}_{0.75}$: Η διάμεσος του τμήματος του δείγματος που περιλαμβάνει τις τιμές $x_i \geq \hat{x}_{0.5}$. | — |
| Διατεταρτημοριακό εύρος | $\hat{\delta}_X = \hat{x}_{0.75} - \hat{x}_{0.25}$ | — |

Τα συνοπτικά χαρακτηριστικά του διατεταγμένου δείγματος μπορούν να απεικονιστούν και με τη μορφή απλού διαγράμματος, της *συνοπτικής παράστασης δείγματος* (στη διεθνή βιβλιογραφία αναφέρεται ως *box plot*), όπως αυτό του Σχ. 5.1 (σ. 111). Ένα τέτοιο διάγραμμα περιλαμβάνει ένα κεντρικό ορθογώνιο “κουτί” και δύο κατακόρυφες “κεραίες” πάνω και κάτω από το κουτί. Όλα αυτά τα στοιχεία απεικονίζονται σε κατάλληλη κατακόρυφη κλίμακα. Για τη σχεδίαση ακολουθούνται οι εξής οδηγίες (Hirsch et al., 1993, σ. 17.10):

1. Η μεσαία γραμμή του κουτιού αντιπροσωπεύει τη διάμεσο του δείγματος.
2. Η κάτω γραμμή (βάση) του κουτιού αντιπροσωπεύει το κάτω τεταρτημόριο του δείγματος.
3. Η πάνω γραμμή (οροφή) του κουτιού αντιπροσωπεύει το άνω τεταρτημόριο του δείγματος.
4. Ορίζεται το βοηθητικό μέγεθος *βήμα* το οποίο είναι 1.5 φορές το διατεταρτημοριακό πλάτος.
5. Η κάτω κεραία εκτείνεται από τη βάση του κουτιού μέχρι τη μικρότερη τιμή του δείγματος που βρίσκεται μέσα στα όρια ενός βήματος από τη βάση του κουτιού.
6. Η άνω κεραία εκτείνεται από την οροφή του κουτιού μέχρι τη μεγαλύτερη τιμή του δείγματος που βρίσκεται μέσα στα όρια ενός βήματος από την οροφή του κουτιού.
7. Τιμές του δείγματος που βρίσκονται έξω από τα όρια των κεραιών και απέχουν 1-2 βήματα από τη βάση ή την οροφή του κουτιού χαρακτηρίζονται ως *εξωτερικές* και σημειώνονται στο διάγραμμα με το σύμβολο \times .
8. Τιμές του δείγματος που βρίσκονται έξω από τα όρια των κεραιών και απέχουν περισσότερο από 2 βήματα από τη βάση ή την οροφή του κουτιού χαρακτηρίζονται ως *μακρινές εξωτερικές* και σημειώνονται στο διάγραμμα με το σύμβολο \circ .

Σύμφωνα με τα παραπάνω, η ελάχιστη και η μέγιστη τιμή του δείγματος σημειώνονται στη συνοπτική παράσταση δείγματος είτε ως όρια των κεραιών, εφόσον δεν απέχουν περισσότερο από ένα βήμα από τα όρια

του κουτιού, είτε διαφορετικά ως εξωτερικές ή μακρινές εξωτερικές τιμές.

Η συνοπτική παράσταση ενός δείγματος δίνει με πολύ σαφή τρόπο τη γενική στατιστική εικόνα του, απεικονίζοντας ταυτόχρονα χαρακτηριστικά θέσης (διάμεσος), διασποράς (διατεταρτημοριακό πλάτος) και ασυμμετρίας. Η συμμετρία ή ασυμμετρία του δείγματος αναγνωρίζεται από θέση της μεσαίας γραμμής σε σχέση με τις βάσεις του κουτιού, καθώς και από τη σύγκριση των μηκών των κεραιών. Ακόμη το διάγραμμα δίνει πληροφορίες για το πόσο κοντά στην κανονική κατανομή είναι ένα δείγμα. Αν η μεταβλητή του δείγματος είναι κανονική, τότε αναμένεται κατ' αρχήν συμμετρική εικόνα του διαγράμματος, και επίσης δεν αναμένονται εξωτερικές και μακρινές εξωτερικές τιμές, παρά μόνο με συχνότητες 1 στα 100 και 1 στα 300 000 σημεία, αντίστοιχα.

Εφαρμογή 5.1

Στον Πίν. 5.2 δίνεται το δείγμα των ετήσιων απορροών της λεκάνης του ποταμού Ευήνου ανάντη του υδρομετρικού σταθμού Πόρος Ρηγανίου. Ζητείται ο υπολογισμός των στατιστικών χαρακτηριστικών του και η σχεδίαση της συνοπτικής παράστασής του.*

Πίν. 5.2 Δείγμα ετήσιου όγκου απορροής (σε hm^3)[†] στη θέση Πόρος Ρηγανίου του ποταμού Ευήνου.

| Υδρολογ. έτος | Όγκος απορροής | Υδρολογ. έτος | Όγκος απορροής | Υδρολογ. έτος | Όγκος απορροής |
|------------------|-------------------|------------------|-------------------|------------------|-------------------|
| 1970-71 | 807 | 1977-78 | 715 | 1984-85 | 588 |
| 1971-72 | 695 | 1978-79 | 1064 | 1985-86 | 874 |
| 1972-73 | 788 | 1979-80 | 942 | 1986-87 | 552 |
| 1973-74 | 705 | 1980-81 | 1042 | 1987-88 | 529 |
| 1974-75 | 462 | 1981-82 | 1037 | 1988-89 | 469 |
| 1975-76 | 580 | 1982-83 | 674 | 1989-90 | 217 |
| 1976-77 | 807 | 1983-84 | 906 | 1990-91 | 772 |

* Η θέση Πόρος Ρηγανίου βρίσκεται σε αρκετή απόσταση κατάντη του φράγματος Αγίου Δημητρίου, το οποίο εντάσσεται στο σύστημα υδατικών πόρων που χρησιμοποιείται για την ύδρευση της Αθήνας. Το δείγμα των απορροών του Πόρου Ρηγανίου χρησιμοποιήθηκε (μετά από κατάλληλη αναγωγή) για τη μελέτη του ταμιευτήρα Αγίου Δημητρίου.

[†] Υπενθυμίζεται ότι η μονάδα hm^3 σημαίνει κυβικά εκατόμετρα ($1 \text{ hm}^3 = (100 \text{ m})^3 = 1\,000\,000 \text{ m}^3$).

α. Δειγματικές ροπές και παράγωγα χαρακτηριστικά

Αν και σήμερα ο υπολογισμός των ροπών γίνεται αυτόματα με υπολογιστές ή αριθμομηχανές, εδώ εκθέτουμε για διδακτικούς λόγους τον υπολογισμό των στατιστικών χαρακτηριστικών με το χέρι.

Πίν. 5.3 Βοηθητικοί υπολογισμοί για την Εφαρμογή 5.1.

| i | x_i | x_i^2 | x_i^3 |
|----------|--------|------------|---------------|
| 1 | 807 | 651 249 | 525 557 943 |
| 2 | 695 | 483 025 | 335 702 375 |
| 3 | 788 | 620 944 | 489 303 872 |
| 4 | 705 | 497 025 | 350 402 625 |
| 5 | 462 | 213 444 | 98 611 128 |
| 6 | 580 | 336 400 | 195 112 000 |
| 7 | 807 | 651 249 | 525 557 943 |
| 8 | 715 | 511 225 | 365 525 875 |
| 9 | 1064 | 1 132 096 | 1 204 550 144 |
| 10 | 942 | 887 364 | 835 896 888 |
| 11 | 1042 | 1 085 764 | 1 131 366 088 |
| 12 | 1037 | 1 075 369 | 1 115 157 653 |
| 13 | 674 | 454 276 | 306 182 024 |
| 14 | 906 | 820 836 | 743 677 416 |
| 15 | 588 | 345 744 | 203 297 472 |
| 16 | 874 | 763 876 | 667 627 624 |
| 17 | 552 | 304 704 | 168 196 608 |
| 18 | 529 | 279 841 | 148 035 889 |
| 19 | 469 | 219 961 | 103 161 709 |
| 20 | 217 | 47 089 | 10 218 313 |
| 21 | 772 | 595 984 | 460 099 648 |
| Άθροισμα | 15 225 | 11 977 465 | 9 983 241 237 |

Ο υπολογισμός των αθροισμάτων $\sum x$, $\sum x^2$ και $\sum x^3$ γίνεται στον Πίν. 5.3. Έχουμε $\sum x = 15\,225$, $\sum x^2 = 11\,977\,465$ και $\sum x^3 = 9\,983\,241\,237$. Η μέση τιμή του δείγματος είναι

$$\bar{x} = \sum x / n = 15\,225 / 21 = 725.0 \text{ hm}^3$$

Η διασπορά είναι

$$s_x^2 = \sum x^2 / n - \bar{x}^2 = 11\,977\,465 / 21 - 725.0^2 = 44\,730.5 \text{ (hm}^3)^2$$

η τυπική απόκλιση

$$s_X = \sqrt{44\,730.5} = 211.5 \text{ hm}^3$$

και ο συντελεστής μεταβλητότητας

$$\hat{C}_{v_X} = s_X / \bar{x} = 211.5 / 725.0 = 0.29$$

Η τρίτη κεντρική ροπή είναι

$$\begin{aligned} \hat{\mu}_X^{(3)} &= \sum x^3 / n - 3 \bar{x} s_X^2 - \bar{x}^3 = \\ &= 9\,983\,241\,237 / 21 - 3 \times 725.0 \times 44\,730.5 - 725.0^3 \\ &= -2\,974\,523 \text{ (hm}^3)^3 \end{aligned}$$

και ο συντελεστής ασυμμετρίας

$$\hat{C}_{s_X} = \hat{\mu}_X^{(3)} / s_X^3 = -2\,974\,523 / 211.5^3 = -0.31$$

Οι συντελεστές διόρθωσης μεροληψίας είναι: (α) για τη διασπορά

$$n / (n - 1) = 21 / 20 = 1.05.$$

(β) για την τυπική απόκλιση και το συντελεστή μεταβλητότητας (κατά προσέγγιση)

$$\sqrt{n / (n - 1)} = \sqrt{1.05} = 1.025$$

και (γ) για την τρίτη κεντρική ροπή (και κατά προσέγγιση για το συντελεστή ασυμμετρίας, βλ. εξίσωση 3.28)

$$n^2 / [(n-1)(n-2)] = 21^2 / (20 \times 19) = 1.16$$

β. Συνοπτικά χαρακτηριστικά διατεταγμένου δείγματος

Το δείγμα, διατεταγμένο σε φθίνουσα σειρά, φαίνεται στον Πίν. 5.5. Από τον πίνακα αυτό υπολογίζονται άμεσα τα συνοπτικά χαρακτηριστικά διατεταγμένου δείγματος, τα οποία δίνονται στον Πίν. 5.4. Η διάμεσος είναι η 11η στη σειρά (μεσαία) τιμή του διατεταγμένου δείγματος, ενώ το κάτω και άνω τεταρτημόριο είναι η 16η και 6η στη σειρά τιμή, αντίστοιχα.

Πίν. 5.4 Στατιστικά χαρακτηριστικά (ροπές και παράγωγα χαρακτηριστικά) του δείγματος του ετήσιου όγκου απορροής (σε hm^3) στη θέση Πόρος Ρηγανίου του ποταμού Ευήνου.

| Στατιστικός δείκτης | Απλή εκτίμηση | Συντελεστής διόρθωσης μεροληψίας | Αμερόληπτη εκτίμηση |
|----------------------------|---|---|-------------------------------------|
| Μέση τιμή | $\bar{x} = \sum x / n = 725.0$ | — | $\bar{x} = 725.0$ |
| Διασπορά | $s_X^2 = \sum x^2 / n - \bar{x}^2 = 44\,730$ | $\frac{n}{n-1} = 1.05$ | $s_X^{*2} = 46\,967$ |
| Τυπική απόκλιση | $s_X = 211.5$ | $\approx \sqrt{\frac{n}{n-1}} = 1.025$ | $s_X^* \approx 216.7$ |
| Συντελεστής μεταβλητότητας | $\hat{C}_{v_X} = s_X / \bar{x} = 0.29$ | $\approx \sqrt{\frac{n}{n-1}} = 1.025$ | $\hat{C}_{v_X}^* \approx 0.29$ |
| Τρίτη κεντρική ροπή | $\hat{\mu}_X^{(3)} = \sum x^3 / n - 3\bar{x}s_X^2 - \bar{x}^3 = -2\,974\,523$ | $\frac{n^2}{(n-1)(n-2)} = 1.16$ | $\hat{\mu}_X^{*(3)} = -3\,542\,012$ |
| Συντελεστής ασυμμετρίας | $\hat{C}_{s_X} = \hat{\mu}_X^{(3)} / s_X^3 = -0.31$ | $\approx \frac{n^2}{(n-1)(n-2)} = 1.16$ | $\hat{C}_{s_X}^* \approx -0.36$ |

Πίν. 5.5 Διατεταγμένο (σε φθίνουσα σειρά) δείγμα ετήσιου όγκου απορροής (σε hm^3) στη θέση Πόρος Ρηγανίου του ποταμού Ευήνου.

| A/A | Όγκος απορροής | A/A | Όγκος απορροής | A/A | Όγκος απορροής |
|-----|----------------|-----|----------------|-----|----------------|
| 1 | 1064 | 8 | 807 | 15 | 588 |
| 2 | 1042 | 9 | 788 | 16 | 580 |
| 3 | 1037 | 10 | 772 | 17 | 552 |
| 4 | 942 | 11 | 715 | 18 | 529 |
| 5 | 906 | 12 | 705 | 19 | 469 |
| 6 | 874 | 13 | 695 | 20 | 462 |
| 7 | 807 | 14 | 674 | 21 | 217 |

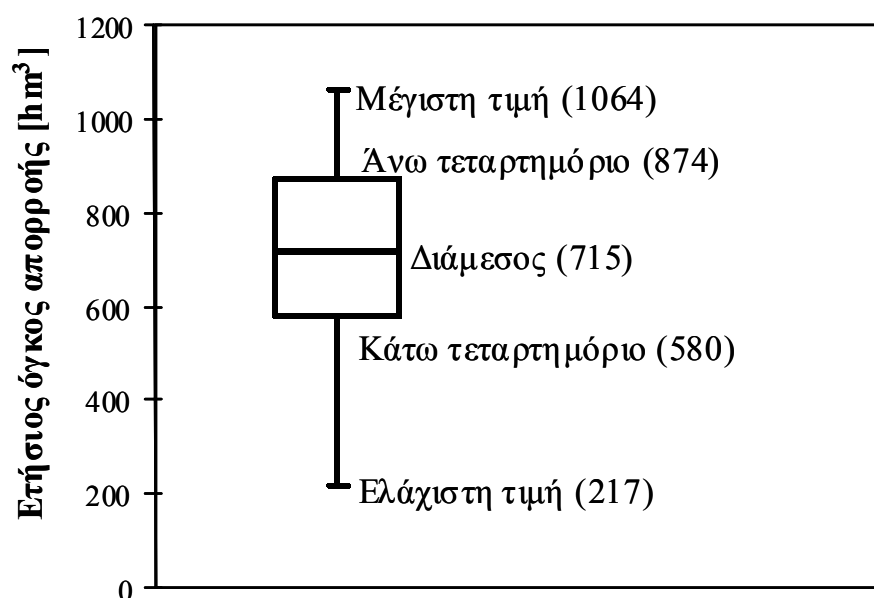
γ. Σχεδίαση συνοπτικής παράστασης δείγματος

Ακολουθώντας τη διαδικασία που περιγράφηκε στην ενότητα 5.1 και χρησιμοποιώντας τα συνοπτικά χαρακτηριστικά διατεταγμένου δείγματος (Πίν. 5.4) σχεδιάζουμε εύκολα το διάγραμμα του Σχ. 5.1. Το βήμα της συνοπτικής παράστασης είναι $1.5 \times 294 = 441 \text{ hm}^3$ και επομένως η μέγιστη τεταγμένη της άνω κεραίας είναι $874 + 441 = 1315$

hm^3 . Δεδομένου, όμως, ότι η μέγιστη τιμή του δείγματος είναι 1064 hm^3 , η άνω κεραία πρέπει να εκτείνεται μέχρι αυτή την τιμή. Αντίστοιχα, η ελάχιστη τεταγμένη της κάτω κεραίας είναι $580 - 441 = 139 \text{ hm}^3$. Δεδομένου, όμως, ότι η ελάχιστη τιμή του δείγματος είναι 217 hm^3 , η κάτω κεραία πρέπει να εκτείνεται μέχρι αυτή την τιμή.

Πίν. 5.6 Συνοπτικά χαρακτηριστικά του διατεταγμένου δείγματος του ετήσιου όγκου απορροής (σε hm^3) στη θέση Πόρος Ρηγανίου του ποταμού Ευήνου.

| Στατιστικός δείκτης | Εκτίμηση |
|-------------------------|---|
| Ελάχιστη τιμή | $\hat{x}_{\max} = \min(x_1, \dots, x_n) = 217$ |
| Μέγιστη τιμή | $\hat{x}_{\max} = \max(x_1, \dots, x_n) = 1064$ |
| Διάμεσος | $\hat{x}_{0.5} = x_{(11)} = 715$ |
| Κάτω τεταρτημόριο | $\hat{x}_{0.25} = x_{(16)} = 580$ |
| Άνω τεταρτημόριο | $\hat{x}_{0.75} = x_{(6)} = 874$ |
| Διατεταρτημοριακό εύρος | $\hat{d}_X = \hat{x}_{0.75} - \hat{x}_{0.25} = 294$ |



Σχ. 5.1 Συνοπτική παράσταση δείγματος (box plot) για το παράδειγμα της εφαρμογής 5.1

5.2 Ιστογράμματα

Ένα άλλο τρόπο γραφικής παράστασης ενός δείγματος παρέχουν τα ιστογράμματα, τα οποία προϋποθέτουν την κατάταξη των τιμών του δείγμα-

τος σε k διαστήματα μεγέθους Δ . Έστω ότι το i διάστημα έχει όρια c_i και $c_{i+1} = c_i + \Delta$ και ότι ο αριθμός των τιμών του δείγματος που βρίσκονται στο i διάστημα, δηλαδή αυτών που ικανοποιούν τη σχέση $c_i \leq x < c_{i+1}$, είναι n_i . Ορίζουμε ως ιστόγραμμα την κλιμακωτή συνάρτηση

$$\varphi(x) = \frac{n_i}{n\Delta} \quad c_i \leq x < c_{i+1} \quad i = 1, \dots, k \quad (5.1)$$

της οποίας η γραφική παράσταση έχει τη μορφή του Σχ. 5.2. Συχνά η συνάρτηση του ιστογράμματος ορίζεται με απλούστερους τρόπους ως $\varphi(x) = n_i/n$, ή και $\varphi(x) = n_i$. Για τις τελευταίες χρησιμοποιούμε τους όρους *ιστόγραμμα σχετικής συχνότητας* και *ιστόγραμμα απόλυτης συχνότητας*, αντίστοιχα. Προς αποφυγή σύγχυσης, για τη συνάρτηση της (5.1) χρησιμοποιούμε και τον όρο *ιστόγραμμα πυκνότητας συχνότητας*.

Προκειμένου να καταρτίσουμε το ιστόγραμμα, αρχικά επιλέγουμε τον αριθμό των διαστημάτων k . Κατά κανόνα λαμβάνουμε $k = \ln n / \ln 2$ και την τιμή που προκύπτει τη στρογγυλεύουμε προς τα πάνω. Το μέγεθος των διαστημάτων Δ το παίρνουμε ίσο για όλα τα διαστήματα (αν και για το ιστόγραμμα πυκνότητας συχνότητας είναι επιτρεπτά και άνισα διαστήματα).

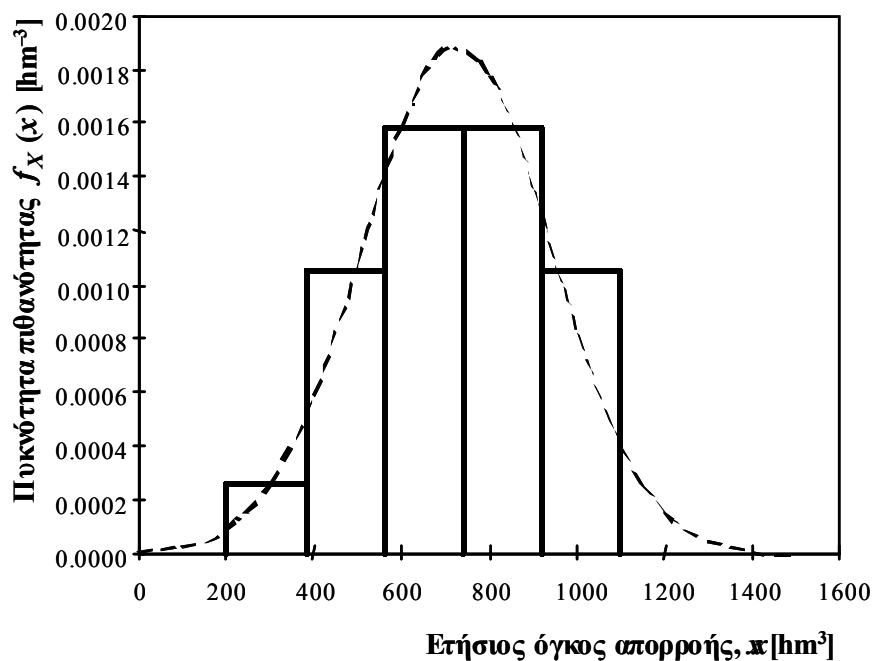
Εφαρμογή 5.2

Να σχεδιαστεί το ιστόγραμμα για το δείγμα της Εφαρμογής 5.1.

Ο αριθμός των κλάσεων θα πρέπει να ληφθεί $k = \ln 21 / \ln 2 = 4.4$. Στρογγυλεύοντας προς τα πάνω, επιλέγουμε 5 κλάσεις. Το διάστημα μεταβολής των τιμών της μεταβλητής στο δείγμα είναι το [217, 1064]. Στρογγυλεύοντας, παίρνουμε για διάστημα μεταβολής το [200, 1100] με $\Delta = (1100 - 200) / 5 = 180$. Οι υπόλοιποι υπολογισμοί δίνονται σε πινακοποιημένη μορφή στον Πίν. 5.7 και το ιστόγραμμα παρουσιάζεται στο Σχ. 5.2. Για λόγους σύγκρισης έχει χαραχτεί και η θεωρητική συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής (βλ. ενότητα 5.4).

Πίν. 5.7 Υπολογισμοί για το ιστόγραμμα της Εφαρμογής 5.2.

| A/A | Όρια κλάσεων | Απόλυτη συχνότητα n_i | Σχετική συχνότητα n_i / n | Πυκνότητα συχνότητας $\phi = n_i / (n \Delta)$ |
|-----|--------------|-------------------------|-----------------------------|--|
| | 200 | | | |
| 1 | 380 | 1 | 0.048 | 0.00026 |
| 2 | 560 | 4 | 0.190 | 0.00106 |
| 3 | 740 | 6 | 0.286 | 0.00159 |
| 4 | 920 | 6 | 0.286 | 0.00159 |
| 5 | 1100 | 4 | 0.190 | 0.00106 |



Σχ. 5.2 Ιστόγραμμα για το παράδειγμα της εφαρμογής 5.2. Με διακεκομμένη γραμμή έχει χαραχτεί (για σύγκριση) η πυκνότητα πιθανότητας της κανονικής κατανομής $N(725, 211.5)$.

5.3 Εμπειρική συνάρτηση κατανομής

Όπως το ιστόγραμμα είναι το εμπειρικό αντίστοιχο της συνάρτησης πυκνότητας πιθανότητας, έτσι έχουμε και εμπειρικό αντίστοιχο της συνάρτησης κατανομής, το οποίο λέγεται *εμπειρική συνάρτηση κατανο-*

μής. Μια τέτοια εμπειρική συνάρτηση μπορεί κατ' αρχήν να κατασκευαστεί από ένα ιστόγραμμα, με ολοκλήρωσή του ως προς x , οπότε προκύπτει μια αύξουσα τεθλασμένη που αντιστοιχεί σε κάποιο είδος συνάρτησης κατανομής. Ωστόσο, η εισαγωγή της εμπειρικής συνάρτησης κατανομής μπορεί να γίνει με αμεσότερο και αντικειμενικότερο τρόπο, χωρίς τη μεσολάβηση του ιστογράμματος, το οποίο έχει κάποιο βαθμό υποκειμενικότητας λόγω της αυθαίρετης επιλογής των διαστημάτων και των ορίων τους.

5.3.1 Κλασική εμπειρική συνάρτηση κατανομής

Έστω η τυχαία μεταβλητή X με συνάρτηση κατανομής $F(x)$ και το δείγμα της X_1, X_2, \dots, X_n . Συμβολίζουμε με $X_{(1)}$ τη μεγαλύτερη από τις μεταβλητές του δείγματος, $X_{(2)}$ την αμέσως μικρότερη, κοκ., έτσι ώστε οι μεταβλητές $X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(n)}$ να παριστάνουν το δείγμα διατεταγμένο σε φθίνουσα σειρά. Οι $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ λέγονται και *διατεταγμένες στατιστικές συναρτήσεις*.

Σύμφωνα με τον κλασικό τρόπο, η εμπειρική συνάρτηση κατανομής είναι κλιμακωτή και ορίζεται από τη σχέση

$$\hat{F}(x) = \frac{n_x}{n} \quad (5.2)$$

όπου n_x είναι ο αριθμός των τιμών του δείγματος που δεν υπερβαίνουν την τιμή x . Η $\hat{F}(x)$ αποτελεί σημειακή εκτίμηση της άγνωστης συνάρτησης κατανομής του πληθυσμού $F(x)$.

5.3.2 Θέσεις σχεδίασης

Ονομάζουμε *θέση σχεδίασης* (plotting position) q_i της τιμής $x_{(i)}$ του διατεταγμένου δείγματος την εμπειρική πιθανότητα υπέρβασης της τιμής αυτής. Με βάση τον κλασικό τρόπο ορισμού της εμπειρικής κατανομής, για $x = x_{(1)}$ θα έχουμε $n_x = n$, και γενικότερα για $x = x_{(i)}$ θα έχουμε $n_x = n + 1 - i$. Κατά συνέπεια η εμπειρική συνάρτηση κατανομής είναι

$$\hat{F}(x_{(i)}) = \frac{n+1-i}{n} \quad i = 1, \dots, n \quad (5.3)$$

οπότε η θέση σχεδίασης, δηλαδή η εμπειρική πιθανότητα υπέρβασης είναι

$$q_i = \hat{F}_1(x_{(i)}) = 1 - \hat{F}(x_{(i)}) = \frac{i-1}{n} \quad i = 1, \dots, n \quad (5.4)$$

Παρατηρούμε ότι για $n = 1$ η πιο πάνω εξίσωση δίνει μηδενική πιθανότητα υπέρβασης. Έτσι, αν για παράδειγμα έχουμε ένα δείγμα ετήσιων βροχοπτώσεων με μέγιστη τιμή $x_{(1)} = 1800$ mm, οδηγούμαστε στο συμπέρασμα ότι η πιθανότητα ετήσιας βροχόπτωσης μεγαλύτερης από 1800 mm είναι μηδενική. Αυτό φυσικά είναι εσφαλμένο συμπέρασμα.

Για την άρση του παραπάνω προβλήματος χρησιμοποιούμε την τυχαία μεταβλητή

$$U_i = F_1(X_{(i)}) = 1 - F(X_{(i)}) \quad (5.5)$$

Μια σημειακή εκτίμηση[†] αυτής της μεταβλητής αποτελεί ταυτόχρονα και εκτίμηση της q_i . Κατ' αρχήν δεν μπορούμε να υπολογίσουμε τιμές της U_i από το δείγμα δεδομένου ότι η $F(x)$ είναι άγνωστη συνάρτηση. Ωστόσο, αποδεικνύεται[‡] ότι η κατανομή της U_i είναι ανεξάρτητη της $F(x)$ και έχει μέση τιμή*

$$E[U_i] = \frac{i}{n+1} \quad (5.6)$$

και διασπορά

$$\text{Var}[U_i] = \frac{i(n-i+1)}{(n+1)^2(n+2)} \quad (5.7)$$

[†] Ακριβέστερα και σύμφωνα με την ορολογία του κεφαλαίου 3, πρόκειται για πρόγνωση της μεταβλητής, εφόσον η U_i είναι τυχαία μεταβλητή και όχι παράμετρος.

[‡] Προκύπτει από τη συνάρτηση κατανομής των διατεταγμένων στατιστικών συναρτήσεων (βλ. Papoulis, 1990, σ. 207-208) μετά από κατάλληλη αλλαγή μεταβλητής.

* Για την ακρίβεια η U_i έχει συνάρτηση κατανομής βήτα (βλ. εδάφιο 6.4.1 με παραμέτρους i και $n - i + 1$).

Η απλούστερη εκτίμηση της U_i είναι η μέση τιμή της, δηλαδή

$$q_i = \frac{i}{n+1} \quad (5.8)$$

η οποία είναι γνωστή στη βιβλιογραφία ως *θέση σχεδίασης Weibull* ή *αμερόληπτη ως προς την πιθανότητα υπέρβασης θέση σχεδίασης*. Ο δεύτερος χαρακτηρισμός οφείλεται στην ιδιότητά της: $q_i = E[U_i] = E[F(X_{(i)})]$. Παρατηρούμε ότι με αυτή την εκτίμηση έχει αρθεί το πρόβλημα του μη-δενισμού της q_i για $i = 1$. Πράγματι για $i = 1$ έχουμε $q_i = 1 / (n + 1)$ και για $i = n$, $q_i = n / (n + 1)$.

Η εξίσωση (5.8) είναι η πιο διαδεδομένη για την εκτίμηση εμπειρικών πιθανοτήτων υπέρβασης στην τεχνική υδρολογία, αλλά δεν είναι η μοναδική. Άλλες παρόμοιες εξισώσεις έχουν αναπτυχθεί με στόχο να δίνουν αμερόληπτες εκτιμήσεις ποσοστημορίων, δηλαδή να ικανοποιούν (κατά προσέγγιση) τη συνθήκη

$$F^{-1}(q_i) = E[X_{(i)}] = E[F^{-1}(U_i)] \quad (5.9)$$

Στην περίπτωση αυτή η εκτίμηση, αντίθετα από αυτήν της (5.8), δεν είναι ανεξάρτητη της συνάρτησης κατανομής $F(x)$. Πάντως στην πράξη οι διάφορες εξισώσεις που έχουν αναπτυχθεί εκφράζονται με το γενικό τύπο

$$q_i = \frac{i - a}{n + 1 - 2a} \quad (5.10)$$

όπου a σταθερά (< 1). Ο τύπος αυτός είναι αντισυμμετρικός με την έννοια ότι $q_i = 1 - q_{n+1-i}$ και εμπεριέχει, ως ειδική περίπτωση ($a = 0$), και την (5.8). Στον Πίν. 5.8 δίνονται οι συνηθέστεροι στην τεχνική υδρολογία τύποι υπολογισμού της θέσης σχεδίασης μαζί με τις αντίστοιχες τιμές της σταθεράς a . Η εφαρμογή των διάφορων σχέσεων δίνει πρακτικά τα ίδια αποτελέσματα, εκτός από τις πολύ μικρές του i και ιδίως για $i = 1$, όπου διαφοροποιούνται αισθητά. Η τιμή για $i = 1$ έχει μεγάλη σημασία για την τεχνική υδρολογία, γιατί δίνει την εμπειρική περίοδο επαναφοράς της μεγαλύτερης τιμής του δείγματος, δηλαδή την $T_1 = 1 / q_1$. Η διαφοροποίηση του T_1 για τους διάφορους τύπους φαίνεται στον Πίν. 5.8.

Πίν. 5.8 Εναλλακτικοί τύποι υπολογισμού εμπειρικών πιθανοτήτων υπέρβασης *

| Όνομα | Τύπος $q_i =$ | Σταθερά $a =$ | Περίοδος επα- ναφοράς μέγ. τιμής $T_1 =$ | Εφαρμοσιμότητα |
|------------|--------------------------|------------------|---|---|
| Weibull | $\frac{i}{n+1}$ | 0 | $n+1$ | Όλες οι κατανομές, αμερόληπτη εκτίμηση πιθανότητας υπέρβασης |
| Blom | $\frac{i-0.375}{n+0.25}$ | 0.375 | $1.6n+0.4$ | Κανονική κατανομή, αμερόληπτη εκτίμηση ποσοστημορίων |
| Cunnane | $\frac{i-0.4}{n+0.2}$ | 0.4 | $1.667n+0.33$ | Μεγάλο εύρος κατανομών, κατά προσέγγιση αμερόληπτη εκτίμηση ποσοστημορίων |
| Gringorten | $\frac{i-0.44}{n+0.12}$ | 0.44 | $1.786n+0.21$ | Κατανομή Gumbel † |
| Hazen | $\frac{i-0.5}{n}$ | 0.5 | $2n$ | Εμπειρική εκτίμηση ιστορικής αξίας· σήμερα τείνει να εγκαταλειφθεί |

5.3.3 Χαρτιά πιθανότητας

Έχοντας εκτιμήσει τη θέση σχεδίασης κάθε τιμής του δείγματος με κάποιον από τους παραπάνω τρόπους, διαθέτουμε ένα σύνολο από n σημεία $(x_{(i)}, q_i)$ ή $(x_{(i)}, 1 - q_i)$, τα οποία μπορούν να παρασταθούν γραφικά ώστε να πάρουμε μια εικόνα της συνάρτησης κατανομής. Αυτό μπορεί κατ' αρχήν να γίνει σε κοινό ή μιλιμετρέ χαρτί, οπότε θα προκύψει ένα γράφημα παρόμοιο με αυτό του Σχ. 2.1 ή του Σχ. 2.3β, με τη διαφορά ότι, αντί της κλιμακωτής ή της συνεχούς γραμμής, θα έχουμε απλώς ένα σύνολο από σημεία. Όμως, στην τεχνική υδρολογία, επειδή η πληροφορία που παρέχει ένα τέτοιο διάγραμμα είναι πολύ ουσιαστική, είμαστε πιο συστηματικοί στη σχεδίασή του. Συγκεκριμένα, επιδιώκουμε, μέσω κατάλληλων μετασχηματισμών των αξόνων, να δώσουμε ευθύγραμμη

* Βλ. και Stedinger et al. (1993) όπου δίνονται και άλλοι τύποι.

† Βλ. κεφάλαιο 6.

διάταξη στα σημεία. Αυτό εξυπηρετεί διάφορους σκοπούς, όπως την ευκολότερη σχεδίαση, τη σαφέστερη σύγκριση θεωρητικής και εμπειρικής κατανομής, την ευκολότερη γραφική επέκταση έξω από τα όρια του δείγματος κ.ά. Χαρτιά στα οποία έχουν σχεδιαστεί οι άξονες με κατάλληλους μετασχηματισμούς, ώστε τα γραφήματα συγκεκριμένων συναρτήσεων κατανομής να παριστάνονται ως ευθείες, λέγονται *χαρτιά πιθανότητας*. Τέτοια χαρτιά συνήθως υπάρχουν έτοιμα (όπως το λογαριθμικό χαρτί) αλλά μπορούν και να κατασκευαστούν εύκολα.

Ας πάρουμε για παράδειγμα την κανονική κατανομή $N(\mu, \sigma)$. Αν παραστήσουμε γραφικά τη συνάρτηση $F(x)$ με οριζόντιο άξονα $h = F$ και κατακόρυφο $v = x$, τότε θα προκύψει ένα γράφημα μορφής \mathcal{J} . Γνωρίζουμε όμως ότι $x = \mu + \sigma z_F$, όπου z_F το F -ποσοστημόριο της τυποποιημένης κανονικής κατανομής $N(0, 1)$. Κατά συνέπεια, αν θέσουμε στον οριζόντιο άξονα $h = z_F$, τότε η εξίσωση του γραφήματος θα είναι $v = \mu + \sigma h$, η οποία παριστάνει ευθεία. Αυτό ισοδυναμεί με το μετασχηματισμό του οριζόντιου άξονα $h = z_F = F_0^{-1}(F)$, όπου $F_0^{-1}(\cdot)$ η αντίστροφη της τυποποιημένης κανονικής συνάρτησης κατανομής. Με άλλους κατάλληλους μετασχηματισμούς του οριζόντιου ή/και του κατακόρυφου άξονα μπορούμε να πετύχουμε ευθειοποίηση των γραφημάτων άλλων συναρτήσεων κατανομής, όπως αναλυτικότερα θα δούμε στο κεφάλαιο 6.

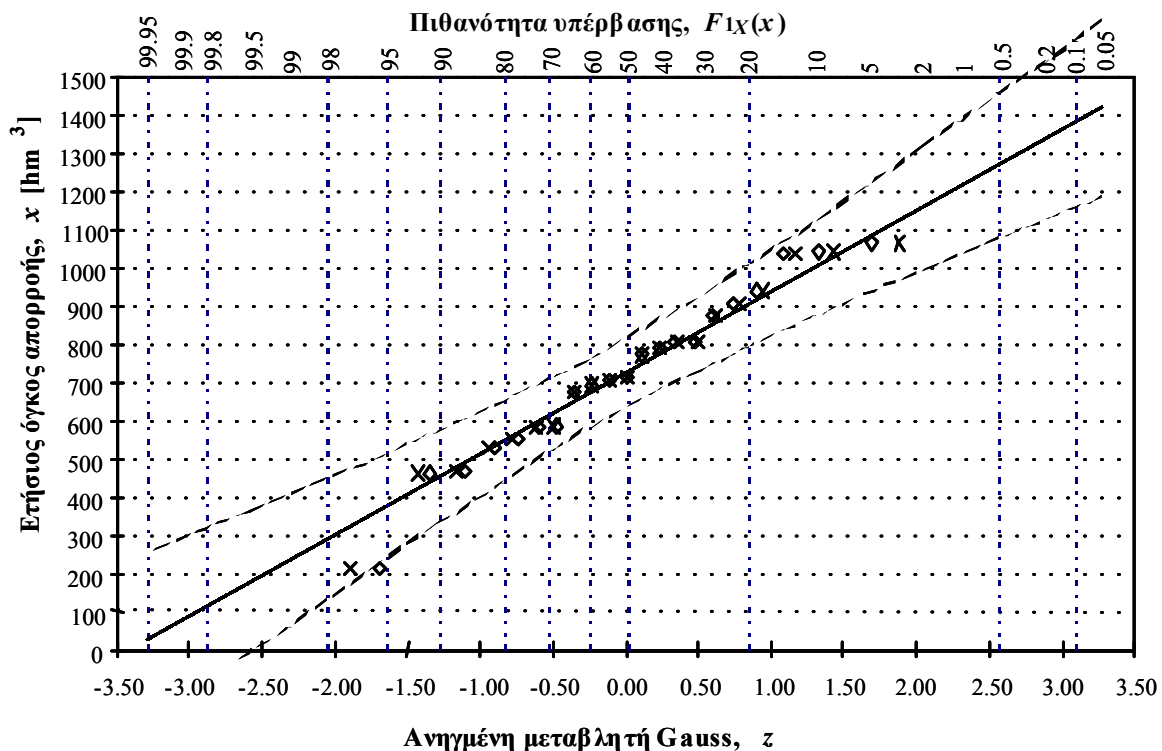
Επειδή υπάρχει ένα-προς-ένα αντιστοιχία μεταξύ των μεγεθών F και z_F , η βαθμονόμηση του οριζόντιου άξονα μπορεί να γίνει σε μονάδες F αντί z_F , πράγμα που διευκολύνει τη διαδικασία κατάρτισης της γραφικής παράστασης. Ακόμη, η βαθμονόμηση του οριζόντιου άξονα μπορεί να γίνει ως προς την πιθανότητα υπέρβασης $F_1 = 1 - F$ ή ως προς την περίοδο επαναφοράς $T = 1 / F_1$. Ένα παράδειγμα χαρτιού κανονικής κατανομής φαίνεται στο Σχ. 5.3, όπου μάλιστα φαίνονται ταυτόχρονα δύο βαθμονομήσεις του οριζόντιου άξονα, ως προς z_F και ως προς F_1 .

Η γραφική παράσταση του σημειοσυνόλου της εμπειρικής συχνότητας υπέρβασης $(x_{(i)}, q_i)$ σε χαρτί κανονικής κατανομής (δηλαδή $h = z_{1-q_i}$, $v = x_{(i)}$) θα δίνει μια περίπου ευθύγραμμη διάταξη των σημείων αν η κατανομή του πληθυσμού X είναι κανονική. Έτσι η γραφική παράσταση αυτή μας προσφέρει ένα γραφικό τρόπο ελέγχου της κανονικότητας ως προς την κατανομή ενός δείγματος. Τα παραπάνω διευκρινίζονται με το παράδειγμα που ακολουθεί.

Εφαρμογή 5.3

Να υπολογιστούν και να σχεδιαστούν σε χαρτί κανονικής κατανομής οι εμπειρικές πιθανότητες υπέρβασης για το δείγμα της Εφαρμογής 5.1.

Για τον υπολογισμό των εμπειρικών πιθανοτήτων υπέρβασης θα χρησιμοποιήσουμε τους τύπους Weibull (αμερόληπτη εκτίμηση πιθανότητας υπέρβασης) και Blom (αμερόληπτη εκτίμηση ποσοστημορίων κανονικής κατανομής, βλ. και Πίν. 5.8). Οι υπολογισμοί είναι απλούστατοι και προϋποθέτουν την κατάταξη του δείγματος σε φθίνουσα σειρά (βλ. Πίν. 5.5). Πινακοποίησή τους δίνεται στον Πίν. 5.9. Οι δύο τελευταίες στήλες κανονικά δεν είναι απαραίτητες αλλά έχουν συμπεριληφθεί για λόγους πληρότητας. Χρειάζονται μόνο στην περίπτωση που δεν διατίθεται κατάλληλα γραμμογραφημένο χαρτί κανονικής κατανομής (όπως αυτό του Σχ. 5.3), οπότε απεικονίζονται οι παρατηρημένες τιμές της μεταβλητής x_i συναρτήσει των τιμών της τυποποιημένης κανονικής μεταβλητής z_{1-q_i} . Οι τελευταίες είτε βρίσκονται από τον πίνακα της κανονικής κατανομής (Πίν. Π1), είτε υπολογίζονται με αριθμητικές μεθόδους (βλ. εδάφιο 6.1.1 και Παράρτημα 6.Α). Η απεικόνιση των εμπειρικών πιθανοτήτων υπέρβασης του δείγματος έγινε στο Σχ. 5.3. Για λόγους σύγκρισης έχει χαραχτεί και η θεωρητική συνάρτηση κανονικής κατανομής (βλ. ενότητα 5.5.4).



Σχ. 5.3 Εμπειρική συνάρτηση κατανομής κατά Weibull (ρόμβοι) και Blom (σύμβολα \times) σε χαρτί κανονικής κατανομής, για το παράδειγμα της Εφαρμογής 5.3. Με συνεχή γραμμή απεικονίζεται η κανονική συνάρτηση κατανομής $N(725, 211.5)$ και με διακεκομμένες τα αντίστοιχα όρια εμπιστοσύνης 95% (βλ. και Εφαρμογή 5.4 καθώς και Εφαρμογή 5.6).

Πίν. 5.9 Υπολογισμοί εμπειρικών πιθανοτήτων υπέρβασης για την Εφαρμογή 5.3.

| A/A | Τιμή μεταβλητής | Εμπειρική πιθανότητα υπέρβασης | | Τιμή τυποποιημένης κανονικής μεταβλητής | |
|--------|-----------------|--------------------------------|--------------------------------|---|-------------|
| | | Weibull | Blom | Weibull | Blom |
| i | x_i | $q_i = \frac{i}{n+1}$ | $q_i = \frac{i-0.375}{n+0.25}$ | z_{1-q_i} | z_{1-q_i} |
| 1 | 1064 | 0.045 | 0.029 | 1.691 | 1.890 |
| 2 | 1042 | 0.091 | 0.076 | 1.335 | 1.429 |
| 3 | 1037 | 0.136 | 0.124 | 1.097 | 1.158 |
| 4 | 942 | 0.182 | 0.171 | 0.908 | 0.952 |
| 5 | 906 | 0.227 | 0.218 | 0.748 | 0.780 |
| 6 | 874 | 0.273 | 0.265 | 0.605 | 0.629 |
| 7 | 807 | 0.318 | 0.312 | 0.473 | 0.491 |
| 8 | 807 | 0.364 | 0.359 | 0.349 | 0.362 |
| 9 | 788 | 0.409 | 0.406 | 0.230 | 0.238 |
| 10 | 772 | 0.455 | 0.453 | 0.114 | 0.118 |
| 11 | 715 | 0.500 | 0.500 | 0.000 | 0.000 |
| 12 | 705 | 0.545 | 0.547 | -0.114 | -0.118 |
| 13 | 695 | 0.591 | 0.594 | -0.230 | -0.238 |
| 14 | 674 | 0.636 | 0.641 | -0.349 | -0.362 |
| 15 | 588 | 0.682 | 0.688 | -0.473 | -0.491 |
| 16 | 580 | 0.727 | 0.735 | -0.605 | -0.629 |
| 17 | 552 | 0.773 | 0.782 | -0.748 | -0.780 |
| 18 | 529 | 0.818 | 0.829 | -0.908 | -0.952 |
| 19 | 469 | 0.864 | 0.876 | -1.097 | -1.158 |
| 20 | 462 | 0.909 | 0.924 | -1.335 | -1.429 |
| $n=21$ | 217 | 0.955 | 0.971 | -1.691 | -1.890 |

5.4 Επιλογή και προσαρμογή θεωρητικής συνάρτησης κατανομής

Αφού εκτελέσουμε τα προηγούμενα, τα οποία αφορούν στη συμπύκνωση του δείγματος, είμαστε σε θέση να κάνουμε και κάποιες στατιστικές εκτιμήσεις ή προγνώσεις. Για παράδειγμα από ένα δείγμα ετήσιων παροχών μήκους 20 ετών μπορούμε να βρούμε τη μέση υπερετήσια παροχή (ίση με το μέσο όρο του δείγματος) ή την παροχή περιόδου επαναφοράς 10

ετών (περίπου ίση με τη δεύτερη σε μέγεθος τιμή του δείγματος), κτλ. Ωστόσο η εμβέλεια αυτών των εκτιμήσεων ή προγνώσεων είναι περιορισμένη. Έτσι δεν μπορούμε, στο παράδειγμά μας, να εκτιμήσουμε την παροχή περιόδου επαναφοράς 100 ετών, αφού δεν υπάρχει τρόπος επέκτασης έξω από τα όρια των 20 ετών που καλύπτει το δείγμα. Επίσης δεν μπορούμε να κάνουμε εκτιμήσεις διαστήματος για τα μεγέθη που μας ενδιαφέρουν.

Τα παραπάνω προβλήματα δεν θα υπήρχαν αν ξέραμε τη συνάρτηση κατανομής του πληθυσμού. Γενικά, η κατανομή του πληθυσμού μπορεί να είναι οποιαδήποτε συνάρτηση που υπακούει στους περιορισμούς (2.4). Η ακριβής γνώση της συνάρτησης κατανομής του πληθυσμού θα απαιτούσε να έχουμε μετρήσει τον ίδιο τον πληθυσμό στο σύνολό του, ή πρακτικώς να έχουμε ένα δείγμα πολύ μεγαλύτερο από τη χρονική εμβέλεια των εκτιμήσεων που πρόκειται να κάνουμε. Βεβαίως αυτό δεν μπορεί ποτέ να πραγματοποιηθεί, και έτσι η λύση που μένει είναι να υποθέσουμε ένα πιθανοτικό μοντέλο για τον πληθυσμό. Με τον όρο πιθανοτικό μοντέλο εννοούμε μια από τις τυπικές συναρτήσεις κατανομής της θεωρίας πιθανοτήτων που έχουν ένα συγκεκριμένο και σχετικώς απλό μαθηματικό τύπο. Το πιο τυπικό παράδειγμα είναι η κανονική κατανομή με συνάρτηση που δίνεται από τη (2.61). Άλλα παραδείγματα θα δοθούν στο κεφάλαιο 6. Βεβαίως, η χρήση ενός πιθανοτικού μοντέλου αποτελεί πάντα μια προσέγγιση της πραγματικότητας. Οι κατανομές των υδρολογικών μεταβλητών ποτέ δεν ταυτίζονται με τις απλές κατανομές της θεωρίας πιθανοτήτων.

Στην επιλογή του κατάλληλου μοντέλου μας οδηγούν τα ακόλουθα:

1. *Η πιθανοθεωρία.* Σε μερικές περιπτώσεις υπάρχουν θεωρητικοί λόγοι για τους οποίους μια συγκεκριμένη υδρολογική μεταβλητή αναμένεται να ακολουθεί συγκεκριμένο τύπο κατανομής. Για παράδειγμα, σύμφωνα με το κεντρικό οριακό θεώρημα, η ετήσια βροχόπτωση μιας περιοχής με πλούσιες βροχοπτώσεις αναμένεται να ακολουθεί κανονική κατανομή (βλ. και εδάφιο 2.8.1).
2. *Η γενική υδρολογική εμπειρία.* Σε πολλές περιπτώσεις η υδρολογική εμπειρία έχει δείξει ότι συγκεκριμένες υδρολογικές μεταβλητές ακολουθούν συγκεκριμένους τύπους κατανομής, αν και δεν υπάρχουν σαφείς πιθανοθεωρητικοί λόγοι που να οδηγούν στις τελευταίες. Για

παράδειγμα, οι μηνιαίες απορροές πολύ συχνά ακολουθούν κατανομές γάμα ή λογαριθμοκανονικές (βλ. κεφάλαιο 6).

3. *Η εξέταση του συγκεκριμένου δείγματος.* Τα στατιστικά χαρακτηριστικά του διαθέσιμου ιστορικού δείγματος μας βοηθούν στην επιλογή ή την απόρριψη συγκεκριμένου τύπου κατανομής. Για παράδειγμα, αν ο δειγματικός συντελεστής ασυμμετρίας έχει τιμή κοντά στο μηδέν, τότε μπορεί να επιλεγεί η κανονική (ή κάποια άλλη συμμετρική) κατανομή. Αντίθετα, αν ο συντελεστής ασυμμετρίας διαφέρει αρκετά από το μηδέν, θα πρέπει να αποκλειστεί η χρήση της κανονικής κατανομής.

Βεβαίως, η καταλληλότητα ενός συγκεκριμένου τύπου κατανομής δεν εξασφαλίζεται με τα παραπάνω κριτήρια, τα οποία αποτελούν απλώς ενδείξεις καταλληλότητας. Ο ουσιαστικός έλεγχος της καταλληλότητας της κατανομής γίνεται εκ των υστέρων. Αφού υπολογιστούν και οι παράμετροί της, ελέγχεται ο βαθμός προσαρμογής της στην εμπειρική συνάρτηση κατανομής του δείγματος. Αυτό μπορεί να γίνει κατ' αρχήν εμπειρικά, με βάση τη γραφική απεικόνιση της εμπειρικής και θεωρητικής συνάρτησης κατανομής στο κατάλληλο χαρτί πιθανότητας. Όμως, τα πλέον αξιόπιστα συμπεράσματα επιτυγχάνονται με τη βοήθεια των τυπικών στατιστικών ελέγχων, όπως διεξοδικά αναλύεται στην επόμενη ενότητα.

Ενδείξεις καταλληλότητας της κανονικής κατανομής για υδρολογικές μεταβλητές

Μέχρι τώρα έχουμε αναφερθεί αρκετές φορές σε ενδεικτικούς ελέγχους της καταλληλότητας της κανονικής κατανομής για την περιγραφή μιας υδρολογικής μεταβλητής. Παρακάτω κωδικοποιούμε όλες τις ενδείξεις καταλληλότητας της κανονικής κατανομής.

1. *Αξιοποίηση κεντρικού οριακού θεωρήματος.* Ελέγχεται κατά πόσον η υπό έλεγχο μεταβλητή μπορεί να θεωρηθεί ως άθροισμα πολλών επιμέρους συνιστωσών, οι οποίες θα πρέπει να εκπληρώνουν (έστω προσεγγιστικά) τις προϋποθέσεις του κεντρικού οριακού θεωρήματος. Το κριτήριο αυτό είναι θεωρητικό και δεν απαιτεί αριθμητικούς υπολογισμούς
2. *Αριθμητικός έλεγχος με βάση το συντελεστή ασυμμετρίας.* Αν ο δειγματικός συντελεστής ασυμμετρίας είναι περίπου μηδενικός,

αυτό αποτελεί μια σοβαρή ένδειξη για την καταλληλότητα της κανονικής κατανομής.

3. *Αριθμητικός έλεγχος με βάση το συντελεστή μεταβλητότητας.* Έστω X μια τυχούσα υδρολογική μεταβλητή. Κατά κανόνα η X παίρνει μόνο θετικές τιμές ή μηδέν, γιατί οι αρνητικές τιμές δεν έχουν φυσικό νόημα. Αντίθετα, η κανονική κατανομή προβλέπει ότι η X μπορεί να πάρει και αρνητικές τιμές. Θεωρητικά, λοιπόν, η κανονική κατανομή δεν μπορεί να αντιπροσωπεύει υδρολογικές μεταβλητές, παρά μόνο προσεγγιστικά. Για να είναι ικανοποιητική η προσέγγιση θα πρέπει η πιθανότητα $P(X < 0)$ να είναι πολύ μικρή, ώστε να μπορεί να αγνοηθεί, ήτοι $P(X < 0) \leq q$ όπου q μια αποδεκτά μικρή τιμή πιθανότητας, π.χ. $q < 0.02$. Αν $Z = (X - \mu_X) / \sigma_X$ είναι η αντίστοιχη τυποποιημένη κανονική μεταβλητή τότε θα πρέπει $P(Z < -\mu_X / \sigma_X) \leq q$. Αν z_q είναι το q -ποσοστημόριο της τυποποιημένης κανονικής κατανομής, τότε ισοδύναμα $C_{vX} = \sigma_X / \mu_X \leq -1/z_q$. Για $q = 0.02$ έχουμε $z_q \approx -2$ οπότε $C_{vX} \leq 0.5$. Αντίστοιχα, για $q = 0.00005$ έχουμε $z_q \approx -4$ οπότε $C_{vX} \leq 0.25$. Καταλήγουμε λοιπόν στο συμπέρασμα ότι αν $C_{vX} \leq 0.25$ έχουμε μια πολύ σοβαρή ένδειξη καταλληλότητας της κανονικής κατανομής. Αν $C_{vX} > 0.5$ τότε θα πρέπει να αποκλειστεί η χρήση της κανονικής κατανομής. Για ενδιάμεσες τιμές του συντελεστή μπορεί να είναι αποδεκτή η κανονική κατανομή αλλά με μικρότερο βαθμό προσέγγισης.
4. *Γραφικός έλεγχος με βάση τη συνοπτική παράσταση δείγματος.* Όπως αναφέρθηκε στο εδάφιο 5.1 τυχόν συμμετρικό διάγραμμα της συνοπτική παράστασης δείγματος, χωρίς αδικαιολόγητα μεγάλο αριθμό εξωτερικών ή μακρινών τιμών, αποτελεί ένδειξη καταλληλότητας της κανονικής κατανομής
5. *Γραφικός έλεγχος με βάση την εμπειρική συνάρτηση κατανομής.* Η ευθύγραμμη διάταξη της σημειοσειράς της εμπειρικής συνάρτησης κατανομής, σχεδιασμένης σε χαρτί κανονικής κατανομής, αποτελεί σοβαρή ένδειξη της καταλληλότητας της κανονικής κατανομής.

Τονίζουμε και πάλι ότι τα κριτήρια αυτά είναι απλές ενδείξεις και δεν μπορούν να θεωρηθούν ως στατιστική απόδειξη της καταλληλότητας της κανονικής κατανομής.

Εφαρμογή 5.4

Να προσαρμοστεί η κανονική κατανομή στο δείγμα της Εφαρμογής 5.1.

Η λύση αυτού του προβλήματος είναι πολύ απλή. Οι παράμετροι της κατανομής είναι (βλ. Εφαρμογή 5.1) $\mu = \bar{x} = 725.0$, $\sigma = s_X = 211.5$ (αποδεκτή είναι και η τιμή $\sigma = s_X^* = 216.7$). Η κανονική κατανομή με αυτές τις παραμέτρους έχει απεικονιστεί στο Σχ. 5.3 και η αντίστοιχη πυκνότητα πιθανότητας στο Σχ. 5.2.

Ο αναγνώστης μπορεί να διαπιστώσει ότι στο παράδειγμα που εξετάζουμε υπάρχουν όλες οι ενδείξεις καταλληλότητας της κανονικής κατανομής που αναφέρθηκαν στο παραπάνω ένθετο εδάφιο. Στην Εφαρμογή 5.5 θα δώσουμε τον πληρέστερο στατιστικό έλεγχο της καταλληλότητας της κανονικής κατανομής.

5.5 Έλεγχος προσαρμογής συνάρτησης κατανομής

Αφού υιοθετήσουμε μια συγκεκριμένη συνάρτηση κατανομής για την περιγραφή μιας φυσικής μεταβλητής και υπολογίσουμε τις παραμέτρους της, ελέγχουμε στη συνέχεια την προσαρμογή της κατανομής αυτής στο συγκεκριμένο δείγμα που διαθέτουμε. Ο έλεγχος βασίζεται στη στατιστική θεωρία ελέγχου υποθέσεων, την οποία περιγράψαμε πολύ συνοπτικά στην ενότητα 3.5. Έχουν αναπτυχθεί διάφοροι τύποι στατιστικών δοκιμών που μπορούν να εφαρμοστούν στον έλεγχο προσαρμογής μιας κατανομής. Θα αναφερθούμε στην πιο κλασική από αυτές, τη δοκιμή χ^2 . Άλλες στατιστικές δοκιμές με σημαντική εφαρμογή στην τεχνική υδρολογία είναι η δοκιμή *Kolmogorov-Smirnov* (βλ. π.χ. Benjamin and Cornell, 1970, σ. 466· Kottegod, 1980, σ. 89) και η πιο πρόσφατη δοκιμή του συντελεστή συσχέτισης εμπειρικής πιθανότητας (βλ. π.χ. Stedinger et al., 1993, σ. 18.27).

5.5.1 Η δοκιμή χ^2 .

Η δοκιμή χ^2 στηρίζεται στη σύγκριση της θεωρητικής συνάρτησης κατανομής με την κλασική εμπειρική συνάρτηση κατανομής. Η σύγκριση γίνεται σε ένα πεπερασμένο σύνολο επιλεγμένων σημείων x_j , και όχι σε κάθε παρατηρημένη τιμή x_i του δείγματος. Η μηδενική υπόθεση H_0 και η εναλλακτική H_1 είναι

$$H_0: F(x_j) = F_0(x_j) \text{ για κάθε } j \quad H_1: F(x_j) \neq F_0(x_j) \text{ για κάποια } j \quad (5.11)$$

όπου $F(x)$ η άγνωστη πραγματική συνάρτηση κατανομής και $F_0(x)$ η κατανομή που έχει υποτεθεί. Η $F_0(x)$ μπορεί να είναι γνωστή πλήρως, και ως προς τη μαθηματική της μορφή και ως προς τις τιμές των παραμέτρων της, πριν από την εξέταση του συγκεκριμένου δείγματος. Στην περίπτωση αυτή η μηδενική υπόθεση λέγεται *τέλεια*. Συνηθέστερα όμως, οι τιμές των παραμέτρων υπολογίζονται από το δείγμα, οπότε μιλούμε για *ατελή μηδενική υπόθεση*.

Τα σημεία ελέγχου $x_j, j = 0, \dots, k$ χωρίζουν το πεδίο μεταβολής της τυχαίας μεταβλητής σε k κλάσεις, δηλαδή διαστήματα της μορφής $(x_0, x_1], (x_1, x_2], \dots, (x_{k-1}, x_k]$. Για τη δεδομένη συνάρτηση κατανομής $F_0(x)$ η πιθανότητα να βρεθεί ένα τυχαία επιλεγμένο σημείο στην κλάση $K_j = (x_{j-1}, x_j]$ προφανώς είναι

$$p_j = F_0(x_j) - F_0(x_{j-1}) \quad (5.12)$$

και κατά συνέπεια το αναμενόμενο πλήθος σημείων του δείγματος που θα βρίσκονται στην εν λόγω κλάση είναι $l_j = n p_j$, όπου n το μέγεθος του δείγματος. Το μέγεθος l_j λέγεται θεωρητικό δυναμικό της κλάσης, ενώ το πραγματικό πλήθος των σημείων του δείγματος, n_j , που βρίσκονται μέσα στα όρια της κλάσης, λέγεται πραγματικό δυναμικό της κλάσης. Προφανώς, η μικρή απόσταση μεταξύ των μεγεθών n_j και l_j , δηλαδή η μικρή διαφορά $|n_j - n p_j|$, αποτελεί συνηγορία υπέρ της καταλληλότητας της κατανομής $F_0(x_j)$ και της μη απόρριψης της μηδενικής υπόθεσης. Η *στατιστική συνάρτηση του Pearson* που ορίζεται από την εξίσωση

$$q := \sum_{j=1}^k \frac{(n_j - n p_j)^2}{n p_j} \quad (5.13)$$

συνεκτιμά τις διαφορές πραγματικού και θεωρητικού δυναμικού σε όλες τις κλάσεις. Αν η μηδενική υπόθεση είναι τέλεια, τότε η κατανομή της είναι η χ^2 με $k - 1$ βαθμούς ελευθερίας. Στη συνηθέστερη περίπτωση της ατελούς μηδενικής υπόθεσης, ο αριθμός των βαθμών ελευθερίας είναι

$k - r - 1$, όπου r είναι ο αριθμός των παραμέτρων που εκτιμώνται από το δείγμα.*

Η παραλλαγή της δοκιμής χ^2 που αναπτύσσουμε εδώ είναι αυτή των *ισοπίθανων κλάσεων*, σύμφωνα με την οποία το θεωρητικό δυναμικό είναι ίδιο σε όλες τις κλάσεις. Σε αυτή την περίπτωση η εξίσωση (5.13) απλοποιείται σε

$$q := \frac{k}{n} \sum_{j=1}^k n_j^2 - n \quad (5.14)$$

Η έκδοση αυτή έχει το πλεονέκτημα ότι δίνει μονοσήμαντα καθορισμένα όρια κλάσεων για δεδομένο αριθμό κλάσεων k . Για την επιλογή του αριθμού των κλάσεων k ισχύουν οι ακόλουθοι δύο αλληλοαντικρουόμενοι κανόνες:

- Υποχρεωτικά θα πρέπει $k \geq r + 2$, όπου r είναι ο αριθμός των παραμέτρων της κατανομής που εκτιμώνται από το δείγμα.
- Γενικά συνιστάται (βλ. π.χ. Benjamin and Cornell, 1970, σ. 465· Kottegoda, 1980, σ. 88) το θεωρητικό δυναμικό των κλάσεων είναι μεγαλύτερο από 5, πράγμα που συνεπάγεται ότι $k \leq n / 5$.

Για μικρά δείγματα είναι δυνατό οι κανόνες αυτοί να μη μπορούν να ικανοποιηθούν ταυτόχρονα, οπότε ικανοποιούμε μόνο τον πρώτο.†

Ο αλγόριθμος εφαρμογής της δοκιμής χ^2 περιγράφεται από τα ακόλουθα βήματα:

1. Επιλέγουμε τον αριθμό των κλάσεων k σύμφωνα με τους παραπάνω κανόνες.

* Κανονικά, στην περίπτωση αυτή οι εκτιμήσεις των παραμέτρων πρέπει να γίνονται με τη μέθοδο της μέγιστης πιθανοφάνειας· ωστόσο, ο όρος αυτός δεν τηρείται στις εφαρμογές.

† Η επιλογή του κατάλληλου αριθμού κλάσεων γίνεται με τον τύπο (Mann and Wald, 1942· Williams, 1955· βλ. και Kottegoda, 1980, σ. 88):

$$k = 2^{1.2} [(n - 1) / z_{1-\alpha}]^{0.4}$$

όπου $z_{1-\alpha}$ το $(1 - \alpha)$ -ποσοστημόριο της κανονικής κατανομής και α το επίπεδο σημαντικότητας του ελέγχου. Οι Kendall and Stuart (1973, σ. 455) δίνουν αναλυτικότερο τρόπο υπολογισμού του αριθμού κλάσεων, ο οποίος όμως ισχύει για μεγάλα δείγματα που σπάνια είναι διαθέσιμα στην τεχνική υδρολογία.

2. Χωρίζουμε το διάστημα πιθανότητας $[0, 1]$ σε k ίσα υποδιαστήματα με όρια $u_j = j / k$, ($j = 0, \dots, k$).
3. Υπολογίζουμε τα όρια x_j των κλάσεων (η τιμή x_j είναι το u_j -ποσοστημόριο της μεταβλητής).
4. Υπολογίζουμε το πραγματικό δυναμικό n_j της κάθε κλάσης (το βήμα αυτό απλοποιείται αν το δείγμα έχει προηγουμένως καταταγεί σε φθίνουσα ή αύξουσα σειρά).
5. Από την (5.14) (ή την (5.13)) υπολογίζουμε την τιμή της στατιστικής συνάρτησης q του Pearson.
6. Για ένα αποδεκτό επίπεδο σημαντικότητας α υπολογίζουμε την κρίσιμη τιμή της στατιστικής συνάρτησης ελέγχου $q_c = q_{1-\alpha}$. Για το σκοπό αυτό χρησιμοποιούμε την κατανομή χ^2 για $k - r - 1$ βαθμούς ελευθερίας όπου r είναι ο αριθμός των παραμέτρων της κατανομής που εκτιμώνται από το δείγμα (βλ. Πίν. Π2 στο Παράρτημα).
7. Απορρίπτουμε τη μηδενική υπόθεση αν $q > q_c$.

Ο αλγόριθμος διευκρινίζεται στο παράδειγμα που ακολουθεί.

Εφαρμογή 5.5

Να ελεγχθεί η καταλληλότητα της κανονικής κατανομής για το δείγμα της Εφαρμογής 5.1.

Στο συγκεκριμένο δείγμα προσαρμόστηκε ήδη (βλ. Εφαρμογή 5.4) η κανονική κατανομή με παραμέτρους $\mu = \bar{x} = 725.0$, $\sigma = s_X = 211.5$.

Ο αριθμός των παραμέτρων της κατανομής είναι $r = 2$ και το μέγεθος του δείγματος $n = 21$. Σύμφωνα με τα παραπάνω, ο αριθμός των κλάσεων k για τη δοκιμή χ^2 θα πρέπει να ικανοποιεί τις σχέσεις

$$k \geq 2 + 2 = 4, \quad k \leq 21 / 5 = 4.2$$

οι οποίες συναληθεύουν για $k = 4$. Κατά συνέπεια επιλέγουμε $k = 4$.

Οι υπολογισμοί των βημάτων 2-4 του παραπάνω αλγορίθμου κωδικοποιούνται στον Πίν. 5.10. Για τον υπολογισμό των ορίων της μεταβλητής χρησιμοποιήσαμε τον Πίν. Π1 (βλ. Παράρτημα) της κανονικής κατανομής. Για παράδειγμα, το άνω όριο της πρώτης κλάσης είναι

$$x_1 = 725.0 - 0.675 \times 211.5 = 528.3$$

Στο Σχ. 5.4 δίνουμε για διδακτικούς λόγους γραφική απεικόνιση των κλάσεων και του πραγματικού δυναμικού τους πάνω σε χαρτί κανονικής κατανομής.

Πίν. 5.10 Βασικοί υπολογισμοί για τη δοκιμή χ^2 της Εφαρμογής 5.5.

| Κλάση | 1 | 2 | 3 | 4 | |
|---------------------|-----------|-------|-------|-------|-----------|
| Όρια πιθανότητας | 0 | 0.25 | 0.5 | 0.75 | 1.0 |
| Όρια μεταβλητής | $-\infty$ | 582.3 | 725.0 | 867.7 | $+\infty$ |
| Πραγματικό δυναμικό | 6 | 5 | 4 | 6 | |

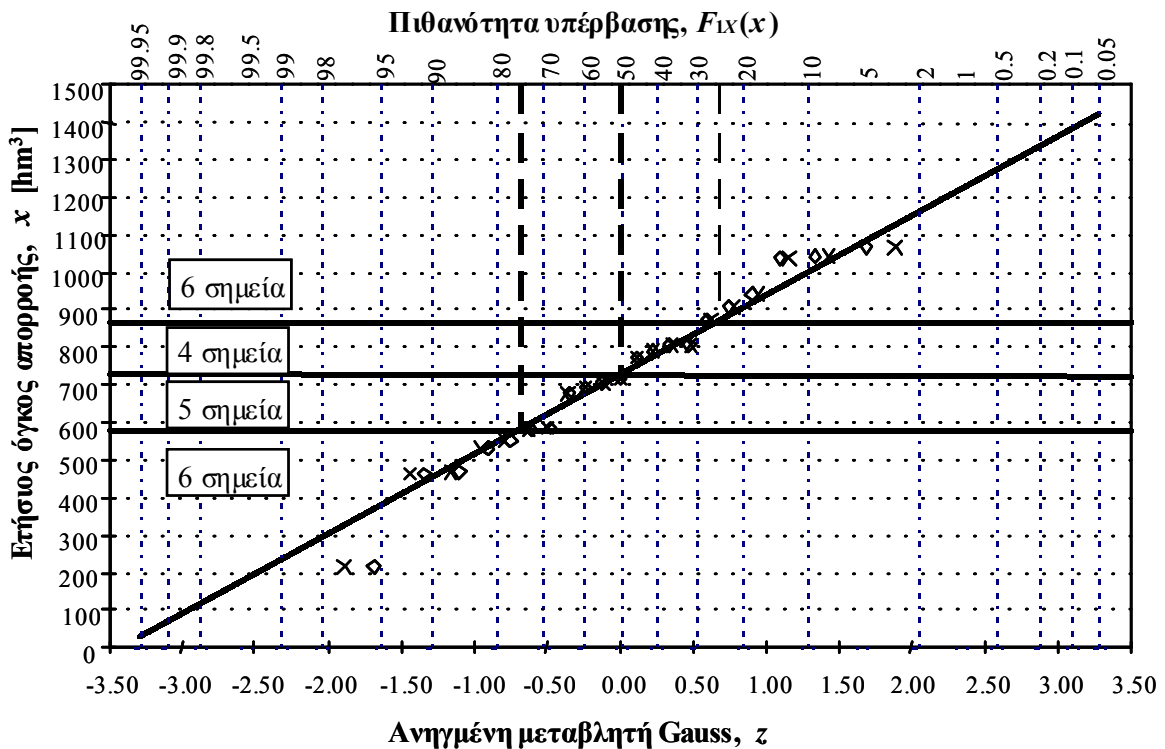
Από την (5.14) παίρνουμε

$$q = (4/21) \times (6^2 + 5^2 + 4^2 + 6^2) - 21 = 0.52$$

Εξ άλλου για επίπεδο σημαντικότητας $\alpha = 0.05$ η κρίσιμη τιμή της μεταβλητής είναι

$$q_c = \chi_{0.95}^2(1) = 3.84$$

(όπως προκύπτει από τον Πίν. Π2 για $u = 1 - \alpha = 0.95$ και βαθμούς ελευθερίας $= 4 - 2 - 1 = 1$). Κατά συνέπεια ισχύει $q < q_c$ και δεν απορρίπτεται η κανονική κατανομή.



Σχ. 5.4 Επεξηγηματικό σκαρίφημα για το παράδειγμα της Εφαρμογής 5.5.

5.6 Στατιστική πρόγνωση

Η στατιστική πρόγνωση στην τεχνική υδρολογία αφορά στην εκτίμηση της τιμής ενός υδρολογικού μεγέθους για δεδομένη πιθανότητα υπέρβασης (ή περίοδο επαναφοράς). Η πρόγνωση αυτή είναι αρκετά εύκολη από υπολογιστική άποψη, εφόσον έχει ήδη υιοθετηθεί και προσαρμοστεί ένα συγκεκριμένο πιθανοτικό μοντέλο για το μέγεθος που ενδιαφέρει. Για το σκοπό αυτό εφαρμόζονται όσα έχουν αναφερθεί στο κεφάλαιο 3. Ειδικότερα, η πρόγνωση μπορεί να είναι είτε σημειακή, είτε διαστήματος, όπως αναλυτικά φαίνεται στο παράδειγμα που ακολουθεί.

Εφαρμογή 5.6

Να υπολογιστεί ο μέγιστος και ο ελάχιστος ετήσιος όγκος απορροής εκατονταετίας στη λεκάνη του ποταμού Ευήνου ανάντη του υδρομετρικού σταθμού Πόρος Ρηγανίου, καθώς και τα όρια εμπιστοσύνης του 95% (βλ. και προηγούμενες Εφαρμογές αυτού του κεφαλαίου).

Εφαρμόζουμε την ίδια διαδικασία που είχαμε εφαρμόσει στην Εφαρμογή 3.3.4. Σημειώνουμε ότι το μέγεθος του δείγματος είναι πολύ μικρό σε σχέση με τον υπολογιστικό ορίζοντα της εκατονταετίας και έτσι δεν περιμένουμε πολύ καλό βαθμό ακρίβειας στους υπολογισμούς μας.

Θα υπολογίσουμε πρώτα τις σημειακές εκτιμήσεις. Για τον μέγιστο όγκο απορροής εκατονταετίας η πιθανότητα μη υπέρβασης είναι $u = 1 - 1/100 = 0.99$ και $z_u = 2.326$ (από τον Πίν. Π1 της κανονικής κατανομής του Παραρτήματος). Άρα η σημειακή εκτίμηση είναι

$$x_u = 725.0 + 2.326 \times 211.5 = 1216.9$$

Αντίστοιχα, για τον ελάχιστο όγκο απορροής εκατονταετίας η πιθανότητα μη υπέρβασης είναι $u = 1 / 100 = 0.01$ και $z_u = -2.326$, άρα

$$x_u = 725.0 - 2.326 \times 211.5 = 233.1$$

Προχωρούμε τώρα στον υπολογισμό των ορίων εμπιστοσύνης. Για $\gamma = 95\%$ και $z_{(1+\gamma)/2} = 1.96$, τα όρια για τον μέγιστο όγκο απορροής εκατονταετίας θα είναι:

$$\hat{x}_{u1} = 1216.9 - 1.96 \sqrt{1 + \frac{2.326^2}{2}} \frac{211.5}{\sqrt{21}} = 1042.8$$

$$\hat{x}_{u2} = 1216.9 + 1.96 \sqrt{1 + \frac{2.326^2}{2}} \frac{211.5}{\sqrt{21}} = 1391.0$$

Αντίστοιχα, τα όρια για τον ελάχιστο όγκο απορροής εκατονταετίας θα είναι:

$$\hat{x}_{u1} = 233.1 - 1.96 \sqrt{1 + \frac{2.326^2}{2}} \frac{211.5}{\sqrt{21}} = 59.0$$

$$\hat{x}_{u2} = 233.1 + 1.96 \sqrt{1 + \frac{2.326^2}{2}} \frac{211.5}{\sqrt{21}} = 407.2$$

Παρατηρούμε ότι ο μέγιστος και ο ελάχιστος όγκος απορροής που έχουν καταγραφεί στο δείγμα (Πίν. 5.5) βρίσκονται ανάμεσα στα αντίστοιχα όρια εμπιστοσύνης του μέγιστου και ελάχιστου όγκου εκατονταετίας. Ειδικότερα, ο ελάχιστος όγκος απορροής των 217 hm³ είναι μικρότερος από τη σημειακή πρόγνωση εκατονταετίας που δίνει η κανονική κατανομή (233.1 hm³), γεγονός που αντανakλά το σημαντικό μέγεθος της ξηρασίας του υδρολογικού έτους 1989-90.

Στο Σχ. 5.3 έχουμε απεικονίσει τα πιο πάνω αποτελέσματα, καθώς και μια σειρά σημειακών εκτιμήσεων και ορίων εμπιστοσύνης για άλλες τιμές της περιόδου επαναφοράς. Οι υπολογισμοί έγιναν με τον ίδιο τρόπο όπως παραπάνω. Συνδέοντας τις τιμές των ορίων εμπιστοσύνης στο διάγραμμα αποκτούμε τις καμπύλες εμπιστοσύνης 95% της κατανομής.

Κεφάλαιο 6 Τυπικές συναρτήσεις κατανομής στην τεχνική υδρολογία

Στο κεφάλαιο αυτό περιγράφουμε τις τρεις βασικές οικογένειες συναρτήσεων κατανομής που χρησιμοποιούνται στην τεχνική υδρολογία. Η πρώτη περιλαμβάνει την κανονική κατανομή και τις παράγωγες κατανομές που προκύπτουν από λογαριθμικό μετασχηματισμό της. Η δεύτερη είναι η οικογένεια κατανομών γάμα που περιλαμβάνει την εκθετική κατανομή, τις κατανομές γάμα δύο και τριών παραμέτρων, και την παράγωγη κατανομή Log-Pearson III που προκύπτει από λογαριθμικό μετασχηματισμό της τελευταίας. Η τρίτη οικογένεια περιλαμβάνει τις κατανομές ακροτάτων με τυπικές εκπροσώπους τις κατανομές Gumbel και Weibull. Στο τέλος του κεφαλαίου δίνονται ορισμένες άλλες κατανομές με δευτερεύουσα σημασία στην τεχνική υδρολογία, όπως οι κατανομές βήτα και Pareto.

6.1 Κανονική κατανομή και μετασχηματισμοί της

6.1.1 Κανονική κατανομή

Μέχρι τώρα έχουμε αναφερθεί αναλυτικά και σε έκταση στην κανονική κατανομή και στη χρήση της στη στατιστική αλλά και στην τεχνική υδρολογία. Συγκεκριμένα, η κανονική κατανομή έχει εισαχθεί, ως συνέπεια του κεντρικού οριακού θεωρήματος, στην ενότητα 2.8 μαζί με δύο στενά συνδεδεμένες κατανομές, τις χ^2 και Student (ή t), οι οποίες, αν και

δεν χρησιμοποιούνται στην περιγραφή υδρολογικών μεταβλητών, είναι βασικής σημασίας για τις στατιστικές εκτιμήσεις. Σε πολλά σημεία του κεφαλαίου 3 έχει χρησιμοποιηθεί η κανονική κατανομή στην θεωρητική εξαγωγή στατιστικών εκτιμήσεων. Τέλος, στο κεφάλαιο 5 έχει περιγραφεί σε λεπτομέρεια η χρήση της κανονικής κατανομής για την περιγραφή υδρολογικών μεταβλητών.

Στο τυπολόγιο του Πίν. 6.1 συνοψίζονται οι βασικότερες μαθηματικές σχέσεις που συνδέονται με την κανονική κατανομή, οι οποίες έχουν αναλυθεί και σε προηγούμενες ενότητες.

Πίν. 6.1 Τυπολόγιο κανονικής κατανομής.

| | |
|----------------------------------|--|
| Συνάρτηση πυκνότητας πιθανότητας | $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ |
| Συνάρτηση κατανομής | $F_X(x) = \int_{-\infty}^x f_X(s)ds$ |
| Τιμές μεταβλητής | $-\infty < x < \infty$ (συνεχής) |
| Παράμετροι | μ : παράμετρος θέσης (= μέση τιμή) $\sigma > 0$: παράμετρος κλίμακας (= τυπική απόκλιση) |
| Μέση τιμή | $\mu_X = \mu$ |
| Διασπορά | $\sigma_X^2 = \sigma^2$ |
| Τρίτη κεντρική ροπή | $\mu_X^{(3)} = 0$ |
| Τέταρτη κεντρική ροπή | $\mu_X^{(4)} = 3\sigma^4$ |
| Συντελεστής ασυμμετρίας | $C_{s_X} = 0$ |
| Συντελεστής κύρτωσης | $C_{k_X} = 3$ |
| Πιθανότερη τιμή | $x_p = \mu$ |
| Διάμεσος | $x_{0.5} = \mu$ |

Συμπερασματικά, πρόκειται για συμμετρική κατανομή δύο παραμέτρων με κωδωνοειδές σχήμα. Η συμμετρία της κατανομής και η δυνατότητα της μεταβλητής X να παίρνει αρνητικές τιμές έρχονται σε αντίθεση με τη φυσική απαίτηση, σύμφωνα με την οποία οι υδρολογικές μεταβλη-

τές κατά κανόνα έχουν αποκλειστικά θετικές τιμές. Το πρόβλημα αυτό έχει ήδη συζητηθεί διεξοδικά στο ένθετο εδάφιο της ενότητας 5.4.

Μια βασική ιδιότητα της κατανομής είναι ότι είναι κλειστή ως προς την πρόσθεση. Έτσι, το άθροισμα (και γενικότερα ο γραμμικός συνδυασμός) δύο μεταβλητών που ακολουθούν κανονική κατανομή, ακολουθεί επίσης κανονική κατανομή.

Τυπικοί υπολογισμοί

Οι τυπικοί υπολογισμοί αφορούν στον υπολογισμό της τιμής $u = F_X(x_u)$ της συνάρτησης κατανομής για δεδομένη τιμή x_u της μεταβλητής, ή, αντίστροφα, στον υπολογισμό του u -ποσοστημορίου της μεταβλητής, δηλαδή τον υπολογισμό της τιμής x_u όταν είναι γνωστή η τιμή u . Το γεγονός ότι το ολοκλήρωμα που υπεισέρχεται στον ορισμό της κανονικής συνάρτησης κατανομής (Πίν. 6.1) δεν υπολογίζεται αναλυτικά, δημιουργεί ορισμένες δυσκολίες στους τυπικούς υπολογισμούς. Η απλούστερη λύση στηρίζεται στην πινακοποίηση των τιμών της τυποποιημένης κανονικής μεταβλητής $z = (x - \mu) / \sigma$. Η τελευταία ακολουθεί κανονική κατανομή με μέση τιμή 0 και τυπική απόκλιση 1 (βλ. ενότητα 2.5), πράγμα που διευκολύνει την πινακοποίηση. Η πινακοποίηση αυτή περιέχεται σε όλα τα βιβλία θεωρίας πιθανοτήτων και στατιστικής, καθώς και στο Πίν. Π1 του Παραρτήματος (στο τέλος του βιβλίου).

Έτσι, ο υπολογισμός του u -ποσοστημορίου (x_u) γίνεται από την εξίσωση

$$x_u = \mu + z_u \sigma \quad (6.1)$$

αφού προηγουμένως προσδιοριστεί από τον Πίν. Π1 η τιμή z_u που αντιστοιχεί στη δεδομένη τιμή της συνάρτησης κατανομής $u = F_Z(z_u)$. Αντίστροφα, για δεδομένο x_u υπολογίζεται από την (6.1) το z_u και από τον Πίν. Π1 προκύπτει το $u = F_Z(z_u)$.

Στη βιβλιογραφία δίνονται διάφορες αριθμητικές προσεγγίσεις της κανονικής συνάρτησης κατανομής, με βάση τις οποίες μπορούμε να αποφύγουμε τη χρήση πινάκων (Press et al., 1987· Stedinger et al., 1993). Η χρήση των προσεγγίσεων αυτών είναι πλεονεκτικότερη όταν οι υπολογισμοί γίνονται με ηλεκτρονικούς υπολογιστές ή αριθμομηχανές. Στο Παράρτημα 6.A δίνονται οι απλούστερες προσεγγίσεις.

Εκτίμηση παραμέτρων

Όπως έχουμε δει στην ενότητα 3.4, και η μέθοδος των ροπών και η μέθοδος της μέγιστης πιθανοφάνειας δίνουν τις ίδιες εκτιμήσεις για τις παραμέτρους της κανονικής κατανομής, οι οποίες είναι

$$\mu = \bar{x} \quad \sigma = s_X \quad (6.2)$$

Υπενθυμίζεται ότι το μέγεθος s_X στη σχέση (6.2) κανονικά είναι η μεροληπτική εκτίμηση της τυπικής απόκλισης. Ωστόσο συχνά χρησιμοποιείται η αμερόληπτη εκτίμηση αντί της μεροληπτικής.

Τυπικό σφάλμα ποσοστημορίου και όρια εμπιστοσύνης

Στο εδάφιο 3.3.4 ορίσαμε την έννοια του τυπικού σφάλματος και των ορίων εμπιστοσύνης στην εκτίμηση του ποσοστημορίου, και δώσαμε τις αντίστοιχες εξισώσεις για το ποσοστημόριο της κανονικής κατανομής. Ανακεφαλαιώνοντας και συνοψίζοντας, αναφέρουμε ότι η σημειακή εκτίμηση του u -ποσοστημορίου της κανονικής κατανομής είναι

$$\hat{x}_u = \bar{x} + z_u s_X \quad (6.3)$$

το τυπικό σφάλμα της εκτίμησης είναι

$$\varepsilon_u = \frac{s_X}{\sqrt{n}} \sqrt{1 + \frac{z_u^2}{2}} \quad (6.4)$$

και τα αντίστοιχα όρια εμπιστοσύνης για βαθμό εμπιστοσύνης γ είναι

$$\hat{x}_{u_{1,2}} \approx (\bar{x} + z_u s_X) \pm z_{(1+\gamma)/2} \frac{s_X}{\sqrt{n}} \sqrt{1 + \frac{z_u^2}{2}} = \hat{x}_u \pm z_{(1+\gamma)/2} \frac{s_X}{\sqrt{n}} \sqrt{1 + \frac{z_u^2}{2}} \quad (6.5)$$

Χαρτί κανονικής κατανομής

Όπως έχει περιγραφεί στο εδάφιο 5.3.3 στο ειδικό χαρτί της κανονικής κατανομής η συνάρτηση κατανομής ευθειοποιείται. Η απεικόνιση στο χαρτί κανονικής κατανομής είναι ισοδύναμη με την απεικόνιση της μεταβλητής x (κατακόρυφος άξονας) συναρτήσει της τυποποιημένης κανονικής μεταβλητής z (οριζόντιος άξονας).

6.1.2 Λογαριθμοκανονική κατανομή δύο παραμέτρων

Η λογαριθμοκανονική κατανομή δύο παραμέτρων προκύπτει από την κανονική κατανομή και το μετασχηματισμό

$$y = \ln x \Leftrightarrow x = e^y \quad (6.6)$$

Έτσι, λέμε ότι η μεταβλητή X ακολουθεί λογαριθμοκανονική κατανομή δύο παραμέτρων αν η Y ακολουθεί κανονική κατανομή $N(\mu_Y, \sigma_Y)$. Στο τυπολόγιο του Πίν. 6.2 συνοψίζονται οι βασικότερες μαθηματικές σχέσεις που συνδέονται με τη λογαριθμοκανονική κατανομή δύο παραμέτρων.

Πίν. 6.2 Τυπολόγιο της λογαριθμοκανονικής κατανομής δύο παραμέτρων.

| | |
|----------------------------------|---|
| Συνάρτηση πυκνότητας πιθανότητας | $f_X(x) = \frac{1}{x\sqrt{2\pi\sigma_Y}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu_Y}{\sigma_Y}\right)^2}$ |
| Συνάρτηση κατανομής | $F_X(x) = \int_0^x f_X(s)ds$ |
| Τιμές μεταβλητής | $0 < x < \infty$ (συνεχής) |
| Παράμετροι | μ_Y : παράμετρος κλίμακας $\sigma_Y > 0$: παράμετρος σχήματος |
| Μέση τιμή | $\mu_X = e^{\mu_Y + \frac{\sigma_Y^2}{2}}$ |
| Διασπορά | $\sigma_X^2 = e^{2\mu_Y + \sigma_Y^2} (e^{\sigma_Y^2} - 1)$ |
| Τρίτη κεντρική ροπή | $\mu_X^{(3)} = e^{3\mu_Y + \frac{3\sigma_Y^2}{2}} (e^{3\sigma_Y^2} - 3e^{\sigma_Y^2} + 2)$ |
| Συντελεστής μεταβλητότητας | $C_{v_X} = \sqrt{e^{\sigma_Y^2} - 1}$ |
| Συντελεστής ασυμμετρίας | $C_{s_X} = 3C_{v_X} + C_{v_X}^3$ |
| Πιθανότερη τιμή | $x_p = e^{\mu_Y - \sigma_Y^2}$ |
| Διάμεσος | $x_{0.5} = e^{\mu_Y}$ |

Ως άμεση συνέπεια της (6.6), η μεταβλητή X είναι θετική. Από τον Πίν. 6.2 προκύπτει ότι η κατανομή έχει πάντα θετική ασυμμετρία. Έτσι, το σχήμα της συνάρτησης πυκνότητας πιθανότητας είναι πάντα κωδωνοειδές και θετικά ασύμμετρο. Αυτές οι βασικές ιδιότητες της λογαριθμοκανονικής κατανομής είναι συμβατές με τις παρατηρημένες ιδιότητες πολλών υδρολογικών μεταβλητών, γι' αυτό και η χρήση της κατανομής είναι συχνή στην τεχνική υδρολογία. Σημειώνεται ακόμη ότι το γινόμενο δύο μεταβλητών που ακολουθούν λογαριθμοκανονική κατανομή δύο παραμέτρων, ακολουθεί επίσης λογαριθμοκανονική κατανομή. Αυτή η ιδιότητα σε συνδυασμό με το κεντρικό οριακό θεώρημα έχει από πολλούς θεωρηθεί ως θεωρητική τεκμηρίωση της χρήσης της κατανομής στην υδρολογία, αφού σε πολλές περιπτώσεις οι μεταβλητές μπορούν να θεωρηθούν ως αποτέλεσμα πολλών πολλαπλασιαστικών, παρά αθροιστικών, παραγόντων.

Τυπικοί υπολογισμοί

Οι τυπικοί υπολογισμοί της λογαριθμοκανονικής κατανομής βασίζονται στους αντίστοιχους υπολογισμούς της κανονικής κατανομής. Έτσι, συνδυάζοντας τις εξισώσεις (6.1) και (6.6) παίρνουμε

$$y_u = \mu_Y + z_u \sigma_Y \Leftrightarrow x_u = e^{\mu_Y + z_u \sigma_Y} \quad (6.7)$$

όπου z_u το u -ποσοστημόριο της τυποποιημένης κανονικής μεταβλητής. Το τελευταίο μπορεί να βρεθεί από πίνακες ή να υπολογιστεί αριθμητικά σύμφωνα με όσα περιγράφηκαν στο εδάφιο 6.1.1.

Εκτίμηση παραμέτρων

Όπως εύκολα προκύπτει από τις εξισώσεις του Πίν. 6.2, η μέθοδος των ροπών δίνει

$$\sigma_Y = \sqrt{\ln(1 + s_X^2 / \bar{x}^2)} \quad \mu_Y = \ln \bar{x} - \sigma_Y^2 / 2 \quad (6.8)$$

Η μέθοδος της μέγιστης πιθανοφάνειας δίνει τις ακόλουθες εκτιμήσεις (βλ. π.χ. Kite, 1988, σ. 57)

$$\mu_Y = \sum_{i=1}^n \ln x_i / n = \bar{y} \quad \sigma_Y = \sqrt{\sum_{i=1}^n (\ln x_i - \mu_Y)^2 / n} = s_Y \quad (6.9)$$

Παρατηρούμε ότι οι δύο μέθοδοι όχι μόνο δίνουν διαφορετικές εκτιμήσεις αλλά στηρίζονται και σε διαφορετικά δειγματικά χαρακτηριστικά. Έτσι, η μέθοδος των ροπών στηρίζεται στη μέση τιμή και την (μεροληπτική) τυπική απόκλιση της μεταβλητής X , ενώ η μέθοδος της μέγιστης πιθανοφάνειας στηρίζεται στη μέση τιμή και την (μεροληπτική) τυπική απόκλιση του λογαρίθμου της μεταβλητής X .

Τυπικό σφάλμα ποσοστημορίου και όρια εμπιστοσύνης

Εφόσον για την εκτίμηση των παραμέτρων της κατανομής χρησιμοποιείται η μέθοδος της μέγιστης πιθανοφάνειας, η σημειακή εκτίμηση των u -ποσοστημορίων των y και x είναι

$$\hat{y}_u = \ln(\hat{x}_u) = \bar{y} + z_u s_Y \Rightarrow \hat{x}_u = e^{\bar{y} + z_u s_Y} \quad (6.10)$$

όπου z_u το u -ποσοστημόριο της τυποποιημένης κανονικής μεταβλητής. Το τετράγωνο του τυπικού σφάλματος εκτίμησης της Y κατά τα γνωστά είναι

$$\varepsilon_Y^2 = \text{Var}(\hat{Y}_u) = \text{Var}(\ln \hat{X}_u) = \frac{s_Y^2}{n} \left(1 + \frac{z_u^2}{2} \right) \quad (6.11)$$

Συνδυάζοντας τα παραπάνω, παίρνουμε την ακόλουθη προσεγγιστική σχέση υπολογισμού των ορίων εμπιστοσύνης της x_u για βαθμό εμπιστοσύνης γ

$$\begin{aligned} \hat{x}_{u,1,2} &\approx \exp \left[(\bar{y} + z_u s_Y) \pm z_{(1+\gamma)/2} \frac{s_Y}{\sqrt{n}} \sqrt{1 + \frac{z_u^2}{2}} \right] = \\ &= \hat{x}_u \exp \left[\pm z_{(1+\gamma)/2} \frac{s_Y}{\sqrt{n}} \sqrt{1 + \frac{z_u^2}{2}} \right] \end{aligned} \quad (6.12)$$

όπου $z_{(1+\gamma)/2}$ το $[(1+\gamma)/2]$ -ποσοστημόριο της τυποποιημένης κανονικής μεταβλητής.

Στην περίπτωση που ο υπολογισμός των παραμέτρων γίνεται με τη μέθοδο των ροπών το τυπικό σφάλμα και τα αντίστοιχα όρια εμπιστοσύνης είναι διαφορετικά. Ο ενδιαφερόμενος αναγνώστης παραπέμπεται στον Kite (1988, σ. 60).

Χαρτί λογαριθμοκανονικής κατανομής

Το ειδικό χαρτί της κανονικής κατανομής (βλ. προηγούμενο εδάφιο) εύκολα μετασχηματίζεται σε χαρτί λογαριθμοκανονικής κατανομής, στο οποίο η αντίστοιχη συνάρτηση κατανομής ευθειοποιείται. Αρκεί να χρησιμοποιηθεί λογαριθμικός κατακόρυφος άξονας. Η απεικόνιση στο χαρτί λογαριθμοκανονικής κατανομής είναι ισοδύναμη με την απεικόνιση του λογαρίθμου της μεταβλητής $\ln x$ (κατακόρυφος άξονας) συναρτήσει της τυποποιημένης κανονικής μεταβλητής z (οριζόντιος άξονας).

Εφαρμογή 6.1.2

Στον Πίν. 6.3 δίνεται το δείγμα των μηνιαίων όγκων απορροής της λεκάνης του ποταμού Ευήνου ανάντη του υδρομετρικού σταθμού Πόρος Ρηγαίου, για το μήνα Ιανουάριο. Ζητείται η προσαρμογή της λογαριθμοκανονικής κατανομής 2 παραμέτρων και η εκτίμηση της απορροής πεντηκονταετίας.

Πίν. 6.3 Δείγμα μηνιαίου όγκου απορροής (σε hm^3) στη θέση Πόρος Ρηγαίου του ποταμού Ευήνου για το μήνα Ιανουάριο.

| Υδρολ. έτος | Όγκος απορροής | Υδρολ. έτος | Όγκος απορροής | Υδρολ. έτος | Όγκος απορροής |
|----------------|-------------------|----------------|-------------------|----------------|-------------------|
| 1970-71 | 102 | 1977-78 | 121 | 1984-85 | 178 |
| 1971-72 | 74 | 1978-79 | 317 | 1985-86 | 185 |
| 1972-73 | 78 | 1979-80 | 213 | 1986-87 | 101 |
| 1973-74 | 48 | 1980-81 | 111 | 1987-88 | 57 |
| 1974-75 | 31 | 1981-82 | 82 | 1988-89 | 24 |
| 1975-76 | 48 | 1982-83 | 61 | 1989-90 | 22 |
| 1976-77 | 114 | 1983-84 | 133 | 1990-91 | 51 |

Η μέση τιμή του δείγματος είναι

$$\bar{x} = \sum x / n = 102.4 \text{ hm}^3$$

Η τυπική απόκλιση (μεροληπτική εκτίμηση) είναι

$$s_X = \left(\sum x^2 / n - \bar{x}^2 \right)^{1/2} = 70.4 \text{ hm}^3$$

και ο συντελεστής μεταβλητότητας

$$\hat{C}_{v_X} = s_X / \bar{x} = 70.4 / 102.4 = 0.69$$

Ο συντελεστής ασυμμετρίας (μεροληπτική εκτίμηση) προκύπτει

$$\hat{C}_{s_X} = 1.4$$

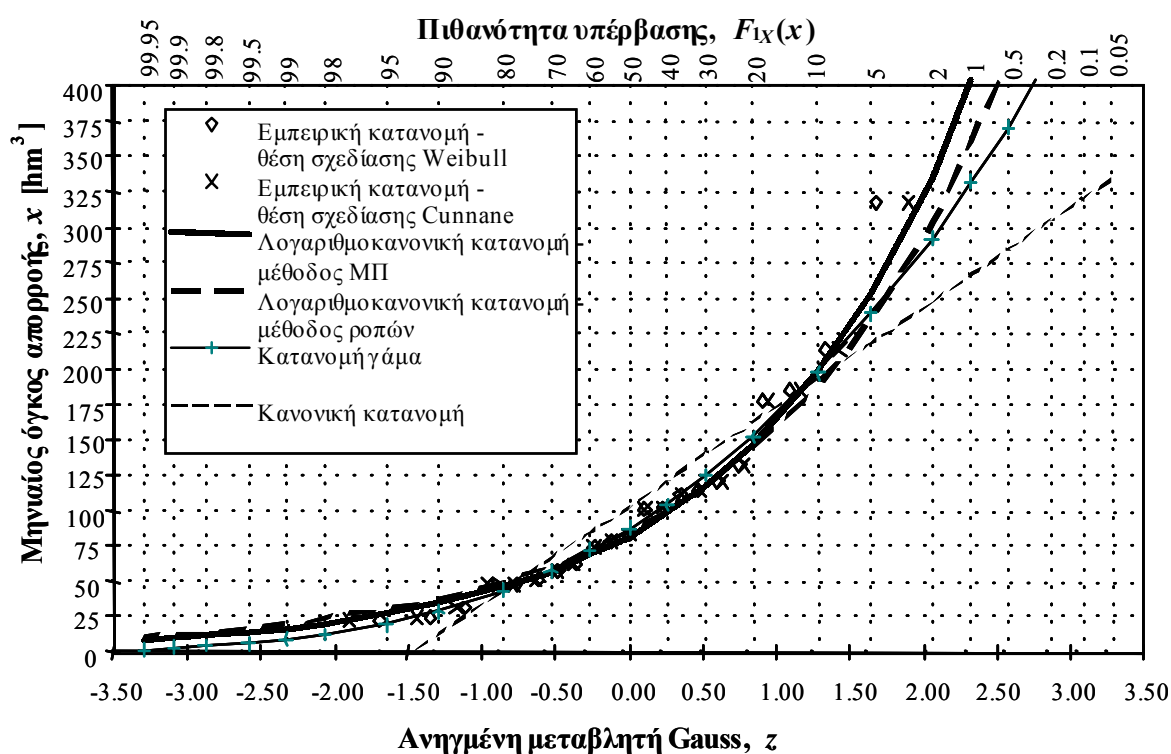
Αυτές οι τιμές των συντελεστών μεταβλητότητας και ασυμμετρίας δείχνουν μεγάλη απόκλιση από την κανονική κατανομή.

Η μέθοδος των ροπών δίνει

$$\sigma_Y = \sqrt{\ln(1 + s_X^2 / \bar{x}^2)} = 0.622, \quad \mu_Y = \ln \bar{x} - \sigma_Y^2 / 2 = 4.435$$

Η μέθοδος της μέγιστης πιθανοφάνειας δίνει

$$\mu_Y = \sum \ln x / n = 4.404 \text{ hm}^3, \quad \sigma_Y = \sqrt{\sum (\ln x)^2 / n - \mu_Y^2} = 0.687$$



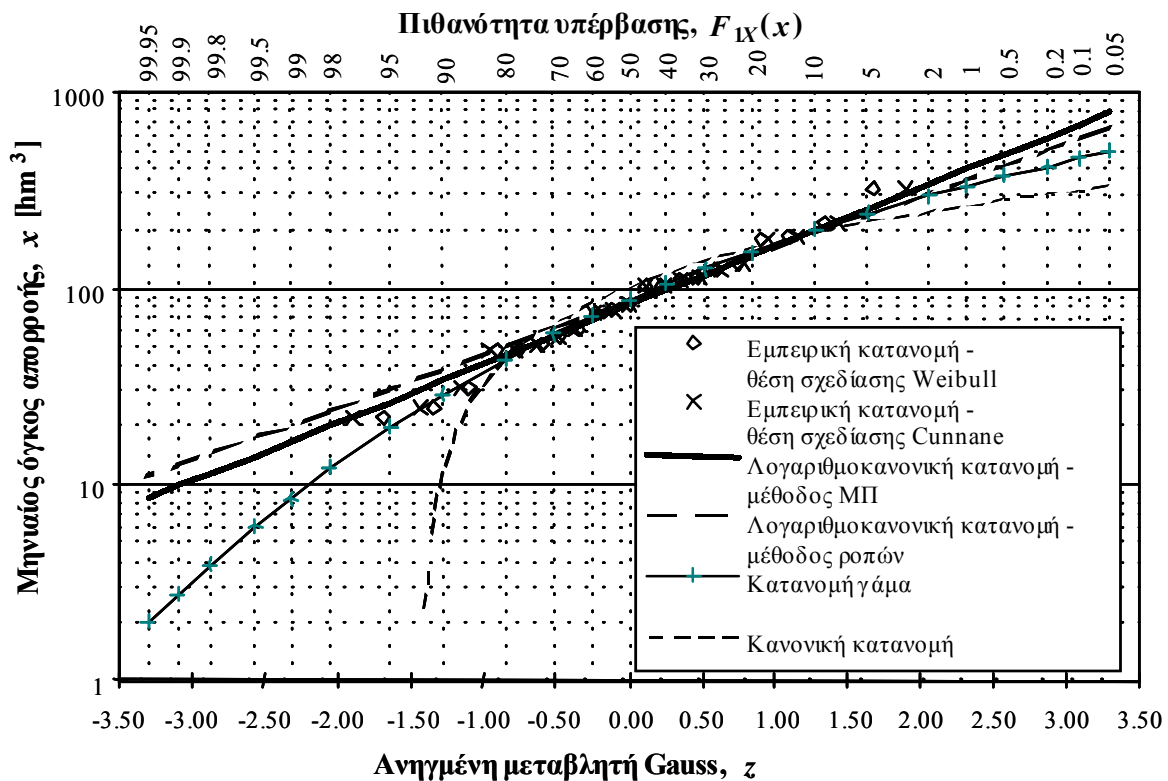
Σχ. 6.1 Εμπειρική και θεωρητικές συναρτήσεις κατανομής του όγκου απορροής του Ιανουαρίου στον Πόρο Ρηγανίου (Εφαρμογή 6.1.2) σε χαρτί κανονικής κατανομής.

Ο μηνιαίος όγκος απορροής πενήνταετίας δίνεται από την εξίσωση $x_u = \exp(\mu_Y + z_u \sigma_Y)$ όπου $u = 1 - 1/50 = 0.98$ και $z_u = 2.054$ (Πίν Π1). Με τις παραμέτρους της μεθόδου των ροπών προκύπτει $x_{0.98} = 302.7 \text{ hm}^3$, ενώ με τις παραμέτρους της μεθόδου μέγιστης πιθανοφάνειας προκύπτει $x_u = 335.1$. Στην τελευταία περίπτωση τα όρια εμπιστοσύνης 95% της τιμής αυτής είναι (με βάση την (6.12), για $z_u = 2.054$ και $z_{(1+\gamma)/2} = 1.96$)

$$\hat{x}_{u_{1,2}} \approx \exp \left[4.404 + 2.054 * 0.687 \pm 1.96 * \frac{0.687}{\sqrt{21}} * \sqrt{1 + \frac{2.054^2}{2}} \right]$$

$$= \exp(5.815 \pm 0.518) = \begin{cases} 562.8 \\ 199.7 \end{cases}$$

Παρατηρούμε ότι το διάστημα εμπιστοσύνης είναι πολύ ευρύ, γεγονός που αντανακλά τη μικρή αξιοπιστία στην πρόγνωση της απορροής πενήνταετίας του Ιανουαρίου. Η μείωση της μεγάλης αβεβαιότητας στην πρόγνωση αυτή προϋποθέτει σημαντικά μεγαλύτερο μέγεθος δείγματος.



Σχ. 6.2 Εμπειρική και θεωρητικές συναρτήσεις κατανομής του όγκου απορροής του Ιανουαρίου στον Πόρο Ρηγανίου (Εφαρμογή 6.1.2) σε χαρτί λογαριθμοκανονικής κατανομής.

Η καταλληλότητα της λογαριθμοκανονικής κατανομής μπορεί να ελεγχθεί με τη δοκιμή χ^2 (βλ εδάφιο 5.5.1). Αντί του ελέγχου αυτού δίνουμε στα Σχ. 6.1 και Σχ. 6.2 γραφική σύγκριση της εμπειρικής συνάρτησης κατανομής με τη λογαριθμοκανονική κατανομή. Το Σχ. 6.1 σχεδιάστηκε σε χαρτί κανονικής κατανομής, ενώ το Σχ. 6.2 σε χαρτί λογαριθμοκανονικής κατανομής. Για την εμπειρική κατανομή χρησιμοποιήθηκαν οι θέσεις σχεδίασης Weibull και Cunnane (βλ. Πίν. 5.7). Για τη λογαριθμοκανονική κατανομή χρησιμοποιήθηκαν και οι

δύο παραπάνω ομάδες παραμέτρων. Παρατηρούμε ότι η μέθοδος της μέγιστης πιθανοφάνειας δίνει καλύτερη προσαρμογή στην περιοχή των μικρών πιθανοτήτων υπέρβασης, που ενδιαφέρουν περισσότερο. Για σύγκριση έχει σχεδιαστεί και η κανονική κατανομή, η ακαταλληλότητα της οποίας για το συγκεκριμένο δείγμα είναι εμφανής, καθώς και η κατανομή γάμα (βλ. Εφαρμογή 6.2.2).

6.1.3 Λογαριθμοκανονική κατανομή τριών παραμέτρων (Galton)

Με συνδυασμό της κανονικής κατανομής και του τροποποιημένου λογαριθμικού μετασχηματισμού

$$y = \ln(x - c) \Leftrightarrow x = c + e^y \quad (6.13)$$

παίρνουμε τη λογαριθμοκανονική κατανομή τριών παραμέτρων ή κατανομή Galton*. Η κατανομή αυτή διαθέτει μία επιπλέον παράμετρο από την απλούστερη λογαριθμοκανονική κατανομή δύο παραμέτρων, την παράμετρο θέσης c , η οποία αποτελεί και το κάτω όριο της μεταβλητής. Η τρίτη αυτή παράμετρος επιτρέπει την καλύτερη προσαρμογή της κατανομής στα δεδομένα. Έτσι, αν χρησιμοποιούμε τη μέθοδο των ροπών, η τρίτη παράμετρος επιτρέπει τη διατήρηση του συντελεστή ασυμμετρίας της μεταβλητής. Στο τυπολόγιο του Πίν. 6.4 συνοψίζονται οι βασικότερες μαθηματικές σχέσεις που συνδέονται με τη λογαριθμοκανονική κατανομή τριών παραμέτρων.

Τυπικοί υπολογισμοί

Ο χειρισμός της κατανομής είναι παρόμοιος με αυτόν της λογαριθμοκανονικής κατανομής δύο παραμέτρων και γίνεται με βάση τη σχέση

$$y_u = \mu_Y + z_u \sigma_Y \Leftrightarrow x_u = c + e^{\mu_Y + z_u \sigma_Y} \quad (6.14)$$

όπου z_u το u -ποσοστημόριο της τυποποιημένης κανονικής μεταβλητής.

* Sir Francis Galton: Γενετιστής και βιοστατιστικός (1822-1911), γνωστός περισσότερο από τον ομώνυμο νόμο στον οποίο στηρίζεται η θεωρία της γραμμικής παλινδρόμησης (regression).

Εκτίμηση παραμέτρων

Οι ακόλουθες εξισώσεις της μεθόδου των ροπών προκύπτουν μετά από πράξεις από τις εκφράσεις των ροπών που δίνονται στον Πίν. 6.4. Η παράμετρος σ_Y υπολογίζεται από την

$$\sigma_Y = \sqrt{\ln(1 + \phi^2)} \quad (6.15)$$

όπου

$$\phi = \frac{1 - \omega^{2/3}}{\omega^{1/3}} \quad \omega = \frac{-\hat{C}_{s_X} + \sqrt{\hat{C}_{s_X}^2 + 4}}{2} \quad (6.16)$$

Πίν. 6.4 Τυπολόγιο της λογαριθμοκανονικής κατανομής τριών παραμέτρων.

| | |
|----------------------------------|--|
| Συνάρτηση πυκνότητας πιθανότητας | $f_X(x) = \frac{1}{(x-c)\sqrt{2\pi\sigma_Y}} e^{-\frac{1}{2}\left(\frac{\ln(x-c)-\mu_Y}{\sigma_Y}\right)^2}$ |
| Συνάρτηση κατανομής | $F_X(x) = \int_c^x f_X(s) ds$ |
| Τιμές μεταβλητής | $c < x < \infty$ (συνεχής) |
| Παράμετροι | c : παράμετρος θέσης μ_Y : παράμετρος κλίμακας $\sigma_Y > 0$: παράμετρος σχήματος |
| Μέση τιμή | $\mu_X = c + e^{\mu_Y + \frac{\sigma_Y^2}{2}}$ |
| Διασπορά | $\sigma_X^2 = e^{2\mu_Y + \sigma_Y^2} (e^{\sigma_Y^2} - 1)$ |
| Τρίτη κεντρική ροπή | $\mu_X^{(3)} = e^{3\mu_Y + \frac{3\sigma_Y^2}{2}} (e^{3\sigma_Y^2} - 3e^{\sigma_Y^2} + 2)$ |
| Συντελεστής ασυμμετρίας | $C_{s_X} = 3(e^{\sigma_Y^2} - 1)^{1/2} + (e^{\sigma_Y^2} - 1)^{3/2}$ |
| Πιθανότερη τιμή | $x_p = c + e^{\mu_Y - \sigma_Y^2}$ |
| Διάμεσος | $x_{0.5} = c + e^{\mu_Y}$ |

Οι άλλες δύο παράμετροι υπολογίζονται από τις εξισώσεις

$$\mu_Y = \ln(s_X / \phi) - \sigma_Y^2 / 2 \quad c = \bar{x} - \frac{s_X}{\phi} \quad (6.17)$$

Η μέθοδος της μέγιστης πιθανοφάνειας στηρίζεται στις εξισώσεις (βλ. π.χ. Kite, 1988, σ. 74)

$$\mu_Y = \sum_{i=1}^n \ln(x_i - c) / n \quad \sigma_Y^2 = \sum_{i=1}^n [\ln(x_i - c) - \mu_Y]^2 / n \quad (6.18)$$

$$(\mu_Y - \sigma_Y^2) \sum_{i=1}^n \frac{1}{x_i - c} = \sum_{i=1}^n \frac{\ln(x_i - c)}{x_i - c} \quad (6.19)$$

οι οποίες μόνο αριθμητικά μπορούν να επιλυθούν.

Η εκτίμηση των ορίων εμπιστοσύνης για την εν λόγω κατανομή απαιτεί πολύπλοκους υπολογισμούς. Ο ενδιαφερόμενος αναγνώστης παραπέμπεται στον Kite (1988, σ. 77).

6.2 Η ομάδα των κατανομών γάμα

6.2.1 Εκθετική κατανομή

Μια από τις απλούστερες αλλά και πολύ χρήσιμες κατανομές της στατιστικής είναι η εκθετική, τα βασικά χαρακτηριστικά της οποίας συνοψίζονται στον Πίν. 6.5.

Στην απλούστερη μορφή της, που έχουμε ήδη συναντήσει στην Εφαρμογή 2.4 έχει μόνο μία παράμετρο, την παράμετρο κλίμακας λ (η δεύτερη παράμετρος c είναι ίση με 0). Η συνάρτηση πυκνότητας πιθανότητας της εκθετικής κατανομής είναι παντού φθίνουσα, δηλαδή έχει σχήμα ανεστραμμένου J.

Όπως ήδη έχουμε αναφέρει (ενότητα 2.4, ένθετο εδάφιο και Εφαρμογή 2.4) η εκθετική κατανομή χρησιμοποιείται για την περιγραφή υδρολογικών μεταβλητών σε μικρή χρονική κλίμακα, όπως για παράδειγμα των ωριαίων ή ημερήσιων υψών βροχής. Από ένα θεώρημα της θεωρίας πιθανοτήτων προκύπτει ότι οι αποστάσεις μεταξύ διαδοχικών τυχαίων σημείων στο χρόνο ακολουθούν εκθετική κατανομή. Αυτό το θεώρημα έχει συχνή εφαρμογή στην τεχνική υδρολογία, όπου, για παράδειγμα, οι

χρόνοι που μεσολαβούν ανάμεσα στις εμφανίσεις διαδοχικών επεισοδίων βροχής ακολουθούν εκθετική κατανομή. Επιπλέον, και οι διάρκειες των επεισοδίων βροχής ακολουθούν πολύ συχνά εκθετική κατανομή.

Πίν. 6.5 Τυπολόγιο της εκθετικής κατανομής.

| | |
|----------------------------------|---|
| Συνάρτηση πυκνότητας πιθανότητας | $f_X(x) = \lambda e^{-\lambda(x-c)}$ |
| Συνάρτηση κατανομής | $F_X(x) = 1 - e^{-\lambda(x-c)}$ |
| Τιμές μεταβλητής | $c < x < \infty$ (συνεχής) |
| Παράμετροι | c : παράμετρος θέσης $\lambda > 0$: παράμετρος κλίμακας |
| Μέση τιμή | $\mu_X = c + \frac{1}{\lambda}$ |
| Διασπορά | $\sigma_X^2 = \frac{1}{\lambda^2}$ |
| Τρίτη κεντρική ροπή | $\mu_X^{(3)} = \frac{2}{\lambda^3}$ |
| Τέταρτη κεντρική ροπή | $\mu_X^{(4)} = \frac{9}{\lambda^4}$ |
| Συντελεστής μεταβλητότητας | $C_{v_X} = \frac{1}{c\lambda + 1}$ |
| Συντελεστής ασυμμετρίας | $C_{s_X} = 2$ |
| Συντελεστής κύρτωσης | $C_{k_X} = 9$ |
| Πιθανότερη τιμή | $x_p = c$ |
| Διάμεσος τιμή | $x_{0.5} = c - \frac{\ln 0.5}{\lambda}$ |

6.2.2 Κατανομή γάμα δύο παραμέτρων

Η κατανομή γάμα είναι από τις πιο διαδεδομένες κατανομές της τεχνικής υδρολογίας. Τα βασικά χαρακτηριστικά της δίνονται στο τυπολόγιο του Πίν. 6.6. Όπως και η λογαριθμική κατανομή δύο παραμέτρων είναι θετικά ασύμμετρη και ορίζεται μόνο για θετικές τιμές της μεταβλητής. Οι

ιδιότητες αυτές την κάνουν συμβατή με τις πιο χαρακτηριστικές υδρολογικές μεταβλητές, όπως μηνιαίες και ετήσιες παροχές ή βροχές.

Πίν. 6.6 Τυπολόγιο της κατανομής γάμα 2 παραμέτρων.

| | |
|----------------------------------|--|
| Συνάρτηση πυκνότητας πιθανότητας | $f_X(x) = \frac{\lambda^\kappa}{\Gamma(\kappa)} x^{\kappa-1} e^{-\lambda x}$ |
| Συνάρτηση κατανομής | $F_X(x) = \int_0^x f_X(s) ds$ |
| Τιμές μεταβλητής | $0 < x < \infty$ (συνεχής) |
| Παράμετροι | $\lambda > 0$: παράμετρος κλίμακας $\kappa > 0$: παράμετρος σχήματος |
| Μέση τιμή | $\mu_X = \frac{\kappa}{\lambda}$ |
| Διασπορά | $\sigma_X^2 = \frac{\kappa}{\lambda^2}$ |
| Τρίτη κεντρική ροπή | $\mu_X^{(3)} = \frac{2\kappa}{\lambda^3}$ |
| Τέταρτη κεντρική ροπή | $\mu_X^{(4)} = \frac{3\kappa(\kappa+2)}{\lambda^4}$ |
| Συντελεστής μεταβλητότητας | $C_{v_X} = \frac{1}{\sqrt{\kappa}}$ |
| Συντελεστής ασυμμετρίας | $C_{s_X} = \frac{2}{\sqrt{\kappa}} = 2C_{v_X}$ |
| Συντελεστής κύρτωσης | $C_{k_X} = 3 + \frac{6}{\kappa} = 3 + 6C_{v_X}^2$ |
| Πιθανότερη τιμή | $x_p = (\kappa - 1) / \lambda$ (για $\kappa > 1$) |

Η κατανομή γάμα έχει δύο παραμέτρους, την παράμετρο κλίμακας λ και την παράμετρο σχήματος κ .^{*} Για $\kappa = 1$ η κατανομή ταυτίζεται με την εκθετική, η οποία είναι ειδική περίπτωση της γάμα. Για $\kappa > 1$ η συνάρτηση πυκνότητας πιθανότητας της κατανομής εμφανίζει κωδωνοειδές σχήμα, ενώ για $\kappa < 1$ το σχήμα της γίνεται ανεστραμμένο J, με άπειρη

^{*} Για ακέραια τιμή του κ , η κατανομή γάμα πολλές φορές αναφέρεται ως κατανομή Erlang.

τεταγμένη στη θέση $x = 0$. Για μεγάλες τιμές του κ (πάνω από 15-30) η κατανομή γάμα πλησιάζει προς την κανονική.

Η κατανομή γάμα, όπως και η κανονική, είναι κλειστή ως προς την πρόσθεση, αλλά μόνο όταν υπάρχει στοχαστική ανεξαρτησία και κοινή παράμετρος κλίμακας των προσθετέων. Έτσι, το άθροισμα δύο ανεξάρτητων μεταβλητών που ακολουθούν κατανομές γάμα με κοινή παράμετρο κλίμακας λ , ακολουθεί επίσης κατανομή γάμα.

Τέλος, υπενθυμίζουμε ότι η κατανομή χ^2 , η οποία έχει εξεταστεί στο εδάφιο 2.8.3, είναι ειδική περίπτωση της κατανομής γάμα.

Τυπικοί υπολογισμοί

Όπως συμβαίνει και με την κανονική κατανομή, το ολοκλήρωμα της συνάρτησης κατανομής γάμα δεν υπολογίζεται αναλυτικά, πράγμα που δημιουργεί δυσκολίες στους υπολογισμούς. Η απλούστερη λύση στηρίζεται στην πινακοποίηση των τιμών της τυποποιημένης μεταβλητής $k = (x - \mu_X) / \sigma_X$, όπου μ_X και σ_X η μέση τιμή και η τυπική απόκλιση της X , αντίστοιχα. Τέτοια πινακοποίηση περιέχεται σε διάφορα βιβλία θεωρίας πιθανοτήτων, καθώς και στους Πίν. Π4α και Π4β του Παραρτήματος (στο τέλος του βιβλίου).

Έτσι, ο υπολογισμός του u -ποσοστημορίου (x_u) γίνεται από την εξίσωση

$$x_u = \mu_X + k_u \sigma_X \quad (6.20)$$

αφού προηγουμένως προσδιοριστεί από τον Πίν. Π4α ή Π4β η τιμή k_u που αντιστοιχεί στη δεδομένη τιμή της συνάρτησης κατανομής $u = F_K(k_u)$. Αντίστροφα, για δεδομένο x_u υπολογίζεται από την (6.1) το k_u και από τον Πίν. Π4α ή Π4β το $u = F_K(k_u)$. Στον Πίν. Π4α ή Π4β κάθε στήλη αντιστοιχεί σε συγκεκριμένη τιμή της παραμέτρου κ (ή, ισοδύναμα, του συντελεστή ασυμμετρίας $C_{sX} = 2 / \sqrt{\kappa} = 2\sigma_X / \bar{x}$). Έτσι θα πρέπει να επιλεγεί η κατάλληλη στήλη, ή να γίνει κατάλληλη παρεμβολή των τιμών γειτονικών στηλών.

Στη βιβλιογραφία δίνονται διάφορες αριθμητικές προσεγγίσεις της συνάρτησης κατανομής γάμα, με βάση τις οποίες μπορούμε να αποφύγουμε τη χρήση πινάκων (Press et al., 1987· Stedinger et al., 1993). Η χρήση των προσεγγίσεων αυτών είναι πλεονεκτικότερη όταν οι υπολογι-

σμοί γίνονται με ηλεκτρονικούς υπολογιστές ή αριθμομηχανές. Στο Παράρτημα 6.B δίνονται οι απλούστερες προσεγγίσεις.

Εκτίμηση παραμέτρων

Η μέθοδος των ροπών δίνει άμεσα τις ακόλουθες απλές εκτιμήσεις των παραμέτρων της κατανομής γάμα:

$$\kappa = \frac{\bar{x}^2}{s_x^2} \quad \lambda = \frac{\bar{x}}{s_x^2} \quad (6.21)$$

Οι εκτιμήσεις της μεθόδου της μέγιστης πιθανοφάνειας είναι πολυπλοκότερες. Βασίζονται στην επίλυση των εξισώσεων (βλ. π.χ. Bobée and Ashkar, 1991)

$$\ln \kappa - \psi(\kappa) = \ln \bar{x} - \frac{1}{n} \sum_{i=1}^n \ln x_i \quad \lambda = \frac{\kappa}{\bar{x}} \quad (6.22)$$

όπου $\psi(\kappa) = d \ln \Gamma(\kappa) / d\kappa$ η λεγόμενη συνάρτηση δίζαμα (παράγωγος του λογαρίθμου της συνάρτησης γάμα), πινακοποιημένες τιμές της οποίας δίνονται σε μαθηματικά συγγράμματα. Οι Masuyama and Kuroiwa (1951) έδωσαν την ακόλουθη προσεγγιστική σχέση για την επίλυση ως προς κ της πρώτης από τις παραπάνω εξισώσεις:

$$\kappa = \begin{cases} \frac{0.5000876 + 0.1648852\phi - 0.054427\phi^2}{\phi} & 0 \leq \phi \leq 0.5772 \\ \frac{8.898919 + 9.05995\phi + 0.9775373\phi^2}{\phi(17.7928 + 11.968477\phi + \phi^2)} & 0.5772 \leq \phi \leq 17.0 \end{cases} \quad (6.23)$$

όπου $\phi = \ln \kappa - \psi(\kappa)$. Το σφάλμα της προσέγγισης αυτής είναι πρακτικά αμελητέο (0.0088% για την πρώτη εξίσωση και 0.0054% για τη δεύτερη).

Τυπικό σφάλμα ποσοστημορίου και όρια εμπιστοσύνης

Η σημειακή εκτίμηση του u -ποσοστημορίου της κατανομής γάμα είναι

$$\hat{x}_u = \bar{x} + k_u s_x \quad (6.24)$$

Στην περίπτωση που οι παράμετροι υπολογίζονται με τη μέθοδο των ροπών, το τετραγωνικό τυπικό σφάλμα της εκτίμησης είναι (βλ. π.χ. Bobée and Ashkar, 1991, σ. 50)

$$\varepsilon_u^2 = \frac{s_X^2}{n} \left[(1 + k_u C_{v_X})^2 + \frac{1}{2} \left(k_u + 2C_{v_X} \frac{\partial k_u}{\partial C_{s_X}} \right)^2 (1 + C_{v_X})^2 \right] \quad (6.25)$$

Το μέγεθος $\partial k_u / \partial C_{s_X}$ μπορεί να εκτιμηθεί προσεγγιστικά από κάποια προσέγγιση της κατανομής γάμα, π.χ. από την προσέγγιση Wilson-Hilferty. Συχνά, σε πρώτη προσέγγιση, ο αντίστοιχος όρος παραλείπεται οπότε η παραπάνω έκφραση απλοποιείται σημαντικά:

$$\varepsilon_u^2 = \frac{s_X^2}{n} \left[1 + 2C_{v_X} k_u + \frac{1}{2} (1 + 3C_{v_X}^2) k_u^2 \right] \quad (6.26)$$

Έτσι, τα προσεγγιστικά όρια εμπιστοσύνης για βαθμό εμπιστοσύνης α είναι

$$\hat{x}_{u_{1,2}} \approx (\bar{x} + k_u s_X) \pm z_{(1+\gamma)/2} \frac{s_X}{\sqrt{n}} \sqrt{1 + 2C_{v_X} k_u + \frac{1}{2} (1 + 3C_{v_X}^2) k_u^2} \quad (6.27)$$

Αν χρησιμοποιείται η μέθοδος της μέγιστης πιθανοφάνειας ο υπολογισμός των ορίων εμπιστοσύνης είναι πιο πολύπλοκος. Ο ενδιαφερόμενος αναγνώστης παραπέμπεται στους Bobée and Ashkar (1991, σ. 46).

Χαρτί κατανομής γάμα

Δεν μπορεί να κατασκευαστεί χαρτί κατανομής γάμα, τέτοιο που να ευθειοποιεί κάθε συνάρτηση κατανομής γάμα. Ωστόσο, μπορεί να κατασκευαστεί χαρτί γάμα αλλά για δεδομένη τιμή της παραμέτρου σχήματος κ . Κάτι τέτοιο, βεβαίως, δεν είναι πρακτικό και γι' αυτό η απεικόνιση της κατανομής γάμα γίνεται συνήθως σε χαρτί κανονικής κατανομής ή σε χαρτί κατανομής Weibull. Σε αυτή την περίπτωση, όμως, η κατανομή δεν απεικονίζεται ως ευθεία, αλλά ως καμπύλη.

Εφαρμογή 6.2.2

Ζητείται η προσαρμογή της κατανομής γάμα 2 παραμέτρων και η εκτίμηση της απορροής πενηντηκονταετίας για το δείγμα των μηνιαίων όγκων απορροής της λεκάνης του ποταμού Ευήνου ανάντη του υδρομετρικού

σταθμού Πόρος Ρηγανίου, για το μήνα Ιανουάριο (Εφαρμογή 6.1.2, Πίν. 6.3).

Για μέση τιμή 102.4 hm^3 και τυπική απόκλιση 70.4 hm^3 η μέθοδος των ροπών δίνει τις ακόλουθες τιμές των παραμέτρων της κατανομής:

$$\kappa = 102.4^2 / 70.4^2 = 2.11, \lambda = 102.4 / 70.4^2 = 0.0207.$$

Για $T = 50$ ή ισοδύναμα $F = 0.98$ από τον Πίν Π4α* προκύπτει $k_{0.98} = 2.70$ και

$$x_u = 102.4 + 2.70 \times 70.4 = 292.5 \text{ hm}^3.$$

Παίρνοντας από τον Πίν. Π1 τις τιμές του k_u για διάφορες τιμές του u υπολογίζουμε μια σειρά ποσοστημορίων της κατανομής, οπότε είμαστε σε θέση να απεικονίσουμε τη συνάρτηση κατανομής. Η απεικόνιση αυτή δίνεται (σε σύγκριση και με άλλες κατανομές) στα Σχ. 6.1 (σε χαρτί κανονικής κατανομής) και Σχ. 6.2 (σε χαρτί λογαριθμοκανονικής κατανομής). Παρατηρούμε ότι η κατανομή γάμα γενικά προσεγγίζει την λογαριθμοκανονική κατανομή, αλλά στην περιοχή των μικρών πιθανοτήτων υπέρβασης η προσαρμογή της προς την εμπειρική κατανομή του συγκεκριμένου δείγματος είναι χειρότερη από αυτήν της λογαριθμοκανονικής κατανομής.

* Εναλλακτικά, αντί να χρησιμοποιήσουμε τον Πίν Π4α, μπορούμε να αξιοποιήσουμε την προσέγγιση της εξίσωσης (6.64) (βλ. Παράρτημα 6.B). Σε αυτή την περίπτωση, οι υπολογισμοί είναι οι ακόλουθοι:

$$\mu = 0.6 \times (2.11^{0.5} - 1) - (1/2.11^{0.5} - 1) = 0.583$$

$$\nu = 0.6 \times (2.11^{0.5} - 1) + 0.01 \times (2.11 - 1) + 1 = 1.283$$

$$a = 0.6 / 2.11^{0.5} + 0.08 = 0.493$$

$$\beta = 0.0234 \ln(2.11) = 0.0175$$

$$c = 1 + 31 \times \exp[-11.6 \times (2.11 - 1)^{-0.25}] = 1.0$$

Έτσι, η (6.64) γίνεται

$$\begin{aligned} x_u &= [0.583 / (0.0207 \times 0.493)] u^{0.493} + [1.283 / (0.0207 \times 0.0175)] [1 - (1-u)^{0.0175}] \\ &= 57.12 u^{0.493} + 3541.8 [1 - (1-u)^{0.0175}] \end{aligned}$$

Για $u = 0.98$ η παραπάνω εξίσωση δίνει

$$x_u = 56.6 + 234.4 = 291.0 \text{ hm}^3$$

Για τον υπολογισμό των ορίων εμπιστοσύνης 95% της απορροής 50ετίας χρησιμοποιούμε την προσεγγιστική σχέση (6.27), η οποία για $z_{(1+\gamma)/2} = 1.96$, $k_u = 2.70$ και $C_{vX} = 0.69$ δίνει

$$\begin{aligned}\hat{x}_{u,2} &\approx 292.5 \pm 1.96 * \frac{70.4}{\sqrt{21}} * \sqrt{1 + 2 * 0.69 * 2.70 + \frac{1}{2} (1 + 3 * 0.69^2) * 2.70^2} \\ &\approx 292.5 \pm 110.9 = \begin{cases} 403.4 \\ 181.6 \end{cases}\end{aligned}$$

6.2.3 Κατανομή γάμα τριών παραμέτρων (Pearson III)

Με την προσθήκη μιας παραμέτρου θέσης στην κατανομή γάμα δύο παραμέτρων, παίρνουμε την κατανομή γάμα τριών παραμέτρων, πιο γνωστή ως κατανομή Pearson τύπου III.

Η παράμετρος θέσης c , η οποία αποτελεί και το κάτω όριο του μεταβλητής επιτρέπει την καλύτερη προσαρμογή της κατανομής στα δεδομένα. Έτσι, αν χρησιμοποιούμε τη μέθοδο των ροπών, η τρίτη παράμετρος επιτρέπει τη διατήρηση του συντελεστή ασυμμετρίας της μεταβλητής. Στο τυπολόγιο του Πίν. 6.7 συνοψίζονται οι βασικότερες μαθηματικές σχέσεις που συνδέονται με την κατανομή Pearson III.

Οι ιδιότητες της κατανομής είναι παρόμοιες με αυτές της κατανομής γάμα δύο παραμέτρων. Για τους τυπικούς υπολογισμούς της κατανομής χρησιμοποιείται και εδώ η εξίσωση (6.20). Μπορούν επίσης να χρησιμοποιηθούν οι προσεγγίσεις που δόθηκαν για την κατανομή γάμα δύο παραμέτρων. Διαφορετικές είναι οι εξισώσεις για την εκτίμηση των παραμέτρων της κατανομής. Έτσι, η μέθοδος των ροπών δίνει

$$\kappa = \frac{4}{\hat{C}_{s_X}^2} \quad \lambda = \frac{\sqrt{\kappa}}{s_X} \quad c = \bar{x} - \frac{\kappa}{\lambda} \quad (6.28)$$

Η μέθοδος της μέγιστης πιθανοφάνειας δίνει πιο πολύπλοκες εξισώσεις, τις οποίες ο ενδιαφερόμενος αναγνώστης μπορεί να βρει π.χ. στα βιβλία των Bobée and Ashkar (1991, σ. 59) και Kite (1988, σ. 117). Στα ίδια βιβλία υπάρχουν και οι τύποι για την εκτίμηση του τυπικού σφάλματος και των ορίων εμπιστοσύνης των ποσοστημορίων της κατανομής.

Πίν. 6.7 Τυπολόγιο της κατανομής Pearson III.

| | |
|----------------------------------|---|
| Συνάρτηση πυκνότητας πιθανότητας | $f_X(x) = \frac{\lambda^\kappa}{\Gamma(\kappa)} (x-c)^{\kappa-1} e^{-\lambda(x-c)}$ |
| Συνάρτηση κατανομής | $F_X(x) = \int_c^x f_X(s) ds$ |
| Τιμές μεταβλητής | $c < x < \infty$ (συνεχής) |
| Παράμετροι | c : παράμετρος θέσης $\lambda > 0$: παράμετρος κλίμακας $\kappa > 0$: παράμετρος σχήματος |
| Μέση τιμή | $\mu_X = c + \frac{\kappa}{\lambda}$ |
| Διασπορά | $\sigma_X^2 = \frac{\kappa}{\lambda^2}$ |
| Τρίτη κεντρική ροπή | $\mu_X^{(3)} = \frac{2\kappa}{\lambda^3}$ |
| Τέταρτη κεντρική ροπή | $\mu_X^{(4)} = \frac{3\kappa(\kappa+2)}{\lambda^4}$ |
| Συντελεστής ασυμμετρίας | $C_{s_X} = \frac{2}{\sqrt{\kappa}}$ |
| Συντελεστής κύρτωσης | $C_{k_X} = 3 + \frac{6}{\kappa}$ |
| Πιθανότερη τιμή | $x_p = c + (\kappa - 1) / \lambda$ (για $\kappa > 1$) |

6.2.4 Κατανομή Log-Pearson III

Η κατανομή Log-Pearson III προκύπτει από την κατανομή Pearson III και το μετασχηματισμό

$$y = \ln x \Leftrightarrow x = e^y \quad (6.29)$$

Έτσι, λέμε ότι η μεταβλητή X ακολουθεί κατανομή Log-Pearson III αν η Y ακολουθεί κατανομή Pearson III. Στο τυπολόγιο του (Πίν. 6.8) συνοψίζονται οι βασικότερες μαθηματικές σχέσεις που συνδέονται με την κατανομή Log-Pearson III.

Πίν. 6.8 Τυπολόγιο της κατανομής Log Pearson III.

| | |
|----------------------------------|---|
| Συνάρτηση πυκνότητας πιθανότητας | $f_X(x) = \frac{\lambda^\kappa}{x\Gamma(\kappa)} (\ln x - c)^{\kappa-1} e^{-\lambda(\ln x - c)}$ |
| Συνάρτηση κατανομής | $F_X(x) = \int_{e^c}^x f_X(s) ds$ |
| Τιμές μεταβλητής | $e^c < x < \infty$ (συνεχής) |
| Παράμετροι | c : παράμετρος κλίμακας $\lambda > 0$: παράμετρος σχήματος $\kappa > 0$: παράμετρος σχήματος |
| Μέση τιμή | $\mu_X = e^c \left(\frac{\lambda}{\lambda - 1} \right)^\kappa$ |
| Διασπορά | $\sigma_X^2 = e^{2c} \left[\left(\frac{\lambda}{\lambda - 2} \right)^\kappa - \left(\frac{\lambda}{\lambda - 1} \right)^{2\kappa} \right]$ |
| Ροπή περί την αρχή τάξης r | $m_X^{(r)} = e^{rc} \left(\frac{\lambda}{\lambda - r} \right)^\kappa$ |

Η συνάρτηση πυκνότητας πιθανότητας της κατανομής Log-Pearson III μπορεί να πάρει διάφορα σχήματα, όπως κωδωνοειδές, ανεστραμμένο J, U, κ.ά. Από τον Πίν. 6.8 προκύπτει ότι η τρίτη ροπή της κατανομής μπορεί να γίνει ακόμη και άπειρη, για $\lambda \leq 3$. Αυτό δείχνει ότι η κατανομή μπορεί να έχει πολύ μεγάλο συντελεστή ασυμμετρίας. Σε αυτή την ιδιότητα της οφείλεται η ευρεία διάδοσή της στην τεχνική υδρολογία. Κυρίως έχει χρησιμοποιηθεί για την περιγραφή πλημμυρικών παροχών, παράλληλα με τις ασυμπτωτικές κατανομής μεγίστων που θα εξεταστούν στην επόμενη ενότητα. Ειδικά στις ΗΠΑ έχει υιοθετηθεί ως η τυπική κατανομή για πλημμύρες από όλες τις κρατικές υπηρεσίες.

Τυπικοί υπολογισμοί

Οι τυπικοί υπολογισμοί της κατανομής Log-Pearson III βασίζονται στους αντίστοιχους υπολογισμούς της κατανομής Pearson III. Έτσι, συνδυάζοντας τις εξισώσεις (6.20) και (6.29) παίρνουμε

$$y_u = \mu_Y + k_u \sigma_Y \Leftrightarrow x_u = e^{\mu_Y + k_u \sigma_Y} \quad (6.30)$$

όπου η τιμή του k_u μπορεί να βρεθεί από πίνακες (Πίν. Π4α ή Π4β). Η τιμή του y_u μπορεί ακόμη να υπολογιστεί αριθμητικά σύμφωνα με όσα περιγράφηκαν στο εδάφιο 6.2.2.

Εκτίμηση παραμέτρων

Η εκτίμηση των παραμέτρων είτε με τη μέθοδο των ροπών, είτε με τη μέθοδο της μέγιστης πιθανοφάνειας απαιτεί μια αρκετά πολύπλοκη διαδικασία (Bobée and Ashkar, 1991, σ. 85· Kite, 1988, σ. 138). Εδώ θα περιοριστούμε στην αναφορά της απλούστερης *έμμεσης μεθόδου των ροπών*: Σύμφωνα με αυτή από το διαθέσιμο δείγμα υπολογίζονται οι τιμές $y_i = \ln x_i$, στη συνέχεια υπολογίζονται τα στατιστικά χαρακτηριστικά των y_i και τέλος εφαρμόζονται οι εξισώσεις της μεθόδου των ροπών για τη μεταβλητή Y , δηλαδή

$$\kappa = \frac{4}{\hat{C}_{sY}^2} \quad \lambda = \frac{\sqrt{\kappa}}{s_Y} \quad c = \bar{y} - \frac{\kappa}{\lambda} \quad (6.31)$$

Όπως και στην περίπτωση της κατανομής Pearson III, έτσι και εδώ η εκτίμηση των ορίων εμπιστοσύνης είναι αρκετά πολύπλοκη.

Χαρτί κατανομής Log-Pearson III

Δεν μπορεί να κατασκευαστεί χαρτί κατανομής γάμα, τέτοιο που να ευθειοποιεί κάθε συνάρτηση κατανομής Log-Pearson III. Ωστόσο, μπορεί να κατασκευαστεί χαρτί για δεδομένη τιμή της παραμέτρου σχήματος κ . Ένα τέτοιο χαρτί έχει άξονα πιθανοτήτων ταυτόσημο με αυτό του χαρτιού γάμα και άξονα τιμών της μεταβλητής λογαριθμικό. Κάτι τέτοιο, βεβαίως, δεν είναι πρακτικό και γι' αυτό η απεικόνιση της κατανομής Log-Pearson III γίνεται συνήθως σε χαρτί λογαριθμοκανονικής κατανομής ή σε χαρτί κατανομής Gumbel. Σε αυτή την περίπτωση, όμως, η κατανομή δεν απεικονίζεται ως ευθεία, αλλά ως καμπύλη.

6.3 Ασυμπτωτικές κατανομές ακροτάτων

Με τον όρο *ασυμπτωτική κατανομή ακροτάτων* εννοούμε την οριακή κατανομή της ακρότατης (δηλαδή της μεγαλύτερης ή, εναλλακτικά, της μικρότερης) από k ισόνομες μεταβλητές, όταν ο αριθμός k τείνει στο

άπειρο. Συμβολικά, αν Y_1, \dots, Y_k είναι μια ακολουθία ισόνομων τυχαίων μεταβλητών και

$$X_k = \max(Y_1, \dots, Y_k) \quad (6.32)$$

τότε η αντίστοιχη ασυμπτωτική κατανομή μεγίστων είναι η

$$F_X(x) = \lim_{k \rightarrow \infty} F_{X_k}(x) = \lim_{k \rightarrow \infty} P(X_k \leq y) \quad (6.33)$$

Αντίστοιχα ορίζεται και η ασυμπτωτική κατανομή ελαχίστων. Στην περίπτωση που οι τυχαίες μεταβλητές Y_i είναι ανεξάρτητες, ο προσδιορισμός της $F_X(x)$ απλοποιείται αρκετά, δεδομένου ότι

$$F_{X_k}(x) = [F_Y(x)]^k \quad (6.34)$$

όπως εύκολα μπορεί να διαπιστωθεί παίρνοντας υπόψη την ανεξαρτησία των μεταβλητών.

Σε πολλές περιπτώσεις η παραπάνω ασυμπτωτική κατανομή δεν εξαρτάται από το ακριβές σχήμα της αρχικής συνάρτησης κατανομής $F_Y(y)$. Οι περιπτώσεις αυτές, που έχουν μεγάλο πρακτικό ενδιαφέρον, έχουν μελετηθεί σε λεπτομέρεια από τον Gumbel (1958). Αναφέρονται σε τυχαίες μεταβλητές Y_i ανεξάρτητες και ισόνομες, των οποίων οι κατανομές ικανοποιούν ορισμένες γενικές συνθήκες.

Η σημασία των κατανομών ακροτάτων στην τεχνική υδρολογία προκύπτει από την ομοιότητα της έννοιας των ακροτάτων, όπως ορίζεται πιο πάνω, με την ανέλιξη ακροτάτων, όπως έχει οριστεί στο εδάφιο 4.1.2. Έτσι, για παράδειγμα, αν συμβολίσουμε με Y_i την ημερήσια παροχή σε μια διατομή ποταμού κατά την ημέρα i και με X_{365} την μέγιστη ημερήσια παροχή κατά τη διάρκεια ενός έτους, είναι προφανής η σύνδεση των μεταβλητών αυτών με την (6.32). Στα πρακτικά προβλήματα αντιπλημμυρικού σχεδιασμού αυτό που ενδιαφέρει είναι η κατανομή μεταβλητών όπως η X_{365} του παραδείγματος (παρά η κατανομή της Y_i). Ωστόσο, οι αυστηρές προϋποθέσεις, κάτω από τις οποίες προκύπτουν θεωρητικά οι κατανομές ακροτάτων, σπάνια ικανοποιούνται από τις φυσικές υδρολογικές μεταβλητές. Στο παραπάνω παράδειγμα, οι διάφορες Y_i ούτε ανεξάρτητες ούτε ισόνομες μπορούν να θεωρηθούν. Εξ άλλου η σύγκλιση προς την οριακή κατανομή είναι κατά κανόνα πολύ αργή. Για όλους αυτούς τους λόγους δεν είναι ποτέ αυτονόητο ότι μια

συγκεκριμένη μέγιστη ή ελάχιστη υδρολογική μεταβλητή ακολουθεί την ασυμπτωτική κατανομή που προβλέπεται θεωρητικά. Η υιοθέτηση της συγκεκριμένης κατανομής θα πρέπει να γίνεται μετά από έλεγχο προσαρμογής στα πραγματικά δεδομένα.

Στα παρακάτω εδάφια εξετάζονται οι δύο πιο διαδεδομένες στην τεχνική υδρολογία κατανομές ακροτάτων. Στην τεχνική υδρολογία χρησιμοποιούνται επιτυχώς και άλλοι τύποι κατανομών για την περιγραφή ακροτάτων, πέρα από τις ασυμπτωτικές κατανομές. Για παράδειγμα, η λογαριθμοκανονική κατανομή, οι κατανομές γάμα δύο και τριών παραμέτρων και η κατανομή Log Pearson III πολύ συχνά χρησιμοποιούνται για την περιγραφή υδρολογικών μεγίστων (πλημμυρικών παροχών, καταιγίδων κτλ.)

6.3.1 Κατανομή μεγίστων τύπου I (Gumbel)

Η κατανομή μεγίστων τύπου I προκύπτει όταν οι μεταβλητές Y_i είναι ισόνομες και ανεξάρτητες με διάστημα τιμών το $(-\infty, +\infty)$ και η κοινή συνάρτηση πιθανότητας υπέρβασής τους εμφανίζει, τουλάχιστον από μια τιμή της μεταβλητής και πάνω, εκθετική μείωση, δηλαδή μπορεί να εκφραστεί ως

$$F_{1Y}(y) = 1 - F_Y(y) = e^{-g(y)} \quad (6.35)$$

όπου η $g(y)$ είναι αύξουσα συνάρτηση του y . Αυτός ο τελευταίος όρος ικανοποιείται από τις πιο κοινές κατανομές (π.χ. κανονική, γάμα, κτλ.).

Τα διάφορα χαρακτηριστικά της κατανομής φαίνονται στον Πίν. 6.9. Η κατανομή έχει απλή μαθηματική έκφραση και διαθέτει δύο παραμέτρους. Το μέγεθος γ που εμφανίζεται στην έκφραση της μέσης τιμής της μεταβλητής είναι η σταθερά του Euler.*

Τυπικοί υπολογισμοί

Λόγω της απλής μαθηματικής έκφρασης της συνάρτησης κατανομής, οι τυπικοί υπολογισμοί είναι άμεσοι και δεν προϋποθέτουν τη χρήση πινά-

* Η σταθερά του Euler ορίζεται από τη σχέση

$$\gamma := \lim_{n \rightarrow \infty} \left(1 + \frac{1}{2} + \dots + \frac{1}{n} - \ln n \right) \approx 0.5772156649\dots$$

κων ή αριθμητικών μεθόδων. Η συνάρτηση κατανομής υπολογίζεται άμεσα αν είναι γνωστή η τιμή της μεταβλητής. Επίσης, η αντίστροφη συνάρτηση υπολογίζεται αναλυτικά, και έτσι το u -ποσοστημόριο της κατανομής δίνεται από την

$$x_u = c - \frac{\ln(-\ln u)}{\lambda} \quad (6.36)$$

Πίν. 6.9 Τυπολόγιο της κατανομής Gumbel μεγίστων.

| | |
|----------------------------------|--|
| Συνάρτηση πυκνότητας πιθανότητας | $f_X(x) = \lambda e^{-\lambda(x-c) - e^{-\lambda(x-c)}}$ |
| Συνάρτηση κατανομής | $F_X(x) = e^{-e^{-\lambda(x-c)}}$ |
| Τιμές μεταβλητής | $-\infty < x < \infty$ (συνεχής) |
| Παράμετροι | c : παράμετρος θέσης $\lambda > 0$: παράμετρος κλίμακας |
| Μέση τιμή | $\mu_X = c + \frac{\gamma}{\lambda} = c + \frac{0.5772}{\lambda}$ |
| Διασπορά | $\sigma_X^2 = \frac{\pi^2}{6\lambda^2} = \frac{1.645}{\lambda^2}$ |
| Τρίτη κεντρική ροπή | $\mu_X^{(3)} = \frac{2.404}{\lambda^3}$ |
| Τέταρτη κεντρική ροπή | $\mu_X^{(4)} = \frac{14.6}{\lambda^4}$ |
| Συντελεστής ασυμμετρίας | $C_{s_X} = 1.1396$ |
| Συντελεστής κύρτωσης | $C_{k_X} = 5.4$ |
| Πιθανότερη τιμή | $x_p = c$ |
| Διάμεσος τιμή | $x_{0.5} = c - \frac{\ln(-\ln 0.5)}{\lambda} = c + \frac{0.3665}{\lambda}$ |

Εκτίμηση παραμέτρων

Όπως προκύπτει άμεσα από τις εκφράσεις ροπών του Πίν. 6.9, οι εκτιμήσεις των παραμέτρων της κατανομής με τη μέθοδο των ροπών είναι:

$$\lambda = \frac{\pi}{\sqrt{6}s_X} = \frac{1}{0.78s_X} \quad c = \bar{x} - \frac{\gamma}{\lambda} = \bar{x} - \frac{0.577}{\lambda} = \bar{x} - 0.45s_X \quad (6.37)$$

Ανάλογες εξισώσεις προκύπτουν και από μια άλλη μέθοδο (Gumbel, 1958, σ. 227), η οποία βασίζεται στην προσαρμογή ελαχίστων τετραγώνων της θεωρητικής συνάρτησης κατανομής προς την εμπειρική κατανομή, όπως δίνεται από τη σχέση Weibull. Στις εξισώσεις αυτής της μεθόδου υπεισέρχονται ορισμένες εκφράσεις του μεγέθους του δείγματος n , πινακοποίηση των οποίων έχει δοθεί από τον Gumbel (1958, σ. 228). Αντί των αυθεντικών αυτών εξισώσεων δίνουμε τις ακόλουθες τροποποιημένες προσεγγιστικές εξισώσεις, με τις οποίες αποφεύγεται η χρήση πινάκων:

$$\lambda = \frac{1}{0.78} - \frac{1.57}{(n+1)^{0.65}} \quad c = \bar{x} - \frac{0.577 - \frac{0.53}{(n+2.5)^{0.74}}}{\lambda} \quad (6.38)$$

Το σφάλμα προσέγγισης, σε σχέση με τις αυθεντικές εξισώσεις, είναι μικρότερο του 0.25% για την πρώτη εξίσωση και του 0.10% για τη δεύτερη (για $n \geq 10$).

Για μικρές πιθανότητες υπέρβασης, η δεύτερη αυτή μέθοδος (εξισώσεις (6.38)) δίνει δυσμενέστερες προβλέψεις σε σχέση με αυτές της μεθόδου των ροπών (εξισώσεις (6.37)). Άλλες μέθοδοι εκτίμησης, στις οποίες περιλαμβάνεται και η αρκετά πολυπλοκότερη μέθοδος της μέγιστης πιθανοφάνειας, επισκοπούνται από τον Kite (1988, σ. 96).

Τυπικό σφάλμα ποσοστημορίου και όρια εμπιστοσύνης

Στην περίπτωση που οι παράμετροι της κατανομής υπολογίζονται με τη μέθοδο των ροπών, η σημειακή εκτίμηση του u -ποσοστημορίου της κατανομής Gumbel μπορεί να γραφεί και με την ακόλουθη μορφή, ισόδυναμη της (6.36):

$$\hat{x}_u = \bar{x} - 0.5772 / \lambda - \ln(-\ln u) / \lambda = \bar{x} + k_u s_X \quad (6.39)$$

όπου,

$$k_u = \frac{-0.5772 - \ln(-\ln u)}{\lambda s_X} = -0.45 - 0.78 \ln(-\ln u) \quad (6.40)$$

Σε αυτή την περίπτωση αποδεικνύεται (Gumbel, 1958, σ. 228· Kite, 1988, σ. 103) ότι το τετραγωνικό τυπικό σφάλμα εκτίμησης είναι

$$\varepsilon_X^2 = \text{Var}(\hat{X}_u) = \frac{s_X^2}{n} (1 + 1.1396 k_u + 1.1 k_u^2) \quad (6.41)$$

Κατά συνέπεια, τα όρια εμπιστοσύνης του u -ποσοστημορίου για βαθμό εμπιστοσύνης γ είναι κατά προσέγγιση

$$\hat{x}_{u_{1,2}} = (\bar{x} + k_u s_X) \pm z_{(1+\gamma)/2} \frac{s_X}{\sqrt{n}} \sqrt{1 + 1.1396 k_u + 1.1 k_u^2} \quad (6.42)$$

Χαρτί κατανομής Gumbel

Μπορεί εύκολα να κατασκευαστεί ειδικό χαρτί της κατανομής Gumbel, στο οποίο η συνάρτηση κατανομής ευθειοποιείται. Αρκεί η διαγράμμιση του οριζόντιου άξονα της πιθανότητας να γίνει με βάση το μέγεθος $h = -\ln(-\ln F)$. Η διαγράμμιση του κατακόρυφου άξονα που απεικονίζει τη μεταβλητή είναι κοινή δεκαδική, χωρίς κανένα μετασχηματισμό. Όπως προκύπτει από την (6.36), η απεικόνιση της συνάρτησης κατανομής σε τέτοιους άξονες είναι ευθεία.

Εφαρμογή 6.3.1

Στον Πίν. 6.10 δίνεται το δείγμα των μέγιστων ημερήσιων παροχών του ποταμού Ευήνου στη θέση του υδρομετρικού σταθμού Πόρος Ρηγαίου. Ζητείται η προσαρμογή της κατανομής Gumbel (μεγίστων), καθώς και η εκτίμηση της μέγιστης παροχής εκατονταετίας.

Η μέση τιμή του δείγματος είναι

$$\bar{x} = \sum x / n = 385.1 \text{ m}^3/\text{s}$$

Η τυπική απόκλιση είναι

$$s_X = \sqrt{\sum x^2 / n - \bar{x}^2} = 181.5 \text{ m}^3/\text{s}$$

και ο συντελεστής μεταβλητότητας

Πίν. 6.10 Δείγμα μέγιστης ημερήσιας παροχής (σε m³/s) του ποταμού Ευήνου στη θέση Πόρος Ρηγαίου.

| Υδρολ. έτος | Μέγιστη παροχή | Υδρολ. έτος | Μέγιστη παροχή | Υδρολ. έτος | Μέγιστη παροχή |
|-------------|----------------|-------------|----------------|-------------|----------------|
| 1970-71 | 884 | 1977-78 | 365 | 1984-85 | 317 |
| 1971-72 | 305 | 1978-79 | 502 | 1985-86 | 374 |
| 1972-73 | 215 | 1979-80 | 381 | 1986-87 | 188 |
| 1973-74 | 378 | 1980-81 | 387 | 1987-88 | 192 |
| 1974-75 | 176 | 1981-82 | 525 | 1988-89 | 448 |
| 1975-76 | 430 | 1982-83 | 412 | 1989-90 | 70 |
| 1976-77 | 713 | 1983-84 | 439 | | |

$$\hat{C}_{v_X} = s_X / \bar{x} = 181.5 / 385.1 = 0.47$$

Ο συντελεστής ασυμμετρίας προκύπτει

$$\hat{C}_{s_X} = 0.94$$

δηλαδή δεν απέχει πολύ από τη θεωρητική τιμή 1.14 της κατανομής Gumbel.

Η μέθοδος των ροπών δίνει

$$\lambda = 1 / (0.78 \times 181.5) = 0.00706, c = 385.1 - 0.45 \times 181.5 = 303.4$$

Η μέγιστη ημερήσια παροχή για $T = 100$, ή ισοδύναμα για $u = 1 - 1/100 = 0.99$, είναι

$$x_{0.99} = 303.4 - \ln[-\ln(0.99)] / 0.00706 = 955.0$$

Με βάση την (6.42), για

$$k_u = (955.0 - 385.1) / 181.5 = 3.16, z_{(1+\gamma)/2} = 1.96$$

υπολογίζουμε τα όρια εμπιστοσύνης 95% της τιμής αυτής, τα οποία είναι:

$$\begin{aligned} \hat{x}_{u,2} &\approx 955.0 \pm 1.96 * \frac{181.5}{\sqrt{20}} * \sqrt{1 + 1.1396 * 3.16 + 1.1 * 3.16^2} \\ &\approx 955.0 \pm 313.1 = \begin{cases} 1268.1 \\ 641.9 \end{cases} \end{aligned}$$

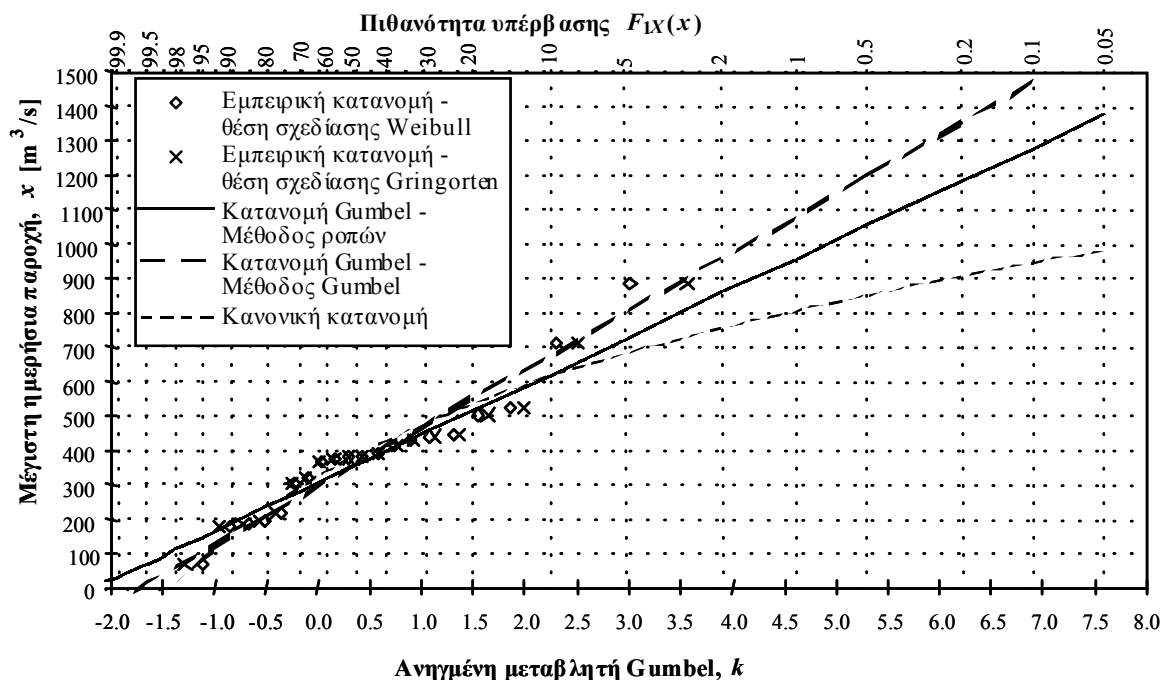
Η μέθοδος του Gumbel με εφαρμογή των προσεγγιστικών εξισώσεων (6.38), για $n = 20$, δίνει

$$\lambda = 0.00587, c = 295.7$$

Η μέγιστη ημερήσια παροχή για $T = 100$ είναι

$$x_{0.99} = 295.7 - \ln[-\ln(0.99)] / 0.00587 = 1079.4$$

Στο Σχ. 6.3 δίνουμε τη γραφική σύγκριση της εμπειρικής συνάρτησης κατανομής και της κατανομής Gumbel, σε χαρτί κατανομής Gumbel. Για την εμπειρική κατανομή χρησιμοποιήθηκαν οι θέσεις σχεδίασης Weibull και Gringorten (βλ. Πίν. 5.7). Παρατηρούμε ότι η μέθοδος εκτίμησης παραμέτρων του Gumbel δίνει καλύτερη προσαρμογή στην περιοχή των μικρών πιθανοτήτων υπέρβασης, που ενδιαφέρουν περισσότερο. Για σύγκριση έχει σχεδιαστεί και η κανονική κατανομή, η ακαταλληλότητα της οποίας για το συγκεκριμένο δείγμα είναι εμφανής.



Σχ. 6.3 Εμπειρική και θεωρητική συνάρτηση κατανομής της μέγιστης ημερήσιας παροχής του Ευήνου στον Πόρο Ρηγανίου (Εφαρμογή 6.3.1) σε χαρτί κατανομής Gumbel (μεγίστων).

6.3.2 Κατανομή ελαχίστων τύπου I (Gumbel)

Κατά ανάλογο τρόπο, η κατανομή ελαχίστων τύπου I προκύπτει όταν οι μεταβλητές Y_i είναι ισόνομες και ανεξάρτητες με διάστημα τιμών το $(-\infty, +\infty)$ και η κοινή συνάρτηση κατανομής τους εμφανίζει, τουλάχιστον από μια τιμή της μεταβλητής και κάτω, εκθετική μείωση.

Τα διάφορα χαρακτηριστικά της κατανομής, που μοιάζουν πολύ με αυτά της αντίστοιχης κατανομής μεγίστων, φαίνονται στον Πίν. 6.11.

Πίν. 6.11 Τυπολόγιο της κατανομής Gumbel ελαχίστων.

| | |
|----------------------------------|--|
| Συνάρτηση πυκνότητας πιθανότητας | $f_X(x) = \lambda e^{\lambda(x-c) - e^{\lambda(x-c)}}$ |
| Συνάρτηση κατανομής | $F_X(x) = 1 - e^{-e^{\lambda(x-c)}}$ |
| Τιμές μεταβλητής | $-\infty < x < \infty$ (συνεχής) |
| Παράμετροι | c : παράμετρος θέσης $\lambda > 0$: παράμετρος κλίμακας |
| Μέση τιμή | $\mu_X = c - \frac{\gamma}{\lambda} = c - \frac{0.5772}{\lambda}$ |
| Διασπορά | $\sigma_X^2 = \frac{\pi^2}{6\lambda^2} = \frac{1.645}{\lambda^2}$ |
| Τρίτη κεντρική ροπή | $\mu_X^{(3)} = -\frac{2.404}{\lambda^3}$ |
| Τέταρτη κεντρική ροπή | $\mu_X^{(4)} = \frac{14.6}{\lambda^4}$ |
| Συντελεστής ασυμμετρίας | $C_{s_X} = -1.1396$ |
| Συντελεστής κύρτωσης | $C_{k_X} = 5.4$ |
| Πιθανότερη τιμή | $x_p = c$ |
| Διάμεσος τιμή | $x_{0.5} = c + \frac{\ln(-\ln 0.5)}{\lambda} = c - \frac{0.3665}{\lambda}$ |

Ο υπολογιστικός χειρισμός της κατανομής είναι παρόμοιος με αυτόν της αντίστοιχης κατανομής μεγίστων. Η αντίστροφη συνάρτηση υπολογίζεται αναλυτικά, και έτσι το u -ποσοστημόριο της κατανομής δίνεται από την

$$x_u = c + \frac{\ln[-\ln(1-u)]}{\lambda} \quad (6.43)$$

Παρόμοιος είναι και ο τρόπος εκτίμησης παραμέτρων. Για παράδειγμα η μέθοδος των ροπών δίνει:

$$\lambda = \frac{\pi}{\sqrt{6}s_X} = \frac{1}{0.78s_X} \quad c = \bar{x} + \frac{\gamma}{\lambda} = \bar{x} + \frac{0.577}{\lambda} = \bar{x} + 0.45s_X \quad (6.44)$$

Γραφικά μπορεί να παρασταθεί στο χαρτί κατανομής Gumbel μεγίστων, με αντιμετάθεση της πιθανότητας υπέρβασης και της πιθανότητας μη υπέρβασης. Μπορεί, όμως, εύκολα να κατασκευαστεί ειδικό χαρτί της κατανομής Gumbel ελαχίστων, αρκεί η διαγράμμιση του οριζόντιου άξονα της πιθανότητας να γίνει με βάση το μέγεθος $h = \ln[-\ln(1-F)]$. Η διαγράμμιση του κατακόρυφου άξονα που απεικονίζει τη μεταβλητή είναι κοινή δεκαδική, χωρίς κανένα μετασχηματισμό.

6.3.3 Κατανομή ελαχίστων τύπου III (Weibull)

Η κατανομή ελαχίστων τύπου III προκύπτει όταν οι μεταβλητές Y_i είναι ισόνομες και ανεξάρτητες με διάστημα τιμών το $(c, +\infty)$ και η κοινή συνάρτηση κατανομής τους μπορεί να εκφραστεί στη γειτονιά του c ως

$$F_Y(y) = \rho (y - c)^{\kappa} \quad (6.45)$$

όπου ρ και κ θετικές σταθερές. Αυτός ο τελευταίος όρος ικανοποιείται από διάφορες κατανομές όπως η γάμα. Αποδεικνύεται ότι σε αυτή την περίπτωση η ασυμπτωτική κατανομή των ελαχίστων* είναι η

$$F_X(x) = 1 - e^{-\left(\frac{x-c}{\alpha-c}\right)^{\kappa}} \quad (6.46)$$

Η κατανομή έχει τρεις παραμέτρους. Στον Πίν. 6.12 φαίνονται τα διάφορα χαρακτηριστικά της κατανομής για την περίπτωση που η παράμετρος θέσης c έχει τιμή μηδέν. Αυτή η απλουστευμένη διπαραμετρική κατανομή είναι η πιο συνηθισμένη στην υδρολογία (π.χ. για την περιγραφή παροχών ξηρασίας), είναι δε γνωστή και ως κατανομή Weibull.

* Η αντίστοιχη κατανομή μεγίστων τύπου III ορίζεται κατ' αναλογία και έχει συνάρτηση κατανομής

$$F_X(x) = e^{-\left(\frac{c-x}{c-\alpha}\right)^{\kappa}}$$

Τυπικοί υπολογισμοί

Λόγω της απλής μαθηματικής έκφρασης της συνάρτησης κατανομής, οι τυπικοί υπολογισμοί είναι άμεσοι και δεν προϋποθέτουν τη χρήση πινάκων ή αριθμητικών μεθόδων. Η συνάρτηση κατανομής υπολογίζεται άμεσα αν είναι γνωστή η τιμή της μεταβλητής. Επίσης, η αντίστροφη συνάρτηση υπολογίζεται αναλυτικά, και έτσι το u -ποσοστημόριο της κατανομής δίνεται από την

$$x_u = \alpha[-\ln(1-u)]^{1/\kappa} \quad (6.47)$$

Πίν. 6.12 Τυπολόγιο της κατανομής Weibull (δύο παραμέτρων).

| | |
|----------------------------------|---|
| Συνάρτηση πυκνότητας πιθανότητας | $f_X(x) = \frac{\kappa}{\alpha} \left(\frac{x}{\alpha}\right)^{\kappa-1} e^{-(x/\alpha)^\kappa}$ |
| Συνάρτηση κατανομής | $F_X(x) = 1 - e^{-(x/\alpha)^\kappa}$ |
| Τιμές μεταβλητής | $0 < x < \infty$ (συνεχής) |
| Παράμετροι | $\alpha > 0$: παράμετρος κλίμακας $\kappa > 0$: παράμετρος σχήματος |
| Μέση τιμή | $\mu_X = \alpha \Gamma\left(1 + \frac{1}{\kappa}\right)$ |
| Διασπορά | $\sigma_X^2 = \alpha^2 \left\{ \Gamma\left(1 + \frac{2}{\kappa}\right) - \left[\Gamma\left(1 + \frac{1}{\kappa}\right) \right]^2 \right\}$ |
| Τρίτη ροπή περί την αρχή | $m_X^{(3)} = \alpha^3 \Gamma\left(1 + \frac{3}{\kappa}\right)$ |
| Συντελεστής μεταβλητότητας | $C_{v_x} = \Gamma\left(1 + \frac{2}{\kappa}\right) / \left[\Gamma\left(1 + \frac{1}{\kappa}\right) \right]^2 - 1$ |
| Πιθανότερη τιμή | $x_p = \alpha(1 - 1/\kappa)^{1/\kappa}$ (για $\kappa > 1$) |
| Διάμεσος | $x_{0.5} = \alpha(\ln 2)^{1/\kappa}$ |

Εκτίμηση παραμέτρων

Όπως προκύπτει άμεσα από τις εκφράσεις ροπών του Πίν. 6.12, οι εξισώσεις της μεθόδου των ροπών είναι:

$$\frac{\Gamma(1+2/\kappa)}{\Gamma^2(1+1/\kappa)} = \frac{s_X^2}{\bar{x}^2} + 1 \quad \alpha = \frac{\bar{x}}{\Gamma(1+1/\kappa)} \quad (6.48)$$

Η επίλυση της πρώτης εξίσωσης ως προς κ γίνεται μόνο αριθμητικά.

Μια απλούστερη μέθοδος προκύπτει αν χρησιμοποιηθεί ο μετασχηματισμός $Y = \ln X$. Εύκολα συμπεραίνουμε ότι η συνάρτηση κατανομής της Y είναι

$$F_Y(y) = 1 - e^{-\kappa(y - \ln \alpha)} \quad (6.49)$$

δηλαδή είναι η συνάρτηση κατανομής ελαχίστων τύπου I με παράμετρο θέσης $\ln \alpha$ και παράμετρο κλίμακας κ . Επομένως, μπορούμε να χρησιμοποιήσουμε τις εξισώσεις της τελευταίας προκειμένου να εκτιμήσουμε τις άγνωστες παραμέτρους. Έτσι παίρνουμε

$$\kappa = \frac{1}{0.78 s_Y} \quad \alpha = e^{\bar{y} + 0.45 s_Y} \quad (6.50)$$

όπου \bar{y} και s_Y η δειγματική μέση τιμή και τυπική απόκλιση, αντίστοιχα, των λογαρίθμων της μεταβλητής.

Για τις εξισώσεις της τριπαραμετρικής κατανομής ελαχίστων τύπου III ο ενδιαφερόμενος αναγνώστης παραπέμπεται στον Kite (1988, σ. 154).

Χαρτί κατανομής Weibull

Μπορεί εύκολα να κατασκευαστεί ειδικό χαρτί της κατανομής Weibull, στο οποίο η συνάρτηση κατανομής ευθειοποιείται. Αρκεί η διαγράμμιση του οριζόντιου άξονα της πιθανότητας να γίνει με βάση το μέγεθος $h = \ln[-\ln(1-F)]$ (όπως και στο χαρτί Gumbel ελαχίστων) και του κατακόρυφου άξονα που απεικονίζει τη μεταβλητή με βάση το μέγεθος $v = \ln x$ (λογαριθμική κλίμακα). Όπως προκύπτει από την (6.47), η απεικόνιση της συνάρτησης κατανομής σε τέτοιους άξονες είναι ευθεία. Το χαρτί Weibull χρησιμοποιείται ακόμη για την παράσταση της εκθετικής κατανομής, η οποία είναι ειδική περίπτωση και της γάμα και της Weibull.

Εφαρμογή 6.3.3

Στον Πίν. 6.13 δίνεται το δείγμα των ελάχιστων ημερήσιων παροχών του ποταμού Ευήνου στη θέση του υδρομετρικού σταθμού Πόρος Ρηγα-

νίου. Ζητείται η προσαρμογή των κατανομών Gumbel (ελαχίστων) και Weibull, καθώς και η εκτίμηση της ελάχιστης παροχής εικοσαετίας.

Πίν. 6.13 Δείγμα ελάχιστης ημερήσιας παροχής (σε m³/s) του ποταμού Ευήνου στη θέση Πόρος Ρηγανίου.

| Υδρολ. έτος | Ελάχιστη παροχή | Υδρολ. έτος | Ελάχιστη παροχή | Υδρολ. έτος | Ελάχιστη παροχή |
|-------------|-----------------|-------------|-----------------|-------------|-----------------|
| 1970-71 | 0.00 | 1977-78 | 2.14 | 1984-85 | 0.54 |
| 1971-72 | 2.19 | 1978-79 | 2.00 | 1985-86 | 0.54 |
| 1972-73 | 2.66 | 1979-80 | 1.93 | 1986-87 | 1.70 |
| 1973-74 | 2.13 | 1980-81 | 2.29 | 1987-88 | 1.70 |
| 1974-75 | 1.28 | 1981-82 | 2.66 | 1988-89 | 0.32 |
| 1975-76 | 0.56 | 1982-83 | 2.87 | 1989-90 | 1.37 |
| 1976-77 | 0.13 | 1983-84 | 1.88 | | |

Η μέση τιμή του δείγματος είναι

$$\bar{x} = \sum x / n = 1.545 \text{ m}^3/\text{s}$$

Η τυπική απόκλιση είναι

$$s_X = \sqrt{\sum x^2 / n - \bar{x}^2} = 0.878 \text{ m}^3/\text{s}$$

και ο συντελεστής μεταβλητότητας

$$\hat{C}_{v_X} = s_X / \bar{x} = 0.878 / 1.545 = 0.57$$

Ο συντελεστής ασυμμετρίας προκύπτει

$$\hat{C}_{s_X} = -0.40$$

Η αρνητική αυτή τιμή του συντελεστή ασυμμετρίας είναι αναμενόμενη για ένα δείγμα ελαχίστων παροχών.

Για την κατανομή Gumbel, η μέθοδος των ροπών δίνει

$$\lambda = 1 / (0.78 \times 0.878) = 1.460, c = 1.545 + 0.45 \times 0.878 = 1.940$$

. Η ελάχιστη ημερήσια παροχή για $T = 20$, ή ισοδύναμα για $u = 1/20 = 0.05$, είναι

$$x_{0.05} = 1.940 + \ln[-\ln(1 - 0.05)] / 1.460 = -0.09$$

Φυσικά, η αρνητική τιμή δεν έχει νόημα και γι' αυτό θα θεωρήσουμε ότι το ζητούμενο μέγεθος είναι μηδέν.

Για την κατανομή Weibull η μέθοδος των ροπών δίνει

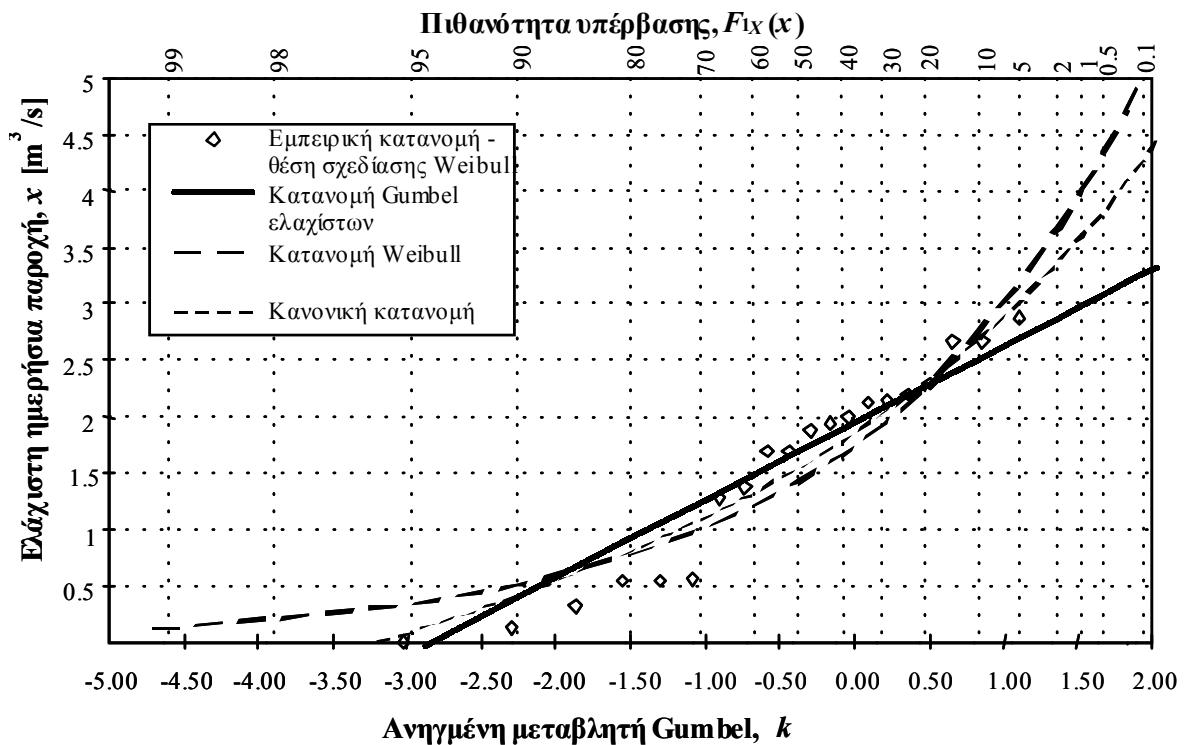
$$\kappa = 1.826, \alpha = 1.738$$

Η εύρεση της τιμής του κ προϋποθέτει αριθμητική επίλυση της πρώτης από τις εξισώσεις (6.48). Για έλεγχο δίνουμε τις τιμές των μεγεθών που υπεισέρχονται στην εν λόγω εξίσωση:

$$\Gamma(1 + 2/\kappa) = \Gamma(2.095) = 1.044, \Gamma(1 + 1/\kappa) = \Gamma(1.548) = 0.889$$

Η μέθοδος εκτίμησης παραμέτρων με λογαριθμικό μετασχηματισμό της μεταβλητής εδώ δεν είναι εφαρμόσιμη, λόγω της παρουσίας της μηδενικής τιμής στο συγκεκριμένο δείγμα. Η ελάχιστη ημερήσια παροχή για $T = 20$ είναι

$$x_{0.05} = 1.738 \times [-\ln(1-0.05)]^{1/1.826} = 0.342 \text{ m}^3/\text{s}$$



Σχ. 6.4 Εμπειρική και θεωρητική συνάρτηση κατανομής της ελάχιστης ημερήσιας παροχής του Ευήνου στον Πόρο Ρηγανίου (Εφαρμογή 6.0) σε χαρτί κατανομής Gumbel (ελαχίστων).

Στο Σχ. 6.4 δίνουμε τη γραφική σύγκριση της εμπειρικής συνάρτησης κατανομής και των δύο παραπάνω θεωρητικών κατανομών, σε χαρτί κατανομής Gumbel ελαχίστων. Για την εμπειρική κατανομή χρησιμοποιήθηκε η θέση σχεδίασης Weibull. Παρατηρούμε ότι καμία από τις δύο θεωρητικές κατανομές δεν προσαρμόζεται πολύ καλά στο

δείγμα, αλλά πάντως η κατανομή Gumbel είναι σχετικά καλύτερη, ιδίως στην περιοχή των μικρών πιθανοτήτων μη υπέρβασης, που ενδιαφέρουν περισσότερο. Το γεγονός ότι η κατανομή Weibull ορίζεται μόνο για θετικές τιμές της μεταβλητής αποτελεί κατ' αρχήν συγκριτικό θεωρητικό πλεονέκτημα της κατανομής αυτής. Ωστόσο, για το συγκεκριμένο δείγμα αυτό είναι εμφανώς μειονέκτημα γιατί, όπως φαίνεται στο Σχ. 6.4, οδηγεί σε μεγάλη απόκλιση από την εμπειρική συνάρτηση κατανομής. Καλύτερη συμπεριφορά έχει η κατανομή Gumbel, με την προϋπόθεση όμως ότι απαλείφουμε εκείνο το τμήμα της που αντιστοιχεί σε αρνητικές τιμές της μεταβλητής. Για σύγκριση έχει σχεδιαστεί και η κανονική κατανομή, η οποία τοποθετείται ανάμεσα στις δύο κατανομές ελαχίστων.

6.4 Άλλες χρήσιμες κατανομές

6.4.1 Κατανομή βήτα δύο παραμέτρων

Η κατανομή βήτα είναι μια πολύ βασική κατανομή της θεωρίας πιθανοτήτων, η οποία χρησιμοποιείται αρκετά συχνά στην υδρολογία, κυρίως για δεσμευμένες κατανομές υδρολογικών μεταβλητών. Βασική της ιδιότητα είναι ότι το πεδίο ορισμού της μεταβλητής είναι το φραγμένο διάστημα $[0, 1]$ (σε αντίθεση με όλες τις άλλες κατανομές που εξετάστηκαν, των οποίων το πεδίο ορισμού εκτείνεται μέχρι το $+\infty$).

Η κατανομή έχει δύο παραμέτρους σχήματος α και β (εύκολα μπορεί να εισαχθεί και μια παράμετρος κλίμακας). Ανάλογα με τις τιμές των παραμέτρων η συνάρτηση πυκνότητας πιθανότητας μπορεί να αποκτήσει διάφορα σχήματα. Για $\alpha = \beta = 1$ η κατανομή μεταπίπτει στην ομοιόμορφη, ενώ για $\alpha = 1$ και $\beta = 2$ (ή $\alpha = 2$ και $\beta = 1$) μεταπίπτει στην τριγωνική με αρνητική (θετική) ασυμμετρία. Για $\alpha < 1$ (ή $\beta < 1$) η συνάρτηση πυκνότητας πιθανότητας απειρίζεται στο σημείο $x = 0$ ($x = 1$). Για $\alpha > 1$ και $\beta > 1$ η συνάρτηση πυκνότητας πιθανότητας αποκτά κωδωνοειδές σχήμα. Στον Πίν. 6.14 δίνονται τα διάφορα χαρακτηριστικά της κατανομής.

Πίν. 6.14 Τυπολόγιο της κατανομής βήτα 2 παραμέτρων.

| | |
|----------------------------------|---|
| Συνάρτηση πυκνότητας πιθανότητας | $f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$ |
| Συνάρτηση κατανομής | $F_X(x) = \int_0^x f_X(s) ds$ |
| Τιμές μεταβλητής | $0 < x < 1$ (συνεχής) |
| Παράμετροι | $\alpha, \beta > 0$: παράμετροι σχήματος |
| Μέση τιμή | $\mu_X = \frac{\alpha}{\alpha + \beta}$ |
| Διασπορά | $\sigma_X^2 = \frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$ |
| Τρίτη ροπή περί την αρχή | $m_X^{(3)} = \frac{\alpha(\alpha + 1)(\alpha + 2)}{(\alpha + \beta)(\alpha + \beta + 1)(\alpha + \beta + 2)}$ |
| Συντελεστής μεταβλητότητας | $C_{v_X} = \sqrt{\frac{\beta}{\alpha(\alpha + \beta + 1)}}$ |
| Πιθανότερη τιμή | $x_p = \frac{\alpha - 1}{\alpha + \beta - 2}$ (για $\alpha, \beta > 1$) |

6.4.2 Κατανομή Pareto

Η γενικευμένη κατανομή Pareto* είναι μια απλή και χρήσιμη κατανομή που έχει χρησιμοποιηθεί και στην τεχνική υδρολογία για την περιγραφή μεταβλητών που ξεπερνούν ένα δεδομένο κατώφλι. Τα κύρια χαρακτηριστικά της κατανομής φαίνονται στον Πίν. 6.15.

* Την κατανομή αυτή εισήγαγε ο ιταλός οικονομολόγος V. Pareto (1848-1923) για τη μελέτη του ατομικού εισοδήματος.

Πίν. 6.15 Τυπολόγιο της κατανομής Pareto.

| | |
|---|---|
| Συνάρτηση πυκνότητας πιθανότητας | $f_X(x) = \frac{1}{\alpha} \left(1 - \kappa \frac{x-c}{\alpha} \right)^{\frac{1}{\kappa}-1}$ |
| Συνάρτηση κατανομής | $F_X(x) = 1 - \left(1 - \kappa \frac{x-c}{\alpha} \right)^{\frac{1}{\kappa}}$ |
| Τιμές μεταβλητής | Για $\kappa < 0$, $c \leq x < \infty$ Για $\kappa > 0$, $c \leq x < c + \alpha / \kappa$ (συνεχής) |
| Παράμετροι | c : παράμετρος θέσης $\alpha > 0$: παράμετρος κλίμακας κ : παράμετρος σχήματος |
| Μέση τιμή | $\mu_X = c + \frac{\alpha}{1 + \kappa}$ |
| Διασπορά (ορίζεται για $\kappa > -0.5$) | $\sigma_X^2 = \frac{\alpha^2}{(1 + \kappa)^2 (1 + 2\kappa)}$ |
| Τρίτη κεντρική ροπή (ορίζεται για $\kappa > -0.33$) | $\mu_X^{(3)} = \frac{2\alpha^3(1 - \kappa)}{(1 + \kappa)^3 (1 + 2\kappa)(1 + 3\kappa)}$ |
| Συντελεστής ασυμμετρίας | $C_{s_X} = \frac{2(1 - \kappa)\sqrt{1 + 2\kappa}}{1 + 3\kappa}$ |
| Πιθανότερη τιμή | $x_p = c$ |
| Διάμεσος | $x_{0.5} = c + \frac{\alpha}{\kappa} (1 - 0.5^\kappa)$ |

Παράρτημα 6.Α: Αριθμητικές προσεγγίσεις της κανονικής κατανομής

Παίρνοντας υπόψη ότι οι τιμές του $u = F_Z(z)$ κυμαίνονται στο διάστημα $[0, 1]$ και ότι η κανονική κατανομή είναι συμμετρική, μπορούμε να εκφράσουμε προσεγγιστικά την παράγωγο της συνάρτησης $z(u)$ (δηλαδή της αντίστροφης της συνάρτησης κατανομής) με την εξίσωση

$$z'(u) = \frac{dz}{du} = \kappa u^{\alpha-1} + \kappa (1-u)^{\alpha-1} \quad (6.51)$$

όπου κ και α σταθερές προς προσδιορισμό. Μετά την ολοκλήρωση ως προς u και τον προσδιορισμό των σταθερών προκύπτει η ζητούμενη έκφραση του u -ποσοστημορίου της τυποποιημένης κανονικής μεταβλητής. Έτσι, οι Stedinger et al. (1993) προτείνουν την προσεγγιστική εξίσωση

$$z_u = \frac{u^{0.135} - (1-u)^{0.135}}{0.1975} \quad (6.52)$$

Στο κείμενο αυτό προτείνεται η εφαρμογή της ακόλουθης, ελαφρώς διαφοροποιημένης, προσεγγιστικής εξίσωσης, η οποία, μετά από αριθμητική διερεύνηση, προέκυψε να δίνει μεγαλύτερη ακρίβεια:

$$z_u = 5.53 \left[u^{0.12} - (1-u)^{0.12} \right] \quad (6.53)$$

Η (6.53) δίνει καλές προσεγγίσεις για $0.0001 \leq u = F_Z(z_u) \leq 0.9999$, δίνοντας απόλυτο σφάλμα (οριζόμενο ως $e = |z_u - \tilde{z}_u|$, όπου \tilde{z}_u η ορθή τιμή) μικρότερο της τιμής 0.03 σε οποιοδήποτε σημείο του παραπάνω πεδίου εφαρμογής της προσέγγισης. Η (6.53) πρέπει να συνδυαστεί με την (6.1) προκειμένου να υπολογιστεί το u -ποσοστημόριο της x .

Για το αντίστροφο πρόβλημα, δηλαδή για τον υπολογισμό του u για δεδομένο x ή z , μπορούμε και πάλι να κάνουμε χρήση της (6.53), αλλά ωστόσο απαιτούνται δοκιμές, δεδομένου ότι δεν επιλύεται άμεσα ως προς u . Για τη συστηματοποίηση των υπολογισμών προτείνεται η χρήση της ακόλουθης εξίσωσης που αποτελεί κατάλληλη τροποποίηση της εξίσωσης Newton-Raphson:

$$u = \begin{cases} u_0 e^{t_1} & \text{αν } t_1 < 0 \\ 1 - (1 - u_0) e^{-t_2} & \text{διαφορετικά} \end{cases} \quad (6.54)$$

όπου

$$t_1 = \frac{z - z_0}{z'_0 u_0} \quad t_2 = \frac{z - z_0}{z'_0 (1 - u_0)} \quad (6.55)$$

ενώ u_0 είναι μια αρχική προσέγγιση του u και z'_0 η παράγωγος του z στη θέση u_0 , που υπολογίζεται από την εξίσωση

$$z'_0 = 0.664 [u_0^{-0.88} + (1 - u_0)^{-0.88}] \quad (6.56)$$

Η σειρά των υπολογισμών είναι η ακόλουθη:

1. Υποτίθεται μια αρχική τιμή $u_0 = 0.5$.
2. Υπολογίζεται η τιμή του z'_0 από την (6.56).
3. Υπολογίζεται το t_1 από την (6.55), και αν προκύψει θετικό υπολογίζεται και το t_2 .
4. Υπολογίζεται η νέα προσέγγιση του u από την (6.54).
5. Επαναλαμβάνονται τα βήματα 2-4 θέτοντας ως u_0 τη νέα προσέγγιση, μέχρι η διαφορά $u - u_0$ να γίνει αμελητέα (συνήθως αυτό επιτυγχάνεται σε 2-3 δοκιμές).

Παράρτημα 6.B: Αριθμητικές προσεγγίσεις της κατανομής γάμα

Μια πολύ συνηθισμένη προσέγγιση είναι αυτή του μετασχηματισμού Wilson-Hilferty. Σύμφωνα με αυτή, ο υπολογισμός της τυποποιημένης μεταβλητής γάμα k_u ανάγεται στον υπολογισμό της τυποποιημένης κανονικής μεταβλητής z_u , σύμφωνα με την εξίσωση

$$k_u = \frac{2}{C_{s_x}} \left[\left(1 + \frac{C_{s_x} z_u}{6} - \frac{C_{s_x}^2}{36} \right)^3 - 1 \right] \quad (6.57)$$

Ο Kirby (1972) ανέπτυξε την ακόλουθη τροποποιημένη έκφραση του μετασχηματισμού Wilson-Hilferty, η οποία έχει τύχει ευρείας εφαρμογής στην τεχνική υδρολογία

$$k_u = A \left[\max \left(H, 1 + \frac{Gz_u}{6} - \frac{G^2}{36} \right)^3 - B \right] \quad (6.58)$$

όπου

$$A = \max(2 / C_{s_x}, 0.4) \quad (6.59)$$

$$B = 1 + 0.0144 \max(0, C_{s_x} - 2.25)^2 \quad (6.60)$$

$$G = C_{s_x} - 0.063 \max(0, C_{s_x} - 1)^{1.85} \quad (6.61)$$

$$H = \left[B - (2 / C_{s_x}) / A \right]^{1/3} \quad (6.62)$$

Η εν λόγω προσέγγιση θεωρείται ικανοποιητική για συντελεστή ασυμμετρίας στο διάστημα $0.25 \leq C_{s_x} \leq 9.75$. Μειονέκτημα και του αρχικού και του τροποποιημένου μετασχηματισμού Wilson-Hilferty είναι η σύνδεσή τους με την κανονική μεταβλητή z_u , ο υπολογισμός της οποίας δεν είναι άμεσος αλλά απαιτεί νέους αριθμητικούς υπολογισμούς ή χρήση πινάκων (βλ. εδάφιο 6.6.1.1).

Παρακάτω παρουσιάζεται μια νέα μέθοδος η οποία μπορεί να υπολογίσει άμεσα το u -ποσοστημόριο της κατανομής γάμα. Η μέθοδος αυτή θεμελιώνεται στην ακόλουθη προσέγγιση (ανάλογη της (6.51)) της παραγώγου της $x(u)$

$$x'(u) = \frac{dx}{du} = \frac{\mu}{\lambda} u^{\alpha-1} + \frac{\nu}{\lambda} (1-u)^{\beta-1} \quad (6.63)$$

όπου λ είναι η παράμετρος κλίμακας της κατανομής ενώ μ , ν , α , β και ξ είναι συντελεστές προς προσδιορισμό που εξαρτώνται από την παράμε-

τρο κ . Ολοκληρώνοντας παίρνουμε την ακόλουθη σχέση προσδιορισμού του u -ποσοστημορίου

$$x_u = \begin{cases} -\frac{\ln(1-u)}{\lambda} & \kappa = 1 \\ \frac{\mu}{\lambda\alpha} u^\alpha + \frac{\nu}{\lambda\beta} [\xi - (1-u)^\beta] & \kappa \neq 1 \end{cases} \quad (6.64)$$

Για τον προσδιορισμό των συντελεστών μ , ν , α , β και ξ προτείνουμε τις ακόλουθες σχετικά απλές εξισώσεις

$$\mu = 0.6(\sqrt{\kappa} - 1) - (1/\sqrt{\kappa} - 1) \quad (6.65)$$

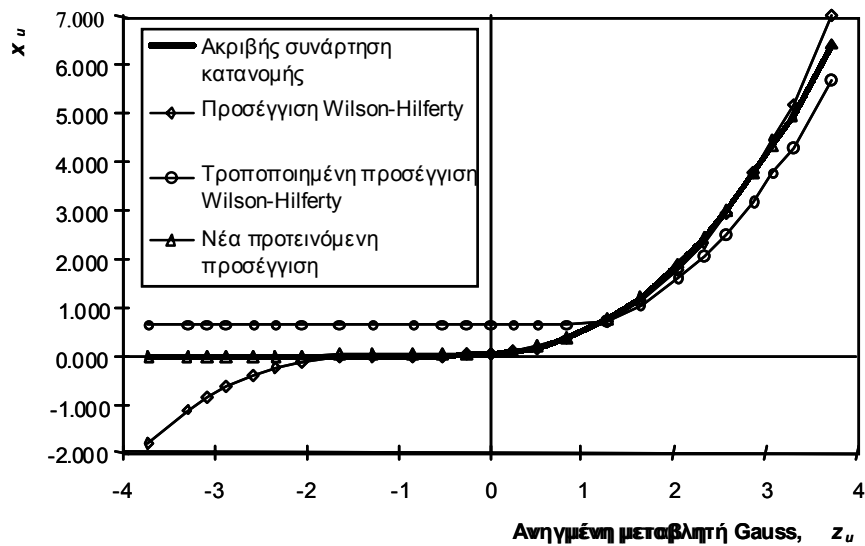
$$\nu = 0.6(\sqrt{\kappa} - 1) + 0.01(\kappa - 1) + 1 \quad (6.66)$$

$$\alpha = 0.6 / \sqrt{\kappa} + 0.08 \quad (6.67)$$

$$\beta = 0.0234 \ln \kappa \quad (6.68)$$

$$\xi = \begin{cases} 1 & \kappa \leq 1 \\ 31e^{-11.6(\kappa-1)^{-0.25}} & \kappa > 1 \end{cases} \quad (6.69)$$

Η παραπάνω προσέγγιση είναι ικανοποιητική για $0.2 \leq \kappa \leq 100$ ($0.2 \leq C_s \leq 4.5$) και $0.0001 \leq F(x) \leq 0.9999$. Το σφάλμα που προκύπτει από την εφαρμογή της (6.64), οριζόμενο ως $e = |x_u - \hat{x}_u| / \sigma_x$, δεν υπερβαίνει την τιμή 0.11 σε κανένα σημείο του πεδίου εφαρμογής της προσέγγισης (δηλαδή του χώρου που ορίζεται από τις πιο πάνω ανισώσεις). Το σφάλμα αυτό είναι μικρότερο από τα αντίστοιχα σφάλματα των προσεγγίσεων Wilson-Hilferty (τυπικής και τροποποιημένης, εξισώσεις (6.57) και (6.58), αντίστοιχα) για $\kappa \leq 4$ ($C_s \geq 1$) (βλ. Σχ. 6.5).



Σχ. 6.5 Σύγκριση τριών προσεγγίσεων της κατανομής γάμα για $\kappa = 0.25$ ($C_s = 4$) και $\lambda = 1$. Η πλησιέστερη προς την ακριβή συνάρτηση κατανομής νέα προτεινόμενη προσέγγιση δίνεται από την (6.64).

Το αντίστροφο πρόβλημα, δηλαδή ο υπολογισμός του u για δεδομένο x , είναι απλό στην περίπτωση που $\kappa = 1$, δεδομένου ότι

$$u = 1 - e^{-\lambda x} \quad \kappa = 1 \quad (6.70)$$

Για $\kappa \neq 1$ μπορούμε να κάνουμε χρήση της (6.64), αλλά ωστόσο απαιτούνται δοκιμές, αφού δεν επιλύεται άμεσα ως προς u . Για τη συστηματοποίηση των υπολογισμών προτείνεται η χρήση της ακόλουθης εξίσωσης, που αποτελεί κατάλληλη τροποποίηση της εξίσωσης Newton-Raphson:

$$u = \begin{cases} u_0 e^{t_1} & \text{αν } t_1 < -1/\kappa \text{ ή } t_2 < \ln(1-u_0) \\ 1 - (1 - u_0) e^{-t_2} & \text{διαφορετικά} \end{cases} \quad (6.71)$$

όπου

$$t_1 = \frac{x - x_0}{x'_0 u_0} \quad t_2 = \frac{x - x_0}{x'_0 (1 - u_0)} \quad (6.72)$$

u_0 είναι μια αρχική προσέγγιση της u_0 και x'_0 η παράγωγος του x στη θέση u_0 , που υπολογίζεται από την εξίσωση (6.63). Η σειρά των υπολογισμών είναι η ακόλουθη:

1. Υποτίθεται μια αρχική τιμή $u_0 = 0.5$.

2. Υπολογίζεται η τιμή του z'_0 από την (6.63).
3. Υπολογίζονται τα t_1 και t_2 από την (6.72).
4. Υπολογίζεται η νέα προσέγγιση του u από την (6.71).
5. Επαναλαμβάνονται τα βήματα 2-4 θέτοντας ως u_0 τη νέα προσέγγιση, μέχρι η διαφορά $u - u_0$ να γίνει αμελητέα (συνήθως αυτό επιτυγχάνεται σε 2-5 δοκιμές).

Κεφάλαιο 7 Ανάλυση δύο τυχαίων μεταβλητών - Εκτιμήσεις ελάχιστων τετραγώνων

Στα κεφάλαια 5 και 6 δώσαμε τη μεθοδολογία για την ανάλυση μίας υδρολογικής μεταβλητής καθώς και τις πιο συνηθισμένες συναρτήσεις κατανομής που χρησιμοποιούνται ως μοντέλα για τις μεταβλητές της τεχνικής υδρολογίας. Η ανάλυση μίας μεταβλητής αποτελεί το θεμέλιο της ανάλυσης διακινδύνευσης των έργων και των διαδικασιών που σχετίζονται με τους υδατικούς πόρους και την προστασία από υδρολογικούς κινδύνους. Σε αυτό το κεφάλαιο θα επεκταθούμε στην ταυτόχρονη ανάλυση δύο (ή και περισσότερων) τυχαίων μεταβλητών. Η ανάλυση δύο μεταβλητών δεν χρησιμοποιείται τόσο στην ανάλυση διακινδύνευσης. Κυρίως χρησιμοποιείται για τη διαπίστωση και τη μαθηματική διατύπωση της σχέσης που συνδέει δύο υδρολογικές μεταβλητές X και Y , η οποία αξιοποιείται κυρίως σε εφαρμογές που έχουν σχέση με την επεξεργασία της υδρολογικής πληροφορίας. Για παράδειγμα, οι μεταβλητές X και Y μπορεί να συμβολίζουν το ύψος βροχής σε δύο διαφορετικά σημεία μιας λεκάνης, ή η μεταβλητή X μπορεί να συμβολίζει την επιφανειακή βροχή μιας λεκάνης και η Y την αντίστοιχη απορροή, κοκ. Μια μαθηματική σχέση ανάμεσα στις μεταβλητές X και Y μας επιτρέπει να συμπληρώσουμε τυχόν κενά στη σειρά των μετρήσεων της Y , αξιοποιώντας τις μετρήσεις της X . Μας επιτρέπει επίσης να ελέγξουμε τη συνέπεια των ταυτόχρονων μετρήσεων των X και Y και να εντοπίσουμε πιθανά σφάλματα στις μετρήσεις.

Η μαθηματική περιγραφή αυτού του είδους των σχέσεων είναι στατιστικού και όχι φυσικού χαρακτήρα, δηλαδή βασίζεται στα διαθέσιμα δείγματα των μεταβλητών και όχι στο φυσικό περιεχόμενο των μεταβλητών και των αλληλοσυνδέσεών τους. Οι σχέσεις που θα μελετήσουμε είναι κατ' αρχήν γραμμικές, δηλαδή της μορφής

$$Y = a + bX \quad (7.1)$$

όπου a και b αριθμητικές σταθερές, γνωστές ως *περίλημμα* (intercept) και *κλίση* (slope) της ευθείας, αντίστοιχα. Θα δώσουμε όμως αρκετά στοιχεία και για μη γραμμικές σχέσεις. Επίσης, θα δώσουμε γενικά στοιχεία για την μαθηματική διατύπωση σχέσεων ανάμεσα σε περισσότερες από δύο μεταβλητές.

Στην απλούστερη περίπτωση, το πρόβλημα της ανάλυσης δύο μεταβλητών περιλαμβάνει (α) την προσαρμογή του μοντέλου (7.1), δηλαδή την εκτίμηση των παραμέτρων a και b , (β) τον έλεγχο του πόσο ισχυρό είναι το μοντέλο, και (γ) την εφαρμογή του μοντέλου για την εκτίμηση τιμών της Y από τιμές της X . Η προσαρμογή και ο έλεγχος του μοντέλου στηρίζονται σε ένα παρατηρημένο δείγμα n ζευγών (ταυτόχρονων) τιμών των μεταβλητών (x_i, y_i) . Για την καλύτερη κατανόηση της μεθόδου ξεκινούμε την παρουσίαση δίνοντας την προσδιοριστική (γεωμετρική και αναλυτική) αντιμετώπιση του προβλήματος και στη συνέχεια προχωρούμε στην πιθανοτική προσέγγισή του. Η πρώτη αντιμετωπίζει τα (x_i, y_i) ως ζεύγη γνωστών αριθμών, χωρίς να τα συναρτά με τις τυχαίες μεταβλητές X και Y . Όπως θα δούμε, τα αποτελέσματα και των δύο προσεγγίσεων συνδέονται στενά.

Θα πρέπει να πούμε ότι και η πιθανοτική προσέγγιση στο πρόβλημα δεν είναι μονοσήμαντη αλλά έχει διάφορες ερμηνείες και όψεις. Στο κείμενο αυτό περιοριζόμαστε σε μια συνοπτική περιγραφή του αντικειμένου, δίνοντας πάντως στοιχεία από τρεις διαφορετικές όψεις (στις ενότητες 7.1, 7.4 και 7.5, αντίστοιχα). Ο αναγνώστης που ενδιαφέρεται για πληρέστερη παρουσίασή του παραπέμπεται σε συγγράμματα γενικής στατιστικής (π.χ. Papoulis, 1990, σσ. 388-413· Benjamin and Cornell, 1970, σσ. 419-440· Spiegel, 1977, σσ. 258-305) αλλά και σε εξειδικευμένα συγγράμματα (π.χ. Draper and Smith, 1981).

7.1 Τυπική γραμμική παλινδρόμηση

7.1.1 Προσδιοριστική αντιμετώπιση

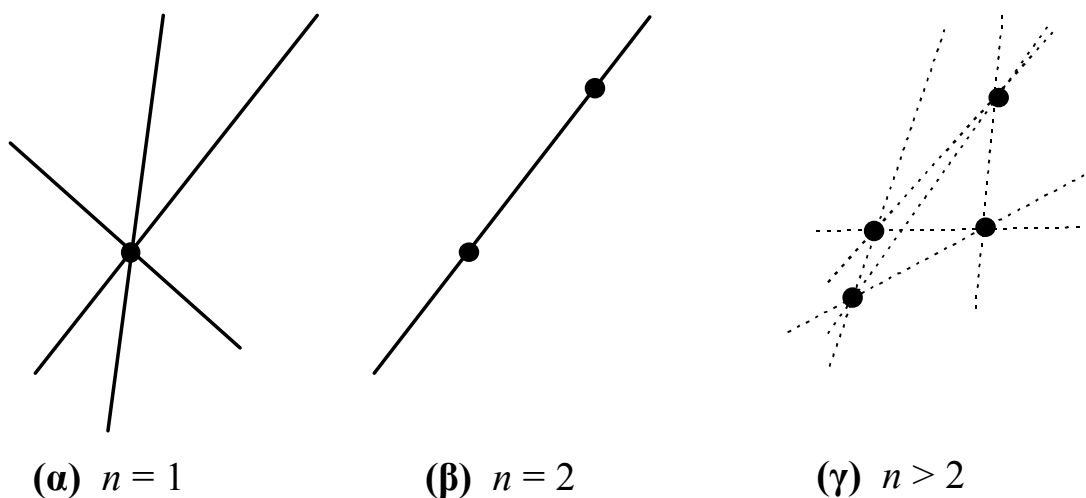
Από προσδιοριστική άποψη το πρόβλημα τίθεται κατ' αρχήν ως εξής: Δίνονται n ζεύγη αριθμών (x_i, y_i) και ζητείται να προσαρμοστεί σε αυτά μια ευθεία της μορφής

$$y = a + b x \quad (7.2)$$

Τα ζεύγη αριθμών (x_i, y_i) μπορούν να θεωρηθούν ως σημεία A_i στο επίπεδο, πράγμα που μας επιτρέπει να δώσουμε μια πρώτη γεωμετρική ερμηνεία στο πρόβλημα. Η γεωμετρική ερμηνεία παρατίθεται για λόγους αμεσότερης κατανόησης και όχι για να χρησιμοποιηθεί στην πράξη.

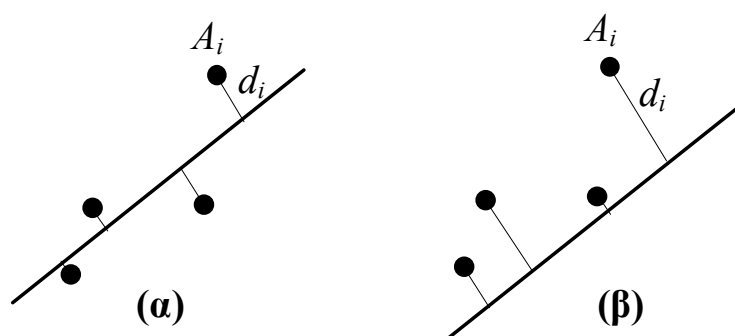
Γεωμετρική ερμηνεία

Ο όρος *προσαρμογή ευθείας* δεν είναι ακόμη αρκετά σαφής. Ας θεωρήσουμε κατ' αρχήν ότι σημαίνει την κατασκευή μιας ευθείας που περνά από όλα τα σημεία. Στην περίπτωση αυτή, όπως είναι γνωστό, το πρόβλημα έχει μία και μοναδική λύση για $n = 2$ και μόνο (Σχ. 7.1(β)). Για $n = 1$ υπάρχουν άπειρες λύσεις (Σχ. 7.1(α)), ενώ για $n > 2$ το πρόβλημα δεν έχει λύση, δηλαδή δεν υπάρχει ευθεία που να περνά ταυτόχρονα απ' όλα τα σημεία (Σχ. 7.1(γ)), εκτός από την ειδική περίπτωση που όλα τα σημεία βρίσκονται σε ευθεία.



Σχ. 7.1 Ακριβής γεωμετρική λύση στο πρόβλημα της κατασκευής ευθείας που περνά από n σημεία.

Φυσικά, η περίπτωση που ενδιαφέρει από πρακτική άποψη είναι η τελευταία. Για να την αντιμετωπίσουμε θα πρέπει να εγκαταλείψουμε την απαίτηση ακριβούς λύσης, δηλαδή την αναζήτηση ευθείας που να περνά απ' όλα τα σημεία, και να δεχτούμε μια προσεγγιστική λύση, δηλαδή μια ευθεία που να μην απέχει πολύ από τα σημεία. Μπορούμε να συγκεκριμενοποιήσουμε και πάλι το πρόβλημα, θέτοντας ως απαίτηση, το άθροισμα των αποστάσεων των σημείων από την ευθεία να είναι το ελάχιστο δυνατό.



Σχ. 7.2 Προσεγγιστική γεωμετρική λύση στο πρόβλημα της κατασκευής ευθείας που περνά από $n > 2$ σημεία.

Στο Σχ. 7.2 έχουμε σχεδιάσει δύο τυχαίες ευθείες και τις αποστάσεις τους d_i από τα σημεία $A_i(x_i, y_i)$. Οπτικά διακρίνουμε ότι στην πρώτη ευθεία (Σχ. 7.2(α)) το άθροισμα των αποστάσεων d_i είναι μικρότερο απ' ότι στη δεύτερη (Σχ. 7.2(β)). Εμπειρικά, θα μπορούσαμε να δοκιμάσουμε πολλές τυχαίες ευθείες, να βρούμε για κάθε μια το άθροισμα των αποστάσεων και να επιλέξουμε αυτή με το μικρότερο άθροισμα. Μια ακριβής γεωμετρική κατασκευή της ευθείας, με την αυστηρή έννοια, είναι πολύπλοκη. Άλλωστε δεν έχει και πρακτικό νόημα γιατί στην πραγματικότητα αυτό που στις εφαρμογές ενδιαφέρει είναι η εξίσωση (7.2) και όχι η γεωμετρική απεικόνισή της. Για το λόγο αυτό προχωρούμε στην αναλυτική αντιμετώπιση, κάνοντας και κάποια τροποποίηση ως προς το στόχο του προβλήματος.

Αναλυτική προσέγγιση

Σύμφωνα με την αναλυτική προσέγγιση, εφαρμόζοντας την εξίσωση (7.2) για καθένα από τα σημεία (x_i, y_i) παίρνουμε ένα σύστημα από n εξισώσεις τις μορφής

$$y_i = a + b x_i \quad (7.3)$$

Οι άγνωστοι του συστήματος των εξισώσεων είναι οι παράμετροι a και b . Το σύστημα έχει μοναδική ακριβή λύση μόνο για $n = 2$, ενώ για $n > 2$ γενικά δεν έχει λύση, αλλά είναι ένα *υπερκαθορισμένο σύστημα*. Για να βρούμε τα a και b στην περίπτωση αυτή δεχόμαστε ότι η εξίσωση (7.2) δεν θα επαληθεύεται ακριβώς από τα σημεία (x_i, y_i) , αλλά θα υπάρχουν σφάλματα

$$w_i = y_i - (a + b x_i) \quad (7.4)$$

τα οποία προσπαθούμε να ελαχιστοποιήσουμε με κάποια έννοια.

Το απλούστερο υπολογιστικό σχήμα ελαχιστοποιεί το άθροισμα των τετραγώνων των επιμέρους σφαλμάτων

$$q = \sum_{i=1}^n w_i^2 = \sum_{i=1}^n [y_i - (a + b x_i)]^2 \quad (7.5)^*$$

Η προκύπτουσα μέθοδος ονομάζεται μέθοδος των *ελάχιστων τετραγώνων*. Η ελαχιστοποίηση του μεγέθους q διαφέρει από την ελαχιστοποίηση του αθροίσματος των αποστάσεων των σημείων από την ευθεία που συζητήθηκε στο προηγούμενο εδάφιο.

Η εύρεση των άγνωστων παραμέτρων που ελαχιστοποιούν το q είναι απλή. Παρατηρούμε ότι το q είναι συνάρτηση των a και b (τα x_i και y_i είναι γνωστοί αριθμοί) και κατά συνέπεια παίρνει την ελάχιστη τιμή του όταν

$$\frac{\partial q}{\partial a} = -2 \sum_{i=1}^n [y_i - (a + b x_i)] = 0 \quad (7.6)$$

* Η άθροιση γίνεται προφανώς για να συνυπολογίσουμε τα σφάλματα σε όλα τα επιμέρους σημεία. Η ύψωση στο τετράγωνο γίνεται για να απαλλαγούμε από το πρόσημο των επιμέρους σφαλμάτων. Εναλλακτικά θα μπορούσαμε να χρησιμοποιήσουμε τις απόλυτες τιμές των σφαλμάτων αλλά αυτό δημιουργεί υπολογιστική δυσκολία.

$$\frac{\partial q}{\partial b} = -2 \sum_{i=1}^n [y_i - (a + bx_i)] x_i = 0 \quad (7.7)$$

Ισοδύναμα οι εξισώσεις αυτές γράφονται

$$na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad (7.8)$$

$$a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \quad (7.9)$$

Η λύση του συστήματος είναι

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (7.10)$$

$$a = \bar{y} - b \bar{x} \quad (7.11)$$

όπου \bar{x} και \bar{y} οι μέσες τιμές των x_i και y_i , αντίστοιχα, δηλαδή

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (7.12)$$

Συντελεστής προσδιορισμού

Οι εξισώσεις (7.10) και (7.11) δίνουν τις ζητούμενες παραμέτρους b και a , χωρίς όμως να δίνουν καμιά πληροφορία σχετικά με το πόσο καλά είναι διατεταγμένα τα σημεία (x_i, y_i) γύρω από την ευθεία που υπολογίστηκε. Αυτή την πληροφορία τη δίνει το μέγεθος q , το οποίο με αντικατάσταση των (7.10) και (7.11) στην (7.5) γίνεται μετά από πράξεις

$$q = \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{\left[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right]^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (7.13)$$

Συνήθως χρησιμοποιούμε την ακόλουθη αδιαστατοποιημένη μορφή του σφάλματος

$$d = 1 - \frac{q}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7.14)$$

Το μέγεθος d λέγεται *συντελεστής προσδιορισμού* (determination coefficient). Παρατηρούμε η μέγιστη τιμή του d είναι 1 και αντιστοιχεί σε μηδενική τιμή του σφάλματος q . Η ελάχιστη τιμή του είναι 0 και αντιστοιχεί σε τιμή του σφάλματος

$$q = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (7.15)$$

Εύκολα μπορούμε να αποδείξουμε ότι το ελάχιστο τετραγωνικό σφάλμα δεν μπορεί να είναι μεγαλύτερο από την παραπάνω τιμή. Αν ήταν, τότε η ευθεία $y = a + bx$ με $a = \bar{y}$ και $b = 0$ θα έδινε μικρότερο σφάλμα, το οποίο είναι άτοπο, αφού οι (7.10) και (7.11) δίνουν το ελάχιστο σφάλμα. Κατά συνέπεια

$$0 \leq d \leq 1 \quad (7.16)$$

Αν αντικαταστήσουμε στην (7.14) την (7.13) παίρνουμε την τελική έκφραση του d

$$d = \frac{\left(n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right)^2}{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] \left[n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right]} = \frac{\left[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right]^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \quad (7.17)$$

Η τετραγωνική του ρίζα

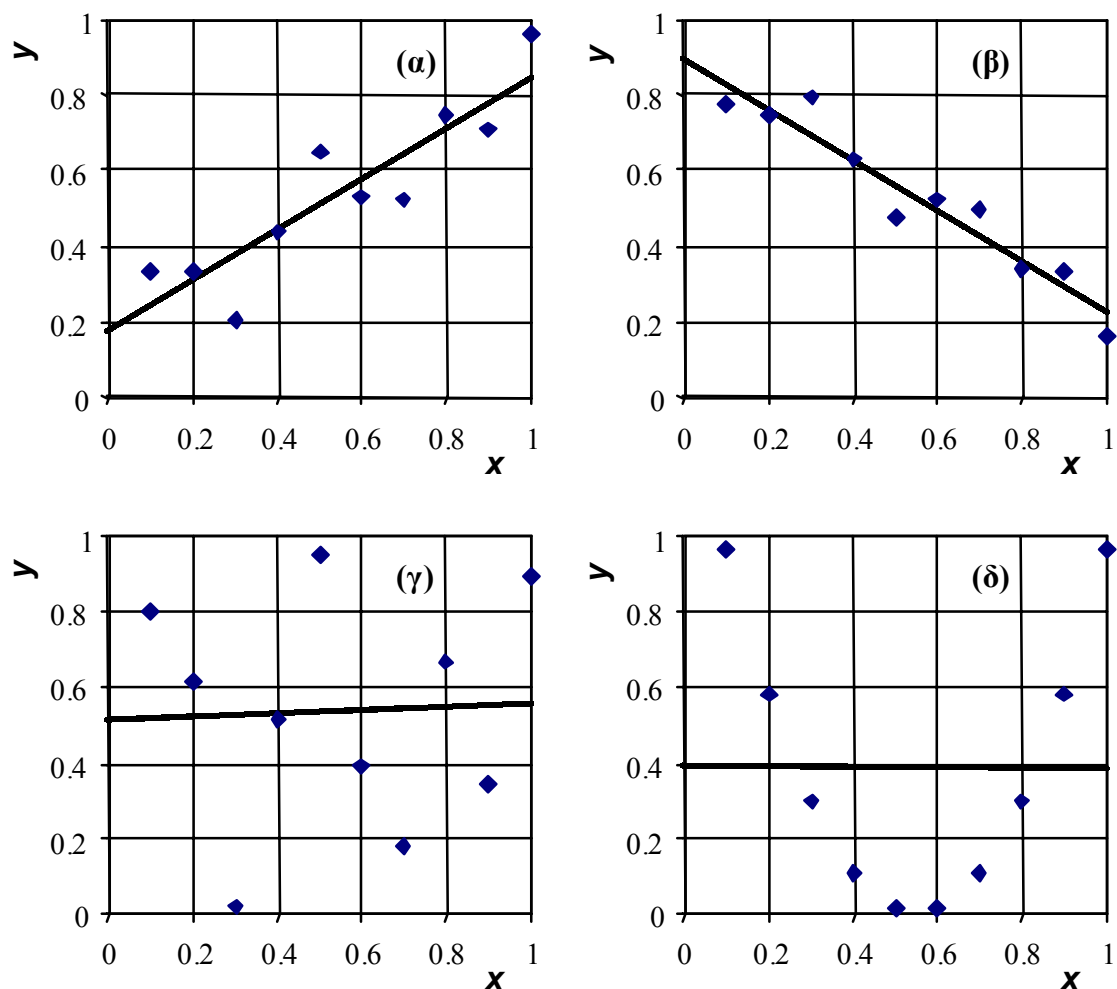
$$r = \sqrt{d} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] \left[n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right]}} \quad (7.18)$$

$$= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

είναι ο γνωστός μας συντελεστής συσχέτισης, με τιμές

$$-1 \leq r \leq 1 \quad (7.19)$$

Συγκεκριμένα, το r παίρνει τιμή κοντά στο 0 όταν η διάταξη των σημείων δεν είναι ευθύγραμμη, οπότε το προκύπτον σφάλμα είναι αρκετά μεγάλο, κοντά στο 1 όταν η διάταξη των σημείων είναι ευθύγραμμη και η κλίση b της ευθείας είναι θετική, και κοντά στο -1 όταν η διάταξη των σημείων είναι ευθύγραμμη αλλά η κλίση b της ευθείας είναι αρνητική (Σχ. 7.3).



Σχ. 7.3 Παραδείγματα ευθειών ελάχιστων τετραγώνων για 10 σημεία (x_i, y_i) και για τέσσερις περιπτώσεις διάταξης σημείων: (α) περίπου ευθύγραμμη διάταξη σημείων με θετική κλίση ($a = 0.17, b = 0.68, r = 0.90$); (β) περίπου ευθύγραμμη διάταξη σημείων με αρνητική κλίση ($a = 0.90, b = -0.67, r = -0.96$); (γ) τυχαία μη ευθύγραμμη διάταξη σημείων ($a = 0.52, b = 0.03, r = 0.03$); (δ) παραβολική διάταξη σημείων ($a = 0.39, b = 0.00, r = 0.00$). Παρατηρούμε ότι στις περιπτώσεις (γ) και (δ) ο συντελεστής συσχέτισης είναι πρακτικώς μηδέν παρόλο που υπάρχει ουσιαστική διαφορά στη διάταξη των σημείων. Συγκεκριμένα στην περίπτωση (γ) η διάταξη των σημείων είναι τελείως τυχαία, ενώ στη (δ) τα σημεία υπακούουν σε ένα παραβολικό νόμο.

Αντιστροφή του γραμμικού νόμου

Αφού προσδιοριστεί η ευθεία $y = a + bx$, φαίνεται φυσική η άμεση αντιστροφή του νόμου, δηλαδή η έκφραση του x ως προς y με άμεση επίλυση, η οποία δίνει $x = (y - a) / b$. Ωστόσο, η έκφραση αυτή έχει προκύψει με ελαχιστοποίηση του σφάλματος του γραμμικού νόμου ως προς τη

μεταβλητή y . Αν θέλουμε την έκφραση του γραμμικού νόμου ως προς x , τότε θα πρέπει να ελαχιστοποιήσουμε το σφάλμα του νόμου ως προς x , δηλαδή την ποσότητα

$$q' = \sum_{i=1}^n w_i'^2 = \sum_{i=1}^n [x_i - (a' + b'y_i)]^2 \quad (7.20)$$

όπου έχουμε υποθέσει ότι η νέα έκφραση του γραμμικού νόμου είναι η $x = a' + b'y$.

Σε αυτή την περίπτωση οι παράμετροι b και a δίνονται από τις ακόλουθες εξισώσεις, ανάλογες με τις (7.10) και (7.11):

$$b' = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7.21)$$

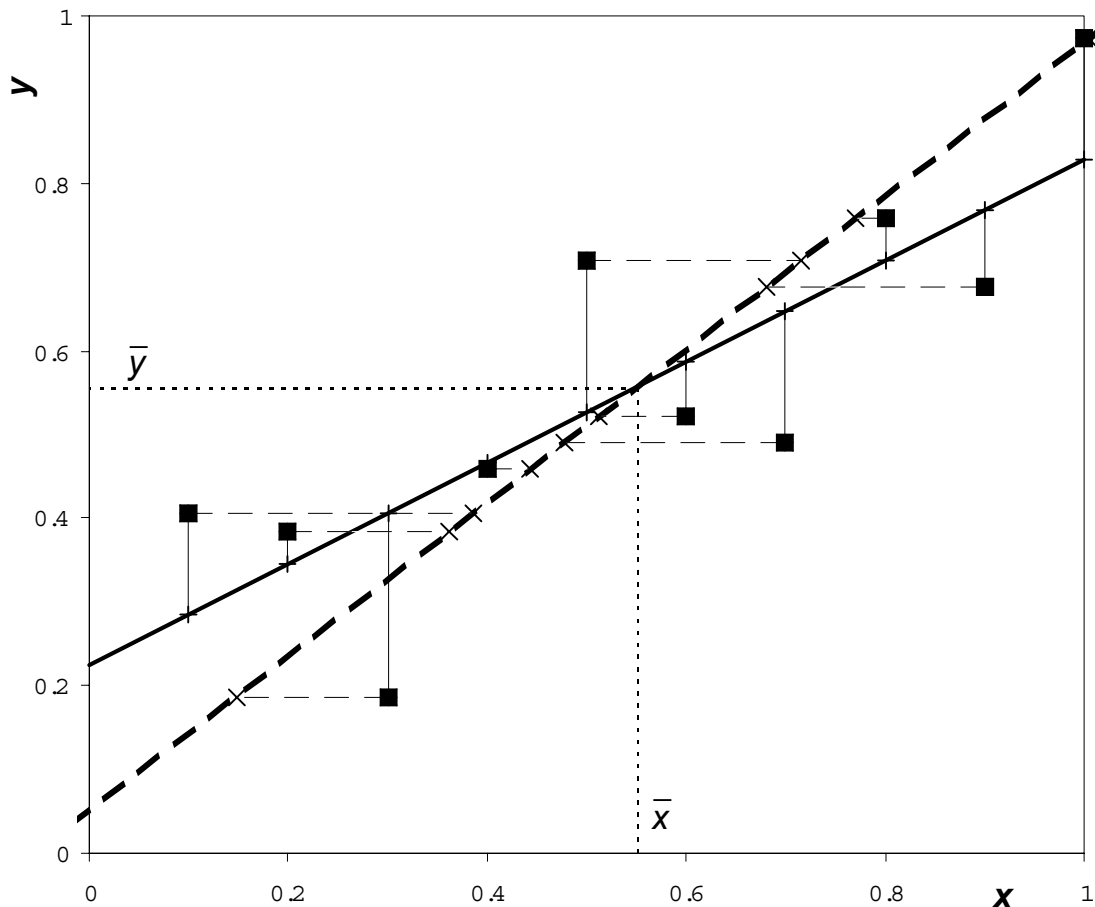
$$a' = \bar{x} - b' \bar{y} \quad (7.22)$$

Οι εξισώσεις (7.17) και (7.18) που δίνουν τους συντελεστές προσδιορισμού και συσχέτισης, αντίστοιχα, οι οποίες είναι συμμετρικές ως προς x και y , ισχύουν ως έχουν και για αυτή την περίπτωση. Επισημαίνουμε ότι οι παράμετροι που προκύπτουν από τις (7.21) και (7.22) ορίζουν διαφορετική ευθεία από αυτή που ορίζουν οι παράμετροι που προκύπτουν από τις (7.10) και (7.11) (βλ. Σχ. 7.4). Οι δύο ευθείες ταυτίζονται μόνο όταν ο συντελεστής προσδιορισμού είναι ίσος με 1.

Εύκολα μπορούμε να διαπιστώσουμε χρησιμοποιώντας τις παραπάνω σχέσεις ότι οι εξισώσεις των δύο ευθειών μπορεί να γραφούν εναλλακτικά με την ακόλουθη μορφή

$$y - \bar{y} = b(x - \bar{x}) \quad x - \bar{x} = b'(y - \bar{y}) \quad (7.23)$$

απ' όπου προκύπτει ότι και οι δύο ευθείες περνούν από το σημείο (\bar{x}, \bar{y}) (Σχ. 7.4).



Σχ. 7.4 Επεξηγηματικό παράδειγμα για τις δύο ευθείες ελάχιστων τετραγώνων που ορίζονται από μια σημειοσειρά (x_i, y_i) . Τα 10 σημεία παριστάνονται με ■. Η έντονη συνεχής ευθεία με εξίσωση $y = 0.22 + 0.60x$ αντιστοιχεί στην ελαχιστοποίηση του τετραγωνικού σφάλματος q (εξίσωση (7.5)). Τα επιμέρους σφάλματα w_i , δηλαδή οι κατακόρυφες αποστάσεις των σημείων από αυτή την ευθεία έχουν απεικονιστεί με λεπτή συνεχή γραμμή. Η έντονη διακεκομμένη ευθεία με εξίσωση $x = -0.05 + 1.09y$ αντιστοιχεί στην ελαχιστοποίηση του τετραγωνικού σφάλματος q' (εξίσωση (7.20)). Τα επιμέρους σφάλματα w'_i , δηλαδή οι οριζόντιες αποστάσεις των σημείων από αυτή την ευθεία έχουν απεικονιστεί με λεπτή διακεκομμένη γραμμή. Ο συντελεστής συσχέτισης στο παράδειγμα είναι 0.81. Παρατηρούμε ότι οι δύο ευθείες τέμνονται στο σημείο (\bar{x}, \bar{y}) .

7.1.2 Πιθανοτική αντιμετώπιση

Εξισώσεις εκτίμησης παραμέτρων

Σύμφωνα με μία από τις πιθανοτικές προσεγγίσεις τα ζεύγη (x_i, y_i) θεωρούνται πραγματοποιήσεις δύο τυχαίων μεταβλητών (X, Y) και το πρό-

βλημα είναι η εκτίμηση της Y συναρτήσει της X . Στην περίπτωση γραμμικού νόμου ανάμεσα στις X και Y θα ισχύει

$$Y = a + bX + W \quad (7.24)$$

όπου a και b άγνωστες αριθμητικές παράμετροι και W τυχαία μεταβλητή που εκφράζει το σφάλμα της εκτίμησης

$$\hat{Y} = a + bX \quad (7.25)$$

δηλαδή $W = Y - \hat{Y} = Y - (a + bX)$. Οι παράμετροι a και b υπολογίζονται έτσι ώστε η (7.25) να αποτελεί την εκτίμηση ελάχιστου μέσου τετραγωνικού σφάλματος, με την έννοια ότι ελαχιστοποιείται το μέγεθος

$$\begin{aligned} E[W^2] &= E[(Y - \hat{Y})^2] = E[(Y - (a + bX))^2] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y - (a + bx)]^2 f_{XY}(x, y) dx dy \end{aligned} \quad (7.26)$$

όπου με $E[\]$ συμβολίζεται η αναμενόμενη τιμή. Παραγωγίζοντας την (7.26) ως προς a και b παίρνουμε μετά από πράξεις (βλ. π.χ. Papoulis, 1990, σ. 150) ένα σύστημα εξισώσεων, το οποίο επιλυόμενο δίνει

$$b = \frac{\sigma_{XY}}{\sigma_X^2} = \rho_{XY} \frac{\sigma_Y}{\sigma_X} \quad (7.27)$$

$$a = \mu_Y - b \mu_X \quad (7.28)$$

όπου μ_X και μ_Y οι μέσες τιμές των X και Y , αντίστοιχα, σ_X και σ_Y οι αντίστοιχες τυπικές αποκλίσεις και σ_{XY} και ρ_{XY} η συνδιασπορά και ο συντελεστής συσχέτισης, αντίστοιχα, των X και Y . Τα μεγέθη αυτά συνήθως είναι άγνωστα (παράμετροι πληθυσμού). Αν στη θέση τους τοποθετήσουμε τις εκτιμήσεις τους από το δείγμα των n σημείων (x_i, y_i) , τότε παίρνουμε πάλι τις εξισώσεις (7.10) και (7.11), οι οποίες είχαν εξαχθεί με την προσδιοριστική προσέγγιση.

Η εξίσωση (7.25), αν αντικαταστήσουμε τις παραπάνω τιμές των a και b , παίρνει την ακόλουθη μορφή

$$\frac{\hat{Y} - \mu_Y}{\sigma_Y} = \rho_{XY} \frac{X - \mu_X}{\sigma_X} \quad (7.29)$$

Κατά ανάλογο τρόπο προκύπτουν και οι εξισώσεις της αντίστροφης ευθείας ελάχιστων τετραγώνων, οι οποίες είναι

$$b' = \frac{\sigma_{XY}}{\sigma_Y^2} = \rho_{XY} \frac{\sigma_X}{\sigma_Y} \quad (7.30)$$

$$a' = \mu_X - b' \mu_Y \quad (7.31)$$

Η έννοια της παλινδρόμησης

Ο όρος *παλινδρόμηση* (regression) χρησιμοποιείται ως συνώνυμος του όρου *μέθοδος ελάχιστων τετραγώνων*. Ο όρος (αν και ανεπιτυχής) έχει καθιερωθεί από τον Galton*, ο οποίος παρατήρησε ότι “Τα πληθυσμιακά ακρότατα οπισθοχωρούν (regress)† προς τη μέση τιμή τους”. Η παρατήρηση αυτή προέρχεται από τη μελέτη του ύψους των παιδιών σχετικά με το ύψος των γονέων. Από τη μελέτη φάνηκε ότι τα παιδιά ψηλών (αντίστροφα, κοντών) γονέων είναι κατά μέσο όρο κοντότερα (αντίστροφα, ψηλότερα) από τους γονείς τους. Αυτό μπορεί να ερμηνευτεί στατιστικά με βάση την ευθεία παλινδρόμησης (7.29), όπου οι μεταβλητές X και Y συμβολίζουν το ύψος των γονέων και των παιδιών, αντίστοιχα. Στη συγκεκριμένη περίπτωση ισχύει $\mu_X = \mu_Y = \mu$ και $\sigma_X = \sigma_Y$, οπότε η εξίσωση παλινδρόμησης παίρνει τη μορφή $\hat{Y} - \mu = \rho_{XY} (X - \mu)$. Δεδομένου ότι για τα εξεταζόμενα μεγέθη $0 < \rho_{XY} < 1$, θα είναι $\hat{Y} - \mu < X - \mu$, πράγμα που δείχνει την τάση μείωσης της διαφοράς από τη μέση τιμή (οπισθοχώρησης προς τη μέση τιμή).

Ας σημειωθεί ότι ο όρος παλινδρόμηση δεν χρησιμοποιείται μόνο για την εξεταζόμενη περίπτωση γραμμικής σχέσης ανάμεσα σε δύο μεταβλη-

* Sir Francis Galton, γενετιστής και βιοστατιστικός (1822-1911).

† Οι αγγλικές λέξεις *regress* και *regression* στην ελληνική επιστημονική ορολογία καθιερώθηκε να αποδίδονται ως *παλινδρομώ* και *παλινδρόμηση*, αν και στη συγκεκριμένη περίπτωση έχουν το νόημα της οπισθοχώρησης.

τές, αλλά για οποιασδήποτε μορφής σχέση ανάμεσα σε οποιοδήποτε αριθμό μεταβλητών (βλ. και ενότητες 7.3 και 7.4).

Ροπές του σφάλματος εκτίμησης

Σύμφωνα με την πιθανοτική θεώρηση που αναπτύσσουμε εδώ, το σφάλμα W θεωρείται ως τυχαία μεταβλητή. Συνδυάζοντας τις (7.24) και (7.28) βρίσκουμε ότι

$$W = (Y - \mu_Y) - b(X - \mu_X) \quad (7.32)$$

Παίρνοντας αναμενόμενες τιμές στην παραπάνω καταλήγουμε ότι

$$\mu_W = 0 \quad (7.33)$$

δηλαδή η αναμενόμενη τιμή του σφάλματος είναι μηδενική.

Εξ άλλου υψώνοντας τα δύο μέλη της (7.32) στο τετράγωνο και παίρνοντας πάλι αναμενόμενες τιμές βρίσκουμε ότι η διασπορά του W είναι

$$\sigma_W^2 = \sigma_Y^2 + b^2 \sigma_X^2 - 2b \sigma_{XY} \quad (7.34)$$

Αντικαθιστώντας στην παραπάνω το b από την (7.27) παίρνουμε

$$E[W^2] = \sigma_W^2 = \sigma_Y^2 (1 - \rho_{XY}^2) \quad (7.35)$$

Τέλος, πολλαπλασιάζοντας τα δύο μέλη της (7.32) με $(X - \mu_X)$ και παίρνοντας στη συνέχεια αναμενόμενες τιμές βρίσκουμε ότι

$$\sigma_{WX} = \sigma_{YX} - b \sigma_X^2 \quad (7.36)$$

Αν αντικαταστήσουμε στην παραπάνω εξίσωση το b από την (7.27) βρίσκουμε ότι

$$\sigma_{WX} = 0 \quad (7.37)$$

πράγμα που σημαίνει ότι τα σφάλματα W είναι ασυσχέτιστα με τα δεδομένα X . Το συμπέρασμα αυτό επεκτείνεται (Paroulis, 1990, σ. 409-411) για οποιαδήποτε καμπύλη ελάχιστων τετραγώνων (απλή ή πολλαπλή,

γραμμική ή μη γραμμική). Έτσι αποδεικνύεται π.χ. ότι $E[Wg(X)] = 0$ όπου $g(X)$ είναι οποιαδήποτε συνάρτηση της X .

Ροπές της εκτιμήτριας

Οι ροπές της μεταβλητής $\hat{Y} = a + bX$, η οποία αποτελεί την εκτιμήτρια της Y για δεδομένη τιμή της X , προκύπτουν εύκολα με αξιοποίηση των παραπάνω σχέσεων και είναι

$$\mu_{\hat{Y}} = \mu_Y \quad (7.38)$$

$$\sigma_{\hat{Y}}^2 = \frac{\sigma_{XY}^2}{\sigma_X^2} = \rho_{XY}^2 \sigma_Y^2 \quad (7.39)$$

Παρατηρούμε ότι η εκτιμήτρια \hat{Y} διατηρεί τη μέση τιμή (έχει την ίδια μέση τιμή με την Y , αλλά όχι και τη διασπορά. Δεδομένου ότι $\rho_{XY} < 1$, η διασπορά της \hat{Y} είναι πάντα μικρότερη από αυτήν της Y .

Συνδυάζοντας τις εξισώσεις (7.39) και (7.35) βρίσκουμε ότι

$$\sigma_Y^2 = \sigma_{\hat{Y}}^2 + \sigma_W^2 \quad (7.40)$$

Τα μεγέθη σ_Y^2 , $\sigma_{\hat{Y}}^2$ και σ_W^2 συχνά αποκαλούνται *ολική*, *παλινδρομική* και *υπόλοιπη* διασπορά, αντίστοιχα.

Συντελεστής προσδιορισμού

Ο συντελεστής προσδιορισμού σύμφωνα με την πιθανοτική προσέγγιση γενικά ορίζεται ως

$$\delta = 1 - \frac{E[W^2]}{\sigma_Y^2} \quad (7.41)$$

Ο ορισμός αυτός βρίσκεται σε αντιστοιχία με αυτόν της προσδιοριστικής προσέγγισης (σχέση (7.14)). Πράγματι, αν στην (7.41) αντικατασταθούν οι θεωρητικές ροπές με τις δειγματικές εκτιμήσεις τους, προκύπτει η (7.14). Στην εξεταζόμενη περίπτωση της γραμμικής παλινδρόμησης συμβαίνει να ισχύει

$$\delta = \rho_{XY}^2 = \frac{\sigma_{\hat{Y}}^2}{\sigma_Y^2} \quad (7.42)$$

δηλαδή ο συντελεστής προσδιορισμού είναι ταυτόχρονα ίσος με το τετράγωνο του συντελεστή συσχέτισης των δύο μεταβλητών, καθώς και με το λόγο της παλινδρομικής προς την ολική διασπορά.

Αυτή η ιδιότητα δεν ισχύει για κάθε παλινδρόμηση (βλ. π.χ. επόμενο εδάφιο). Λόγω της τελευταίας ισότητας ο συντελεστής προσδιορισμού, συχνά εκφρασμένος ως ποσοστό στα %, ονομάζεται και *ποσοστό εξηγούμενης διασποράς* του μοντέλου παλινδρόμησης.

Εφαρμογή 7.1

Η λεκάνη απορροής του Αχελώου ανάντη του φράγματος Κρεμαστών έχει ολική έκταση 3584 km², από τα οποία τα 80.6 km² (κατά μέγιστο) καταλαμβάνει ο ταμιευτήρας, ωφέλιμης χωρητικότητας 3300 hm³. Στον Πίν. 7.1 δίνονται για την περίοδο 1967-68 μέχρι 1991-92 (25 χρόνια) σε ετήσια βάση (1) η επιφανειακή βροχόπτωση της λεκάνης, P, (2) το ισοδύναμο ύψος επιφανειακής απορροής της λεκάνης, Q, και (3) η εξάτμιση από την επιφάνεια του ταμιευτήρα κατά Penman, E.

α. Να διερευνηθεί στατιστικά με βάση τα δείγματα αν συσχετίζονται γραμμικά τα μεγέθη P-Q και E-Q και να προσδιοριστούν οι εξισώσεις εφόσον έχουν νόημα.

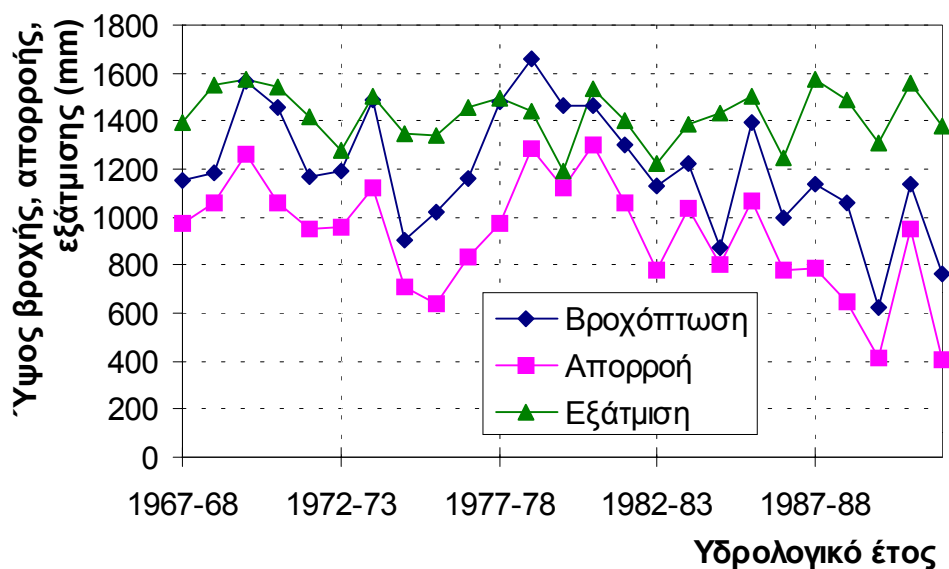
Μια πρώτη εμπειρική εικόνα για την ύπαρξη ή όχι συσχέτισης ανάμεσα στις μεταβλητές P, Q και E παρέχεται στο Σχ. 7.5, όπου έχουν απεικονιστεί οι τρεις χρονοσειρές συναρτήσει του υδρολογικού έτους. Είναι εμφανές ότι οι υπάρχει έντονη συσχέτιση ανάμεσα στη βροχόπτωση και την απορροή της λεκάνης, αλλά όχι ανάμεσα στην εξάτμιση του ταμιευτήρα και την απορροή της λεκάνης. Άλλωστε, αυτό είναι και φυσικά αναμενόμενο και εξηγήσιμο. Σε μια περιοχή σημαντικής υδροφορίας, όπως είναι η βορειοδυτική Ελλάδα, στην οποία ανήκει η λεκάνη του Αχελώου, το μεγαλύτερο τμήμα της βροχόπτωσης απορρέει επιφανειακά, πράγμα που εξηγεί την ισχυρή συσχέτιση ανάμεσα στη βροχόπτωση και την επιφανειακή απορροή. Αντίθετα, δεν περιμένουμε να υπάρχει συσχέτιση ανάμεσα στην απορροή της λεκάνης και στο ύψος εξάτμισης από τον ταμιευτήρα κατά Penman. Το τελευταίο εξαρτάται από μετεωρολογικές μεταβλητές της περιοχής (θερμοκρασία, σχετική υγρασία, ηλιοφάνεια, ταχύτητα ανέμου) και δε συναρτάται με τη βροχόπτωση ή την απορροή.

Πίν. 7.1 Χρονοσειρές ετήσιου ύψους βροχόπτωσης (P) και απορροής (Q) της λεκάνης Αχελώου ανάντη των Κρεμαστών, και εξάτμισης (E) του ταμιευτήρα Κρεμαστών.

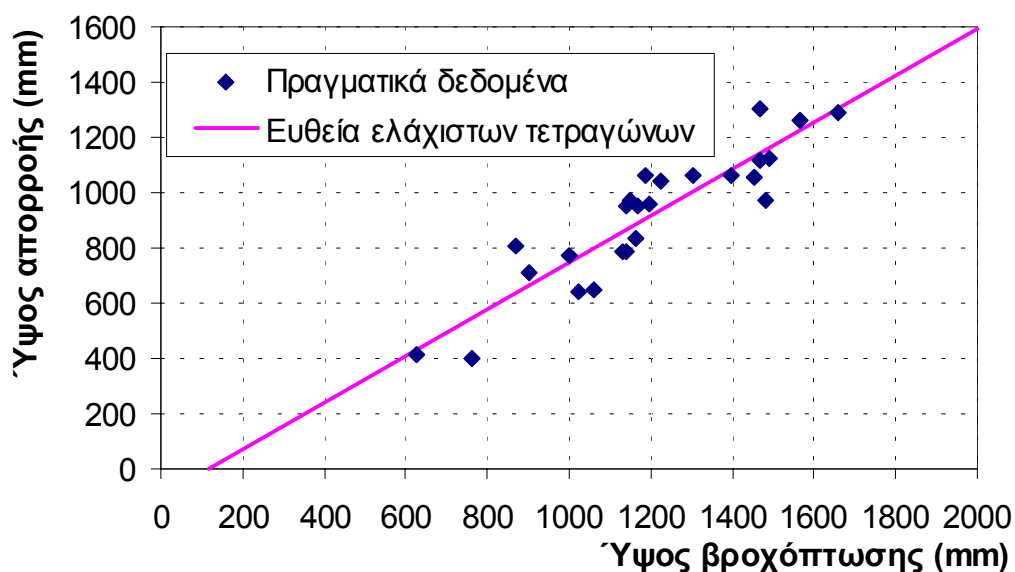
| Υδρολ. έτος | P (mm) | Q (mm) | E (mm) | Υδρολ. έτος | P (mm) | Q (mm) | E (mm) |
|----------------|-------------|-------------|-------------|----------------|-------------|-------------|-------------|
| 1967-68 | 1150.6 | 972.8 | 1392.0 | 1980-81 | 1466.3 | 1301.0 | 1535.2 |
| 1968-69 | 1185.3 | 1060.7 | 1549.2 | 1981-82 | 1304.9 | 1063.5 | 1399.9 |
| 1969-70 | 1563.5 | 1260.1 | 1576.2 | 1982-83 | 1131.5 | 782.8 | 1226.3 |
| 1970-71 | 1454.6 | 1057.1 | 1545.5 | 1983-84 | 1223.2 | 1038.1 | 1390.0 |
| 1971-72 | 1165.9 | 948.9 | 1417.3 | 1984-85 | 871.2 | 805.9 | 1434.6 |
| 1972-73 | 1194.6 | 956.4 | 1279.4 | 1985-86 | 1398.2 | 1065.0 | 1501.5 |
| 1973-74 | 1490.7 | 1120.7 | 1504.7 | 1986-87 | 1000.5 | 775.8 | 1245.0 |
| 1974-75 | 901.7 | 710.7 | 1345.9 | 1987-88 | 1141.3 | 787.6 | 1572.7 |
| 1975-76 | 1021.9 | 641.9 | 1342.7 | 1988-89 | 1058.6 | 647.8 | 1488.0 |
| 1976-77 | 1162.9 | 833.9 | 1460.2 | 1989-90 | 626.4 | 411.7 | 1305.7 |
| 1977-78 | 1482.9 | 972.2 | 1492.4 | 1990-91 | 1139.3 | 950.7 | 1560.2 |
| 1978-79 | 1661.2 | 1286.6 | 1440.1 | 1991-92 | 761.3 | 403.2 | 1382.8 |
| 1979-80 | 1467.0 | 1118.5 | 1190.9 | | | | |

Στα ίδια συμπεράσματα καταλήγουμε μελετώντας τα Σχ. 7.6 και Σχ. 7.7. Συγκεκριμένα, στο Σχ. 7.6 έχουμε απεικονίσει το ύψος επιφανειακής απορροής συναρτήσει του ύψους βροχόπτωσης. Είναι εμφανής η συσχέτιση ανάμεσα στα δύο μεγέθη, η οποία μάλιστα φαίνεται να είναι γραμμική, κάτι που είναι επίσης αναμενόμενο και εξηγήσιμο, δεδομένου ότι στις περιοχές μεγάλης υδροφορίας η ετήσια βροχή και απορροή ακολουθούν από κοινού κανονική κατανομή (βλ. ενότητα 7.4). Αντίστοιχα, στο Σχ. 7.7 έχουμε απεικονίσει το ύψος επιφανειακής απορροής συναρτήσει του ύψους εξάτμισης από τον ταμιευτήρα. Εδώ είναι εμφανής η ανυπαρξία συσχέτισης, γραμμικής ή άλλης, ανάμεσα στα δύο μεγέθη.

Βεβαίως οι παραπάνω παρατηρήσεις είναι ποιοτικού χαρακτήρα, αλλά μπορούν εύκολα να ποσοτικοποιηθούν με τη χρήση στατιστικής. Η στατιστική παράμετρος που ενδιαφέρει είναι ο συντελεστής συσχέτισης (ή ο συντελεστής προσδιορισμού της γραμμικής παλινδρόμησης). Σύμφωνα με όσα έχουν αναφερθεί στο εδάφιο 3.5.3 (εξίσωση 3.74), για να είναι στατιστικά σημαντικός (διάφορος του μηδενός) ο συντελεστής συσχέτισης θα πρέπει να είναι μεγαλύτερος του $r_c = 2/\sqrt{n}$, όπου n το μέγεθος του δείγματος. Στην προκειμένη περίπτωση $r_c = 2/\sqrt{25} = 0.4$.

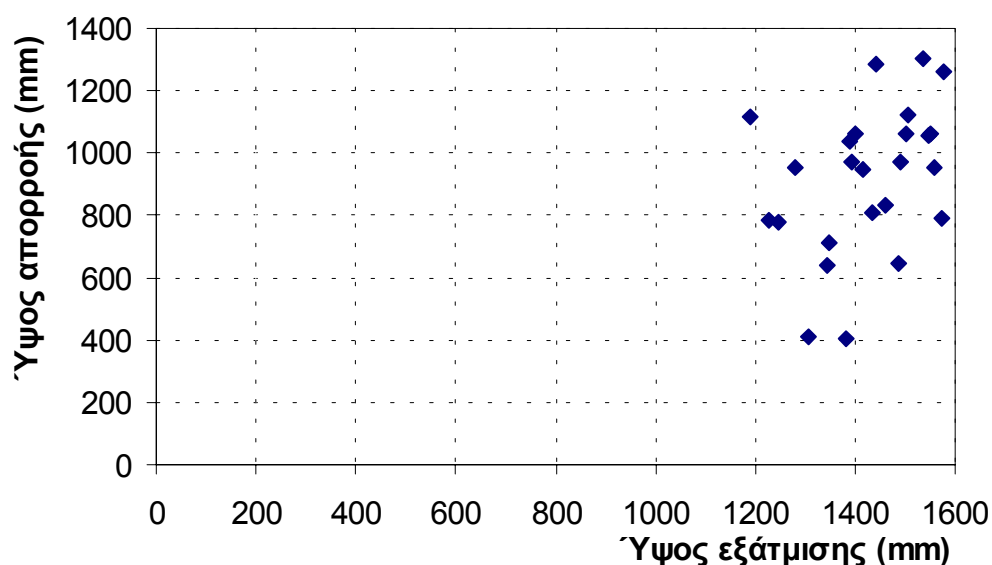


Σχ. 7.5 Απεικόνιση των χρονοσειρών ετήσιου ύψους βροχόπτωσης και απορροής της λεκάνης Αχελώου ανάντη των Κρεμαστών, και της εξάτμισης του ταμιευτήρα Κρεμαστών.



Σχ. 7.6 Ετήσιο ύψος επιφανειακής απορροής συναρτήσει του ετήσιου ύψους βροχόπτωσης στη λεκάνη ανάντη Κρεμαστών.

Η εκτίμηση του δειγματικού συντελεστή συσχέτισης δίνεται από την εξίσωση (7.18). Αν και ο υπολογισμός του γίνεται αυτόματα από υπολογιστές ή αριθμομηχανές, εδώ εκθέτουμε για διδακτικούς λόγους τον υπολογισμό του με το χέρι. Στους παρακάτω συμβολισμούς των αθροισμάτων παραλείπουμε για ευκολία τους δείκτες i στα δείγματα $x_i y_i$.



Σχ. 7.7 Ετήσιο ύψος επιφανειακής απορροής στη λεκάνη ανάντη Κρεμαστών συναρτήσει του ετήσιου ύψους εξάτμισης από τον ταμιευτήρα Κρεμαστών.

Για τις μεταβλητές $P (\equiv X)$ και $Q (\equiv Y)$ έχουμε:

$$n = 25, \sum x = 30\,025.5, \sum y = 22\,973.6, \sum xy = 28\,941\,151,$$

$$\sum x^2 = 37\,655\,556, \sum y^2 = 22\,487\,374$$

Κατά συνέπεια από την εξίσωση (7.18) βρίσκουμε

$$r = \frac{25 \times 28\,941\,151 - 30\,025.5 \times 22\,973.6}{\sqrt{(25 \times 37\,655\,556 - 30\,025.5^2)(25 \times 22\,487\,374 - 22\,973.6^2)}}$$

απ' όπου τελικά προκύπτει $r = 0.911$. Η τιμή αυτή είναι μεγαλύτερη από την κρίσιμη τιμή $r_c = 0.4$, πράγμα που μας οδηγεί στην απόρριψη της υπόθεσης ότι ο συντελεστής γραμμικής συσχέτισης ανάμεσα στις δύο μεταβλητές είναι μηδενικός. Επιβεβαιώνεται δηλαδή και στατιστικά η ύπαρξη γραμμικής συσχέτισης ανάμεσα στη βροχόπτωση και την απορροή της λεκάνης.

Αντίστοιχα, για τις μεταβλητές $E (\equiv X)$ και $R (\equiv Y)$ έχουμε:

$$n = 25, \sum x = 35\,578.4, \sum y = 22\,973.6, \sum xy = 32\,935\,017,$$

$$\sum x^2 = 50\,940\,067, \sum y^2 = 22\,487\,374$$

Κατά συνέπεια από την εξίσωση (7.18) βρίσκουμε, όπως παραπάνω, ότι $r = 0.370$. Η τιμή αυτή είναι μικρότερη από την κρίσιμη τιμή $r_c = 0.4$, πράγμα που μας οδηγεί στην μη απόρριψη της υπόθεσης ότι ο συντελεστής γραμμικής συσχέτισης ανάμεσα στις δύο μεταβλητές είναι μηδενικός.

Για την περίπτωση των μεταβλητών $P (\equiv X)$ και $Q (\equiv Y)$, όπου όπως είδαμε έχει νόημα η γραμμική συσχέτιση, υπολογίζουμε τους συντελεστές b και a της γραμμικής παλινδρόμησης, χρησιμοποιώντας τις εξισώσεις (7.10) και (7.11), αντίστοιχα. Έτσι, από την (7.10) παίρνουμε

$$b = \frac{25 \times 28\,941\,151 - 30\,025.5 \times 22\,973.6}{25 \times 37\,655\,556 - 30\,025.5^2} = 0.846$$

και από την (7.11)

$$a = \frac{22\,973.6}{25} - 0.846 \times \frac{30\,025.5}{25} = -97.1$$

Η ευθεία με εξίσωση $y = -97.1 + 0.846x$ έχει παρασταθεί γραφικά στο Σχ. 7.6.

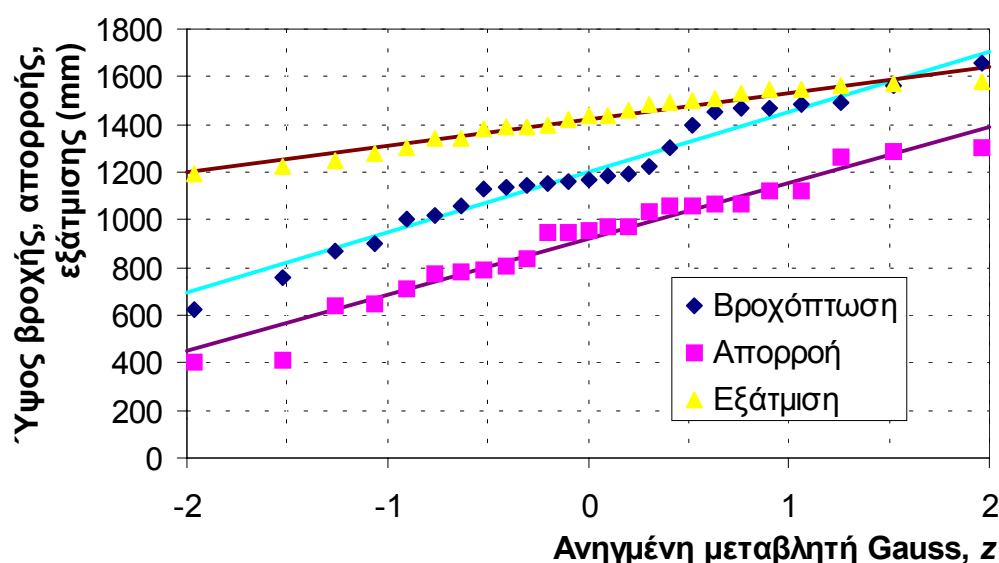
β. Να υπολογιστεί η περίοδος επαναφοράς των τιμών των τριών μεταβλητών για το υδρολογικό έτος 1978-79, θεωρητικά και εμπειρικά. Αν υποτεθεί ότι έλειπε η τιμή της απορροής του έτους αυτού, ποια θα ήταν η εκτίμησή του και σε ποια περίοδο επαναφοράς θα αντιστοιχούσε;

Για να μπορέσουμε να υπολογίσουμε θεωρητικές πιθανότητες υπέρβασης είναι απαραίτητο να υποθέσουμε μια συγκεκριμένη συνάρτηση κατανομής σε καθεμιά από τις μεταβλητές. Ήδη έχουμε υπαινιχθεί ότι η βροχόπτωση και η απορροή ακολουθούν κανονικές κατανομές. Το ίδιο θα πρέπει να υποθέσουμε και για την εξάτμιση από τον ταμιευτήρα, δεδομένου ότι αναφερόμαστε σε ετήσια χρονική κλίμακα. Η προσαρμογή (εκτίμηση παραμέτρων) και ο έλεγχος (με τη δοκιμή χ^2) των κανονικών συναρτήσεων κατανομής γίνεται σύμφωνα με όσα έχουμε αναπτύξει στο κεφάλαιο 5. Ειδικά για τη δοκιμή χ^2 επιλέγουμε 5 κλάσεις, οπότε η κρίσιμη τιμή του στατιστικού ελέγχου για επίπεδο σημαντικότητας 5% και για $5 - 2 - 1 = 2$ βαθμούς ελευθερίας είναι $\chi_{0.95}^2(2) = 5.99$. Τα αποτελέσματα της διαδικασίας προσαρμογής και ελέγχου δίνονται συνοπτικά στον Πίν. 7.2. Παρατηρούμε ότι για καμία από τις τρεις μεταβλητές δεν απορρίπτεται η υπόθεση ότι ακολουθεί κανονική κατανομή. Στο Σχ. 7.8 φαίνεται και γραφικά η

προσαρμογή των κανονικών συναρτήσεων κατανομής στις αντίστοιχες εμπειρικές.

Πίν. 7.2 Παράμετροι κανονικής κατανομής και αποτελέσματα της δοκιμής χ^2 για τις τρεις μεταβλητές της Εφαρμογής 7.1.

| | μ | σ | χ^2 |
|------------|--------|----------|----------|
| Βροχόπτωση | 1201.0 | 252.5 | 4.8 |
| Απορροή | 918.9 | 234.6 | 0.0 |
| Εξάτμιση | 1423.1 | 110.8 | 0.4 |



Σχ. 7.8 Εμπειρικές και θεωρητικές (Gauss) συναρτήσεις κατανομής του ετήσιου ύψους βροχόπτωσης και απορροής της λεκάνης Αχελώου ανάντη των Κρεμαστών, και της εξάτμισης του ταμιευτήρα Κρεμαστών.

Το υδρολογικό έτος 1978-79 η τιμή της βροχόπτωσης είναι 1661.2 mm, και είναι η πρώτη σε μέγεθος τιμή του δείγματος των 25 ετών. Κατά συνέπεια η εμπειρική περίοδος επαναφοράς της, σύμφωνα με τη σχέση Blom (βλ. Πίν. 5.7), είναι

$$T = (25 + 0.25) / (1 - 0.375) = 40.4$$

Η ανηγμένη μεταβλητή είναι

$$z = (1661.2 - 1201.0) / 252.5 = 1.823$$

που, για την υιοθετημένη κανονική κατανομή, αντιστοιχεί σε $F = 0.9658$. Άρα, η θεωρητική περίοδος επαναφοράς είναι

$$T = 1 / (1 - 0.9658) = 29.2.$$

Για το ίδιο υδρολογικό έτος η τιμή της απορροής είναι 1286.6 mm, και είναι η δεύτερη σε μέγεθος τιμή του δείγματος των 25 ετών. Κατά συνέπεια η εμπειρική περίοδος επαναφοράς της είναι

$$T = (25 + 0.25) / (2 - 0.375) = 15.5$$

Η ανηγμένη μεταβλητή είναι

$$z = (1286.6 - 918.9) / 234.6 = 1.567$$

που αντιστοιχεί σε $F = 0.9415$. Άρα, η θεωρητική περίοδος επαναφοράς είναι

$$T = 1 / (1 - 0.9415) = 17.1.$$

Τέλος, για το ίδιο υδρολογικό έτος η τιμή της εξάτμισης είναι 1440.1 mm, και είναι η δωδέκατη σε μέγεθος τιμή του δείγματος των 25 ετών. Κατά συνέπεια η εμπειρική περίοδος επαναφοράς της είναι

$$T = (25 + 0.25) / (12 - 0.375) = 2.17$$

Η ανηγμένη μεταβλητή είναι

$$z = (1440.1 - 1426.1) / 110.8 = 0.153$$

που αντιστοιχεί σε $F = 0.5608$. Άρα, η θεωρητική περίοδος επαναφοράς είναι

$$T = 1 / (1 - 0.5608) = 2.28$$

Αν έλειπε η τιμή της απορροής για το υδρολογικό έτος 1978-79, τότε θα την εκτιμούσαμε από την αντίστοιχη τιμή της βροχόπτωσης, δεδομένου ότι, όπως είδαμε, υπάρχει σημαντική γραμμική συσχέτιση ανάμεσα στα δύο μεγέθη. Βέβαια, στην περίπτωση αυτή οι συντελεστές b και a της γραμμικής εξίσωσης είναι ελαφρά διαφοροποιημένοι, επειδή υπολογίζονται από τα δείγματα των 24 και όχι των 25 ετών. Επαναλαμβάνοντας την ίδια διαδικασία, όπως παραπάνω, βρίσκουμε ότι η εξίσωση της παλινδρόμησης είναι, στην περίπτωση των 24 ετών, $y = -105.7 + 0.854x$, οπότε, για $x = 1661.2$ βρίσκουμε

$$y = -105.7 + 0.854 \times 1661.2 = 1313.0 \text{ mm.}$$

Στην τελευταία περίπτωση η εκτιμημένη τιμή του έτους 1978-79 είναι η μεγαλύτερη στο δείγμα των 25 ετών και επομένως η εμπειρική περίοδος επαναφοράς της θα είναι $T = 40.4$. Για τη θεωρητική περίοδο επαναφοράς χρησιμοποιούμε τη μέση τιμή και τυπική απόκλιση των 24 ετών που είναι $\mu = 903.6$ mm και $\sigma = 226.9$ mm. Έτσι, η ανηγμένη μεταβλητή είναι

$$z = (1313.0 - 903.6) / 226.9 = 1.804$$

οπότε $F = 0.9644$ και $T = 28.1$. Παρατηρούμε λοιπόν ότι η εμπειρική περίοδος επαναφοράς της εκτιμημένης τιμής της απορροής ταυτίζεται με αυτήν της βροχής (πρόκειται βέβαια για σύμπτωση, παρά για γενικό κανόνα), ενώ είναι η θεωρητική είναι ελαφρά μικρότερη από την αντίστοιχη της βροχής. Και τα δύο μεγέθη είναι αισθητά μεγαλύτερα από αυτά που υπολογίσαμε για την πραγματική τιμή της απορροής του έτους 1978-79 (1286.6 mm έναντι της εκτιμημένης τιμής 1313.0 mm).

γ. Να υπολογιστεί η μεταβολή αποθέματος του ταμιευτήρα για ένα έτος στο οποίο (1) πραγματοποιείται ελάχιστη βροχόπτωση περιόδου επαναφοράς 25 ετών (2) η απορροή και η εξάτμιση έχουν τις τιμές που στατιστικά αναμένονται για την εν λόγω βροχόπτωση (3) οι ολικές απολήψεις από τον ταμιευτήρα είναι 80% του μέσου ετήσιου όγκου απορροής (4) η μέση επιφάνεια του ταμιευτήρα είναι ίση με 65% της μέγιστης και (5) δεν πραγματοποιούνται υπερχειλίσεις από τον ταμιευτήρα.

Η ελάχιστη βροχόπτωση περιόδου επαναφοράς 25 ετών είναι $1201.0 - 1.7507 \times 252.5 = 758.9$ mm (όπου 1.7507 είναι η τιμή της ανηγμένης κανονικής μεταβλητής z για $F = 1 / T = 1/25 = 0.04$). Αν και η τιμή αυτή αναφέρεται στο σύνολο της λεκάνης, θα θεωρήσουμε ότι την ίδια τιμή έχει και η βροχόπτωση πάνω από τον ταμιευτήρα. Η στατιστικά αναμενόμενη τιμή της απορροής για αυτή την τιμή της βροχής προκύπτει από την εξίσωση παλινδρόμησης του ερωτήματος α. και είναι

$$y = -97.1 + 0.846 \times 758.9 = 544.9 \text{ mm}$$

Η στατιστικά αναμενόμενη τιμή της εξάτμισης είναι ίση με τη μέση τιμή της, δεδομένου ότι δεν υπάρχει συσχέτιση βροχόπτωσης και εξάτμισης, δηλαδή είναι 1423.1 mm.

Η μέση έκταση της επιφάνειας του ταμιευτήρα κατά το εν λόγω έτος είναι

$$A_T = 0.65 \times 80.6 = 52.4 \text{ km}^2$$

Επομένως η έκταση της λεκάνης που δεν καταλαμβάνεται από τον ταμιευτήρα είναι

$$A_\Lambda = 3584 - 52.4 = 3531.6 \text{ km}^2.$$

Κατά συνέπεια ο όγκος απορροής είναι

$$V_Q = 0.5449 \times 3531.6 \times 10^6 = 1924.4 \times 10^6 \text{ m}^3 = 1924.4 \text{ hm}^3$$

ο όγκος βροχόπτωσης

$$V_P = 0.7589 \times 52.4 \times 10^6 = 39.8 \times 10^6 \text{ m}^3 = 39.8 \text{ hm}^3$$

και ο όγκος εξάτμισης

$$V_E = 1.4231 \times 52.4 \times 10^6 = 74.6 \times 10^6 \text{ m}^3 = 74.6 \text{ hm}^3$$

Ο μέσος ετήσιος όγκος απορροής είναι

$$E[V_Q] \approx 0.9189 \times 3531.6 \times 10^6 = 3245.2 \times 10^6 \text{ m}^3 = 3245.2 \text{ hm}^3$$

και κατά συνέπεια η απώληση του υπόψη έτους είναι

$$V_D = 0.80 \times 3245.2 \text{ hm}^3 = 2596.1 \text{ hm}^3$$

Κατά συνέπεια, η μεταβολή του αποθέματος του ταμιευτήρα είναι

$$\begin{aligned} \Delta S &= V_Q + V_P - V_E - V_D = \\ &= 1924.4 + 39.8 - 74.6 - 2596.1 = -706.5 \text{ hm}^3 \end{aligned}$$

7.2 Άλλες γραμμικές εκτιμήσεις δύο μεταβλητών

7.2.1 Ομογενής ευθεία

Σε πολλές περιπτώσεις είναι επιθυμητή η αγνόηση του σταθερού όρου a στην εξίσωση της ευθείας, οπότε οι εξισώσεις (7.24) και (7.25) παίρνουν τη μορφή

$$Y = bX + W \quad (7.43)$$

$$\hat{Y} = bX \quad (7.44)$$

Η τελευταία εξίσωση λέγεται *ομογενής ευθεία*. Η παράμετρος b υπολογίζεται και πάλι έτσι ώστε η (7.44) να αποτελεί την εκτίμηση ελάχιστου μέσου τετραγωνικού σφάλματος, με την έννοια ότι ελαχιστοποιείται το μέγεθος

$$E[W^2] = E[(Y - \hat{Y})^2] = E[(Y - bX)^2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - bx)^2 f_{XY}(x, y) dx dy \quad (7.45)$$

Για την ελαχιστοποίηση παραγωγίζουμε την παραπάνω ως προς b και εξισώνουμε την παράγωγο με 0, οπότε διαδοχικά βρίσκουμε

$$\frac{\partial E[W^2]}{\partial b} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} -2(y - bx) x f_{XY}(x, y) dx dy = 0 \quad (7.46)$$

$$E[(Y - bX)X] = E[XY] - bE[X^2] = 0 \quad (7.47)$$

απ' όπου προκύπτει

$$b = \frac{E[XY]}{E[X^2]} = \frac{\sigma_{XY} + \mu_X \mu_Y}{\sigma_X^2 + \mu_X^2} \quad (7.48)$$

όπου μ_X και μ_Y οι μέσες τιμές των X και Y , αντίστοιχα, σ_X και σ_Y οι αντίστοιχες τυπικές αποκλίσεις και σ_{XY} και ρ_{XY} η συνδιασπορά και ο συντελεστής συσχέτισης, αντίστοιχα, των X και Y . Τα μεγέθη αυτά συνήθως είναι άγνωστα (παράμετροι πληθυσμού). Αν στη θέση των $E[XY]$ και $E[X^2]$ τοποθετήσουμε τις (αμερόληπτες) εκτιμήσεις τους από το δείγμα των n σημείων (x_i, y_i) , τότε παίρνουμε την τελική έκφραση

$$b = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \quad (7.49)$$

η οποία, όπως παρατηρούμε είναι πολύ απλούστερη από την αντίστοιχη εξίσωση (7.10) της πλήρους γραμμικής μορφής.

Εύκολα μπορεί να αποδειχτεί ότι η ομογενής μορφή της εκτιμήτριας \hat{Y} δε διατηρεί ούτε τη μέση τιμή ούτε τη διασπορά, δηλαδή $\mu_{\hat{Y}} \neq \mu_Y$ και $\sigma_{\hat{Y}}^2 \neq \sigma_Y^2$. Μάλιστα η διασπορά της \hat{Y} δεν είναι πάντα μικρότερη από αυτήν της Y , αλλά μπορεί να είναι και μεγαλύτερη. Ως συνέπεια της μη διατήρησης της μέσης τιμής της Y , το μέσο σφάλμα δεν είναι μηδενικό, δηλαδή $\mu_W \neq 0$ και επομένως $\sigma_W^2 \neq E[W^2]$. Ωστόσο ισχύει $E[WX] = 0$.

Ο συντελεστής προσδιορισμού στην περίπτωση αυτή ορίζεται από τη γενική σχέση (7.41). Εύκολα μπορεί να αποδειχτεί ότι στην περίπτωση της ομογενούς ευθείας

$$\delta \neq \rho_{XY}^2 \neq \frac{\sigma_{\hat{Y}}^2}{\sigma_Y^2} \quad (7.50)$$

Ακόμη, επειδή είναι δυνατό να ισχύει $\sigma_{\hat{Y}}^2 < \sigma_W^2$, ο συντελεστής προσδιορισμού είναι δυνατό να πάρει και αρνητικές τιμές. Γενικά πάντως ισχύει $\delta \leq 1$ και η τιμή του δ κοντά στο 1 δείχνει καλή προσαρμογή της ομογενούς ευθείας στα δεδομένα.

Με μια εναλλακτική μέθοδο εκτίμησης της παραμέτρου b της ομογενούς ευθείας είναι δυνατό να διατηρηθεί η μέση τιμή της Y . Πράγματι, παίρνοντας αναμενόμενες τιμές στην (7.44) και θεωρώντας $\mu_{\hat{Y}} = \mu_Y$, βρίσκουμε

$$b = \frac{E[Y]}{E[X]} = \frac{\mu_Y}{\mu_X} \quad (7.51)$$

Είναι αυτονόητο ότι το τετραγωνικό σφάλμα $E[W^2]$ που προκύπτει με αυτό τον τρόπο εκτίμησης της παραμέτρου b δεν είναι το ελάχιστο δυνατό.

7.2.2 Οργανική συσχέτιση

Όπως είδαμε στα προηγούμενα εδάφια, τόσο η πλήρης όσο και η ομογενής γραμμική παλινδρόμηση αποτυγχάνουν στη διατήρηση της διασποράς της εξαρτημένης μεταβλητής Y (ειδικά η ομογενής μπορεί να μη διατηρεί ούτε τη μέση τιμή). Ωστόσο, στην τεχνική υδρολογία είναι συχνά επιθυμητή η διατήρηση της διασποράς, ιδίως στην περίπτωση επέκτασης δειγμάτων (βλ. ενότητα 7.6). Αν εγκαταλείψουμε την απαίτηση του ελάχιστου τετραγωνικού σφάλματος, τότε είναι δυνατό να εκτιμήσουμε τις παραμέτρους της (7.25) σε τρόπο ώστε να διατηρείται και η μέση τιμή και η διασπορά. Αυτός ο τρόπος εκτίμησης των παραμέτρων είναι γνωστός ως *οργανική συσχέτιση* ή και ως *επέκταση διατήρησης διασποράς* (maintenance of variance extension - MOVE.1).

Συγκεκριμένα, η (7.25) σε συνδυασμό με τις απαιτήσεις $\mu_{\hat{Y}} = \mu_Y$ και $\sigma_{\hat{Y}}^2 = \sigma_Y^2$ δίνει

$$\mu_Y = a + b \mu_X \quad (7.52)$$

$$\sigma_Y^2 = b^2 \sigma_X^2 \quad (7.53)$$

απ' όπου προκύπτει

$$b = \text{sgn}(\rho_{XY}) \frac{\sigma_Y}{\sigma_X} \quad (7.54)$$

$$a = \mu_Y - b \mu_X \quad (7.55)$$

όπου $\text{sgn}(\rho_{XY})$ είναι το πρόσημο του συντελεστή συσχέτισης (+1 ή -1). Ο όρος προσήμου έχει τεθεί για να είναι συνεπής η εκτίμηση με την πραγματικότητα, δηλαδή για θετικά συσχετισμένες μεταβλητές να προκύπτει θετική τιμή του b και αντίστροφα (η (7.53) επιτρέπει και τα δύο πρόσημα). Παρατηρούμε ότι η (7.55) ταυτίζεται με την αντίστοιχή της (7.28), ενώ η (7.54) διαφέρει από την (7.27) κατά το ότι το ρ_{XY} έχει αντικατασταθεί από το πρόσημό του. Αν στη θέση των θεωρητικών ροπών του πληθυσμού τοποθετήσουμε τις εκτιμήσεις τους από το δείγμα των n σημείων (x_i, y_i) , τότε παίρνουμε

$$b = \text{sgn}(r_{XY}) \frac{s_Y}{s_X} \quad (7.56)$$

$$a = \bar{y} - b \bar{x} \quad (7.57)$$

Οι ροπές του σφάλματος $W = Y - \hat{Y} = Y - (a + bX)$ προκύπτουν με ανάλογο τρόπο όπως προηγουμένως και είναι

$$\mu_W = 0 \quad (7.58)$$

$$\sigma_W^2 = 2 \sigma_Y^2 (1 - |\rho_{XY}|) \quad (7.59)$$

Όπως αναμέναμε, το μέσο τετραγωνικό σφάλμα είναι γενικά μεγαλύτερο από αυτό της τυπικής γραμμικής παλινδρόμησης. Η μέγιστη τιμή του, ίση με $2 \sigma_Y^2$, εμφανίζεται για $\rho_{XY} = 0$. Η τιμή αυτή είναι διπλάσια από την

αντίστοιχη της τυπικής γραμμικής παλινδρόμησης. Η ελάχιστη τιμή του, ίση με 0 όπως και στην τυπική γραμμική παλινδρόμηση, εμφανίζεται όταν $|\rho_{XY}| = 1$.

Το σφάλμα W στην περίπτωση της οργανικής παλινδρόμησης δεν είναι ασυσχέτιστο με τη μεταβλητή X . Η αντίστοιχη συνδιασπορά είναι

$$\sigma_{WX} = -\text{sgn}(\rho_{XY}) (1 - |\rho_{XY}|) \sigma_X \sigma_Y \quad (7.60)$$

Ο συντελεστής προσδιορισμού είναι

$$\delta = 2|\rho_{XY}| - 1 \quad (7.61)$$

Παρατηρούμε ότι ο συντελεστής προσδιορισμού γίνεται ίσος με 1 για $|\rho_{XY}| = 1$, μηδενίζεται για $|\rho_{XY}| = 0.5$ και παίρνει αρνητικές τιμές για $|\rho_{XY}| < 0.5$. Κατά συνέπεια δεν έχει πρακτικό νόημα η εφαρμογή της οργανικής συσχέτισης για $|\rho_{XY}| < 0.5$.

Η εξίσωση (7.25), αν αντικαταστήσουμε τις τιμές των a και b από τις (7.55) και (7.54), παίρνει την ακόλουθη μορφή (για $\rho_{XY} > 0$)

$$\frac{\hat{Y} - \mu_Y}{\sigma_Y} = \frac{X - \mu_X}{\sigma_X} \quad (7.62)$$

Σε περίπτωση που η X ακολουθεί κανονική κατανομή, την ίδια κατανομή θα ακολουθεί και η \hat{Y} , οπότε η παραπάνω εξίσωση ισοδυναμεί με την

$$F_{\hat{Y}}(\hat{y}) = F_X(x), \quad (7.63)$$

η οποία σημαίνει ότι η εκτίμηση \hat{Y} είναι ισοπίθανη με την αντίστοιχη τιμή της X .

Τέλος, στην περίπτωση της οργανικής συσχέτισης η αντίστροφη εξίσωση που εκφράζει την εκτίμηση της \hat{X} για δεδομένη Y δεν είναι άλλη από αυτή που προκύπτει με τη συνήθη αλγεβρική αντιστροφή, δηλαδή η

$$\hat{X} = (Y - a) / b \quad (7.64)$$

Δηλαδή, εδώ δεν έχουμε δύο διαφορετικές ευθείες όπως στην τυπική γραμμική παλινδρόμηση.

Εφαρμογή 7.2

α. Να βρεθούν οι εξισώσεις ανάμεσα στις μεταβλητές P και Q της Εφαρμογής 7.1, χρησιμοποιώντας την ομογενή ευθεία και την οργανική συσχέτιση.

Υπενθυμίζουμε ότι τα χαρακτηριστικά αθροίσματα των δειγμάτων X ($\equiv P$) και Y ($\equiv Q$) της Εφαρμογής 7.1 είναι:

$$n = 25, \sum x = 30\,025.5, \sum y = 22\,973.6, \sum xy = 28\,941\,151, \\ \sum x^2 = 37\,655\,556, \sum y^2 = 22\,487\,374$$

Τα στατιστικά χαρακτηριστικά των μεταβλητών είναι

$$\bar{x} = 1201.0, \bar{y} = 918.9, s_X = 252.5, s_Y = 234.6, r_{XY} = 0.911,$$

$$s_{XY} = r_{XY} s_X s_Y = 0.911 \times 252.5 \times 234.6 = 53\,964$$

Εφαρμόζοντας την εξίσωση (7.49) υπολογίζουμε ότι ο συντελεστής της ομογενούς εξίσωσης ελάχιστου τετραγωνικού σφάλματος είναι

$$b = \sum xy / \sum x^2 = 28\,941\,151 / 37\,655\,556 = 0.769$$

(έναντι 0.846 της πλήρους ευθείας). Εναλλακτικά, ο συντελεστής που διατηρεί τη μέση τιμή στην ομογενή ευθεία δίνεται από την (7.51) και είναι

$$b = \bar{y} / \bar{x} = 918.9 / 1201.0 = 0.765$$

Οι συντελεστές b και a της οργανικής συσχέτισης δίνονται από τις εξισώσεις (7.54) και (7.55), και είναι

$$b = +s_Y / s_X = 234.6 / 252.5 = 0.929$$

και

$$a = \bar{y} - b \bar{x} = 918.9 - 0.929 \times 1201.0 = -196.8$$

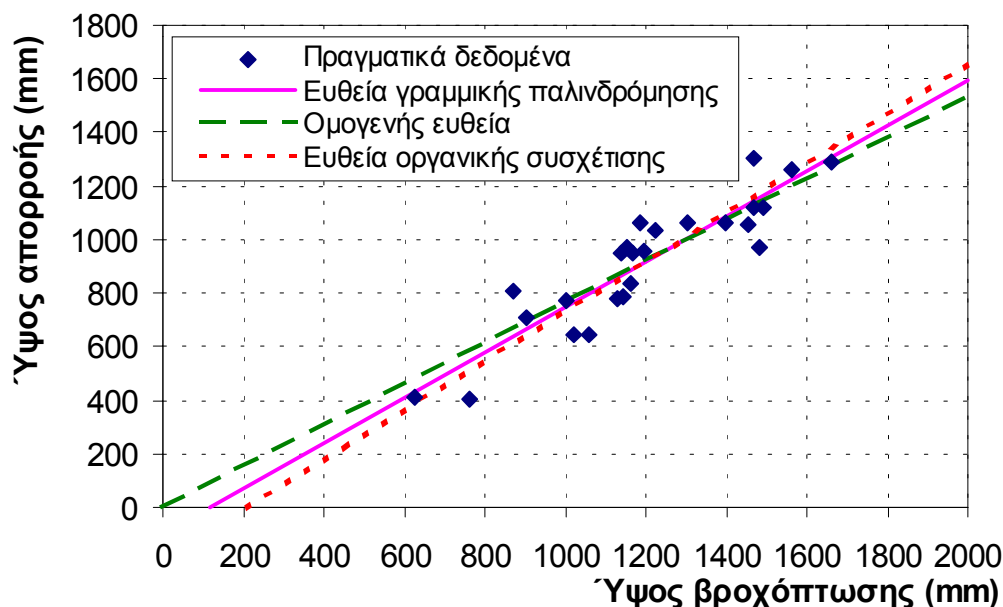
Οι διάφορες εξισώσεις που έχουν προσδιοριστεί παραπάνω καθώς και στην Εφαρμογή 7.1 φαίνονται στο Σχ. 7.9.

β. Να επανεξεταστεί το ερώτημα β. της Εφαρμογής 7.1, χρησιμοποιώντας την οργανική συσχέτιση.

Αν δε συμπεριλάβουμε το έτος 1978-79 στα δείγματα της βροχής και απορροής τότε τα στατιστικά χαρακτηριστικά των δειγμάτων γίνονται

$$\bar{x} = 1181.8, \bar{y} = 903.6, s_X = 239.2, s_Y = 226.9$$

και η εξίσωση οργανικής συσχέτισης γίνεται $y = -217.0 + 0.948 x$. Εφαρμόζοντας αυτή την εξίσωση για $x = 1661.2$ βρίσκουμε $y = 1358.2$ (έναντι 1313.0 που έχει δώσει η τυπική γραμμική παλινδρόμηση).



Σχ. 7.9 Ετήσιο ύψος επιφανειακής απορροής συναρτήσει του ετήσιου ύψους βροχόπτωσης στη λεκάνη ανάντη Κρεμαστών: Πραγματικά δεδομένα και εναλλακτικές γραμμικές εξισώσεις.

Με τα παραπάνω στατιστικά χαρακτηριστικά των 24 ετών η ανηγμένη μεταβλητή για την απορροή είναι

$$z = (1358.2 - 903.6) / 226.9 = 2.00$$

που αντιστοιχεί σε $F = 0.9774$ και $T = 1 / (1 - 0.9774) = 44.3$. Την ίδια τιμή της περιόδου επαναφοράς βρίσκουμε και για τη βροχόπτωση, αν χρησιμοποιήσουμε τα στατιστικά χαρακτηριστικά των 25 ετών, δεδομένου ότι

$$z = (1661.2 - 1181.8) / 239.2 = 2.00$$

Αν χρησιμοποιήσουμε τα στατιστικά χαρακτηριστικά των 25 ετών, που για τη βροχόπτωση είναι

$$\bar{x} = 1201.0, s_X = 252.5$$

ενώ για την απορροή γίνονται (με την τιμή 1358.2 για το έτος 1978-79)

$$\bar{y} = 921.8, s_Y = 239.5$$

η ανηγμένη μεταβλητή για την απορροή είναι

$$z = (1358.2 - 921.8) / 239.5 = 1.82$$

που αντιστοιχεί σε $F = 0.9657$ και $T = 1 / (1 - 0.9657) = 29.2$. Την ίδια τιμή της περιόδου επαναφοράς βρίσκουμε και για τη βροχόπτωση, δεδομένου ότι

$$z = (1661.2 - 1201.0) / 252.5 = 29.2$$

(βλ. και Εφαρμογή 7.1). Έτσι, επιβεβαιώνουμε και αριθμητικά την ταύτιση των περιόδων επαναφοράς του αρχικού και εκτιμημένου μεγέθους, στην περίπτωση που χρησιμοποιείται η οργανική συσχέτιση και η κατανομή των δύο μεταβλητών είναι κανονική. Βεβαίως αυτό δεν σημαίνει ότι η οργανική συσχέτιση είναι πιο κοντά στην πραγματικότητα από την τυπική γραμμική παλινδρόμηση. Αντίθετα, στο παράδειγμα που εξετάσαμε η εκτίμηση της οργανικής συσχέτισης διαφέρει περισσότερο από την πραγματική τιμή της απορροής του έτους 1978-79, απ' ότι διαφέρει η αντίστοιχη εκτίμηση της τυπικής γραμμικής παλινδρόμησης.

7.3 Γενίκευση της γραμμικής παλινδρόμησης

7.3.1 Γραμμική παλινδρόμηση πολλών μεταβλητών

Η γραμμική παλινδρόμηση δύο μεταβλητών που εξετάστηκε στην προηγούμενη ενότητα γενικεύεται εύκολα για περισσότερες από δύο μεταβλητές. Θεωρούμε ότι οι $m + 1$ μεταβλητές X_1, X_2, \dots, X_m, Y , συνδέονται με το γραμμικό νόμο

$$Y = a_0 + a_1X_1 + \dots + a_mX_m + W \quad (7.65)$$

όπου a_0, a_1, \dots, a_m άγνωστες αριθμητικές παράμετροι και W τυχαία μεταβλητή που εκφράζει το σφάλμα της εκτίμησης

$$\hat{Y} = a_0 + a_1X_1 + \dots + a_mX_m \quad (7.66)$$

δηλαδή $W = Y - \hat{Y} = Y - (a_0 + a_1X_1 + \dots + a_mX_m)$. Οι παράμετροι υπολογίζονται έτσι ώστε να ελαχιστοποιείται το μέσο τετραγωνικό σφάλμα, δηλαδή το μέγεθος

$$\begin{aligned}
E[W^2] &= E[(Y - \hat{Y})^2] = E[(Y - (a_0 + a_1X_1 + \dots + a_mX_m))^2] \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y - (a_0 + a_1x_1 + \dots + a_mx_m)]^2 f_{XY}(x, y) dx dy \quad (7.67)
\end{aligned}$$

Παραγωγίζοντας την (7.26) ως προς a_i παίρνουμε μετά από πράξεις (βλ. π.χ. Papoulis, 1990, σ. 450) ένα γραμμικό σύστημα $m + 1$ εξισώσεων, το οποίο υπό μορφή μητρώων γράφεται

$$\mathbf{c}_{XX} \mathbf{a} = \mathbf{c}_{XY} \quad (7.68)$$

όπου το \mathbf{c}_{XX} είναι συμμετρικό τετραγωνικό μητρώο συντελεστών διάστασης $(m + 1) \times (m + 1)$ και το \mathbf{c}_{XY} είναι το διάνυσμα σταθερών όρων διάστασης $(m + 1)$:

$$\mathbf{c}_{XX} = \begin{bmatrix} 1 & E[X_1] & \dots & E[X_m] \\ E[X_1] & E[X_1^2] & \dots & E[X_1X_m] \\ \vdots & \vdots & \ddots & \vdots \\ E[X_m] & E[X_mX_1] & \dots & E[X_m^2] \end{bmatrix} \quad \mathbf{c}_{XY} = \begin{bmatrix} E[Y] \\ E[X_1Y] \\ \vdots \\ E[X_mY] \end{bmatrix} \quad (7.69)$$

ενώ \mathbf{a} είναι το διάνυσμα των αγνώστων διάστασης $(m + 1)$:

$$\mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} \quad (7.70)$$

Η επίλυση του συστήματος δίνει τις τιμές των αγνώστων παραμέτρων. Για την εφαρμογή αντικαθιστούμε τις θεωρητικές αναμενόμενες τιμές με τις εκτιμήσεις τους από το δείγμα των n σημείων $(x_{1i}, x_{2i}, \dots, x_{mi}, y_i)$:

$$E[X_s] = \bar{x}_s, \quad E[Y] = \bar{y}, \quad E[X_sX_r] = \frac{1}{n} \sum_{i=1}^n x_{si} x_{ri}, \quad E[X_sY] = \frac{1}{n} \sum_{i=1}^n x_{si} y_i \quad (7.71)$$

όπου $s, r = 1, \dots, m$.

Στην περίπτωση της ομογενούς εξίσωσης $\hat{Y} = a_1X_1 + \dots + a_mX_m$ (χωρίς το σταθερό όρο a_0) η επίλυση είναι παρόμοια. Η μαθηματική έκφραση παραμένει ίδια, με τη διαφορά ότι διαγράφεται η πρώτη σειρά και η πρώτη στήλη του μητρώου \mathbf{c}_{XX} και η πρώτη σειρά των διανυσμάτων \mathbf{c}_{XY} και \mathbf{a} .

Για τον υπολογισμό του συντελεστή προσδιορισμού, ο οποίος δείχνει το μέτρο της προσαρμογής της γραμμικής εξίσωσης προς τα δεδομένα εφαρμόζεται η γενικευμένη εξίσωση (7.41) ή η αντίστοιχη της (7.14), αφού προηγουμένως υπολογιστεί το τετραγωνικό σφάλμα.

7.3.2 Μη γραμμική παλινδρόμηση με γραμμικούς συντελεστές

Το παραπάνω γενικευμένο γραμμικό μοντέλο μπορεί να εφαρμοστεί άμεσα και για μη γραμμικό ως προς X_i νόμο $Y = \varphi(X_1, X_2, \dots) + W$, με την προϋπόθεση ότι ο νόμος είναι γραμμικός ως προς τις άγνωστες παραμέτρους a_0, a_1, \dots . Για παράδειγμα η πολυωνυμική παλινδρόμηση μιας μεταβλητής βαθμού m :

$$Y = a_0 + a_1X + a_2X^2 + \dots + a_mX^m + W \quad (7.72)$$

ανάγεται άμεσα στην γραμμική παλινδρόμηση m μεταβλητών (7.65) αν τεθεί $X_1 = X, X_2 = X^2, \dots, X_m = X^m$.

7.3.3 Γραμμικοποίηση με μετασχηματισμούς μεταβλητών

Η παραπάνω τεχνική δεν μπορεί να εφαρμοστεί όταν η συνάρτηση $\varphi(\)$ δεν είναι γραμμική ως προς τις παραμέτρους. Σε τέτοιες περιπτώσεις είναι μερικές φορές δυνατό να γίνει γραμμικοποίηση με κατάλληλο μετασχηματισμό των μεταβλητών. Παραδείγματα αυτού του τύπου αποτελούν η εκθετική συνάρτηση και η συνάρτηση δύναμης μιας ή περισσότερων μεταβλητών. Στον Πίν. 7.3 δίνονται οι απαιτούμενοι μετασχηματισμοί μεταβλητών και παραμέτρων για τις συναρτήσεις αυτές. Οι μετασχηματισμοί προκύπτουν με λογαρίθμιση και των δύο μελών των αρχικών εξισώσεων.

Τα στατιστικά χαρακτηριστικά που απαιτούνται για τον υπολογισμό των παραμέτρων αναφέρονται στις μετασχηματισμένες μεταβλητές και όχι στις αρχικές. Επίσης το μέσο τετραγωνικό σφάλμα και ο συντελεστής προσδιορισμού υπολογίζονται για τις μετασχηματισμένες μεταβλητές.

Πίν. 7.3 Τυπικές συναρτήσεις που επιδέχονται γραμμικοποίηση με μετασχηματισμό μεταβλητών και αντίστοιχοι μετασχηματισμοί.

| Τύπος αρχικής συνάρτησης | Εκφραση αρχικής συνάρτησης | Γραμμικοποιημένη συνάρτηση | Μετασχηματισμοί μεταβλητών | Μετασχηματισμοί παραμέτρων |
|----------------------------|--|--|--|--|
| Εκθετική δύο μεταβλητών | $\hat{Y} = \kappa e^{\lambda X}$ | $\hat{Y}' = a + bX'$ | $\hat{Y}' = \ln \hat{Y}$ $X' = X$ | $a = \ln \kappa$ $b = \lambda$ |
| Δύναμης δύο μεταβλητών | $\hat{Y} = \kappa X^\lambda$ | $\hat{Y}' = a + bX'$ | $\hat{Y}' = \ln \hat{Y}$ $X' = \ln X$ | $a = \ln \kappa$ $b = \lambda$ |
| Εκθετική πολλών μεταβλητών | $\hat{Y} = \kappa e^{\lambda_1 X_1 + \lambda_2 X_2 + \dots}$ | $\hat{Y}' = a_0 + a_1 X_1' + a_2 X_2' + \dots$ | $\hat{Y}' = \ln \hat{Y}$ $X_1' = X_1,$ $X_2' = X_2, \dots$ | $a = \ln \kappa$ $b_1 = \lambda_1,$ $b_2 = \lambda_2, \dots$ |
| Δύναμης πολλών μεταβλητών | $\hat{Y} = \kappa X_1^{\lambda_1} X_2^{\lambda_2} \dots$ | $\hat{Y}' = a_0 + a_1 X_1' + a_2 X_2' + \dots$ | $\hat{Y}' = \ln \hat{Y}$ $X_1' = \ln X_1,$ $X_2' = \ln X_2, \dots$ | $a = \ln \kappa$ $b_1 = \lambda_1,$ $b_2 = \lambda_2, \dots$ |

Εφαρμογή 7.3.3

Στον υδρομετρικό σταθμό Αχλαδόκαστρο του ποταμού Ευήνου έχουν γίνει συστηματικές μετρήσεις στάθμης και παροχής του ποταμού. 35 από αυτές που αναφέρονται σε διάρκεια 2.5 ετών, κατά την οποία δεν υπήρχαν ουσιαστικές μεταβολές της γεωμετρίας και των χαρακτηριστικών της κοίτης του ποταμού, φαίνονται στον Πίν. 7.4. Με βάση τις μετρήσεις αυτές, να περιγραφεί μαθηματικά η σχέση στάθμης και παροχής του ποταμού για την υπόψη περίοδο.

Η εφαρμογή αυτή αποτελεί ένα από τα πιο τυπικά προβλήματα της τεχνικής υδρολογίας. Η κατασκευή καμπύλης στάθμης-παροχής είναι απαραίτητη προϋπόθεση για την εξαγωγή της χρονοσειράς της παροχής σε κάθε θέση υδρομέτρησης, με βάση την αντίστοιχη χρονοσειρά της στάθμης.

Πίν. 7.4 Ταυτόχρονες μετρήσεις στάθμης και παροχής του ποταμού Ευήνου στη θέση Αχλαδόκαστρο, για την περίοδο από Μάρτιο 1974 μέχρι Αύγουστο 1976.

| A/A | Ημερο- μηνία | Στάθμη (m) | Παροχή (m ³ /s) | A/A | Ημερο- μηνία | Στάθμη (m) | Παροχή (m ³ /s) |
|-----|-----------------|---------------|-------------------------------|-----|-----------------|---------------|-------------------------------|
| 1 | 19/03/74 | 1.65 | 32.174 | 19 | 19/07/75 | 0.58 | 1.447 |
| 2 | 29/03/74 | 1.51 | 21.537 | 20 | 31/07/75 | 0.68 | 2.448 |
| 3 | 29/04/74 | 1.37 | 23.413 | 21 | 07/08/75 | 0.77 | 3.646 |
| 4 | 27/05/74 | 1.12 | 17.157 | 22 | 12/08/75 | 0.66 | 2.200 |
| 5 | 24/06/74 | 0.79 | 4.252 | 23 | 18/09/75 | 0.49 | 1.030 |
| 6 | 16/07/74 | 0.63 | 1.894 | 24 | 24/09/75 | 0.48 | 0.920 |
| 7 | 02/09/74 | 0.56 | 1.067 | 25 | 08/01/76 | 1.05 | 12.066 |
| 8 | 20/09/74 | 0.54 | 0.891 | 26 | 16/02/76 | 1.35 | 25.899 |
| 9 | 15/10/74 | 0.69 | 1.544 | 27 | 19/03/76 | 1.08 | 14.655 |
| 10 | 13/11/74 | 1.35 | 25.314 | 28 | 13/04/76 | 1.49 | 40.192 |
| 11 | 29/11/74 | 1.42 | 34.201 | 29 | 21/05/76 | 0.99 | 7.484 |
| 12 | 16/01/75 | 1.13 | 15.769 | 30 | 02/06/76 | 0.87 | 5.350 |
| 13 | 21/01/75 | 1.06 | 10.907 | 31 | 11/06/76 | 0.83 | 4.296 |
| 14 | 07/03/75 | 1.19 | 18.037 | 32 | 26/06/76 | 0.75 | 3.491 |
| 15 | 17/04/75 | 1.05 | 10.494 | 33 | 13/07/76 | 0.67 | 2.566 |
| 16 | 28/04/75 | 0.99 | 8.080 | 34 | 30/07/76 | 0.63 | 2.076 |
| 17 | 09/06/75 | 0.81 | 4.600 | 35 | 18/08/76 | 0.54 | 1.243 |
| 18 | 10/07/75 | 0.60 | 1.872 | | | | |

Στο Σχ. 7.10 έχει απεικονιστεί το σύνολο των μετρήσεων του Πίν. 7.4 υπό μορφή διαγράμματος παροχής συναρτήσεως της στάθμης. Έχει καθιερωθεί (για λόγους εποπτικότερης παρουσίασης) σε τέτοια διαγράμματα η στάθμη z να απεικονίζεται στον κατακόρυφο άξονα, παρόλο που αποτελεί την ανεξάρτητη μεταβλητή του προβλήματος (που κατά κανόνα τοποθετείται στον οριζόντιο άξονα). Παρατηρούμε ότι υπάρχει σαφής συσχέτιση στάθμης και παροχής, η οποία όμως δεν είναι γραμμική. Άλλωστε, από την υδραυλική γνωρίζουμε ότι η σχέση στάθμης-παροχής περιγράφεται κατά προσέγγιση από μια εξίσωση δύναμης, δηλαδή της μορφής $q = \kappa z^{\lambda}$. Όπως είδαμε στον Πίν. 7.3, η εξίσωση αυτής της μορφής γραμμικοποιείται με λογαριθμικό μετασχηματισμό και των δύο μεταβλητών. Ισοδύναμα, το σημειοσύνολο ευθειοποιείται αν παρασταθεί σε διπλό λογαριθμικό χαρτί. Πράγματι, αυτό έχει γίνει στο Σχ. 7.11, όπου πράγματι διαπιστώνουμε την γραμμικότητα της σχέσης ανάμεσα στους λογαρίθμους των μεταβλητών.

Χρησιμοποιώντας τις μετασχηματισμένες μεταβλητές $X \equiv \ln Z$ και $Y \equiv \ln Q$, υπολογίζουμε τα αθροίσματα που θα μας επιτρέψουν τον υπολογισμό των παραμέτρων της γραμμικής σχέσης. Έτσι έχουμε:

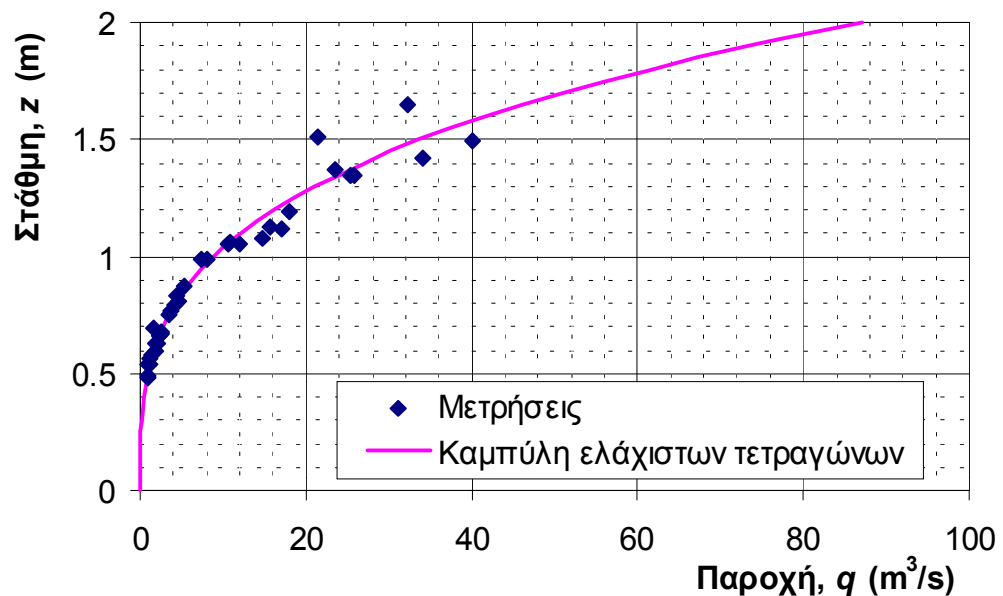
$$n = 35, \sum x = -4.933, \sum y = 60.107, \sum xy = 6.065,$$

$$\sum x^2 = 5.107, \sum y^2 = 152.323$$

Κατά συνέπεια από την εξίσωση (7.18) βρίσκουμε $r = 0.988$. Η τιμή αυτή είναι μεγαλύτερη από την κρίσιμη τιμή

$$r_c = 2/\sqrt{35} = 0.34$$

πράγμα που επιβεβαιώνει και στατιστικά η ύπαρξη ισχυρής γραμμικής συσχέτισης ανάμεσα στους λογαρίθμους της στάθμης και της παροχής.



Σχ. 7.10 Παροχή συναρτήσει της στάθμης στη θέση Αχλαδόκαστρο του ποταμού Ευήνου: μετρήσεις και προσαρμοσμένη με τη μέθοδο ελάχιστων τετραγώνων καμπύλη δύναμης.

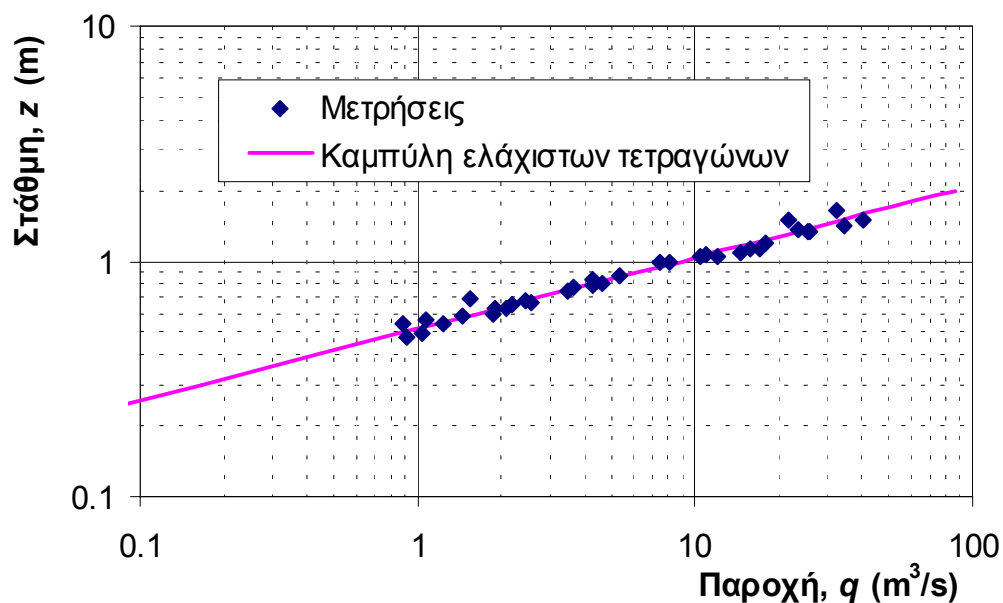
Στη συνέχεια υπολογίζουμε τους συντελεστές b και a της γραμμικής παλινδρόμησης $y = a + b x$, χρησιμοποιώντας τις εξισώσεις (7.10) και (7.11), αντίστοιχα. Έτσι, από την (7.10) παίρνουμε $b = 3.296$ και από την (7.11) $a = 2.182$, οπότε η ζητούμενη ευθεία ελάχιστων τετραγώνων είναι η

$$y = 2.182 + 3.296 x$$

Υψώνοντας τα δύο μέλη της στην e , παίρνουμε την τελική έκφραση

$$q = 8.86 z^{3.296}$$

η οποία έχει παρασταθεί γραφικά στο Σχ. 7.10 καθώς και στο Σχ. 7.11.



Σχ. 7.11 Παροχή συναρτήσει της στάθμης στη θέση Αχλαδόκαστρο του ποταμού Ευήνου σε διπλό λογαριθμικό χαρτί: μετρήσεις και προσαρμοσμένη με τη μέθοδο ελαχίστων τετραγώνων καμπύλη δύναμης.

Τελειώνοντας το παράδειγμά μας, θα πρέπει να σημειώσουμε ότι η κατάρτιση καμπυλών στάθμης-παροχής δεν είναι τόσο απλό πρόβλημα όσο φαίνεται στην εν λόγω εφαρμογή. Οι δυσκολίες προκύπτουν από τις μεταβαλλόμενες ιδιότητες της διατομής του ποταμού (γεωμετρία, τραχύτητα, κλίση), που οδηγούν σε μετατοπιζόμενες με το χρόνο καμπύλες στάθμης-παροχής. Μια επιπρόσθετη δυσκολία εισάγουν τα συχνά σφάλματα των υδρομετρήσεων.

7.4 Γενική παλινδρόμηση

7.4.1 Μεταβλητές με τυχούσα συνάρτηση κατανομής

Στις προηγούμενες ενότητες μελετήθηκαν διάφορα μοντέλα παλινδρόμησης, τα οποία στην περίπτωση των δύο μεταβλητών γράφονται με τη γενική έκφραση

$$\hat{Y} = \phi(X) \quad (7.73)$$

Η συνάρτηση $\phi(X)$ μπορεί να είναι π.χ. γραμμική, παραβολική, πολυωνυμική m βαθμού κτλ. Σε όλες τις περιπτώσεις που εξετάστηκαν η μορφή

της συνάρτησης είχε προεπιλεγεί πριν από την εφαρμογή της μεθόδου ελάχιστων τετραγώνων, κάτι το οποίο γίνεται συχνά στην πράξη, μετά από μια αρχική εξερεύνηση των δεδομένων. Εύλογα, λοιπόν, τίθεται το ερώτημα εάν υπάρχει βέλτιστη συνάρτηση $\phi(X)$ και πώς αυτή μπορεί να προσδιοριστεί. Η βέλτιστη συνάρτηση, αν υπάρχει, θα είναι αυτή για την οποία το μέσο τετραγωνικό σφάλμα είναι μικρότερο από ό,τι σε κάθε άλλη συνάρτηση.

Αρχικά διαπιστώνουμε ότι το ερώτημα αυτό δεν έχει ουσιαστικό νόημα στην περίπτωση που αντιμετωπίζουμε το πρόβλημα με την προσδιοριστική προσέγγιση. Πράγματι, αν δοθεί ένα σύνολο n σημείων (x_i, y_i) τότε υπάρχουν άπειρες καμπύλες (δηλαδή συναρτήσεις) που περνούν από όλα τα σημεία (x_i, y_i) , δίνοντας έτσι μηδενικό τετραγωνικό σφάλμα. Μια απ' αυτές είναι το ταυτοτικό πολυώνυμο βαθμού $n - 1$. Μια άλλη είναι η τεθλασμένη που συνδέει όλα τα σημεία κατά τέτοιο τρόπο ώστε να αποτελεί συνάρτηση (δηλαδή κατά σειρά μεγέθους της τετμημένης x).

Το ερώτημα αποκτά νόημα όταν αντιμετωπιστεί με την πιθανοτική προσέγγιση, οπότε η απάντηση δεν μπορεί να εξαρτάται από το συγκεκριμένο σημειοσύνολο που δίνεται, αλλά από την από κοινού συνάρτηση πυκνότητας πιθανότητας των δύο μεταβλητών $f_{XY}(x, y)$. Αποδεικνύεται, λοιπόν, (π.χ. Papoulis, 1990, σ. 183) ότι υπάρχει μια μοναδική (βέλτιστη) συνάρτηση που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα, η οποία δίνεται από τη σχέση

$$\phi(x) = E[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(x, y) dy \quad (7.74)$$

Υπενθυμίζουμε ότι $E[Y|X = x]$ είναι η δεσμευμένη μέση τιμή της Y για δεδομένη τιμή της $X = x$ (βλ. εξ. (2.37)), ενώ $f_{Y|X}(x, y) = f_{XY}(x, y) / f_X(x)$ είναι η δεσμευμένη συνάρτηση πυκνότητας πιθανότητας της Y για δεδομένη τιμή της $X = x$ (βλ. εξ. (2.32)).

Ένα επεξηγηματικό σκαρίφημα για τη γενική παλινδρόμηση δίνεται στο Σχ. 7.12.

7.4.2 Μεταβλητές με κανονική κατανομή

Εισαγωγικές έννοιες για τη διδιάστατη κανονική κατανομή

Δύο μεταβλητές X και Y λέμε ότι έχουν από κοινού κανονική συνάρτηση κατανομής (ή διδιάστατη κανονική κατανομή), όταν κάθε γραμμικός συνδυασμός των μεταβλητών $Z = \alpha X + \beta Y$ ακολουθεί κανονική κατανομή. Αποδεικνύεται (βλ. π.χ. Papoulis, 1990, σ. 162) ότι η από κοινού συνάρτηση πυκνότητας πιθανότητας των μεταβλητών σε αυτή την περίπτωση είναι

$$f_{XY}(x, y) = \frac{\exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right]}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \quad (7.75)$$

Η κατανομή έχει πέντε παραμέτρους, τις μ_X , μ_Y , σ_X , σ_Y και ρ .

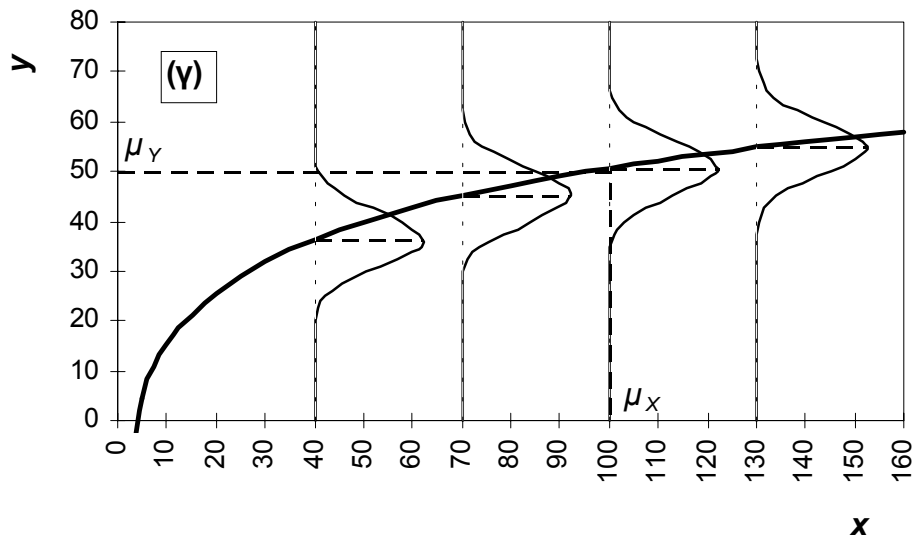
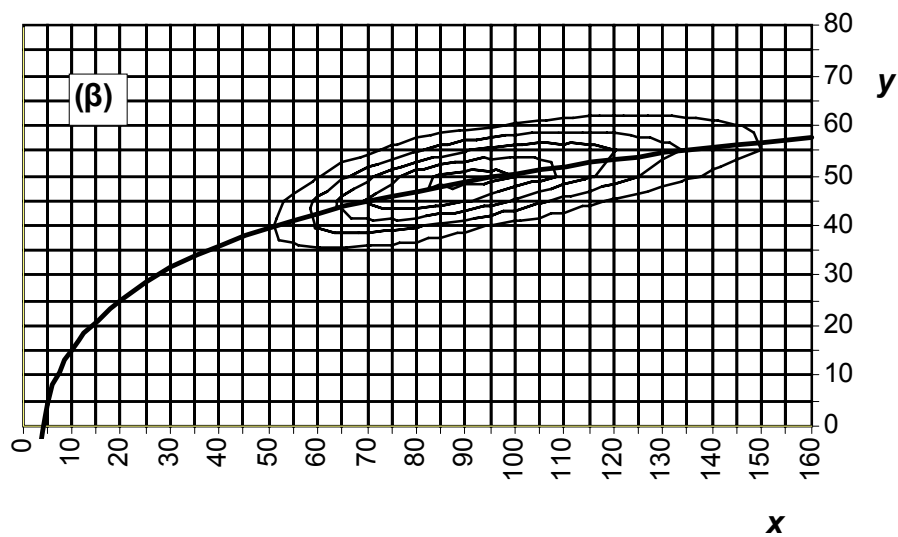
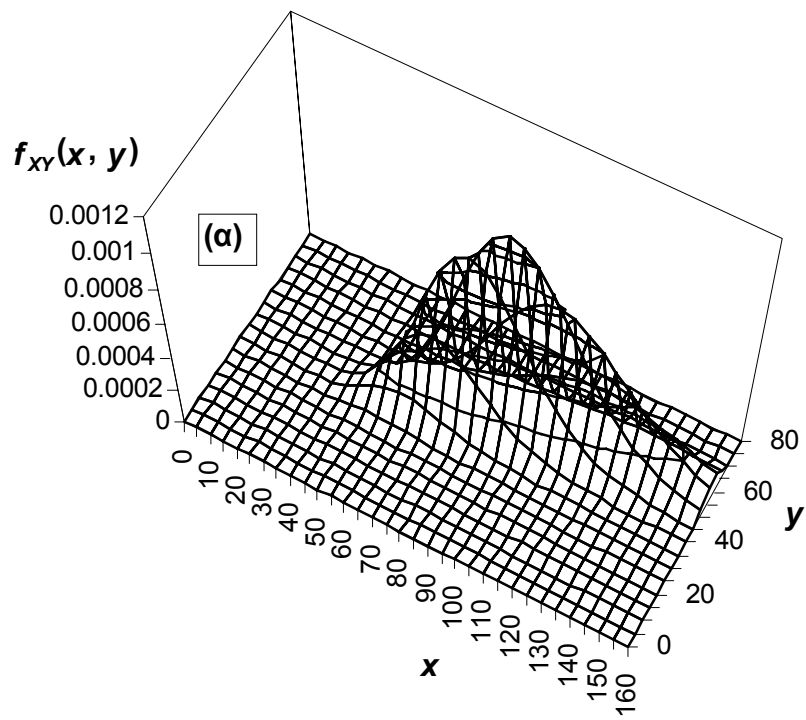
Όπως προκύπτει άμεσα από τον ορισμό οι περιθώριες κατανομές των X και Y είναι κανονικές και οι δύο. Ειδικότερα, αποδεικνύεται ότι η μέση τιμή της X είναι μ_X και η τυπική απόκλισή της είναι σ_X . Επίσης, η μέση τιμή της Y είναι μ_Y και η τυπική απόκλισή της σ_Y . Τέλος ο συντελεστής συσχέτισης των δύο μεταβλητών, σύμφωνα με τον ορισμό της (2.35), είναι ρ .

Η διδιάστατη κανονική κατανομή έχει ορισμένες ενδιαφέρουσες ιδιότητες σχετικά με τη συνεπαγόμενη βέλτιστη παλινδρόμηση, οι οποίες μελετώνται στο επόμενο εδάφιο.

Η γραμμική παλινδρόμηση ως συνέπεια της γενικής παλινδρόμησης

Συνδυάζοντας την εξίσωση (7.75) και τον ορισμό της δεσμευμένης συνάρτησης πυκνότητας πιθανότητας (εξίσωση (2.32)) βρίσκουμε εύκολα ότι η τελευταία για μεταβλητές X , Y που ακολουθούν διδιάστατη κανονική κατανομή έχει την έκφραση

$$f_{Y|X}(x, y) = \frac{1}{\sigma_Y\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{\left[y - \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)\right]^2}{2\sigma_Y^2(1-\rho^2)}\right\} \quad (7.76)$$



1. Η δεσμευμένη συνάρτηση κατανομής της Y για δεδομένη τιμή της $X = x$ είναι κανονική. Αυτό διαπιστώνεται από τη σύγκριση της παραπάνω εξίσωσης με την πυκνότητα πιθανότητας της κανονικής κατανομής (εξίσωση (2.60)).
2. Η δεσμευμένη μέση τιμή της Y για δεδομένη τιμή της $X = x$ είναι

$$\mu_{Y|X} = E[Y|X = x] = \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X) \quad (7.77)$$

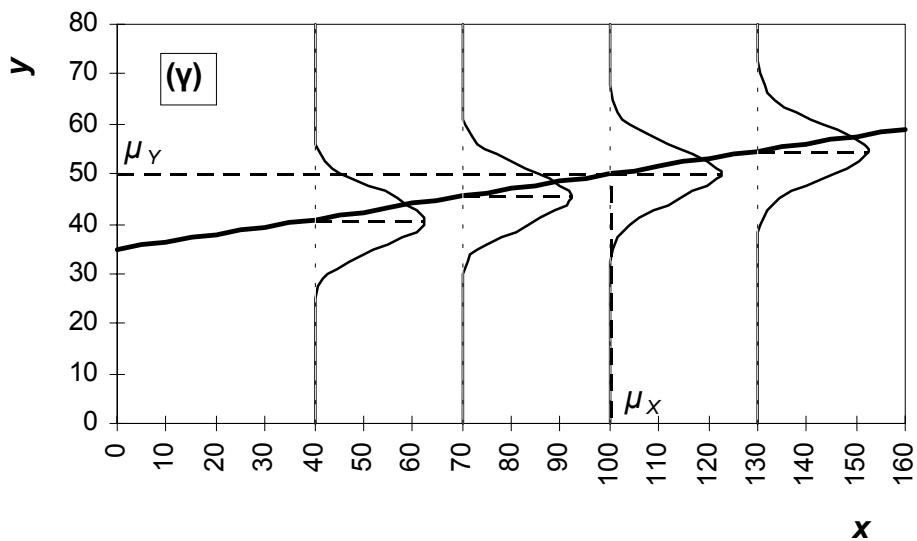
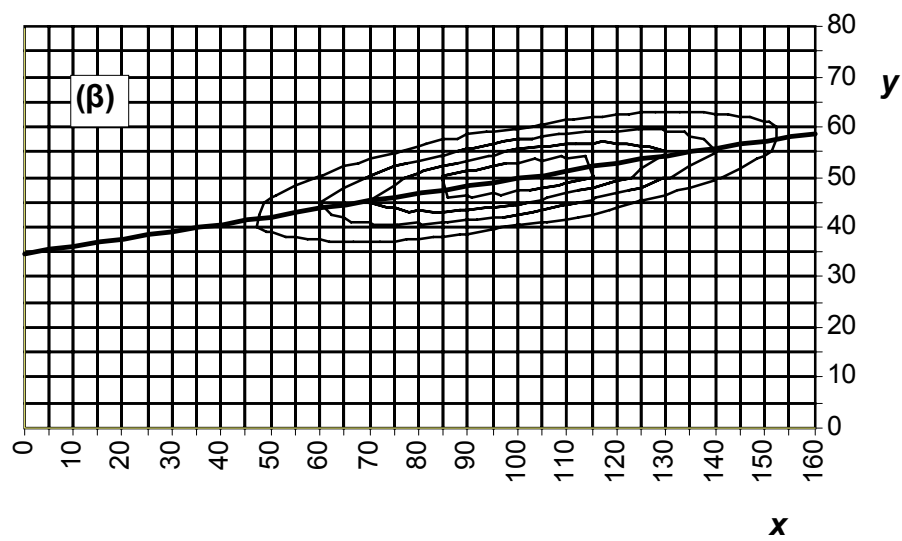
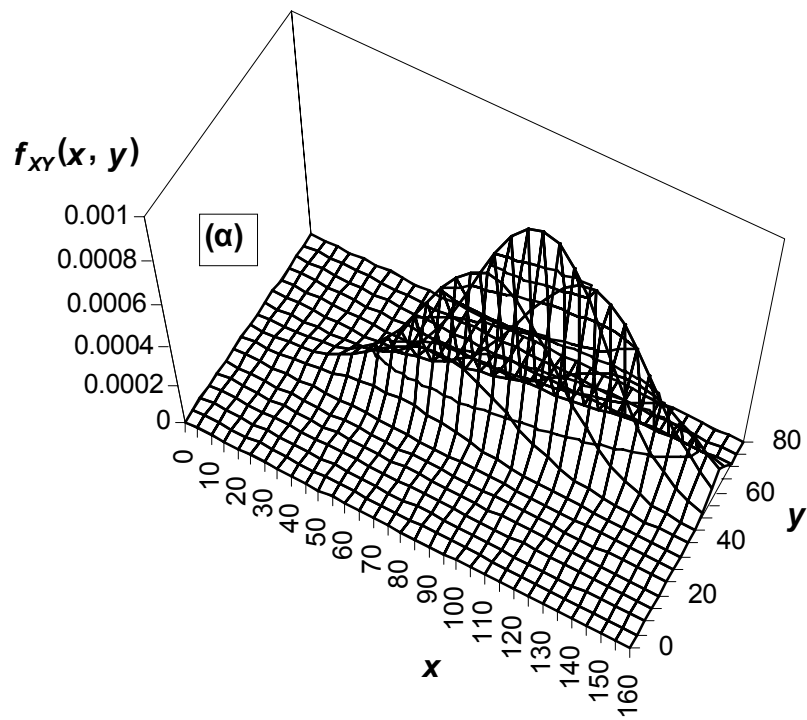
Παρατηρούμε ότι η δεσμευμένη μέση τιμή είναι γραμμική συνάρτηση της x . Εξ άλλου, από την (7.74) προκύπτει ότι η γενική (βέλτιστη) παλινδρόμηση $\phi(x)$ ταυτίζεται με την παραπάνω γραμμική συνάρτηση. Κατά συνέπεια στην κανονική κατανομή η βέλτιστη παλινδρόμηση είναι η γραμμική. Τέλος, συγκρίνοντας τις παραπάνω εξισώσεις με τις (7.54) και (7.55) που είχαν εξαχθεί για τη γραμμική παλινδρόμηση χωρίς καμιά υπόθεση για την κατανομή που ακολουθούν οι μεταβλητές, διαπιστώνουμε ότι υπάρχει ταύτιση.

3. Η δεσμευμένη διασπορά της Y για δεδομένη τιμή της $X = x$ είναι

$$\sigma_{Y|X}^2 = \text{Var}[Y|X = x] = \sigma_Y^2 (1 - \rho^2) \quad (7.78)$$

Παρατηρούμε ότι η διασπορά αυτή είναι ανεξάρτητη της τιμής x , δηλαδή είναι σταθερή για όλα τα x . Το μέγεθος αυτό δεν είναι άλλο από το μέσο τετραγωνικό σφάλμα της εκτίμησης, όπως επιβεβαιώνουμε συγκρίνοντας με την εξίσωση (7.35).

Σχ. 7.12 (Απέναντι σελίδα) Εποπτικό σκαρίφημα για τη γενική παλινδρόμηση. Στο (α) απεικονίζεται σε προοπτικό διάγραμμα η από κοινού συνάρτηση πυκνότητας πιθανότητας των μεταβλητών X και Y . Στο (β) απεικονίζεται επίσης η ίδια συνάρτηση υπό μορφή καμπυλών ίσης πυκνότητας (ισοδιάσταση 0.0002), καθώς και η γενική καμπύλη παλινδρόμησης, η οποία ορίζεται από την εξίσωση (7.74) (παχιά συνεχής γραμμή). Στο (γ) απεικονίζεται και πάλι η γενική καμπύλη παλινδρόμησης (παρατηρούμε ότι περνά από το σημείο (μ_X, μ_Y)) και η δεσμευμένη συνάρτηση πυκνότητας πιθανότητας $f_{Y|X}(y, x)$ για τέσσερις διαφορετικές τιμές του x (κλίμακα αυθαίρετη, μέγιστη τεταγμένη 0.075 και στις τέσσερις θέσεις). Οι παράμετροι των περιθώριων κατανομών των X και Y για το παράδειγμα που απεικονίζεται είναι: $\mu_X = 100$, $\sigma_X = 30$, $\mu_Y = 50$, $\sigma_Y = 7$.



Ομοσκεδαστικότητα

Η ιδιότητα του παραπάνω σημείου 3, σύμφωνα με την οποία το τετραγωνικό σφάλμα είναι σταθερό για όλα τις τιμές του x λέγεται *ομοσκεδαστικότητα*. Στη γραμμική παλινδρόμηση μεταξύ μεταβλητών που ακολουθούν κανονική κατανομή η ομοσκεδαστικότητα είναι δεδομένη. Σε περίπτωση που οι μεταβλητές ακολουθούν άλλες κατανομές, η ομοσκεδαστικότητα δεν είναι εξασφαλισμένη, αλλά ωστόσο αποτελεί μια επιθυμητή ιδιότητα. Πρακτικά η ομοσκεδαστικότητα μπορεί να ελεγχθεί με τη γραφική απεικόνιση του σφάλματος εκτίμησης w συναρτήσει των τιμών της μεταβλητής x . Αν τα σημεία (x_i, w_i) κατανέμονται με τυχαίο τρόπο σχηματίζοντας ένα νέφος γύρω από την οριζόντια $w = 0$, τότε θεωρούμε ότι υπάρχει ομοσκεδαστικότητα. Αν υπάρχει συστηματικότητα στη διάταξη των σημείων (π.χ. τα σημεία σχηματίζουν κάποια καμπύλη, ή τα σημεία αποκλίνουν λιγότερο από την οριζόντια $w = 0$ για μικρές τιμές του x και περισσότερο για μεγάλες τιμές του x), τότε μιλούμε για *ετεροσκεδαστικότητα*. Η ετεροσκεδαστικότητα δείχνει ακαταλληλότητα του μοντέλου παλινδρόμησης που έχει επιλεγεί. Για το λόγο αυτό, αν διαπιστωθεί ετεροσκεδαστικότητα, τότε επαναλαμβάνεται η παλινδρόμηση με άλλο μοντέλο, π.χ. προσθέτοντας ένα μη γραμμικό όρο στην εξίσωση παλινδρόμησης ή κάνοντας κατάλληλους μετασχηματισμούς των μεταβλητών, σαν αυτούς που συζητήθηκαν στο εδάφιο 7.3.3.

Εφαρμογή 7.4.2

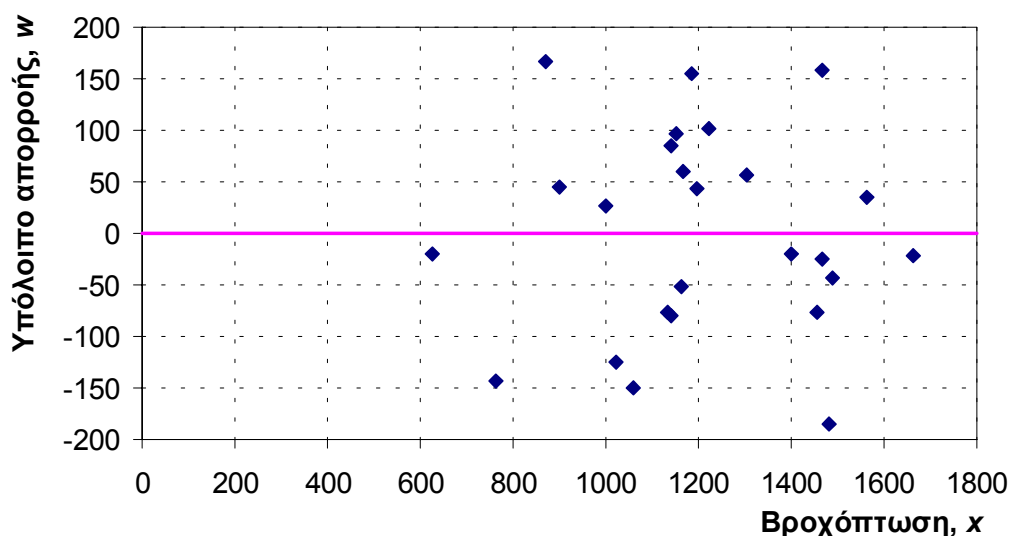
Να εξεταστεί η ομοσκεδαστικότητα των υπολοίπων στις Εφαρμογές 7.1 και 7.3.3.

Υπενθυμίζεται ότι στην Εφαρμογή 7.1 ασχοληθήκαμε (μεταξύ άλλων) με τη διατύπωση μιας γραμμικής σχέσης ανάμεσα στη βροχόπτωση και την απορροή της λεκάνης ανάντη Κρεμαστών του Αχελώου, ενώ στην Εφαρμογή 7.3.3 εξερευνήσαμε τη σχέση ανάμεσα στη στάθμη και την παροχή στη θέση Αχλαδόκαστρο του Ευήνου.

Ο έλεγχος της ομοσκεδαστικότητας των υπολοίπων παλινδρόμησης

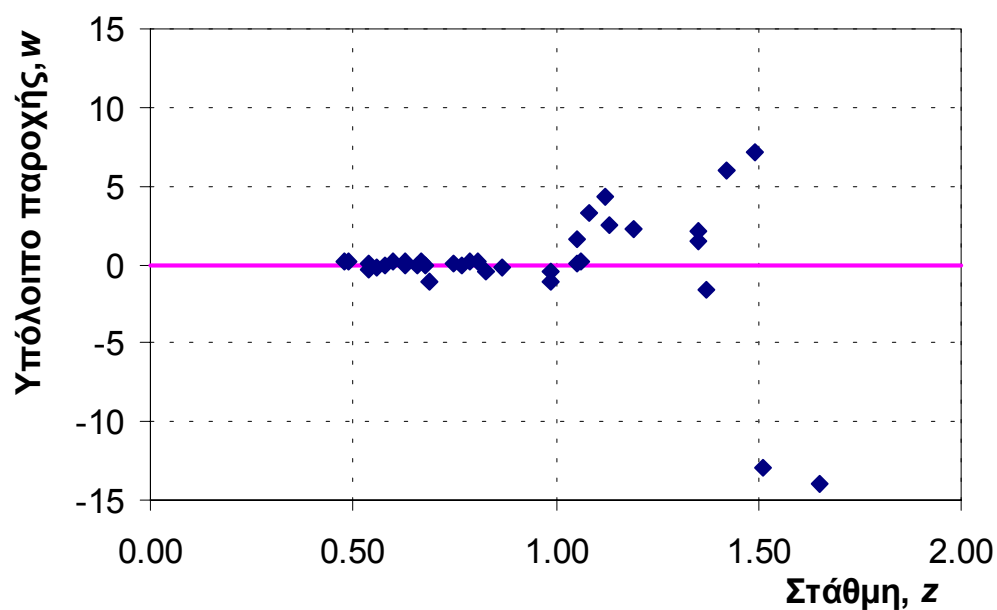
Σχ. 7.13 (Απέναντι σελίδα) Επεξηγηματικό σκαρίφημα για τη γραμμική παλινδρόμηση ως συνέπεια της διδιάστατης κανονικής κατανομής. Το σχήμα είναι παρόμοιο με το Σχ. 7.12, με τη μόνη διαφορά ότι εδώ η από κοινού συνάρτηση κατανομής των X και Y είναι κανονική. Οι παράμετροι της κατανομής των X και Y για το παράδειγμα που απεικονίζεται είναι: $\mu_X = 100$, $\sigma_X = 30$, $\mu_Y = 50$, $\sigma_Y = 7$ (όπως και στο Σχ. 7.12) και $\rho_{XY} = 0.65$.

γίνεται με γραφικό τρόπο. Στο Σχ. 7.14 φαίνονται τα υπόλοιπα παλινδρόμησης της Εφαρμογής 7.1. Τα υπόλοιπα αυτά υπολογίζονται από την εξίσωση $w_i = y_i - (a + b x_i)$ και απεικονίζονται συναρτήσει της ανεξάρτητης μεταβλητής (στην προκειμένη περίπτωση της βροχόπτωσης) x . Η τυχαία διάταξη των σημείων γύρω από τη γραμμή $w = 0$ δείχνει ότι υπάρχει ομοσκεδαστικότητα των υπολοίπων.

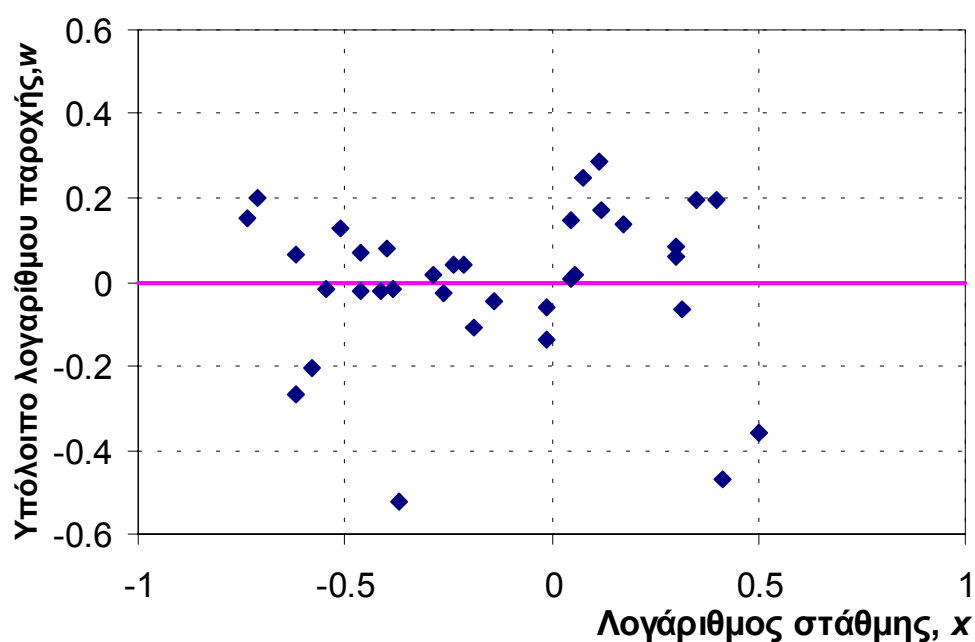


Σχ. 7.14 Διάγραμμα για τον έλεγχο της ομοσκεδαστικότητας των υπολοίπων παλινδρόμησης στην εκτίμηση της απορροής συναρτήσει της βροχόπτωσης στα Κρεμαστά (βλ. Εφαρμογή 7.1).

Στο Σχ. 7.15 έχει γίνει ένα παρόμοιο διάγραμμα για την Εφαρμογή 7.3.3. Εδώ η παροχή δίνεται συναρτήσει της στάθμης από την εξίσωση δύναμης $q = \kappa z^\lambda$, οπότε τα υπόλοιπα προκύπτουν από την εξίσωση $w_i = q_i - \kappa z_i^\lambda$. Στο διάγραμμα είναι έκδηλη η ετεροσκεδαστικότητα των υπολοίπων: τα σημεία αποκλίνουν λιγότερο από την οριζόντια γραμμή $w = 0$ για μικρές τιμές της στάθμης και περισσότερο για μεγάλες τιμές. Ωστόσο, υπενθυμίζουμε ότι, στην εν λόγω εφαρμογή, η παλινδρόμηση δεν έγινε απ' ευθείας ανάμεσα στη στάθμη και την παροχή, αλλά ανάμεσα στους λογαρίθμους τους. Η εξίσωση που δίνει το υπόλοιπο παλινδρόμησης σε αυτή την περίπτωση είναι $w_i = y_i - (a + b x_i)$, όπου x και y είναι οι (νεπέριοι) λογάριθμοι της στάθμης και παροχής, αντίστοιχα. Τα τελευταία υπόλοιπα παλινδρόμησης απεικονίζονται στο Σχ. 7.16, όπου είναι εμφανής η ομοσκεδαστικότητα των υπολοίπων. Το γεγονός αυτό επικυρώνει την ορθότητα του λογαριθμικού μετασχηματισμού, ο οποίος προηγήθηκε της παλινδρόμησης.



Σχ. 7.15 Διάγραμμα για τον έλεγχο της ομοσκεδαστικότητας των υπολοίπων στην εκτίμηση της παροχής συναρτήσει της στάθμης στο Αχλαδόκαστρο του Ευήνου (βλ. Εφαρμογή 7.3.3).



Σχ. 7.16 Διάγραμμα για τον έλεγχο της ομοσκεδαστικότητας των υπολοίπων παλινδρόμησης στην εκτίμηση του λογαρίθμου της παροχής συναρτήσει του λογαρίθμου της στάθμης στο Αχλαδόκαστρο του Ευήνου (βλ. Εφαρμογή 7.3.3).

7.5 Όρια εμπιστοσύνης και δοκιμές σημαντικότητας στη γραμμική παλινδρόμηση

Όλα όσα εκτέθηκαν στις προηγούμενες ενότητες αναφέρονται ουσιαστικά στην εξαγωγή σημειακών εκτιμήσεων της μεταβλητής Y για δεδομένη τιμή της X . Επανερχόμενοι στην γραμμική παλινδρόμηση, στην ενότητα αυτή θα δώσουμε τα απαραίτητα στοιχεία για την εξαγωγή εκτιμήσεων διαστήματος (ορίων εμπιστοσύνης). Διευκρινίζεται ότι όσα αναφέρονται σε αυτή την ενότητα αφορούν στην τυπική γραμμική παλινδρόμηση δύο μεταβλητών και δεν εφαρμόζονται στις ειδικές μορφές (ομογενής ευθεία ή οργανική συσχέτιση) ούτε σε μη γραμμικές μορφές παλινδρόμησης.

Οι πηγές αβεβαιότητας στις εκτιμήσεις της παλινδρόμησης, οι οποίες παίρνονται υπόψη για την εξαγωγή των εκτιμήσεων διαστήματος είναι δύο. Η πρώτη είναι η εγγενής αβεβαιότητα, που εισάγεται από τον όρο σφάλματος W της εξίσωσης παλινδρόμησης (7.24). Η πηγή αυτή αναιρείται μόνο όταν η διασπορά του σφάλματος (εξίσωση (7.35)) είναι μηδενική ή, ισοδύναμα, ο συντελεστής προσδιορισμού είναι 1. Η δεύτερη πηγή είναι η στατιστική αβεβαιότητα, η οποία προκύπτει από το πεπερασμένο μέγεθος του δείγματος (x_i, y_i) . Η αβεβαιότητα αυτή εισάγεται στην εκτίμηση των παραμέτρων a και b της εξίσωσης παλινδρόμησης και μεταφέρεται και στην εκτίμηση της Y . Αίρεται μόνο αν είναι γνωστές οι παράμετροι $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ και σ_{XY} του πληθυσμού, οι οποίες υπεισέρχονται στην εκτίμηση των a και b (εξισώσεις (7.28) και (7.27)).

Η εξαγωγή των εκτιμήσεων διαστήματος απαιτεί μια διαφορετική θεώρηση του προβλήματος της παλινδρόμησης, η οποία είναι γνωστή ως στατιστική θεώρηση. Σύμφωνα με αυτή το μοντέλο παλινδρόμησης γράφεται με τη μορφή

$$Y_i = a + bx_i + W \quad (7.79)$$

η οποία διαφέρει από την μορφή (7.24) στο ότι οι τιμές x_i του δείγματος δεν θεωρούνται ως πραγματοποιήσεις μιας τυχαίας μεταβλητής X , αλλά ως γνωστές αριθμητικές τιμές. Στο πρόβλημα αυτό αναζητούνται οι εκτιμήτριες \hat{A} και \hat{B} των παραμέτρων του μοντέλου a και b . Οι εκτιμήτριες αυτές είναι συναρτήσεις των τυχαίων μεταβλητών Y_i .

Δεν θα περιγράψουμε αναλυτικά τη διαδικασία εύρεσης των παραπάνω εκτιμητριών, την οποία ο ενδιαφερόμενος αναγνώστης μπορεί να βρει σε βιβλία στατιστικής (π.χ. Papoulis, 1990, σσ. 402-407· Benjamin and Cornell, 1970, σσ. 428-439). Θα δώσουμε μόνο τα αποτελέσματα της ανάλυσης σε πινακοποιημένη μορφή (Πίν. 7.5) και συγκεκριμένα τις μέσες τιμές και διασπορές των εκτιμητριών \hat{A} και \hat{B} καθώς και των μεταβλητών \hat{Y}_x και Y_x , που ορίζονται από τις σχέσεις

$$\hat{Y}_x = \hat{A} + \hat{B}x \quad Y_x = \hat{Y}_x + W = \hat{A} + \hat{B}x + W \quad (7.80)$$

Η \hat{Y}_x εκφράζει τη μέση τιμή της εκτίμησης για δεδομένη τιμή x , ενώ η Y_x εκφράζει μια απλή εκτίμηση της παλινδρόμησης, πάλι για δεδομένη τιμή x .

Πίν. 7.5 Στατιστικά χαρακτηριστικά των εκτιμητριών της γραμμικής παλινδρόμησης.

| Μέγεθος προς εκτίμηση | Στατιστική συνάρτηση ή μεταβλητή | Μέση τιμή | Διασπορά |
|--|----------------------------------|---|--|
| Κλίση ευθείας, b | \hat{B} | \hat{b} , όπως το b στην εξίσωση (7.10) | $\sigma_{\hat{B}}^2 = \frac{\sigma_w^2}{n s_x^2}$ |
| Περίλημμα ευθείας, a | \hat{A} | \hat{a} , όπως το a στην εξίσωση (7.11) | $\sigma_{\hat{A}}^2 = \frac{\sigma_w^2}{n} \left(1 + \frac{\bar{x}^2}{s_x^2} \right)$ |
| Μέση εκτίμηση (= τεταγμένη της ευθείας) για δεδομένο x | \hat{Y}_x | $\hat{y}_x = \hat{a} + \hat{b}x$ | $\sigma_{\hat{Y}_x}^2 = \frac{\sigma_w^2}{n} \left[1 + \frac{(x - \bar{x})^2}{s_x^2} \right]$ |
| Απλή εκτίμηση για δεδομένο x | Y_x | $y_x = \hat{y}_x = \hat{a} + \hat{b}x$ | $\sigma_{Y_x}^2 = \frac{\sigma_w^2}{n} \left[n + 1 + \frac{(x - \bar{x})^2}{s_x^2} \right]$ |

Στις εξισώσεις του Πίν. 7.5 τα μεγέθη \bar{x} και s_x^2 ορίζονται ως εξής:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \quad (7.81)$$

δηλαδή ταυτίζονται αριθμητικά με τη δειγματική μέση τιμή και τη (μεροληπτική) δειγματική διασπορά των x_i .^{*} Εξ άλλου το μέγεθος σ_W^2 είναι η θεωρητική διασπορά του σφάλματος W . Δεδομένου ότι το μέγεθος αυτό δεν είναι γνωστό, χρησιμοποιείται στη θέση του η αμερόληπτη εκτίμησή του από το δείγμα, η οποία είναι (Papoulis, 1990, σ. 405):

$$\hat{\sigma}_W^2 = \frac{1}{n-2} \sum_{i=1}^n [y_i - (\hat{a} + \hat{b} x_i)]^2 = \frac{n}{n-2} s_Y^2 (1 - r_{XY}^2) \quad (7.82)$$

Στην τελευταία εξίσωση τα s_Y^2 και r_{XY} εννοούνται ως μεροληπτικές εκτιμήσεις.

Ο υπολογισμός των ορίων εμπιστοσύνης για οποιοδήποτε από τα μεγέθη του Πίν. 7.5 (συμβολικά Ξ) βασίζεται στη συμβολική σχέση

$$\left(\hat{\xi} - t_{(1+\gamma)/2} \sigma_{\Xi}, \hat{\xi} + t_{(1+\gamma)/2} \sigma_{\Xi} \right) \quad (7.83)$$

όπου τα $\hat{\xi}$ και σ_{Ξ} δίνονται, για κάθε μέγεθος, στην τρίτη και τέταρτη στήλη του Πίν. 7.5, αντίστοιχα, ενώ $t_{(1+\gamma)/2}$ είναι το $[(1+\gamma)/2]$ -ποσοστημόριο της κατανομής Student για $n-2$ βαθμούς ελευθερίας και γ είναι ο συντελεστής εμπιστοσύνης.

Οι παραπάνω τύποι μπορούν να χρησιμοποιηθούν και για την εκτέλεση στατιστικών ελέγχων, σχετικά με τις τιμές των παραμέτρων ή των εκτιμήσεων. Η εφαρμογή είναι άμεση και ακολουθεί τη μεθοδολογία της ενότητας 3.5. Ωστόσο, ο βασικότερος έλεγχος αφορά στη σημαντικότητα του συντελεστή γραμμικής συσχέτισης ρ_{XY} . Ο έλεγχος αυτός έχει περιγραφεί αναλυτικά στο εδάφιο 3.5.3.

^{*} Γι' αυτό και χρησιμοποιούνται τα σύμβολα \bar{x} και s_x^2 , παρόλο που, όπως αναφέρθηκε παραπάνω, τα x_i δεν εκλαμβάνονται στην εξεταζόμενη στατιστική προσέγγιση ως πραγματοποιήσεις μιας τυχαίας μεταβλητής X .

Εφαρμογή 7.5

Να υπολογιστούν τα όρια εμπιστοσύνης των συντελεστών a και b , και των εκτιμήσεων της απορροής με βάση τη βροχόπτωση για τα δεδομένα της Εφαρμογής 7.1, για συντελεστή εμπιστοσύνης 98%.

Υπενθυμίζεται ότι τα στατιστικά χαρακτηριστικά των δειγμάτων της βροχόπτωσης (X) και της απορροής (Y) της Εφαρμογής 7.1 ήταν:

$$n = 25, \bar{x} = 1201.0, \bar{y} = 918.9, s_X = 252.5, s_Y = 234.6, r_{XY} = 0.911,$$

$$s_{XY} = r_{XY} s_X s_Y = 0.911 \times 252.5 \times 234.6 = 53\,964$$

Κατά συνέπεια, η τυπική απόκλιση των υπολοίπων (βλ. εξίσωση (7.82)) είναι

$$\begin{aligned} \hat{\sigma}_w &= \sqrt{n/(n-2)} s_Y \sqrt{1-r_{XY}^2} \\ &= \sqrt{25/23} \times 234.6 \times \sqrt{1-0.911^2} = 100.9 \text{ mm} \end{aligned}$$

Οι εκτιμήσεις των παραμέτρων της γραμμικής παλινδρόμησης είναι

$$\hat{b} = 0.846, \hat{a} = -97.1$$

Το $[(1+\gamma)/2]$ ποσοστημόριο της κατανομής Student για συντελεστή εμπιστοσύνης $\gamma = 0.98$ και βαθμούς ελευθερίας $25 - 2 = 23$ είναι 2.50 (βλ. Πίν. Π3 Παραρτήματος πινάκων).

Για την κλίση b έχουμε

$$\hat{b} = 0.846, \sigma_{\hat{B}} = \sigma_w / (\sqrt{n} s_X) = 100.9 / (\sqrt{25} \times 252.5) = 0.0799$$

και επομένως τα όρια εμπιστοσύνης είναι

$$(0.846 - 2.50 \times 0.0799, 0.846 + 2.50 \times 0.0799) = (0.646, 1.045)$$

Για το περίλημμα a έχουμε

$$\begin{aligned} \hat{a} &= -97.1, \sigma_{\hat{A}} = (\sigma_w / \sqrt{n}) \sqrt{1 + \bar{x}^2 / s_X^2} = \\ &= (100.9 / \sqrt{25}) \times \sqrt{1 + 1201.0^2 / 252.5^2} = 98.1 \end{aligned}$$

και επομένως τα όρια εμπιστοσύνης είναι

$$(-97.1 - 2.50 \times 98.1, -97.1 + 2.50 \times 98.1) = (-342.4, 148.2)$$

Για τη μέση εκτίμηση (= τεταγμένη της ευθείας παλινδρόμησης) για δεδομένο x έχουμε

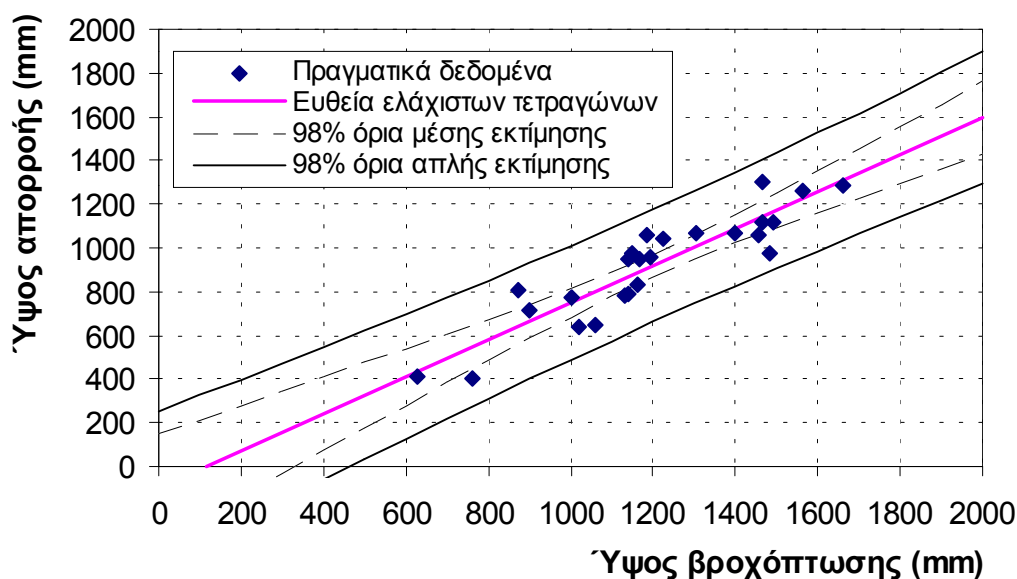
$$\hat{y}_x = \hat{a} + \hat{b}x = -97.1 + 0.846x$$

και

$$\begin{aligned}\sigma_{\hat{Y}_x} &= (\sigma_w/\sqrt{n}) \sqrt{1 + (x - \bar{x})^2 / s_x^2} \\ &= (100.9 / \sqrt{25}) \times \sqrt{1 + (x - 1201.0)^2 / 252.5^2} \\ &= 20.2 \times \sqrt{1 + 1.568 \times 10^{-5} \times (x - 1201.0)^2}\end{aligned}$$

και επομένως τα όρια εμπιστοσύνης είναι

$$\begin{aligned}-97.1 + 0.846 x \pm 2.50 \times 20.2 \times \sqrt{1 + 1.568 \times 10^{-5} \times (x - 1201.0)^2} \\ = -97.1 + 0.846 x \pm 50.5 \times \sqrt{1 + 1.568 \times 10^{-5} \times (x - 1201.0)^2}\end{aligned}$$



Σχ. 7.17 Όρια εμπιστοσύνης 98% της εκτίμησης του ετήσιου ύψους επιφανειακής απορροής συναρτήσει του ετήσιου ύψους βροχόπτωσης στη λεκάνη ανάντη Κρεμαστών.

Για την απλή εκτίμηση για δεδομένο x έχουμε

$$y_x = \hat{a} + \hat{b}x = -97.1 + 0.846 x$$

και

$$\begin{aligned}\sigma_{\hat{Y}_x} &= (\sigma_w/\sqrt{n}) \sqrt{n + 1 + (x - \bar{x})^2 / s_x^2} \\ &= (100.9 / \sqrt{25}) \times \sqrt{26 + (x - 1201.0)^2 / 252.5^2} \\ &= 20.2 \times \sqrt{26 + 1.568 \times 10^{-5} \times (x - 1201.0)^2}\end{aligned}$$

και επομένως τα όρια εμπιστοσύνης είναι

$$-97.1 + 0.846 x \pm 50.5 \times \sqrt{26 + 1.568 \times 10^{-5} \times (x - 1201.0)^2}$$

Τα παραπάνω όρια εμπιστοσύνης της μέσης και της απλής εκτίμησης, τα οποία είναι συναρτήσεις της ανεξάρτητης μεταβλητής x , έχουν παρασταθεί γραφικά στο Σχ. 7.17. Παρατηρούμε ότι, όσο απομακρυνόμαστε από τις μέσες τιμές, τόσο διευρύνονται τα διαστήματα εμπιστοσύνης των εκτιμήσεων, πράγμα που σημαίνει ότι αυξάνει η αβεβαιότητα των εκτιμήσεων.

7.6 Συμπλήρωση και επέκταση δειγμάτων

Η τυπική χρήση της παλινδρόμησης στην τεχνική υδρολογία αφορά στη *συμπλήρωση* των ελλείψεων ενός υδρολογικού δείγματος με βάση ένα πληρέστερο δείγμα, το οποίο συσχετίζεται με το πρώτο, ή και την *επέκταση* του πρώτου δείγματος σε μια μεγαλύτερη διάρκεια που καλύπτεται από το δεύτερο δείγμα. Το πρώτο δείγμα θα το λέμε *δείγμα μελέτης* και το δεύτερο *δείγμα αναφοράς* ή *βάσης*. Στην ενότητα αυτή αναφερόμαστε σε ένα μοναδικό δείγμα αναφοράς, και κατά συνέπεια σε απλή παλινδρόμηση, αν και η μεθοδολογία που εφαρμόζεται μπορεί γενικά να περιλαμβάνει και περισσότερα από ένα δείγματα αναφοράς με πολλαπλή παλινδρόμηση.

Τα δείγματα μελέτης και αναφοράς μπορεί να είναι περιέχουν πρωτογενείς σημειακές μετρήσεις μιας φυσικής διεργασίας. Για παράδειγμα μπορεί να αποτελούνται από τις μετρήσεις της σημειακής βροχόπτωσης σε δύο γειτονικούς σταθμούς. Μπορεί επίσης να περιέχουν επεξεργασμένα δεδομένα της ίδιας φυσικής διεργασίας, όπως για παράδειγμα τα επιφανειακά ύψη βροχής δύο κοντινών υδρολογικών λεκανών. Τέλος, μπορεί να περιέχουν δεδομένα από διαφορετικές φυσικές διεργασίες, για τις οποίες υπάρχουν βάσιμοι λόγοι να συσχετίζονται. Για παράδειγμα μπορεί το δείγμα μελέτης να αποτελείται από την ετήσια απορροή μιας λεκάνης με σημαντική υδροφορία και το δείγμα αναφοράς να περιέχει την επιφανειακή βροχόπτωση της ίδιας λεκάνης, ή τη σημειακή βροχόπτωση ενός αντιπροσωπευτικού σταθμού της λεκάνης.

Η συμπλήρωση και η επέκταση είναι έννοιες πολύ κοντινές και δεν διακρίνονται με αυστηρό τρόπο μεταξύ τους. Συνήθως χρησιμοποιούμε τον πρώτο όρο σε περιπτώσεις που το δείγμα μελέτης παρουσιάζει λίγα και σποραδικά κενά (π.χ. 1-3), τα οποία πρέπει να συμπληρωθούν, και το δεύτερο για περιπτώσεις που το δείγμα μελέτης εμφανίζει συστηματικές,

συνήθως πολυετείς, ελλείψεις κατά τη διάρκεια της λειτουργίας του, ή η λειτουργία του εκτείνεται σε σημαντικά μικρότερο διάστημα σε σχέση με αυτό του σταθμού αναφοράς. Αυτό που επιδιώκουμε με την επέκταση και συμπλήρωση είναι η απόκτηση ενός δείγματος μελέτης με μεγαλύτερο μήκος από το αρχικό και κατά συνέπεια μεγαλύτερη αξιοπιστία εκτιμήσεων. Θα πρέπει όμως από τώρα να ξεκαθαρίσουμε ότι η αξιοπιστία του διευρυμένου δείγματος δεν αντιστοιχεί στο ονομαστικό διευρυμένο μήκος του, αλλά σε ένα ενδιάμεσο μήκος ανάμεσα στο αρχικό και το διευρυμένο (μόνο οι άμεσες μετρήσεις δίνουν αξιοπιστία αντίστοιχη με το μήκος). Για παράδειγμα αν έχουμε ένα παρατηρημένο δείγμα μελέτης με 20 ετήσιες τιμές και το επεκτείνουμε, με βάση ένα δείγμα αναφοράς, για άλλα 10 χρόνια, τότε η αξιοπιστία του νέου δείγματος δεν είναι ίδια με αυτή που θα είχε ένα παρατηρημένο (με μετρήσεις) δείγμα 30 ετών, αλλά θα αντιστοιχεί σε ένα παρατηρημένο δείγμα ισοδύναμου μήκους n' όπου $20 \leq n' \leq 30$. Είναι προφανές ότι, όσο πιο έντονη είναι η συσχέτιση του δείγματος μελέτης με το δείγμα αναφοράς, τόσο πιο πολύ κοντά στο 30 θα είναι το ισοδύναμο μήκος.

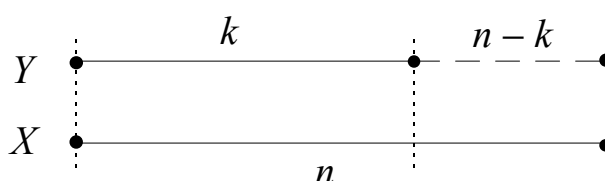
Για τον ίδιο λόγο, οι εκτιμήσεις των στατιστικών χαρακτηριστικών ενός διευρυμένου δείγματος δεν μπορεί να γίνονται με τους απλούς τύπους που χρησιμοποιούμε για ένα παρατηρημένο δείγμα. Το θέμα απαιτεί μεγάλη προσοχή, ιδίως στις περιπτώσεις σημαντικού μήκους επέκτασης δείγματος, οπότε και τα σφάλματα, αν η εκτίμηση γίνεται με τους τύπους του παρατηρημένου δείγματος, είναι μεγάλα.

Η ενότητα αυτή καλύπτει κατά βάση την περίπτωση που η συμπλήρωση η επέκταση των τιμών της Y γίνεται με μια γραμμική εξίσωση ως προς X , σε μία από τις διάφορες μορφές που έχουν εξεταστεί προηγουμένως. Μπορούμε να θεωρήσουμε ότι έμμεσα καλύπτει και τις περιπτώσεις γραμμικοποιημένων εξισώσεων μέσω μετασχηματισμών των μεταβλητών (εδάφιο 7.3.3), δεδομένου ότι αυτές ανάγονται σε γραμμικές εξισώσεις. Η γραμμική μορφή εξίσωσης είναι η πιο διαδεδομένη, όχι μόνο λόγω της απλότητάς της. Όπως είδαμε στο εδάφιο 7.4.2, η γραμμική παλινδρόμηση είναι η βέλτιστη παλινδρόμηση για μεταβλητές που ακολουθούν κανονική κατανομή. Μάλιστα, όπως είδαμε και σε άλλα κεφάλαια, στην τεχνική υδρολογία η κανονική κατανομή είναι πολύ συχνά κατάλληλη για μεταβλητές που αναφέρονται σε μεγάλες χρονικές κλίμακες, ώστε σε κάθε χρονικό διάστημα να αντιστοιχεί μεγάλος αριθμός

υδρολογικών επεισοδίων. Η καταλληλότητα της κανονικής κατανομής σε αυτές τις περιπτώσεις εξηγείται από το κεντρικό οριακό θεώρημα.

Ωστόσο, η χρήση της γραμμικής παλινδρόμησης δεν περιορίζεται στην κανονική κατανομή. Δεν υπάρχει λόγος που να αποκλείει τη γραμμική παλινδρόμηση από μη κανονικές μεταβλητές. Αντίθετα, και σε μη κανονικές μεταβλητές, αν δεν υπάρχει θεωρητικός λόγος που να μας οδηγεί σε άλλο τύπο παλινδρόμησης, ξεκινάμε την ανάλυσή μας δοκιμάζοντας την απλούστερη γραμμική παλινδρόμηση. Προηγουμένως, ελέγχουμε το συντελεστή γραμμικής συσχέτισης, ο οποίος θα πρέπει να είναι σημαντικά (με τη στατιστική έννοια) διάφορος του μηδενός. Εκ των υστέρων ελέγχουμε την ομοσκεδαστικότητα των υπολοίπων (βλ. αντίστοιχο εδάφιο), η οποία, αν υπάρχει, αποτελεί ένδειξη για την επιτυχία της επιλογής του συγκεκριμένου τύπου παλινδρόμησης.

Στα επόμενα εδάφια θα αναφερθούμε διεξοδικά στις διάφορες μεθόδους συμπλήρωσης, στα αντίστοιχα προβλήματα και στον τρόπο αντιμετώπισής τους, ενώ στην Εφαρμογή 7.6 δίνουμε ένα πλήρες παράδειγμα για τη χρήση των διάφορων μεθόδων. Σε όλες τις περιπτώσεις χρησιμοποιούμε το συμβολισμό που διευκρινίζεται στο Σχ. 7.18. Θεωρούμε, δηλαδή, ότι το δείγμα της μεταβλητής Y διαθέτει k παρατηρημένα δεδομένα, ενώ το δείγμα της μεταβλητής X διαθέτει $n > k$ παρατηρημένα δεδομένα. Επίσης θεωρούμε ότι για τις k μετρήσεις της μεταβλητής Y υπάρχουν οι ταυτόχρονες μετρήσεις της X . Το πρόβλημα που μελετάμε είναι η εκτίμηση των $(n - k)$ τιμών της Y με βάση τις αντίστοιχες της X .



Σχ. 7.18 Διευκρινιστικό σκαρίφημα για το συμβολισμό στην επέκταση δείγματος.

Επίσης, τα στατιστικά χαρακτηριστικά που αναφέρονται στην κοινή περίοδο, μεγέθους k , συμβολίζονται με άτονα γράμματα, π.χ. τα \bar{x} και \bar{y} συμβολίζουν τις δειγματικές μέσες τιμές των X και Y , εκτιμημένες από τα δείγματα μεγέθους k , και s^{*X} και s^{*Y} τις αντίστοιχες αμερόληπτες δειγματικές διασπορές. Τα αντίστοιχα μεγέθη για την ολική περίοδο μεγέθους n

συμβολίζονται με τονούμενα γράμματα, π.χ. με \bar{x}' και \bar{y}' συμβολίζονται οι εκτιμήσεις των μέσων τιμών από τα δείγματα μεγέθους n , με s'^2_{X} και s'^2_{Y} οι αντίστοιχες εκτιμήσεις διασπορών κτλ.

7.6.1 Χρήση της γραμμικής παλινδρόμησης χωρίς όρο σφάλματος

Η συμπλήρωση ή επέκταση ενός δείγματος Y , με βάση το αντίστοιχο πληρέστερο δείγμα της μεταβλητής X , γίνεται τις περισσότερες φορές με βάση την εξίσωση (7.25) ($\hat{Y} = a + bX$), στην οποία δεν παίρνεται υπόψη ο όρος σφάλματος W . Η μέθοδος αυτή οδηγεί, όπως είδαμε, στην ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος και γι' αυτό είναι η πιο διαδεδομένη. Ωστόσο έχει σημαντικά μειονεκτήματα. Το διευρυμένο δείγμα που παράγεται με εφαρμογή της μεθόδου αυτής δίνει αμερόληπτη εκτίμηση της μέσης τιμής αλλά μεροληπτική εκτίμηση της διασποράς. Συγκεκριμένα, η χρήση του τύπου της δειγματικής διασποράς παρατηρημένου δείγματος (εξίσωση 3.15) υπεκτιμά τη διασπορά της Y . Η διασπορά αυτή θα πρέπει να διορθωθεί με πολλαπλασιασμό επί το συντελεστή

$$\varphi = \frac{n-1}{k + (n-k)\rho^2_{XY} - 1} \quad (7.84)$$

ο οποίος είναι μεγαλύτερος ή ίσος του 1. Συγκεκριμένα, ο συντελεστής διόρθωσης παίρνει την τιμή 1 μόνο όταν $|\rho_{XY}| = 1$, ενώ για $\rho_{XY} = 0$ παίρνει τη μέγιστη τιμή του $(n-1)/(k-1)$. Η παραγωγή της εξίσωσης (7.84) δίνεται αναλυτικά στο Παράρτημα 7Α. Παρατηρούμε ότι στην εξίσωση υπεισέρχεται η θεωρητική τιμή του συντελεστή συσχέτισης ρ_{XY} , η οποία όμως στην πραγματικότητα είναι άγνωστη. Στην εφαρμογή μπορεί να χρησιμοποιηθεί στη θέση της η δειγματική τιμή r_{XY} που υπολογίζεται από τις k ταυτόχρονες μετρήσεις των X και Y . Δεν είναι σωστό να χρησιμοποιηθεί η τιμή του συντελεστή συσχέτισης που προκύπτει από τα n δεδομένα, μετά τη συμπλήρωση των $n-k$ τιμών της X , γιατί και σε αυτή την τιμή υπάρχει σοβαρή μεροληψία. Συγκεκριμένα η τιμή αυτή είναι σημαντικά υπερεκτιμημένη.

Η μέση τιμή και η διασπορά του διευρυμένου δείγματος μπορούν να εκτιμηθούν και από τις ακόλουθες εξισώσεις, οι οποίες δεν κάνουν χρήση των εκτιμημένων $n-k$ δεδομένων της Y :

$$\bar{y}' = \bar{y} + b(\bar{x}' - \bar{x}) \quad (7.85)$$

$$s'^2_{*Y} = s^2_{*Y} + b^2 \left(s'^2_{*X} - s^2_{*X} \right) \quad (7.86)$$

Εύκολα διαπιστώνουμε ότι οι παραπάνω εκτιμήσεις είναι αμερόληπτες εφόσον είναι γνωστή η θεωρητική τιμή της παραμέτρου $b = \rho_{XY} \sigma_Y / \sigma_X$. Στην πραγματικότητα, βέβαια, αυτή δεν είναι γνωστή, οπότε χρησιμοποιείται στη θέση της η αντίστοιχη εκτίμηση από τα δείγματα μεγέθους k , $\hat{b} = r_{XY} s_{*Y} / s_{*X}$. Αυτό εισάγει κάποια μεροληψία στις εκτιμήσεις, η οποία πάντως δεν είναι σημαντική.

Αντίστοιχα, μια κατά προσέγγιση αμερόληπτη εκτίμηση για το συντελεστή συσχέτισης, η οποία αναφέρεται στην περίοδο μήκους n , είναι

$$r'_{XY} = r_{XY} \frac{s_{*Y} \ s'_{*X}}{s_{*X} \ s'_{*Y}} \quad (7.87)$$

όπου η τυπική απόκλιση s'_{*Y} θα πρέπει να υπολογιστεί με βάση την (7.86). Η εξαγωγή των εξισώσεων (7.85) έως (7.87) δίνεται αναλυτικά στο Παράρτημα 7Α.

Προκειμένου να αποκτήσουμε μια εικόνα του βαθμού βελτίωσης της εκτίμησης της μέσης τιμής στο διευρυμένο δείγμα σε σχέση με την αρχική, μας χρειάζεται η διασπορά της εκτιμήτριας της μέσης τιμής του διευρυμένου δείγματος. Στο Παράρτημα 7Α αποδεικνύεται ότι αυτή δίνεται από τον τύπο

$$\text{Var}[\bar{Y}'] = \frac{\sigma_Y^2}{k} \left(1 - \frac{n-k}{n} \rho_{XY}^2 \right) \quad (7.88)$$

Παίρνοντας υπόψη ότι η διασπορά της αρχικής εκτίμησης είναι

$$\text{Var}[\bar{Y}] = \frac{\sigma_Y^2}{k} \quad (7.89)$$

η βελτίωση στην εκτίμηση δίνεται από το λόγο

$$\frac{\text{Var}[\bar{Y}']}{\text{Var}[\bar{Y}]} = 1 - \frac{n-k}{n} \rho_{XY}^2 \quad (7.90)$$

ο οποίος παίρνει τη βέλτιστη (ελάχιστη) τιμή του k/n για $|\rho_{XY}| = 1$, ενώ είναι ίσος με 1 (που σημαίνει μηδενική βελτίωση) για $\rho_{XY} = 0$.

Στην εξίσωση (7.88), όπως και σε προηγούμενες εξισώσεις, υποτίθεται ότι είναι γνωστή η θεωρητική τιμή του συντελεστή συσχέτισης ρ_{XY} . Στην πραγματικότητα είναι γνωστή η εκτίμηση r_{XY} . Για αυτή την περίπτωση ο τύπος είναι ελαφρά διαφοροποιημένος (βλ. Yevjevich, 1972, σ. 260· Salas, 1993, σ. 19.44):

$$\text{Var}[\bar{Y}'] = \frac{\sigma_Y^2}{k} \left[1 - \frac{n-k}{n} \left(r_{XY}^2 - \frac{1-r_{XY}^2}{k-3} \right) \right] \quad (7.91)$$

οπότε η βελτίωση στην εκτίμηση δίνεται από το λόγο

$$\frac{\text{Var}[\bar{Y}']}{\text{Var}[\bar{Y}]} = 1 - \frac{n-k}{n} \left(r_{XY}^2 - \frac{1-r_{XY}^2}{k-3} \right) \quad (7.92)$$

Γενικά θεωρείται ότι το ισοδύναμο μήκος του διευρυμένου δείγματος n' δίνεται από τη σχέση

$$\frac{k}{n'} = \frac{\text{Var}[\bar{Y}']}{\text{Var}[\bar{Y}]} \quad (7.93)$$

Από την εξίσωση (7.92) προκύπτει ότι, για να είναι ο λόγος $\text{Var}[\bar{Y}'] / \text{Var}[\bar{Y}]$ μικρότερος από 1 και, κατά συνέπεια, για να υπάρχει βελτίωση στην εκτίμηση της μέσης τιμής, θα πρέπει

$$|r_{XY}| > \frac{1}{\sqrt{k-2}} \quad (7.94)$$

Η τελευταία ανισότητα δίνει μια εκτίμηση του κάτω ορίου του συντελεστή συσχέτισης για να έχει νόημα η επέκταση του δείγματος με γραμμική παλινδρόμηση.

Κατά μία άλλη λογική, την οποία συστήνουμε, για να έχει νόημα η παλινδρόμηση, θα πρέπει ο συντελεστής συσχέτισης r_{XY} να είναι σημαντικά διάφορος του μηδενός. Θεωρώντας επίπεδο σημαντικότητας $\alpha = 5\%$, η λογική αυτή οδηγεί στην ισχυρότερη ανισότητα

$$|r_{XY}| \gtrsim \frac{2}{\sqrt{k}} \quad (7.95)$$

Η τελευταία ανισότητα έχει παραχθεί στο εδάφιο 3.5.3 (βλ. εξίσωση 3.74).

7.6.2 Χρήση της οργανικής συσχέτισης

Η χρήση της οργανικής συσχέτισης είναι ίδια όπως της γραμμικής παλινδρόμησης χωρίς όρο σφάλματος. Η διαφορά εδώ είναι ότι η παράμετρος b εκτιμάται από την (7.54) αντί της (7.27).

Όπως έχουμε αναφέρει προηγουμένως, η μέθοδος της οργανικής συσχέτισης δεν οδηγεί σε μεροληψία στις εκτιμήσεις των μέσων τιμών και διασπορών της Y και κατά συνέπεια αυτές μπορούν να υπολογίζονται κανονικά από το διευρυμένο δείγμα, χωρίς να απαιτούνται διορθώσεις. Ωστόσο, η μέθοδος εισάγει σοβαρή μεροληψία στην εκτίμηση του συντελεστή συσχέτισης και γι' αυτό ο τελευταίος δεν πρέπει να υπολογίζεται άμεσα από το διευρυμένο δείγμα. Μπορεί να χρησιμοποιηθεί για τον προσεγγιστικό υπολογισμό του η εξίσωση (7.87).

Κατ' αναλογία με όσα αναφέρθηκαν στο προηγούμενο εδάφιο (εξίσωση (7.95)) και παίρνοντας υπόψη και όσα αναφέρθηκαν στο εδάφιο 7.2.2 (εξίσωση (7.61)) οδηγούμαστε στο συμπέρασμα ότι για να έχει νόημα η εφαρμογή της οργανικής συσχέτισης θα πρέπει

$$|r_{XY}| \gtrsim \max \left\{ 0.5, \frac{2}{\sqrt{k}} \right\} \quad (7.96)$$

7.6.3 Χρήση της ομογενούς ευθείας

Η χρήση της ομογενούς ευθείας για επέκταση δείγματος θα πρέπει γενικά να αποφεύγεται, διότι, όπως ήδη αναφέρθηκε, οδηγεί σε μεροληψία και ως προς τη μέση τιμή και ως προς τη διασπορά.

Ωστόσο, σε περιπτώσεις συμπλήρωσης λίγων τιμών, οι οποίες δεν επηρεάζουν τα στατιστικά χαρακτηριστικά ενός δείγματος, είναι πιθανό η ομογενής ευθεία να είναι πλεονεκτική. Ας θεωρήσουμε σαν παράδειγμα το πρόβλημα της συμπλήρωσης λίγων ελλείψεων ενός πολυετούς δείγματος ημερήσιων βροχοπτώσεων. Αν χρησιμοποιηθεί η τυπική γραμμική παλινδρόμηση ή η οργανική συσχέτιση, τότε είναι δυνατό να προκύψουν μερικές αρνητικές τιμές, επειδή ο σταθερός όρος a της γραμμικής εξίσωσης μπορεί να είναι αρνητικός. Με την ομογενή ευθεία αποφεύγουμε αυτό το πρόβλημα. Πάντως και σε αυτή την περίπτωση είναι εξεταστέα και η εφαρμογή της εξίσωσης δύναμης ($Y = \kappa X^{\lambda}$, βλ. εδάφιο 7.3.3), η οποία αποφεύγει επίσης τις αρνητικές τιμές, διατηρώντας παράλληλα τη μέση τιμή του λογαρίθμου της μεταβλητής.

7.6.4 Χρήση της γραμμικής παλινδρόμησης με όρο σφάλματος

Όπως είδαμε στα προηγούμενα εδάφια, τόσο η τυπική γραμμική παλινδρόμηση, όσο και η οργανική συσχέτιση, παρουσιάζουν διάφορα προβλήματα, τα οποία οδηγούν σε μεροληπτικές εκτιμήσεις διάφορων στατιστικών χαρακτηριστικών του διευρυμένου δείγματος. Η “πηγή” όλων αυτών των προβλημάτων είναι το γεγονός ότι τα σημεία (x_i, y_i) του διευρυμένου δείγματος βρίσκονται όλα πάνω στην ευθεία παλινδρόμησης (ή οργανικής συσχέτισης, βλ. Σχ. 7.19), γεγονός που αντίκειται στη φυσική πραγματικότητα. Τα σημεία (x_i, y_i) που προέρχονται από μετρήσεις ποτέ δεν είναι συνευθειακά (εκτός της ακραίας περίπτωσης όπου $|\rho_{XY}| = 1$). Φυσικά αυτό το πρόβλημα δεν οφείλεται στη γραμμικότητα της εξίσωσης. Και σε μη γραμμική εξίσωση παλινδρόμησης το πρόβλημα θα υπάρχει, μόνο που αντί να είναι συνευθειακά τα σημεία της επέκτασης θα ανήκουν σε μια λεία καμπύλη. Το πρόβλημα προκύπτει από την αγνόηση του σφάλματος γύρω από το γραμμικό (ή μη) νόμο, που στη φύση πάντα υπάρχει.

Οι διάφορες μέθοδοι που εκτέθηκαν παραπάνω για την άρση των σφαλμάτων στον υπολογισμό των στατιστικών χαρακτηριστικών, είναι ικανοποιητικές μεν σε ότι αφορά αυτά τα ίδια τα στατιστικά χαρακτηριστικά, αλλά δεν λύνουν κάθε συνέπεια της μη ρεαλιστικής διάταξης των σημείων. Πολύ συχνά ένα δείγμα μιας μεταβλητής επεκτείνεται προκειμένου να χρησιμοποιηθεί ως (πιο επαρκής) είσοδος σε ένα άλλο μοντέλο, το οποίο πραγματοποιεί κάποιο μετασχηματισμό της εισόδου αυτής (π.χ.

μοντέλο υδατικού ισοζυγίου, μοντέλο βροχής-απορροής, μοντέλο λειτουργίας ταμιευτήρα κτλ.). Στην περίπτωση αυτή δεν ενδιαφέρουν μόνο τα συγκεκριμένα στατιστικά χαρακτηριστικά της μεταβλητής, αλλά όλη η στατιστική δομή του δείγματος, η οποία ασφαλώς έχει διαταραχθεί με τη διεύρυνση.

Η μόνη ρεαλιστική λύση για αυτές τις περιπτώσεις, που ενδιαφέρει η όλη δομή του δείγματος, είναι η χρήση της πλήρους μορφής της γραμμικής εξίσωσης (7.24) που συμπεριλαμβάνει και τον όρο σφάλματος W ($Y = a + bX + W$). Στην εφαρμογή αυτής της εξίσωσης ο όρος σφάλματος δεν εκτιμάται (η εκτίμησή του θα ήταν ίση π.χ. με τη μέση τιμή του, δηλαδή μηδέν) αλλά γεννάται ή προσομοιώνεται. Χρησιμοποιείται, δηλαδή, μια γεννήτρια τυχαίων αριθμών, η οποία γεννά τις απαραίτητες $n - k$ τυχαίες τιμές του σφάλματος.

Μια ακολουθία αριθμών x_i λέγεται ακολουθία τυχαίων αριθμών δεδομένης κατανομής $F(x)$ αν αποτελεί δείγμα της τυχαίας μεταβλητής X , η οποία έχει συνάρτηση κατανομής $F(x)$. Για κάθε συνάρτηση κατανομής μπορεί να κατασκευαστεί μία γεννήτρια τυχαίων αριθμών (ή και περισσότερες). Η γεννήτρια είναι ένας αλγόριθμος, συνήθως αναδρομικός, ο οποίος μπορεί να παράγει διαδοχικά οσουςδήποτε όρους της τυχαίας ακολουθίας. Στην πράξη βέβαια υπάρχει ένα αρκετά μεγάλο, αλλά πάντως πεπερασμένο, όριο τυχαίων αριθμών που μπορεί να δώσει η ακολουθία. Πάνω από αυτό το όριο γίνεται επανάληψη των ίδιων αριθμών, δηλαδή η ακολουθία γίνεται περιοδική. Πάντως σε κάθε περίπτωση οι τυχαίοι αριθμοί δεν γεννώνται στην τύχη ο ένας μετά τον άλλον, αλλά βάσει ενός αυστηρά προσδιοριστικού αλγορίθμου, ο οποίος οδηγεί στην ίδια ακολουθία αριθμών, αν ξεκινήσει με τις ίδιες αρχικές συνθήκες. Για το λόγο αυτό τους τυχαίους αριθμούς μερικοί τους ονομάζουν ψευδοτυχαίους. Πάντως, αν αλλάξουμε τις αρχικές συνθήκες παίρνουμε άλλη τυχαία ακολουθία (ακριβέστερα άλλο τμήμα της ίδιας περιοδικής ακολουθίας). Στο επόμενο ένθετο εδάφιο δίνονται γεννήτριες τυχαίων αριθμών με ομοιόμορφη και με κανονική κατανομή. Η γεννήτρια της κανονικής κατανομής μπορεί να καλύψει τη συχνότερη περίπτωση επέκτασης δειγμάτων που ακολουθούν κανονική κατανομή (αν οι μεταβλητές X και Y ακολουθούν κανονική κατανομή τότε και η W ακολουθεί κανονική κατανομή). Ο αναγνώστης που ενδιαφέρεται για γεννήτριες άλλων

συναρτήσεων κατανομής παραπέμπεται π.χ. στον Papoulis, 1990, σσ. 251-372.

Ανακεφαλαιωτικά, στην περίπτωση που στη διεύρυνση ενός δείγματος λαμβάνουμε υπόψη το τυχαίο σφάλμα, δεν κάνουμε εκτίμηση των (βέλτιστων) τιμών της Y αλλά προσομοιώσή τους. Αυτή η διαδικασία έχει ένα πολύ σοβαρό μειονέκτημα: δεν δίνει μοναδικές τιμές y_i για τις μεταβλητές που προσομοιώνει. Με εφαρμογή άλλης γεννήτριας τυχαίων αριθμών ή της ίδιας γεννήτριας με διαφορετικές αρχικές συνθήκες θα προκύψουν άλλες τιμές y_i . Ωστόσο, έχει το σοβαρό πλεονέκτημα να μη διαταράσσει τη στατιστική δομή του δείγματος. Έτσι, οι εκτιμήσεις της μέσης τιμής, της διασποράς, του συντελεστή συσχέτισης κτλ., μπορούν να γίνουν από το διευρυμένο δείγμα με εφαρμογή των κοινών τύπων του παρατηρημένου δείγματος, χωρίς καμιά διόρθωση. Βέβαια, ανακύπτει και στην εκτίμηση αυτών των στατιστικών χαρακτηριστικών το πρόβλημα ότι αυτά δεν ορίζονται με μονοσήμαντο τρόπο.

Συμπερασματικά, δεν υπάρχει καμιά τέλεια μέθοδος για τη συμπλήρωση και επέκταση ενός δείγματος. Οι τρεις μέθοδοι που εξετάστηκαν αναλυτικά παραπάνω, δηλαδή η εκτίμηση με τυπική γραμμική παλινδρόμηση χωρίς τυχαίο σφάλμα, η εκτίμηση με οργανική συσχέτιση, και η προσομοίωση με γραμμική παλινδρόμηση και συνυπολογισμό του τυχαίου σφάλματος, έχουν πλεονεκτήματα και μειονεκτήματα. Η επιλογή της κατάλληλης κατά περίπτωση μεθόδου διεύρυνσης του δείγματος εναπόκειται στο μελετητή υδρολόγο.

Γεννήτριες τυχαίων αριθμών ομοιόμορφης και κανονικής κατανομής

Παραθέτουμε εδώ χωρίς απόδειξη τους πιο διαδεδομένους αλγορίθμους για τη γέννηση τυχαίων αριθμών ομοιόμορφης και κανονικής κατανομής. Ο αναγνώστης που ενδιαφέρεται για την τεκμηρίωση των αλγορίθμων αυτών παραπέμπεται στον Papoulis, 1990, σσ. 251-372. Οι τυχαίοι αριθμοί ομοιόμορφης κατανομής αποτελούν τη βάση για τη γέννηση τυχαίων αριθμών από άλλες κατανομές. Σημειώνουμε ότι όλοι οι μεταφραστές (compilers) προγραμμάτων ηλεκτρονικών υπολογιστών διαθέτουν ενσωματωμένη γεννήτρια τυχαίων αριθμών ομοιόμορφης κατανομής, η οποία μπορεί να χρησιμοποιηθεί αντί αυτής που δίνεται εδώ. Επίσης, πολλά έτοιμα προγράμματα υπολογισμών (λογιστικά φύλλα κτλ.) διαθέτουν ενσωματωμένη τέτοια γεννήτρια, καθώς και γεννήτριες για άλλες κατανομές.

α. Γεννήτρια τυχαίων αριθμών ομοιόμορφης κατανομής

Οι ακέραιοι αριθμοί q_i που υπολογίζονται από τον αναδρομικό τύπο

$$q_i = (k q_{i-1} + c) \bmod m \quad (7.97)$$

όπου k , c και m κατάλληλες ακέραιες σταθερές, αποτελούν τυχαίους αριθμούς ομοιόμορφα κατανομημένους στο διάστημα $[1, m - 1]$. Στην παραπάνω εξίσωση ο συμβολισμός $a = b \bmod \gamma$ σημαίνει ότι ο αριθμός a είναι το υπόλοιπο της διαίρεσης του ακέραιου αριθμού b με τον ακέραιο αριθμό γ . Κατά συνέπεια η ακολουθία αριθμών

$$u_i = q_i / m \quad (7.98)$$

αποτελεί πρακτικά ακολουθία τυχαίων αριθμών συνεχούς τύπου στο διάστημα $(0, 1)$.

Οι ακέραιοι αριθμοί k και m επιλέγονται έτσι ώστε η περιοδικότητα της ακολουθίας των q_i να είναι κατά το δυνατό μεγαλύτερη (δηλαδή $m - 1$). Μια επιτυχημένη επιλογή είναι η

$$k = 16\,807, \quad c = 0, \quad m = 2^{31} - 1 = 2\,147\,483\,647 \quad (7.99)$$

Ωστόσο, η επιλογή αυτή δύσκολα υλοποιείται σε συνήθεις υπολογιστές, οι οποίοι εκτελούν πράξεις ακεραίων μεγέθους μέχρι 32 bits. Μια άλλη επιλογή, που υλοποιείται σε υπολογιστές με αριθμητική ικανότητα 32 bits, είναι η

$$k = 4096, \quad c = 150\,889, \quad m = 714\,025 \quad (7.100)^*$$

Η αρχική τιμή q_0 που χρειάζεται για να λειτουργήσει η γεννήτρια μπορεί να είναι οποιοσδήποτε ακέραιος μεταξύ του 1 και του $m - 1$.

α. Γεννήτρια τυχαίων αριθμών κανονικής κατανομής

Αν οι αριθμοί u_i και v_i είναι διαδοχικοί όροι ακολουθίας τυχαίων αριθμών με ομοιόμορφη κατανομή στο διάστημα $(0, 1)$, τότε αποδεικνύεται ότι οι αριθμοί

$$w_i = \sqrt{-2 \ln v_i} \cos \pi(2 - u_i), \quad z_i = \sqrt{-2 \ln v_i} \sin \pi(2 - u_i) \quad (7.101)$$

αποτελούν διαδοχικούς όρους ακολουθίας τυχαίων αριθμών με κανονική κατανομή $N(0, 1)$. Κατά συνέπεια οι αριθμοί

$$w'_i = w_i \sigma + \mu, \quad z'_i = z_i \sigma + \mu \quad (7.102)$$

* Για περισσότερες επιλογές καθώς και για αλγορίθμους κατάλληλους για υπολογιστές με μικρότερη αριθμητική ακρίβεια βλ. Press et al., 1987, σ. 198.

αποτελούν διαδοχικούς όρους ακολουθίας τυχαίων αριθμών με κανονική κατανομή $N(\mu, \sigma)$. Έτσι μια οποιαδήποτε ακολουθία τυχαίων αριθμών ομοιόμορφης κατανομής μπορεί να μετασχηματιστεί εύκολα σε (ισοπληθή) ακολουθία τυχαίων αριθμών κανονικής κατανομής.

Εφαρμογή 7.6

Στη δεύτερη και τρίτη στήλη του Πίν. 7.6 δίνονται τα ετήσια ύψη βροχής X και Y δύο κοντινών βροχομετρικών σταθμών της Δυτικής Ελλάδας για περίοδο 40 ετών. Για λόγους διερεύνησης και σύγκρισης, να αγνοηθούν τα παρατηρημένα δεδομένα της δεύτερης εικοσαετίας του δείγματος Y και στη συνέχεια, με βάση τα δεδομένα της πρώτης εικοσαετίας να γίνει επέκταση του δείγματος Y με τις μεθόδους (α) γραμμικής παλινδρόμησης χωρίς τυχαίο σφάλμα, (β) οργανικής συσχέτισης και (γ) γραμμικής παλινδρόμησης με τυχαίο σφάλμα. Να εκτιμηθούν τα στατιστικά χαρακτηριστικά της Y και ο συντελεστής συσχέτισης των X και Y για τις διάφορες περιπτώσεις επέκτασης.

Τα στατιστικά χαρακτηριστικά των δειγμάτων της πρώτης εικοσαετίας είναι:

$$k = 20, \bar{x} = 1492.4, \bar{y} = 1116.5, s_{*X} = 357.3, s_{*Y} = 301.2, r_{XY} = 0.644$$

Τα στατιστικά χαρακτηριστικά των συνολικών δειγμάτων των 40 ετών είναι:

$$n = 40, \bar{x}' = 1513.8, \bar{y}' = 1163.4, s'_{*X} = 359.6, s'_{*Y} = 317.3, r'_{XY} = 0.691$$

Εφαρμόζουμε κατ' αρχήν την τυπική γραμμική παλινδρόμηση. Οι παράμετροι της γραμμικής εξίσωσης $y = a + b x$, υπολογισμένες από τα δείγματα της πρώτης εικοσαετίας (εξ. (7.27) και (7.28)), είναι:

$$b = 0.644 \times 301.2 / 357.3 = 0.543$$

$$a = 1116.5 - 0.543 \times 1492.4 = 306.1$$

Εφαρμόζοντας τη γραμμική εξίσωση αυτή για τα δεδομένα x της δεύτερης εικοσαετίας, παίρνουμε τις εκτιμήσεις που φαίνονται στη στήλη $Y^{(1)}$ του Πίν. 7.6). Αν τώρα υπολογίσουμε τα στατιστικά χαρακτηριστικά των 40 ετών της στήλης $Y^{(1)}$ βρίσκουμε

$$\bar{y}' = 1128.1, s'_{*Y} = 253.0, r'_{XY} = 0.772.$$

Πίν. 7.6 Δεδομένα και αποτελέσματα των υπολογισμών της Εφαρμογής 7.6. Όλες οι στήλες εκτός της πρώτης περιέχουν ύψη βροχής σε mm. Η δεύτερη και η τρίτη στήλη περιέχουν τα πραγματικά δείγματα και οι υπόλοιπες τα δείγματα που έχουν προκύψει με επέκταση στη δεύτερη εικοσαετία του δείγματος Y της πρώτης εικοσαετίας με βάση τις εξεταζόμενες μεθόδους.

| A/A | X | Y | $Y^{(1)}$ | $Y^{(2)}$ | $Y^{(3)}$ |
|-----|--------|--------|-----------|-----------|-----------|
| 1 | 1767.7 | 1334.1 | 1334.1 | 1334.1 | 1334.1 |
| 2 | 1280.3 | 606.8 | 606.8 | 606.8 | 606.8 |
| 3 | 1618.2 | 1652.1 | 1652.1 | 1652.1 | 1652.1 |
| 4 | 1143.1 | 983.9 | 983.9 | 983.9 | 983.9 |
| 5 | 1964.1 | 1620.7 | 1620.7 | 1620.7 | 1620.7 |
| 6 | 1288.9 | 1022.6 | 1022.6 | 1022.6 | 1022.6 |
| 7 | 1979.6 | 1290.5 | 1290.5 | 1290.5 | 1290.5 |
| 8 | 1196.0 | 942.2 | 942.2 | 942.2 | 942.2 |
| 9 | 1354.1 | 1002.2 | 1002.2 | 1002.2 | 1002.2 |
| 10 | 968.3 | 972.5 | 972.5 | 972.5 | 972.5 |
| 11 | 1670.9 | 907.0 | 907.0 | 907.0 | 907.0 |
| 12 | 2052.9 | 1237.1 | 1237.1 | 1237.1 | 1237.1 |
| 13 | 910.5 | 506.9 | 506.9 | 506.9 | 506.9 |
| 14 | 980.4 | 792.7 | 792.7 | 792.7 | 792.7 |
| 15 | 2006.1 | 1086.9 | 1086.9 | 1086.9 | 1086.9 |
| 16 | 1616.0 | 1186.3 | 1186.3 | 1186.3 | 1186.3 |
| 17 | 1490.7 | 1090.9 | 1090.9 | 1090.9 | 1090.9 |
| 18 | 1511.6 | 1238.1 | 1238.1 | 1238.1 | 1238.1 |
| 19 | 1357.0 | 1473.3 | 1473.3 | 1473.3 | 1473.3 |
| 20 | 1691.1 | 1383.9 | 1383.9 | 1383.9 | 1383.9 |
| 21 | 1307.9 | 763.7 | 1016.3 | 961.0 | 1376.7 |
| 22 | 2078.1 | 1565.2 | 1434.5 | 1610.2 | 1694.6 |
| 23 | 1657.3 | 1241.8 | 1206.0 | 1255.5 | 970.6 |
| 24 | 659.0 | 815.8 | 663.9 | 413.9 | 362.2 |
| 25 | 1752.7 | 1278.4 | 1257.8 | 1335.9 | 987.7 |
| 26 | 1437.0 | 1021.9 | 1086.4 | 1069.8 | 740.1 |
| 27 | 1762.8 | 844.2 | 1263.3 | 1344.4 | 1126.2 |
| 28 | 1132.2 | 682.9 | 920.9 | 812.8 | 949.4 |
| 29 | 847.1 | 475.7 | 766.1 | 572.5 | 655.6 |
| 30 | 1629.8 | 1306.6 | 1191.1 | 1232.3 | 1243.9 |
| 31 | 1449.3 | 1560.0 | 1093.1 | 1080.2 | 1542.5 |
| 32 | 1434.3 | 1360.4 | 1084.9 | 1067.5 | 1166.8 |
| 33 | 1444.0 | 1279.2 | 1090.2 | 1075.7 | 1395.3 |
| 34 | 1880.6 | 1360.2 | 1327.3 | 1443.7 | 980.2 |
| 35 | 1438.6 | 1379.5 | 1087.3 | 1071.1 | 1104.8 |
| 36 | 1352.5 | 1187.0 | 1040.5 | 998.6 | 1051.8 |
| 37 | 2111.6 | 1601.1 | 1452.7 | 1638.5 | 1264.7 |
| 38 | 1760.2 | 1432.4 | 1261.9 | 1342.2 | 1319.5 |
| 39 | 1884.4 | 1403.5 | 1329.3 | 1446.9 | 1414.4 |
| 40 | 1685.7 | 1646.1 | 1221.4 | 1279.4 | 1216.0 |

Παρατηρούμε ότι υπάρχει σημαντική υπεκτίμηση της τυπικής απόκλισης (κατά 20% περίπου σε σχέση με την πραγματική τιμή 317.3) και υπερεκτίμηση του συντελεστή συσχέτισης (κατά 12% περίπου σε σχέση με την πραγματική τιμή 0.691).

Η τυπική απόκλιση μπορεί να διορθωθεί με πολλαπλασιασμό επί τον συντελεστή $\sqrt{\varphi}$ της εξίσωσης (7.84), όπου

$$\varphi = (40 - 1) / [20 + (40 - 20) \times 0.644^2 - 1] = 1.429$$

οπότε βρίσκουμε

$$s'_{*Y} = 253.0 \times \sqrt{1.429} = 302.4$$

τιμή πολύ κοντύτερα στην πραγματική. Εναλλακτικά, μπορούμε να εφαρμόσουμε την εξίσωση (7.86), οπότε βρίσκουμε

$$s'^2_{*Y} = 301.2^2 + 0.543^2 \times (359.6^2 - 357.3^2) = 91\,208$$

και $s'_{*Y} = 302.0$. Εξ άλλου, η περίπου αμερόληπτη εκτίμηση του συντελεστή συσχέτισης του διευρυμένου δείγματος δίνεται από την εξίσωση (7.87) και είναι

$$r'_{XY} = 0.644 \times (301.2 / 357.3) \times (359.6 / 302.0) = 0.646$$

Τέλος, το ισοδύναμο μήκος του διευρυμένου δείγματος δίνεται από την (7.92), ήτοι

$$k / n' = 1 - (20/40) \times [0.644^2 - (1 - 0.644^2) / (20 - 3)] = 0.81$$

άρα

$$n' = 20 / 0.810 \approx 25$$

(έναντι $n = 40$).

Στη συνέχεια εφαρμόζουμε τη μέθοδο της οργανικής συσχέτισης. Οι παράμετροι της γραμμικής εξίσωσης $y = a + b x$, υπολογισμένες από τα δείγματα της πρώτης εικοσαετίας (εξ. (7.54) και (7.55)), είναι:

$$b = 301.2 / 357.3 = 0.843, a = 1116.5 - 0.843 \times 1492.4 = -141.6$$

Εφαρμόζοντας τη γραμμική εξίσωση αυτή για τα δεδομένα x της δεύτερης εικοσαετίας, παίρνουμε τις εκτιμήσεις που φαίνονται στη στήλη $Y^{(2)}$ του Πίν. 7.6. Αν τώρα υπολογίσουμε τα στατιστικά χαρακτηριστικά των 40 ετών της στήλης $Y^{(2)}$ βρίσκουμε

$$\bar{y}' = 1134.6, s'_{*Y} = 303.1, r'_{XY} = 0.829.$$

Παρατηρούμε ότι η εκτίμηση της τυπικής απόκλισης είναι στην περίπτωση αυτή ικανοποιητική και δεν χρειάζεται αναγωγή. Αντίθετα, υπάρχει σημαντική υπερεκτίμηση του συντελεστή συσχέτισης (κατά 20% περίπου σε σχέση με την πραγματική τιμή 0.691). Η περίπου αμερόληπτη εκτίμηση του συντελεστή συσχέτισης του διευρυμένου δείγματος μπορεί να υπολογιστεί και πάλι από την εξίσωση (7.87), οπότε θα είναι

$$r'_{XY} = 0.644 \times (301.2 / 357.3) \times (359.6 / 303.1) = 0.643$$

Τέλος εφαρμόζουμε τη μέθοδο της γραμμικής παλινδρόμησης με όρο σφάλματος. Χρησιμοποιούμε την εξίσωση $y = a + b x + w$, όπου οι παράμετροι έχουν τις τιμές που έχουν υπολογιστεί στην τυπική γραμμική παλινδρόμηση, δηλαδή $b = 0.543$ και $a = 306.1$. Ο όσος σφάλματος έχει μέση τιμή 0 και τυπική απόκλιση που δίνεται από την εξίσωση (7.82), δηλαδή

$$\begin{aligned} \hat{\sigma}_w &= \sqrt{n/(n-2)} s_Y \sqrt{1 - r_{XY}^2} = \\ &= \sqrt{20/18} \times (301.2 \times \sqrt{19/20}) \times \sqrt{1 - 0.644^2} = 236.7 \end{aligned}$$

Πριν εφαρμόσουμε την πιο πάνω γραμμική εξίσωση γεννούμε 20 τυχαίους αριθμούς από την τυποποιημένη κανονική κατανομή. Εδώ έχουμε χρησιμοποιήσει τη βοήθεια ρουτίνας ηλεκτρονικού υπολογιστή και οι 20 τυχαίοι αριθμοί που γεννήσαμε είναι οι ακόλουθοι:

$$\begin{aligned} &+1.52281 \quad +1.09894 \quad -0.99478 \quad -1.27496 \quad -1.14112 \\ &-1.46306 \quad -0.57916 \quad +0.12054 \quad -0.46660 \quad +0.22329 \\ &+1.89879 \quad +0.34577 \quad +1.28891 \quad -1.46615 \quad +0.07426 \\ &+0.04751 \quad -0.79445 \quad +0.24338 \quad +0.35920 \quad -0.02287 \end{aligned}$$

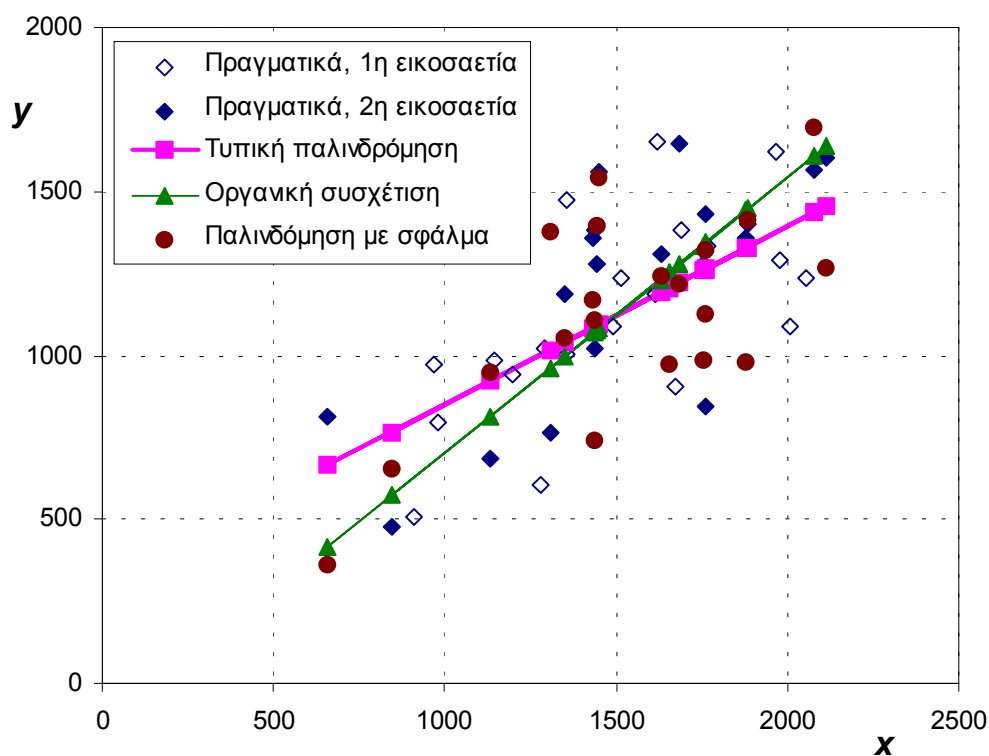
Για να βρούμε τους αριθμούς w_i που είναι απαραίτητοι, πολλαπλασιάζουμε τους παραπάνω τυχαίους αριθμούς επί την τυπική απόκλιση $\hat{\sigma}_w = 236.7$. Εφαρμόζοντας τη γραμμική εξίσωση $y = a + b x + w$, για τα δεδομένα x_i της δεύτερης εικοσαετίας, παίρνουμε τις τιμές που φαίνονται στη στήλη $Y^{(3)}$ του Πίν. 7.6). Αν υπολογίσουμε τα στατιστικά χαρακτηριστικά των 40 ετών της στήλης $Y^{(3)}$ βρίσκουμε

$$\bar{y}' = 1122.3, \quad s'_{*Y} = 302.8, \quad r'_{XY} = 0.654$$

Παρατηρούμε ότι αυτά τα στατιστικά χαρακτηριστικά βρίσκονται κοντά στις πραγματικές τιμές και δεν χρειάζονται αναγωγές ή διορθώσεις.

Τα τρία διευρυμένα δείγματα της μεταβλητής Y , που έχουμε υπολογίσει παραπάνω, τα έχουμε απεικονίσει στο Σχ. 7.19, συναρτήσει των

τιμών της X . Είναι εμφανές ότι το δείγμα της τρίτης μεθόδου είναι το πιο ρεαλιστικό, με την έννοια ότι τα σημεία του γενικά διατάσσονται με παρόμοιο τρόπο, όπως οι πραγματικές μετρήσεις. Αυτό βέβαια δεν σημαίνει ότι κάθε σημείο πλησιάζει στην πραγματική του θέση περισσότερο απ' ό,τι τα σημεία των άλλων δύο μεθόδων. Το βασικό μειονέκτημα της τρίτης μεθόδου είναι ότι τα σημεία της δεν ορίζονται με μονοσήμαντο τρόπο, αφού αν χρησιμοποιούσαμε άλλους τυχαίους αριθμούς, θα είχαμε διαφορετικά αποτελέσματα.



Σχ. 7.19 Γραφική απεικόνιση των τιμών της Y συναρτήσει των τιμών της X στα δεδομένα της Εφαρμογής 7.6.

Παράρτημα 7Α: Παραγωγή των εκτιμήσεων του διευρυμένου δείγματος και σχετικές αποδείξεις

Θα εξαγάγουμε πρώτα την εκτιμήτρια της μέσης τιμής του διευρυμένου δείγματος. Από την εξίσωση (7.24), παίρνοντας υπόψη και την (7.28), για το τυχόν ζεύγος μεταβλητών (X_i, Y_i) έχουμε

$$Y_i = \mu_Y + b(X_i - \mu_X) + W_i \quad (7.103)$$

Γράφοντας την (7.103) για όλα τα ζεύγη (X_i, Y_i) , αθροίζοντας και διαιρώντας διά του μεγέθους του δείγματος k , βρίσκουμε

$$\frac{1}{k} \sum_{i=1}^k Y_i = \mu_Y + b \left(\frac{1}{k} \sum_{i=1}^k X_i - \mu_X \right) + \frac{1}{k} \sum_{i=1}^k W_i \quad (7.104)$$

ή, ισοδύναμα,

$$\bar{Y} = \mu_Y + b(\bar{X} - \mu_X) + \bar{W} \quad (7.105)$$

Όμοια, για το διευρυμένο δείγμα, μεγέθους n , έχουμε

$$\bar{Y}' = \mu_Y + b(\bar{X}' - \mu_X) + \bar{W}' \quad (7.106)$$

Αφαιρώντας τις (7.104) και (7.105) παίρνουμε

$$\bar{Y}' - \bar{Y} = b(\bar{X}' - \bar{X}) + \bar{W}' - \bar{W} \quad (7.107)$$

Αν στην τελευταία εξίσωση παραλείψουμε τους όρους σφάλματος, για καθέναν από τους οποίους η αναμενόμενη τιμή είναι μηδενική, και στη θέση των εκτιμητριών βάλουμε τις εκτιμήσεις, παίρνουμε την (7.85).

Ερχόμαστε τώρα στην εκτιμήτρια της διασποράς. Αφαιρώντας από την (7.105) την (7.103) έχουμε

$$Y_i - \bar{Y} = b(X_i - \bar{X}) + W_i - \bar{W} \quad (7.108)$$

και υψώνοντας στο τετράγωνο παίρνουμε

$$(Y_i - \bar{Y})^2 = b^2(X_i - \bar{X})^2 + (W_i - \bar{W})^2 + 2b(X_i - \bar{X})(W_i - \bar{W}) \quad (7.109)$$

Γράφοντας την παραπάνω για όλα τα ζεύγη (X_i, Y_i) , αθροίζοντας και διαιρώντας διά $k - 1$, βρίσκουμε

$$S_{*Y}^2 = b^2 S_{*X}^2 + S_{*W}^2 + 2b S_{*XW} \quad (7.110)$$

Γράφοντας την ίδια εξίσωση για το διευρυμένο δείγμα και στη συνέχεια αφαιρώντας από αυτή την (7.110) παίρνουμε

$$S'_{*Y}{}^2 = S_{*Y}^2 + b^2 \left(S'_{*X}{}^2 - S_{*X}^2 \right) + \left(S'_{*W}{}^2 - S_{*W}^2 \right) + 2b \left(S'_{*XW} - S_{*XW} \right) \quad (7.111)$$

Αν στην τελευταία εξίσωση παραλείψουμε τους όρους σφάλματος και στη θέση των εκτιμητριών βάλουμε τις εκτιμήσεις, παίρνουμε την (7.86). Η παράλειψη αυτή δεν εισάγει μεροληψία δεδομένου ότι

$$E[S'_{*W}{}^2] = E[S_{*W}^2] = \sigma_W^2, \quad E[S'_{*XW}] = E[S_{*XW}] = 0 \quad (7.112)$$

και επομένως η συνολική συνεισφορά τους στην αναμενόμενη τιμή του δεξιού μέλους της (7.111) είναι μηδενική.

Στη συνέχεια θα εξετάσουμε την εκτιμήτρια του συντελεστή συσχέτισης του διευρυμένου δείγματος. Πολλαπλασιάζοντας τα δύο μέλη της (7.108) επί $(X_i - \bar{X})$ παίρνουμε

$$(Y_i - \bar{Y})(X_i - \bar{X}) = b(X_i - \bar{X})^2 + (W_i - \bar{W})(X_i - \bar{X}) \quad (7.113)$$

και εργαζόμενοι όπως προηγουμένως καταλήγουμε ότι

$$S_{*XY} = b S_{*X}^2 + S_{*XW} \quad (7.114)$$

Παραλείποντας τον τελευταίο όρο, ο οποίος έχει μηδενική συνεισφορά σε αναμενόμενη τιμή, δεδομένου ότι $E[S_{*XW}] = 0$, η (7.114) απλοποιείται, χωρίς εισαγωγή μεροληψίας, και γίνεται

$$S_{*XY} = b S_{*X}^2 \quad (7.115)$$

Υπό αυτές τις συνθήκες, η εκτιμήτρια του συντελεστή συσχέτισης είναι

$$R_{XY} = \frac{S^*_{XY}}{S^*_X S^*_Y} = b \frac{S^*_X}{S^*_Y} \quad (7.116)$$

Γράφοντας την ίδια εξίσωση για το διευρυμένο δείγμα και στη συνέχεια διαιρώντας τη με τη (7.116) παίρνουμε

$$R'_{XY} = R_{XY} \frac{S^*_Y}{S^*_X} \frac{S'_{*X}}{S'_{*Y}} \quad (7.117)$$

Αν στην (7.117) αντικαταστήσουμε τις εκτιμήτριες με τις εκτιμήσεις παίρνουμε την (7.87).

Θα υπολογίσουμε τώρα τη διασπορά της εκτιμήτριας της μέσης τιμής του διευρυμένου δείγματος, όταν η μέση τιμή αυτή δίνεται από την εξίσωση (7.85). Η εξίσωση (7.85) γράφεται ισοδύναμα

$$\begin{aligned} (\bar{Y}' - \mu_Y) &= (\bar{Y} - \mu_Y) + b(\bar{X}' - \mu_X) - b(\bar{X} - \mu_X) \\ &= \frac{1}{k} \sum_{i=1}^k (Y_i - \mu_Y) + \frac{b}{n} \sum_{i=1}^n (X_i - \mu_X) - \frac{b}{k} \sum_{i=1}^k (X_i - \mu_X) \end{aligned} \quad (7.118)$$

Υψώνουμε την παραπάνω στο τετράγωνο και στη συνέχεια παίρνουμε αναμενόμενες τιμές. Παρατηρούμε ότι λόγω της ανεξαρτησίας των διαφόρων ζευγών μεταβλητών (X_i, Y_i) ισχύουν τα ακόλουθα:

$$E[(Y_i - \mu_Y)(Y_j - \mu_Y)] = \begin{cases} 0 & i \neq j \\ \sigma_Y^2 & i = j \end{cases} \quad (7.119)$$

$$E[(X_i - \mu_X)(X_j - \mu_X)] = \begin{cases} 0 & i \neq j \\ \sigma_X^2 & i = j \end{cases} \quad (7.120)$$

$$E[(X_i - \mu_X)(Y_j - \mu_Y)] = \begin{cases} 0 & i \neq j \\ \sigma_{XY} & i = j \end{cases} \quad (7.121)$$

Έτσι βρίσκουμε

$$\begin{aligned} \text{Var}[\bar{Y}'] &= \frac{1}{k^2} k \sigma_Y^2 + \frac{b^2}{n^2} n \sigma_X^2 + \frac{b^2}{k^2} k \sigma_X^2 \\ &+ 2 \frac{1}{k} \frac{b}{n} k \sigma_{XY} - 2 \frac{1}{k} \frac{b}{k} k \sigma_{XY} - 2 \frac{b}{n} \frac{b}{k} k \sigma_X^2 \end{aligned} \quad (7.122)$$

Παίρνοντας υπόψη ότι $b^2 \sigma_X^2 = b \sigma_{XY} = \rho^2 \sigma_Y^2$, βρίσκουμε ότι

$$\text{Var}[\bar{Y}'] = \frac{\sigma_Y^2}{k} \left[1 + \rho_{XY}^2 \left(\frac{k}{n} + 1 + 2 \frac{k}{n} - 2 - 2 \frac{k}{n} \right) \right] \quad (7.123)$$

και τελικά, κάνοντας τις πράξεις παίρνουμε την εξίσωση (7.88).

Τέλος, θα δώσουμε την απόδειξη για την εξίσωση (7.84), η οποία δίνει το συντελεστή διόρθωσης της μεροληψίας για την περίπτωση που η διασπορά υπολογίζεται από το διευρυμένο δείγμα, χρησιμοποιώντας τον τύπο της δειγματικής διασποράς παρατηρημένου δείγματος. Η δειγματική διασπορά αυτή γράφεται

$$S_{*Y}^{\prime\prime 2} = \frac{1}{n-1} \left[\sum_{i=1}^k (Y_i - \bar{Y}')^2 + \sum_{i=k+1}^n (\hat{Y}_i - \bar{Y}')^2 \right] \quad (7.124)$$

όπου ο πρώτος όρος μέσα στην τετραγωνική αγκύλη αναφέρεται στις παρατηρημένες τιμές της μεταβλητής Y και ο δεύτερος στις εκτιμημένες μέσω της μεταβλητής X . Για τις πρώτες ισχύει

$$Y_i - \mu_Y = (Y_i - \bar{Y}') + (\bar{Y}' - \mu_Y) \quad (7.125)$$

και κατά συνέπεια

$$(Y_i - \mu_Y)^2 = (Y_i - \bar{Y}')^2 + (\bar{Y}' - \mu_Y)^2 + 2 (Y_i - \bar{Y}')(\bar{Y}' - \mu_Y) \quad (7.126)$$

Για τις δεύτερες ισχύει

$$Y_i - \mu_Y = (\hat{Y}_i - \bar{Y}') + (\bar{Y}' - \mu_Y) + W_i \quad (7.127)$$

και κατά συνέπεια

$$(Y_i - \mu_Y)^2 = (\hat{Y}_i - \bar{Y}')^2 + (\bar{Y}' - \mu_Y)^2 + 2 (\hat{Y}_i - \bar{Y}')(\bar{Y}' - \mu_Y) + W_i^2 + 2 W_i (\hat{Y}_i - \mu_Y) \quad (7.128)$$

Σχηματίζουμε την (7.126) για $i = 1$ μέχρι k και την (7.128) για $i = k + 1$ μέχρι n , προσθέτουμε τις n αυτές εξισώσεις και παίρνουμε αναμενόμενες τιμές. Στο αριστερό μέλος της εξίσωσης θα προκύψει n φορές η διασπορά της Y , δεδομένου ότι

$$E[(Y_i - \mu_Y)^2] = \sigma_Y^2 \quad (7.129)$$

Στο δεξιό μέλος της εξίσωσης θα προκύψει κατ' αρχήν το άθροισμα που βρίσκεται μέσα στην αγκύλη της εξίσωσης (7.124) και οι υπόλοιποι όροι για τους οποίους ισχύουν

$$E\left[\sum_{i=1}^n (\bar{Y}' - \mu_Y)^2\right] = n \text{Var}[\bar{Y}'] = \frac{n}{k} \left(1 - \frac{n-k}{n} \rho_{XY}^2\right) \quad (7.130)$$

(λόγω της (7.88)),

$$\sum_{i=1}^k (Y_i - \bar{Y}') + \sum_{i=k+1}^n (\hat{Y}_i - \bar{Y}') = 0 \quad (7.131)$$

και

$$E[W_i^2] = (1 - \rho_{XY}^2)\sigma_Y^2, \quad E[W_i (\hat{Y}_i - \mu_Y)] = 0 \quad (7.132)$$

Κατά συνέπεια βρίσκουμε

$$n\sigma_Y^2 = (n-1)E[S_{*Y}''^2] + \frac{n}{k} \left(1 - \frac{n-k}{n} \rho_{XY}^2\right) \sigma_Y^2 + (n-k) (1 - \rho_{XY}^2) \sigma_Y^2 \quad (7.133)$$

Λύνοντας την τελευταία ως προς $E[S_{*Y}''^2]$ παίρνουμε

$$E[S_{*Y}''^2] = \sigma_Y^2 \frac{k + (n - k) \rho_{XY}^2 - 1 - (1 - \rho_{XY}^2) \frac{n - k}{k}}{n - 1} \quad (7.134)$$

Ο τελευταίος όρος στον αριθμητή του δεξιού μέλους της (7.134) μπορεί να παραλειφθεί χωρίς ουσιαστικό σφάλμα, δεδομένου ότι είναι πολύ μικρότερος από το άθροισμα των άλλων όρων. Παραλείποντας τον όρο αυτό και αντιστρέφοντας το κλάσμα παίρνουμε το συντελεστή διόρθωσης της (7.84).

Κεφάλαιο 8 Ανάλυση τυχαίας μεταβλητής εξαρτημένης από παράμετρο - Όμβριες καμπύλες

Στο κεφάλαιο αυτό θα εφαρμόσουμε τις γνώσεις όλων των προηγούμενων κεφαλαίων σε ένα τυπικό πρόβλημα της τεχνικής υδρολογίας, το οποίο περιλαμβάνει την πιθανοτική ανάλυση περισσότερων από μίας τυχαίων μεταβλητών, καθενιά από τις οποίες αντιστοιχεί σε μια δεδομένη τιμή μιας παραμέτρου. Το πρόβλημα αυτό το αντιμετωπίζουμε στην ανάλυση των ισχυρών βροχοπτώσεων μιας περιοχής, η οποία καταλήγει στην κατάρτιση των αποκαλούμενων όμβριων καμπυλών. Οι *όμβριες καμπύλες*, όπως απλούστερα έχει καθιερωθεί να αποκαλούνται στην ελληνική τεχνική ορολογία οι *καμπύλες έντασης-διάρκειας-περιόδου επαναφοράς* βροχόπτωσης, αποτελούν ένα από τα βασικότερα εργαλεία του υδρολόγου μηχανικού για το σχεδιασμό αντιπλημμυρικών έργων. Πρόκειται για απλές αναλυτικές ή γραφικές εκφράσεις της μέγιστης έντασης βροχής i συναρτήσει της διάρκειας d και της περιόδου επαναφοράς T . Για την κατάρτιση των όμβριων καμπυλών πρέπει να είναι διαθέσιμες ιστορικές σειρές μέγιστων εντάσεων βροχής για ένα σύνολο k διαρκειών d_j , $j = 1, \dots, k$, ξεκινώντας από την ελάχιστη διάρκεια που επιτρέπει η ευκρίνεια των παρατηρήσεων (π.χ. 5 min - 1 h) και φθάνοντας μέχρι τη μέγιστη διάρκεια βροχής που ενδιαφέρει στα τυπικά προβλήματα του μηχανικού (π.χ. 24 - 48 h). Αναλυτικά, ο ορισμός των μεγεθών αυτών και ο τρόπος κατασκευής της σειράς τιμών της μεταβλη-

τής i από τη διαθέσιμη πλήρη χρονοσειρά εντάσεων βροχής φαίνεται στην ενότητα 8.1.

Η χρησιμότητα των όμβριων καμπυλών είναι άμεση σε όλα τα προβλήματα υδρολογικού σχεδιασμού που αφορούν στην αντιπλημμυρική προστασία. Ουσιαστικά η χρήση τους συνίσταται στην πρόγνωση της έντασης βροχής i , για δεδομένη διάρκεια d , η οποία έχει σχέση με τα χαρακτηριστικά του υδατορεύματος ή αγωγού που μελετάται, και για δεδομένη περίοδο επαναφοράς T , που έχει σχέση με τη σημασία του έργου που μελετάται. Οι όμβριες καμπύλες αποτελούν τη βασική είσοδο σε όλα τα μοντέλα μετασχηματισμού της βροχόπτωσης σε πλημμυρική απορροή, ξεκινώντας από την απλή ορθολογική μέθοδο ($Q = c i A$) και φτάνοντας σε πιο σύνθετες μεθόδους, όπως αυτές του μοναδιαίου υδρογραφήματος, και των υδρολογικών και υδραυλικών μοντέλων διόδευσης. Στο κείμενο αυτό δεν θα αναφερθούμε στην εφαρμογή των όμβριων καμπυλών σε υδρολογικά προβλήματα, αλλά στον τρόπο κατάρτισής τους. Για τις εφαρμογές ο αναγνώστης παραπέμπεται σε πληρέστερα βιβλία υδρολογίας και υδραυλικών έργων, όπως στους Ξανθόπουλο, 1990, Μιμίκου, 1990 και Κουτσογιάννη, 1993 (για να περιοριστούμε στην ελληνική βιβλιογραφία).

Το πρόβλημα της κατάρτισης όμβριων καμπυλών είναι κάπως ιδιόμορφο. Κατ' αρχήν δεν μπορεί να αντιμετωπιστεί ως πρόβλημα μιας τυχαίας μεταβλητής, αφού περιλαμβάνει την ταυτόχρονη αντιμετώπιση της έντασης i και της διάρκειας d . Ούτε ως πρόβλημα δύο τυχαίων μεταβλητών μπορεί να θεωρηθεί αφού η ένταση i αντιστοιχεί πράγματι σε μια τυχαία μεταβλητή I , αλλά η διάρκεια d δεν αποτελεί τυχαία μεταβλητή αλλά παράμετρο (βλ. ενότητα 8.1). Έτσι έχουμε να αντιμετωπίσουμε μια οικογένεια τυχαίων μεταβλητών $I(d)$, όπου η παράμετρος d παίρνει (θεωρητικά άπειρες) τιμές από ένα πραγματικό διάστημα. Η οικογένεια τυχαίων μεταβλητών $I(d)$ δεν αποτελεί τυπική στοχαστική ανέλιξη (βλ. εδάφιο 4.1.2) αφού η μεταβλητή I δεν αποτελεί πραγματική ένταση, αλλά μέγιστη μέση ένταση για διάρκεια d , αλλά και η διάρκεια d δεν συμβολίζει συνεχή (ή διακριτό) χρόνο, αλλά χρονικό διάστημα (βλ. ενότητα 8.1). Όπως θα δούμε παρακάτω, οι μεταβλητές $I(d)$ κατασκευάζονται με κάποιο ειδικό τρόπο από την ανέλιξη της πραγματικής έντασης βροχής σε συνεχή χρόνο. Στην πραγματικότητα, δεν είναι απαραίτητη η χρήση της θεωρίας στοχαστικών ανελιξεων στο πρόβλημα που αντιμετωπί-

ζουμε, γιατί δεν μας ενδιαφέρει η πολυδιάστατη κατανομή των μεταβλητών $I(d)$. Αντίθετα, μας ενδιαφέρει μόνο η συνάρτηση κατανομής πρώτης τάξης της μεταβλητής $I(d)$, δηλαδή η $F(i; d) = P(I(d) < i)$, η οποία αποτελεί και τον τελικό στόχο της ανάλυσης αυτού του κεφαλαίου. Πράγματι, η συνάρτηση $F(i; d)$ μετατρέπεται άμεσα σε σχέση μεταξύ των μεγεθών i, d, T , δεδομένου ότι $T = 1 / (1 - F)$.

Για την κατάστρωση αυτής της σχέσης έχουν αναπτυχθεί διάφορες μέθοδοι στην τεχνική υδρολογία, από καθαρά εμπειρικές χωρίς στατιστική θεμελίωση, μέχρι συνεπείς στατιστικές. Στο κεφάλαιο αυτό δεν θα περιοριστούμε σε μία μοναδική μέθοδο, αλλά θα δώσουμε μια σειρά μεθόδων, παραλείποντας μόνο τις καθαρά εμπειρικές.

Τελειώνοντας αυτή την εισαγωγή θα πρέπει να τονίσουμε ότι η κατάρτιση όμβριων καμπυλών προϋποθέτει την ύπαρξη μετρήσεων της βροχής με υψηλή χρονική ευκρίνεια (π.χ. από μερικά λεπτά μέχρι μία ώρα). Τέτοια ευκρίνεια παρέχεται από τις ταινίες αυτογραφικών οργάνων (βροχογράφων) ή από τους πιο σύγχρονους ψηφιακούς αισθητήρες του ύψους βροχής. Τα δεδομένα από βροχόμετρο, που κανονικά έχουν ευκρίνεια μίας ημέρας δεν προσφέρονται για την εξαγωγή όμβριων καμπυλών. Ωστόσο, και στην περίπτωση του βροχομέτρου, αν συναξιολογηθούν και τα δεδομένα γειτονικών βροχογράφων, είναι δυνατή η εξαγωγή όμβριων καμπυλών με έμμεσο τρόπο. Μια ενδεικτική μεθοδολογία για τη συναξιολόγηση των δεδομένων χαμηλής χρονικής ευκρίνειας δίνεται στην ενότητα 8.5· σημειώνουμε, ωστόσο, ότι δεν υπάρχει ενιαία καθιερωμένη μεθοδολογία για αυτό το σκοπό. Είναι βεβαίως προφανές ότι στην περίπτωση που λείπουν τα άμεσα δεδομένα υψηλής ευκρίνειας στην υπό εξέταση θέση, η αξιοπιστία των όμβριων καμπυλών είναι περιορισμένη.

8.1 Έννοια και τρόπος κατασκευής των μεταβλητών

Ο απλούστερος τρόπος για να διευκρινίσουμε την έννοια των μεταβλητών i και d που υπεισέρχονται στο πρόβλημα είναι να δώσουμε τον τρόπο κατασκευής της σειράς τιμών της μεταβλητής i από πλήρη χρονοσειρά εντάσεων βροχής $\zeta(t)$, βασισμένη στα βροχογραφήματα ενός συγκεκριμένου σταθμού. Η κατασκευή αυτή περιλαμβάνει τα ακόλουθα βήματα (βλ. και ενότητα 4.1):

1. Επιλέγουμε μια δεδομένη διάρκεια d (συνήθως από μερικά λεπτά της ώρας μέχρι μερικές ώρες), η οποία λειτουργεί ως “χρονικό παράθυρο” μέσα από το οποίο βλέπουμε τη συνεχή χρονοσειρά $\zeta(t)$.
2. Μετακινώντας το χρονικό παράθυρο μήκους d κατά τη διάρκεια του χρόνου για τον οποίο διατίθενται μετρήσεις της $\zeta(t)$, υπολογίζουμε την σειρά κινούμενων μέσων όρων $\zeta_d(t)$, ήτοι

$$\zeta_d(t) = \frac{1}{d} \int_{t-d}^t \zeta(s) ds \quad (8.1)$$

Στην πραγματικότητα, επειδή ποτέ δεν έχουμε μετρήσεις της στιγμιαίας έντασης βροχής $\zeta(t)$ αλλά της μέσης έντασης $\zeta_\delta(t)$ για μια δεδομένη ευκρίνεια δ (συνήθως 5-10 min μέχρι 1 h), η παραπάνω σχέση τροποποιείται και γίνεται

$$\zeta_d(t) = \frac{\delta}{d} \sum_{i=0}^{N-1} \zeta_\delta(t - i \delta) \quad (8.2)$$

όπου υποτίθεται ότι η διάρκεια d είναι ακέραιο πολλαπλάσιο της ευκρίνειας δ , ήτοι $d = N \delta$.

3. Στη συνέχεια, επιλέγοντας τις αιχμές της χρονοσειράς κινούμενων μέσων όρων $\zeta_d(t)$ σχηματίζουμε μια σειρά μέγιστων εντάσεων $i_l(d)$ ($l = 1, \dots, n$) η οποία περιλαμβάνει n τιμές, όπου n είναι το πλήθος των υδρολογικών ετών για τα οποία διατίθενται μετρήσεις. Αυτό μπορεί να γίνει με δύο τρόπους:

- 3.1 Σύμφωνα με τον πρώτο τρόπο σχηματίζουμε τη σειρά ετήσιων μεγίστων, παίρνοντας την αιχμή κάθε υδρολογικού έτους, δηλαδή

$$i_l(d) := \max_{\Gamma < t < \Gamma^+} \{\zeta_d(t)\} \quad (8.3)$$

όπου Γ και Γ^+ είναι ο χρόνος έναρξης και λήξης του l υδρολογικού έτους.

3.2 Σύμφωνα με το δεύτερο τρόπο σχηματίζουμε τη σειρά υπεράνω κατώφλιου (γνωστή και ως σειρά μερικής διάρκειας) παίρνοντας όσες τιμές υπερβαίνουν ένα δεδομένο κατώφλι φ . Για εξασφάλιση της στοχαστικής ανεξαρτησίας μεταξύ των διαδοχικών τιμών αυτής της σειράς θέτουμε και ένα όριο τ ελάχιστης χρονικής απόστασης για τις διαδοχικές τιμές (π.χ. μία ημέρα ή περισσότερο). Τέλος, επιλέγουμε (μετά από δοκιμές) το κατώφλι φ έτσι ώστε η σειρά $\{i_l(d)\}$ να περιλαμβάνει ακριβώς n τιμές (όσα είναι τα υδρολογικά έτη). Συμβολικά, η σειρά υπεράνω κατώφλιου είναι

$$\{i_l(d), l = 1, \dots, n\} := \left\{ \zeta_d(t_l) \mid \zeta_d(t_l) > \varphi, t_l > t_{l-1} + \tau, \zeta_d(t_l) = \max_{t_l - \tau < t < t_l + \tau} \{ \zeta_d(t) \} \right\} \quad (8.4)$$

όπου οι τρεις συνθήκες του δεξιού μέλους πρέπει να ικανοποιούνται ταυτόχρονα, διαφορετικά δεν επιλέγεται το σημείο t_l και η αντίστοιχη ένταση $\zeta_d(t_l)$.

Στην πραγματικότητα οι παραπάνω εργασίες γίνονται ταυτόχρονα για ένα σύνολο k διαρκειών $d_j, j = 1, \dots, k$, ξεκινώντας από ελάχιστη διάρκεια ίση με την ευκρίνεια δ των παρατηρήσεων (π.χ. 5-10 min για βροχογράφο ημερήσιας ταινίας και 1 h για βροχογράφο εβδομαδιαίας ταινίας) και φθάνοντας μέχρι τη μέγιστη διάρκεια βροχής που ενδιαφέρει στα τυπικά προβλήματα του μηχανικού (π.χ. 24 ή 48 h). Κανονικά όλες οι k σειρές θα πρέπει να έχουν τον ίδιο αριθμό δεδομένων n , αλλά, λόγω των ελλείψεων που συχνά υπάρχουν στα πρωτογενή δεδομένα, είναι δυνατό ο αριθμός αυτός (n_j) να διαφέρει από διάρκεια σε διάρκεια.

Ο πιο πάνω τρόπος κατασκευής των σειρών μέγιστων εντάσεων μας επιτρέπει να κάνουμε τις ακόλουθες διευκρινιστικές παρατηρήσεις:

1. Η διάρκεια d δεν είναι τυχαία μεταβλητή αλλά παράμετρος. Δεν έχει σχέση με την πραγματική διάρκεια των επεισοδίων βροχής, αλλά εκφράζει τη χρονική διάρκεια για την οποία εξάγεται η μέση ένταση βροχής.
2. Η σειρά μέγιστων εντάσεων $i(d)$ που κατασκευάστηκε με έναν από τους δύο παραπάνω τρόπους δεν περιλαμβάνει στιγμιαίες εντάσεις, αλλά μέσες εντάσεις για διάρκεια d . Η σειρά αυτή θεωρείται ότι απο-

τελεί τυχαίο δείγμα της τυχαίας μεταβλητής $I(d)$. Χάριν συντομίας κάθε τιμή $i(d)$ την αποκαλούμε *μέγιστη ένταση βροχής για διάρκεια d αντί μέγιστη μέση ένταση για διάρκεια d* , όπως θα ήταν ακριβέστερο.

3. Από τον τρόπο κατασκευής των σειρών $i(d)$ είναι προφανές ότι για δύο διάρκειες $d_1 < d_2$ ισχύει

$$i_l(d_1) \geq i_l(d_2) \geq (d_1 / d_2) i_l(d_1) \quad (8.5)$$

Η σχέση αυτή μας επιτρέπει να ελέγξουμε (χονδροειδώς) τις σειρές διάφορων διαρκειών, οι οποίες συχνά δίνονται έτοιμες από τις αρμόδιες υπηρεσίες, και να εντοπίσουμε τυχόν σφάλματα της συγκεκριμένης μορφής.

Εφαρμογή 8.1

Στον Πίν. 8.1 (στήλες 1 και 2) φαίνονται οι καταγραφές του ύψους βροχής ανά δεκάλεπτο του αυτόματου τηλεμετρικού μετεωρολογικού σταθμού του ΕΜΠ στην Πολυτεχνειούπολη Ζωγράφου, για το επεισόδιο βροχής της 31/05/1994. Να υπολογιστούν οι μέγιστες μέσες εντάσεις βροχής για διάρκειες 10 min, 20 min, 30 min, 1 h, 2h και 4 h. Σημειώνεται ότι οι προηγούμενες και επόμενες καταγραφές έξω από τις ώρες που δίνονται στον Πίν. 8.1 είναι μηδενικές για χρονικά διαστήματα αρκετών ημερών.

Η λύση του προβλήματος είναι πολύ απλή και δίνεται στον Πίν. 8.1 (στήλες 3-8). Στη στήλη 3 έχουμε απλώς μετατρέψει τα ύψη βροχής της στήλης 2 σε εντάσεις διαιρώντας διά της ευκρίνειας δ , η οποία στην περίπτωσή μας είναι $\delta = 10 \text{ min} = (1/6) \text{ h}$ (άρα $1 / \delta = 6 \text{ h}^{-1}$). Στις υπόλοιπες στήλες έχουμε εφαρμόσει την εξίσωση (8.2). Για παράδειγμα ας δούμε πως υπολογίζεται η τιμή της στήλης της 1 h που αντιστοιχεί στο χρόνο $t = 31/05/1994 \text{ 19:53}$. Εδώ θα έχουμε $\delta / d = (1/6) / 1 = 1/6$ και το άθροισμα υπολογίζεται από $t - d + \delta = t - 50 \text{ min} = 31/05/1994 \text{ 19:03}$ μέχρι t . Επομένως η ζητούμενη τιμή είναι $(1/6) (0 + 0 + 0 + 0 + 1.8 + 49.8) = 8.6 \text{ mm/h}$, δηλαδή ταυτίζεται με το μέσο όρο των 6 προηγούμενων τιμών των δεκάλεπτων εντάσεων βροχής.

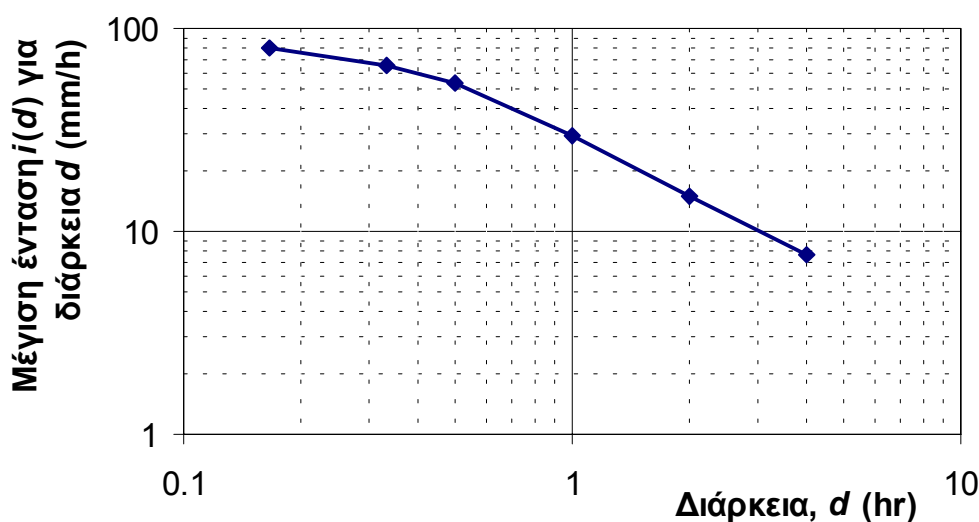
Πίν. 8.1 Πινακοποιημένο βροχογράφημα και υπολογισμός μέγιστων εντάσεων βροχής του επεισοδίου της 31/05/1994 στην Αθήνα (Δεδομένα από τον αυτόματο τηλεμετρικό μετεωρολογικό σταθμό του ΕΜΠ - Πολυτεχνειούπολη Ζωγράφου).

| Ημερομηνία-ώρα | Ύψος βροχής | Μέση ένταση βροχής $\zeta_d(t)$ σε mm/h για $d =$ | | | | | |
|------------------|-------------|---|--------|--------|--------|--------|-------|
| | | 10 min | 20 min | 30 min | 1 h | 2 h | 4 h |
| 31/05/1994 19:03 | 0 | 0.000 | | | | | |
| 31/05/1994 19:13 | 0 | 0.000 | 0.000 | | | | |
| 31/05/1994 19:23 | 0 | 0.000 | 0.000 | 0.000 | | | |
| 31/05/1994 19:33 | 0 | 0.000 | 0.000 | 0.000 | | | |
| 31/05/1994 19:43 | 0.3 | 1.800 | 0.900 | 0.600 | | | |
| 31/05/1994 19:53 | 8.3 | 49.800 | 25.800 | 17.200 | 8.600 | | |
| 31/05/1994 20:03 | 13.5 | 81.000 | 65.400 | 44.200 | 22.100 | | |
| 31/05/1994 20:13 | 5.1 | 30.600 | 55.800 | 53.800 | 27.200 | | |
| 31/05/1994 20:23 | 0.7 | 4.200 | 17.400 | 38.600 | 27.900 | | |
| 31/05/1994 20:33 | 0.9 | 5.400 | 4.800 | 13.400 | 28.800 | | |
| 31/05/1994 20:43 | 0.8 | 4.800 | 5.100 | 4.800 | 29.300 | | |
| 31/05/1994 20:53 | 0 | 0.000 | 2.400 | 3.400 | 21.000 | 14.800 | |
| 31/05/1994 21:03 | 0.3 | 1.800 | 0.900 | 2.200 | 7.800 | 14.950 | |
| 31/05/1994 21:13 | 0.1 | 0.600 | 1.200 | 0.800 | 2.800 | 15.000 | |
| 31/05/1994 21:23 | 0 | 0.000 | 0.300 | 0.800 | 2.100 | 15.000 | |
| 31/05/1994 21:33 | 0 | 0.000 | 0.000 | 0.200 | 1.200 | 15.000 | |
| 31/05/1994 21:43 | 0 | 0.000 | 0.000 | 0.000 | 0.400 | 14.850 | |
| 31/05/1994 21:53 | 0 | 0.000 | 0.000 | 0.000 | 0.400 | 10.700 | |
| 31/05/1994 22:03 | 0 | 0.000 | 0.000 | 0.000 | 0.100 | 3.950 | |
| 31/05/1994 22:13 | 0 | 0.000 | 0.000 | 0.000 | 0.000 | 1.400 | |
| 31/05/1994 22:23 | 0 | 0.000 | 0.000 | 0.000 | 0.000 | 1.050 | |
| 31/05/1994 22:33 | 0 | 0.000 | 0.000 | 0.000 | 0.000 | 0.600 | |
| 31/05/1994 22:43 | 0 | 0.000 | 0.000 | 0.000 | 0.000 | 0.200 | |
| 31/05/1994 22:53 | 0.1 | 0.600 | 0.300 | 0.200 | 0.100 | 0.250 | 7.525 |
| 31/05/1994 23:03 | 0.1 | 0.600 | 0.600 | 0.400 | 0.200 | 0.150 | 7.550 |
| 31/05/1994 23:13 | 0.1 | 0.600 | 0.600 | 0.600 | 0.300 | 0.150 | 7.575 |
| 31/05/1994 23:23 | 0 | 0.000 | 0.300 | 0.400 | 0.300 | 0.150 | 7.575 |
| 31/05/1994 23:33 | 0.1 | 0.600 | 0.300 | 0.400 | 0.400 | 0.200 | 7.600 |
| 31/05/1994 23:43 | 0.2 | 1.200 | 0.900 | 0.600 | 0.600 | 0.300 | 7.575 |
| 31/05/1994 23:53 | 0.2 | 1.200 | 1.200 | 1.000 | 0.700 | 0.400 | 5.550 |
| 01/06/1994 00:03 | 0.1 | 0.600 | 0.900 | 1.000 | 0.700 | 0.450 | 2.200 |
| 01/06/1994 00:13 | 0 | 0.000 | 0.300 | 0.600 | 0.600 | 0.450 | 0.925 |
| 01/06/1994 00:23 | 0 | 0.000 | 0.000 | 0.200 | 0.600 | 0.450 | 0.750 |
| 01/06/1994 00:33 | 0 | 0.000 | 0.000 | 0.000 | 0.500 | 0.450 | 0.525 |
| 01/06/1994 00:43 | 0 | 0.000 | 0.000 | 0.000 | 0.300 | 0.450 | 0.325 |
| 01/06/1994 00:53 | 0 | 0.000 | 0.000 | 0.000 | 0.100 | 0.400 | 0.325 |
| Μέγιστη τιμή | | 81.000 | 65.400 | 53.800 | 29.300 | 15.000 | 7.600 |

Στην τελευταία σειρά του πίνακα έχουμε εξαγάγει τις μέγιστες μέσες εντάσεις που έχουν παρατηρηθεί στη διάρκεια του επεισοδίου.

Αν ο πίνακας επεκτεινόταν στη διάρκεια ενός ολόκληρου υδρολογικού έτους, τότε τα μεγέθη που θα προέκυπταν με αυτό τον τρόπο θα ήταν τα $i(d)$ που φαίνονται στην (8.3). Για εποπτικότερη παρουσίαση των μέγιστων εντάσεων της τελευταίας στήλης του πίνακα, είτε σε βάση επεισοδίου βροχής, είτε σε βάση υδρολογικού έτους, χρησιμοποιούμε τη γραφική απεικόνιση σε λογαριθμικούς άξονες $d - i(d)$, όπως φαίνεται στο Σχ. 8.1.

Τέλος, παρατηρούμε ότι οι μέγιστες εντάσεις που υπολογίστηκαν υπακούουν στη σχέση (8.5), πράγμα που αποτελεί ένα χονδροειδή έλεγχο της ορθότητάς τους.



Σχ. 8.1 Μέγιστες εντάσεις που παρατηρήθηκαν κατά τη διάρκεια του επεισοδίου της 31/05/1994 στην Αθήνα (Δεδομένα από τον αυτόματο τηλεμετρικό μετεωρολογικό σταθμό του ΕΜΠ - Πολυτεχνειούπολη Ζωγράφου).

8.2 Συμβατική στατιστική μέθοδος εξαγωγής όμβριων καμπυλών

Μετά την κατασκευή των σειρών μέγιστων εντάσεων ακολουθεί η επεξεργασία τους με στόχο την κατάρτιση των όμβριων καμπυλών. Το πρώτο βήμα, σύμφωνα με τη συμβατική στατιστική μέθοδο, την οποία θα περιγράψουμε στην παρούσα ενότητα, είναι η επιλογή και προσαρμογή μιας συνάρτησης κατανομής σε καθεμιά από τις σειρές, η οποία αναφέρεται σε δεδομένη διάρκεια. Δεδομένου ότι πρόκειται για σειρές μέγιστων τιμών, η τυπική κατάλληλη συνάρτηση κατανομής είναι η Gumbel

μεγίστων. Λιγότερο συχνά χρησιμοποιείται και η κατανομή Log-Pearson III, καθώς και άλλες κατανομές (βλ. και εδάφιο 8.3.1 πιο κάτω).

Ουσιαστικά, η προσαρμογή μιας συνάρτησης κατανομής για μια δεδομένη διάρκεια d , ισοδυναμεί με τον καθορισμό μιας παραμετρικής σχέσης ανάμεσα στη μέγιστη ένταση και την περίοδο επαναφοράς, της μορφής $i = g_d(T)$. Η σχέση αυτή καθορίζεται πλήρως από την συνάρτηση κατανομής που έχει επιλεγεί και τις παραμέτρους της. Αν δοθεί η διάρκεια d και η περίοδος επαναφοράς T , τότε από αυτή τη σχέση μπορεί να γίνει η πρόγνωση της έντασης i . Ανάλογα με την συνάρτηση κατανομής, η σχέση αυτή μπορεί να έχει απλή αναλυτική έκφραση (όπως στην περίπτωση της κατανομής Gumbel· βλ. Εφαρμογή 8.2.1) ή πιο σύνθετη, οπότε ο υπολογισμός της έντασης γίνεται με χρήση αριθμητικών μεθόδων, πινάκων ή διαγραμμάτων.

Ωστόσο, η οικογένεια των καμπυλών $i = g_d(T)$ δεν είναι τόσο εύχρηστη, δεδομένου ότι μπορεί να χρησιμοποιηθεί μόνο εφόσον η διάρκεια d ταυτίζεται με μία από τις δεδομένες διάρκειες d_j , για τις οποίες υπάρχουν δείγματα εντάσεων. Στην πράξη, όμως, απαιτείται ο υπολογισμός της έντασης για διάρκειες διαφορετικές από τις d_j . Στην περίπτωση αυτή απαιτείται κάποια παρεμβολή ή επέκταση. Είναι προφανές ότι η παρεμβολή ή επέκταση δεν μπορεί να είναι γραμμική, αλλά πρέπει να ακολουθεί κάποιο μη γραμμικό νόμο. Αυτός ο μη γραμμικός νόμος μπορεί να προσδιοριστεί με συστηματικό τρόπο χρησιμοποιώντας τις εντάσεις που αντιστοιχούν σε όλες τις δεδομένες διάρκειες d_j για μια δεδομένη περίοδο επαναφοράς. Ο νόμος αυτός μπορεί να διατυπωθεί με τη μορφή μιας σχέσης της μορφής $i = g'_T(d)$. Κατά συνέπεια, το πρόβλημα που αντιμετωπίζουμε είναι η συναγωγή του νόμου $i = g'_T(d)$, στον οποίο η διάρκεια d είναι η ανεξάρτητη μεταβλητή, ενώ η περίοδος επαναφοράς T είναι η παράμετρος, ξεκινώντας από τον γνωστό νόμο $i = g_d(T)$, στον οποίο η διάρκεια d είναι η παράμετρος, ενώ η περίοδος επαναφοράς T είναι η ανεξάρτητη μεταβλητή.

8.2.1 Μεμονωμένες καμπύλες για κάθε διάρκεια

Στην απλούστερη περίπτωση, για μια δεδομένη περίοδο επαναφοράς T , οι όμβριες καμπύλες $i = g'_T(d)$ εκφράζονται ως υπερβολικές συναρτήσεις της διάρκειας, σε μια από τις ακόλουθες διαδοσόμενες μορφές

$$i = \frac{\omega}{(d + \theta)^\eta} \quad (8.6)$$

$$i = \frac{\omega}{d^\eta} \quad (8.7)$$

$$i = \frac{\omega}{d + \theta} \quad (8.8)$$

όπου ω , θ και η είναι παράμετροι που οι τιμές τους εξαρτώνται από την περίοδο επαναφοράς. Είναι προφανές ότι η δεύτερη και τρίτη από τις παραπάνω εκφράσεις αποτελούν ειδικές περιπτώσεις της πρώτης (για $\theta = 0$ και $\eta = 1$, αντίστοιχα). Στη βιβλιογραφία συναντάται επίσης και η εξίσωση

$$i = \frac{\omega}{d^\eta + \theta} \quad (8.9)$$

η οποία, αν και στην αναλυτική της έκφραση διαφέρει από τις παραπάνω, αριθμητικά περιγράφεται με αμελητέο σφάλμα από την (8.6), και γι' αυτό δεν μελετάται ξεχωριστά στο κείμενο αυτό.

Η εξάρτηση των παραμέτρων των (8.6) - (8.8) από την περίοδο επαναφοράς εισάγει έμμεσα την περίοδο επαναφοράς στην εκτίμηση της έντασης i .

Η μέθοδος εφαρμόζεται σε δύο ξεχωριστά στάδια. Το πρώτο στάδιο αντιστοιχεί στην προσαρμογή μιας συνάρτησης κατανομής και το δεύτερο στην εκτίμηση των παραμέτρων της (8.6) (ή των (8.7) ή (8.8)). Αναλυτικότερα, στο πρώτο στάδιο επιλέγεται, προσαρμόζεται και ελέγχεται με βάση τις γνωστές μεθόδους της στατιστικής, μια κατάλληλη συνάρτηση κατανομής (π.χ. Gumbel ή Log Pearson III), ξεχωριστά για κάθε δείγμα δεδομένης διάρκειας. Ο προσδιορισμός των παραμέτρων ω , θ και η για μια δεδομένη περίοδο επαναφοράς T γίνεται στο δεύτερο στάδιο χρησιμοποιώντας τη μέθοδο ελάχιστων τετραγώνων, όπως διευκρινίζεται στο παράδειγμα που ακολουθεί.

Εφαρμογή 8.2.1

Στον Πίν. 8.2 δίνονται οι τιμές των ετήσιων μέγιστων εντάσεων βροχής στο σταθμό Ελληνικό (αεροδρόμιο Αθήνας) της Εθνικής Μετεωρολογικής Υπηρεσίας, όπως έχουν προκύψει μετά από επεξεργασία βροχογρα-

φημάτων 30 ετών, για διάρκειες βροχής από 5 min μέχρι 24 h (με ορισμένες ελλείψεις σε μερικές διάρκειες). Ζητείται ο υπολογισμός των όμβριων καμπυλών για περιόδους επαναφοράς $T = 5$ και $T = 50$.

Πίν. 8.2 Καταγραμμένες ετήσιες μέγιστες εντάσεις βροχής στο σταθμό Ελληνικό (Πηγή Εθνική Μετεωρολογική Υπηρεσία).

| Υδρολ. Έτος | Μέγιστη ένταση βροχής i (mm/h) για διάρκεια $d =$ | | | | | | | |
|----------------|---|---------|--------|--------|--------|--------|-------|-------|
| | 5 min | 10 min | 30 min | 1 h | 2 h | 6 h | 12 h | 24 h |
| 1957-58 | 81.600 | 66.000 | 53.200 | 35.000 | 26.900 | 8.967 | 6.267 | |
| 1958-59 | 58.800 | 48.000 | 33.000 | 21.500 | 11.200 | 6.750 | 3.983 | |
| 1959-60 | 39.600 | 34.800 | 20.000 | 11.600 | 6.850 | 2.400 | 1.817 | |
| 1960-61 | 54.000 | 34.800 | 18.400 | 11.000 | 6.650 | 3.617 | 2.275 | |
| 1961-62 | 120.000 | 85.800 | 41.800 | 24.800 | 19.300 | 7.317 | 3.733 | |
| 1962-63 | 67.200 | 60.000 | 23.600 | 13.800 | 7.200 | 3.033 | 1.942 | |
| 1963-64 | 78.000 | 48.000 | 27.800 | 14.300 | 8.500 | 3.517 | 2.717 | |
| 1964-65 | 96.000 | 63.000 | 28.000 | 15.500 | 10.650 | 4.283 | 2.167 | |
| 1965-66 | 38.400 | 36.000 | 23.000 | 12.000 | 6.550 | 2.450 | 1.692 | |
| 1966-67 | 74.400 | 63.600 | 28.400 | 15.100 | 7.550 | 4.883 | 2.458 | |
| 1967-68 | 36.000 | 24.600 | 16.600 | 10.200 | 6.650 | 3.650 | 2.750 | 1.583 |
| 1968-69 | 126.000 | 69.000 | 43.200 | 26.800 | 15.150 | 5.933 | 2.967 | 1.483 |
| 1969-70 | 82.800 | 64.200 | 41.600 | 24.500 | 12.450 | 5.450 | 2.750 | 1.763 |
| 1970-71 | 42.000 | 42.000 | 25.200 | 17.700 | 8.950 | 3.700 | 3.092 | 1.546 |
| 1971-72 | 117.600 | 85.200 | 65.200 | 35.900 | 19.750 | 10.017 | 5.008 | 2.925 |
| 1972-73 | 68.400 | 49.800 | 39.400 | 33.500 | 17.750 | 6.783 | 5.267 | 2.679 |
| 1973-74 | 60.000 | 42.000 | 28.600 | 15.200 | 9.850 | 4.200 | 3.467 | 2.004 |
| 1974-75 | 48.000 | 48.000 | 30.600 | 15.900 | 8.300 | 4.267 | 2.600 | 1.300 |
| 1975-76 | 120.000 | 120.000 | 74.000 | 40.900 | 21.500 | 7.383 | 4.542 | 2.271 |
| 1976-77 | 115.200 | 87.600 | 41.400 | 23.200 | 14.900 | 6.117 | 3.300 | 1.650 |
| 1977-78 | 56.400 | 46.200 | 38.600 | 32.700 | 20.150 | 6.733 | 3.367 | 1.683 |
| 1978-79 | 78.000 | 66.600 | 47.600 | 30.000 | 19.550 | 11.933 | 6.117 | 3.371 |
| 1979-80 | 67.200 | 40.800 | 17.200 | 13.300 | 8.600 | 4.217 | 2.808 | 1.621 |
| 1980-81 | 58.800 | 56.400 | 30.400 | 19.400 | 11.100 | 5.583 | 3.267 | 1.950 |
| 1981-82 | 67.200 | 64.800 | 40.600 | 24.700 | 13.050 | 4.350 | 2.275 | 1.138 |
| 1982-83 | 141.600 | 79.800 | 49.600 | 36.200 | 22.900 | 7.633 | 4.517 | 2.292 |
| 1983-84 | 102.000 | 69.000 | 50.400 | 29.000 | 17.700 | 7.033 | 3.633 | 1.817 |
| 1984-85 | 40.800 | 31.800 | 16.400 | 12.900 | 12.150 | 9.867 | 6.000 | 3.400 |
| 1985-86 | 74.400 | 66.000 | 29.200 | 15.600 | 9.400 | 3.133 | 1.567 | 0.833 |
| 1986-87 | | | 32.200 | 29.100 | 18.550 | 9.500 | 7.242 | 3.846 |

Πριν προχωρήσουμε στην λύση του προβλήματος, παρατηρούμε ότι όλες οι τιμές του Πίν. 8.2 υπακούουν στην (8.5), πράγμα που εξασφαλίζει την απουσία χονδροειδών λαθών. Ακόμη, παρατηρούμε ότι οι ελλείψεις δεν είναι επιτρεπτό να συμπληρωθούν, π.χ. με εκτίμηση μέσω παλινδρόμησης από άλλες διάρκειες, ή από γειτονικούς

σταθμούς, και έτσι τα δείγματα θα πρέπει να θεωρηθούν ως έχουν (βλ. και εδάφιο 8.4.1).

Για λόγους που έχουν συζητηθεί στην ενότητα 6.3, περιμένουμε ότι η κατανομή μεγίστων Gumbel θα αποτελεί κατάλληλο μοντέλο για τις μέγιστες εντάσεις βροχής, πράγμα που μπορούμε να ελέγξουμε στατιστικά εκ των υστέρων. Στον Πίν. 8.3 έχουν εξαχθεί τα στατιστικά χαρακτηριστικά των δειγμάτων των επιμέρους διαρκειών και έχουν γίνει οι αναγκαίοι υπολογισμοί προσαρμογής της συνάρτησης κατανομής Gumbel με τη μέθοδο των ροπών (εξ. 6.37).

Πίν. 8.3 Στατιστικά χαρακτηριστικά και παράμετροι κατανομής Gumbel των μεγίστων εντάσεων βροχής στο σταθμό Ελληνικό.

| Παρά- μετρος | Διάρκεια $d =$ | | | | | | | |
|--------------------|----------------|--------|--------|--------|--------|--------|--------|--------|
| | 5 min | 10 min | 30 min | 1 h | 2 h | 6 h | 12 h | 24 h |
| n | 29 | 29 | 30 | 30 | 30 | 30 | 30 | 20 |
| \bar{x} | 76.221 | 58.407 | 35.173 | 22.043 | 13.325 | 5.823 | 3.520 | 2.058 |
| s_x | 29.144 | 20.318 | 13.877 | 8.889 | 5.660 | 2.433 | 1.464 | 0.786 |
| λ | 0.0440 | 0.0631 | 0.0924 | 0.1442 | 0.2265 | 0.5270 | 0.8758 | 1.6310 |
| c | 63.104 | 49.263 | 28.928 | 18.043 | 10.778 | 4.728 | 2.861 | 1.704 |
| $\psi = \lambda c$ | 2.776 | 3.108 | 2.672 | 2.602 | 2.441 | 2.492 | 2.505 | 2.779 |

Μια βασική παρατήρηση στον Πίν. 8.3 είναι ότι η αδιάστατη παράμετρος ψ που προκύπτει ως γινόμενο των παραμέτρων θέσης (c) και κλίμακας (λ) της κατανομής παραμένει περίπου σταθερή για όλες τις διάρκειες, πράγμα που θα το εκμεταλλευτούμε σε επόμενες ενότητες.

Η δοκιμή χ^2 οδηγεί σε μη απόρριψη της κατανομής Gumbel για όλες τις διάρκειες. Οπτικά, η καταλληλότητα της κατανομής διαπιστώνεται στο Σχ. 8.2 για τις μικρές διάρκειες βροχής (5 min - 1 h) και στο Σχ. 8.3 για τις μεγάλες διάρκειες βροχής (2 h - 24 h). Για την εμπειρική συνάρτηση κατανομής έχει χρησιμοποιηθεί η σχέση Gringorten $F_i = (i - 0.44) / (n + 0.12)$ (βλ. Πίν. 5.7), ενώ η ανηγμένη μεταβλητή Gumbel είναι $k = -\ln(-\ln F)$.

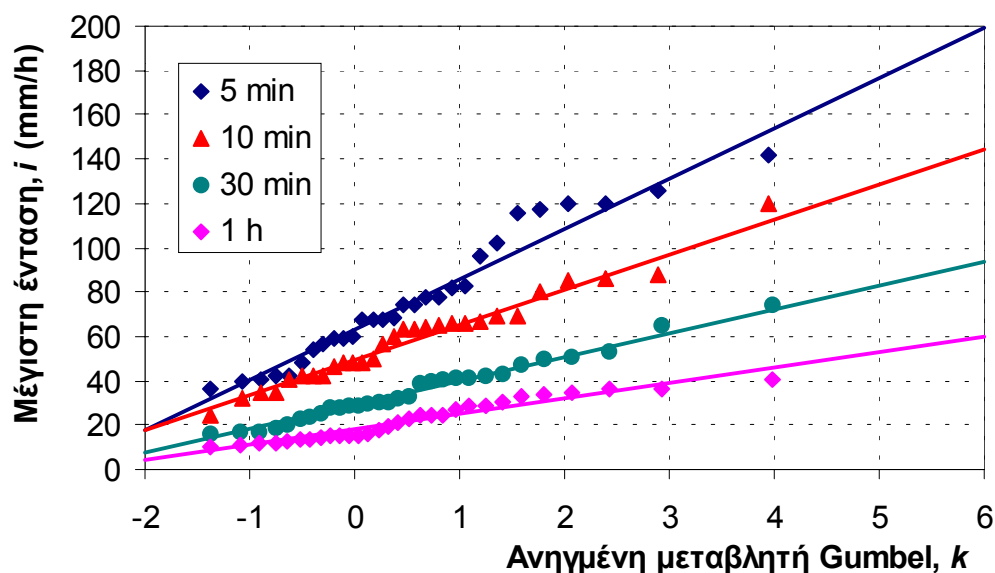
Έχοντας εκτιμήσει τις παραμέτρους των κατανομών Gumbel μπορούμε να υπολογίσουμε για καθεμιά από τις δεδομένες διάρκειες την μέγιστη ένταση που αντιστοιχεί σε κάποια περίοδο επαναφοράς T . Για το σκοπό αυτό εφαρμόζουμε την εξίσωση (6.36):

$$i = c - \ln[-\ln(1 - 1/T)] / \lambda = (1 / \lambda) \{ \psi - \ln[-\ln(1 - 1/T)] \}$$

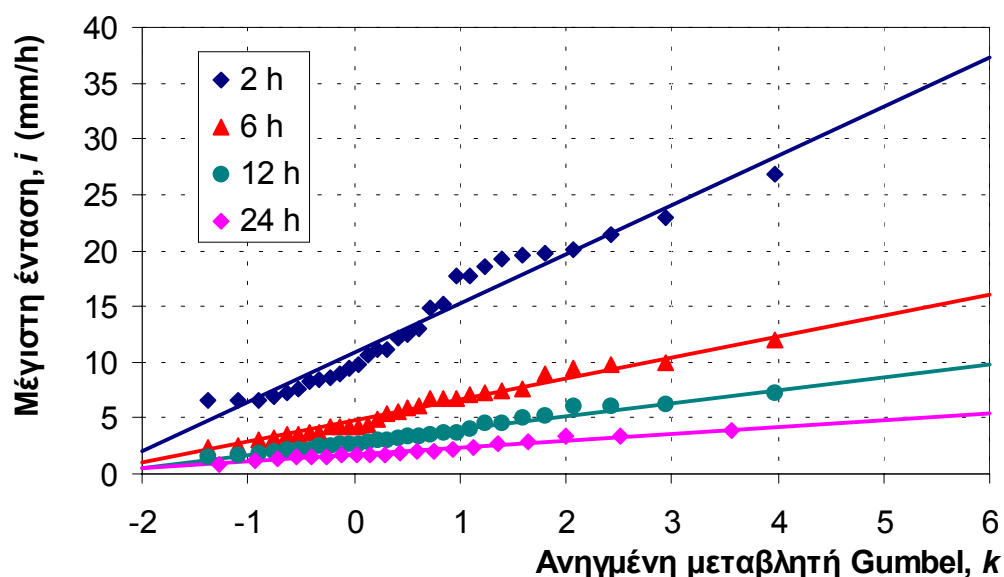
Για παράδειγμα, για $d = 5$ min και $T = 5$ έχουμε

$$i = (1 / 0.0440) \{ 2.776 - \ln[-\ln(1 - 1/5)] \} = 97.180 \text{ mm/h}$$

(Η τιμή αυτή προκύπτει και γραφικά από τη συνεχή καμπύλη του Σχ. 8.2 για $k = -\ln[-\ln(1-1/5)] = 1.5$.) Με τον ίδιο τρόπο έχουν υπολογιστεί και δίνονται στον Πίν. 8.4 οι τιμές της μέγιστης έντασης βροχής για όλες τις δεδομένες διάρκειες και για $T = 5$ και 50.



Σχ. 8.2 Εμπειρικές και θεωρητικές (Gumbel) συναρτήσεις κατανομής των μέγιστων εντάσεων βροχής στο Ελληνικό για διάρκειες βροχής από 5 min μέχρι 1 h.



Σχ. 8.3 Εμπειρικές και θεωρητικές (Gumbel) συναρτήσεις κατανομής των μέγιστων εντάσεων βροχής στο Ελληνικό για διάρκειες βροχής από 2 h μέχρι 1 h.

Πίν. 8.4 Μέγιστες εντάσεις βροχής κατά Gumbel στο σταθμό Ελληνικό για περιόδους επαναφοράς $T = 5$ και 50.

| Περίοδος επαναφ. | Διάρκεια $d =$ | | | | | | | |
|---------------------|----------------|---------|--------|--------|--------|--------|-------|-------|
| | 5 min | 10 min | 30 min | 1 h | 2 h | 6 h | 12 h | 24 h |
| 5 | 97.180 | 73.026 | 45.151 | 28.446 | 17.399 | 7.575 | 4.573 | 2.624 |
| 50 | 151.771 | 111.093 | 71.147 | 45.104 | 28.004 | 12.133 | 7.316 | 4.096 |

Στο Σχ. 8.4 έχουμε απεικονίσει σε λογαριθμικό διάγραμμα, με οριζόντιο άξονα τη διάρκεια d και κατακόρυφο τη μέγιστη ένταση i , τα σημεία που προκύπτουν από τον Πίν. 8.4. Ειδικότερα, με τετράγωνα έχουμε συμβολίσει τα σημεία που αντιστοιχούν σε περίοδο επαναφοράς $T = 5$ και με κύκλους τα σημεία που αντιστοιχούν σε περίοδο επαναφοράς $T = 50$. Παρατηρούμε ότι και οι δύο σημειοσειρές διατάσσονται σχεδόν ευθύγραμμα στο λογαριθμικό διάγραμμα, πράγμα που μας επιτρέπει να προσεγγίσουμε τη συνάρτηση $i(d)$ με μια εξίσωση δύναμης, δηλαδή να υιοθετήσουμε την εξίσωση (8.7). Όπως έχουμε δει και στο προηγούμενο κεφάλαιο, η εξίσωση δύναμης απεικονίζεται ως ευθεία σε λογαριθμικό διάγραμμα. Η εκτίμηση των παραμέτρων της (8.7) γίνεται με τον τρόπο που έχει περιγραφεί στο εδάφιο 7.3.3. Συγκεκριμένα λογαριθμίζουμε τις τιμές της διάρκειας και έντασης του Πίν. 8.4 και κατά τα γνωστά (βλ. και Εφαρμογή 7.3.3) για $T = 5$ βρίσκουμε την εξίσωση

$$\ln i = 3.182 - 0.649 \ln d$$

με αρκετά υψηλό συντελεστή συσχέτισης: $r = -0.9938$. Απολογαριθμίζοντας, βρίσκουμε την τελική όμβρια καμπύλη

$$i = 24.09 / d^{0.649} \quad (i \text{ σε mm/h, } d \text{ σε h}), \quad T = 5 \quad (\alpha)$$

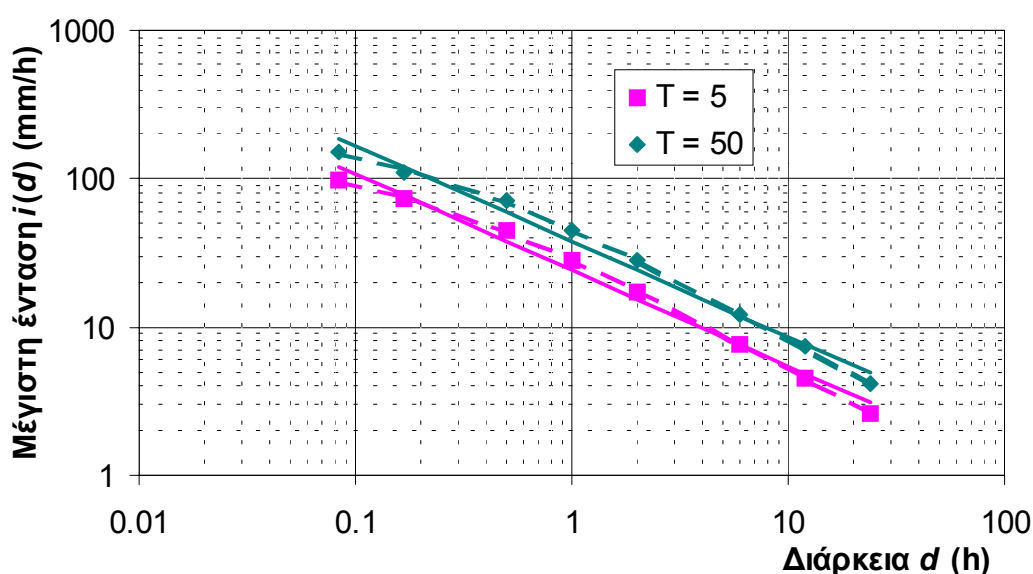
Αντίστοιχα, για $T = 50$ βρίσκουμε την εξίσωση $\ln i = 3.635 - 0.644 \ln d$, με συντελεστή συσχέτισης $r = -0.9931$. Απολογαριθμίζοντας, βρίσκουμε την τελική όμβρια καμπύλη

$$i = 37.91 / d^{0.644} \quad (i \text{ σε mm/h, } d \text{ σε h}), \quad T = 50 \quad (\beta)$$

Οι δύο παραπάνω εξισώσεις έχουν παρασταθεί στο λογαριθμικό διάγραμμα του Σχ. 8.4 με συνεχείς ευθείες καθώς στο Σχ. 8.5 (γραμμικοί άξονες) με συνεχείς καμπύλες.

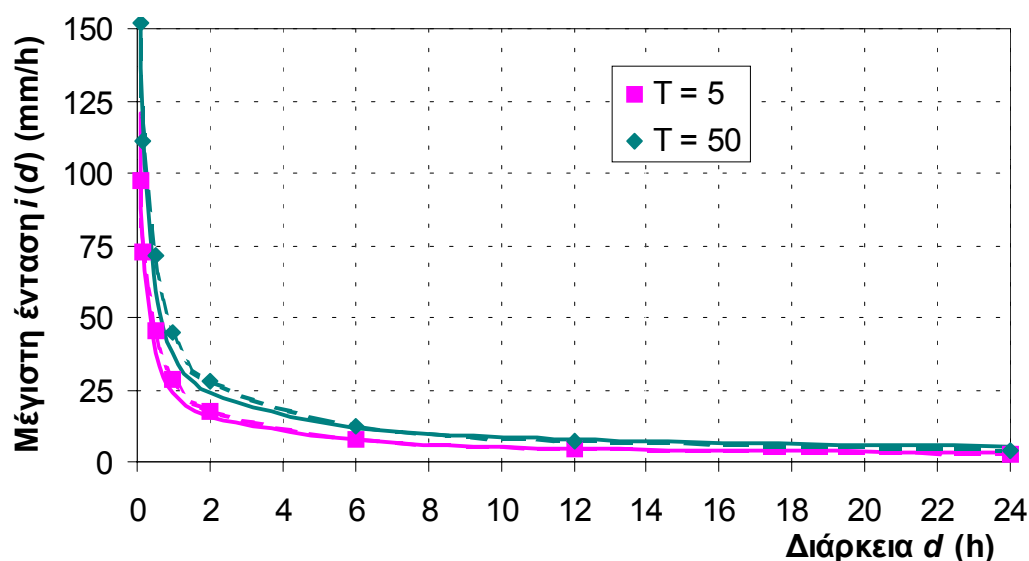
Μια σημαντική παρατήρηση στα παραπάνω αποτελέσματα είναι η ακόλουθη: Οι εκθέτες των δύο εξισώσεων ελαφρώς διαφέρουν, πράγμα που σημαίνει ότι οι κλίσεις των δύο ευθειών στο Σχ. 8.4 είναι διαφορετικές, ή αλλιώς, οι δύο ευθείες δεν είναι παράλληλες. Αυτό θεωρητικά δεν είναι σωστό, γιατί το σημείο τομής (i_τ, d_τ) των δύο

ευθειών θα αντιστοιχεί σε δύο περιόδους επαναφοράς, $T = 5$ και $T = 50$, πράγμα το οποίο είναι άτοπο. Ωστόσο, στην πράξη αυτό δεν έχει και τόσο μεγάλη σημασία αν οι εξισώσεις χρησιμοποιούνται για διάρκειες μέσα στο εύρος διακύμανσης των διαρκειών για τις οποίες υπάρχουν δεδομένα (στο συγκεκριμένο παράδειγμα 5 min - 24 h). Πράγματι, εύκολα μπορεί να βρεθεί ότι το σημείο τομής των δύο παραπάνω εξισώσεων αντιστοιχεί σε διάρκεια $d_\tau = 2.2 \times 10^{-43}$ h, δηλαδή πρακτικώς μηδενική. Παρόλα αυτά, η μαθηματική συνέπεια επιβάλλει την ταύτιση των δύο εκθετών, πράγμα που μπορεί εύκολα να επιτευχθεί με άλλες μεθόδους κατάρτισης όμβριων καμπυλών που αναπτύσσονται στη συνέχεια.



Σχ. 8.4 Όμβριες καμπύλες για περιόδους επαναφοράς $T = 5$ και 50 στο Ελληνικό, υπολογισμένες από ξεχωριστές εξισώσεις για κάθε περίοδο επαναφοράς (λογαριθμικοί άξονες). Με συνεχείς και διακεκομμένες γραμμές απεικονίζονται οι καμπύλες που αντιστοιχούν σε $\theta = 0$ και $\theta \neq 0$, αντίστοιχα.

Στο Σχ. 8.4 είναι εμφανές ότι η προσαρμογή των ευθειών στα σημεία δίνει κάποιες αποκλίσεις που δεν είναι αμελητέες. Πράγματι, η παραπάνω εξίσωση όμβριας καμπύλης (β) για $T = 50$ και για $d = 5$ min δίνει ένταση $i = 188.00$ mm/h ενώ η αντίστοιχη τιμή του Πίν. 8.4 είναι 151.77 mm/h. Για να αποφύγουμε αυτές τις αποκλίσεις είναι προτιμότερο να υιοθετήσουμε για την περιγραφή των όμβριων καμπυλών την γενικότερη εξίσωση (8.6) αντί της (8.7), δεδομένου ότι η πρώτη επιτρέπει την καλύτερη προσαρμογή στα δεδομένα, αφού περιέχει μια επιπλέον παράμετρο. Βεβαίως, αυτό ενέχει κάπως μεγαλύτερη υπολογιστική δυσκολία, επειδή η (8.6) δεν γραμμικοποιείται και ως προς τις τρεις παραμέτρους της.



Σχ. 8.5 Όμβριες καμπύλες για περιόδους επαναφοράς $T = 5$ και 50 στο Ελληνικό, υπολογισμένες από ξεχωριστές εξισώσεις για κάθε περίοδο επαναφοράς (γραμμικοί άξονες). Με συνεχείς και διακεκομμένες γραμμές απεικονίζονται οι καμπύλες που αντιστοιχούν σε $\theta = 0$ και $\theta \neq 0$, αντίστοιχα.

Η μέθοδος που προτείνουμε εδώ για την εκτίμηση των παραμέτρων της (8.6) είναι η μερική γραμμικοποίηση με λογαρίθμησή της, που δίνει

$$\ln i = \ln \omega - \eta \ln (d + \theta) \quad (\gamma)$$

Ο στόχος είναι να μεγιστοποιηθεί η απόλυτη τιμή του συντελεστή συσχέτισης μεταξύ των $\ln i$ και $\ln (d + \theta)$. Η αριθμητική ανάλυση βοηθά στην επίτευξη αυτού του στόχου. Επίσης τα σύγχρονα κοινά υπολογιστικά πακέτα (π.χ. λογιστικά φύλλα) αυτοματοποιούν τη σχετική υπολογιστική διαδικασία. Πάντως η επαναληπτική διαδικασία μπορεί να γίνει και με απλές δοκιμές, υποθέτοντας διάφορες τιμές της παραμέτρου θ και ελέγχοντας το συντελεστή συσχέτισης. Οι επαναλήψεις συνεχίζονται μέχρι να βρεθεί η μέγιστη απόλυτη τιμή του συντελεστή συσχέτισης.

Στο συγκεκριμένο παράδειγμα, οι τελικές όμβριες καμπύλες της μορφής (8.6) που μεγιστοποιούν το συντελεστή συσχέτισης είναι οι ακόλουθες:

$$i = 32.03 / (d + 0.166)^{0.785} \quad (i \text{ σε mm/h, } d \text{ σε h), \quad T = 5 \quad (\delta)$$

$$i = 51.68 / (d + 0.185)^{0.791} \quad (i \text{ σε mm/h, } d \text{ σε h), \quad T = 5 \quad (\epsilon)$$

με αντίστοιχους συντελεστές συσχέτισης -0.99988 και -0.99974 . Οι εξισώσεις αυτές έχουν παρασταθεί με διακεκομμένες γραμμές στα Σχ. 8.4 και Σχ. 8.5, όπου είναι εμφανής η καλύτερη (σε σχέση με τις (α) και (β)) προσαρμογή τους στα αρχικά σημεία.

8.2.2 Ενιαία ημιεμπειρική έκφραση όμβριων καμπυλών

Όπως συζητήσαμε και στην προηγούμενη εφαρμογή η παράμετρος η της εξίσωσης (8.7) θα πρέπει να είναι ανεξάρτητη της περιόδου επαναφοράς. Διαφορετικά θα ήταν δυνατό για δεδομένη διάρκεια d και για δύο διαφορετικές περιόδους επαναφοράς να προκύψει η ίδια τιμή της έντασης βροχής i , οπότε δεν υπάρχει αμφιμονοσήμαντη αντιστοιχία έντασης και περιόδου επαναφοράς. Στις εκφράσεις (8.6) και (8.8) δεν είναι κατ' αρχήν απαραίτητο οι παράμετροι θ και η να έχουν σταθερές τιμές, ανεξάρτητες της περιόδου επαναφοράς (δεν προκύπτει για κάθε συνδυασμό παραμέτρων η παραπάνω άτοπη συνέπεια). Ωστόσο, και πάλι η υιοθέτηση σταθερών τιμών (ανεξάρτητων από την περίοδο επαναφοράς) των παραμέτρων αυτών διευκολύνει την ανάλυση και αποκλείει την πιθανότητα υιοθέτησης εσφαλμένου συνδυασμού παραμέτρων (δηλαδή, τέτοιου που να καταργεί την αμφιμονοσήμαντη αντιστοιχία έντασης και περιόδου επαναφοράς).

Καταλήγουμε λοιπόν στην απλουστευτική παραδοχή, σύμφωνα με την οποία μόνο ο αριθμητής ω στις εκφράσεις (8.6) - (8.8) επιτρέπεται να εξαρτάται από την περίοδο επαναφοράς T , οπότε οδηγούμαστε στην παρακάτω έκφραση όμβριων καμπυλών (βλ. και Koutsoyiannis, 1994)

$$i = \frac{a(T)}{b(d)} \quad (8.10)$$

όπου οι $a(T)$ και $b(d)$ είναι συναρτήσεις της περιόδου επαναφοράς και διάρκειας, αντίστοιχα. Σε αντιστοιχία με τις (8.6) - (8.8) η $b(d)$ παίρνει μια από τις ακόλουθες μορφές:

$$b(d) = (d + \theta)^n \quad b(d) = d^n \quad b(d) = d + \theta \quad (8.11)$$

Κατά μία εμπειρική θεώρηση, η έκφραση της συνάρτησης $a(T)$ μπορεί να πάρει μια από τις ακόλουθες μορφές:

$$a(T) = \lambda T^\kappa \quad a(T) = \frac{1}{\lambda} (\psi + \ln T) \quad (8.12)$$

Συνδυάζοντας την πρώτη από τις (8.11) με την πρώτη από τις (8.14), παίρνουμε την συχνότερα χρησιμοποιούμενη ενιαία ημιεμπειρική έκφραση όμβριων καμπυλών

$$i = \frac{\lambda T^\kappa}{(d + \theta)^\eta} \quad (8.13)$$

με την οποία θα ασχοληθούμε ιδιαίτερα στο εδάφιο αυτό. Η αντιμετώπιση ενιαίων εκφράσεων που προκύπτουν με βάση τη δεύτερη από τις εξισώσεις (8.12) είναι πιο πολύπλοκη.

Όπως και στην περίπτωση των μεμονωμένων καμπυλών για κάθε διάρκεια (εδάφιο 8.2.1), και εδώ η μέθοδος περιλαμβάνει δύο ξεχωριστά στάδια. Το πρώτο στάδιο αντιστοιχεί στην προσαρμογή μιας συνάρτησης κατανομής και το δεύτερο στην εκτίμηση των παραμέτρων της (8.13).

Αναλυτικότερα, στο πρώτο στάδιο επιλέγεται, προσαρμόζεται και ελέγχεται με βάση τις γνωστές μεθόδους της στατιστικής, μια κατάλληλη συνάρτηση κατανομής (π.χ. Gumbel ή Log Pearson III), ξεχωριστά για κάθε δείγμα δεδομένης διάρκειας. Στο στάδιο αυτό δεν γίνεται συσχέτισμός με τη διάρκεια βροχής ούτε προσπάθεια ενοποιημένης αντιμετώπισης των επιμέρους δειγμάτων. Κατά συνέπεια στο στάδιο αυτό ακολουθούμε την τυπική μεθοδολογία στατιστικής ανάλυσης μιας μεταβλητής (όπως και στην περίπτωση των μεμονωμένων καμπυλών του εδαφίου 8.2.1).

Στο δεύτερο στάδιο εφαρμόζουμε τη μέθοδο ελάχιστων τετραγώνων για την εκτίμηση των παραμέτρων της, χρησιμοποιώντας όχι τις αρχικές μετρημένες εντάσεις βροχής, αλλά αυτές που προκύπτουν από τις συναρτήσεις κατανομής του πρώτου σταδίου για ένα καθορισμένο διάστημα μεταβολής της περιόδου επαναφοράς. Η μέθοδος ελάχιστων τετραγώνων είναι εύκολο να εφαρμοστεί δεδομένου ότι η (8.13) γραμμικοποιείται με λογαρίθμηση και γίνεται

$$\ln i = \ln \lambda + \kappa \ln T - \eta \ln(d + \theta) \quad (8.14)$$

πράγμα που επιτρέπει τον άμεσο υπολογισμό των παραμέτρων λ , κ και η , ενώ προϋποθέτει επαναλήψεις ως προς την παράμετρο θ .

Συγκεκριμένα η εφαρμογή του δεύτερου σταδίου περιλαμβάνει τα ακόλουθα βήματα:

1. Επιλέγουμε κάποιες περιόδους επαναφοράς μέσα στο επιθυμητό εύρος μεταβολής.
2. Από τις συναρτήσεις κατανομής του πρώτου σταδίου υπολογίζουμε για κάθε δεδομένη διάρκεια d_j και για κάθε τιμή της περιόδου επαναφοράς T (του βήματος 1) τη θεωρητική τιμή της έντασης βροχής i .
3. Υποθέτουμε μια δοκιμαστική τιμή της παραμέτρου θ .
4. Προσαρμόζουμε τη σχέση (8.14) στις θεωρητικές εντάσεις βροχής που προέκυψαν από το βήμα 3, υπολογίζοντας έτσι τις τιμές των παραμέτρων λ , κ και η .
5. Υπολογίζουμε το συντελεστή προσδιορισμού που αντιστοιχεί στην (8.14)
6. Επαναλαμβάνουμε τα βήματα 3 έως 5 με διαφορετικές τιμές της παραμέτρου θ μέχρι να επιτύχουμε τη μέγιστη τιμή του συντελεστή προσδιορισμού.

Η μέθοδος της ενιαίας ημιεμπειρικής έκφρασης των όμβριων καμπυλών έχει το πλεονέκτημα της διατύπωσης της σχέσης μεταξύ της έντασης, της διάρκειας και της περιόδου επαναφοράς της βροχόπτωσης με μια μοναδική εξίσωση. Ωστόσο, έχει ένα σοβαρό μειονέκτημα: την υποκατάσταση της πραγματικής συνάρτησης κατανομής της έντασης με μια εμπειρική εξίσωση δύναμης. Για το λόγο αυτό οι παράμετροι κ και λ αυτής της εξίσωσης δεν μπορεί να θεωρηθούν σταθερές, αλλά εξαρτώνται από το διάστημα μεταβολής της περιόδου επαναφοράς T (βλ. και εφαρμογή που ακολουθεί). Όπως θα δούμε στην επόμενη ενότητα, υπάρχουν τρόποι να αποφύγουμε αυτό το μειονέκτημα και να οδηγηθούμε σε ενιαία έκφραση των όμβριων καμπυλών με πιο συνεπή τρόπο και με σταθερές παραμέτρους.

Εφαρμογή 8.2.2

Να καταρτιστεί ενιαία έκφραση όμβριων καμπυλών της μορφής (8.13) με βάση τα δεδομένα του σταθμού Ελληνικό, που δίνονται στον Πίν. 8.2 (βλ. Εφαρμογή 8.2.1).

Στην Εφαρμογή 8.2.1 έχουμε δει ότι οι μέγιστες εντάσεις ακολουθούν κατανομές Gumbel και έχουμε εξαγάγει τις παραμέτρους λ και ψ αυτών των κατανομών για τις διάφορες διάρκειες στον Πίν. 8.3. Έτσι, για κάθε διάρκεια και για δεδομένη περίοδο επαναφοράς T μπορούμε να βρούμε τη μέγιστη ένταση από την εξίσωση

$$i = (1/\lambda) \{ \psi - \ln[-\ln(1 - 1/T)] \}$$

Ας υποθέσουμε ότι ενδιαφερόμαστε για το διάστημα διακύμανσης της περιόδου επαναφοράς από 2 μέχρι 50 χρόνια. Διαλέγουμε στο διάστημα αυτό τις χαρακτηριστικές τιμές $T = 2, 5, 10, 20$ και 50 . Για κάθε συνδυασμό αυτών των τιμών και των δεδομένων διάρκειών υπολογίζουμε από την πιο πάνω εξίσωση την ένταση i και δίνουμε στον Πίν. 8.5 το λογάριθμό της. Στον ίδιο πίνακα δίνουμε και τους λογαρίθμους των μεγεθών T και $d + \theta$, υποθέτοντας κατ' αρχήν ότι $\theta = 0$. Με οδηγό την εξίσωση (8.14), κάνουμε πολλαπλή γραμμική παλινδρόμηση ανάμεσα στα μεγέθη

$$y = \ln i, x_1 = \ln T, x_2 = \ln(d + \theta)$$

εφαρμόζοντας τη μέθοδο που περιγράφεται στο εδάφιο 7.3.1. Προκύπτει τελικά η ακόλουθη εξίσωση

$$i = 15.755 T^{0.237} / d^{0.648} \quad (i \text{ σε mm/h, } d \text{ σε h}) \quad (\alpha)$$

με συντελεστή προσδιορισμού $r^2 = 0.9865$. Για σύγκριση με τις εξισώσεις που έχουν εξαχθεί στην Εφαρμογή 8.2.1 θέτουμε στην παραπάνω $T = 2$ και $T = 50$, οπότε προκύπτει

$$i = 23.06 / d^{0.648} \quad (i \text{ σε mm/h, } d \text{ σε h}), \quad T = 5 \quad (\beta)$$

$$i = 39.79 / d^{0.648} \quad (i \text{ σε mm/h, } d \text{ σε h}), \quad T = 50 \quad (\gamma)$$

Παρατηρούμε ότι οι παραπάνω εξισώσεις, οι οποίες παριστάνονται με συνεχείς γραμμές στο Σχ. 8.6, δεν απέχουν από αυτές της Εφαρμογής 8.2.1.

Επαναλαμβάνουμε την παραπάνω διαδικασία για διάφορες δοκιμαστικές δοκιμές του θ , κάνοντας την κατάλληλη αλλαγή στη στήλη $\ln(d + \theta)$ του Πίν. 8.5 και υπολογίζοντας εκ νέου τους συντελεστές της πολλαπλής γραμμικής παλινδρόμησης, καθώς και το συντελεστή προσδιορισμού. Καταλήγουμε ότι τη μέγιστη τιμή του συντελεστή προσδιορισμού ($r^2 = 0.9984$) δίνει η τιμή $\theta = 0.170$. Η αντίστοιχη ενιαία έκφραση είναι η ακόλουθη:

Πίν. 8.5 Βοηθητικός πίνακας υπολογισμών για την πολλαπλή γραμμική παλινδρόμηση των λογαρίθμων των μεγεθών T , $d + \theta$ και i με στόχο την εξαγωγή ενιαίας έκφρασης όμβριων καμπυλών ($2 \leq T \leq 50$, $\theta = 0$).

| T | d | $\ln T$ | $\ln (d + \theta)$ | $\ln i$ | T | d | $\ln T$ | $\ln (d + \theta)$ | $\ln i$ | |
|-----|------|---------|--------------------|---------|-----|-------|---------|--------------------|---------|-------|
| 2 | 5/60 | 0.693 | -2.485 | 4.269 | 10 | 2 | 2.303 | 0.693 | 3.031 | |
| | 1/4 | 0.693 | -1.792 | 4.008 | | 6 | 2.303 | 1.792 | 2.197 | |
| | 1/2 | 0.693 | -0.693 | 3.493 | | 12 | 2.303 | 2.485 | 1.692 | |
| | 1 | 0.693 | 0.000 | 3.025 | | 24 | 2.303 | 3.178 | 1.126 | |
| | 2 | 0.693 | 0.693 | 2.517 | | 20 | 5/60 | 2.996 | -2.485 | 4.872 |
| | 6 | 0.693 | 1.792 | 1.691 | | | 1/4 | 2.996 | -1.792 | 4.568 |
| | 12 | 0.693 | 2.485 | 1.187 | | | 1/2 | 2.996 | -0.693 | 4.112 |
| | 24 | 0.693 | 3.178 | 0.657 | | | 1 | 2.996 | 0.000 | 3.654 |
| 5 | 5/60 | 1.609 | -2.485 | 4.577 | 2 | 2.996 | 0.693 | 3.173 | | |
| | 1/4 | 1.609 | -1.792 | 4.291 | 6 | 2.996 | 1.792 | 2.338 | | |
| | 1/2 | 1.609 | -0.693 | 3.810 | 12 | 2.996 | 2.485 | 1.833 | | |
| | 1 | 1.609 | 0.000 | 3.348 | 24 | 2.996 | 3.178 | 1.260 | | |
| | 2 | 1.609 | 0.693 | 2.856 | 50 | 5/60 | 3.912 | -2.485 | 5.022 | |
| | 6 | 1.609 | 1.792 | 2.025 | | 1/4 | 3.912 | -1.792 | 4.710 | |
| | 12 | 1.609 | 2.485 | 1.520 | | 1/2 | 3.912 | -0.693 | 4.265 | |
| | 24 | 1.609 | 3.178 | 0.965 | | 1 | 3.912 | 0.000 | 3.809 | |
| 10 | 5/60 | 2.303 | -2.485 | 4.738 | 2 | 3.912 | 0.693 | 3.332 | | |
| | 1/4 | 2.303 | -1.792 | 4.442 | 6 | 3.912 | 1.792 | 2.496 | | |
| | 1/2 | 2.303 | -0.693 | 3.975 | 12 | 3.912 | 2.485 | 1.990 | | |
| | 1 | 2.303 | 0.000 | 3.516 | 24 | 3.912 | 3.178 | 1.410 | | |

$$i = 21.064 T^{0.237} / (d + 0.170)^{0.785} \quad (i \text{ σε mm/h, } d \text{ σε h}) \quad (\delta)$$

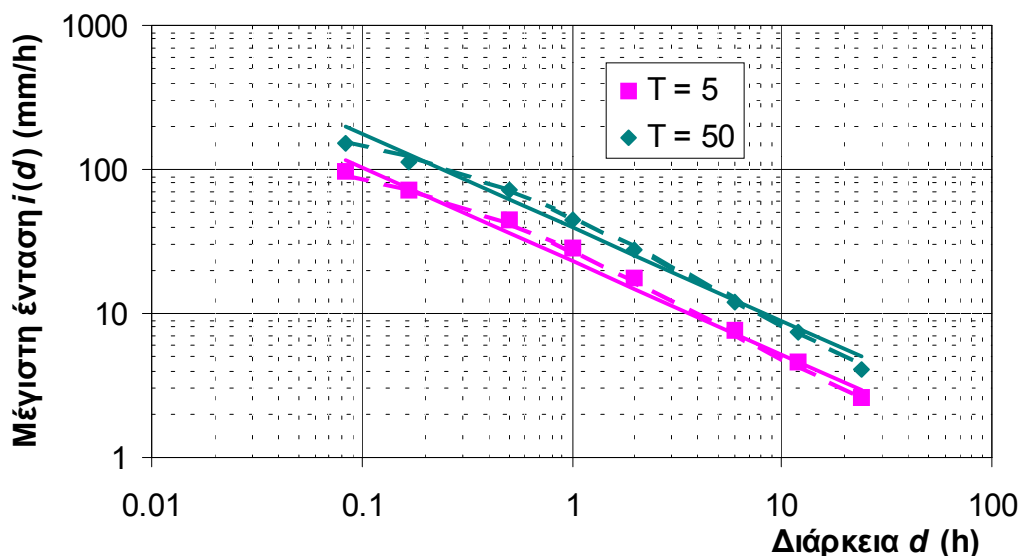
Η παραπάνω για $T = 2$ και $T = 50$ δίνει

$$i = 30.83 / (d + 0.170)^{0.785} \quad (i \text{ σε mm/h, } d \text{ σε h}), \quad T = 5 \quad (\epsilon)$$

$$i = 53.20 / (d + 0.170)^{0.785} \quad (i \text{ σε mm/h, } d \text{ σε h}), \quad T = 50 \quad (\sigma\tau)$$

Και πάλι διαπιστώνουμε ότι οι παραπάνω εξισώσεις, οι οποίες παριστάνονται με διακεκομμένες γραμμές στο Σχ. 8.6, δεν απέχουν από αυτές της Εφαρμογής 8.2.1.

Για λόγους διερεύνησης έχουμε υπολογίσει ενιαίες ημιεμπειρικές εκφράσεις της μορφής (8.13) και για ένα άλλο διάστημα μεταβολής της περιόδου επαναφοράς: το διάστημα $100 \leq T \leq 10\,000$ με χαρακτηριστικές τιμές $T = 100, 200, 500, 1000, 2\,000, 5\,000$ και $10\,000$. Για $\theta = 0$ προκύπτει τελικά η εξίσωση



Σχ. 8.6 Όμβριες καμπύλες για περιόδους επαναφοράς $T = 5$ και 50 στο Ελληνικό, υπολογισμένες από ενιαία ημιεμπειρική έκφραση (8.13) (λογαριθμικοί άξονες). Με συνεχείς και διακεκομμένες γραμμές απεικονίζονται οι καμπύλες που αντιστοιχούν σε $\theta = 0$ και $\theta \neq 0$, αντίστοιχα.

$$i = 26.15 T^{0.106} / d^{0.642} \quad (i \text{ σε mm/h, } d \text{ σε h}) \quad (\zeta)$$

με συντελεστή προσδιορισμού $r^2 = 0.9850$. Το μέγιστο συντελεστή προσδιορισμού ($r^2 = 0.9992$) δίνει η τιμή $\theta = 0.188$, με τελική έκφραση

$$i = 36.14 T^{0.106} / (d + 0.196)^{0.794} \quad (i \text{ σε mm/h, } d \text{ σε h}) \quad (\eta)$$

Παρατηρούμε ότι οι εξισώσεις (ζ) και (η) διαφέρουν σημαντικά από τις αντίστοιχές τους (α) και (δ), ιδιαίτερα ως προς τους συντελεστές κ και λ . Αυτό είναι, όπως έχουμε πει και παραπάνω, ένα σοβαρό μειονέκτημα της μεθόδου.

8.3 Συνεπείς μέθοδοι εξαγωγής όμβριων καμπυλών

Η γενική σχέση (8.10) που διατυπώθηκε στο προηγούμενο εδάφιο αποτελεί χωρίς αμφιβολία μια ορθή βάση για την έκφραση των όμβριων καμπυλών. Ωστόσο, όπως είδαμε, στη μεθοδολογία που αναπτύχθηκε παραπάνω τέθηκαν ορισμένες παραδοχές, οι οποίες αποτελούν ασυνέπειες ως προς την φύση του προβλήματος που μελετάμε. Η πρώτη ασυνέ-

πεια είναι η υιοθέτηση μιας εμπειρικής έκφρασης της συνάρτησης $a(T)$, ενώ η έκφραση κανονικά συνδέεται με τη συνάρτηση κατανομής της μέγιστης έντασης βροχής. Επακόλουθη αυτής της ασυνέπειας είναι η μη σταθερότητα των παραμέτρων της ενιαίας έκφρασης. Η δεύτερη ασυνέπεια συνδέεται με την εκτίμηση των παραμέτρων με πολλαπλή παλινδρόμηση, στην οποία συσχετίζεται το μέγεθος i , το οποίο αντιστοιχεί σε μια τυχαία μεταβλητή, με τα μεγέθη d και T , τα οποία δεν αποτελούν τυχαίες μεταβλητές. Ασφαλώς, αυτού του είδους η συσχέτιση δεν έχει θεωρητική συνέπεια. Στην παρούσα ενότητα θα δούμε ότι τα παραπάνω προβλήματα μπορούν να παρακαμφθούν χρησιμοποιώντας μια διαφορετική προσέγγιση, η οποία μπορεί να οδηγήσει σε ενιαία σχέση με σταθερές παραμέτρους. Η προσέγγιση αυτή αρχικά θεμελιώθηκε από τον Koutsoyiannis (1993) και διερευνήθηκε από τους Δέα (1994), Κοζώνη (1995) και Κουτσογιάννη και Μανέτα (1995).

Από τη γενική σχέση (8.10) συνάγουμε εύκολα ότι η συνάρτηση $a(T)$ είναι πρακτικώς η αντίστροφη της συνάρτησης κατανομής της τυχαίας μεταβλητής $Y = I b(d)$, δηλαδή $a(T) \equiv y_T$, όπου y_T είναι το ποσοστημόριο της κατανομής της Y που αντιστοιχεί σε περίοδο επαναφοράς T . Κατά συνέπεια η $a(T)$ εξαρτάται άμεσα από τον τύπο της συνάρτησης κατανομής που υιοθετείται για την ένταση βροχής I . Αφού η διάρκεια d στην ανάλυσή μας δεν αποτελεί τυχαία μεταβλητή, αλλά παράμετρο, το μέγεθος $b(d)$ δεν επηρεάζει τον τύπο της συνάρτησης κατανομής, παρά μόνο την παράμετρο κλίμακας της συνάρτησης κατανομής. Στην επόμενη ενότητα εξετάζονται οι συνηθέστεροι τύποι συνάρτησης κατανομής που είναι κατάλληλες για την πιθανοτική περιγραφή της έντασης I και υπολογίζεται για κάθε τύπο η αντίστοιχη έκφραση της συνάρτησης $a(T)$.

8.3.1 Εναλλακτικοί τύποι συνάρτησης κατανομής

Κατανομή Gumbel

Από τη φύση της η κατανομή μεγίστων τύπου I (Gumbel) είναι κατάλληλη για την περιγραφή μεγίστων μεγεθών, γι' αυτό και είναι η πιο διαδεδομένη κατανομή για τη μοντελοποίηση των μεγίστων εντάσεων βροχής. Η μαθηματική της έκφραση είναι

$$F_X(x) = \exp(-e^{-\lambda x + \psi}) \quad (8.15)$$

όπου λ και ψ είναι οι παράμετροι κλίμακας και θέσης, αντίστοιχα. Ας σημειωθεί ότι η εξίσωση (8.15) έχει γραφεί ελαφρώς διαφοροποιημένη από την καθιερωμένη της μορφή (βλ. Πίν. 6.9) και συγκεκριμένα στον ανώτερο εκθέτη αντί της παράστασης $-\lambda(x - c)$ έχει γραφεί $-\lambda x + \psi$, όπου $\psi = \lambda c$. Η μεταβλητή X συμβολίζει είτε την ένταση βροχής I είτε το γινόμενο $Y = I b(d)$ (το μόνο που αλλάζει είναι η τιμή της παραμέτρου κλίμακας). Στην τελευταία περίπτωση, παίρνοντας υπόψη ότι

$$F_Y(y) = F_I(i) = 1 - \frac{1}{T} \quad (8.16)$$

και επιλύοντας την (8.15) ως προς y , παίρνουμε

$$y_T \equiv a(T) = \frac{1}{\lambda} \left\{ \psi - \ln \left[-\ln \left(1 - \frac{1}{T} \right) \right] \right\} \quad (8.17)$$

όπου οι παράμετροι λ και ψ αναφέρονται στη μεταβλητή Y .

Για την εκτίμηση των παραμέτρων λ και ψ από οποιοδήποτε δείγμα μπορεί να χρησιμοποιηθούν οι μέθοδοι που συζητούνται στο εδάφιο 6.3.1. Για παράδειγμα, η μέθοδος ροπών δίνει

$$\lambda = 1 / (0.78 s_X) \quad \psi = \lambda \bar{x} - 0.577 \quad (8.18)$$

Κατανομή γάμα

Μια άλλη κατανομή που έχει χρησιμοποιηθεί, αν και όχι τόσο συχνά, για την περιγραφή μέγιστων εντάσεων βροχής είναι η κατανομή γάμα (Pearson III) δύο παραμέτρων, η οποία περιγράφεται από τις εξισώσεις του Πίν. 6.6. Λόγω της πολύπλοκης αναλυτικής έκφρασης της συνάρτησης κατανομής δεν είναι κατ' αρχήν δυνατή η ρητή έκφραση της συνάρτησης $a(T)$ για τη συγκεκριμένη κατανομή. Ωστόσο, χρησιμοποιώντας την προσέγγιση που έχει αναπτυχθεί στο Παράρτημα 6.B μπορούμε να γράψουμε

$$a(T) \approx \frac{\mu}{\lambda \alpha} \left(1 - \frac{1}{T} \right)^\alpha + \frac{\nu}{\lambda \beta} \left[\xi - \left(\frac{1}{T} \right)^\beta \right] \kappa \neq 1 \quad (8.19)$$

Για την περίπτωση που $\kappa = 1$, η κατανομή μεταπίπτει στην απλούστερη εκθετική κατανομή (βλ. αντίστοιχο εδάφιο πιο κάτω). Στην παραπάνω

εξίσωση, το λ είναι η παράμετρος κλίμακας της κατανομής, ενώ τα μ , ν , α , β και ξ είναι σταθεροί συντελεστές, οι τιμές των οποίων εξαρτώνται από την παράμετρο σχήματος κ της κατανομής και δίνονται από τις εξισώσεις (6.65)-(6.69). Η απλούστερη μέθοδος εκτίμησης των παραμέτρων κ και λ της κατανομής γάμα από το δείγμα είναι η μέθοδος των ροπών (εξισώσεις (6.21)).

Κατανομή Log Pearson III

Μια άλλη κατανομή που έχει χρησιμοποιηθεί για την περιγραφή μέγιστων εντάσεων βροχής, ιδίως όταν εμφανίζονται πολύ υψηλές εντάσεις για μεγάλες περιόδους επαναφοράς, είναι η κατανομή Log Pearson III, οι εξισώσεις της οποίας δίνονται στον Πίν. 6.8. Η κατανομή διαθέτει τρεις παραμέτρους: την παράμετρο κλίμακας c , και τις παραμέτρους σχήματος κ και λ . Λόγω της πολύπλοκης αναλυτικής έκφρασης της συνάρτησης κατανομής δεν είναι κατ' αρχήν δυνατή η ρητή έκφραση της συνάρτησης $a(T)$ για τη συγκεκριμένη κατανομή. Ωστόσο, χρησιμοποιώντας την ίδια προσέγγιση όπως στην κατανομή γάμα μπορούμε να γράψουμε

$$a(T) = \exp \left\{ c + \frac{\mu}{\lambda\alpha} \left(1 - \frac{1}{T} \right)^\alpha + \frac{\nu}{\lambda\beta} \left[\xi - \left(\frac{1}{T} \right)^\beta \right] \right\} \kappa \neq 1 \quad (8.20)$$

Για την περίπτωση που $\kappa = 1$, η κατανομή μεταπίπτει στην απλούστερη κατανομή Pareto (βλ. αντίστοιχο εδάφιο πιο κάτω). Στην παραπάνω εξίσωση, τα μ , ν , α , β και ξ είναι σταθεροί συντελεστές, οι τιμές των οποίων εξαρτώνται από την παράμετρο σχήματος κ και δίνονται και πάλι από τις εξισώσεις (6.65)-(6.69). Οι εκτιμήσεις των παραμέτρων κ , λ και c της κατανομής log Pearson III με την έμμεση μέθοδο των ροπών δίνονται από τις εξισώσεις (6.31).

Κατανομή Pareto

Η συνάρτηση κατανομής Pareto δύο παραμέτρων δίνεται από την ακόλουθη εξίσωση, η οποία προκύπτει ως ειδική περίπτωση των εξισώσεων του Πίν. 6.15 (αν τεθεί στη θέση του κ το $-\kappa$ για θεωρηθεί $c = \lambda = \alpha / \kappa$)

$$F_X(x) = 1 - \left(\frac{\lambda}{x} \right)^{1/\kappa} \quad x \geq \lambda \quad (8.21)$$

όπου κ και λ είναι οι παράμετροι σχήματος και κλίμακας, αντίστοιχα, της κατανομής. Άμεσα προκύπτει από την (8.21) ότι

$$a(T) = \lambda T^\kappa \quad (8.22)$$

Η μέθοδος ροπών δίνει τις ακόλουθες εκτιμήτριες των παραμέτρων κ και λ της κατανομής:

$$\kappa = \sqrt{\hat{C}_{v_x}^2 \left(1 + \hat{C}_{v_x}^2\right) - \hat{C}_{v_x}^2} \quad \lambda = \bar{x} (1 - \kappa) \quad (8.23)$$

όπου \bar{x} και $\hat{C}_{v_x}^2$ η δειγματική μέση τιμή και ο δειγματικός συντελεστής μεταβλητότητας, αντίστοιχα.

Παρατηρούμε ότι η εξίσωση (8.22) ταυτίζεται με την πρώτη από τις εμπειρικές εκφράσεις της (8.12), πράγμα που σημαίνει ότι η υιοθέτηση της εν λόγω εμπειρικής έκφρασης ουσιαστικά ισοδυναμεί με υιοθέτηση της κατανομής Pareto για τις μέγιστες εντάσεις βροχής. Ωστόσο, η εμπειρία δείχνει ότι η εν λόγω κατανομή δεν είναι κατάλληλη ως μοντέλο για τις μέγιστες εντάσεις βροχής. Η διαδεδομένη χρήση της (8.22) για την έκφραση όμβριων καμπυλών οφείλεται αποκλειστικά στην απλή μαθηματική έκφραση της και όχι στην καταλληλότητα της κατανομής Pareto. Έτσι δεν συνιστούμε την κατανομή Pareto για άμεση χρήση, παρά μόνο μετά από προσεκτική εξέταση. Παρόλα αυτά η εξίσωση (8.22) μπορεί να χρησιμοποιείται ως προσέγγιση άλλων συνθετότερων εκφράσεων για ένα ορισμένο διάστημα μεταβολής της περιόδου επαναφοράς T . Σε αυτή όμως την περίπτωση δεν είναι σκόπιμη η χρήση των εξισώσεων (8.23) για την εκτίμηση των παραμέτρων κ και λ , αλλά πρέπει να χρησιμοποιείται η ημιεμπειρική μέθοδος που περιγράφηκε στο εδάφιο 8.2.2.

Εκθετική κατανομή

Η εκθετική συνάρτηση κατανομής δύο παραμέτρων δίνεται από την εξίσωση

$$F_X(x) = 1 - e^{-\lambda x + \psi} \quad x \geq \psi / \lambda \quad (8.24)$$

όπου ψ και λ είναι οι παράμετροι θέσης και κλίμακας, αντίστοιχα, της κατανομής. Άμεσα προκύπτει από την (8.21) ότι

$$a(T) = \frac{1}{\lambda} (\psi + \ln T) \quad (8.25)$$

Ας σημειωθεί ότι η εξίσωση (8.24) έχει γραφεί ελαφρώς διαφοροποιημένη από την καθιερωμένη μορφή της (βλ. Πίν. 6.5) και συγκεκριμένα στον εκθέτη αντί της παράστασης $-\lambda(x - c)$ έχει γραφεί $-\lambda x + \psi$, όπου $\psi = \lambda c$. Η μέθοδος ροπών δίνει τις ακόλουθες εκτιμήτριες των παραμέτρων ψ και λ της κατανομής:

$$\lambda = \frac{1}{s_x} \quad \psi = \lambda \bar{x} - 1 \quad (8.26)$$

Όπως συμβαίνει και με την αντίστοιχη έκφραση της κατανομής Pareto, η λογαριθμική έκφραση (8.25) είναι επίσης πολύ συνηθισμένη για την έκφραση όμβριων καμπυλών (βλ. και εξίσωση (8.12)). Αυτό οφείλεται σε δύο λόγους: (α) στην απλή μαθηματική έκφραση της (8.25) και (β) στην καταλληλότητα της εκθετικής κατανομής για την περιγραφή των σειρών υπεράνω κατωφλίου (βλ. και εδάφιο 8.4.3).

Για την περίπτωση των σειρών ετήσιων μεγίστων ισχύουν και για την εκθετική κατανομή όσες παρατηρήσεις γράφονται για την κατανομή Pareto ως προς τις επιφυλάξεις άμεσης χρήσης της και ως προς τη σύσταση του έμμεσου τρόπου υπολογισμού των παραμέτρων της ψ και λ αντί της χρήσης των εξισώσεων (8.26). Ωστόσο, στην περίπτωση που οι μέγιστες εντάσεις βροχής ακολουθούν κατανομή Gumbel, μπορεί άμεσα να χρησιμοποιηθεί η έκφραση (8.25) ως προσέγγιση της (8.17) για μεγάλες τιμές της περιόδου επαναφοράς T , με παραμέτρους ψ και λ , όπως υπολογίζονται από την (8.18). Πράγματι, για μεγάλες τιμές του T (π.χ. $T \geq 50$) ισχύει $\ln [1 - (1/T)] = -(1/T) - (1/T)^2 - \dots \approx -(1/T)$, οπότε η (8.17) μεταπίπτει στην (8.25).

8.3.2 Εκτίμηση παραμέτρων

Το σύνολο των παραμέτρων των όμβριων καμπυλών περιλαμβάνει αφενός τις παραμέτρους θ και η της συνάρτησης διάρκειας $b(d)$, και αφετέρου τις παραμέτρους της κατανομής που χρησιμοποιείται, οι οποίες ταυτίζονται με τις παραμέτρους της συνάρτησης $a(T)$. Οι τελευταίες εξαρτώνται από τη συγκεκριμένη κατανομή που κάθε φορά υιοθετείται. Στην ενότητα αυτή δίνουμε δύο διαφορετικές μεθόδους για την εκτίμηση

του συνόλου των παραμέτρων των όμβριων καμπυλών. Σε κάθε περίπτωση υποτίθεται ότι διατίθενται k δείγματα εντάσεων βροχής που το καθένα αναφέρεται σε ξεχωριστή διάρκεια βροχής. Συμβολίζουμε με d_1, \dots, d_k τις διάρκειες βροχής των δειγμάτων αυτών και με n_1, \dots, n_k τα αντίστοιχα μεγέθη τους.

Εκτίμηση με ενοποίηση διαρκειών

Η μέθοδος εκτίμησης με ενοποίηση διαρκειών υπολογίζει το σύνολο των παραμέτρων των όμβριων καμπυλών σε δύο στάδια. Στο πρώτο στάδιο υπολογίζονται οι παράμετροι της συνάρτησης $b(d)$ και στο δεύτερο αυτές της $a(T)$.

Από την (8.10) προκύπτει άμεσα ότι η τυχαία μεταβλητή $Y = I b(d)$ έχει συνάρτηση κατανομής ανεξάρτητη της διάρκειας d , η οποία καθορίζεται πλήρως από τη συνάρτηση $a(T)$. Πρέπει λοιπόν οι παράμετροι θ και η να υπολογιστούν έτσι ώστε να ικανοποιούν αυτή τη συνθήκη.

Αν υποθέσουμε ότι είναι γνωστές οι τιμές των παραμέτρων θ και η , τότε μπορούν να υπολογιστούν οι τιμές $y_{jl} = i_{jl} b(d_j)$. Ενοποιώντας όλα τα δείγματα που περιέχουν τις τιμές y_{jl} αποκτούμε ένα συνολικό δείγμα μεγέθους

$$m = \sum_{j=1}^k n_j \quad (8.27)$$

Με βάση το δείγμα αυτό, καταταγμένο σε φθίνουσα σειρά, μπορούμε να αντιστοιχίσουμε αύξοντες αριθμούς ή βαθμούς (ranks) r_{jl} σε όλες τις m τιμές y_{jl} (Για την περίπτωση που έχουμε ταυτόσημες τιμές y_{jl} χρησιμοποιούμε το μέσο όρο των αντίστοιχων βαθμών). Επανερχόμενοι στα αρχικά επιμέρους δείγματα των ξεχωριστών διαρκειών υπολογίζουμε για κάθε τιμή το μέσο βαθμό

$$\bar{r}_j = \frac{1}{n_j} \sum_{l=1}^{n_j} r_{jl} \quad (8.28)$$

Αν όλα τα επιμέρους δείγματα έχουν την ίδια κατανομή τότε κάθε \bar{r}_j θα πρέπει να βρίσκεται πολύ κοντά στην τιμή $(m + 1) / 2$, διαφορετικά οι τιμές \bar{r}_j θα διαφέρουν σημαντικά μεταξύ τους. Αυτό μας οδηγεί στη χρήση της στατιστικής παραμέτρου Kruskal-Wallis (βλ. π.χ. Hirsch et al.,

1993, σ. 17.25), η οποία που συνδυάζει τους μέσους βαθμούς από όλα τα επιμέρους δείγματα:

$$h = \frac{12}{m(m+1)} \sum_{j=1}^k n_j \left(\bar{r}_j - \frac{m+1}{2} \right)^2 \quad (8.29)$$

Κατά συνέπεια το πρόβλημα του προσδιορισμού των παραμέτρων θ και η μπορεί να αναχθεί στην ελαχιστοποίηση της στατιστικής παραμέτρου h .^{*} Η αναλυτική ελαχιστοποίηση δεν είναι δυνατή και γι' αυτό θα πρέπει να χρησιμοποιηθεί αριθμητική μέθοδος, η οποία ελέγχει με συστηματικό τρόπο δοκιμαστικές τιμές των παραμέτρων. Για παράδειγμα μπορεί να χρησιμοποιηθεί για τη βελτιστοποίηση η τυπική μέθοδος διχοτόμησης.

Για λόγους καλύτερης προσαρμογής της συνάρτησης $b(d)$ στην περιοχή των υψηλότερων εντάσεων, είναι σκόπιμο να μη χρησιμοποιείται σε αυτό το πρώτο στάδιο υπολογισμού το σύνολο των δεδομένων κάθε επιμέρους δείγματος, αλλά ένα μέρος αυτών των δεδομένων. Για παράδειγμα, μπορεί να χρησιμοποιείται μόνο το πρώτο 1/3 των δεδομένων από κάθε διάρκεια, αφού τα δεδομένα καταταγούν σε φθίνουσα σειρά.

Αφού προσδιοριστούν οι παράμετροι θ και η , είναι απλή υπόθεση η εκτίμηση των παραμέτρων της συνάρτησης $a(T)$, η οποία γίνεται στο δεύτερο στάδιο υπολογισμού. Συγκεκριμένα, οι τελευταίες παράμετροι εκτιμώνται με τις τυπικές μεθόδους της στατιστικής (π.χ. με τη μέθοδο ροπών), όπως αυτές περιγράφονται στο κεφάλαιο 6 και συνοψίζονται στην ενότητα 8.3.1, χρησιμοποιώντας το ενοποιημένο δείγμα που περιέ-

* Αν τα επιμέρους δείγματα είναι ανεξάρτητα, τότε η στατιστική συνάρτηση H , της οποίας η σημειακή εκτίμηση είναι η παραπάνω τιμή h , ακολουθεί κατανομή χ^2 με $k - 1$ βαθμούς ελευθερίας. Στην περίπτωση αυτή είναι δυνατός ο έλεγχος της υπόθεσης $H = 0$, που ισοδυναμεί με την υπόθεση ότι όλα τα δείγματα προέρχονται από τον ίδιο πληθυσμό. Ο έλεγχος αυτός είναι μη παραμετρικός με την έννοια ότι δεν κάνει καμιά υπόθεση σχετικά με την κατανομή που ακολουθεί η μεταβλητή Y . Ωστόσο, στην περίπτωση που εξετάζουμε τα επιμέρους δείγματα που αναφέρονται σε διαφορετικές διάρκειες δεν είναι ανεξάρτητα, αλλά, αντίθετα, ισχυρώς συσχετισμένα. Έτσι δεν είναι γνωστή η κατανομή της H και δεν είναι δυνατός ο στατιστικός έλεγχος. Πάντως, ο στόχος της ελαχιστοποίησης της τιμής h εξακολουθεί να έχει νόημα και σε αυτή την περίπτωση.

χει όλα τα m δεδομένα y_{jl} . Είναι βέβαια αυτονόητο ότι σε αυτό το δεύτερο στάδιο υπολογισμού πρέπει να χρησιμοποιείται το σύνολο των δεδομένων, και όχι ένα τμήμα τους.

Καθολική εκτίμηση

Η μέθοδος αυτή εκτιμά ταυτόχρονα το σύνολο των παραμέτρων των όμβριων καμπυλών ελαχιστοποιώντας το συνολικό σφάλμα των όμβριων καμπυλών σε σχέση με τα ιστορικά δεδομένα.

Προϋπόθεση για τη χρήση αυτής της μεθόδου είναι να αντιστοιχιστεί σε κάθε στοιχείο κάθε δείγματος μια συγκεκριμένη περίοδος επαναφοράς. Για το σκοπό αυτό χρησιμοποιούμε την εμπειρική συνάρτηση κατανομής, οπότε για το στοιχείο (ένταση βροχής) l του δείγματος j , διατεταγμένου σε φθίνουσα σειρά (συμβολικά i_{jl}), η περίοδος επαναφοράς είναι

$$T_{jl} = \frac{n_j + 0.12}{l - 0.44} \quad (8.30)$$

αν δεχτούμε την έκφραση κατά Gringorten, ή

$$T_{jl} = \frac{n_j + 0.2}{l - 0.4} \quad (8.31)$$

αν δεχτούμε την έκφραση κατά Cunnane. Κατά συνέπεια, κάθε στοιχείο δείγματος περιγράφεται από μια τριάδα αριθμών (i_{jl} , T_{jl} , d_j). Αν υποθέσουμε ότι είναι γνωστό το σύνολο των παραμέτρων των όμβριων καμπυλών, τότε από την (8.10) για δεδομένα T_{jl} και d_j υπολογίζεται η αντίστοιχη θεωρητική (μοντελοποιημένη) ένταση

$$\hat{i}_{jl} = \frac{a(T_{jl})}{b(d_j)} \quad (8.32)$$

και το αντίστοιχο σφάλμα

$$e_{jl} = \ln i_{jl} - \ln \hat{i}_{jl} = \ln (i_{jl} / \hat{i}_{jl}) \quad (8.33)$$

Στην τελευταία εξίσωση έχουμε χρησιμοποιήσει λογαριθμικό μετασχηματισμό των εντάσεων, προκειμένου να κρατήσουμε ισορροπία ανάμεσα στα σφάλματα των εντάσεων για μικρές και μεγάλες διάρκειες (δεδομένου ότι στις πρώτες οι εντάσεις είναι μεγαλύτερες). Το καθολικό μέσο σφάλμα υπολογίζεται από την ακόλουθη εξίσωση

$$E = \frac{1}{k} \sum_{j=1}^k \frac{1}{n_j} \sum_{l=1}^{n_j} e_{jl}^2 \quad (8.34)$$

Η ελαχιστοποίηση του παραπάνω καθολικού σφάλματος αποτελεί το στόχο της διαδικασίας εκτίμησης παραμέτρων. Λόγω των πολύπλοκων εκφράσεων των όμβριων καμπυλών, ιδιαίτερα της συνάρτησης $a(T)$, δεν μπορεί να κατασκευαστεί γενική αναλυτική λύση του προβλήματος ελαχιστοποίησης. Για το λόγο αυτό πρέπει να καταφύγουμε σε αριθμητικές μεθόδους. Μία τέτοια μέθοδος περιγράφεται από τους Κουτσογιάννη και Μανέτα (1995).

Εφαρμογή 8.3.2

Να εξαχθούν συνεπείς ενιαίες σχέσεις όμβριων καμπυλών για τα δεδομένα του σταθμού Ελληνικό (Πίν. 8.2) με τη μέθοδο της ενοποίησης διάρκειών.

Όπως εξηγήσαμε στην Εφαρμογή 8.2.1, για την πιθανοτική περιγραφή των εντάσεων υιοθετούμε την κατανομή Gumbel, οπότε η συνάρτηση $a(T)$ θα δίνεται από την (8.17), και για τη συνάρτηση διάρκειας υιοθετούμε την τριπαραμετρική εξίσωση $b(d) = (d + \theta)^n$. Κατά συνέπεια, η έκφραση των όμβριων καμπυλών θα είναι

$$i = \frac{a(T)}{b(d)} = \frac{\frac{1}{\lambda} \left\{ \psi - \ln \left[-\ln \left(1 - \frac{1}{T} \right) \right] \right\}}{(d + \theta)^n} \quad (\alpha)$$

Ισοδύναμα, η έκφραση μπορεί να γραφεί

$$i = \frac{1}{l(d)} \left\{ \psi - \ln \left[-\ln \left(1 - \frac{1}{T} \right) \right] \right\} \quad (\beta)$$

όπου

$$l(d) = \lambda (d + \theta)^n \quad (\gamma)$$

Ας σημειωθεί ότι η (β) ουσιαστικά ταυτίζεται με μία (αντίστροφη) συνάρτηση κατανομής Gumbel, στην οποία η παράμετρος κλίμακας είναι συνάρτηση της διάρκειας $l(d)$, ενώ η αδιάστατη παράμετρος ψ είναι σταθερή, ανεξάρτητη της διάρκειας. Το γεγονός ότι η παράμετρος ψ είναι σταθερή επιβεβαιώνεται (κατά προσέγγιση) και από τις μεμονωμένες συναρτήσεις κατανομής που προσαρμόσαμε στα επιμέρους δείγματα στην Εφαρμογή 8.2.1 (Πίν. 8.3).

Σύμφωνα με τη γενική μεθοδολογία που αναπτύχθηκε πιο πάνω, στο πρώτο στάδιο εκτιμάμε τις παραμέτρους θ και η . Όπως είδαμε, διαδικασία περιλαμβάνει επαναλήψεις, οι οποίες απαιτούν μακροσκελείς υπολογισμούς που για λόγους χώρου δεν μπορούν να παρατεθούν εδώ στο σύνολό τους. Θα παραθέσουμε μόνο τους υπολογισμούς της τελικής, βέλτιστης λύσης, $\theta = 0.189$ και $\eta = 0.796$.

Στον Πίν. 8.6 έχουν υπολογιστεί οι τιμές $y = i b(d)$ για κάθε διάρκεια, όπου οι τιμές i έχουν ληφθεί από τον Πίν. 8.2 και οι τιμές $b(d)$ είναι ίσες με $(d + 0.189)^{0.796}$. Τα δείγματα των επιμέρους διαρκειών έχουν καταταγεί σε φθίνουσα σειρά.

Για την εξαγωγή των βαθμών των δεδομένων του Πίν. 8.6 έχει ληφθεί το $1/3$ των δεδομένων από κάθε διάρκεια. Έτσι για τις διάρκειες 5 και 10 min που διαθέτουν 29 δεδομένα παίρνουμε τα δέκα πρώτα ($29/3 = 9.67 \approx 10$), για τις διάρκειες 30 min - 12 h παίρνουμε επίσης τα δέκα πρώτα δεδομένα, και για τη διάρκεια των 24 h παίρνουμε τα πρώτα 7 δεδομένα ($20/3 = 6.67 \approx 7$). Η τεθλασμένη γραμμή που έχει χαραχτεί στον Πίν. 8.6 δείχνει ακριβώς ποια δεδομένα παίρνονται υπόψη στην εξαγωγή των βαθμών. Για τα δεδομένα αυτά έχουν εξαχθεί οι βαθμοί στον Πίν. 8.7. Για παράδειγμα, ο βαθμός 1 δίνεται στην μεγαλύτερη απ' όλες τις τιμές, (55.012), ο βαθμός 2 στη δεύτερη (53.002) κ.ο.κ. Στην τελευταία σειρά του Πίν. 8.7 έχουν εξαχθεί οι μέσοι βαθμοί για κάθε διάρκεια.

Ο συνολικός αριθμός δεδομένων είναι $m = 10 \times 7 + 7 = 77$. Ο θεωρητικά αναμενόμενος μέσος βαθμός για κάθε διάρκεια είναι $(m + 1) / 2 = 39$. Το άθροισμα του δεξιού μέλους της (8.29) είναι

$$10 \times (32.9 - 39)^2 + 10 \times (49 - 39)^2 + 10 \times (40.1 - 39)^2 + \\ 10 \times (36.1 - 39)^2 + 10 \times (36.6 - 39)^2 + 10 \times (41.4 - 39)^2 + \\ 10 \times (36.1 - 39)^2 + 7 \times (39.7 - 39)^2 = 1671.2.$$

Κατά συνέπεια η στατιστική παράμετρος h είναι

$$h = [12 / (77 \times 78)] \times 1671.2 = 3.34$$

Μπορεί να διαπιστωθεί ότι με οποιοδήποτε άλλο συνδυασμό παραμέτρων θ και η προκύπτει μεγαλύτερη τιμή της h .*

* Για ενδεικτικούς λόγους περιγράφουμε και τη διαδικασία στατιστικού ελέγχου αν και όπως είδαμε στην υποσημείωση της σ. 277, δεν έχει νόημα, λόγω της ισχυρής εξάρτησης των επιμέρους δειγμάτων. Οι βαθμοί ελευθερίας είναι $k - 1 = 8 - 1 = 7$. Η κρίσιμη τιμή της στατιστικής δοκιμής για επίπεδο σημαντικότητας 5% είναι $\chi_{0.95}^2(7) = 14.07 > 3.34$, γεγονός που θα μας οδηγούσε στη μη απόρριψη της μηδενικής υπόθεσης ότι τα δείγματα της μεταβλητής Y προέρχονται από τον ίδιο πληθυσμό.

Πίν. 8.6 Δείγματα της μεταβλητής $Y = I b(d)$ καταταγμένα σε φθίνουσα σειρά, για $\theta = 0.189$ και $\eta = 0.796$. Εντάσεις βροχής από το σταθμό Ελληνικό (Πίν. 8.2).

| A/A | Μεταβλητή $y = i b(d)$ για διάρκεια $d =$ | | | | | | | |
|-----|---|--------|--------|--------|--------|--------|--------|--------|
| | 5 min | 10 min | 30 min | 1 h | 2 h | 6 h | 12 h | 24 h |
| 1 | 50.281 | 52.700 | 55.012 | 46.943 | 50.187 | 50.919 | 53.002 | 48.570 |
| 2 | 44.742 | 38.471 | 48.470 | 41.548 | 42.724 | 42.743 | 45.866 | 42.938 |
| 3 | 42.611 | 37.681 | 39.549 | 41.204 | 40.112 | 42.103 | 44.768 | 42.571 |
| 4 | 42.611 | 37.417 | 37.467 | 40.171 | 37.593 | 40.537 | 43.912 | 36.939 |
| 5 | 41.759 | 35.046 | 36.873 | 38.449 | 36.847 | 38.263 | 38.547 | 33.832 |
| 6 | 40.907 | 30.303 | 35.386 | 37.531 | 36.474 | 32.571 | 36.652 | 28.945 |
| 7 | 36.219 | 30.303 | 32.115 | 34.432 | 36.007 | 31.504 | 33.241 | 28.680 |
| 8 | 34.089 | 29.249 | 31.074 | 33.399 | 34.608 | 31.222 | 33.058 | 25.308 |
| 9 | 29.402 | 28.985 | 30.925 | 33.285 | 33.116 | 30.010 | 29.150 | 24.626 |
| 10 | 28.975 | 28.985 | 30.777 | 30.760 | 33.022 | 28.944 | 27.320 | 22.946 |
| 11 | 27.697 | 28.458 | 30.182 | 28.464 | 28.265 | 28.803 | 26.589 | 22.264 |
| 12 | 27.697 | 28.195 | 29.290 | 28.349 | 27.799 | 28.730 | 25.374 | 21.254 |
| 13 | 26.419 | 27.931 | 28.695 | 28.120 | 24.347 | 26.102 | 24.642 | 20.837 |
| 14 | 26.419 | 27.668 | 24.532 | 26.628 | 23.228 | 25.317 | 24.152 | 20.471 |
| 15 | 24.288 | 26.350 | 23.937 | 24.676 | 22.668 | 23.823 | 23.910 | 19.991 |
| 16 | 23.862 | 24.769 | 22.748 | 22.266 | 20.896 | 23.256 | 22.629 | 19.524 |
| 17 | 23.862 | 21.871 | 22.599 | 20.315 | 20.709 | 20.836 | 21.714 | 18.728 |
| 18 | 23.862 | 21.080 | 21.707 | 18.249 | 19.869 | 18.562 | 20.551 | 16.417 |
| 19 | 21.305 | 21.080 | 21.261 | 17.905 | 18.377 | 18.276 | 20.126 | 14.372 |
| 20 | 20.879 | 21.080 | 21.113 | 17.790 | 17.537 | 18.208 | 20.126 | 10.520 |
| 21 | 20.879 | 20.290 | 20.815 | 17.446 | 16.698 | 17.994 | 19.885 | |
| 22 | 20.027 | 18.445 | 20.667 | 17.331 | 16.045 | 17.922 | 19.028 | |
| 23 | 19.175 | 18.445 | 18.734 | 16.413 | 15.858 | 15.788 | 17.989 | |
| 24 | 17.044 | 18.445 | 17.544 | 15.839 | 15.485 | 15.575 | 16.650 | |
| 25 | 14.914 | 17.918 | 17.098 | 15.265 | 14.086 | 15.434 | 16.650 | |
| 26 | 14.914 | 15.283 | 14.868 | 14.806 | 13.433 | 15.007 | 15.860 | |
| 27 | 14.488 | 15.283 | 13.679 | 13.773 | 12.780 | 13.369 | 14.213 | |
| 28 | 14.062 | 13.966 | 12.786 | 13.314 | 12.407 | 12.942 | 13.298 | |
| 29 | 12.783 | 10.804 | 12.340 | 12.625 | 12.407 | 10.454 | 12.383 | |
| 30 | | | 12.192 | 11.707 | 12.220 | 10.241 | 11.468 | |

Ακολουθεί η εκτίμηση των παραμέτρων λ και ψ της κατανομής Gumbel. Η μέση τιμή του συνόλου των δεδομένων του Πίν. 8.6 (όλων των στηλών) είναι 25.701 και η τυπική απόκλιση 10.208. Κατά συνέπεια η μέθοδος ροπών δίνει

$$\lambda = 1 / (0.78 \times 10.208) = 0.1256, \psi = 0.1256 \times 25.710 - 0.577 = 2.652$$

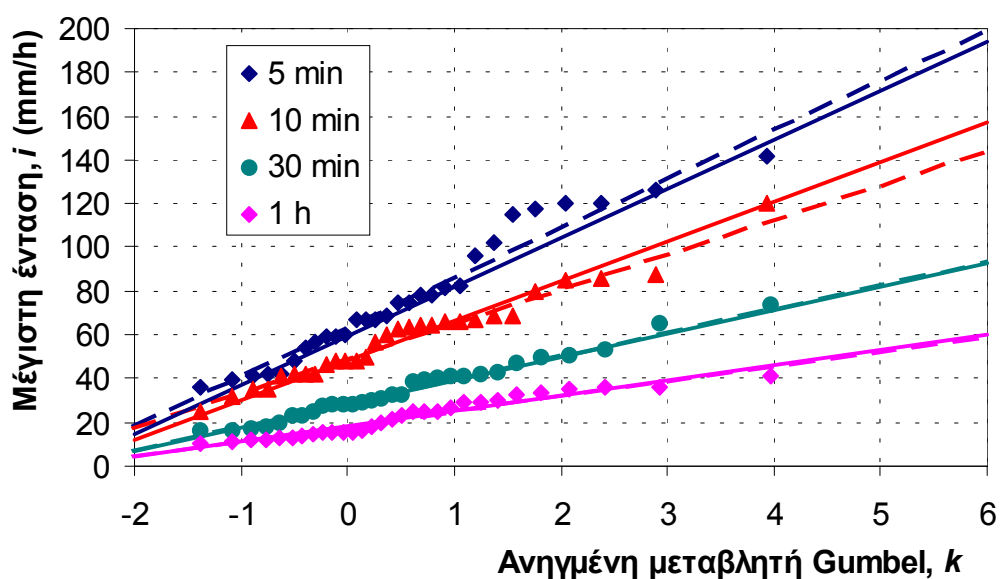
Έτσι η τελική εξίσωση όμβριων καμπυλών είναι

$$i = \frac{1}{0.1256 (d + 0.189)^{0.796}} \left\{ 2.652 - \ln \left[-\ln \left(1 - \frac{1}{T} \right) \right] \right\} \quad (\delta)$$

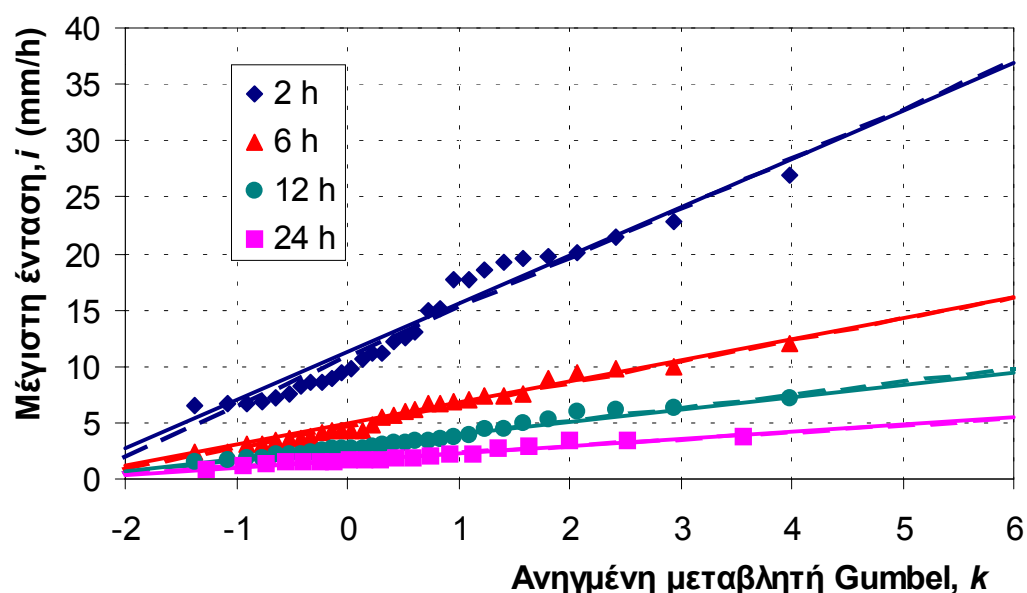
Η απεικόνιση και η σύγκριση της παραπάνω εξίσωσης γίνεται σε δύο ομάδες σχημάτων. Στο Σχ. 8.7 και στο Σχ. 8.8 απεικονίζονται σε χαρτί Gumbel οι συναρτήσεις κατανομής των μέγιστων εντάσεων βροχής στο Ελληνικό ξεχωριστά για κάθε διάρκεια. Με σημεία παριστάνονται οι εμπειρικές συναρτήσεις κατανομής (κατά Gringorten), με συνεχείς γραμμές οι συναρτήσεις Gumbel της συνεπούς ενιαίας έκφρασης (δ) και με διακεκομμένες γραμμές οι μεμονωμένες οι συναρτήσεις Gumbel που έχουν εξαχθεί στην Εφαρμογή 8.2.1. Παρατηρούμε την πολύ καλή συμφωνία των δύο ομάδων καμπυλών, τόσο μεταξύ τους, όσο και με τις σημειοσειρές των εμπειρικών κατανομών. Στο Σχ. 8.9 παριστάνονται οι όμβριες καμπύλες όπως προκύπτουν από την παραπάνω εξίσωση (δ) για περιόδους επαναφοράς από 5 μέχρι 5000. Για σύγκριση έχουν απεικονιστεί και τα σημεία που προκύπτουν από τις μεμονωμένες συναρτήσεις Gumbel που έχουν εξαχθεί στην Εφαρμογή 8.2.1. Άξια προσοχής είναι η πολύ καλή συμφωνία των καμπυλών με τα σημεία για όλο το φάσμα των περιόδων επαναφοράς (τρεις τάξεις μεγέθους). Αυτό επιβεβαιώνει τον ισχυρισμό μας ότι η παραπάνω συνεπής μέθοδος καλύπτει όλο το φάσμα μεταβολής διαρκειών και περιόδων επαναφοράς με ένα μοναδικό σύνολο παραμέτρων.

Πίν. 8.7 Αύξοντες αριθμοί (βαθμοί) του 1/3 των δεδομένων του Πίν. 8.6.

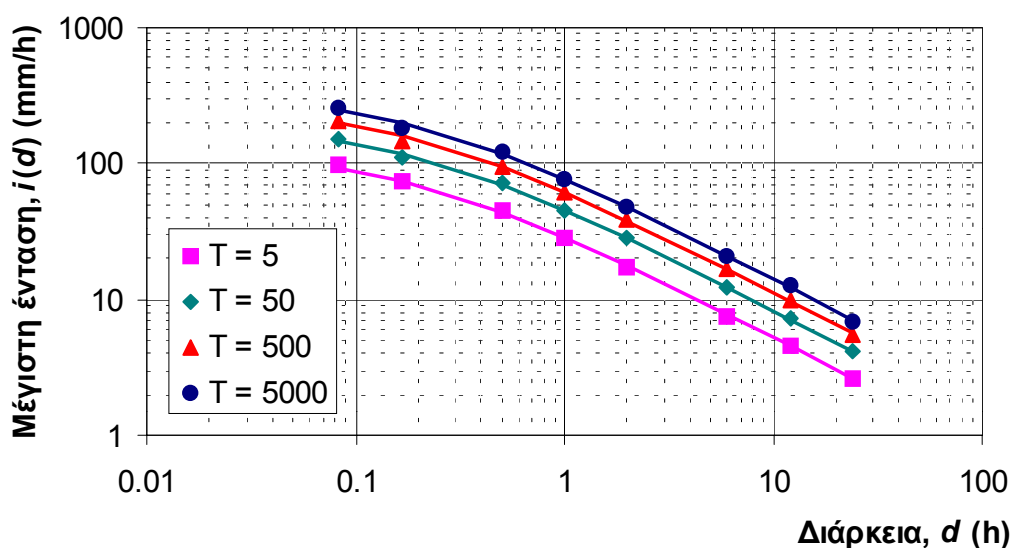
| A/A | Βαθμοί των δεδομένων κάθε επιμέρους δείγματος διάρκειας $d =$ | | | | | | | |
|--------------|---|--------|--------|------|------|------|------|------|
| | 5 min | 10 min | 30 min | 1 h | 2 h | 6 h | 12 h | 24 h |
| 1 | 5 | 3 | 1 | 9 | 6 | 4 | 2 | 7 |
| 2 | 12 | 30 | 8 | 22 | 16 | 15 | 10 | 14 |
| 3 | 17 | 33 | 28 | 23 | 27 | 20 | 11 | 19 |
| 4 | 17 | 37 | 36 | 26 | 34 | 25 | 13 | 38 |
| 5 | 21 | 46 | 39 | 31 | 40 | 32 | 29 | 50 |
| 6 | 24 | 65 | 45 | 35 | 42 | 57 | 41 | 74 |
| 7 | 43 | 65 | 58 | 48 | 44 | 59 | 53 | 76 |
| 8 | 49 | 69 | 61 | 51 | 47 | 60 | 55 | |
| 9 | 68 | 71 | 62 | 52 | 54 | 67 | 70 | |
| 10 | 73 | 71 | 63 | 64 | 56 | 75 | 77 | |
| Μέσος βαθμός | 32.9 | 49.0 | 40.1 | 36.1 | 36.6 | 41.4 | 36.1 | 39.7 |



Σχ. 8.7 Συναρτήσεις κατανομής των μέγιστων εντάσεων βροχής στο Ελληνικό για διάρκειες βροχής από 5 min μέχρι 1 h. Με σημεία παριστάνονται οι εμπειρικές συναρτήσεις κατανομής, με συνεχείς γραμμές οι συναρτήσεις Gumbel της συνεπούς ενιαίας έκφρασης και με διακεκομμένες γραμμές οι μεμονωμένες οι συναρτήσεις Gumbel ανά διάρκεια.



Σχ. 8.8 Συναρτήσεις κατανομής των μέγιστων εντάσεων βροχής στο Ελληνικό για διάρκειες βροχής από 2 μέχρι 24 h. Με σημεία παριστάνονται οι εμπειρικές συναρτήσεις κατανομής, με συνεχείς γραμμές οι συναρτήσεις Gumbel της συνεπούς ενιαίας έκφρασης και με διακεκομμένες γραμμές οι μεμονωμένες οι συναρτήσεις Gumbel ανά διάρκεια.



Σχ. 8.9 Όμβριες καμπύλες για περιόδους επαναφοράς $T = 5, 50, 500$ και 5000 στο Ελληνικό, υπολογισμένες από τη συνεπή ενιαία εξίσωση με παραμέτρους εκτιμημένες με τη μέθοδο ενοποίησης των διαρκειών (λογαριθμικοί άξονες).

8.4 Συμπληρωματικά στοιχεία

8.4.1 Ενοποίηση δειγμάτων διαφορετικών σταθμών

Μερικές φορές λειτουργούν δύο ή περισσότεροι βροχογραφικοί σταθμοί σε πολύ κοντινές αποστάσεις μεταξύ τους. Είναι αναμενόμενο ότι οι σχέσεις έντασης-διάρκειας-περιόδου επαναφοράς στους σταθμούς αυτούς θα είναι ταυτόσημες. Με άλλα λόγια μπορεί να υποτεθεί ότι τα δείγματα των μέγιστων εντάσεων βροχής που έχουν ληφθεί στους διαφορετικούς σταθμούς προέρχονται από τον ίδιο πληθυσμό ή αποτελούν δείγματα της ίδιας τυχαίας μεταβλητής. Αν είναι αληθής η υπόθεση αυτή, τα δείγματα των διαφορετικών σταθμών για κάθε δεδομένη διάρκεια μπορούν να ενοποιηθούν και να αποκτηθούν έτσι δείγματα μεγαλύτερου μήκους. Οι παράμετροι των όμβριων καμπυλών μπορούν στη συνέχεια να εκτιμηθούν από τα ενοποιημένα δείγματα, με έναν από τους τρόπους που περιγράφηκαν παραπάνω. Η ενοποίηση των δειγμάτων διαφορετικών σταθμών είναι γνωστή με τον όρο *μέθοδος σταθμών-ετών*.

Η υπόθεση ότι τα επιμέρους δείγματα εκφράζουν την ίδια τυχαία μεταβλητή πρέπει να ελέγχεται με δύο τρόπους: εμπειρικά και στατιστικά:

1. Ο εμπειρικός έλεγχος αφορά στη διαπίστωση ότι οι θέσεις των μετρητικών σταθμών βρίσκονται στο ίδιο μικροκλιματικό περιβάλλον: ίδια υδρολογική λεκάνη, κοντινές αποστάσεις, περίπου ίδια υψόμετρα, παρόμοιες ορογραφικές συνθήκες, κτλ.
2. Ο στατιστικός έλεγχος αφορά στην εφαρμογή της δοκιμής Kruskal-Wallis (βλ. εξίσωση (8.29)) για κάθε ομάδα επιμέρους δειγμάτων δεδομένης διάρκειας. Για να μην απορριφθεί η μηδενική υπόθεση και να καταστεί έτσι επιτρεπτή η ενοποίηση, θα πρέπει οι στατιστικές παράμετροι h να είναι μικρότερες της κρίσιμης τιμής h_c για όλες τις διάρκειες.

Για να έχει νόημα ο παραπάνω στατιστικός έλεγχος θα πρέπει τα δείγματα των διαφορετικών σταθμών, που αναφέρονται στην ίδια διάρκεια, να είναι στατιστικώς ανεξάρτητα. Αυτό ελέγχεται κατά προσέγγιση και πάλι με δύο τρόπους: εμπειρικά και στατιστικά.

1. Ο εμπειρικός έλεγχος αφορά στη διαπίστωση ότι η μέγιστη ένταση βροχής σε καθέναν από τους σταθμούς έχει καταγραφεί σε διαφορετική ημερομηνία του ίδιου έτους, δηλαδή ότι δεν πρόκειται για καταγραφές ίδιων επεισοδίων βροχής σε διαφορετικές χωρικές θέσεις.
2. Ο στατιστικός έλεγχος εξετάζει κατά κανόνα το κατά πόσο ο συντελεστής συσχέτισης μπορεί να θεωρηθεί στατιστικώς μηδέν.

Η μέθοδος σταθμών ετών έχει χρησιμοποιηθεί πολύ συχνά με στόχο την απόκτηση ενός μεγάλου σε μήκος δείγματος από δύο ή περισσότερα μικρά δείγματα και, άρα, την αύξηση της αξιοπιστίας των εκτιμήσεων ή, ισοδύναμα, τον περιορισμό των ορίων εμπιστοσύνης τους. Θα διατυπώσουμε, ωστόσο, ορισμένες επιφυλάξεις ως προς την φύση της μεθόδου, οι οποίες θα πρέπει να είναι εις γνώση του μηχανικού που χρησιμοποιεί τη μέθοδο.

1. Η θετική έκβαση του ελέγχου της ανεξαρτησίας των μικρών δειγμάτων που πρόκειται να ενοποιηθούν, δεν σημαίνει απαραίτητα και την ανεξαρτησία των αντίστοιχων μεταβλητών. Πράγματι, σε μέτρια σε μέγεθος επεισόδια βροχής φαίνεται να υπάρχει ανεξαρτησία των

καταγραφών σε γειτονικούς σταθμούς. Ωστόσο, η πείρα έχει δείξει ότι τα πολύ εξαιρετικά επεισόδια καταιγίδων, δίνουν ταυτόχρονα (άρα εξαρτημένα) μέγιστα σε γειτονικούς σταθμούς. Κατά συνέπεια, αν διαθέταμε μακροχρόνια δείγματα (π.χ. 100 ετών) μέγιστων εντάσεων στους γειτονικούς σταθμούς είναι πολύ πιθανό ότι θα είχαμε εξαρτημένες καταγραφές στην περιοχή των πολύ υψηλών εντάσεων βροχής. Άρα η ανεξαρτησία που διαπιστώνουμε όταν διαθέτουμε μικρά δείγματα, είναι ενδεχομένως παραπλανητική.

2. Όσο μεγαλώνει η χρονική διάρκεια αναφοράς d τόσο η ανεξαρτησία τίθεται εν αμφιβόλω, ακόμη και σε μικρά δείγματα. Έτσι, για διάρκειες βροχής π.χ. 24 ή 48 ωρών τα ίδια επεισόδια βροχής δίνουν συνήθως μέγιστες εντάσεις σε αρκετούς σταθμούς μιας περιοχής.
3. Όσο μικραίνει η απόσταση ανάμεσα σε δύο γειτονικούς σταθμούς τόσο περισσότερο εξαρτημένα θα είναι τα δείγματα. Αλλά, μια αρκετά μεγάλη απόσταση ανάμεσα στους σταθμούς, που θα εξασφάλιζε την ανεξαρτησία, ακυρώνει την ουσιαστική προϋπόθεση της μεθόδου, ότι ανήκουν στο ίδιο μικροκλιματικό περιβάλλον.

Κατά συνέπεια, δεν συνιστάται η γενική και ανεπιφύλακτη εφαρμογή της μεθόδου σταθμών-ετών. Η εφαρμογή της μπορεί να είναι απαραίτητη μόνο σε ειδικές περιπτώσεις ανεπαρκών σε μέγεθος δειγμάτων, αλλά και πάλι χρειάζεται ιδιαίτερη προσοχή, κυρίως στην εκτίμηση των ορίων εμπιστοσύνης, όπου παίζει καθοριστικό ρόλο το μέγεθος του δείγματος.

8.4.2 Επίδραση της χρονικής ευκρίνειας

Η χρονική ευκρίνεια των πρωτογενών δεδομένων (βροχογραφήμάτων ή ψηφιακών μετρήσεων) είναι προφανές ότι επηρεάζει τη σειρά των μέγιστων εντάσεων βροχής. Εκ των πραγμάτων κατά την εξαγωγή της σειράς μέγιστων εντάσεων, αντί της ακριβούς εξίσωσης (8.1) χρησιμοποιείται η προσεγγιστική εξίσωση (8.2), στην οποία υπεισέρχεται και η χρονική ευκρίνεια των παρατηρήσεων. Αυτό έχει συνέπεια την υπεκτίμηση των μέγιστων εντάσεων. Είναι προφανές ότι το μέγεθος του σφάλματος εξαρτάται από το λόγο διάρκειας προς ευκρίνεια (d/δ) και αν ο λόγος αυτός είναι αρκετά μεγάλος τότε το σφάλμα γίνεται αμελητέο. Για την άρση του σφάλματος για μικρές τιμές του λόγου d/δ , συνήθως γίνεται αναγωγή των τιμών $i(d)$, με πολλαπλασιασμό επί ένα συντελεστή που εξαρτάται από το λόγο d/δ . Τιμές αυτού του συντελεστή έχουν βρεθεί από έρευνες

στην Αμερική και δίνονται στη βιβλιογραφία (π.χ. Linsley et al., 1975, σ. 357), απ' όπου προέρχεται και ο Πίν. 8.8 που δίνεται παρακάτω.

Πίν. 8.8 Τιμές του συντελεστή άρσης του σφάλματος διακριτοποίησης

| Λόγος διάρκειας προς ευκρίνεια (d/δ) | Συντελεστής άρσης του σφάλματος διακριτοποίησης |
|--|--|
| 1 | 1.13 |
| 2 | 1.04 |
| 3-4 | 1.03 |
| 5-8 | 1.02 |
| 9-24 | 1.01 |

8.4.3 Σχέση σειρών ετήσιων μεγίστων και σειρών υπεράνω κατωφλίου

Είναι προφανές ότι ο τρόπος κατάρτισης της σειράς μέγιστων εντάσεων βροχής επηρεάζει τις τελικές εξισώσεις των όμβριων καμπυλών. Συγκεκριμένα, για την ίδια περίοδο επαναφοράς, η σειρά των ετήσιων μεγίστων θα δώσει, μετά από τη στατιστική επεξεργασία της, χαμηλότερη τιμή της έντασης βροχής από την αντίστοιχη τιμή που θα προκύψει από την επεξεργασία της σειράς υπεράνω κατωφλίου. Πάντως, η διαφορά αυτή γίνεται αμελητέα για μεγάλες τιμές της περιόδου επαναφοράς ($T > 10$).

Πιο αντιπροσωπευτικές για τη μελέτη έργων με μικρή περίοδο επαναφοράς σχεδιασμού είναι οι όμβριες καμπύλες που προκύπτουν από τις σειρές υπεράνω κατωφλίου. Ωστόσο, στην πράξη χρησιμοποιούνται περισσότερο οι σειρές ετήσιων μεγίστων, λόγω του ευκολότερου τρόπου κατασκευής τους. Μπορεί πάντως και εκ των υστέρων να γίνει αναγωγή των αποτελεσμάτων που προέρχονται από επεξεργασία σειρών ετήσιων μεγίστων σε τρόπο ώστε να αντιστοιχούν προσεγγιστικά σε αυτά που προέρχονται από επεξεργασία σειρών υπεράνω κατωφλίου. Η απλούστερη μέθοδος για το σκοπό αυτό αφορά στην αναγωγή της περιόδου επαναφοράς σύμφωνα με τον τύπο (Raudkivi, 1979, σ. 411)

$$T = \frac{1}{1 - \exp(-1/T')} \Leftrightarrow T' = \frac{1}{-\ln(1 - 1/T)} \quad (8.35)$$

όπου T' η περίοδος επαναφοράς για δεδομένη ένταση βροχής όπως προκύπτει από τη σειρά μεγίστων υπεράνω κατωφλίου και T η αντίστοιχη

περίοδος επαναφοράς για τη σειρά ετησίων μεγίστων. Η πιο πάνω σχέση προσεγγίζεται με ακρίβεια δεύτερου δεκαδικού ψηφίου από την απλούστερη σχέση

$$T = T' + 0.5 \quad (8.36)$$

Πρακτικώς η αναγωγή δεν είναι απαραίτητη για $T > 10$.

Σε μερικά βιβλία υδρολογίας δίνεται σε πινακοποιημένη μορφή η αντιστοιχία των μεγεθών T' και T (π.χ. Linsley et al., 1975, σ. 356· Haan, 1977, σ. 134· Viessman et al., 1989, σ. 742). Οι τιμές των πινάκων συμφωνούν με την εξίσωση (8.35). Σε άλλα βιβλία δίνονται συντελεστές αναγωγής των εντάσεων αντί των περιόδων επαναφοράς οι οποίοι κυμαίνονται από 0.88 για $T = 2$ μέχρι 0.99 για $T = 10$ (Haan, 1977, σ. 155· Chow et al., 1988, σ. 456). Ωστόσο η μέθοδος που στηρίζεται στην αναγωγή των περιόδων επαναφοράς (εξίσωση (8.35)) συνιστάται ως πλέον ορθολογιστική.

Ας σημειωθεί ότι η (8.35) επιτρέπει την άμεση αναγωγή των εξισώσεων των όμβριων καμπυλών σε τρόπο ώστε να εκφράζονται συναρτήσει της T' αντί της T , παρόλο που στην όλη ανάλυση χρησιμοποιείται η σειρά των ετήσιων μεγίστων. Για παράδειγμα, στην περίπτωση της κατανομής Gumbel, η εξίσωση (8.17) μετασχηματίζεται σε

$$a(T) = \frac{1}{\lambda} (\psi + \ln T') \quad (8.37)$$

όπου οι παράμετροι λ και ψ εκτιμώνται από τη σειρά ετήσιων μεγίστων.

8.4.4 Όρια εμπιστοσύνης όμβριων καμπυλών

Στην περίπτωση που χρησιμοποιείται η μεμονωμένη θεώρηση των δειγμάτων των διάφορων διαρκειών d_j (εδάφιο 8.2.1) τα όρια (και κατ' επέκταση οι καμπύλες) εμπιστοσύνης των μέγιστων εντάσεων βροχής για καθεμιά από τις διάρκειες d_j , για τις οποίες υπάρχουν δείγματα όμβριων καμπυλών μπορούν να εκτιμώνται με τις μεθόδους του κεφαλαίου 6. Πρόβλημα ανακύπτει για ενδιάμεσες διάρκειες, όπου χρειάζεται να γίνει κάποια παρεμβολή.

Η γενικευμένη έκφραση των όμβριων καμπυλών με ενσωμάτωση της θεωρητικής συνάρτησης κατανομής σε αυτή (ενότητα 8.3), διευκολύνει την εκτίμηση των διαστημάτων εμπιστοσύνης. Η μέθοδος εκτίμησης με

ενοποίηση διαρκειών προσφέρει την μαθηματικά συνεπέστερη βάση για την εκτίμηση ορίων εμπιστοσύνης, για το λόγο ότι σε αυτή τη μέθοδο οι παράμετροι της συνάρτησης $a(T)$ εκτιμώνται με βάση καθιερωμένες στατιστικές μεθόδους (όπως π.χ. η μέθοδος ροπών) για τις οποίες υπάρχει γνωστή μεθοδολογία εκτίμησης ορίων εμπιστοσύνης. Ωστόσο υπάρχει ένα σημείο που χρειάζεται προσοχή, το οποίο έχει σχέση με την τιμή του μεγέθους δείγματος που θα χρησιμοποιηθεί για τον υπολογισμό των ορίων εμπιστοσύνης. Συγκεκριμένα, θα είναι λάθος να θεωρήσει κανείς ότι το μέγεθος του δείγματος είναι ίσο με το πλήθος των τιμών του ενοποιημένου δείγματος m , δεδομένου ότι τα δείγματα των επιμέρους διαρκειών που ενοποιούνται δεν είναι μεταξύ τους ανεξάρτητα, αλλά αντίθετα έχουν ισχυρή στατιστική εξάρτηση. Αλλά ούτε και μπορεί να θεωρηθεί ίσο με το μέγεθος του κάθε επιμέρους δείγματος n , δεδομένου ότι η ενοποίηση των δειγμάτων αυξάνει την αξιοπιστία της εκτίμησης, πράγμα που ισοδυναμεί με αύξηση του μεγέθους δείγματος. Πρακτικά, περιμένει κανείς ότι το μέγεθος δείγματος που πρέπει να χρησιμοποιηθεί στην εκτίμηση ορίων εμπιστοσύνης βρίσκεται κάπου ανάμεσα στις τιμές n και m . Ο ακριβής προσδιορισμός χρειάζεται εκτεταμένη θεωρητική και εμπειρική ανάλυση. Στην πράξη συνιστάται η χρήση της τιμής n , η οποία δίνει εκτιμήσεις προς την πλευρά της ασφάλειας.

8.5 Εξαγωγή όμβριων καμπυλών με χρήση δεδομένων από βροχόμετρα

Στην Ελλάδα αλλά και σε άλλες χώρες η γενική κατάσταση των βροχομετρικών δικτύων χαρακτηρίζεται από (α) ένα σχετικά αραιό δίκτυο σταθμών εξοπλισμένων με βροχογράφο και (β) ένα πολύ πυκνότερο δίκτυο σταθμών με βροχόμετρο, στο οποίο τυπικά η μέτρηση γίνεται μία φορά ημερησίως (συνήθως στις 08:00, αν και υπάρχουν ορισμένοι σταθμοί της ΕΜΥ με πυκνότερες μετρήσεις, π.χ. δύο μετρήσεις ημερησίως, στις 08:00 και 20:00). Οι ταινίες των βροχογράφων, μετά από ψηφιοποίησή τους, παρέχουν επαρκή βάση για την κατάρτιση όμβριων καμπυλών, ιδίως στην περίπτωση που είναι ημερήσιες, οπότε μπορεί να έχουν ευκρίνεια 5-10 min. Οι μετρήσεις των βροχομέτρων δεν μπορούν να χρησιμοποιηθούν άμεσα για κατάρτιση όμβριων καμπυλών.

Πολλές φορές οι μελετητές, προκειμένου να αποκτήσουν μια εικόνα των εντάσεων για διάρκειες βροχής μικρότερες από 24 h, χρησιμοποιούν τις συμπληρωματικές ενδείξεις των παρατηρητών των βροχομέτρων, οι οποίες αναφέρονται στις ώρες έναρξης και λήξης της βροχής. Ωστόσο, η μέθοδος αυτή δεν είναι ικανοποιητική για την κατάρτιση όμβριων καμυλών, για δύο λόγους: Πρώτον, γιατί οι ενδείξεις που αναγράφονται στα φύλλα παρατηρήσεων είναι συχνά εσφαλμένες, ιδίως στις περιπτώσεις που οι βροχοπτώσεις πραγματοποιούνται κατά τη διάρκεια της νύχτας. Δεύτερο και κυριότερο, η συνολική διάρκεια της βροχής, έστω και αν είναι ακριβής, δεν μπορεί να δώσει πληροφορία για τις μέγιστες εντάσεις βροχής που εμφανίστηκαν σε μικρές διάρκειες κατά τη διάρκεια του επεισοδίου βροχής, οι οποίες μπορεί να είναι πολλαπλάσιες της μέσης έντασης. Έτσι, οι καταγραφές των χρόνων έναρξης και λήξης της βροχής δεν προσφέρουν επαρκή βάση για την κατάρτιση όμβριων καμυλών.

Μια άλλη μέθοδος που έχει τύχει ευρείας εφαρμογής στην Ελλάδα είναι αυτή του U. S. Army Corps of Engineers (1965· βλ. και Ξανθόπουλος, 1975). Η μέθοδος αυτή χρησιμοποιεί ως βάση τη μέση τιμή του ύψους των μέγιστων ετήσιων βροχοπτώσεων 24ώρου, όπως αυτή προκύπτει από τις καταγραφές των βροχομέτρων, και ορισμένα βοηθητικά στοιχεία, όπως το μέσο ετήσιο ύψος βροχής, το μέσο ετήσιο αριθμό ημερών καταιγίδας και το μέσο ετήσιο αριθμό ημερών με ύψος βροχής μεγαλύτερο του 1 mm ή των 0.25 mm. Η εφαρμογή της μεθόδου γίνεται με βάση ένα νομογράφημα, το οποίο δίνει ως εξαγόμενο τη μέγιστη βροχόπτωση περιόδου επαναφοράς $T = 2$ για διάρκεια βροχής 1 h, στην οποία μπορεί στη συνέχεια να θεμελιωθεί η κατάρτιση των όμβριων καμυλών. Αν και σε ορισμένες περιπτώσεις η εν λόγω μέθοδος έδωσε ικανοποιητικά αποτελέσματα στον Ελλαδικό χώρο, σε άλλες περιπτώσεις έχει δώσει τελείως απαράδεκτες όμβριες καμύλες. Για το λόγο αυτό δεν συνιστάται για εφαρμογή στην Ελλάδα.

Παρακάτω προτείνουμε μια νέα μεθοδολογία, η οποία έχει ήδη δοκιμαστεί με ενθαρρυντικά αποτελέσματα στη Στερεά Ελλάδα (Κοζώνης, 1995). Η μεθοδολογία αυτή αναφέρεται στον τρόπο εξαγωγής όμβριων καμυλών σε περιοχή που καλύπτεται από βροχόμετρο, αλλά όχι από βροχογράφο. Τα δεδομένα που απαιτεί είναι:

- Μέγιστα ύψη βροχής 24ωρης ή και 48ωρης διάρκειας από το βροχομετρικό σταθμό της περιοχής μελέτης, ο οποίος παρακάτω θα αναφέρεται ως σταθμός μελέτης.
- Μέγιστα ύψη (ή αντίστοιχες εντάσεις) για μεγάλο φάσμα διαρκειών (π.χ. με κατώτερη διάρκεια από μερικά λεπτά μέχρι 1 h και ανώτερη διάρκεια 24 ή 48 h) από τους πλησιέστερους 2-3 βροχομετρικούς σταθμούς οι οποίοι διαθέτουν και βροχογράφο επαρκούς αξιοπιστίας και περιόδου λειτουργίας.

Διευκρινίζεται ότι η μεθοδολογία αναφέρεται σε σημειακές και όχι επιφανειακές όμβριες καμπύλες (η αναγωγή των σημειακών υψών σε επιφανειακά μπορεί να γίνει με τις γνωστές μεθόδους της βιβλιογραφίας). Αν και η μεθοδολογία δεν μπορεί να τυποποιηθεί πλήρως, σε μορφή μονοσήμαντης πορείας υπολογισμών, σε γενικές γραμμές μπορούμε να θεωρήσουμε ότι περιλαμβάνει τα ακόλουθα βήματα:

1. *Κατάρτιση των όμβριων καμπυλών από τα δεδομένα των βροχογράφων.*

Με βάση τα δεδομένα των 2-3 κοντινών σταθμών καταρτίζονται όμβριες καμπύλες με τη μορφή π.χ. (8.6), ξεχωριστά για καθένα από αυτούς. Η έκφραση που υιοθετείται για τις όμβριες καμπύλες πρέπει να είναι κοινή για όλους τους σταθμούς. Γίνεται προσπάθεια να περιοριστεί ο αριθμός των παραμέτρων της σχέσης όμβριων καμπυλών, π.χ. επιλέγοντας $\theta = 0$ ή $\eta = 1$ στην έκφραση της συνάρτησης διάρκειας $b(d)$ και υιοθετώντας διπαραμετρική και όχι τριπαραμετρική συνάρτηση κατανομής.

2. *Συγκριτική μελέτη των παραμέτρων των παραπάνω καμπυλών και επιλογή παραμέτρων του σταθμού μελέτης.*

Εξετάζονται συγκριτικά οι παράμετροι των παραπάνω καμπυλών των βροχογράφων και κατά περίπτωση υιοθετούνται ορισμένες από αυτές για τον βροχομετρικό σταθμό μελέτης, ενώ άλλες υπολογίζονται στο επόμενο βήμα 3. Οι παράμετροι του σταθμού μελέτης που υιοθετούνται στο βήμα αυτό περιλαμβάνουν όλες τις παραμέτρους της συνάρτησης διάρκειας $b(d)$ και ενδεχομένως κάποιες από τις παραμέτρους της συνάρτησης περιόδου επαναφοράς $a(T)$. Πιο συγκεκριμένα:

- 2.1 Σε περίπτωση που το σύνολο των παραμέτρων δεν διαφέρει αισθητά από σταθμό σε σταθμό επαναλαμβάνεται ο υπολογισμός των παραμέτρων με στόχο την εκτίμηση ενός μοναδικού συνόλου παραμέτρων για όλους τους γειτονικούς σταθμούς. Αυτό γίνεται με ενοποίηση των δειγμάτων όλων των σταθμών σύμφωνα με τη μέθοδο σταθμών-ετών. Από το τελικό σύνολο παραμέτρων υιοθετούνται για τον βροχομετρικό σταθμό μελέτης οι παράμετροι της συνάρτησης διάρκειας $b(d)$. Επίσης μπορεί να υιοθετηθούν και παράμετροι της συνάρτησης περιόδου επαναφοράς $a(T)$, αφήνοντας όμως τουλάχιστον μία ελεύθερη, η οποία θα προσδιοριστεί από τα δεδομένα του σταθμού μελέτης.
- 2.2 Αν οι παράμετροι της συνάρτησης διάρκειας $b(d)$ δεν διαφέρουν αισθητά από σταθμό σε σταθμό, αλλά υπάρχουν διαφορές σε κάποιες από τις παραμέτρους της συνάρτησης περιόδου επαναφοράς $a(T)$, τότε εξάγεται ο μέσος όρος των τιμών των (αντίστοιχων) παραμέτρων που δεν έχουν μεγάλες διαφορές, ο οποίος και υιοθετείται ως αντιπροσωπευτική τιμή για την αντίστοιχη παράμετρο του σταθμού μελέτης. Και πάλι θα πρέπει να αφηθεί τουλάχιστον μία ελεύθερη παράμετρος της συνάρτησης περιόδου επαναφοράς $a(T)$, η οποία θα προσδιοριστεί από τα δεδομένα του σταθμού μελέτης.
- 2.3 Αν υπάρχουν σημαντικές διαφορές από σταθμό σε σταθμό στις παραμέτρους της συνάρτησης διάρκειας $b(d)$ τότε υιοθετείται για το σταθμό μελέτης ένα από τα σύνολα παραμέτρων της $b(d)$ και συγκεκριμένα αυτό που δίνει δυσμενέστερες εντάσεις βροχής (για λόγους ασφάλειας). Για παράδειγμα, αν η $b(d)$ είναι της μορφής d^n τότε υιοθετείται για το σταθμό μελέτης η μεγαλύτερη από τις τιμές n των γειτονικών βροχομετρικών σταθμών.
3. Τελική φάση εκτίμησης παραμέτρων.
Εκτιμώνται κατά περίπτωση οι υπόλοιπες παράμετροι της συνάρτησης περιόδου επαναφοράς $a(T)$ που δεν έχουν καθοριστεί στο βήμα 2. Για το σκοπό αυτό:
- 3.1 Αν πρόκειται να εκτιμηθεί σε αυτό το βήμα το σύνολο των παραμέτρων της συνάρτησης περιόδου επαναφοράς $a(T)$, τότε εφαρ-

μόζονται οι γνωστές μέθοδοι εκτίμησης παραμέτρων κατανομών για τη μεταβλητή $Y = I b(d)$.

- 3.2 Αν ορισμένες παράμετροι της συνάρτησης περιόδου επαναφοράς $a(T)$ έχουν ήδη καθοριστεί στο βήμα 2, τότε αναπροσαρμόζονται κατάλληλα οι εξισώσεις εκτίμησης παραμέτρων για τη μεταβλητή $Y = I b(d)$.

Ως παράδειγμα αναπροσαρμογής των εξισώσεων εκτίμησης παραμέτρων για το βήμα 3.2 αναφέρουμε τη συνηθέστερη περίπτωση που χρησιμοποιείται η κατανομή Gumbel για την περιγραφή των μέγιστων εντάσεων βροχής. Στην περίπτωση αυτή χρησιμοποιείται η εξίσωση (8.17) για την έκφραση της $a(T)$. Αν υποθέσουμε ότι η αδιάστατη παράμετρος ψ έχει καθοριστεί στο βήμα 2 (υποπερίπτωση 2.1 ή 2.2) τότε η παράμετρος λ που απομένει εκτιμάται από τη δεύτερη από τις εξισώσεις (8.18), η οποία οδηγεί στη διατήρηση της μέσης τιμής των ανηγμένων εντάσεων $Y = I b(d)$. Επιλύοντας ως προς λ παίρνουμε

$$\lambda = \frac{\psi + 0.577}{\bar{y}} \quad (8.38)$$

όπου \bar{y} είναι η μέση τιμή των ανηγμένων εντάσεων του ενοποιημένου δείγματος του σταθμού μελέτης.

8.6 Γεωγραφική μεταβλητότητα των όμβριων καμπυλών

Σε πολλές χώρες το θέμα της κατάρτισης όμβριων καμπυλών έχει αντιμετωπιστεί μαζικά για εκτεταμένες γεωγραφικές περιοχές και κατασκευάστηκαν έτσι χάρτες που δίνουν έτοιμα στοιχεία για κάθε περιοχή, βάσει των οποίων μπορεί να καταρτιστούν εύκολα όμβριες καμπύλες σε οποιοδήποτε σημείο, χωρίς να απαιτείται να ανατρέξει κανείς στα πρωτογενή ιστορικά δεδομένα. Για παράδειγμα, στις ΗΠΑ έχουν κατασκευαστεί τέτοιοι χάρτες από το U. S. Weather Bureau που χρονολογούνται από το 1961 (Hershfield, 1961) οι οποίοι έχουν αναπαραχθεί σε πολλά υδρολογικά συγγράμματα (π.χ. Linsley et al., 1975, σ. 358· Viessman et al., 1989, σ. 337· Wanielista, 1990, σ. 59). Οι χάρτες αυτοί απεικονίζουν ισουέτιες καμπύλες μέγιστων υψών βροχής για διάφορες διάρκειες βρο-

χής (30 min - 24 h) και διάφορες περιόδους επαναφοράς (1 - 100). Νεότεροι χάρτες για τις ΗΠΑ έχουν κατασκευαστεί από τη NOAA (Miller et al., 1973 για τις δυτικές ΗΠΑ· Frederick et al., 1977, για τις ανατολικές και κεντρικές ΗΠΑ) και έχουν αναδημοσιευτεί από τον Smith (1993). Για τη Μεγάλη Βρετανία και την Ιρλανδία ανάλογοι χάρτες έχουν κατασκευαστεί από το Institute of Hydrology (NERC, 1975) και έχουν αναδημοσιευτεί και σε διάφορα υδρολογικά συγγράμματα (π.χ. Wilson, 1990, σσ. 278-338). Σε αυτή την περίπτωση οι χάρτες δίνουν αφενός τη βροχόπτωση 2 ημερών περιόδου επαναφοράς 5 ετών και αφετέρου το ποσοστό της βροχόπτωσης 1 ώρας περιόδου επαναφοράς 5 ετών προς την αντίστοιχη των 2 ημερών. Αλλά και σε άλλες χώρες, όπως για παράδειγμα τις Ινδίες, έχουν κατασκευαστεί ανάλογοι χάρτες που δίνουν τη μέγιστη βροχόπτωση 1 ώρας για δεδομένη περίοδο επαναφοράς, π.χ. 50 ετών (UNESCO, 1974· βλ. και Subramanya, 1984, σ. 40).

Στην Ελλάδα έχει γίνει μια πρώτη πιλοτική διερεύνηση προς την κατεύθυνση της κατασκευής ανάλογων χαρτών (Κοζώνης, 1995) με εφαρμογή στη Στερεά Ελλάδα. Η εφαρμογή αυτή αποσκοπούσε περισσότερο στον έλεγχο της προτεινόμενης μεθοδολογίας της ενότητας 8.5, η οποία αφορά στην αξιοποίηση και των δεδομένων από βροχόμετρα, και λιγότερο στην κατασκευή τελικών χαρτών που θα μπορούσαν να χρησιμοποιηθούν επιχειρησιακά. Ο χάρτης που για ενδεικτικούς λόγους αναδημοσιεύεται από την εργασία του Κοζώνη (1995) (Σχ. 8.10) δεν μπορεί να θεωρηθεί ως τελικός επιχειρησιακά αξιοποιήσιμος αλλά ως ένα πρώτο βήμα προς την κατεύθυνση της κατασκευής ενός άτλαντα της χώρας. Πάντως, τα αποτελέσματα της εφαρμογής έδειξαν ότι (α) η προτεινόμενη μέθοδος είναι ικανοποιητική, (β) οι περισσότερες από τις παραμέτρους των όμβριων καμπυλών μπορούν να θεωρηθούν σταθερές σε μεγάλες γεωγραφικές περιοχές και να εκτιμηθούν από το σχετικώς αραιό δίκτυο βροχογράφων, (γ) οι υπόλοιπες παράμετροι μπορούν να εξαχθούν με βάση δεδομένα από το πυκνότερο δίκτυο βροχομέτρων, και (δ) η προτεινόμενη μέθοδος βοηθά στην κατασκευή χαρτών για την έμμεση κατάρτιση όμβριων καμπυλών χωρίς αναδρομή στα πρωτογενή δεδομένα.

Στην εν λόγω πιλοτική εργασία χρησιμοποιήθηκαν τα δεδομένα 13 σταθμών με βροχογράφο, σχεδόν ομοιόμορφα κατανομημένων στα υδατικά διαμερίσματα της Στερεάς Ελλάδας, και 58 σταθμών με βροχόμε-

τρο. Οι διάρκειες που επιλέχθηκαν για τα δείγματα των βροχογράφων είναι 1, 2, 6, 12, 24 και 48 h και για τα δείγματα των βροχομέτρων 24 και 48 h. Στα μέγιστα ύψη βροχής των βροχομέτρων εφαρμόστηκε η διόρθωση του σφάλματος διακριτοποίησης, όπως περιγράφεται στο εδάφιο 8.4.1.

Για την έκφραση των όμβριων καμπυλών χρησιμοποιήθηκε η κατανομή Gumbel, η οποία έδειξε να προσαρμόζεται ικανοποιητικά στο σύνολο των δειγμάτων. Υιοθετήθηκε η ακόλουθη έκφραση όμβριας

$$i = \frac{a(T)}{b(d)} = \frac{\frac{1}{\lambda} \left\{ \psi - \ln \left[-\ln \left(1 - \frac{1}{T} \right) \right] \right\}}{d^\eta} \quad (8.39)$$

που διαθέτει τρεις μόνο παραμέτρους, τις η , ψ και λ .

Στο βήμα 2 της παραπάνω μεθοδολογίας (υποπερίπτωση 2.2) φάνηκε ότι οι παράμετροι η και ψ μπορούν να θεωρηθούν σταθερές ανά ζώνη (υδατικό διαμέρισμα), με τιμές που φαίνονται στο Σχ. 8.10. Προχωρώντας στο βήμα 3 της πιο πάνω μεθοδολογίας (υποπερίπτωση 3.2) υπολογίστηκε για καθέναν από τους βροχομετρικούς σταθμούς η παράμετρος λ (εξίσωση (8.38)) και στη συνέχεια ελέγχθηκε η προσαρμογή της εξίσωσης (8.39) στα δεδομένα, είτε γραφικά είτε με στατιστικές δοκιμές (χ^2). Τα αποτελέσματα του ελέγχου προσαρμογής ήταν γενικώς ικανοποιητικά.

Εναλλακτικά εξετάστηκε και η περίπτωση καθορισμού μιας σταθερής παραμέτρου, της η , από το βήμα 2 (υποπερίπτωση 2.2) και εκτίμησης των άλλων δύο παραμέτρων ψ και λ στο βήμα 3 (υποπερίπτωση 3.1). Τα αποτελέσματα σε αυτή την περίπτωση δεν είχαν ουσιαστικές διαφορές από αυτά της προηγούμενης, γι' αυτό και τελικά υιοθετήθηκε η πρώτη απλούστερη εκδοχή των δύο σταθερών παραμέτρων ανά ζώνη (υδατικό διαμέρισμα).

Αφού τελικά σε κάθε υδατικό διαμέρισμα υπάρχει μόνο μία παράμετρος, η λ , που μεταβάλλεται από θέση σε θέση αρκεί ένας μόνος χάρτης για την απεικόνιση της γεωγραφικής μεταβολής των όμβριων καμπυλών. Ο χάρτης αυτός αποφασίστηκε να δοθεί με τη μορφή των ισοϋετίων για διάρκεια βροχής 24 h και περίοδο επαναφοράς 5 (Σχ. 8.10). Για να κατασκευαστούν οι ισοϋέτιες υπολογίστηκε προηγουμένως σε κάθε θέση βροχομετρικού σταθμού το ύψος της μέγιστης 24ωρης βροχόπτωσης για $T =$

5, με βάση την (8.39) και στη συνέχεια χαράχτηκαν οι ισοϋέτιες με τη βοήθεια συστήματος γεωγραφικής πληροφορίας.

Χρησιμοποιώντας το χάρτη του Σχ. 8.10 μπορούμε να εφαρμόσουμε την αντίστροφη διαδικασία και να εκτιμήσουμε την όμβρια καμπύλη σε οποιοδήποτε σημείο της Στερεάς Ελλάδας. Συγκεκριμένα, εντοπίζουμε το σημείο στο χάρτη του Σχ. 8.10 και βρίσκουμε την τιμή $h_5(24)$ του μέγιστου 24ωρου ύψους βροχής για περίοδο επαναφοράς $T = 5$, ενώ από τον ίδιο χάρτη, ανάλογα με την ένταξη του σημείου σε μία από τις τρεις ζώνες, αποκτούμε τις τιμές των παραμέτρων η και ψ . Τέλος, υπολογίζουμε την παράμετρο λ από την ακόλουθη σχέση, η οποία είναι συνέπεια της (8.39):

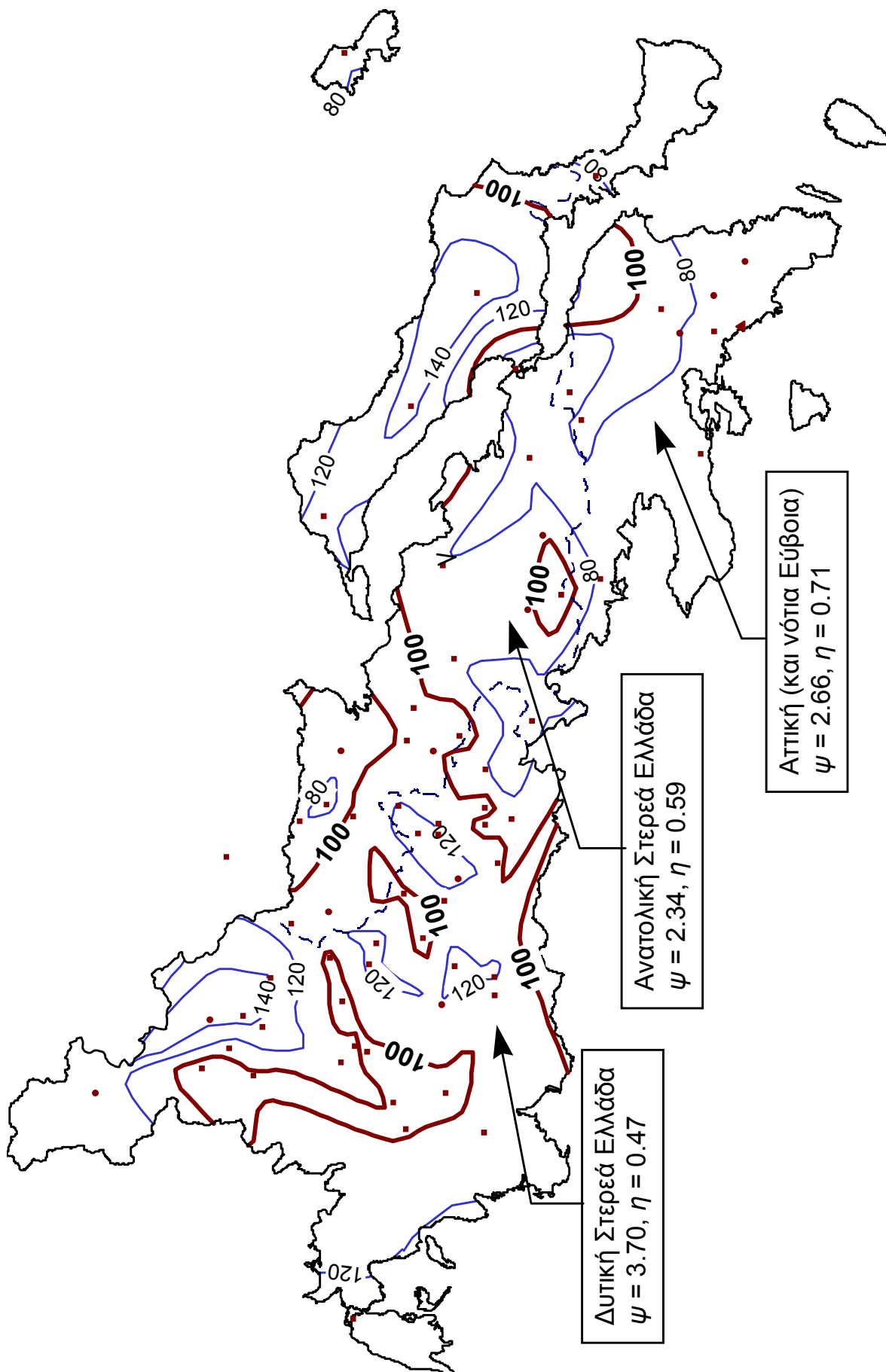
$$\lambda = \frac{24^{1-\eta}}{h_5(24)} [\psi - \ln(-\ln 0.8)] \quad (8.40)$$

Η τελική έκφραση της όμβριας καμπύλης για το υπόψη σημείο προκύπτει με αντικατάσταση στην (8.39) των τιμών των παραμέτρων που καθορίστηκαν με τον παραπάνω τρόπο. Στην έκφραση αυτή η περίοδος επαναφοράς T αντιστοιχεί στη σειρά ετήσιων μεγίστων. Η αντίστοιχη εξίσωση για την περίοδο επαναφοράς T' της σειράς υπεράνω κατωφλίου προκύπτει άμεσα, σύμφωνα με όσα αναφέρθηκαν στο εδάφιο 8.4.3, και είναι

$$i = \frac{\frac{1}{\lambda} (\psi + \ln T')}{d^\eta} \quad (8.41)$$

όπου οι παράμετροι η , ψ και λ έχουν τις ίδιες τιμές όπως στην (8.39).

Σχ. 8.10 (Επόμενη σελίδα) Χάρτης της Στερεάς Ελλάδας με πληροφορίες για την έμμεση εξαγωγή όμβριων καμπυλών. Οι συνεχείς γραμμές είναι οι ισοϋέτιες της μέγιστης 24ωρης βροχόπτωσης (σε mm) για περίοδο επαναφοράς $T = 5$. Οι διακεκομμένες γραμμές παριστάνουν τα όρια των υποπεριοχών με σταθερές παραμέτρους η και ψ . Οι κύκλοι και τα τετράγωνα δείχνουν τις θέσεις των βροχογραφικών και βροχομετρικών σταθμών, αντίστοιχα.



Παράρτημα **Στατιστικοί πίνακες**

Πίν. Π1 Τιμές της τυποποιημένης κανονικής κατανομής.

| z | $F(z)$ | $F_1(z)$ | z | $F(z)$ | $F_1(z)$ | z | $F(z)$ | $F_1(z)$ |
|--------|-----------|----------|--------|-----------|----------|--------|---------------|-----------|
| 0 | 0.5 | 0.5 | 1.35 | 0.91149 | 0.08851 | 2.70 | 0.99653 | 0.00347 |
| 0.05 | 0.51994 | 0.48006 | 1.40 | 0.91924 | 0.08076 | 2.75 | 0.99702 | 0.00298 |
| 0.10 | 0.53983 | 0.46017 | 1.45 | 0.92647 | 0.07353 | 2.80 | 0.99744 | 0.00256 |
| 0.15 | 0.55962 | 0.44038 | 1.50 | 0.93319 | 0.06681 | 2.85 | 0.99781 | 0.00219 |
| 0.20 | 0.57926 | 0.42074 | 1.55 | 0.93943 | 0.06057 | 2.8782 | 0.998 | 0.002 |
| 0.25 | 0.59871 | 0.40129 | 1.60 | 0.94520 | 0.05480 | 2.90 | 0.99813 | 0.00187 |
| 0.2533 | 0.6 | 0.4 | 1.6449 | 0.95 | 0.05 | 2.95 | 0.99841 | 0.00159 |
| 0.30 | 0.61791 | 0.38209 | 1.65 | 0.95053 | 0.04947 | 3.00 | 0.99865 | 0.00135 |
| 0.35 | 0.63683 | 0.36317 | 1.70 | 0.95543 | 0.04457 | 3.05 | 0.99841 | 0.00159 |
| 0.40 | 0.65542 | 0.34458 | 1.75 | 0.95994 | 0.04006 | 3.0902 | 0.999 | 0.001 |
| 0.45 | 0.67364 | 0.32636 | 1.80 | 0.96407 | 0.03593 | 3.10 | 0.99886 | 0.00114 |
| 0.50 | 0.69146 | 0.30854 | 1.85 | 0.96784 | 0.03216 | 3.15 | 0.99900 | 0.00100 |
| 0.5244 | 0.7 | 0.3 | 1.90 | 0.97128 | 0.02872 | 3.20 | 0.99903 | 0.00097 |
| 0.55 | 0.70884 | 0.29116 | 1.95 | 0.97441 | 0.02559 | 3.25 | 0.99918 | 0.00082 |
| 0.60 | 0.72575 | 0.27425 | 2.00 | 0.97725 | 0.02275 | 3.2905 | 0.9995 | 0.0005 |
| 0.65 | 0.74215 | 0.25785 | 2.05 | 0.97982 | 0.02018 | 3.30 | 0.99942 | 0.00058 |
| 0.70 | 0.75804 | 0.24196 | 2.0537 | 0.98 | 0.02 | 3.35 | 0.99950 | 0.00050 |
| 0.75 | 0.77337 | 0.22663 | 2.10 | 0.98214 | 0.01786 | 3.40 | 0.99952 | 0.00048 |
| 0.80 | 0.78814 | 0.21186 | 2.15 | 0.98422 | 0.01578 | 3.45 | 0.99960 | 0.00040 |
| 0.8416 | 0.8 | 0.2 | 2.20 | 0.98610 | 0.01390 | 3.50 | 0.99966 | 0.00034 |
| 0.85 | 0.80234 | 0.19766 | 2.25 | 0.98778 | 0.01222 | 3.5402 | 0.9998 | 0.0002 |
| 0.90 | 0.81594 | 0.18406 | 2.30 | 0.98928 | 0.01072 | 3.55 | 0.99977 | 0.00023 |
| 0.95 | 0.82894 | 0.17106 | 2.3263 | 0.99 | 0.01 | 3.60 | 0.99980 | 0.00020 |
| 1.00 | 0.84134 | 0.15866 | 2.35 | 0.99061 | 0.00939 | 3.65 | 0.99981 | 0.00019 |
| 1.05 | 0.85314 | 0.14686 | 2.40 | 0.99180 | 0.00820 | 3.70 | 0.99984 | 0.00016 |
| 1.10 | 0.86433 | 0.13567 | 2.45 | 0.99286 | 0.00714 | 3.7195 | 0.9999 | 10^{-4} |
| 1.15 | 0.87493 | 0.12507 | 2.50 | 0.99379 | 0.00621 | 4.27 | $1 - 10^{-5}$ | 10^{-5} |
| 1.20 | 0.88493 | 0.11507 | 2.55 | 0.99461 | 0.00539 | 4.75 | $1 - 10^{-6}$ | 10^{-6} |
| 1.25 | 0.89435 | 0.10565 | 2.5758 | 0.995 | 0.005 | 5.20 | $1 - 10^{-7}$ | 10^{-7} |
| 1.2816 | 0.9 | 0.1 | 2.60 | 0.99534 | 0.00466 | 5.61 | $1 - 10^{-8}$ | 10^{-8} |
| 1.30 | 0.90320 | 0.09680 | 2.65 | 0.99598 | 0.00402 | 6.00 | $1 - 10^{-7}$ | 10^{-7} |
| $-z$ | $F_1(-z)$ | $F(-z)$ | $-z$ | $F_1(-z)$ | $F(-z)$ | $-z$ | $F_1(-z)$ | $F(-z)$ |

Παραδείγματα: $F(0.80) = 0.78814$ $F(-3.30) = 0.00058$
 $z_{0.8} = 0.8416$ $z_{0.01} = -2.3263$

Πίν. Π2 Ποσοστημόρια $\chi^2_u(n)$ της κατανομής χ^2 για χαρακτηριστικές τιμές του u και για n βαθμούς ελευθερίας.

| $u =$ | 0.005 | 0.01 | 0.025 | 0.05 | 0.1 | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $n=1$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 2.71 | 3.84 | 5.02 | 6.63 | 7.88 |
| 2 | 0.01 | 0.02 | 0.05 | 0.10 | 0.21 | 4.61 | 5.99 | 7.38 | 9.21 | 10.60 |
| 3 | 0.07 | 0.11 | 0.22 | 0.35 | 0.58 | 6.25 | 7.81 | 9.35 | 11.34 | 12.84 |
| 4 | 0.21 | 0.30 | 0.48 | 0.71 | 1.06 | 7.78 | 9.49 | 11.14 | 13.28 | 14.86 |
| 5 | 0.41 | 0.55 | 0.83 | 1.15 | 1.61 | 9.24 | 11.07 | 12.83 | 15.09 | 16.75 |
| 6 | 0.68 | 0.87 | 1.24 | 1.64 | 2.20 | 10.64 | 12.59 | 14.45 | 16.81 | 18.55 |
| 7 | 0.99 | 1.24 | 1.69 | 2.17 | 2.83 | 12.02 | 14.07 | 16.01 | 18.48 | 20.28 |
| 8 | 1.34 | 1.65 | 2.18 | 2.73 | 3.49 | 13.36 | 15.51 | 17.53 | 20.09 | 21.95 |
| 9 | 1.73 | 2.09 | 2.70 | 3.33 | 4.17 | 14.68 | 16.92 | 19.02 | 21.67 | 23.59 |
| 10 | 2.16 | 2.56 | 3.25 | 3.94 | 4.87 | 15.99 | 18.31 | 20.48 | 23.21 | 25.19 |
| 11 | 2.60 | 3.05 | 3.82 | 4.57 | 5.58 | 17.28 | 19.68 | 21.92 | 24.73 | 26.76 |
| 12 | 3.07 | 3.57 | 4.40 | 5.23 | 6.30 | 18.55 | 21.03 | 23.34 | 26.22 | 28.30 |
| 13 | 3.57 | 4.11 | 5.01 | 5.89 | 7.04 | 19.81 | 22.36 | 24.74 | 27.69 | 29.82 |
| 14 | 4.07 | 4.66 | 5.63 | 6.57 | 7.79 | 21.06 | 23.68 | 26.12 | 29.14 | 31.32 |
| 15 | 4.60 | 5.23 | 6.26 | 7.26 | 8.55 | 22.31 | 25.00 | 27.49 | 30.58 | 32.80 |
| 16 | 5.14 | 5.81 | 6.91 | 7.96 | 9.31 | 23.54 | 26.30 | 28.85 | 32.00 | 34.27 |
| 17 | 5.70 | 6.41 | 7.56 | 8.67 | 10.09 | 24.77 | 27.59 | 30.19 | 33.41 | 35.72 |
| 18 | 6.26 | 7.01 | 8.23 | 9.39 | 10.86 | 25.99 | 28.87 | 31.53 | 34.81 | 37.16 |
| 19 | 6.84 | 7.63 | 8.91 | 10.12 | 11.65 | 27.20 | 30.14 | 32.85 | 36.19 | 38.58 |
| 20 | 7.43 | 8.26 | 9.59 | 10.85 | 12.44 | 28.41 | 31.41 | 34.17 | 37.57 | 40.00 |
| 22 | 8.64 | 9.54 | 10.98 | 12.34 | 14.04 | 30.81 | 33.92 | 36.78 | 40.29 | 42.80 |
| 24 | 9.89 | 10.86 | 12.40 | 13.85 | 15.66 | 33.20 | 36.42 | 39.36 | 42.98 | 45.56 |
| 26 | 11.16 | 12.20 | 13.84 | 15.38 | 17.29 | 35.56 | 38.89 | 41.92 | 45.64 | 48.29 |
| 28 | 12.46 | 13.56 | 15.31 | 16.93 | 18.94 | 37.92 | 41.34 | 44.46 | 48.28 | 50.99 |
| 30 | 13.79 | 14.95 | 16.79 | 18.49 | 20.60 | 40.26 | 43.77 | 46.98 | 50.89 | 53.67 |
| 35 | 17.19 | 18.51 | 20.57 | 22.47 | 24.80 | 46.06 | 49.80 | 53.20 | 57.34 | 60.27 |
| 40 | 20.71 | 22.16 | 24.43 | 26.51 | 29.05 | 51.81 | 55.76 | 59.34 | 63.69 | 66.77 |
| 45 | 24.31 | 25.90 | 28.37 | 30.61 | 33.35 | 57.51 | 61.66 | 65.41 | 69.96 | 73.17 |
| 50 | 27.99 | 29.71 | 32.36 | 34.76 | 37.69 | 63.17 | 67.50 | 71.42 | 76.15 | 79.49 |

Παραδείγματα: $\chi^2_{0.05}(5) = 1.15$ $\chi^2_{0.99}(10) = 23.21$

Για $n \geq 50$: $\chi^2_u(n) = \frac{1}{2} \left(z_u + \sqrt{2n-1} \right)^2$

όπου z_u το u -ποσοστημόριο της κανονικής κατανομής

Πίν. Π3 Ποσοστημόρια $t_u(n)$ της κατανομής t για χαρακτηριστικές τιμές του u και για n βαθμούς ελευθερίας.

| $u =$ | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 |
|----------|------|------|-------|-------|-------|
| $n=1$ | 3.08 | 6.31 | 12.71 | 31.82 | 63.66 |
| 2 | 1.89 | 2.92 | 4.30 | 6.96 | 9.92 |
| 3 | 1.64 | 2.35 | 3.18 | 4.54 | 5.84 |
| 4 | 1.53 | 2.13 | 2.78 | 3.75 | 4.60 |
| 5 | 1.48 | 2.02 | 2.57 | 3.36 | 4.03 |
| 6 | 1.44 | 1.94 | 2.45 | 3.14 | 3.71 |
| 7 | 1.41 | 1.89 | 2.36 | 3.00 | 3.50 |
| 8 | 1.40 | 1.86 | 2.31 | 2.90 | 3.36 |
| 9 | 1.38 | 1.83 | 2.26 | 2.82 | 3.25 |
| 10 | 1.37 | 1.81 | 2.23 | 2.76 | 3.17 |
| 11 | 1.36 | 1.80 | 2.20 | 2.72 | 3.11 |
| 12 | 1.36 | 1.78 | 2.18 | 2.68 | 3.05 |
| 13 | 1.35 | 1.77 | 2.16 | 2.65 | 3.01 |
| 14 | 1.35 | 1.76 | 2.14 | 2.62 | 2.98 |
| 15 | 1.34 | 1.75 | 2.13 | 2.60 | 2.95 |
| 16 | 1.34 | 1.75 | 2.12 | 2.58 | 2.92 |
| 17 | 1.33 | 1.74 | 2.11 | 2.57 | 2.90 |
| 18 | 1.33 | 1.73 | 2.10 | 2.55 | 2.88 |
| 19 | 1.33 | 1.73 | 2.09 | 2.54 | 2.86 |
| 20 | 1.33 | 1.72 | 2.09 | 2.53 | 2.85 |
| 22 | 1.32 | 1.72 | 2.07 | 2.51 | 2.82 |
| 24 | 1.32 | 1.71 | 2.06 | 2.49 | 2.80 |
| 26 | 1.31 | 1.71 | 2.06 | 2.48 | 2.78 |
| 28 | 1.31 | 1.70 | 2.05 | 2.47 | 2.76 |
| 30 | 1.31 | 1.70 | 2.04 | 2.46 | 2.75 |
| 35 | 1.31 | 1.69 | 2.03 | 2.44 | 2.72 |
| 40 | 1.30 | 1.68 | 2.02 | 2.42 | 2.70 |
| 45 | 1.30 | 1.68 | 2.01 | 2.41 | 2.69 |
| 50 | 1.30 | 1.68 | 2.01 | 2.40 | 2.68 |
| ∞ | 1.28 | 1.64 | 1.96 | 2.33 | 2.58 |

Παράδειγμα: $t_{0.95}(5) = 2.02$

Για $n \geq 50$: $t_u(n) \approx z_u \sqrt{\frac{n}{n-2}}$

όπου z_u το u -ποσοστημόριο της κανονικής κατανομής

Πίν. Π4α Ποσοστημόρια k_u της τυποποιημένης κατανομής γάμα για χαρακτηριστικές τιμές του συντελεστή ασυμμετρίας $C_s (\leq 2)$ ή της παραμέτρου σχήματος $k (\geq 1)$.

| $u-r$ | $1-u=r-1$ | $C_s=0$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 | 1.7 | 1.8 | 1.9 | 2.0 |
|--------|-----------|-------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | | $\kappa = \infty$ | 400 | 100 | 44.44 | 25.00 | 16.00 | 11.11 | 8.163 | 6.250 | 4.938 | 4.000 | 3.306 | 2.778 | 2.367 | 2.041 | 1.778 | 1.563 | 1.384 | 1.235 | 1.108 | 1.000 |
| 0.0001 | 0.9999 | -3.72 | -3.51 | -3.30 | -3.10 | -2.90 | -2.71 | -2.53 | -2.35 | -2.18 | -2.03 | -1.88 | -1.75 | -1.63 | -1.52 | -1.42 | -1.33 | -1.25 | -1.18 | -1.11 | -1.05 | -1.00 |
| 0.0002 | 0.9998 | -3.54 | -3.35 | -3.16 | -2.98 | -2.80 | -2.63 | -2.46 | -2.30 | -2.14 | -2.00 | -1.86 | -1.73 | -1.62 | -1.51 | -1.41 | -1.33 | -1.25 | -1.17 | -1.11 | -1.05 | -1.00 |
| 0.0005 | 0.9995 | -3.29 | -3.13 | -2.97 | -2.81 | -2.65 | -2.50 | -2.36 | -2.21 | -2.08 | -1.95 | -1.82 | -1.71 | -1.60 | -1.50 | -1.40 | -1.32 | -1.24 | -1.17 | -1.11 | -1.05 | -1.00 |
| 0.001 | 0.999 | -3.09 | -2.95 | -2.81 | -2.67 | -2.53 | -2.40 | -2.27 | -2.14 | -2.02 | -1.90 | -1.79 | -1.68 | -1.58 | -1.48 | -1.39 | -1.31 | -1.24 | -1.17 | -1.11 | -1.05 | -1.00 |
| 0.002 | 0.998 | -2.88 | -2.76 | -2.64 | -2.52 | -2.40 | -2.28 | -2.17 | -2.06 | -1.95 | -1.84 | -1.74 | -1.64 | -1.55 | -1.46 | -1.38 | -1.30 | -1.23 | -1.17 | -1.10 | -1.05 | -1.00 |
| 0.005 | 0.995 | -2.58 | -2.48 | -2.39 | -2.29 | -2.20 | -2.11 | -2.02 | -1.93 | -1.84 | -1.75 | -1.66 | -1.58 | -1.50 | -1.42 | -1.35 | -1.28 | -1.22 | -1.15 | -1.10 | -1.04 | -0.99 |
| 0.01 | 0.99 | -2.33 | -2.25 | -2.18 | -2.10 | -2.03 | -1.95 | -1.88 | -1.81 | -1.73 | -1.66 | -1.59 | -1.52 | -1.45 | -1.38 | -1.32 | -1.26 | -1.20 | -1.14 | -1.09 | -1.04 | -0.99 |
| 0.02 | 0.98 | -2.05 | -2.00 | -1.94 | -1.89 | -1.83 | -1.78 | -1.72 | -1.66 | -1.61 | -1.55 | -1.49 | -1.44 | -1.38 | -1.32 | -1.27 | -1.22 | -1.17 | -1.12 | -1.07 | -1.02 | -0.98 |
| 0.05 | 0.95 | -1.64 | -1.62 | -1.59 | -1.56 | -1.52 | -1.49 | -1.46 | -1.42 | -1.39 | -1.35 | -1.32 | -1.28 | -1.24 | -1.21 | -1.17 | -1.13 | -1.09 | -1.06 | -1.02 | -0.98 | -0.95 |
| 0.1 | 0.9 | -1.28 | -1.27 | -1.26 | -1.25 | -1.23 | -1.22 | -1.20 | -1.18 | -1.17 | -1.15 | -1.13 | -1.11 | -1.09 | -1.06 | -1.04 | -1.02 | -0.99 | -0.97 | -0.94 | -0.92 | -0.89 |
| 0.2 | 0.8 | -0.84 | -0.85 | -0.85 | -0.85 | -0.86 | -0.86 | -0.86 | -0.86 | -0.86 | -0.85 | -0.85 | -0.85 | -0.84 | -0.84 | -0.83 | -0.83 | -0.82 | -0.81 | -0.80 | -0.79 | -0.78 |
| 0.3 | 0.7 | -0.52 | -0.54 | -0.55 | -0.56 | -0.57 | -0.58 | -0.59 | -0.60 | -0.60 | -0.61 | -0.62 | -0.62 | -0.63 | -0.63 | -0.64 | -0.64 | -0.64 | -0.64 | -0.64 | -0.64 | -0.64 |
| 0.4 | 0.6 | -0.25 | -0.27 | -0.28 | -0.30 | -0.31 | -0.33 | -0.34 | -0.36 | -0.37 | -0.38 | -0.39 | -0.41 | -0.42 | -0.43 | -0.44 | -0.45 | -0.46 | -0.47 | -0.48 | -0.48 | -0.49 |
| 0.5 | 0.5 | 0.00 | -0.02 | -0.03 | -0.05 | -0.07 | -0.08 | -0.10 | -0.12 | -0.13 | -0.15 | -0.16 | -0.18 | -0.20 | -0.21 | -0.23 | -0.24 | -0.25 | -0.27 | -0.28 | -0.29 | -0.31 |
| 0.6 | 0.4 | 0.25 | 0.24 | 0.22 | 0.21 | 0.19 | 0.17 | 0.16 | 0.14 | 0.12 | 0.10 | 0.09 | 0.07 | 0.05 | 0.04 | 0.02 | 0.00 | -0.02 | -0.03 | -0.05 | -0.07 | -0.08 |
| 0.7 | 0.3 | 0.52 | 0.51 | 0.50 | 0.49 | 0.47 | 0.46 | 0.44 | 0.43 | 0.41 | 0.40 | 0.38 | 0.36 | 0.35 | 0.33 | 0.31 | 0.30 | 0.28 | 0.26 | 0.24 | 0.22 | 0.20 |
| 0.8 | 0.2 | 0.84 | 0.84 | 0.83 | 0.82 | 0.82 | 0.81 | 0.80 | 0.79 | 0.78 | 0.77 | 0.76 | 0.75 | 0.73 | 0.72 | 0.71 | 0.69 | 0.68 | 0.66 | 0.64 | 0.63 | 0.61 |
| 0.9 | 0.1 | 1.28 | 1.29 | 1.30 | 1.31 | 1.32 | 1.32 | 1.33 | 1.33 | 1.34 | 1.34 | 1.34 | 1.34 | 1.34 | 1.34 | 1.34 | 1.33 | 1.33 | 1.32 | 1.32 | 1.31 | 1.30 |
| 0.95 | 0.05 | 1.64 | 1.67 | 1.70 | 1.73 | 1.75 | 1.77 | 1.80 | 1.82 | 1.84 | 1.86 | 1.88 | 1.89 | 1.91 | 1.92 | 1.94 | 1.95 | 1.96 | 1.97 | 1.98 | 1.99 | 2.00 |
| 0.98 | 0.02 | 2.05 | 2.11 | 2.16 | 2.21 | 2.26 | 2.31 | 2.36 | 2.41 | 2.45 | 2.50 | 2.54 | 2.58 | 2.63 | 2.67 | 2.71 | 2.74 | 2.78 | 2.81 | 2.85 | 2.88 | 2.91 |
| 0.99 | 0.01 | 2.33 | 2.40 | 2.47 | 2.54 | 2.62 | 2.69 | 2.76 | 2.82 | 2.89 | 2.96 | 3.02 | 3.09 | 3.15 | 3.21 | 3.27 | 3.33 | 3.39 | 3.44 | 3.50 | 3.55 | 3.61 |
| 0.995 | 0.005 | 2.58 | 2.67 | 2.76 | 2.86 | 2.95 | 3.04 | 3.13 | 3.22 | 3.31 | 3.40 | 3.49 | 3.58 | 3.66 | 3.75 | 3.83 | 3.91 | 3.99 | 4.07 | 4.15 | 4.22 | 4.30 |
| 0.998 | 0.002 | 2.88 | 3.00 | 3.12 | 3.24 | 3.37 | 3.49 | 3.61 | 3.73 | 3.85 | 3.97 | 4.09 | 4.21 | 4.32 | 4.44 | 4.55 | 4.67 | 4.78 | 4.89 | 5.00 | 5.11 | 5.21 |
| 0.999 | 0.001 | 3.09 | 3.23 | 3.38 | 3.52 | 3.67 | 3.81 | 3.96 | 4.10 | 4.24 | 4.39 | 4.53 | 4.67 | 4.81 | 4.96 | 5.10 | 5.23 | 5.37 | 5.51 | 5.64 | 5.78 | 5.91 |
| 0.9995 | 0.0005 | 3.29 | 3.46 | 3.62 | 3.79 | 3.96 | 4.12 | 4.29 | 4.46 | 4.63 | 4.80 | 4.97 | 5.13 | 5.30 | 5.47 | 5.63 | 5.80 | 5.96 | 6.12 | 6.28 | 6.44 | 6.60 |
| 0.9998 | 0.0002 | 3.54 | 3.73 | 3.93 | 4.13 | 4.33 | 4.53 | 4.73 | 4.93 | 5.13 | 5.33 | 5.53 | 5.73 | 5.94 | 6.14 | 6.34 | 6.54 | 6.74 | 6.93 | 7.13 | 7.32 | 7.52 |
| 0.9999 | 0.0001 | 3.72 | 3.93 | 4.15 | 4.37 | 4.60 | 4.82 | 5.05 | 5.27 | 5.50 | 5.73 | 5.96 | 6.19 | 6.41 | 6.64 | 6.87 | 7.09 | 7.32 | 7.54 | 7.77 | 7.99 | 8.21 |

Σημείωση: $k_u = (x - \mu_X) / \sigma_X \Leftrightarrow x = \mu_X + \sigma_X k_u$ Παράδειγμα: Για $C_s = 0.5 (\kappa = 16)$: $k_{0.98} = 2.31$

Πίν. Π4β Ποσοστημόρια k_u της τυποποιημένης κατανομής γάμμα για χαρακτηριστικές τιμές του συντελεστή ασυμμετρίας $C_s (\geq 2)$ ή της παραμέτρου σχήματος $k (\leq 1)$.

| $u-r$ | $C_s=2$ | 2.1 | 2.2 | 2.3 | 2.4 | 2.5 | 2.6 | 2.7 | 2.8 | 2.9 | 3 | 3.1 | 3.2 | 3.3 | 3.4 | 3.5 | 3.6 | 3.7 | 3.8 | 3.9 | 4 | |
|-----------|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $1-u=r-1$ | $\kappa=1$ | 0.907 | 0.826 | 0.76 | 0.69 | 0.64 | 0.59 | 0.549 | 0.510 | 0.476 | 0.444 | 0.416 | 0.391 | 0.367 | 0.346 | 0.327 | 0.309 | 0.292 | 0.277 | 0.263 | 0.250 | |
| 0.0001 | 0.9999 | -1.00 | -0.95 | -0.91 | -0.87 | -0.83 | -0.80 | -0.77 | -0.74 | -0.71 | -0.69 | -0.67 | -0.65 | -0.62 | -0.61 | -0.59 | -0.57 | -0.56 | -0.54 | -0.53 | -0.51 | -0.50 |
| 0.0002 | 0.9998 | -1.00 | -0.95 | -0.91 | -0.87 | -0.83 | -0.80 | -0.77 | -0.74 | -0.71 | -0.69 | -0.67 | -0.65 | -0.62 | -0.61 | -0.59 | -0.57 | -0.56 | -0.54 | -0.53 | -0.51 | -0.50 |
| 0.0005 | 0.9995 | -1.00 | -0.95 | -0.91 | -0.87 | -0.83 | -0.80 | -0.77 | -0.74 | -0.71 | -0.69 | -0.67 | -0.65 | -0.62 | -0.61 | -0.59 | -0.57 | -0.56 | -0.54 | -0.53 | -0.51 | -0.50 |
| 0.001 | 0.999 | -1.00 | -0.95 | -0.91 | -0.87 | -0.83 | -0.80 | -0.77 | -0.74 | -0.71 | -0.69 | -0.67 | -0.65 | -0.62 | -0.61 | -0.59 | -0.57 | -0.56 | -0.54 | -0.53 | -0.51 | -0.50 |
| 0.002 | 0.998 | -1.00 | -0.95 | -0.91 | -0.87 | -0.83 | -0.80 | -0.77 | -0.74 | -0.71 | -0.69 | -0.67 | -0.65 | -0.62 | -0.61 | -0.59 | -0.57 | -0.56 | -0.54 | -0.53 | -0.51 | -0.50 |
| 0.005 | 0.995 | -0.99 | -0.95 | -0.91 | -0.87 | -0.83 | -0.80 | -0.77 | -0.74 | -0.71 | -0.69 | -0.67 | -0.65 | -0.62 | -0.61 | -0.59 | -0.57 | -0.56 | -0.54 | -0.53 | -0.51 | -0.50 |
| 0.01 | 0.99 | -0.99 | -0.95 | -0.91 | -0.87 | -0.83 | -0.80 | -0.77 | -0.74 | -0.71 | -0.69 | -0.67 | -0.65 | -0.62 | -0.61 | -0.59 | -0.57 | -0.56 | -0.54 | -0.53 | -0.51 | -0.50 |
| 0.02 | 0.98 | -0.98 | -0.94 | -0.90 | -0.86 | -0.83 | -0.80 | -0.77 | -0.74 | -0.71 | -0.69 | -0.67 | -0.65 | -0.62 | -0.61 | -0.59 | -0.57 | -0.56 | -0.54 | -0.53 | -0.51 | -0.50 |
| 0.05 | 0.95 | -0.95 | -0.91 | -0.88 | -0.85 | -0.82 | -0.79 | -0.76 | -0.74 | -0.71 | -0.69 | -0.67 | -0.64 | -0.62 | -0.61 | -0.59 | -0.57 | -0.56 | -0.54 | -0.53 | -0.51 | -0.50 |
| 0.1 | 0.9 | -0.89 | -0.87 | -0.84 | -0.82 | -0.79 | -0.77 | -0.75 | -0.72 | -0.70 | -0.68 | -0.66 | -0.64 | -0.62 | -0.60 | -0.59 | -0.57 | -0.55 | -0.54 | -0.53 | -0.51 | -0.50 |
| 0.2 | 0.8 | -0.78 | -0.76 | -0.75 | -0.74 | -0.72 | -0.71 | -0.70 | -0.68 | -0.67 | -0.65 | -0.64 | -0.62 | -0.61 | -0.59 | -0.58 | -0.56 | -0.55 | -0.54 | -0.52 | -0.51 | -0.50 |
| 0.3 | 0.7 | -0.64 | -0.64 | -0.64 | -0.63 | -0.63 | -0.62 | -0.62 | -0.61 | -0.60 | -0.60 | -0.59 | -0.58 | -0.57 | -0.56 | -0.55 | -0.54 | -0.53 | -0.52 | -0.51 | -0.50 | -0.49 |
| 0.4 | 0.6 | -0.49 | -0.49 | -0.50 | -0.50 | -0.51 | -0.51 | -0.51 | -0.51 | -0.51 | -0.51 | -0.51 | -0.51 | -0.51 | -0.50 | -0.50 | -0.49 | -0.49 | -0.48 | -0.48 | -0.47 | -0.46 |
| 0.5 | 0.5 | -0.31 | -0.32 | -0.33 | -0.34 | -0.35 | -0.36 | -0.37 | -0.38 | -0.38 | -0.39 | -0.40 | -0.40 | -0.40 | -0.41 | -0.41 | -0.41 | -0.41 | -0.41 | -0.41 | -0.41 | -0.41 |
| 0.6 | 0.4 | -0.08 | -0.10 | -0.12 | -0.13 | -0.15 | -0.16 | -0.18 | -0.19 | -0.20 | -0.22 | -0.23 | -0.24 | -0.25 | -0.26 | -0.27 | -0.28 | -0.29 | -0.29 | -0.30 | -0.31 | -0.31 |
| 0.7 | 0.3 | 0.20 | 0.19 | 0.17 | 0.15 | 0.13 | 0.11 | 0.09 | 0.08 | 0.06 | 0.04 | 0.02 | 0.01 | -0.01 | -0.03 | -0.04 | -0.06 | -0.07 | -0.09 | -0.10 | -0.11 | -0.13 |
| 0.8 | 0.2 | 0.61 | 0.59 | 0.57 | 0.56 | 0.54 | 0.52 | 0.50 | 0.48 | 0.46 | 0.44 | 0.42 | 0.40 | 0.38 | 0.36 | 0.34 | 0.32 | 0.30 | 0.28 | 0.26 | 0.24 | 0.23 |
| 0.9 | 0.1 | 1.30 | 1.29 | 1.28 | 1.27 | 1.26 | 1.25 | 1.24 | 1.22 | 1.21 | 1.20 | 1.18 | 1.16 | 1.15 | 1.13 | 1.11 | 1.10 | 1.08 | 1.06 | 1.04 | 1.02 | 1.00 |
| 0.95 | 0.05 | 2.00 | 2.00 | 2.01 | 2.01 | 2.01 | 2.01 | 2.01 | 2.01 | 2.01 | 2.01 | 2.00 | 2.00 | 1.99 | 1.99 | 1.98 | 1.97 | 1.96 | 1.95 | 1.94 | 1.93 | 1.92 |
| 0.98 | 0.02 | 2.91 | 2.94 | 2.97 | 3.00 | 3.02 | 3.05 | 3.07 | 3.09 | 3.11 | 3.13 | 3.15 | 3.17 | 3.19 | 3.20 | 3.21 | 3.23 | 3.24 | 3.25 | 3.26 | 3.27 | 3.27 |
| 0.99 | 0.01 | 3.61 | 3.66 | 3.71 | 3.75 | 3.80 | 3.85 | 3.89 | 3.93 | 3.97 | 4.01 | 4.05 | 4.09 | 4.12 | 4.16 | 4.19 | 4.22 | 4.26 | 4.29 | 4.31 | 4.34 | 4.37 |
| 0.995 | 0.005 | 4.30 | 4.37 | 4.44 | 4.51 | 4.58 | 4.65 | 4.72 | 4.78 | 4.85 | 4.91 | 4.97 | 5.03 | 5.09 | 5.14 | 5.20 | 5.25 | 5.31 | 5.36 | 5.41 | 5.46 | 5.50 |
| 0.998 | 0.002 | 5.21 | 5.32 | 5.42 | 5.53 | 5.63 | 5.73 | 5.83 | 5.92 | 6.02 | 6.11 | 6.21 | 6.30 | 6.39 | 6.47 | 6.56 | 6.65 | 6.73 | 6.81 | 6.89 | 6.97 | 7.05 |
| 0.999 | 0.001 | 5.91 | 6.04 | 6.17 | 6.30 | 6.42 | 6.55 | 6.67 | 6.79 | 6.91 | 7.03 | 7.15 | 7.27 | 7.38 | 7.50 | 7.61 | 7.72 | 7.83 | 7.94 | 8.04 | 8.15 | 8.25 |
| 0.9995 | 0.0005 | 6.60 | 6.76 | 6.91 | 7.07 | 7.22 | 7.37 | 7.52 | 7.67 | 7.82 | 7.96 | 8.11 | 8.25 | 8.39 | 8.53 | 8.67 | 8.81 | 8.94 | 9.08 | 9.21 | 9.34 | 9.47 |
| 0.9998 | 0.0002 | 7.52 | 7.71 | 7.90 | 8.09 | 8.28 | 8.47 | 8.65 | 8.84 | 9.02 | 9.20 | 9.38 | 9.56 | 9.74 | 9.92 | 10.09 | 10.26 | 10.43 | 10.60 | 10.77 | 10.94 | 11.10 |
| 0.9999 | 0.0001 | 8.21 | 8.43 | 8.65 | 8.87 | 9.09 | 9.30 | 9.51 | 9.73 | 9.94 | 10.15 | 10.36 | 10.56 | 10.77 | 10.97 | 11.18 | 11.38 | 11.57 | 11.77 | 11.97 | 12.16 | 12.35 |

Σημείωση: $k_u = (x - \mu_x) / \sigma_x \Leftrightarrow x = \mu_x + \sigma_x k_u$ Παράδειγμα: Για $C_s = 2.5$ ($\kappa = 0.64$): $k_{0.98} = 3.05$

Αναφορές

- Δέας, Ν., Κατάρτιση και διερεύνηση της χωρικής μεταβολής των όμβριων καμυλών στη Στερεά Ελλάδα, Διπλωματική εργασία, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, 1995.
- Κάκουλος, Θ. Ν., *Στοχαστικές Ανελίξεις*, Πανεπιστήμιο Αθηνών, Αθήνα, 1978.
- Κοζώνης, Δ., Κατάρτιση όμβριων καμυλών με ελλιπή δεδομένα, Εφαρμογή στην περιοχή της Στερεάς Ελλάδας, Διπλωματική εργασία, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, 1995.
- Κουτσογιάννης, Δ. και Α. Μανέτας, Λογισμικό κατάρτισης όμβριων καμυλών, *Εκτίμηση και διαχείριση των υδατικών πόρων της Στερεάς Ελλάδας*, Τεύχος 13, Εθνικό Μετσόβιο Πολυτεχνείο, Τομέας Υδατικών Πόρων, Αθήνα, 1995.
- Κουτσογιάννης, Δ., *Σχεδιασμός Αστικών Δικτύων Αποχέτευσης*, έκδοση 3, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, 1993.
- Μιμίκου, Μ., *Τεχνολογία Υδατικών Πόρων*, Εκδόσεις Παπασωτηρίου, Αθήνα, 1995
- Ξανθόπουλος, Θ. Σ., *Μαθήματα Στατιστικής Υδρολογίας*, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Θεσσαλονίκη, 1975
- Ξανθόπουλος, Θ. Σ., *Εισαγωγή στην Τεχνική Υδρολογία*, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, 1990.
- Benjamin, J. R., and C. A. Cornell, *Probability, Statistics and Decision for Civil Engineers*, McGraw-Hill, 1970.

- Bobée, B., and F. Ashkar, *The Gamma Family and Derived Distributions Applied in Hydrology*, Water Resources Publications, Littleton, Colorado, 1991.
- Bobée, B., and R. Robitaille, Correction of bias in the estimation of the coefficient of skewness, *Water Resour. Res.*, 11(6), 851-854, 1975.
- Bras, R. L., and I. Rodriguez-Iturbe, *Random Functions in Hydrology*, Addison-Wesley, 1985.
- Chow, V. T., D. R. Maidment, and L. W. Mays, *Applied Hydrology*, McGraw-Hill, 1988.
- Cooper, N. G., *From Cardinals to Chaos, Reflections on the life and legacy of Stanislaw Ulam*, Cambridge Univ. Press, 1989.
- Draper, N., and H. Smith, *Applied Regression Analysis*, 2nd edition, John Wiley and Sons, New York, 1981.
- Frederick, R. H., V. A. Meyers and E. P. Auciello, Five- to 60-minute precipitation frequency for the eastern and central United States, NOAA Tech. Mem. NWS HYDRO-35, Washington, DC, 1977.
- Freund, J. E., F. J. Williams, and B. M. Perles, *Elementary Business Statistics, The Modern Approach*, Prentice-Hall, 1988.
- Gardner, M., *Aha! Gotcha: Paradoxes to Puzzle and Delight*, W. H. Freeman and Company, 1982 (Ελληνική μετάφραση Τ. Δημητρίου και Γ. Τρουφάκου, *Η Μαγεία των Παραδόξων*, Τροχαλία, 1989).
- Gumbel, E. J., *Statistics of Extremes*, Columbia University Press, New York, 1958.
- Haan, C. T., *Statistical Methods in Hydrology*, The Iowa State University Press, USA, 1977.
- Hershfield, D. M., Rainfall frequency atlas of the United States for durations from 30 minutes to 24 hours and return periods from 1 to 100 years, U. S. Weather Bureau Technical Paper 40, Washington, DC, 1961.
- Hirsch, R. M., D. R. Helsel, T. A. Cohn, and E. J. Gilroy, Statistical analysis of hydrologic data, in *Handbook of Hydrology*, edited by D. R. Maidment, McGraw-Hill, 1993.
- Kendall, M. G., and A. Stuart, *The Advanced Theory of Statistics*, Vol. 1, Distribution theory, Charles Griffin & Co., London, 1963.

- Kendall, M. G., and A. Stuart, *The Advanced Theory of Statistics*, Vol. 2, Inference and relationship, Third edition, Charles Griffin & Co., London, 1973.
- Kirby, W., Algebraic boundness of sample statistics, *Water Resour. Res.*, 10(2), 220-222, 1974.
- Kirby, W., Computer oriented Wilson-Hilferty transformation that preserves the first three moments and the lower bound of the Pearson type 3 distribution, *Water Resour. Res.*, 8(5), 1251-1254, 1972.
- Kite, G. W., *Frequency and Risk Analyses in Hydrology*, Water Resources Publications, Littleton, Colorado, 1988.
- Kottegoda, N. T., *Stochastic Water Resources Technology*, Macmillan Press, London, 1980.
- Koutsoyiannis, D., A stochastic disaggregation method for design storm and flood synthesis, *Journal of Hydrology*, 156, 193-225, 1994.
- Linsley, R. K. Jr., M. A. Kohler and J. L. H. Paulus, *Hydrology for Engineers*, McGraw-Hill, Tokyo, 2nd edition, 1975.
- Masuyama, M. and Y. Kuroiwa, Table for the likelihood solutions of gamma distribution and its medical applications, *Rev. Stat. Appl. Res. Un. Jap. Sci. Engrs.*, 1, 18-23, 1951.
- Miller, J. F., R. H. Frederick and R. J. Tracey, Precipitation frequency analysis of the western United States, NOAA Atlas 2, National Weather Service, NOAA, U. S. Department of Commerce, Silver Spring, Md. 1973.
- NERC (National Environmental Research Council), *Flood Studies Report*, Institute of Hydrology, Wallingford, 1975.
- Papoulis, A., *Probability and Statistics*, Prentice-Hall, 1990.
- Papoulis, A., *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 1965.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, Cambridge University Press, Cambridge, 1987.
- Raudkivi, A. J., *Hydrology, An Advanced Introduction to Hydrological Processes and Modelling*, Pergamon Press, Oxford, 1979.
- Salas, J. D., Analysis and modeling of hydrologic time series, Chapter 19, *Handbook of Hydrology*, edited by D. Maidment, McGraw-Hill, New York, 1993.

- Salas, J. D., Delleur, J. W., Yevjevich, V., and Lane, W. L., *Applied Modelling of Hydrologic Time Series*, Water Resources Publications, Littleton, Co., USA, 1988.
- Schroeder, M., *Fractals, Chaos, Power Laws, Minutes From an Infinite Paradise*, W. H. Freeman and Co., 1990.
- Singh, V. P., and A. K. Rajagopal, A new method of parameter estimation for hydrologic frequency analysis, *Hydrological Science and Technology*, 2(3) 33-44, 1986.
- Shaw, E. M., *Hydrology in Practice*, Van Nostrand Reinhold, Berkshire, UK, 1983.
- Smith, J. A., Precipitation, Ch. 4 in *Handbook of Hydrology*, edited by D. R. Maidment, McGraw-Hill, New York, 1993.
- Spiegel, M. R., *Theory and Problems of Probability and Statistics*, Schaum's Outline Series, McGraw-Hill, New York, 1975 (Ελληνική μετάφραση Σ. Κ. Περσίδη: *Πιθανότητες και Στατιστική*, ΕΣΠΙ, Αθήνα, 1977).
- Stedinger, J. R., R. M. Vogel, and E. Foufoula-Georgiou, Frequency analysis of extreme events, Chapter 18 in *Handbook of Hydrology*, edited by D. R. Maidment, McGraw-Hill, 1993.
- Stewart, I., *Does God Play Dice?*, Penguin, 1990 (Ελληνική μετάφραση Κ. Σαμαρά, *Παίζει ο Θεός Ζάρια;*, Εκδόσεις Κωσταράκη, 1991).
- Subramaya, K., *Engineering Hydrology*, Tata McGraw-Hill, New Delhi, 1984.
- Taylor, H. M., and S. Karlin, *An Introduction to Stochastic Modelling*, Academic Press, 1984.
- UNESCO, National resources of humid tropical Asia - National resources research XII, UNESCO, 1974.
- U. S. Army Corps of Engineers, *Drainage and Erosion Control*, EM-1110-345-281, 1965.
- Viessman, W. Jr., G. L. Lewis and J. W. Knapp, *Introduction to Hydrology*, 3rd edition, Happer & Row, New York, 1989.
- Wallis, J. R., N. C. Matalas, and J. R. Slack, Just a moment!, *Water Resour. Res.*, 10(2), 211-219, 1974.
- Wanielista, M., *Hydrology and Water Quality Control*, John Wiley & Sons, New York, 1990.

Wilson, E. M., *Engineering Hydrology*, 4th edition, Macmillan, London, 1990.

Yevjevich, V., *Probability and Statistics in Hydrology*, Water Resources Publications, Fort Collins, Colorado, 1972.

Σημείωση: Για την απόδοση όρων της αγγλικής βιβλιογραφίας στην ελληνική χρησιμοποιήθηκαν κατά βάση τα λεξικά:

Καλογεροπούλου, Α., Μ. Γκίκας, Δ. Καραγιαννάκης, και Μ. Λάμπρου, *Αγγλοελληνικό Λεξικό Μαθηματικών Όρων*, Τροχαλία, 1992.

Κοκολάκης, Γ., και Ο. Χρυσάφινου, *Λεξικό Αγγλοελληνικό και Ελληνοαγγλικό Στατιστικών Όρων*, Αθήνα, 1988.

