ΕΙΣΑΓΩΓΗ ΣΤΙΣ ΠΙΘΑΝΟΤΗΤΕΣ ΚΑΙ ΤΗ ΣΤΑΤΙΣΤΙΚΗ

Τάκη Παπαϊωάννου

Τὸ ἐξώφυλλο παριστάνει Αἰγύπτιο εὐγενῆ ὁ ὁποῖος παίζει ἐπιτραπέζιο παιχνίδι μὲ τὴ βοήθεια ἀστράγαλου (Αἰγυπτιακὴ ἐπιτάφια ζωγραφική, Oriental Institute, University of Chicago).

εισαγωγη
στις Πιθανοτητες και τη Στατιστικη

ΤΑΚΗ ΠΑΠΑΙΩΑΝΝΟΥ
Καθηγητη Πανεπιστημιου Ιωαννινων

εισαγωγη στις πιθανοτητες και τη στατιστικη

Ιωαννινα 1981

Κάθε γνήσιο ἀντίτυπο φέρει τήν ὑπογραφή τοῦ συγγραφέα.

PRINTED IN IOANNINA GREECE

ΠΕΡΙΕΧΟΜΕΝΑ

ΜΕΡΟΣ Ι: ΠΙΘΑΝΟΤΗΤΕΣ

Κεφ. 1 ΕΙΣΑΓΩΓΗ

Κεφ. 2 ΠΙΘΑΝΟΤΗΤΕΣ - ΒΑΣΙΚΕΣ ΠΡΟΤΑΣΕΙΣ

ΠΡΟΛΟΓΟΣ

Οἱ Πιθανότητες καί ἡ Στατιστική εἶναι ἕνας ἐνδιαφέρον καί γοητευτικός κλάδος τῶν 'Εφηρμοσμένων Μαθηματικῶν μέ μεγάλη ἀνάπτυξη τόσο στή θεωρία ὅσο καί τίς 'Εφαρμογές. Ἕνας δείκτης τῆς ἀνάπτυξης αὐτῆς εἶναι οἱ πολυάριθμες Σχολές (Departments) ἤ 'Ινστιτοῦτα Στατιστικῆς πού ξεπήδησαν τά τελευταῖα εἴκοσι χρόνια ἀπό Μαθηματικά κυρίως Τμήματα σέ πολλά Πανεπιστήμια ἀνά τόν κόσμο. Στήν χώρα μας οἱ Πιθανότητες καί ἡ Στατιστική ἔχουν μπεῖ στήν ὕλη τοῦ Λυκείου καί ἀπό τίς ἀρχές τῆς δεκαετίας τοῦ 1960 διδάσκονται στά Μαθηματικά Τμήματα τῶν Πανεπιστημίων. Διδάσκονται ἐπίσης σέ ἄλλα Τμήματα τῶν Φυσικομαθηματικῶν Σχολῶν, στά Πολυτεχνεῖα καί τίς Νομικές, Οἰκονομικές, Βιομηχανικές, Γεωπονικές, 'Ιατρικές καί Φιλοσοφικές Σχολές τῶν Πανεπιστημίων.

Οἱ Πιθανότητες καί ἡ Στατιστική ἀποτελοῦν ἕνα ἀπαραίτητο ἐργαλεῖο στά χέρια κάθε ἐρευνητῆ. Δημοσκοπήσεις, ἔρευνες ἀγορᾶς, ἰατρικές μελέτες, ποιοτικοί ἔλεγχοι καί γενικά ἀναλύσεις μετρήσεων (δεδομένων) χρησιμοποιοῦν μεθόδους τῶν Πιθανοτήτων καί τῆς Στατιστικῆς. 'Υπάρχουν πολλά βιβλία Πιθανοτήτων καί Στατιστικῆς, κυρίως στήν ξένη βιβλιογραφία. Τά βιβλία αὐτά καλύπτουν πολλά θέματα καί ἀπευθύνονται σέ διάφορα ἀκροατήρια ἀνάλογα μέ τό θέμα πού ἀσχολοῦνται καί τό ἐπίπεδο μαθηματικῆς ὡριμότητας τοῦ ἀκροατηρίου.

Τό βιβλίο αὐτό ἀποτελεῖ μία εἰσαγωγή στίς Πιθανότητες καί τή Στατιστική, δέν ἀπαιτεῖ προηγούμενη γνώση τοῦ θέματος καί ἀπευθύνεται κυρίως στούς φοιτητές Μαθηματικῶν Τμημάτων Πανεπιστημίων. 'Απευθύνεται ἐπίσης καί στούς φοιτητές Πολυτεχνείων καί ἄλλων Σχολῶν πού ἔχουν τό ἀπαιτούμενο ὑπόβαθρο. Τό μαθηματικό ὑπόβαθρο πού χρειάζεται γιά τήν κατανόηση τῶν έννοιῶν καί μεθόδων του εἶναι οἱ γυμνασιακές μαθηματικές γνώσεις. Σέ ὁρισμένα κεφάλαια χρειάζεται γνώση 'Απειροστικοῦ Λογισμοῦ (παράγωγοι καί ὁλοκληρώματα). Κατά τό μεγαλύτερο του μέρος εἶναι αὐτάρκες. Τά ἐπιχειρήματα του καί ἀποδείξεις του στηρίζονται στή μαθηματική λογική καί τή διαίσθηση, τόν μεγάλο αὐτό παράγοντα πού διακρίνει τίς Πιθανότητες καί τή Στατιστική ἀπό τούς ἄλλους κλάδους τῶν Μαθηματικῶν. Ἡ μαθηματική αὐστηρότητα διατηρεῖται ὅπου αὐτό εἶναι δυνατό. Διαφορετικά τά προβλήματα λύνονται μέ τή διαίσθηση.

Ἡ ἀνάλυση τοῦ βιβλίου κατά κεφάλαιο ἔχει ὡς ἑξῆς: Τό Πρῶτο Κεφάλαιο εἶναι εἰσαγωγικό. Περιγράφει σέ γενικές γραμμές τούς κλάδους τῶν Πιθανοτήτων καί τῆς Στατιστικῆς, εἰσάγει τίς ἔννοιες τοῦ πληθυσμοῦ καί τοῦ δείγματος καί παρουσιάζει μία σειρά ἀπό ἀποτελέσματα τῆς Συνδυαστικῆς πού εἶναι ἀπαραίτητα γιά τό λογιστικό μέρος τῶν Πιθανοτήτων. Στό τέλος περιγράφει σέ γενικές γραμμές τίς δειγματοληψίες ἀπό πεπερασμένους πληθυσμούς.

Οἱ πρῶτες ἔννοιες τῆς θεωρίας τῶν Πιθανοτήτων δίνονται στό Δεύτερο Κεφάλαιο. Ἡ ἔννοια τῆς πιθανότητας ἔχει πολλές θεωρήσεις. Δίνονται οἱ ποσοτικές ἐκεῖνες θεωρήσεις πού ὁδηγοῦν σέ τρόπους ὑπολογισμοῦ τῶν πιθανοτήτων διαφόρων ἐνδεχομένων. Τό οἰκοδόμημα τῆς θεωρίας τῶν Πιθανοτήτων χτίζεται σιγά σιγά ἀξιωματικά καί προχωρεῖ μέ τή βοήθεια τῶν βασικῶν νόμων τῶν πιθανοτήτων. Εἰσάγονται καινούργιες ἔννοιες ὅπως ἡ δεσμευμένη πιθανότητα, ἡ ἀνεξαρτησία, ὁ κανόνας τοῦ Bayes κλπ.

Στό Τρίτο Κεφάλαιο εἰσάγονται οἱ θεμελιώδεις ἔννοιες τῆς τυχαίας μεταβλητῆς καί τῆς κατανομῆς πιθανότητας. Οἱ κατανομές πιθανότητας περιγράφονται μέ τίς ἀθροιστικές συναρτήσεις κατανομῆς, τίς συναρτήσεις πιθανότητας καί τίς συναρτήσεις πυκνότητας πιθανότητας. Καί οἱ τρεῖς αὐτές συναρτήσεις παίζουν πρωταρχικό ρόλο στή θεωρία καί τή μεθοδολογία πού παρουσιάζεται στά ἑπόμενα κεφάλαια.

Στό Τέταρτο Κεφάλαιο παρουσιάζονται διάφορες εἰδικές κατανομές καί ἐφαρμογές αὐτῶν. Οἱ κατανομές αὐτές ἀποτελοῦν πρότυπα (μοντέλα) πιθανοτήτων. Μέ τή βοήθεια τους μποροῦμε νά ὑπολογίζουμε τίς πιθανότητες διαφόρων ἐνδεχομένων χωρίς νά χρειάζεται νά ἀνατρέχουμε κάθε φορά στίς βασικές ἀρχές τῆς θεωρίας τῶν Πιθανοτήτων. Πρῶτα παρουσιάζονται οἱ εἰδικές διακριτές κατανομές καί στή συνέχεια οἱ εἰδικές συνεχεῖς κατανομές.

Στό Πέμπτο Κεφάλαιο παρουσιάζεται ἡ ἔννοια τῆς μαθηματικῆς ἐλπίδας μίας τυχαίας μεταβλητῆς. Παράλληλα ἀναπτύσσονται οἱ ἔννοιες τῆς μέσης τιμῆς, διακύμανσης καί ροπῆς μίας τυχαίας μεταβλητῆς καί μελετοῦνται οἱ ἰδιότητες τους. Στό τέλος περιγράφονται διάφορα ἀριθμητικά ἤ ποσοτικά χαρακτηριστικά τῶν κατανομῶν καί τῶν τυχαίων μεταβλητῶν. Ἀπό αὐτά τό πιό σημαντικό εἶναι ἡ ροπογεννήτρια συνάρτηση ἡ ὁποία ἔχει ἀξιόλογες ἰδιότητες καί μᾶς προσφέρει ἕνα πολύτιμο ἐργαλεῖο γιά τή μελέτη τῆς θεωρίας τῶν κατανομῶν πιθανότητας.

Στό Ἕκτο Κεφάλαιο παρουσιάζονται μερικές χρήσιμες ἔννοιες ἀπό τίς πολυδιάστατες τυχαῖες μεταβλητές καί κατανομές. Κατόπιν δίνεται ἡ ἔννοια τῆς ἀνεξαρτησίας τυχαίων μεταβλητῶν καί στή συνέχεια ἀναπτύσσεται τό θέμα τῆς ἀλλαγῆς μεταβλητῶν γιά μονοδιάστατες τυχαῖες μεταβλητές. Τά ἀποτελέσματα αὐτά, μερικά ἀπό τά ὁποῖα δίνονται χωρίς ἀποδείξεις, ὁλοκληρώνουν τήν Εἰσαγωγή στή θεωρία τῶν Πιθανοτήτων καί τῶν τυχαίων μεταβλητῶν καί ἀνοίγουν τό δρόμο γιά τή Στατιστική.

Τό Ἕβδομο, Ὄγδοο καί Ἔνατο Κεφάλαιο ἀποτελοῦν τό δεύτερο μέρος τοῦ βιβλίου, τήν εἰσαγωγή στή Στατιστική. Ἐδῶ παρουσιάζονται καί ἀναπτύσσονται εἰσαγωγικές ἔννοιες, ἡ Περιγραφική Στατιστική καί τά ἀριθμητικά χαρακτηριστικά τοῦ δείγματος. Στό Ὄγδοο Κεφάλαιο δίνονται τά θεωρητικά θεμέλια τῆς Στατιστικῆς. Εἰσάγονται οἱ ἔννοιες τοῦ στατιστικοῦ, τῆς δειγματικῆς κατανομῆς, οἱ κατανομές X^2, t καί F, ἡ δειγματοληψία ἀπό κανονικούς πληθυσμούς καί τό Κεντρικό Ὁριακό Θεώρημα. Τό Ἔνατο Κεφάλαιο ἀποτελεῖ μία εἰσαγωγή στή Στατιστική Συμπερασματολογία. Ἀναπτύσσονται πρῶτα οἱ ἀρχές οἱ μέθοδοι ἐκτιμήσεως στατιστικῶν παραμέτρων καί στή συνέχεια δίνονται ἡ φιλοσοφία καί μεθοδολογία τῶν στατιστικῶν τέστ. Τέλος δίνεται μία σειρά ἐφαρμογῶν ἡ λεγόμενη ἁπλή Στατιστική Συμπερασματολογία γιά ἕνα ἤ δύο πληθυσμούς.

Στό τέλος τοῦ βιβλίου δίνεται ἡ σχετική βιβλιογραφία καθώς καί στατιστικοί πίνακες.

Οἱ ὁρισμοί, τά θεωρήματα, τά πορίσματα, τά παραδείγματα, τά σχήματα καί οἱ ἀσκήσεις ἔχουν χωριστή ἀρίθμηση μέσα σέ κάθε κεφάλαιο. Ἔτσι θεώρημα 4.3 σημαίνει τό τρίτο θεώρημα στό Τέταρτο Κεφάλαιο. Τό σύμβολο ▼ σημαίνει τό τέλος μίας ἀποδείξεως. Τό βιβλίο περιέχει 101 παραδείγματα πού ἀναπτύσσονται μέ κάθε λεπτομέρεια καί 249 ἀσκήσεις, θεωρητικές καί ἐφηρμοσμένες πού ἐμφανίζονται στό τέλος κάθε κεφαλαίου.

Τό βιβλίο αὐτό εἶναι προϊόν τῶν παραδόσεων μου σέ φοιτητές Πανεπιστημίων καί Πολυτεχνείων στήν Ἑλλάδα καί στό ἐξωτερικό γιά πολλά χρόνια. Στόν Ἑλληνικό χῶρο ξεκίνησε ὑπό μορφή σημειώσεων γιά τούς πρωτοετεῖς καί δευτεροετεῖς φοιτητές τοῦ Μαθηματικοῦ Τμήματος τοῦ Πανεπιστημίου Ἰωαννίνων τό 1976. Ἀπό τότε οἱ σημειώσεις αὐτές δουλεύτηκαν καί ζυμώθηκαν μέ τήν βοήθεια πολλῶν εὐστόχων παρατηρήσεων τόσον τῶν ἐπιμελητῶν καί βοηθῶν τοῦ Ἐργαστηρίου Πιθανοτήτων καί Στατιστικῆς τοῦ Πανεπιστημίου Ἰωαννίνων ὅσο καί τῶν φοιτητῶν μου. Σέ ὅλους αὐτούς ἀπευθύνω τίς εὐχαριστίες μου.

Ἰδιαίτερα εὐχαριστῶ τόν Ὑφηγητή κ. Παναγιώτη Βασιλείου καί τούς Ἐπιμελητές κ.κ. Σωτήριο Παπαχρήστου καί Κοσμᾶ Φερεντῖνο. Ἰδιαίτερες εὐχαριστίες ἀπευθύνονται πρός τόν Διδάκτορα τοῦ Πανεπιστημίου τοῦ Bradford κ. Σωτήριο Λουκᾶ ὁ ὁποῖος ἀνέγνωσε μέ προσοχή τά δοκίμια καί συνέταξε τό εὑρετήριο, τόν Ἐπιμελητή κ. Σωτήριο Παπαχρήστου μέ τήν ἐπιμέλεια καί συνεργασία τοῦ ὁποίου βγῆκαν οἱ πρῶτες σημειώσεις τῶν παραδόσεων μου καί τήν παρασκευάστρια τοῦ Ἐργαστηρίου δ. Κατερίνα Χρηστίδη γιά τήν ἀποκρυπτογράφηση τῶν χειρογράφων μου καί τήν προσεκτική δακτυλογράφηση τοῦ κειμένου. Πάνω ἀπό ὅλους εὐχαριστῶ τήν σύζυγο μου Μαρία γιά τήν ἀμέριστο συμπαράσταση της καί τήν ἐπιλογή τοῦ ἐξωφύλλου.

Ἰωάννινα, Ἀπρίλιος 1981 — Τάκης Παπαϊωάννου

Μερος I: ΠΙΘΑΝΟΤΗΤΕΣ

ΚΕΦΑΛΑΙΟ 1

εισαγωγη

1.1 Γενικότητες

Θεωρία Πιθανοτήτων εἶναι ὁ κλάδος ὁ ὁποῖος ἀσχολεῖται μέ τήν ἔννοια τῆς ἀβεβαιότητας (πιθανότητας).

Στατιστική εἶναι ὁ κλάδος ὁ ὁποῖος ἀσχολεῖται μέ τήν σχεδίαση πειραμάτων ἤ μεθόδων δειγματοληψίας , τή συλλογή καί ἀνάλυση ἀριθμητικῶν δεδομένων (μετρήσεων)καί τήν ἐξαγωγή συμπερασμάτων γιά ἕνα σύνολο βάσει τῶν πληροφοριῶν πού περιέχονται σ' ἕνα δεῖγμα ἀπό τό σύνολο αὐτό.

Ὁ Στατιστικός ἐνδιαφέρεται γιά τήν ἐξαγωγή συμπερασμάτων κατά τόν πλέον καλύτερο τρόπο καθώς καί γιά τήν ἀριθμητική ἀξιολόγηση τῆς καταλληλότητας τῆς μεθόδου ἐξαγωγῆς συμπερασμάτων πού χρησιμοποιεῖ.

Ἡ ἐξαγωγή συμπερασμάτων μπορεῖ νά ἐξομοιωθεῖ μέ τήν λήψη ἀποφάσεων γιά ἕνα στατιστικό πρόβλημα. Περιλαμβάνει: τόν ὑπολογισμό (ἐκτίμηση) ἀγνώστων στοιχείων, τόν ἔλεγχο στατιστικῶν ὑποθέσεων καί τήν πρόγνωση ἤ πρόβλεψη (prediction).

Ἡ θεωρία Πιθανοτήτων ἐνδιαφέρεται γιά τή διατύπωση νόμων ἤ προτύπων (μοντέλων) πιθανότητας πού διέπουν διάφορα φαινόμενα. Πολλοί ἀπό τούς νόμους αὐτούς διατυποῦνται βάσει ἐξιδανικεύσεων ἤ γενικῶν ἀρχῶν. Ἡ Στατιστική μέ τά δείγματά της καλεῖται νά ἐπιβεβαιώσει τήν ὀρθότητα ἤ μή τῶν νόμων αὐτῶν. Ἐπί πλέον ἡ θεωρία Πιθανοτήτων προσφέρει τή βάση γιά τή δημιουργία στατιστικῶν μεθόδων καί τήν ἀξιολόγηση τῶν στατιστικῶν συμπερασμάτων.

Ἡ θεωρία Πιθανοτήτων ὡς μαθηματική ἐπιστήμη ἔχει χαρακτήρα παρόμοιο μέ ἐκεῖνον τῆς Γεωμετρίας ἤ Ἀναλυτικῆς Μηχανικῆς. Ὁ χαρακτήρας αὐτός ἔχει τρεῖς μορφές: τό αὐστηρό λογικό περιεχόμενο, τή διαίσθηση καί τίς ἐφαρμογές. Τό λογικό περιεχόμενο ἐμφανίζεται στήν ἀξιωματική θεμελίωση τῆς ἔννοιας τῆς πιθανότητας καί τά διάφορα θεωρητικά ἀποτελέσματα τοῦ κλάδου. Ἡ διαίσθηση εἶναι διάχυτη τόσο στή θεωρία ὅσο καί τίς ἐφαρμογές. Οἱ δέ ἐφαρμογές εἶναι πολλές καί ποικίλες σέ πολύπλευρους κλάδους.

Ἡ θεωρία Πιθανοτήτων ξεκίνησε ἀπό τήν μελέτη προβλημάτων πού παρουσιαζόντουσαν στά τυχερά παιγνίδια ἐπιδιώκοντας κατά κύριο λόγο τόν ὑπολογισμό ὁρισμένων πιθανοτήτων. Σήμερα ἔχει ἐξελειχθεῖ σέ μία γενική θεωρία τῆς ὁποίας στόχος εἶναι ἡ ἀνακάλυψη νέων γενικῶν νόμων καί ἡ κατασκευή ἱκανοποιητικῶν θεωρητικῶν μοντέλων.

Ὁ κλάδος τῆς Στατιστικῆς ἔλαβε μεγάλη ἀνάπτυξη κατά τά τελευταῖα *30-40* χρόνια καί συντέλεσε στήν προώθηση τῆς ποσοτικῆς σκέψης σέ ὅλους τούς κλάδους τῆς πνευματικῆς δραστηριότητας. Πατέρας της θεωρεῖται ὁ Ἄγγλος R. Fisher.

Στό εἰσαγωγικό αὐτό Κεφάλαιο παρουσιάζουμε τίς πρῶτες βασικές ἔννοιες τῆς θεωρίας Πιθανοτήτων καί ὁρισμένα στοιχεῖα ἀπό τή Συνδυαστική τά ὁποῖα θά φανοῦν χρήσιμα στόν ὑπολογισμό τῶν πιθανοτήτων. Βασικές μαθηματικές ἔννοιες ὅπως ἡ συνάρτηση, ἡ ἀκολουθία, τό σύμβολο τοῦ ἀθροίσματος Σ καί στοιχεῖα ἀπό τή θεωρία Συνόλων θεωροῦνται γνωστά.

1.2 Πληθυσμός — Δεῖγμα

Οἱ πρῶτες βασικές ἔννοιες τῆς θεωρίας Πιθανοτήτων εἶναι οἱ ἔννοιες τοῦ πληθυσμοῦ καί τοῦ δείγματος:

Π λ η θ υ σ μ ό ς εἶναι ἕνα ὁποιοδήποτε σύνολο ἀτόμων ἤ ἀντικειμένων ἤ ἄλλων ἀφηρημένων ἤ μή ὀντοτήτων γιά τό ὁποῖο ἐνδιαφερόμαστε νά μελετήσουμε ὁρισμένα χαρακτηριστικά.

Δ ε ῖ γ μ α εἶναι ἕνα μέρος (ὑποσύνολο) τοῦ πληθυσμοῦ τό ὁποῖο συλλέγεται κατά κάποιο τρόπο.

Οἱ τρόποι μέ τούς ὁποίους συλλέγονται (ἐκλέγονται) τά δείγματα εἶναι διάφοροι: συστηματικός, τυχαῖος, μέ πιθανότητες, μέ ἐπανατοποθέτηση, χωρίς ἐπανατοποθέτηση κ.λ.π. Ἐδῶ θά ἀσχοληθοῦμε μόνο μέ τυχαῖα δείγματα τά ὁποῖα ὁρίζονται καί ἀναπτύσσονται στό Κεφάλαιο 7. Οἱ ἄλλοι τρόποι δειγματοληψίας ἀποτελοῦν ἀντικείμενο μελέτης εἰδικοῦ κλάδου τῆς Στατιστικῆς πού λέγεται Θεωρία Δειγματοληψίας.

Παραδείγματα πληθυσμῶν καί δειγμάτων ὑπάρχουν πολυάριθμα: τά σύνολα τῶν ἀτόμων μίας χώρας, πόλης, ἱδρύματος, ἐπαγγέλματος κ.λ.π. γιά τά ὁποῖα ἐνδιαφερόμαστε νά μάθουμε ἄν π.χ. υἱοθετοῦν κάποιο θέμα, ἤ ἄν καπνίζουν καί πάσχουν ἀπό καρκίνο ἤ ποία εἶναι ἡ σύσταση τους ἀπό πλευρᾶς φύλου, ἡλικίας, μορφώσεως κ.λ.π. εἶναι πληθυσμοί. Δείγματα εἶναι τά μέλη τῶν πληθυσμῶν πού χρησιμοποιοῦνται στή δημοσκόπηση. Τό σύνολο τῶν καταστημάτων μίας χώρας εἶναι πληθυσμός καί τά καταστήματα στά ὁποῖα παίρνονται τιμές γιά τόν προσδιορισμό τοῦ τιμαρίθμου εἶναι τό δεῖγμα. Τό σύνολο τῶν φιαλιδίων πενικιλίνης πού προμηθεύεται μία Κρατική Ὑπηρεσία εἶναι πληθυσμός. Ἐδῶ ἐνδιαφερόμαστε ἄν τά φιαλίδια πληροῦν τίς προδιαγραφές. Στό ποιοτικό ἔλεγχο παίρνουμε δείγματα καί τά ἐξετάζουμε ὡς πρός τίς προδιαγραφές καί ἀνάλογα ἀποφασίζουμε ἄν ἡ παρτίδα θά γίνει ἀποδεκτή ἤ ὄχι.

Οἱ ἀριθμοί τῶν τηλεφωνικῶν κλήσεων πού μποροῦν νά φθάσουν σέ ἕνα τηλεφωνικό κέντρο ἕνα ὁρισμένο χρονικό διάστημα ἀποτελοῦν ἕνα πληθυσμό. Οἱ ἀριθμοί αὐτοί θεωρητικά μπορεῖ νά εἶναι $0, 1, 2, \ldots$. Οἱ δυνατοί χρόνοι ζωῆς μίας συσκευῆς εἶναι πληθυσμός. Οἱ χρόνοι αὐτοί θεωρητικά μπορεῖ νά εἶναι ὁποιοσδήποτε ἀριθμός μεταξύ 0 καί ∞.

Τό δεῖγμα εἶναι περισσότερο ἔννοια τῆς Στατιστικῆς καί χρησιμοποιεῖται ὅταν εἶναι ἀδύνατο ἤ δύσκολο ἤ πολυέξοδο νά ἐξεταστοῦν ἕνα-ἕνα ὅλα τά στοιχεῖα τοῦ πληθυσμοῦ γιά νά πάρουμε τίς πληροφορίες πού θέλουμε.

1.3 Συνδυαστική

Θ ε ώ ρ η μ α 1.1: Κανόνας mn.

<u>Ἔστω ἕνα σύνολο</u> $A=\{a_1,a_2,\ldots,a_m\}$ <u>μέ</u> m <u>στοιχεῖα καί ἄλλο ἕνα σύνολο</u> $B=\{b_1,b_2,\ldots,b_n\}$ <u>μέ</u> n <u>στοιχεῖα. Τότε ὁ ἀριθμός ὅλων τῶν δυνατῶν ζευγῶν τύπου</u> (a_i, b_j), $i=1,$

$\ldots, m$, $j=1,\ldots,n$ εἶναι mn.

Ἀπόδειξη: Ἕνα στοιχεῖο a_i τοῦ A μαζί μέ κάθε στοιχεῖο τοῦ B δίδει n ζεύγη $(a_i,b_1),\ldots(a_i,b_n)$. Ἔτσι τά m στοιχεῖα τοῦ A δίδουν mn ζεύγη.▼

Τό ἑπόμενο θεώρημα γενικεύει τόν κανόνα mn γιά περισσότερα ἀπό δύο σύνολα καί ἡ ἀπόδειξη του γίνεται μέ τόν ἴδιο συλλογισμό.

Θεώρημα 1.2: Γενίκευση τοῦ κανόνα mn

Ἔστω τά σύνολα:

$A_1=\{a_{11}, a_{12},\ldots,a_{1n_1}\}$ μέ n_1 στοιχεῖα,

$A_2=\{a_{21}, a_{22},\ldots,a_{2n_2}\}$ μέ n_2 στοιχεῖα, ...,

$A_k=\{a_{k1}, a_{k2},\ldots,a_{kn_k}\}$ μέ n_k στοιχεῖα .

Τότε ὁ ἀριθμός τῶν k-δων τύπου$(a_{1j_1},a_{2j_2},\ldots, a_{kj_k})$ $j_1=1,\ldots,n_1$, $j_2=1,\ldots,n_2,\ldots$, $j_k=1,\ldots,n_k$ εἶναι $n_1n_2\ldots n_k$.

Ὁρισμοί:

1.1 **Μετάθεση** n (διακεκριμένων) στοιχείων λέγεται κάθε τοποθέτησή τους σέ σειρά πάνω σέ εὐθεῖα γραμμή.

1.2 **Διάταξη** n (διακεκριμένων) στοιχείων ἀνά r, $1\le r\le n$, λέγεται κάθε τοποθέτηση r στοιχείων ἀπό τά n σέ σειρά πάνω σέ εὐθεῖα γραμμή.

Εἶναι φανερό ὅτι ἡ μετάθεση n στοιχείων εἶναι εἰδική περίπτωση τῶν διατάξεων n στοιχείων ἀνά r μέ

$r=n$. Σέ πολλά προβλήματα μᾶς ἐνδιαφέρει ὁ ἀριθμός τῶν διατάξεων πού μποροῦν νά γίνουν ἀπό n στοιχεῖα.

Θεώρημα 1.3:

Ἔστω P^n_r (ἤ $(n)_r$ ἤ $[^n_r]$) τό πλῆθος (ἀριθμός) τῶν διατάξεων n διακεκριμένων στοιχείων ἀνά r. Τότε

$$P^n_r = n(n-1)\ldots(n-r+1) = \frac{n!}{(n-r)!}$$

Ἀπόδειξη: Γιά τήν ἀπόδειξη χρησιμοποιεῖται τό θεώρημα 1.2. Μιά ὁποιαδήποτε διάταξη μπορεῖ νά κατασκευασθεῖ ἀφοῦ ἐκλέξουμε τό πρῶτο στοιχεῖο ἀπό τό σύνολο τῶν n στοιχείων, τό δεύτερο ἀπό τό σύνολο τῶν $n-1$ στοιχείων πού ἀπομένουν κ.ο.κ. Ἐφαρμόζοντες τό θεώρημα 1.2 παίρνουμε τόν πρός ἀπόδειξη τύπο. ▼

Πόρισμα 1.1: Τό πλῆθος P^n_n τῶν μεταθέσεων n διακεκριμένων στοιχείων εἶναι

$$P^n_n = n(n-1)\ldots2\cdot1 = n!$$

Ὁρισμός 1.3: Κάθε σύνολο r στοιχείων, $1\leq r\leq n$, ἀπό n (διακεκριμένα) στοιχεῖα λέγεται **συνδυασμός** τῶν n στοιχείων ἀνά r.

Θεώρημα 1.4 : Διωνυμικοί συντελεστές.

Ἔστω $\binom{n}{r}$ (ἤ (C^n_r)) τό πλῆθος (ἀριθμός) τῶν συνδυασμῶν n διακεκριμένων στοιχείων ἀνά r. Τότε

$$\binom{n}{r} = C^n_r = \frac{n!}{r!(n-r)!} = \frac{P^n_r}{r!}$$

'Απόδειξη: 'Από κάθε ένα συνδυασμό μπορούμε νά δημιουργήσουμε $r!$ διατάξεις μέ άντιμετάθεση τῶν r στοιχείων αὐτοῦ. "Ετσι άπό τούς $\binom{n}{r}$ συνδυασμούς τῶν n ἀνά r μπορούμε νά δημιουργήσουμε $\binom{n}{r}.r!$ διατάξεις πού εἶναι καί τό σύνολο ὅλων τῶν δυνατῶν διατάξεων τῶν n στοιχείων ἀνά r . "Ετσι θά ἔχουμε

$$\binom{n}{r}\cdot r! = P_r^n$$

ἀπό ὅπου προκύπτει ὁ πρός ἀπόδειξη τύπος. ▼

Στίς μεταθέσεις καί διατάξεις μᾶς ἐνδιαφέρει τόσο ἡ σειρά ὅσο καί τό ποιά στοιχεῖα εἶναι τοποθετημένα πάνω στή γραμμή. Στούς συνδυασμούς μᾶς ἐνδιαφέρει μόνο τό εἶδος τῶν στοιχείων καί ὄχι ἡ θέση τους. "Ετσι δύο διατάξεις πού ἔχουν τά ἴδια στοιχεῖα σέ διαφορετική θέση (διαφορετική σειρά) θεωροῦνται διάφοροι ἐνῶ δύο συνδυασμοί μέ τά ἴδια στοιχεῖα σέ διαφορετικές θέσεις θεωροῦνται ταυτόσημοι (ἴδιοι).

Οἱ συνδυασμοί μποροῦν νά χρησιμοποιηθοῦν γιά νά εὕρουμε τούς τρόπους μέ τούς ὁποίους μποροῦμε νά χωρίσουμε ἕνα σύνολο ἀπό n στοιχεῖα σέ δύο ὁμάδες, ἡ μία μέ r στοιχεῖα καί ἡ ἄλλη μέ $n-r$ στοιχεῖα. 'Ο ἀριθμός τῶν τρόπων μέ τούς ὁποίους μποροῦμε νά πετύχουμε τόν παραπάνω χωρισμό εἶναι $\binom{n}{r}$.

Μέ τήν βοήθεια τῶν συνδυασμῶν τό γνωστό θεώρημα τοῦ Νεύτωνος (διωνυμικός τύπος) γράφεται

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^r b^{n-r} \tag{1.1}$$

Θεώρημα 1.5:

Γιά ὅλους τούς φυσικούς ἀριθμούς n, r $n \geq r$ ἔχουμε

(a) $\binom{n}{r} = \binom{n}{n-r}$

(b) $\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$

'Απόδειξη: 'Η σχέση (a) εἶναι προφανής. 'Η σχέση (b) μπορεῖ νά ἀποδειχθεῖ εὔκολα ἄν ἀναπτύξουμε τό δεύτερο μέλος καί ἐκτελέσουμε ἁπλές πράξεις. 'Αντί αὐτοῦ θά χρησιμοποιήσουμε μία ἄλλη τεχνική πού συχνά χρησιμποιεῖται στούς διωνυμικούς συντελεστές. Εἶναι γνωστό ὅτι

$$(1+x)^n \equiv (1+x)^{n-1} + x(1+x)^{n-1} .$$

Ἄν ἀναπτύξουμε τό πρῶτο καί δεύτερο μέλος σύμφωνα μέ τόν διωνυμικό τύπο παίρνουμε

$$\sum_{r=0}^{n} \binom{n}{r} x^r \equiv \sum_{r=0}^{n-1} \binom{n-1}{r} x^r + x \sum_{r=0}^{n-1} \binom{n-1}{r} x^r$$

$$\equiv \sum_{r=0}^{n} \binom{n-1}{r} x^r + \sum_{r=0}^{n} \binom{n-1}{r-1} x^r$$

$$\equiv \sum_{r=0}^{n} \left[\binom{n-1}{r} + \binom{n-1}{r-1}\right] x^r$$

καί ἐξισώνοντας τούς συντελεστές τῶν ἴσων δυνάμεων τοῦ x λαμβάνουμε τήν (b): ['Υπενθυμίζουμε ὅτι $\binom{n}{0} = \binom{n}{n} = 1$ καί $\binom{n}{r} = 0$ ἄν $r<0$ ἤ $r>n$]. ▼

Θεώρημα 1.6: Πολυωνυμικοί συντελεστές

"Εστω οἱ ἀκέραιοι καί θετικοί ἀριθμοί $r_1, r_2, \ldots, r_k$ μέ $r_1+r_2+\cdots+r_k = n$. Ὁ ἀριθμός τῶν τρόπων μέ τούς ὁποίους ἕνας πληθυσμός ἀπό n διακεκριμένα στοιχεῖα μπορεῖ νά χωριστεῖ σέ k μέρη, τό πρῶτο μέ r_1 στοιχεῖα, τό δεύτερο μέ r_2, κ.λ.π. εἶναι

$$C^n_{r_1,r_2,\ldots,r_k} = \frac{n!}{r_1!r_2!\ldots r_k!} =$$

$$=\binom{n}{r_1}\binom{n-r_1}{r_2}\ldots\binom{n-r_1-\cdots-r_{k-2}}{r_{k-1}}$$

Ἀπόδειξη: Ἀπό τά n στοιχεῖα, r_1 μποροῦν νά ἐπιλεγοῦν μέ $\binom{n}{r_1}$ τρόπους. Ἀπό τά ὑπόλοιπα $n-r_1$ στοιχεῖα, r_2 στοιχεῖα, γιά τό δεύτερο μέρος, μποροῦν νά ἐπιλεγοῦν μέ $\binom{n-r_1}{r_2}$ τρόπους, κ.ο.κ. Στό τέλος πρέπει νά ἐπιλέξουμε r_{k-1} στοιχεῖα ἀπό τά ὑπόλοιπα $n-r_1-\cdots-r_{k-2}$ καί ὁ χωρισμός ἐπιτυγχάνεται ἀφοῦ ἔτσι ἀπομένουν r_k ἀκριβῶς στοιχεῖα. Ἡ τελευταία ἐπιλογή γίνεται μέ $\binom{n-r_1-\cdots-r_{k-2}}{r_{k-1}}$ τρόπους. Τό ἀποτέλεσμα τοῦ θεωρήματος ἀποδεικνύεται ἐφαρμόζοντες στή συνέχεια τόν γενικευμένο κανόνα $m\cdot n$. ▼

Οἱ ποσότητες $C^n_{r_1,r_2,\ldots,r_k}$ τοῦ θεωρήματος 1.6 λέγονται καί πολυωνυμικοί συντελεστές διότι ἐμφανίζονται στό γνωστό πολυωνυμικό θεώρημα.

$$(a_1+a_2+\ldots+a_k)^n = \sum_{\substack{r_1,r_2,\ldots,r_k \\ r_1+r_2+\ldots+r_k=n}} \frac{n!}{r_1!\ldots r_k!}\, a_1^{r_1} a_2^{r_2}\ldots a_k^{r_k} \tag{1.2}$$

'Εφαρμογή 1.1: Διατάξεις μή μ ή δ ι α κ ε - κ ρ ι μ έ ν α στοιχεῖα.

"Εστω τώρα ἡ περίπτωση n μή διακεκριμένων στοιχείων καί ἄς ὑποθέσουμε ὅτι ἐνδιαφερόμαστε νά εὕρουμε τόν ἀριθμό τῶν δυνατῶν καί διαφορετικῶν διατάξεων τῶν n στοιχείων ἀνά n, δηλαδή τῶν μεταθέσεων. Οἱ μεταθέσεις αὐτές λέγονται καί ἐ π α ν α λ η π τ ι κ έ ς.
"Εστω ὅτι ἀπό τά n στοιχεῖα, r_1 εἶναι ὅμοια μεταξύ τους, τύπου a_1,r_2 ὅμοια ἐπίσης μεταξύ τους, τύπου a_2, ... καί τά τελευταῖα r_k εἶναι ὅμοια μεταξύ τους στοιχεῖα, τύπου a_k, μέ $r_1+r_2+\ldots+r_k = 0$. 'Ο ἀριθμός τῶν διατάξεων τῶν n μή διακεκριμένων στοιχείων ἀνά n εἶναι

$$\frac{n!}{r_1!r_2!\ldots r_k!} \cdot$$

'Η ἀπόδειξη τῆς προτάσεως αὐτῆς εἶναι φανερή ἄν σκεφθεῖ κανείς ὅτι μία διάταξη τοῦ ἀνωτέρω τύπου δέν εἶναι τίποτε ἄλλο παρά ἕνα χώρισμα (διαμέριση) ἑνός πληθυσμοῦ n στοιχείων σέ k μέρη μέ στοιχεῖα $r_1,r_2,\ldots,r_k$ στά μέρη $1,2,\ldots,k$ ἀντίστοιχα. Στή συνέχεια ἐφαρμόζεται τό θεώρημα 1.6. Παράλληλα μποροῦμε νά σκεφθοῦμε καί ὡς ἑξῆς: ἔστω x ὁ ἄγνωστος ἀριθμός τῶν ἀνωτέρω διατάξεων. Κάθε μία ἀπό αὐτές περιέχει r_1 ὅμοια μεταξύ τους στοιχεῖα, r_2 ὅμοια μεταξύ τους στοιχεῖα κ.ο.κ. "Αν κάνουμε ὁποιαδήποτε ἀλλαγή στίς θέσεις τῶν ὁμοίων στοιχείων ένός τύπου χωρίς νά πειράξουμε τά ἄλλα στοιχεῖα τῶν διατάξεων θά φτιάξουμε μία νέα διάταξη ἴδια μέ τήν ἀρχική. Βάσει τοῦ θεωρήματος 1.2 ὁ ἀριθμός τῶν διατάξεων πού γίνονται μέ τόν τρόπο αὐτό καί πού εἶναι ἴδιες μέ τήν ἀρχική εἶναι $r_1!r_2!\ldots r_k!$. Ἄρα $x.\ r_1!r_2!\ldots r_k!$ εἶναι ὁ ἀριθμός τῶν διατάξεων (μεταθέσεων) n διακεκριμένων στοιχείων καί ἰσοῦται μέ $n!$.

Παράδειγμα 1.1:

Νά βρεθῆ ὁ ἀριθμός τῶν λέξεων πού μποροῦν νά σχηματισθοῦν (ἀσχέτως νοήματος ἤ κανόνων γραμματικῆς ἤ συντακτικοῦ) δι'ἀντιμεταθέσεως τῶν γραμμάτων τῆς λέξεως ΜΑΘΗΜΑΤΙΚΑ.

Πρόκειται περί διατάξεων (μεταθέσεων) 10 μή διακεκριμένων στοιχείων μέ $r_1=2$ γιά τό στοιχεῖο M, $r_2=3$ γιά τό στοιχεῖο A, $r_3=1$ γιά τό στοιχεῖο I, $r_4=1$ γιά τό στοιχεῖο H, $r_5=1$ γιά τό στοιχεῖο T, $r_6=1$ γιά τό στοιχεῖο I καί $r_7=1$ γιά τό στοιχεῖο K. Ἔτσι ὁ ζητούμενος ἀριθμός τῶν λέξεων εἶναι:

$$\frac{10!}{2!3!1!1!1!1!1!} = 151.200 \ .$$

Ἐφαρμογή 1.2: Ἔστω ἕνα σύνολο μέ N διακεκριμένα στοιχεῖα ἀπό τά ὁποῖα Np εἶναι τύπου A καί Nq τύπου B ὅπου $0\leq p\leq 1$, $0\leq q\leq 1$ καί $p+q=1$. Ὁ ἀριθμός τῶν συνδυασμῶν τῶν N στοιχείων ἀνά n εἶναι $\binom{N}{n}$. Ἀπό αὐτούς τούς συνδυασμούς πόσοι περιέχουν ἀκριβῶς x στοιχεῖα τύπου A;

Ἡ ἀπάντηση εἶναι

$$\binom{Np}{x} \cdot \binom{Nq}{n-x} .$$

Εὑρίσκεται ἐφαρμόζοντας τόν κανόνα $m \cdot n$ καί τό γεγονός ὅτι οἱ συνδυασμοί Np στοιχείων τύπου A ἀνά x εἶναι $\binom{Np}{x}$ καί οἱ συνδυασμοί Nq στοιχείων τύπου B ἀνά $n-x$ εἶναι $\binom{Nq}{n-x}$.

Θεώρημα 1.7: Σχῆμα: Κυψέλες-Μπαλάκια (διακεκριμένα)

Ἔστω n κυψέλες καί r διακεκριμένα μπαλάκια, τά ὁποῖα ρίπτονται στίς κυψέλες. Τότε ὁ ἀριθμός τῶν τρόπων μέ τούς ὁποίους τά μπαλάκια μποροῦν νά τοποθετηθοῦν στίς κυψέλες εἶναι n^r.

Ἀπόδειξη: Ἐάν εἴχαμε 1 μπαλάκι τότε θά μποροῦσε νά μπεῖ στίς n κυψέλες κατά τούς ἑξῆς n τρόπους.

Τρόπος 1. └*┘ └┘ └┘ . . . └┘
1 2 3 n

Τρόπος 2. └┘ └*┘ └┘ . . . └┘

. . . .

Τρόπος n. └┘ └┘ └┘ └*┘

Ἐάν εἴχαμε 2 μπαλάκια a,b κάθε μπαλάκι θά ἔδιδε ἀπό n τρόπους, ὁπότε, συνολικά μέ τόν κανόνα $m \cdot n$ θά εἴχαμε n^2 τρόπους καί οὕτω καθ'ἑξῆς. Τό τελικό ἀποτέλεσμα θεμελιοῦται μέ ἐπαγωγή.▼

Παράδειγμα 1.2:

Ἔστω τρία μπαλάκια a,b,c καί τρεῖς κυψέλες. Ὁ ἀριθμός τῶν δυνατῶν τοποθετήσεων εἶναι $3^3 = 27$. Οἱ τοποθετήσεις αὐτές εἶναι:

abc			ac	b		ac		b
ab	c		a	bc		a	c	b
ab		c	a	b	c	a		bc
bc	a		c	ab		c	a	b
b	ac			abc			ac	b
b	a	c		ab	c		a	bc
bc		a	c	b	a	c		ab
b	c	a		bc	a		c	ab
b		ab		b	ac			abc .

Οἱ ἐφαρμογές τοῦ σχήματος κυψέλες-μπαλάκια εἶναι πολυάριθμες. Ἀναφέρομε ἐνδεικτικά τά γενέθλια τῶν ἀνθρώπων ὅπου οἱ *365* ἡμέρες τοῦ ἔτους εἶναι οἱ κυψέλες καί οἱ ἄνθρωποι τά μπαλάκια, τά τροχαῖα ἀτυχήματα ἐντός μίας ἑβδομάδας ὅπου οἱ *7* μέρες τῆς ἑβδομάδας εἶναι οἱ κυψέλες καί τά ἀτυχήματα τά μπαλάκια, τό ἀσανσέρ πολυκατοικίας ὅπου οἱ ὄροφοι τῆς πολυκατοικίας εἶναι οἱ κυψέλες καί οἱ ἐπιβάτες τά μπαλάκια, κ.ο.κ. Στή συνέχεια δίδομε μιά ἐφαρμογή στό ΠΡΟ-ΠΟ.

<u>Παράδειγμα</u> <u>1.3</u>:

Νά εὑρεθοῦν ὅλες οἱ δυνατές στῆλες ΠΡΟ-ΠΟ πού πρέπει νά συμπληρωθοῦν γιά νά ἐξασφαλίσουμε ἐπιτυχία ὅταν ἔχουμε νά προβλέψουμε *13* μάτς.

Γιά κάθε μάτς ἔχουμε τρεῖς δυνατές προβλέψεις *1,2,X*. Ἄρα οἱ προβλέψεις εἶναι οἱ κυψέλες καί τά μάτς τά μπαλάκια. Συνεπῶς οἱ δυνατοί συνδυασμοί εἶναι $3^{13} \approx 1.500.000$.

Ἔστω τώρα ὅτι τά r μπαλάκια πού ρίπτονται στίς n κυψέλες εἶναι μή διακεκριμένα. Στή περίπτωση αὐτή ἡ πιό βασική ἰδιότης μίας τοποθετήσεως εἶναι τό πόσα μπαλάκια πηγαίνουν σέ κάθε κυψέλη. Ἔστω $r_1, r_2, \ldots r_n$ οἱ ἀριθμοί τοποθετήσεως *(occupancy numbers)* ὅπου r_i εἶναι ὁ ἀριθμός τῶν σφαιρῶν πού τοποθετοῦνται στήν κυψέλη i. Τότε

$$r_1 + r_2 + \ldots + r_n = r .$$

Μία τοποθέτηση διαφέρει ἀπό μία ἄλλη ἐάν οἱ ἀντίστοιχες διατεταγμένες n-άδες $(r_1,\ldots,r_n)$ εἶναι διάφορες μεταξύ τους.

Θ ε ώ ρ η μ α 1.8: Σχῆμα: Κυψέλες-Μπαλάκια (μ ή δ ι α κ ε κ ρ ι μ έ ν α).

<u>Ἔστω n κυψέλες καί r μή διακεκριμένα (δηλ. ἴδια) μπαλάκια. Ὁ ἀριθμός τῶν τρόπων μέ τούς ὁποίους τά r μπαλάκια μποροῦν νά τοποθετηθοῦν στίς n κυψέλες εἶναι</u>:

$$\binom{n+r-1}{r} = \binom{n+r-1}{n-1} . \qquad (1.3)$$

Ἀπόδειξη: Μία τοποθέτηση μπορεῖ νά παρασταθεῖ σχηματικά μέ $n+1$ κάθετες γραμμές καί r ἀστερίσκους. Π.χ. ἄν $r=8$ καί $n=6$ μιά τοποθέτηση τῆς μορφῆς:

|* * * | *| | | |* * * *|

σημαίνει ότι 3 μπαλάκια είναι στήν πρώτη κυψέλη, 1 στήν δεύτερη, κανένα στήν τρίτη κ.ο.κ. Οἱ ἀριθμοί τοποθετήσεως εἶναι $r_1=3$, $r_2=1$, $r_3=0$, $r_4=0$, $r_5=0$ καί $r_6=4$. Ὁ συμβολισμός αὐτός ἀρχίζει καί τελειώνει μέ κάθετο γραμμή καί οἱ ὑπόλοιπες $(n+1)-2=n-1$ γραμμές καί οἱ r ἀστερίσκοι μποροῦν νά ἐμφανιστοῦν σέ ὁποιαδήποτε σειρά. Ἔτσι ὁ ἀριθμός πού ζητᾶμε εἶναι ὁ αὐτός μέ τούς τρόπους πού μποροῦμε νά συνδυάσουμε $n-1+r$ θέσεις ἀνά r γιά νά τοποθετήσωμε τά r μπαλάκια (εἶναι αὐτονόητο ὅτι οἱ ὑπόλοιπες θέσεις θά ἔχουν τίς κάθετες γραμμές). Δηλαδή ὁ ἀριθμός πού ζητᾶμε εἶναι ὁ

$$\binom{n+r-1}{r} = \binom{n+r-1}{n-1} . \blacktriangledown$$

Παρατηρήσεις:

1. Κάθε τοποθέτηση τῶν r μή διακεκριμένων σφαιρῶν στίς n κυψέλες εἶναι καί μία ἀκέραιη καί θετική λύση τῆς ἐξίσωσης $r_1+r_2+...+r_n=r$ μέ ἀγνώστους τά r_1, $r_2,...,r_n$. Ἄρα ὁ ἀριθμός τῶν λύσεων τῆς ἐξίσωσης αὐτῆς εἶναι $\binom{n+r-1}{r}$.

2. Τό σχῆμα κυψέλες-μπαλάκια μαζί μέ τό θεώρημα 1.2 μπορεῖ νά χρησιμοποιηθεῖ γιά τήν κατασκευή καί ὑπολογισμό τοῦ ἀριθμοῦ τῶν συνδυασμῶν n διακεκριμένων στοιχείων ἀνά r ὅταν κάθε στοιχεῖο μπορεῖ νά ἐπαναληφθεῖ τό πολύ r φορές. Στήν περίπτωση αὐτή τά n στοιχεῖα παριστοῦν τίς κυψέλες στίς ὁποῖες τοποθετοῦμε r μπαλάκια θεωρῶντας ὡς ἴδια τά μπαλάκια πού μπαίνουν σέ

κάθε κυψέλη.

<u>Παράδειγμα 1.4</u>:

Ἔστω τρεῖς κυψέλες καί τρία μή διακεκριμένα μπαλάκια. Τότε ὁ ἀριθμός τῶν δυνατῶν τοποθετήσεων εἶναι:

$$\binom{n+r-1}{n-1} = \binom{5}{2} = 10$$

Ἔχομε δέ τούς κάτωθι *10* σχηματισμούς

1	* * *		
2		* * *	
3			* * *
4	* *	*	
5	* *		*
6	*	* *	
7	*		* *
8		* *	*
9		*	* *
10	*	*	*

1.4 Δειγματοληψίες ἀπό πεπερασμένους πληθυσμούς

Συχνά στήν Στατιστική ἀσχολούμεθα μέ δειγματοληψίες ἀπό πεπερασμένους πληθυσμούς ὅπως π.χ. στήν σφυγμομέτρηση κοινῆς γνώμης, στόν ὑπολογισμό ποσοστοῦ ἀνεργίας ἤ μεγέθους ἐργατικῆς δυνάμεως μίας χώρας, στήν γεωργική παραγωγή, σέ ἔρευνες γιά τό ἐπίπεδο ὑγείας τῶν κατοίκων μίας περιοχῆς ἤ χώρας κ.λ.π. Ἔστω N τό μέγεθος τοῦ πληθυσμοῦ καί n τό μέγεθος τοῦ δείγματος. Κατά τήν δειγματοληψία τά μέλη τοῦ δείγματος ἐκλέγονται διαδοχικά τό ἕνα κατόπιν τοῦ ἄλλου. Σέ κάθε ἐπιλεγόμενο μέλος μετροῦμε τίς τιμές μίας ἤ περισσοτέρων μεταβλητῶν. Ἡ ἐπιλογή μπορεῖ νά γίνει μέ δύο τρόπους: (*i*) Χωρίς ἐπανάθεση ἤ ἐπανατοποθέτηση τοῦ ἐπιλεγέντος μέλους καί (*ii*) μέ ἐπανάθεση ἤ ἐπανατοποθέτηση τοῦ ἐπιλεγέντος μέλους. Ἄλλες φορές ἡ σειρά ἐπιλογῆς τῶν μελῶν τοῦ δείγματος δέν παίζει ρόλο καί τό δεῖγμα εἶναι μή διατεταγμένο καί ἄλλες φορές (σπάνια) ἡ σειρά ἐπιλογῆς παίζει ρόλο καί τό δεῖγμα εἶναι διατεταγμένο.

Στίς τέσσερες παραπάνω περιπτώσεις ὁ ἀριθμός ὅλων τῶν δυνατῶν δειγμάτων πού μποροῦν νά ληφθοῦν ὑπολογίζεται ὡς ἑξῆς:

(1) Δεῖγμα χωρίς ἐπανάθεση-μή διατεραγμένο: ἀριθμός δειγμάτων $= \binom{N}{n}$

(2) Δεῖγμα χωρίς ἐπανάθεση-διατεταγμένο: ἀριθμός δειγμάτων $= P_n^N$

(3) Δεῖγμα μέ ἐπανάθεση-διατεταγμένο: ἀριθμός δειγμάτων $= N^n$

(βάσει τοῦ σχήματος κυψέλες-μπαλάκια τά μέλη τοῦ πληθυσμοῦ εἶναι οἱ κυψέλες καί οἱ θέσεις τοῦ δείγματος τά μπαλάκια).

(4) Δεῖγμα μέ ἐπανάθεση-μή διατεταγμένο: ἀριθμός δειγμάτων $= \binom{N+n-1}{n} = \binom{N+n-1}{N-1}$

(βάσει τοῦ τρόπου μέ τόν ὁποῖον μποροῦμε νά τοποθετήσουμε N μή-διακεκριμένα μπαλάκια σέ n κυψέλες).

Ἔστω ὅτι ὁ πληθυσμός ἀποτελεῖται ἀπό τρία στοιχεῖα $\{a_1, a_2, a_3\}$. Τά τρία μή διατεταγμένα καί χωρίς ἐπανάθεση δείγματα μεγέθους δύο εἶναι τά (a_1, a_2), (a_1, a_3) καί

(a_2,a_3). Τά ἕξι ἤ P_2^3 διατεταγμένα καί χωρίς ἐπανάθεση δείγματα μεγέθους δύο εἶναι: $(a_1,a_2),(a_2,a_1)$, (a_1,a_3) , (a_3,a_1), (a_2,a_3) καί (a_3,a_2). Τά ἐννέα ἤ 3^2 διατεταγμένα καί μέ ἐπανάθεση δείγματα μεγέθους δύο εἶναι: (a_1,a_1), (a_1,a_2), $(a_2,a_2),(a_2,a_1)$, (a_3,a_3), (a_1,a_3) , (a_3,a_1), (a_2,a_3) καί (a_3,a_2). Τά ἕξι ἤ $\binom{3+2-1}{2}$ μή διατεταγμένα καί μέ ἐπανάθεση δείγματα μεγέθους δύο εἶναι: (a_1,a_1), (a_1,a_2), (a_2,a_2), (a_3,a_3), (a_1,a_3) καί (a_2,a_3).

1.5 Διωνυμικοὶ Συντελεστὲς καὶ ὁ τύπος τοῦ Stirling.

Ὁ διωνυμικός συντελεστής $\binom{n}{r}=n!/[r!(n-r)!]$ἔχει ὁριστεῖ μόνο γιά n καί r θετικά καί ἀκέραια καί $r\leq n$. Μποροῦμε ὅμως νά ἐπεκτείνουμε τόν ὁρισμό αὐτό γιά κάθε πραγματικό x καί καθε θετικό καί ἀκέραιο r ὡς ἀκολούθως

$$\binom{x}{r} = \frac{x(x-1)\dots(x-r+1)}{r!} \quad . \tag{1.4}$$

Θέτουμε $\binom{x}{0} = 1$ καί γιά r ἀρνητικό καί ἀκέραιο ὁρίζουμε $\binom{x}{0} = 0$. Ἐπίσης γιά n θετικό καί ἀκέραιο $\binom{n}{r} = 0$ ἄν $r>0$ ἤ $r<0$.

Γιά παράδειγμα:

$$\binom{1/2}{4} = \frac{\frac{1}{2}\left(\frac{1}{2}-1\right)\left(\frac{1}{2}-2\right)\left(\frac{1}{2}-3\right)}{4!} = -\frac{5}{128}$$

$$\binom{-1}{r} = (-1)^r , \qquad \binom{-2}{r} = (-1)^r(r+1) .$$

Μέ τήν παραπάνω επέκταση ισχύει καί ὁ διωνυμικός τύπος τοῦ Νεύτωνος.

$$(1+t)^a = 1+\binom{a}{1}t+\binom{a}{2}t^2+\binom{a}{3}t^3+\dots , \quad -1<t<1 \qquad (1.5)$$

γιά κάθε πραγματικό a.

Τό κατά *Taylor* ἀνάπτυγμα τοῦ φυσικοῦ λογαρίθμου εἶναι

$$ln(1+t) = t- \frac{1}{2} t^2+ \frac{1}{3} t^3- \frac{1}{4} t^4+\dots , \quad -1<t<1 . \qquad (1.6)$$

Ὁ τύπος τοῦ *Stirling* εἶναι

$$n! \sim e^{-n}n^n \sqrt{2n\pi} , \qquad (1.7)$$

ὅπου τό σύμβουλο $\sim$ σημαίνει ὅτι

$$\lim_{n\to\infty} \frac{n!}{n^{-n}n^n \sqrt{2n\pi}} = 1 .$$

Εἶναι χρήσιμο ἀναλυτικό ἐργαλεῖο καί μᾶς ἐπιτρέπει τόν εὔκολο καί μέ καλή προσέγγιση ὑπολογισμό ποσοτήτων ὅπως $10!$, $20!$, $50!$ κ.λ.π.

ΑΣΚΗΣΕΙΣ

1.1. Κατά πόσους τρόπους μπορούμε .. μοιράσουμε 5 διαφορετικές εφημερίδες σέ 5 άτομα;

1.2. Σέ ένα σύνολο 10 άριθμών 6 είναι θετικοί καί οί ύπόλοιποι άρνητικοί. Κατά πόσους τρόπους μποροῦμε νά πάρουμε (χωρίς έπανάθεση) δύο άριθμούς ὥστε (*a*) νά είναι έτερόσημοι καί (*b*) όμόσημοι;

1.3. Ένα διαγώνισμα πολλαπλῆς έκλογῆς περιέχει 15 έρωτήσεις γιά κάθε μία άπό τίς όποῖες δίδονται 5 άπαντήσεις. Ὁ μαθητής πρέπει νά διαλέξει μία άπάντηση άπό τίς πέντε γιά κάθε έρώτηση. Κατά πόσους διαφορετικούς τρόπους μπορεῖ ένας μαθητής νά δώσει άπαντήσεις σ'αύτές τίς έρωτήσεις.

1.4. Κάποιος πρόκειται νά άγοράσει ένα καινούργιο αύτοκίνητο. Ἔχει νά διαλέξει μεταξύ μηχανῶν τριῶν μεγεθῶν (κυβισμοῦ), 5 τύπων άμαξώματος καί 10 χρωμάτων. Πόσα είναι τά διαφορετικά αύτοκίνητα άπό τά όποῖα έχει νά διαλέξει;

1.5. Ἡ Ὀλυμπιακή έχει 3 πτήσεις άπό τό Ἡράκλειο Κρήτης στήν Ἀθήνα καί 9 πτήσεις άπό τήν Ἀθήνα στή Θεσσαλονίκη. Πόσους διαφορετικούς συνδυασμούς πτήσεων μπορεῖ ή Ὀλυμπιακή Ἀεροπορία νά προσφέρει άπό τό Ἡράκλειο στήν Θεσσαλονίκη;

1.6. Πόσοι διαφορετικοί πενταψήφιοι τηλεφωνικοί άριθμοί μποροῦν νά σχηματισθοῦν άν τό πρῶτο ψηφίο πρέπει νά είναι 2 ή 3.

1.7. Κατά πόσους τρόπους μπορεῖ ἕνας φοιτητής νά διαλέξει 2 κατ'ἐπιλογήν μαθήματα ἀπό ἕξι πού προσφέρονται σέ διαφορετικές ὧρες;

1.8. Πόσοι τετραψήφιοι ἀριθμοί μποροῦν νά σχηματιστοῦν ἀπό τά ψηφία $4,5,6,7,8$ καί 9 ἄν (a) κάθε ψηφίο μπορεῖ νά χρησιμοποιηθεῖ μόνο μία φορά, (b) ἐπιτρέπονται ἐπαναλήψεις ψηφίων.

1.9. Τά γράμματα στόν κώδικα τοῦ *Morse* σχηματίζονται μέ τελεῖες καί παῦλες (ἐπαναλήψεις ἐπιτρέπονται). Πόσα διαφορετικά γράμματα μποροῦμε νά σχηματίσουμε ἄν κάθε γράμμα γίνεται μέ τέσσερα τό πολύ σύμβολα.

1.10. Τά ἀρχικά τῶν ἀτόμων σχηματίζονται ἀπό τό ὄνομα,τό πατρώνυμο καί τό ἐπώνυμο. Πόσα διαφορετικά σύνολα ἀρχικῶν μποροῦν νά σχηματιστοῦν ἄν κάθε ἄτομο ἔχει ἕνα μόνο ἐπώνυμο καί (a) ἕνα ἀκριβῶς ὄνομα καί ἕνα ἀκριβῶς πατρώνυμο (b) τό πολύ δύο ὀνόματα καί ἕνα ἀκριβῶς πατρώνυμο.

1.11. Νά δεχθεῖ ὅτι

a) $$1-\binom{n}{1}+\binom{n}{2}-\ldots\ldots+(-1)^n\binom{n}{n} = 0$$

b) $$\binom{n}{1}+2\binom{n}{2}+3\binom{n}{3}+\ldots+n\binom{n}{n} = n2^{n-1}$$

c) $$\binom{n}{1}-2\binom{n}{2}+3\binom{n}{3}-\ldots\ldots+(-1)^n n\binom{n}{n} = 0$$

ΚΕΦΑΛΑΙΟ 2

Πιθανοτητες βασικες προτασεις

Οἱ πρῶτες ἔννοιες τῆς θεωρίας Πιθανοτήτων εἶναι τό τυχαῖο πείραμα, ὁ δειγματικός χῶρος καί τά ἐνδεχόμενα. Ἡ ἔννοια τῆς πιθανότητας ἔχει πολλές θεωρήσεις. Στό κεφάλαιο αὐτό μελετᾶμε τίς ποσοτικές ἐκεῖνες θεωρήσεις πού ὁδηγοῦν σέ τρόπους ὑπολογισμοῦ τῶν πιθανοτήτων διαφόρων ἐνδεχομένων. Τό οἰκοδόμημα τῆς θεωρίας Πιθανοτήτων χτίζεται σιγά σιγά ἀξιωματικά καί προχωρεῖ μέ τή βοήθεια τῶν βασικῶν νόμων τῶν πιθανοτήτων. Εἰσάγονται καινούργιες ἔννοιες ὅπως ἡ δεσμευμένη πιθανότητα, ἡ ἀνεξαρτησία, ὁ κανόνας τοῦ *Bayes* κλπ. Τά θέματα αὐτά ἀποτελοῦν ἀντικείμενο μελέτης τοῦ παρόντος κεφαλαίου.

2.1 Πείραμα — Δειγματικὸς χῶρος — Ἐνδεχόμενα

Ὁ ὅρος π ε ί ρ α μ α μέ τήν κοινή καθομιλούμενη ἔννοια δέν χρειάζεται νά ὁρισθεῖ ἰδιαίτερα. Ἡ θεωρία Πιθανοτήτων καί ἡ Στατιστική χρησιμοποιοῦν τόν ὅρο αὐτό εἴτε μέ τή στενή, πραγματική του ἔννοια ὅπως γιά παράδειγμα τό πείραμα ἑνός ἐργαστηρίου Φυσικῆς, εἴτε μέ μία γενικώτερη ἰδεατή μορφή. 'Η δοκιμασία τῶν Πανελλήνιων 'Εξετάσεων μπορεῖ νά θεωρηθεῖ ὡς πείραμα. 'Ομοίως ἡ θεραπεία μέ ἕνα φάρμακο, ἡ δημοσκόπηση, ἡ καλλιέργεια ἑνός ἀγροῦ μέ δύο ποικιλίες σταριοῦ κλπ.

'Υπάρχουν πειράματα τῶν ὁποίων τά ἀποτελέσματα εἶναι γνωστά ἐκ τῶν προτέρων. 'Υπάρχουν ὅμως καί πειράματα τῶν ὁποίων τά ἀποτελέσματα δέν μποροῦν νά προκαθοριστοῦν. Ἄν π.χ. ἑνώσουμε δύο ἄτομα ὑδρογόνου καί ἕνα ἄτομο ὀξυγόνου τό ἀποτέλεσμα θά εἶναι ὁ σχηματισμός ἑνός μορίου νεροῦ, πράγμα προκαθορισμένο. Ἄν ὅμως ἐκτοξεύσουμε ἕνα βλῆμα σέ κάποιο στόχο τό ἀποτέλεσμα δέν μπορεῖ νά προκαθορισθεῖ γιατί ἐπιδροῦν σ'αὐτό ἄγνωστοι παράγοντες ἤ παράγοντες πού δέν μποροῦν νά ἐλεγθοῦν.

'Η Θεωρία Πιθανοτήτων καί ἡ Στατιστική ἀσχολοῦνται ἰδιαίτερα μέ τά λεγόμενα τυχαῖα πειράματα (ἤ πειράματα τύχης) δηλαδή πειράματα τῶν ὁποίων τά ἀποτελέσματα κυβερνῶνται ἀπό τούς νόμους τῆς τύχης καί ἔτσι δέν μποροῦν νά καθορισθοῦν ἐκ τῶν προτέρων. Συχνά γιά λόγους συντομίας θά χρησιμοποιοῦμε μόνο τόν ὅρο πείραμα. Παραδείγματα τυχαίων πειραμάτων εἶναι: τό ρίψιμο ἑνός νομίσματος μία φορά, τό ρίψιμο ἑνός νομίσματος *100* φορές, τό ρίψιμο τριῶν ζαριῶν, τό μοίρασμα τῶν χαρτιῶν μίας τράπουλας, ἡ παρατήρηση τοῦ χρόνου ζωῆς ἑνός ραδιενεργοῦ ἀτόμου ἤ ἑνός ἀνθρώπου, ἡ ἐπιλογή ἑνός δείγματος ἀνθρώπων καί ἡ παρατήρηση τοῦ φύλου, τοῦ ἀριθμοῦ πτυχιούχων Παν/μιου κλπ., ἡ σπορά καί παραγωγή σίτου σέ ἕνα ἀγρό κλπ.

Κάθε πείραμα ἔχει ἀποτελέσματα. Μπορεῖ νά ἐπαναλαμβάνεται κάτω από τίς ἴδιες συνθῆκες πραγματικά ἤ νοερά πολλές ἤ ἄπειρες φορές. Τά ἀποτελέσματα ἑνός τυχαίου πειράματος λέγονται ἐ ν δ ε χ ό μ ε ν α ἤ γ ε γ ο ν ό - τ α. Τά ἀτομικά ἀδιαίρετα ἀποτελέσματα λέγονται σ τ ο ι - χ ε ι ώ δ η ἤ ἁ π λ ά ἐ ν δ ε χ ό μ ε ν α ἤ δ ε ι γ - μ α τ ι κ ά σ η μ ε ῖ α. Ἕνα ἐνδεχόμενο εἶναι μία συλλογή ἁπλῶν ἐνδεχομένων ἤ δειγματικῶν σημείων. Τό σύνολο τῶν δειγματικῶν σημείων ἑνός τυχαίου πειράματος λέγεται δ ε ι γ μ α τ ι κ ό ς χ ῶ ρ ο ς. Ὁ χῶρος αὐτός θά συμβολίζεται μέ τό S. Τά ἐνδεχόμενα θά συμβολίζονται μέ A, B, Γ, C, D, E κλπ.

Ἕνα ἐνδεχόμενο πραγματοποιεῖται (συμβαίνει) ὅταν τουλάχιστον ἕνα ἀπό τά δειγματικά του σημεῖα πραγματοποιεῖται. Ἔτσι ὁ δειγματικός χῶρος λέγεται καί β έ - β α ι ο ἐ ν δ ε χ ό μ ε ν ο δεδομένου ὅτι περιέχει ὅλα τά δειγματικά σημεῖα ἕνα ἀπό τά ὁποῖα ὁπωσδήποτε πραγματοποιεῖται. Ἡ λογική εἰκόνα τώρα συμπληρώνεται μέ τό ἀ- δ ύ ν α τ ο ἐ ν δ ε χ ό μ ε ν ο δηλαδή τό ἐνδεχόμενο πού δέν μπορεῖ νά πραγματοποιηθεῖ καί πού ἀντιστοιχεῖ μέ τό κενό σύνολο. Συμβολίζεται καί αὐτό μέ τό $\emptyset$.

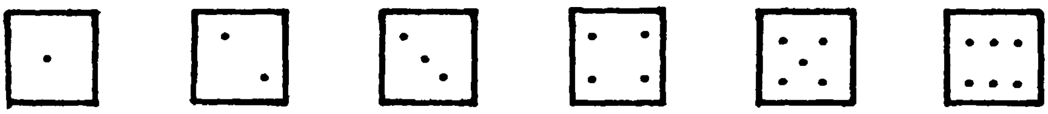
Τά ἐνδεχόμενα διακρίνονται καί σέ ἁ π λ ά καί σ ύ ν θ ε τ α. Ἁπλά ἐνδεχόμενα εἶναι τά δειγματικά σημεῖα. Σύνθετα ἐνδεχόμενα εἶναι τά ὑποσύνολα τοῦ δειγματικοῦ χώρου.

Μέ βάση τούς παραπάνω ὁρισμούς καί ἔννοιες βλέπουμε ὅτι οἱ ἔννοιες τοῦ δειγματικοῦ χώρου καί τοῦ ἐνδεχομένου ἀντιστοιχοῦν στίς ἔννοιες τοῦ συνόλου καί ὑποσυνόλου. Πράγματι κ ά θ ε ὑ π ο σ ύ ν ο λ ο τ ο ῦ S εἶναι ἕνα ἐνδεχόμενο.

Παραδὲίγματα:

2.1: 'Η γέννηση ἑνός παιδιοῦ μπορεῖ νά θεωρηθεῖ ὡς τυχαῖο πείραμα μέ δύο δυνατά ἀποτελέσματα: ἀγόρι (A), κορίτσι (K). Στήν περίπτωση ἁπλῆς γεννήσεως ὁ δειγματικός χῶρος εἶναι $S = \{A,K\}$ καί τά A,K εἶναι ἁπλά ἐνδεχόμενα.

2.2: Τό ζάρι εἶναι ἕνας κύβος μέ τίς ἀκόλουθες ἕξη πλευρές:


.

Οἱ πλευρές αὐτές θά συμβολίζονται μέ E_1, E_2, E_3, E_4, E_5, E_6. Ἕνα ζάρι ρίχνεται πάνω σέ μία ὁριζόντια ἐπιφάνεια. Ἡ ἔνδειξη τῆς πλευρᾶς πού εἶναι στραμμένη πρός τά πάνω ἀποτελεῖ τό ἀποτέλεσμα τοῦ πειράματος. Τό πείραμα ἔχει ἕξη δυνατά ἀποτελέσματα καί ὁ δειγματικός χῶρος εἶναι $S = \{E_1,E_2,E_3,E_4,E_5,E_6\}$. Τά E_i, $i=1,2,\ldots,6$ εἶναι ἁπλά ἐνδεχόμενα. Σύνθετα ἐνδεχόμενα εἶναι τά

$$A = \{E_2,E_4,E_6\} = \{\text{τό ἀποτέλεσμα εἶναι ἄρτιο}\}$$

$$B = \{E_1,E_2,E_3,E_4\} = \{\text{τό ἀποτέλεσμα εἶναι } \leq 4\}$$

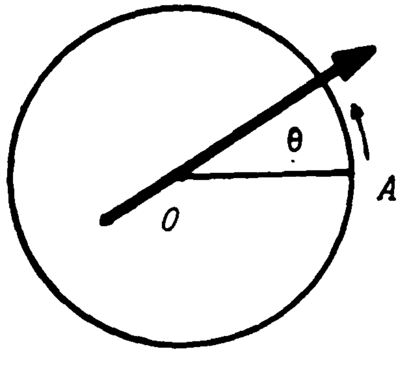
κλπ. Ὅταν τό E_3 π.χ. πραγματοποιεῖται (συμβαίνει) τότε πραγματοποιεῖται καί τό B.

2.3: Ἕνα Κέντρο Ἀμέσου Βοηθείας θέλοντας νά βελτιώσει τήν τηλεφωνική του ἐπικοινωνία μέ τό κοινό παρακολουθεῖ τόν ἀριθμό τῶν τηλεφωνικῶν κλήσεων ἀπό τίς *8* τό βράδυ μέχρι τίς *8* τό πρωΐ. Τά δυνατά ἀποτελέσματα τοῦ πειράματος ἐδῶ εἶναι $0, 1, 2, 3, \ldots$ κλήσεις καί ὁ δειγματικός χῶρος εἶναι $S = \{0, 1, 2, 3, \ldots\}$. Δύο σύνθετα ἐνδεχόμενα εἶναι τά ἀκόλουθα:

$$A = \{\text{τουλάχιστον } 3 \text{ κλήσεις}\} = \{3, 4, 5, \ldots\}$$

$$B = \{\text{τό πολύ } 2 \text{ κλήσεις}\} = \{0, 1, 2\}.$$

2.4: Ἔστω μία βελόνη ἡ ὁποία περιστρέφεται γύρω ἀπό ἕνα ἄξονα παράλληλα σέ ἕνα ἐπίπεδο ὅπως στή μαγνητική πυξίδα.



Γυρίζουμε τή βελόνη μερικές φορές ἀντίθετα ἀπό τήν κίνηση τῶν δεικτῶν τοῦ ὡρολογίου καί τήν ἀφήνουμε νά κινηθεῖ ἐλεύθερα. Τό ἀποτέλεσμα τοῦ πειράματος εἶναι ἡ γωνία θ πού σχηματίζεται μέ μία σταθερή ἀκτίνα *OA*. Ὁ δειγματικός χῶρος εἶναι ὅλες οἱ δυνατές γωνίες μεταξύ *0* καί 2π. $S=\{\theta : 0 \leq \theta \leq 2\pi\}$.

Ὁ δειγματικός χῶρος μπορεῖ νά εἶναι πεπερασμένος ἤ ἀριθμήσιμος ἤ μή ἀριθμήσιμος (συνεχής). Στά Παραδείγματα 2.1 καί 2.2 οἱ δειγματικοί χῶροι εἶναι σύνολα μέ πεπερασμένο πλῆθος στοιχείων, στό 2.3 ἀριθμήσιμος καί στό 2.4 μή ἀριθμήσιμος. Ἐπίσης δύο ἤ περισσότεροι δειγματικοί χῶροι μποροῦν νά προκύψουν ἀπό τό ἴδιο πείραμα.

Παραδείγματα:

2.5: Ἔστω ὅτι ρίχνουμε ἕνα νόμισμα μέ πλευρές K καί Γ δύο φορές. Τά ἀποτελέσματα τοῦ πειράματος αὐτοῦ εἶναι τέσσερα καί συμβολίζονται μέ KK, $K\Gamma$, κλπ., ὅπου KK σημαίνει ὅτι πρῶτα ἐμφανίστηκε ἡ πλευρά K καί κατόπιν πάλι ἡ ἴδια πλευρά, $K\Gamma$ σημαίνει πρῶτα ἡ πλευρά K καί κατόπιν ἡ πλευρά Γ κ.ο.κ. Ὁ δειγματικός χῶρος εἶναι

$$S = \{KK, K\Gamma, \Gamma K, \Gamma\Gamma\}.$$

Ἄν ὅμως ἀγνοήσουμε τή σειρά ἐμφανίσεως καί ἐνδιαφερθοῦμε γιά τόν ἀριθμό τῶν φορῶν πού θά ἐμφανιστεῖ ἡ πλευρά K τότε ἔχουμε ἕνα νέο δειγματικό χῶρο

$$S = \{0K, 1K, 2K\},$$

ὅπου

$$0K = \{\text{μηδέν πλευρές } K\} = \{\Gamma\Gamma\}$$

$$1K = \{\text{μία πλευρά } K\} = \{K\Gamma, \Gamma K\}$$

$$2K = \{\text{δύο πλευρές } K\} = \{KK\}.$$

2.6: Ένα βλήμα εκτοξεύεται μέ σκοπό νά προσβάλλει κάποιο στόχο στήν επιφάνεια τῆς γῆς. Τυχαῖα σφάλματα καί ἄλλοι ἀνεξέλεγκτοι παράγοντες ἐπηρεάζουν τό σημεῖο στό ὁποῖο τό βλῆμα τελικά θά χτυπήσει τόν στόχο. Ἄν μᾶς ἐνδιαφέρει ἁπλῶς ἡ ἀπόσταση τοῦ σημείου προσκρούσεως τοῦ βλήματος ἀπό τόν στόχο τότε τά δυνατά ἀποτελέσματα τοῦ πειράματος ἀποτελοῦνται ἀπό τά σημεῖα ἑνός διαστήματος $[o,a]$ ὅπου $a>o$. Ἄν μᾶς ἐνδιαφέρει ἡ θέση τοῦ σημείου προσκρούσεως τότε τά δυνατά ἀποτελέσματα ἀποτελοῦνται ἀπό τίς τριάδες (x,y,z) ὅπου x,y,z εἶναι οἱ συντεταγμένες κάθε σημείου στό ὁποῖο μπορεῖ τό βλῆμα νά προσκρούσει καί μεταβάλλονται σέ ὁρισμένα διαστήματα.

Ἀνάλογα μέ τό ἀποτέλεσμα πού μᾶς ἐνδιαφέρει ὁδηγούμαστε σέ δύο διαφορετικούς δειγματικούς χώρους. Ὁ πρῶτος εἶναι ἕνα ὑποσύνολο τοῦ R (σύνολο πραγματικῶν ἀριθμῶν) καί ὁ δεύτερος εἶναι ἕνα ὑποσύνολο τοῦ R^3. Καί οἱ δύο χῶροι εἶναι μή ἀριθμήσιμοι συνεχεῖς χῶροι.

Πολλές φορές ἕνα πείραμα μπορεῖ νά θεωρηθεῖ ὅτι ἀποτελεῖται ἀπό δύο ἤ καί περισσότερα ἄλλα ἐπί μέρους πειράματα. Τό πείραμα τότε λέγεται σ ύ ν θ ε τ ο καί ἡ μελέτη του γίνεται εὐκολώτερα διά τοῦ συνδυασμοῦ τῆς μελέτης τῶν ἐπί μέρους πειραμάτων. Γιά παράδειγμα, σύνθετο πείραμα εἶναι μία σειρά ἀπό δέκα ὅμοια τέστ (δοκιμές)ἐργαστηρίου κάθε ἕνα ἀπό τά ὁποῖα μπορεῖ νά ἔχει δύο ἤ περισσότερα δυνατά ἀποτελέσματα. Ἡ ἐκλογή χωρίς ἐπανάθεση ἑνός δείγματος μεγέθους n ἀπό κάποιο πληθυσμό μεγέθους N εἶναι ἕνα σύνθετο πείραμα ἀποτελούμενο ἀπό n ἐπί μέρους πειράματα ὅπου τό πρῶτο ἔχει N δυνατά ἀποτελέσματα, τό δεύτερο $N-1$, τό τρίτο $N-2$, κ.ο.κ .

2. 2. Πράξεις μὲ ἐνδεχόμενα

Τό ἑπόμενο φυσικό βῆμα στό λογικό οἰκοδόμημα πού χτίζομε εἶναι νά ὁρίσουμε τίς πράξεις μεταξύ τῶν ἐνδεχομένων. Ἔστω λοιπόν A,B δύο ἐνδεχόμενα τοῦ δειγματικοῦ χώρου S.

Ὁρισμοί:

2.1: Ἄν ἡ πραγματοποίηση τοῦ ἐνδεχομένου A συνεπάγεται τήν πραγματοποίηση τοῦ B τότε λέμε ὅτι τό A συνεπάγεται τό B καί συμβολίζουμε $A > B$ ἤ $A\ B$.

2.2: Ἄν τό A συνεπάγεται τό B καί τό B συνεπάγεται τό A τότε λέμε ὅτι τά A καί B εἶναι ἰσοδύναμα (ἤ ἴσα) καί γράφουμε $A=B$. Μέ ἄλλα λόγια $A=B$ σημαίνει ὅτι τά ἐνδεχόεμνα A καί B εἴτε πραγματοποιοῦνται εἴτε δέν πραγματοποιοῦνται ταυτόχρονα.

2.3: Τό ἐνδεχόμενο πού συνίσταται στήν πραγματοποίηση τοὐλάχιστον ἑνός ἀπό τά δύο ἐνδεχόμενα A,B ὀνομάζεται ἄθροισμα ἤ ἕνωση τῶν A καί B καί συμβολίζεται μέ $A+B$ ἤ $A\cup B$. Μέ ἄλλα λόγια τό ἄθροισμα $A+B$ σημαίνει τήν πραγματοποίηση τοῦ A ἤ τοῦ B ἤ καί τῶν δύο μαζί.

2.4: Τό ἐνδεχόμενο πού συνίσταται στήν ταυτόχρονη πραγματοποίηση τῶν A καί B ὀνομάζεται τό γινόμενο ἤ ἡ τομή αὐτῶν καί συμβολίζεται μέ AB ἤ $A\cap B$.

2.5: Ἐάν $A \cap B$ εἶναι τό ἀδύνατο ἐνδεχόμενο, τότε τά A καί B λέγονται ἀ σ υ μ β ί β α σ τ α ἐνδεχόμενα. Μέ ἄλλα λόγια τά A καί B δέν μποροῦν νά πραγματοποιηθοῦν ταυτόχρονα.

2.6: Τό ἐνδεχόμενο πού συνίσταται στήν πραγματοποίηση τοῦ B καί τήν μή πραγματοποίηση τοῦ B ὀνομάζεται ἡ δ ι α φ ο ρ ά τῶν A,B καί συμβολίζεται μέ $A-B$.

2.7: Ἀ ν τ ί θ ε τ ο ἤ σ υ μ π λ η ρ ω μ α τ ι κ ό τοῦ ἐνδεχομένου A λέγεται τό ἐνδεχόμενο πού συνισταται στήν μή πραγματοποίηση τοῦ A καί συμβολίζεται μέ A' ἤ A^c ἤ $\overline{A}$.

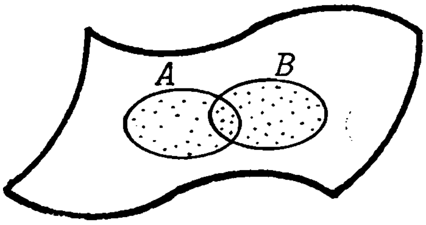
Ἡ πραγματοποίηση τοῦ ἐνδεχόμενου A' σημαίνει τήν μή πραγματοποίηση τοῦ A καί τά ἐνδεχόμενα A καί A' εἶναι ἀσυμβίβαστα.

Τό ἄθροισμα καί τό γινόμενο περισσότερων ἀπό δύο ἀλλά πεπερασμένου πλήθους ἐνδεχομένων ὁρίζονται ἀνάλογα. Ἄν ἔχουμε μία ἀκολουθία ἐνδεχομένων A_n, $n=1,2,\ldots$ δεχόμαστε χωρίς συζήτηση ὅτι τά

$$\bigcup_{n=1}^{\infty} A_n \quad \text{καί} \quad \bigcap_{n=1}^{\infty} A_n$$

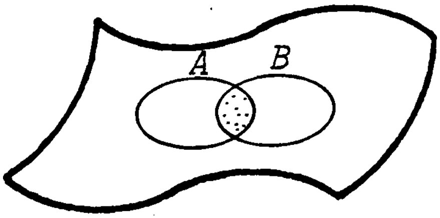
ἀποτελοῦν ἐνδεχόμενα. Τό πρῶτο συνίσταται στήν πραγματοποίηση τουλάχιστον ἑνός ἀπό τά A_n καί τό δεύτερο συνίσταται στήν ταυτόχρονη πραγματοποίηση ὅλων τῶν A_n, $n=1, 2,\ldots$.

Ἡ ἀντιστοιχία ἐνδεχόμενων καί συνόλων μᾶς ἐπιτρέπει τήν ἀκόλουθη, γνωστή ἀπό τή θεωρία Συνόλων, γραφική παρασταση τῶν παραπάνω πράξεων (βλέπε Σχῆμα 2.1).



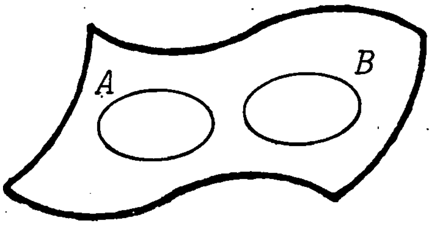
$A+B$

ἄθροισμα ἐνδεχομένων



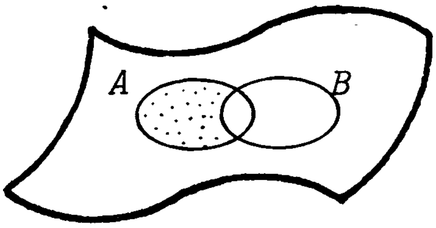
AB

γινόμενο ἐνδεχομένων



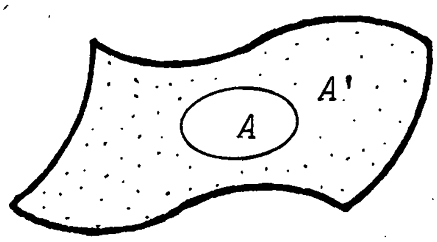
$AB=\emptyset$

ἀσυμβίβαστα ἐνδεχόμενα



$A-B$

διαφορά ἐνδεχομένων



A'

ἀντίθετο ἐνδεχόμενο

Σχῆμα 2.1 Πράξεις ἐνδεχομένων

Παραδείγματα:

2.7: Ἔστω μία τράπουλα 52 χαρτιῶν ἀπό τήν ὁποία τραβάμε στήν τύχη ἕνα χαρτί. Τό πείραμα ἔχει 52 δυνατά ἀποτελέσματα καί ὁ δειγματικός χῶρος S ἀποτελεῖται, ἀπό τά 52 παιγνιδόχαρτα. Ἔστω τώρα τά ἐνδεχόμενα $A=$ {σπαθί}, $B =$ {βαλές}, $C =$ {φιγούρα}. Τότε

$A \cup B =${Ὅλα τά σπαθιά καί οἱ ἄλλοι τρεῖς βαλέδες},

$A-B =${Ὅλα τά σπαθιά ἐκτός ἀπό τόν βαλέ σπαθί},

$A \cap B =${Βαλές σπαθί},

$S-A =${Ὅλα τά καρρώ, κοῦπες καί μπαστούνια},

$(A \cup C)-B =${Ὅλα τά σπαθιά καί οἱ φιγοῦρες ἐκτός ἀπό τούς τέσσερεις βαλέδες},

ὅπου οἱ ἐκφράσεις μέσα στίς παρενθέσεις { } περιγράφουν τά δειγματικά σημεῖα τῶν ἐνδεχομένων.

2.8: Ἔστω ὅτι ὁ δειγματικός χῶρος εἶναι τό σύνολο τῶν πραγματικῶν ἀριθμῶν καί ὅτι τό ἐνδεχόμενο A_n παρίσταται μέ τό διάστημα $(a_n - 1/n,\ a_n+1/n)$. Ἡ ἕνωση $\bigcup_{n=1}^{\infty} A_n$ εἶναι τό ἐνδεχόμενο-διάστημα $(a-1,\ a+1)$, καί ἡ τομή $\bigcap_{n=1}^{\infty} A_n$ εἶναι τό ἐνδεχόμενο $\{a\}$.

2. 3. Πιθανότητα

Ἐρχόμαστε τώρα νά παρουσιάσουμε τήν ἔννοια τῆς πιθανότητας καί ειδικώτερα νά ορίσουμε αυτό πού λέμε πιθανότητα ενός ενδεχομένου. Ἡ ἔννοια τῆς πιθανότητας εἶναι πρωταρχικῆς σημασίας καί ἅπτεται πολλῶν θεωρήσεων πού καλύπτουν ἕνα ευρύ φάσμα: από τή φιλοσοφία καί τή λογική μέχρι τή διαίσθηση, τήν εμπειρία καί τήν πρακτική εφαρμογή. Συνδέεται μέ τήν ἔννοια τῆς αβεβαιότητας, τῆς τύχης, τῆς σχετικῆς συχνότητας καί τῆς υποκειμενικῆς κρίσης.

Δέν θά ασχοληθοῦμε μέ τίς φιλοσοφικές θεωρήσεις τῆς ἔννοιας τῆς πιθανότητας. Οὔτε θά προσπαθήσουμε νά εξηγήσουμε τήν "αληθινή σημασία" τῆς ἔννοιας τῆς πιθανότητας. Δέν θά κάνουμε τίποτε περισσότερο από ὅτι ὁ σύγχρονος φυσικός ὅταν ασχολεῖται μέ τίς ἔννοιες τῆς μάζας καί ενέργειας ἤ ὁ γεωμέτρης ὅταν παρουσιάζει τήν ἔννοια τοῦ σημείου. Ἀντίθετα, θά δώσουμε τρόπους υπολογισμοῦ τῶν πιθανοτήτων ενδεχομένων, θεωρήματα καί εφαρμογές.

Ὑπάρχουν τρεῖς κύριοι ορισμοί τῆς πιθανότητας ενός ενδεχομένου: ὁ κλασσικός, ὁ εμπειρικός καί ὁ αξιωματικός. Ὁ κλασσικός ορισμός ξεκίνησε από τή μελέτη τῶν προβλημάτων πού εμφανίζονται στά τυχερά παιγνίδια. Ὁ εμπειρικός ορισμός στηρίζεται στήν εμπειρία πού απαιτεῖται από επαναλήψεις τυχαίων πειραμάτων. Ὁ αξιωματικός ορισμός εἶναι μαθηματικό δημιούργημα, συμβιβαστό μέ τόν εμπειρικό καί κλασσικό ορισμό. Οἱ ορισμοί αυτοί αναπτύσσονται στά επόμενα εδάφια.

Μεγάλα ονόματα στήν ἱστορία τῆς πιθανότητας εἶναι τά ονόματα τῶν *Bernoulli, de Moivre, Laplace, Gauss, Poisson, Chebychev, Markov* καί *Lyapounov*.

Ἡ εμπειρική ἤ στατιστική θεώρηση τῆς πιθανότητας ἀναπτύχθηκε κυρίως ἀπό τόν *R.A. Fisher* καί *R.Von Mises*. Ἡ ἔννοια τοῦ δειγματικοῦ χώρου ὀφείλεται στόν *Von Mises*. Ἡ ἔννοια αὐτή ἐπέτρεψε τήν δημιουργία μίας καθαρά μαθηματικῆς θεωρίας τῆς πιθανότητας μέ βάση τή θεωρία Μέτρου. Ἡ σύγχρονη ἀξιωματική θεώρηση ὀφείλεται στόν *A.Kolmogorov*.

2.3.1 Κλασσικός Ὁρισμός

Ὁ κλασσικός ὁρισμός ἀναφέρεται σέ πεπερασμένους δειγματικούς χώρους καί στηρίζεται στήν ἔννοια τοῦ"ἰσοπίθανου" τῶν ἀποτελεσμάτων ἑνός πειραματος ἤ τῶν σημείων τοῦ δειγματικοῦ χώρου. Λέγοντας "ἰσοπίθανα" ἀποτελέσματα ἐννοοῦμε ἀποτελέσματα μέ ἴσες πιθανότητες πραγματοποιήσεως. Γιά παράδειγμα οἱ δύο πλευρές ἑνός τέλειου νομίσματος, οἱ ἔξη πλευρές ἑνός τέλειου ζαριοῦ, οἱ πεντάδες στό πόκερ κλπ.

Ἔστω ὅτι τραβᾶμε ἕνα χαρτί ἀπό μία τράπουλα. Τά δυνατά ἀποτελέσματα εἶναι πενήντα δύο. Ἐφ'ὅσον ὑπάρχουν δεκατρία ἀπό τά πενενταδύο ἀποτελέσματα πού εἶναι σπαθιά ὅλοι θά συμφωνήσουν ὅτι, ἄν ἡ τράπουλα εἶναι καλοανακατεμένη, ἡ πιθανότητα νά τραβήξουμε σπαθί εἶναι $13/52$ καί ἡ πιθανότητα νά τραβήξουμε ἕνα ἄσσο εἶναι $4/52$. Αὐτό βέβαια ἰσχύει διότι τά ἀποτελέσματα τοῦ δειγματικοῦ χώρου θεωροῦνται ἰσοπίθανα.

Ἔστω ὅτι ρίχνομε δύο ζάρια. Τά δυνατά ἀποτελέσματα τοῦ πειράματος εἶναι τριάντα ἔξη:

$$\begin{array}{cccccc}
(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\
(2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\
(3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\
(4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\
(5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\
(6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6)
\end{array},$$

ὅπου $(3,4)$ σημαίνει ὅτι τό ἕνα ζάρι ἔφερε 3 καί τό ἄλλο 4. Τό ἀποτέλεσμα $(4,3)$ καταχωρεῖται διότι τά ζάρια θεωροῦνται διαφορετικά τό ἕνα ἀπό τό ἄλλο. Ἄν τά ζάρια εἶναι τέλεια τά τριάντα ἕξη ἀποτελέσματα εἶναι ἰσοπίθανα. Ἔτσι ἡ πιθανότητα νά φέρουμε ἕνα 5(δηλαδή τό ἄθροισμα τῶν ἀποτελεσμάτων σέ κάθε ζάρι νά εἶναι 5) εἶναι $4/36$ διότι ὑπάρχουν 4 ἀποτελέσματα πού φέρνουν τό 5, τά $(2,3)$, $(3,2)$, $(1,4)$ καί $(4,1)$.

Ὁρισμός 2.8: (Κλασσική πιθανότητα). Ἔστω A ἕνα ἐνδεχόμενο πού ἀποτελεῖται ἀπό n στοιχειώδη ἐνδεχόμενα σέ ἕνα δειγματικό χῶρο N ἰσοπίθανων ἐνδεχομένων. Ἡ πιθανότητα $P(A)$ τοῦ A εἶναι

$$P(A) = \frac{n}{N} \tag{2.1}$$

Τά n στοιχειώδη ἐνδεχόμενα πού ἀπαρτίζουν τό A λέγονται καί εὐνοϊκές περιπτώσεις τοῦ A.

Ὁ ὑπολογισμός τῶν n καί N στό τύπο (2.1) ἄλλοτε εἶναι ἁπλός καί ἄμεσος καί ἄλλοτε δύσκολος. Τά ἀποτελέσματα τῆς Συνδυαστικῆς (βλ. Κεφαλ. I §1.3) προσφέρουν πολύτιμη βοήθεια στούς ὑπολογισμούς τῶν κλασσικῶν πιθανοτήτων.

Παραδείγματα:

2.9: Ἕνα τέλειο ζάρι ρίχνεται μία φορά. Νά βρεθεῖ ἡ πιθανότητα τό ἀποτέλεσμα νά εἶναι ἄρτιο.

Ἔστω A={ἀποτέλεσμα ἄρτιο}. Ὑπάρχουν τρεῖς εὐνοϊκές περιπτώσεις γιά τό A οἱ E_2, E_4 καί E_6. Ἔτσι

$$P(A) = \frac{3}{6} = 0,5.$$

2.10: Ἕνα κουτί περιέχει 10 ἄσπρα, 4 μαῦρα καί 2 κόκκινα σφαιρίδια. Ἐάν ληφθοῦν 2 σφαιρίδια χωρίς ἐπανάθεση ἀπό τό κουτί, νά ὑπολογισθεῖ ἡ πιθανότητα (α) καί τά δύο σφαιρίδια νά εἶναι ἄσπρα, (β) καί τά δύο σφαιρίδια νά εἶναι κόκκινα καί (γ) ἕνα σφαιρίδιο νά εἶναι ἄσπρο καί ἕνα μαῦρο.

Δεδομένου ὅτι ἡ δειγματοληψία γίνεται χωρίς ἐπανάθεση θά χρησιμοποιήσουμε συνδυασμούς.

Εἶναι φανερό ὅτι τά ἀποτελέσματα τοῦ πειράματος αὐτοῦ εἶναι ἰσοπίθανα καί τό πλῆθος τους $\binom{16}{2}$. Ἄν A συμβολίζει τό ἄσπρο σφαιρίδιο, M τό μαῦρο καί K τό κόκκινο ἔχουμε

α) $$P\text{ (καί τά δύο σφαιρίδια ἄσπρα)} = P(2A) = \frac{\binom{10}{2}}{\binom{16}{2}}$$

$$= \frac{3}{8} = 0,375.$$

β) P(καί τά δύο σφαιρίδια κόκκινα) $= P(2K) = \dfrac{\binom{2}{2}}{\binom{16}{2}}$

$$= \frac{1}{120} = 0{,}003 \;.$$

γ) P(ένα σφαιρίδιο άσπρο καί ένα μαῦρο)$= P(1A$ καί $1M)$

$$= \frac{\binom{10}{1}\binom{4}{1}}{\binom{16}{2}} = 0{,}333,$$

(βλέπε Κεφάλαιο Ι, 'Εφαρμογή 1.2).

2.11: Δειγματοληψία χωρίς επανάθεση Ι. "Εστω $\{a_1,\ldots a_N\}$ ένας πεπερασμένος πληθυσμός N στοιχείων. n στοιχεῖα εκλέγονται άπό τόν πληθυσμό αὐτό χωρίς επανάθεση, τό ένα κατόπιν τοῦ ἄλλου καί μέ ἴσες πιθανότητες. 'Η πιθανότητα ένα όποιοδήποτε στοιχεῖο a_i τοῦ πληθυσμοῦ νά εκλεγεῖ (συμπεριληφθεῖ) στό (μή διατεταγμένο) δεῖγμα εἶναι

ταγμένο) δεῖγμα είναι

$$P(a_i \epsilon \text{δεῖγμα}) = \binom{N-1}{n-1}/\binom{N}{n} = \frac{n}{N}$$

διότι ὁ ἀριθμός τῶν δειγμάτων εἶναι $\binom{N}{n}$ [βλ. §1.3] καί οἱ τρόποι μέ τούς όποίους μποροῦν νά πληρωθοῦν οἱ $n-1$ θέσεις τοῦ δείγματος (ἡ μία θέση θά πληρωθεῖ μέ τό a_i) μέ τά $N-1$ ἄλλα στοιχεῖα εἶναι $\binom{N-1}{n-1}$.

2.12: Λέξεις τριῶν γραμμάτων κατασκευάζονται (χωρίς ἐπαναλήψεις καί ἀσχέτως νοήματος) ἀπό τά γράμματα τοῦ ἄλφαβήτου A,B,Γ,Δ καί E. Τά ἀποτελέσματα τοῦ δειγματικοῦ χώρου θεωροῦνται ἰσοπίθανα. Νά βρεθεῖ ἡ πιθανότητα μία τέτοια λέξη νά ἀρχίζει ἀπό A ἤ E.

Ὁ ἀριθμός ὅλων τῶν δυνατῶν λέξεων πού μποροῦν νά κατασκευασθοῦν εἶναι $P_3^5=5\cdot 4\cdot 3=60$. Ὁ ἀριθμός τῶν λέξεων πού ἀρχίζει μέ A εἶναι ὁ ἀριθμός τῶν τρόπων μέ τούς ὁποίους μποροῦμε νά γεμίσουμε τά κενά στό A-- μέ τά γράμματα B,Γ,Δ,E, δηλαδή $4\cdot 3=12$. Ἴδιος εἶναι καί ὁ ἀριθμός τῶν λέξεων πού ἀρχίζουν μέ Ε. Ἔτσι συνολικά ὁ ἀριθμός τῶν λέξεων πού ἀρχίζουν μέ A ἤ Ε εἶναι $2\cdot 4\cdot 3=24$. Ἄρα ἡ πιθανότητα εἶναι

$$\frac{2\cdot 4\cdot 3}{5\cdot 4\cdot 3}=\frac{2}{5}.$$

Ὁ κλασσικός ὁρισμός πρέπει νά θεωρηθεῖ ὡς ἕνας τρόπος ὑπολογισμοῦ μίας πιθανότητας καί ὄχι ὡς ὁ γενικός ὁρισμός τῆς πιθανότητας. Κατά πόσον ὁ ὑπολογισμός αὐτός ὁδηγεῖ στή σωστή ἀπάντηση ἐξαρτᾶται ἀπό τίς περιστάσεις καί τίς συνθῆκες. Ἄν ὅλα τά δειγματικά σημεῖα (ἤ δυνατά ἀποτελέσματα τοῦ πειράματος) εἶναι ἰσοπίθανα τότε ὁ κλασσικός ὁρισμός ὁδηγεῖ στή σωστή ἀπάντηση.

2.3.2 Ἐμπειρικός ἤ στατιστικός ὁρισμός.

Εἶναι τά ἀποτελέσματα ἑνός πειράματος πάντοτε ἰσοπίθανα; Ὄχι. Ἄν π.χ. τό νόμισμα δέν εἶναι τέλειο τά δύο ἀποτελέσματα Κ,Γ δέν εἶναι ἰσοπίθανα.

Εἶναι ἐπίσης δυνατό νά ξέρουμε ἐκ τῶν προτέρων τό ἰσοπίθανο τῶν ἀποτελεσμάτων ἑνός πειράματος; Ὄχι.

Βέβαια δεχόμαστε ὅτι ἄν ρίψουμε κάποιο νόμισμα (τέλειο ἤ μή) ὑπάρχουν ὁρισμένες πιθανότητες (ἴσως ἄγνωστες μας) νά πάρουμε Κ ἤ Γ.

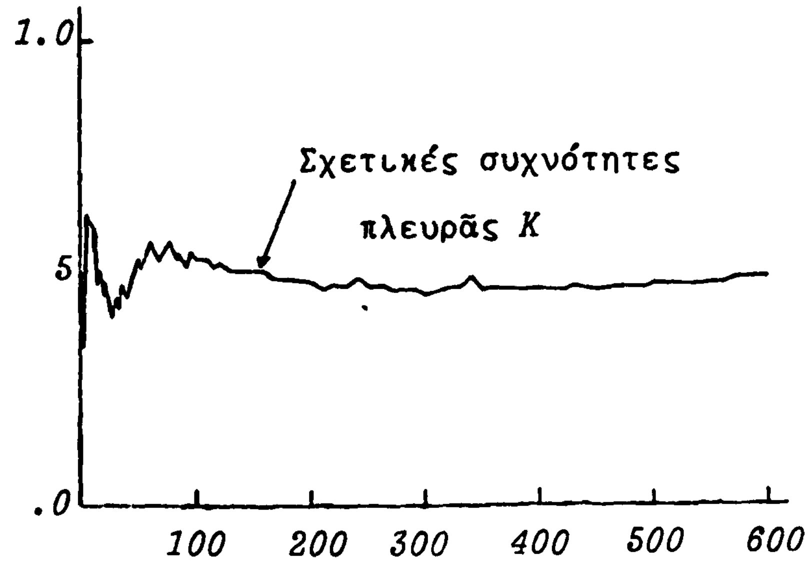
Εἶναι δύσκολο νά ὁρίσουμε ἱκανοποιητικά τίς προηγούμενες πιθανότητες ὅπως εἶναι δύσκολο νά ὁρίσουμε μέ ἀκρίβεια τίς ἔννοιες τῆς μάζας καί δύναμης στή φυσική. Περισσότερο μᾶς ἐνδιαφέρει ὁ τρόπος ὑπολογισμοῦ τῶν πιθανοτήτων αὐτῶν. Ὁ πιό προφανής τρόπος ὑπολογισμοῦ εἶναι νά ἐπαναλάβουμε τό πείραμα ἕνα ὁρισμένο ἀριθμό φορῶν N καί νά ὑπολογίσουμε τό κλάσμα (σχετική συχνότητα) τῶν φορῶν n πού ἐμφανίζεται ἕνα ἐνδεχόμενο. Τό κλάσμα $f=n/N$ εἶναι ἡ σχετική συχνότητα τοῦ ἐνδεχομένου. Ἄν σέ *1000* ρίψεις ἑνός νομίσματος ἡ πλευρά *Κ* ἐμφανιστεῖ *521* φορές ἡ πιθανότητα τοῦ Κ πρέπει νά εἶναι περίπου *52%*.

Ἡ ἐμπειρία ἔχει δείξει ὅτι ἡ σχετική συχνότητα μέ τήν ὁποία ἐμφανίζεται κάποιο ἐνδεχόμενο σέ πολλαπλές ἐπαναλήψεις τοῦ πειράματος τείνει νά σταθεροποιηθεῖ σέ κάποια συγκεκριμένη τιμή τήν ὁποία παίρνουμε ὡς τήν πιθανοτήτα τοῦ ἐνδεχομένου. Αὐτή ἡ παρατήρηση ἀποτελεῖ τή βάση τοῦ ἐμπειρικοῦ ὁρισμοῦ τῆς πιθανότητας καί λέγεται στατιστική ὁμαλότητα.

Ὁρισμός 2.9: (Ἐμπειρική ἤ Στατιστική πιθανότητα). Ἔστω n/N ἡ σχετική συχνότητα ἑνός ἐνδεχομένου σέ N ἐπαναλήψεις ἑνός πειράματος. Ἡ πιθανότητα $P(A)$ τοῦ A εἶναι

$$P(A) = \lim_{N\to\infty} \frac{n}{N} \quad . \qquad (2.2)$$

Τό Σχῆμα 2.2 περιγράφει τά ἀποτελέσματα *600* ρίψεων ἑνός νομίσματος. Ὁ ὁριζόντιος ἄξονας παριστάνει τόν ἀ-



Σχῆμα 2.2 Ρίψεις νομίσματος

ριθμό τῶν ρίψεων τοῦ νομίσματος καί τό ὕψος τῆς καμπύλης σέ ἕνα σημεῖο στόν ἄξονα τῶν x παριστάνει τή σχετική συχνότητα τῆς πλευρᾶς Κ μέχρι τό σημεῖο αὐτό. Ἡ σχετική συχνότητα τείνει πρός τήν τιμή *1/2* καί ἔτσι μποροῦμε νά ποῦμε ὅτι τό νόμισμα εἶναι τέλειο.

Οἱ φυσικές ποσότητες ὅπως π.χ. ἡ μάζα, δύναμη κλπ. μποροῦν συχνά νά ὑπολογιστοῦν μέ τά μαθηματικά. Ἔτσι καί ἡ πιθανότητα ἑνός ἐνδεχομένου. Ἡ πιθανότητα τοῦ κλασσικοῦ ὁρισμοῦ συμφωνεῖ μέ τήν πιθανότητα τοῦ ἐμπειρικοῦ ὁρισμοῦ ὅταν τά ἀποτελέσματα τοῦ πειράματος εἶναι ἰσοπίθανα.

2.3.3 Ἀξιωματικός ὁρισμός

Μέχρι πρόσφατα ἡ θεωρία τῶν Πιθανοτήτων ἦταν μία Μαθηματική Ἐπιστήμη τῆς ὁποίας οἱ βασικές της ἔννοιες δέν εἶχαν ὁρισθεῖ αὐστηρά. Αὐτή ἡ ἔλλειψη ὁδήγησε συχνά σέ μαθηματικά παράδοξα καί διαμάχες μεταξύ τῶν ἐπιστημόνων καί καλλιέργησε τό ἔδαφος γιά μιά ἀξιωματική θεμελίωση τῆς ἔννοιας τῆς πιθανότητας.

Ὁ ἀξιωματικός ὁρισμός τῆς πιθανότητας εἶναι καθαρό μαθηματικό δημιούργημα στηριζόμενο σέ τρία ἀξιώματα πού συμβιβάζονται μέ τήν ἀντίληψη τοῦ κλασσικοῦ καί ἐμπειρικοῦ ὁρισμοῦ. Γιά νά δώσουμε τόν ὁρισμό αὐτό χρειάζεται νά διευκρινήσουμε τίς ἔννοιες τοῦ δειγματικοῦ χώρου S καί τοῦ ἐνδεχομένου A.

Ἐάν ὁ δειγματικός χῶρος εἶναι διακεκριμένος (πεπερασμένος ἤ ἀριθμήσιμος) κάθε ὑποσύνολο του εἶναι ἕνα ἐνδεχόμενο. Ἔτσι ἐάν A_i, $i=1,2,\ldots$ εἶναι μία ἀκολουθία ἐνδεχομένων, $\bigcup_{i=1}^{\infty} A_i$ καί $\bigcap_{i=1}^{\infty} A_i$ εἶναι ἐπίσης ἐνδεχόμενα. Ἐάν ὅμως ὁ δειγματικός χῶρος εἶναι συνεχής, π.χ. ἡ εὐθεῖα γραμμή, ὑπάρχουν ὑποσύνολα πού δέν μποροῦν νά ἐκφρασθοῦν ὡς ἑνώσεις ἤ τομές πεπερασμένου ἤ ἀπείρου πλήθους ἐνδεχομένων π.χ. τῶν διαστημάτων. Γι'αὐτό τό λόγο καί πρός ἀπόκτηση ἑνός ὁμοιόμορφου συστήματος ἀξιωμάτων γιά τήν ἔννοια τῆς πιθανότητας περιορίζουμε τήν ἔννοια τοῦ ὅρου "ἐνδεχόμενο ἤ γεγονός" ἔτσι ὥστε νά μήν ἀναφέρεται σέ κάθε ὑποσύνολο τοῦ S ἀλλά μόνο σ'ἐκεῖνα τά ὑποσύν ο λ α πο ύ μ π ο ρ ο ῦ ν νά ἐ κ φ ρ α σ θ ο ῦ ν ὡ ς ἄ π ε ι ρ ε ς ἤ π ε π ε ρ α σ μ έ ν ε ς ἑ ν ώ - σ ε ι ς ἤ τ ο μ έ ς ἐ ν δ ε χ ο μ έ ν ω ν -"δ ι α - σ τ η μ ά τ ω ν". Καί ἡ οἰκογένεια αὐτή τῶν γεγονότων εἶναι ἀρκετά μεγάλη. Ἀπό πρακτικῆς πλευρᾶς ὁ περιορισμός αὐτός δέν ἔχει καμμιά συνέπεια.

Ἔτσι παράλληλα μέ τόν δειγματικό χῶρο S ἔχομε καί μία οἰκογένεια $\mathcal{A}$ ἀπό ὑποσύνολα-γεγονότα. Τώρα μποροῦμε νά δώσουμε τά ἀξιώματα πού ὁρίζουν τήν μαθηματική πιθανότητα (ἀξιώματα τοῦ *Kolmogorov*).

Ἀ ξ ι ώ μ α 1: Ἡ πιθανότητα ἑνός ἐνδεχομένου εἶναι ἕνας μή ἀρνητικός πραγματικός ἀριθμός, δηλ. $P(A)\geq 0$ γιά κάθε $A \epsilon \mathcal{A}$.

Ἀ ξ ί ω μ α 2: $P(S) = 1$. (2.3)

Ἀ ξ ι ώ μ α 3: Ἐάν $A_1, A_2, \ldots$ εἶναι μία π ε π ε ρ α σ μ έ ν η ἤ ἄ π ε ι ρ η ἀκολουθία ἀ σ υ μ β ί - β α σ τ ω ν ἀ ν ά δ ύ ο ἐνδεχομένων τότε

$$P(\bigcup_i A_i) = \sum_i P(A_i), \qquad (2.4)$$

ὅπου ὁ δείκτης i διατρέχει τούς ἀριθμούς $1,2,\ldots,n$ ἤ $1,2,\ldots$

Μέ ἄλλα λόγια πιθανότητα εἶναι μία συνολοσυνάρτηση πού ὀνομάζεται μ έ τ ρ ο π ι θ α ν ό τ η τ α ς μέ πεδίο ὁρισμοῦ τό $\mathcal{A}$, τιμές τούς μή ἀρνητικούς ἀριθμούς $[0,1]$, ἡ ὁποία ἱκανοποιεῖ τά Ἀξιώματα 2 καί 3. Τά Ἀξιώματα 1 καί 2 ἀνταποκρίνονται στή λογική καί τήν ἀντίληληψη τοῦ κλασσικοῦ καί ἐμπειρικοῦ ὁρισμοῦ. Τό Ἀξίωμα 3 εἶναι λογικό γιά πεπερασμένο πλῆθος ἐνδεχομένων: ἄν n_A καί n_B εἶναι οἱ ἀριθμοί τῶν εὐνοϊκῶν περιπτώσεων τοῦ A καί B ἀντίστοιχα καί τά A καί B εἶναι ἀσυμβίβαστα (δέν ἔχουν κοινά σημεῖα) τότε n_A+n_B εἶναι ὁ ἀριθμός n_{A+B} τῶν εὐνοϊκῶν περιπτώσεων τοῦ $A+B$. Ἔτσι $n_{A+B}/N=n_A/N+n_B/N$ ἤ $P(A+B)=P(A)+P(B)$. Γιά ἄπειρο πλῆθος ἀσυμβίβαστων ἀνά δύο ἐνδεχομένων ὑπάρχει κάποια δυσκολία. Ὑπάρχουν τεχνικοί (θεωρητικοί) λόγοι πού ἐπιβάλουν τό Ἀξίωμα 3 ἐκτός ἀπό τό γεγονός ὅτι δέν ἀντίκειται πρός τή διαίσθηση μας. Σέ μερικούς κλάδους τῆς Φυσικῆς ἀμφισβητεῖται ἡ ἰσχύς τοῦ Ἀξιώματος 2. Τά τρία ἀξιώματα δέν μᾶς λέγουν πῶς ὁρίζονται οἱ πιθανότητες διαφόρων γεγονότων. Ἁπλῶς καθορίζουν τούς νόμους οἱ ὁποῖοι πρέπει νά διέπουν μία τέτοια πράξη. Ὁ μαθηματικός ὁρίζει τίς πιθανότητες κατά κάποιο τρόπο αὐθαίρεται ἀρκεῖ να μήν παραβιάζονται τά παραπάνω ἀξιώματα. Γιά παράδειγμα, ἄν ἔχουμε ἕνα πείραμα μέ 5 διάφορα καί δυνατά ἀποτελέσματα Α,Β,Γ,Δ,Ε ἡ κατανομή

$$P(A)=0,1 \quad P(B)=0,1, \quad P(\Gamma)=0,1, \quad P(\Delta)=0,3, \quad P(E)=0,4$$

ἀποτελεῖ ἕνα σωστό τρόπο καθορισμοῦ τῶν πιθανοτήτων ἐνῶ ὁ τρόπος

$$P(A)=0,2, \quad P(B)=0,3, \quad P(\Gamma)=0,2, \quad P(\Delta)=0,4, \quad P(E)=0,2$$

παραβιάζει τά Ἀξιώματα 2 καί 3 διότι

$$P(S)=P(A+B+\Gamma+\Delta+E)=P(A)+P(B)+P(\Gamma)+P(\Delta)+P(E)$$

$$= 0,2+0,3+0,2+0,4+0,2 = 1/3.$$

2. 4. Ἰδιότητες τῶν πιθανοτήτων

Εἶναι εὔκολο τώρα νά ἀποδείξει κανείς αὐστηρά τίς κάτωθι προτάσεις:

Θεώρημα 2.1:

Ἐάν $A \in \mathcal{A}$ εἶναι ἕνα ὑποσύνολο ἑνός διακεκριμένου δειγματικοῦ χώρου καί $E_1, E_2, \ldots$ μιά πεπερασμένη ἤ ἄπειρη ἀκολουθία στοιχειωδῶν γεγονότων πού ἀντιπροσωπεύουν τό A δηλ. $A=\cup E_i$, τότε

$$P(A)=\Sigma P(E_i) \quad \text{ἤ} \quad P(A)=\sum_{s \in A} P(s), \qquad (2.5)$$

ὅπου $P(s)$ εἶναι ἡ πιθανότητα ἑνός δειγματικοῦ σημείου s.

Ἀπόδειξη: Ἐξ ὁρισμοῦ τά E_i εἶναι ἀσυμβίβαστα καί $A=\cup E_i$. Ἔτσι σύμφωνα μέ τό ἀξίωμα 3 ἔχουμε τό ἀποτέλεσμα. ▼

Θεώρημα 2.2: $P(A) \leq 1$ γιά κάθε $A \in \mathcal{A}$.

Ἀπόδειξη: Ἐπειδή $A \cup \overline{A}=S$, τά $A, \overline{A}$ εἶναι ἀσυμβίβαστα καί $P(S)=1$ ἔχουμε

$$P(A \cup \overline{A})=1=P(A)+P(\overline{A})$$

Ἔτσι

$$P(A)=1-P(\overline{A}) \leq 1. \blacktriangledown$$

Θ ε ώ ρ η μ α 2.3: $P(\emptyset) = 0$.

Ἀπόδειξη: Ἔχουμε $\emptyset \cup S=S$. Ἔτσι

$$P(\emptyset)+P(S)=P(S)$$

καί ἐπειδή $P(S)=1$

$$P(\emptyset)+1=1 \quad \text{ἤ} \quad P(\emptyset)=0. \ \blacktriangledown$$

Θ ε ώ ρ η μ α 2.4: $P(A)+P(\overline{A})=1$.

Ἀπόδειξη: Προφανής ἀπό τό θεώρημα 2.2. ▼

Θ ε ώ ρ η μ α 2.5:

<u>Ἐάν τό ἐνδεχόμενο</u> A <u>συνεπάγεται τό ἐνδεχόμενο</u> B, <u>τότε</u> $P(A)\leq P(B)$. <u>Συμβολικά</u>

$$A \rightsquigarrow B \implies P(A)\leq P(B) .$$

Ἀπόδειξη: $B=A\cup(B-A)$. Ἐπειδή τά ἐνδεχόμενα A καί $B-A$ εἶναι ἀσυμβίβαστα ἔχουμε

$$P(B)=P(A)+P(B-A)$$

Ἔτσι

$$P(B)\geq P(A). \ \blacktriangledown$$

Θ ε ώ ρ η μ α 2.6:

"Αν A καί B εἶναι δύο ἐνδεχόμενα τότε

$$P(A\cup B)=P(A)+P(B)-P(A\cap B) \tag{2.6}$$

'Απόδειξη: Τά ἐνδεχόμενα $A\cup B$ καί B μποροῦν νά γραφοῦν

$$A\cup B = A\cup(B-A\cap B)$$

$$B = (A\cap B)\cup(B-A\cap B)$$

ὅπου τά A, $B-A\cap B$ καί $A\cap B$, $B-A\cap B$ εἶναι ἀσυμβίβαστα. "Ετσι ἔχουμε

$$P(A\cup B) = P(A)+P(B-A\cap B)$$

$$P(B) = P(A\cap B)+P(B-A\cap B).$$

'Από τίς δύο αὐτές σχέσεις προκύπτει τό ἀποτέλεσμα. ▼

Τό παραπάνω θεώρημα λέγεται καί θ ε ώ ρ η μ α τ ῆ ς π ρ ο σ θ έ σ ε ω ς. 'Αφορᾶ ὁποιαδήποτε ἐνδεχόμενα A καί B ὄχι ἀναγκαστικά ἀσυμβίβαστα. Γενικεύεται εὔκολα γιά περισσότερα ἀπό δύο ἐνδεχόμενα. "Ετσι τό θεώρημα τῆς προσθέσεως γιά τρία ἐνδεχόμενα A,B,Γ εἶναι:

$$P(A\cup B\cup\Gamma) = P(A)+P(B)+P(\Gamma)-P(A\cap B)-P(A\cap\Gamma)$$

$$- P(B\cap\Gamma)+P(A\cap B\cap\Gamma).$$

Πόρισμα 2.1: 'Ανισότητα τοῦ *Boole*.

$$P(A \cup B) \leq P(A)+P(B)$$
$$P(A \cup B \cup \Gamma) \leq P(A)+P(B)+P(\Gamma) \quad \text{κλπ.} \tag{2.7}$$

Πόρισμα 2.2: 'Ανισότητα τοῦ *Bonferroni*.

$$P(AB) \geq 1-[P(A)+P(B)]$$
$$P(AB\Gamma) \geq 1-[P(A)+P(B)+P(\Gamma)] \quad \text{κλπ.} \tag{2.8}$$

Παραδείγματα:

2.13: Συνέχεια τοῦ Παραδείγματος 2.10. Νά βρεθοῦν οἱ πιθανότητες (α) τουλάχιστον ἕνα σφαιρίδιο νά εἶναι ἄσπρο, (β) τό πολύ ἕνα σφαιρίδιο νά εἶναι ἄσπρο, (γ) ἀκριβῶς ἕνα σφαιρίδιο νά εἶναι ἄσπρο, (δ) κανένα σφαιρίδιο νά εἶναι κόκκινο καί (ε) κανένα σφαιρίδιο νά εἶναι ἄσπρο.

α) P(τουλάχιστον ἕνα σφαιρίδιο ἄσπρο)

$$= P\{(1A \text{ καί } 1M) \text{ ἤ } (1A \text{ καί } 1K) \text{ ἤ } (2A)\}$$

$$= P(1A \text{ καί } 1M)+P(1A \text{ καί } 1K)+P(2A)$$

$$= \frac{\binom{10}{1}\binom{4}{1}}{\binom{16}{2}} + \frac{\binom{10}{1}\binom{2}{1}}{\binom{16}{2}} + \frac{\binom{10}{2}}{\binom{16}{2}} = \frac{7}{8} = 0{,}875 .$$

Τά ἐνδεχόμενα ($1A$ καί $1M$), ($1A$ καί $1K$), $2A$ εἶναι ἀσυμβίβαστα.

β) P(τό πολύ ἕνα σφαιρίδιο ἄσπρο)

$$= P\{(1A \text{ καί } 1M) \text{ ἤ } (1A \text{ καί } 1K) \text{ ἤ } (2M) \text{ ἤ } (1M \text{ καί } 1K) \text{ ἤ } (2K)\}$$

$$= P(1A \text{ καί } 1M)+P(1A \text{ καί } 1K)+P(2M) +P(1M \text{ καί } 1K)+ P(2K)$$

$$= \frac{\binom{10}{1}\binom{4}{1}}{\binom{16}{2}} + \frac{\binom{10}{1}\binom{2}{1}}{\binom{16}{2}} + \frac{\binom{4}{2}}{\binom{16}{2}} + \frac{\binom{4}{1}\binom{2}{1}}{\binom{16}{2}} + \frac{\binom{2}{2}}{\binom{16}{2}} = \frac{5}{8} = 0,625 \ .$$

Τά ἐπί μέρους ἐνδεχόμενα εἶναι ἐπίσης ἀσυμβίβαστα. Τό ἴδιο ἰσχύει καί γιά τίς ἑπόμενες περιπτώσεις.

γ) P(ἀκριβῶς ἕνα σφαιρίδιο ἄσπρο)

$$= P\{(1A \text{ καί } 1M) \text{ ἤ } (1A \text{ καί } 1K)\}$$

$$= P(1A \text{ καί } 1M) +P(1A \text{ καί } 1K)$$

$$= \frac{\binom{10}{1}\binom{4}{1}}{\binom{16}{2}} + \frac{\binom{10}{1}\binom{2}{1}}{\binom{16}{2}} = \frac{1}{2} = 0,5 \ .$$

δ) P(κανένα σφαιρίδιο κόκκινο)

$$= P\{(2A) \text{ ἤ } (1A \text{ καί } 1M) \text{ ἤ } (2M)\}$$

$$= P(2A)+P(1A \text{ καί } 1M)+P(2M)$$

$$= \frac{\binom{10}{2}}{\binom{16}{2}} + \frac{\binom{10}{1}\binom{4}{1}}{\binom{16}{2}} + \frac{\binom{4}{2}}{\binom{16}{2}} = \frac{91}{120} = 0,758 .$$

ε) Μέ ἀνάλογο τρόπο βρίσκουμε

$$P(\text{κανένα σφαιρίδιο ἄσπρο}) = \frac{1}{8} = 0,125 .$$

2.14: Ἔστω μία συνηθισμένη τράπουλα ἀπό τήν ὁποία τραβᾶμε ἕνα χαρτί. Νά βρεθεῖ ἡ πιθανότητα τό χαρτί νά εἶναι σπαθί ἤ φιγούρα.

Ἔστω A τό ἐνδεχόμενο τό χαρτί νά εἶναι σπαθί καί B τό ἐνδεχόμενο νά εἶναι φιγούρα. Ζητᾶμε τήν $P(A+B)$. Ἔχουμε $P(A)=13/52$, $P(B)=12/52$ διότι στήν τράπουλα ὑπάρχουν 12 φιγοῦρες καί $P(AB)=3/52$. Τά σπαθιά ἔχουν τρεῖς φιγοῦρες, τό ρῆγα, τό βαλέ καί τή ντάμα. Ἔτσι

$$P(A+B) = P(A)+P(B)-P(AB) = \frac{13}{52} + \frac{12}{52} - \frac{3}{52} = \frac{22}{52} .$$

2. 5. Δεσμευμένη πιθανότητα

Ἡ ἔννοια τῆς δεσμευμένης ἤ ὑπό συνθῆκες πιθανότητας στηρίζεται στήν ἔννοια τοῦ δεσμευμένου ἤ ὑπό συνθῆκες ἐνδεχόμενου. Καί οἱ δύο ἔννοιες εἶναι θεμελιώδους σημασίας. Θά ξεκινήσουμε μέ δύο παραδείγματα.

Ἔστω ἕνας πληθυσμός μέ N ἄτομα ἀπό τά ὁποῖα N_A

ἔχουν ἀχρωματοψία, N_B εἶναι γυναῖκες καί N_{AB} εἶναι ὁ ἀριθμός τῶν γυναικῶν πού πάσχουν ἀπό ἀχρωματοψία. Ἐκλέγεται ἕνα ἄτομο στήν τύχη ἀπό τόν πληθυσμό. Ἔστω A τό ἐνδεχόμενο τό ἄτομο πού ἐκλέχτηκε νά ἔχει ἀχρωματοψία καί B τό ἐνδεχόμενο νά εἶναι γυναίκα. Προφανῶς

$$P(A) = \frac{N_A}{N} \quad \text{καί} \quad P(B) = \frac{N_B}{N} \quad .$$

Ἀντί τοῦ συνολικοῦ πληθυσμοῦ ἐξετάζουμε τόν ὑποπληθυσμό τῶν γυναικῶν καί ἐνδαφερόμαστε γιά τήν πιθανότητα ἡ γυναίκα πού ἐκλέγεται στήν τύχη νά ἔχει ἀχρωματοψία. Ἡ πιθανότητα αὐτή εἶναι N_{AB}/N_B . Τίποτε τό καινούργιο δέν ἔχομε συναντήσει μέχρι τώρα. Χρειαζόμαστε ὅμως ἕνα συμβολισμό νά χαρακτηρίσουμε τόν ὑποπληθυσμό πού χρησιμοποιοῦμε. Ἔτσι γράφοντας $A|B$ ἐννοοῦμε τό ἐνδεχόμενο A (ἀχρωματοψία) δοθέντος τοῦ ἐνδεχόμενου B (τό ἄτομο πού ἐκλέχτηκε εἶναι γυναίκα) καί ὀνομάζουμε τό ἐνδεχόμενο αὐτό δ ε σ μ ε υ μ έ ν ο ἤ ὑ π ό σ υ ν θ ῆ κ ε ς. Τό σύμβολο $A|B$ διαβάζεται τό ἐνδεχόμενο A δοθέντος τοῦ B. Προφανῶς

$$P(A|B) = \frac{N_{AB}}{N_B} = \frac{P(AB)}{P(B)} \quad .$$

Ἡ ἀνωτέρω πιθανότητα λέγεται δ ε σ μ ε υ μ έ ν η ἤ ὑ π ό σ υ ν θ ῆ κ ε ς π ι θ α ν ό τ η τ α.

Ἔστω τώρα ἕνα μετεωρολογικό παράδειγμα. Ἄν A εἶναι τό ἐνδεχόμενο βροχῆς αὔριο καί B τό ἐνδεχόμενο συννεφιᾶς αὔριο, εἶναι φανερό ὅτι τό $A|B$={βροχή αὔριο δοθέντος ὅτι θά ἔχει συννεφιά} εἶναι ἕνα διαφορετικό ἐνδεχόμενο. $P(A)$ μπορεῖ νά θεωρηθεῖ ὡς τό κλάσμα τῶν ἡμερῶν μέ βροχή στό σύνολο τῶν ἀντιστοίχων ἡμερῶν γιά ὅλα τά προηγούμενα ἔτη γιά τά ὁποῖα ἡ Μετεωρολογική Ὑπηρεσία κρατάει ἀρχεῖο. $P(A|B)$ εἶναι τό κλάσμα τῶν ἡμερῶν μέ βροχή στό ὑποσύνολο τῶν ἀντιστοίχων ἡμερῶν μέ συννεφιά.

Ὁρισμός 2.10: Ἔστω ἕνας δειγματικός χῶρος S καί ἕνα ἐνδεχόμενο B μέ $P(B)>0$. Ἡ δεσμευμένη πιθανότητα ἑνός οἱουδήποτε ἐνδεχομένου A δοθέντος τοῦ B συμβολίζεται μέ $P(A|B)$ καί ὁρίζεται ὡς

$$P(A|B) = \frac{P(AB)}{P(B)} . \qquad (2.9)$$

Ἐάν $P(B)=0$, οἱ δεσμευμένες πιθανότητες δέν ὁρίζονται. Πρός ἀντιδιαστολή μέ τήν $P(A|B)$ ἡ $P(A)$ λέγεται καί ἀπόλυτος πιθανότητα.

Ἄν θεωρήσουμε τό B σταθερό καί ἐπιτρέψουμε στό A νά μεταβάλλεται στό A τότε ἡ $P(A|B)$ εἶναι ἕνα μέτρο πιθανότητας πού ἱκανοποιεῖ τά Ἀξιώματα 1, 2 καί 3. Δοθέντος τοῦ B ὁ δειγματικός χῶρος περιορίζεται στό B καί τά διάφορα ἐνδεχόμενα A περιορίζονται στά AB. Θά περίμενε κανείς ὅτι $P(A|B)\geq P(A)$. Ὅπως θά δοῦμε πιό κάτω ἡ σχέση αὐτή δέν ἀληθεύει πάντα.

Ἡ δεσμευμένη πιθανότητα ἔχει τίς ἀκόλουθες ἰδιότητες πού ἀποδεικνύονται χωρίς δυσκολία:

$$(i) \quad P(A|A) = 1 \qquad (ii) \quad P(\emptyset|A) = 0 \qquad (2.10)$$

$$(iii) \quad P(A\cup B|C) = P(A|C)+P(B|C)-P(AB|C) . \qquad (2.11)$$

Προσοχή: $P(A|B\cup C) \neq P(A|B)+P(A|C)$. (2.12)

Ἡ δεσμευμένη πιθανότητα τῆς δεσμευμένης πιθανότητας ὁρίζεται μέ τή σχέση

$$P(A|B|C) = P(A|B,C) = \frac{P(ABC)}{P(BC)} \quad . \qquad (2.13)$$

Παράδειγμα 2.15:

Ὁ παρακάτω πίνακας

	Πτυχιοῦχοι Πανεπιστημίου	Μή πτυχιοῦχοι Πανεπιστημίου	Σύνολο
Ἄνδρες	30	20	50
Γυναῖκες	10	40	50
Σύνολο	40	60	

δίνει τόν ἀριθμό τῶν ἀνδρῶν καί γυναικῶν πού εἶναι πτυχιοῦχοι ἤ μή σ'ἕνα σύνολο *100* ἀτόμων. Ἄν ἐκλέξουμε ἕνα ἄτομο στήν τύχη μποροῦμε εὔκολα ἀπό τόν πίνακα νά δοῦμε ὅτι

P(τό ἄτομο νά εἶναι πτυχιοῦχος |ἄνδρας)

$$= \frac{P(\text{τό ἄτομο νά εἶναι ἄνδρας πτυχιοῦχος})}{P(\text{τό ἄτομο νά εἶναι ἄνδρας})} = \frac{30/100}{50/100} = \frac{30}{50}$$

Μέ ἀπ'εὐθείας ὑπολογισμό εὑρίσκουμε ὅτι

$$P(\text{τό ἄτομο νά εἶναι πτυχιοῦχος}|\text{ἄνδρας}) = \frac{30}{50} \;.$$

Ὁμοίως ἔχουμε

$$P(\text{πτυχιοῦχος}|\text{Γυναίκα}) = \frac{10}{50}$$

$$P(\text{Ἄνδρας}|\text{πτυχιοῦχος}) = \frac{30}{40}$$

$$P(\text{Γυναίκα}|\text{Μή πτυχιοῦχος}) = \frac{40}{60} \;.$$

<u>Πολλαπλασιαστικός τύπος</u>.

Ἀπό τόν τύπο τῆς δεσμευμένης πιθανότητας ἔχουμε

$$P(AB)=P(A|B)P(B) = P(B|A)P(A)$$

μέ τήν προϋπόθεση ὅτι $P(A), P(B)>0$. Ἄν ἐφαρμόσουμε τήν μέθοδο τῆς τέλειας ἐπαγωγῆς μποροῦμε νά δοῦμε ὅτι ἄν A_1, $A_2,\dots,A_n$ εἶναι n ἐνδεχόμενα τότε

$$P(A_1A_2\dots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\dots P(A_n|A_1A_2\dots A_{n-1}) \;. \tag{2.14}$$

Ὁ τύπος αὐτός λέγεται π ο λ λ α π λ α σ ι α σ τ ι κ ός τ ύ π ο ς τῆς πιθανότητας.
Εἰδικώτερα γιά τρία ἐνδεχόμενα ἔχουμε

$$P(A_1A_2A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \;.$$

Θά δοῦμε τώρα μία ἁπλή ἐφαρμογή αὐτοῦ τοῦ τύπου. Ἄς ὑποθέσουμε ὅτι ἀπό μία τράπουλα τραβᾶμε διαδοχικά καί χωρίς ἐπανάθεση 5 χαρτιά καί ζητᾶμε τήν πιθανότητα νά τραβήξουμε πέντε σπαθιά.

῎Ας ὀνομάσουμε S_1, S_2, S_3, S_4, S_5 τά ἐνδεχόμενα ἡ πρώτη ἐκλογή εἶναι σπαθί, ἡ δεύτερη ἐκλογή εἶναι σπαθί, ..., ἡ πέμπτη ἐκλογή εἶναι σπαθί. Τότε ζητᾶμε τήν $P(S_1S_2S_3S_4S_5)$. ῎Εχουμε

$$P(S_1S_2S_3S_4S_5) = P(S_1)P(S_2|S_1)P(S_3|S_1S_2)P(S_4|S_1S_2S_3)$$

$$P(S_5|S_1S_2S_3S_4)$$

$$= \frac{13}{52}\,\frac{12}{51}\,\frac{11}{50}\,\frac{10}{49}\,\frac{9}{48} = \frac{\binom{13}{5}}{\binom{52}{5}} \simeq 0,0005 \; .$$

•

Χρειάζεται ἰδιαίτερη προσοχή ὅταν ἐφαρμόζουμε τόν πολλαπλασιαστικό τύπο γιατί ὑπάρχει κάποια σύγχυση στήν ἔκφραση $A_1 A_2 A_3 \cdots A_n$, πού εἶναι τό ἐνδεχόμενο πού συνίσταται στήν ταυτόχρονη πραγματοποίηση τῶν $A_1, A_2, \ldots, A_n$. ῎Αν μᾶς ἐνδιαφέρει ἡ πραγματοποίηση τῶν $A_1, A_2, \ldots, A_n$ νά γίνει μέ τήν σειρά, δηλ. πρῶτο τό A_1 καί ἀφοῦ πραγματοποιηθεῖ αὐτό μετά νά πραγματοποιηθεῖ τό A_2 καί ἀφοῦ πραγματοποιηθοῦν τά A_1, A_2 μετά νά πραγματοποιηθεῖ τό A_3 κ.ο.κ. τότε ἡ $P(A_1A_2 \ldots A_n)$ δίνεται ἀπό τόν πολλαπλασιαστικό νόμο. ῎Αν ὅμως δέν μᾶς ἐνδιαφέρει ἡ σειρά πραγματοποίησης τῶν $A_1, A_2, \ldots, A_n$ τότε γιά νά βροῦμε τήν $P(A_1A_2 \ldots A_n)$ θά πρέπει νά ἀθροίσουμε τίς πιθανότητες πού δίνονται ἀπό τόν πολλαπλασιαστικό νόμο γιά ὅλα τά ἐνδεχόμενα $A_{i_1} A_{i_2} \ldots A_{i_n}$, ὅπου $i_1, i_2, \ldots, i_n$ εἶναι κάθε δυνατή μετάθεση τῶν $1,2,\ldots,n$. Τό παράδειγμα πού ἀκολουθεῖ διασαφηνίζει αὐτή τήν παρατήρηση.

Παραδείγματα:

2.16: Ἔστω ὅτι ἀπό μία τράπουλα μέ 52 τραβάμε 5 χαρτιά διαδοχικά καί χωρίς ἐπανάθεση. Ποιά ἡ πιθανότητα τό πρῶτο χαρτί νά εἶναι σπαθί, τό δεύτερο νά εἶναι σπαθί, τό τρίτο νά εἶναι σπαθί, τό τέταρτο νά εἶναι καρρώ καί τό πέμπτο νά εἶναι καρρώ.

Ἄν καλέσουμε S_1, S_2, S_3 καί K_4, K_5 τά ἐπί μέρους ἀποτελέσματα

$$A = S_1 S_2 S_3 K_4 K_5$$

καί

$$P(A) = P(S_1)P(S_2|S_1)P(S_3|S_1S_2)P(K_4|S_1S_2S_3)$$

$$\cdot\, P(K_5|S_1S_2S_3K_4)$$

$$= \frac{13}{52}\,\frac{12}{51}\,\frac{11}{50}\,\frac{13}{49}\,\frac{12}{48} = 0{,}0009\ .$$

Ἄν ὅμως δέν μᾶς ἐνδιαφέρει ἡ σειρά ἐμφανίσεως ἀλλά ἁπλῶς θέλουμε τρία σπαθιά καί δύο καρρώ τότε, ἄν B εἶναι τό γεγονός αὐτό, ἔχουμε

$$P(B) = \frac{5!}{2!3!}\,P(A) = \frac{\binom{13}{3}\binom{13}{2}}{\binom{52}{5}} = 0{,}009,$$

ὅπου ὁ συντελεστής $5!/(2!3!)$ μᾶς δίνει τίς διάφορες ἐπαναληπτικές μεταθέσεις τῶν συμβόλων, S_1, S_2, S_3, K_4, K_5 ὅπου τά S_1, S_2, S_3 θεωροῦνται μή διακεκριμένα καί τό ἴδιο ἰσχύει καί γιά τά K_4, K_5. Τό A εἶναι ἕνα στοιχειῶδες ἐνδεχόμενο, δηλαδή ἕνα δειγματικό σημεῖο, ἐνῶ τό B εἶναι ἕνα σύνθετο ἐνδεχόμενο μέ δειγματικά σημεῖα τοῦ τύπου A. Γιά νά βροῦμε τήν πιθανότητα του θά πρέπει νά προσθέσουμε τίς πιθανότητες τῶν στοιχειωδῶν ἐνδεχομένων πού τό ἀποτελοῦν.

2.17: Δειγματοληψία χωρίς ἐπανάθεση II. Συνεχίζοντας τό Παράδειγμα 2.11 θά δείξουμε ὅτι στίς ἁπλές τυχαῖες δειγματοληψίες ἀπό πεπερασμένους πληθυσμούς χωρίς ἐπανάθεση ἡ πιθανότητα ἐκλογῆς ἑνός συγκεκριμένου στοιχείου σέ κάποια ἐκλογή εἶναι $1/N$. Ἔστω a_i τό στοιχεῖο καί ν ἡ ἐκλογή, $\nu=1,2,\ldots,n$. Ἔχουμε

P(ἐκλογῆς τοῦ a_i στήν ν-στή ἐκλογή)

$=P$(ὄχι τό a_i στήν 1η ἐκλογή, ὄχι τό a_i στή 2η ἐκλογή, ..., στήν ν-στή ἐκλογή)

$=P(\overline{a}_i$ στήν 1η ἐκλογή$)\, P(\overline{a}_i$ στήν 2η ἐκλογή$\,|\,$ὄχι τό a_i στήν 1η ἐκλογή$)\ldots$

$$=\frac{N-1}{N}\,\frac{N-2}{N-1}\cdots\frac{1}{N-\nu+1}=\frac{1}{N}\;.$$

2.6. Θεώρημα ὁλικῆς πιθανότητας καὶ Κανόνας τοῦ Bayes

Ὑπάρχουν πολλά προβλήματα στά ὁποῖα τό τελικό ἀποτέλεσμα ἐξαρτᾶται ἀπό τό τί γίνεται σέ ἐνδιάμεσες περιπτώσεις. Ἄς θεωρήσουμε τό ἀκόλουθο παράδειγμα. Ἔστω ὅτι σέ ἕνα πληθυσμό *5* ἄνδρες στούς *100* καί *25* γυναῖκες στίς *10000* ἔχουν ἀχρωματοψία. Ἄς ὑποθέσωμε ὅτι τά ποσοστά ἀνδρῶν καί γυναικῶν στόν ὑπό ἐξέταση πληθυσμό εἶναι τά ἴδια. Ἐάν ἐκλέξωμε στήν τύχη ἕνα ἄτομο ποία ἡ πιθανότητα νά ἔχει ἀχρωματοψία;

Ἔστω A τό γεγονός ὅτι τό ἐκλεγόμενο ἄτομο ἔχει ἀχρωματοψία, B τό ὅτι εἶναι ἄνδρας καί Γ τό ὅτι εἶναι γυναίκα. Τότε ἔχουμε

$$P(B) = \frac{1}{2}, \quad P(\Gamma) = \frac{1}{2}, \quad \Gamma = B'$$

$$P(A|B) = 0{,}05 \quad \text{καί} \quad P(A|\Gamma) = 0{,}0025$$

καί

$$A = AB \cup AB'.$$

Ἀλλά τά γεγονότα B, B' ἤ AB, AB' εἶναι ἀσυμβίβαστα καί σύμφωνα μέ τό Ἀξίωμα 3 θά ἔχουμε

$$P(A) = P(AB)+P(AB')$$

καί σύμφωνα μέ τήν πολλαπλασιαστική ἰδιότητα

$$P(A) = P(A|B)P(B)+P(A|B')P(B')$$

ἤ

$$P(A) = 0{,}05\times\frac{1}{2}+0{,}0025\times\frac{1}{2} = \frac{0{,}0525}{2} = 0{,}02525.$$

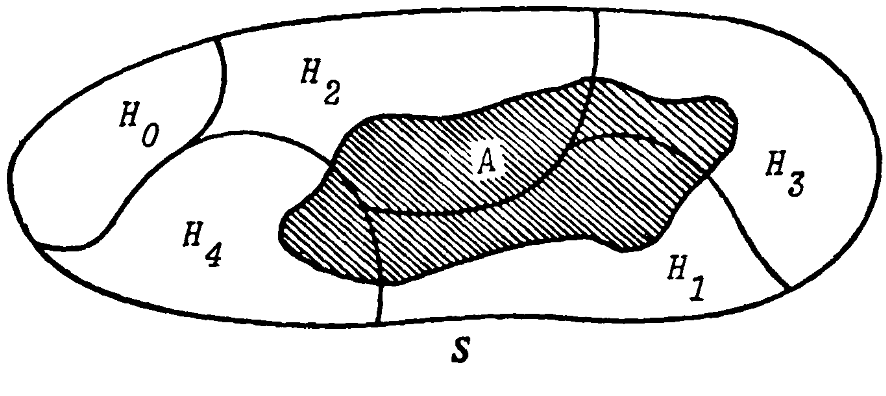
Ἡ γενίκευση τῆς ἀνωτέρω μεθόδου μᾶς δίνει τό θεώρημα τῆς ὁλικῆς πιθανότητας.

Θεώρημα 2.7: (Θεώρημα ὁλικῆς πιθανότητας).

Ἔστω $H_1, H_2, \ldots, H_k$ k ἀσυμβίβαστα μεταξύ τους ἐνδεχόμενα τό ἄθροισμα τῶν ὁποίων καλύπτει ὅλον τόν δειγματικό χῶρο S (δηλ. $S=H_1 U H_2 U \ldots U H_k$) καί $P(H_i)>0$ γιά κάθε $i=1,2,\ldots,k$. Τότε γιά κάθε ἐνδεχόμενο A ἔχουμε

$$P(A) = \sum_{i=1}^{k} P(AH_i) = \sum_{i=1}^{k} P(A|H_i)P(H_i) \qquad (2.15)$$

Ἀπόδειξη: Ἐπειδή τά $H_1, H_2, \ldots, H_k$ εἶναι ἀσυμβίβαστα καί καλύπτουν τό δειγματικό χῶρο ἔχουμε



$$A = (AH_1)U(AH_2)U\ldots U(AH_k).$$

Ἐπί πλέον τά $AH_1, AH_2, \ldots, AH_k$ εἶναι ἀσυμβίβαστα. Ἔτσι σύμφωνα μέ τό Ἀξίωμα

$$P(A) = \sum_{i=1}^{k} P(AH_i)$$

καί ἐπειδή $P(AH_i) = P(A|H_i)P(H_i)$ ἔχουμε τό ἀποτέλεσμα. ▼

Τό θεώρημα ὁλικῆς πιθανότητας μᾶς βοηθᾶ νά ὑπολογίσουμε πιθανότητες ἐνδεχομένων ὅταν γνωρίζουμε δεσμευμένες πιθανότητες. Τά $H_1, H_2,\ldots,H_k$ πρέπει νά εἶναι ἀσυμβίβαστα καί νά καλύπτουν τόν δειγματικό χῶρο. Τέτοια ἐνδεχόμενα ὀνομάζονται καί ἐ ξ α ν τ λ η τ ι κ ά.

Ἄς ἐπανέλθουμε τώρα στό προηγούμενο παράδειγμα καί ἄς ὑποθέσουμε ὅτι τό ἄτομο πού ἐκλέγεται ἔχει ἀχρωματοψία. Ποία ἡ πιθανότητα νά εἶναι ἄνδρας; Ἔχομε λοιπόν ὡς ἐνδεχόμενο τό ὅτι τό ἄτομο εἶναι ἄνδρας δοθέντος ὅτι ἔχει ἀχρωματοψία. Ἄρα

$$P(B|A) = \frac{P(BA)}{P(A)} = \frac{P(A|B)P(B)}{0{,}02625} = \frac{0{,}05\times 1/2}{0{,}02625} = \frac{0{,}05}{0{,}0525} = 0{,}9524.$$

Ἡ γενίκευση τοῦ ἀνωτέρω προβλήματος ἀποτελεῖ τόν κανόνα τοῦ *Bayes*, ὁ ὁποῖος διατυπώνεται ὡς ἑξῆς:

Θ ε ώ ρ η μ α 2.8: (Κανόνας τοῦ Bayes)

Ἔστω $H_1,\ldots,H_k$ k ἀσύμβιβαστα μεταξύ τους γεγονότα καλύπτοντα ὅλον τόν δειγματικό χῶρο S μέ $P(H_i) > 0$ γιά κάθε i. Τότε γιά κάθε ἐνδεχόμενο A μέ $P(A)>0$ ἔχουμε

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{\Sigma P(A|H_i)P(H_i)} \qquad (2.16)$$

γιά $i=1,2,\ldots,k$.

'Απόδειξη:

$$P(H_i|A) = \frac{P(H_i A)}{P(A)} = \frac{P(A|H_i)P(H_i)}{P(A)} .$$

"Αν έφαρμόσουμε τό θεώρημα τῆς όλικῆς πιθανότητας γιά τό A ἔχουμε τόν πρός ἀπόδειξη τύπο. ▼

Παράδειγμα 2.18:

"Εστω δύο δοχεῖα Δ_1 καί Δ_2. Τό πρῶτο περιέχει 2 ἄσπρες καί 8 κόκκινες μπάλες καί τό δεύτερο 6 ἄσπρες καί 4 κόκκινες μπάλες. 'Εκτελοῦμε τό ἑξῆς πείραμα: Ρίχνουμε ἕνα νόμισμα καί έάν τό άποτέλεσμα εἶναι ἡ πλευρά K διαλέγουμε τό δοχεῖο Δ_1, ἐάν εἶναι ἡ πλευρά Γ, διαλέγουμε τό δοχεῖο Δ_2. Μετά διαλέγουμε μία μπάλα ἀπό τό δοχεῖο πού έκλέχτηκε. "Εστω ὅτι ἡ μπάλα εἶναι ἄσπρη καί ἄς καλέσουμε A τό ένδεχόμενο αὐτό. "Αν ξεχάσουμε ποιό δοχεῖο διαλέξαμε νά βρεθοῦν οἱ πιθανότητες νά ἔχουμε διαλέξει τά δοχεῖα Δ_1 καί Δ_2.

"Εστω τά ένδεχόμενα

$\Delta_1|A$={διαλέξαμε τό δοχεῖο Δ_1|έξαχθεῖσα μπάλα ἦταν ἄσπρη}

καί

$\Delta_2|A$={διαλέξαμε τό δοχεῖο Δ_2|έξαχθεῖσα μπάλα ἦταν ἄσπρη}

'Η έκ τῶν προτέρων πιθανότητα νά διαλέξουμε τό δοχεῖο Δ_1 εἶναι

$P(\Delta_1) = 1/2$ καί τό δοχεῖο Δ_2 εἶναι $P(\Delta_2) = 1/2$."Εχουμε

$$P(\Delta_1|A) = \frac{P(\Delta_1 A)}{P(A)} = \frac{P(A|\Delta_1)P(\Delta_2)}{P(A|\Delta_1)P(\Delta_1)+P(A|\Delta_2)P(\Delta_2)}$$

$$= \frac{(2/10)(1/2)}{(2/10)(1/2)+(6/10)(1/2)} = 1/4$$

καί

$$P(\Delta_2|A) = \frac{P(A|\Delta_2)P(\Delta_2)}{(A|\Delta_1)P(\Delta_1)+P(A|\Delta_2)P(\Delta_2)} = 3/4 \ .$$

Οἱ πιθανότητες $P(\Delta_1|A)$ καί $P(\Delta_2|A)$, δηλ. οἱ πιθανότητες μετά τήν ἐκτέλεση τοῦ πειράματος λέγονται καί ἐ κ τ ῶ ν ὑ σ τ έ ρ ω ν π ι θ α ν ό τ η τ ε ς.

Τό θεώρημα τῆς ὁλικῆς πιθανότητας καί ὁ κανόνας τοῦ *Bayes* ἰσχύουν καί γιά ἀριθμήσιμο πλῆθος γεγονότων ἀσυμβίβαστων πού καλύπτουν τό S.

'Ορισμός 2.11: "Εστω ὁ δειγματικός χῶρος S καί $H_1, H_2, \ldots$ ἐνδεχόμενα τέτοια ὥστε $S = \cup H_i$ καί $H_i \cap H_j = \emptyset$, $i \neq j$. Τά ἐνδεχόμενα $H_1, H_2, \ldots$ λέγονται ἀσυμβίβαστα μεταξύ τους καί ἐ ξ α ν τ λ η τ ι κ ά ἤ ἁπλῶς λέμε ὅτι ἀποτελοῦν μία δ ι α μ έ ρ ι σ η τοῦ χώρου S.

'Εφαρμογή - 'Υποκειμενική Πιθανότητα

Οἱ πιθανότητες πού ἔχουμε συναντήσει μέχρι τώρα εἶναι ἀντικειμενικές μέ τήν ἔννοια ὅτι ἀποτελοῦν φυσικά χαρακτηριστικά τῶν ἐνδεχομένων. 'Υπάρχει ὅμως καί μία ἄλλη ὑποκειμενική - προσωπική θεώρηση τῆς πιθανότητας.

Ἄς θεωρήσουμε τήν ὑπόθεση ὅτι ὑπάρχει ζωή στόν Ἄρη. Ἔστω H ἡ ὑπόθεση αὐτή καί H' τό ἀντίθετο της, δηλαδή ὅτι δέν ὑπάρχει ζωή στόν Ἄρη. Ὁ βαθμός πίστης (ἐμπιστοσύνης) στήν H διαφέρει μεταξύ τῶν ἀνθρώπων. Ἄλλοι λένε ὅτι εἶναι ἀπίθανο νά ὑπάρχει ζωή στήν Ἄρη καί ἄλλοι ὅτι εἶναι σχετικά πιθανό νά ὑπάρχει ζωή ἐκεῖ. Ἡ θεωρία τῆς ὑποκειμενικῆς πιθανότητας δέχεται ὅτι εἶναι δυνατόν νά ὁρισθεῖ στό H μία πιθανότητα $P(H)$ ἡ ὁποία ἐκφράζει ἀριθμητικῶς τόν βαθμό πίστης τοῦ ἀτόμου στήν H. Ἡ ἰδέα εἶναι ὅτι ἡ $P(H)$ θά διαφέρει ἀπό ἄτομο σέ ἄτομο διότι ἔχουν διαφορετικές πληροφορίες γιά τό H καί ἐκτιμοῦν τίς πληροφορίες αὐτές κατά διαφορετικούς τρόπους. Ἡ θεώρηση αὐτή τῆς πιθανότητας δέν προϋποθέτει ἰσοπίθανα ἀποτελέσματα ἤ σχετικές συχνότητες πού συγκλίνουν σέ κάποια τιμή.

Ἐδῶ δέν θά ἀσχοληθοῦμε μέ τό πῶς ὁρίζονται οἱ ὑποκειμενικές πιθανότητες ἀλλά μέ τό πῶς τροποποιοῦνται κάτω ἀπό τό πρῖσμα νέας πληροφορίας ἤ δεδομένων.

Ἔστω ὅτι σέ πρόσφατη διαστημική ἀποστολή στόν Ἄρη μία συσκευή ἀνιχνεύσεως ζωῆς μᾶς στέλνει τήν πληροφορία (δεδομένα) ὅτι πράγματι ὑπάρχει ζωή ἐκεῖ. Αὐτή ἡ νέα πληροφορία, ἄς τήν καλέσουμε D, θά ἐπηρεάσει τό πιστεύω μας ὡς πρός τήν ὑπόθεση H. Ξέρουμε βέβαια ὅτι ἡ συσκευή ἀνιχνεύσεως ζωῆς δέν εἶναι ἀλάθητη. Ἀπό προηγούμενα ἐργαστηριακά τέστ γνωρίζουμε τήν πιθανότητα $P(D|H)$ ἡ συσκευή νά μεταδόσει ζωή ὅταν πράγματι ὑπάρχει ζωή καί τήν πιθανότητα $P(D|H')$ νά μεταδόσει ζωή ὅταν στήν πραγματικότητα δέν ὑπάρχει ζωή.

Ἔχουμε τώρα τά στοιχεῖα νά ὑπολογίσουμε μέ τή βοήθεια τοῦ κανόνα τοῦ *Bayes* τ ή ν έ α πιθανότητα $P(H|D)$ τῆς ὑποθέσεως H δοθείσης τῆς πληροφορίας D. Ἡ ἀρχική πιθανότητα $P(H)$ λέγεται ἐ κ τ ῶ ν π ρ ο τ ε ρ ω ν π ι θ α ν ό τ η τ α (*prior probability*) καί ἡ νέα πιθανότητα $P(H|D)$ λέγεται ἐ κ τ ῶ ν ὑ σ τ έ ρ ω ν π ι θ α ν ό τ η τ α (*posterior probability*). Ἡ τελευταία εἶναι ἡ προσωπική ὑποκειμενική πιθανότητα πού διαμορφώθηκε λαμβάνοντας ὑπόψη τήν πληροφορία D.

Ὁ κανόνας τοῦ *Bayes* ἦταν καί εἶναι ἕνα ἐπίμαχο καί ἀμφισβητούμενο θεώρημα. Ἀπό μαθηματικῆς πλευρᾶς δέν ὑπάρχει τίποτε τό διαφιλονικούμενο. Οἱ δυσκολίες ἀφοροῦν τίς ἐφαρμογές του, τήν ἑρμηνεία του καί ἔγκεινται στό κατά πόσον μποροῦμε λογικά νά ὁρίσουμε πιθανότητες σέ ὑποθέσεις. Μερικές φορές χρησιμοποιεῖται λανθασμένα ἀντιστρόφως γιά νά ἀποδειχθεῖ ἡ αἰτία ἀπό τό ἀποτέλεσμα, π.χ. ἕνα ἄτομο ἔχει καρκίνο ἐπειδή καπνίζει. Ἔτσι ὑπολογίζουμε τήν πιθανότητα νά ἔχει καρκίνο δοθέντος ὅτι καπνίζει καί ἄν εἶναι μεγάλη ἀποφαινόμαστε ὅτι τ ό κ ά π ν ι σ μ α προκαλεῖ τόν καρκίνο, δηλαδή ε ἶ ν α ι ἡ α ἰ τ ί α τ ο ῦ κ α ρ κ ί ν ο υ. Προφανῶς μία τέτοια ἑρμηνεία εἶναι ἐπικίνδυνη. Ὁ Πλάτων ἐχρησιμοποίησε ἐπιχειρήματα τοῦ τύπου αὐτοῦ γιά νά δείξει τήν ὕπαρξη τῆς Ἀτλαντίδας καί οἱ φιλόσοφοι ἔκαναν τό ἴδιο γιά νά δείξουν τό γελοῖο τῆς μηχανικῆς τοῦ Νεύτωνα. Οἱ μεταφυσικές ἐφαρμογές τοῦ τύπου τοῦ *Bayes* εἶναι ἐπικίνδυνες.

2. 7. Ἀνεξαρτησία

Ὅπως ἀναφέραμε στά προηγούμενα (§2.4) ὑπάρχουν ζεύγη ἐνδεχομένων γιά τά ὁποῖα ἰσχύει μία ἀπό τίς σχέσεις

$$P(A|B)>P(A) \quad , \quad P(A|B)=P(A) \quad , \quad P(A|B)<P(B) \; .$$

Γιά παράδειγμα, ἄν ρίξουμε ἕνα ζάρι καί πάρουμε τά ἐνδεχόμενα A={ἄρτιο ἀποτελέσμα} , B={ἀποτέλεσμα μικρότερο ἤ ἴσο τοῦ 4}, $Γ$={ἀποτέλεσμα μικρότερο ἤ ἴσο τοῦ 3}
τότε

$$P(A|B)=\frac{1}{2} \quad \text{καί} \quad P(A)=\frac{1}{2} \qquad \text{ἄρα} \qquad P(A|B)=P(A),$$

$$P(A|\Gamma)=\frac{1}{3} \quad \text{καί} \quad P(A)=\frac{1}{2} \qquad \text{ἄρα} \qquad P(A|\Gamma)<P(A),$$

$$P(\Gamma|B)=\frac{3}{4} \quad \text{καί} \quad P(\Gamma)=\frac{1}{2} \qquad \text{ἄρα} \qquad P(\Gamma|B)>P(\Gamma) \; .$$

Ὅταν $P(A|B)=P(A)$ τότε εἶναι φανερό ὅτι ἡ πληροφορία πού δίδεται ἀπό τήν πραγματοποίηση τοῦ B δέν ἐπηρεάζει (αὐξάνει ἤ ἐλαττώνει) τήν πιθανότητα τῆς πραγματοποίησης τοῦ A καί λέμε ὅτι τά δύο ἐνδεχόμενα εἶναι ἀνεξάρτητα.

Ὁρισμός 2.12: Ἔστω δύο ἐνδεχόμενα A καί B μέ $P(A)>0$ καί $P(B)>0$. Τά ἐνδεχόμενα A καί B θά λέγονται στατιστικῶς ἤ στοχαστικῶς ἀνεξάρτητα ἄν καί μόνον ἄν ἰσχύει μία ἀπό τήν παρακάτω σχέσεις:

$$\text{(i)} \quad P(A|B) = P(A)$$

$$\text{(ii)} \quad P(B|A) = P(B) \qquad (2.17)$$

$$\text{(iii)} \quad P(AB) = P(A)P(B).$$

Εἶναι εὔκολο, χρησιμοποιοῦντες τήν ἔννοια τῆς δεσμευμένης πιθανότητας, νά δοῦμε ὅτι οἱ τρεῖς αὐτές σχέσεις εἶναι ἰσοδύναμες.

Παραδείγματα:

2.19: Ἔστω μία συνηθισμένη τράπουλα ἀπό τήν ὁποία τραβάμε ἕνα χαρτί. Γιά λόγους συμμετρίας τά ἐνδεχόμενα "σπαθί" καί "ἄσσος" θά πρέπει νά εἶναι ἀνεξάρτητα. Πράγματι οἱ πιθανότητες εἶναι $1/13$ καί $1/4$ ἀντίστοιχα καί ἡ πιθανότητα ταυτόχρονης πραγματοποίησης εἶναι $1/52 = (1/13)(1/4)$.

2.20: Ἄς θεωρήσουμε τό σύνολο ὅλων τῶν οἰκογενειῶν πού ἔχουν τρία παιδιά. Ἐκλέγουμε μία ὁποιαδήποτε οἰκογένεια καί σημειώνουμε τήν τάξη μέ τήν ὁποία γεννήθηκαν τά παιδιά καί τό φύλο π.χ. τό πρῶτο ἀγόρι, A, τό δεύτερο κορίτσι , K, τό τρίτο κορίτσι K. Ἔστω τά ἐνδεχόμενα

Γ={ἡ οἰκογένεια ἔχει παιδιά καί τῶν δύο φύλων}

Δ={ἡ οἰκογένεια ἔχει τό πολύ ἕνα ἀγόρι}

Εἶναι τά Γ καί Δ ἀνεξάρτητα;

Ὁ δειγματικός χῶρος S ἔχει ὀκτώ σημεῖα δηλαδή

$$S = \{AAA, AAK, AKA, KAA, AKK, KKA, KAK, KKK\}.$$

Ἔτσι εὔκολα μποροῦμε νά δοῦμε ὅτι $P(\Gamma)=6/8$ καί $P(\Delta) = 4/8$ Τό ἐνδεχόμενο $\Gamma\Delta$ σημαίνει ὅτι ἡ οἰκογένεια ἔχει ἀκριβῶς ἕνα ἀγόρι.

Δηλαδή

$$\Gamma\Delta = \{AKK, KKA, KAK\}$$

καί ἔτσι

$$P(\Gamma\Delta) = 3/8 .$$

Ἐξ ἄλλου

$$P(\Gamma)P(\Delta) = 3/8 .$$

Βλέπουμε δηλαδή ὅτι ἰσχύει ἡ σχέση $P(\Gamma\Delta) = P(\Gamma)P(\Delta)$. Συνεπῶς τά Γ καί Δ εἶναι ἀνεξάρτητα. Μποροῦμε νά δοῦμε ὅτι τά Γ καί Δ δέν εἶναι ἀνεξάρτητα σέ οἰκογένειες μέ δύο ἤ τέσσερα παιδιά.

Ὁ ἔλεγχος τῆς ἀνεξαρτησίας ἐνδεχομένων δέν εἶναι πάντοτε εὔκολος. Ἄλλοτε μποροῦμε νά ὑπολογίσουμε τίς πιθανότητες $P(A)$, $P(B)$, $P(AB)$ ἤ $P(A|B)$ ἤ $P(B|A)$ καί νά ἐλέγξουμε μέ τήν (2.17) ἄν τά ἐνδεχόμενα A καί B εἶναι ἀνεξάρτητα. Ἄλλοτε ὅμως ὑ π ο θ έ τ ο υ μ ε ὅτι τά A καί B εἶναι ἀνεξάρτητα στηριζόμενοι στό εἶδος τῶν A καί B καί τά φυσικά χαρακτηριστικά τοῦ πειράματος πού παράγει τά A καί B.

<u>Παράδειγμα</u> <u>2.21</u>:

Δύο ζάρια ρίχνονται δύο φορές. Ἔστω A τό ἐνδεχόμενο ὅτι τό ἀποτέλεσμα (ἄθροισμα) τῆς πρώτης ρίψης εἶναι 9 καί B τό ἀποτέλεσμα τῆς δεύτερης ρίψης εἶναι 4. Εἶναι εὔκολο νά δοῦμε ὅτι $P(A)=4/36$ καί $P(B)=3/36$. 'Επίσης εἶναι φυσικό νά ὑποθέσουμε ὅτι τά ἐνδεχόμενα A καί B εἶναι ἀνεξάρτητα. Τό ἀποτέλεσμα 9 δέν ἐπηρεάζει κατά κανένα τρόπο τή πραγματοποίηση τοῦ 4. Ἔτσι

$$P(AB) = \frac{4}{36}\,\frac{3}{36} = \frac{1}{3} \quad .$$

Βέβαια ἡ ἀνεξαρτησία μπορεῖ νά ἐπιβεβαιωθεῖ ὑπολογίζοντας ἀπ'εὐθείας τήν $P(AB)$. Ρίχνοντας δύο ζάρια δύο φορές εἶναι ἰσοδύναμο μέ τό νά ρίξουμε ἕνα ζάρι τέσσερες φορές. Τά δυνατά ἀποτελέσματα εἶναι $36^4 = 1296$. Τά εὐνοϊκά γιά τά AB εἶναι 12.

'Ορισμός 2.13: Ἔστω $\{A_1, A_2, \ldots, A_n\}$ ἕνα σύνολο n ἐνδεχομένων τοῦ δειγματικοῦ χώρου S. 'Εάν ἰσχύει ἡ σχέση

$$P(A_{i_1} A_{i_2} \ldots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \ldots P(A_{i_k}) \qquad (2.18)$$

γιά κ ά θ ε ὑ π ο σ ύ ν ο λ ο $(i_1, i_2, \ldots, i_k)$ τοῦ συνόλου $\{1, 2, 3, \ldots, n\}$, τότε τά ἐνδεχόμενα $A_1, A_2, \ldots, A_n$ λέγονται στατιστικῶς ἤ στοχαστικῶς ἀνεξάρτητα.

Μέ ἄλλα λόγια ἡ παραπάνω σχέση θά πρέπει νά ἰσχύει γιά δυάδες, τριάδες, τετράδες,... n-άδες ἐνδεχομένων γιά νά λέμε ὅτι τά ἐνδεχόμενα εἶναι ἀνεξάρτητα. Πολλές φορές στήν περίπτωση περισσοτέρων ἀπό δύο ἐνδεχομένων χρησιμοποιεῖται καί ὁ ὅρος "ὁλικῶς ἀνεξάρτητα". Γιά συντομία στήν ἔκφραση θά χρησιμοποιοῦμε ἁπλῶς τόν ὅρο ἀνεξάρτητα ἐνδεχόμενα.

Εἶναι εὔκολο νά δοῦμε ὅτι ὁ ἀριθμός τῶν σχέσεων τῆς μορφῆς (2.18) πού θά πρέπει νά ἰσχύει γιά νά εἶναι n ἐνδεχόμενα ἀνεξάρτητα εἶναι

$$\binom{n}{2} + \binom{n}{3} +\ldots+\binom{n}{n} = 2^n - \binom{n}{1}-\binom{n}{0}$$

$$= 2^n-n-1 \ .$$

Ἰδιότητες ἀνεξαρτησίας

1. Ἄν δύο ἐνδεχόμενα A καί B εἶναι ἀσυμβίβαστα, δηλαδή $AB=\emptyset$, μέ $P(A)>0$ καί $P(B)>0$ τότε τά ἐνδεχόμενα αὐτά δέν μπορεῖ νά εἶναι ἀνεξάρτητα διότι $P(AB) =0$. Μέ ἄλλα λόγια οἱ ἔννοιες ἀνεξαρτησίας καί "ἀσυμβίβαστα" εἶναι ἐκ διαμέτρου ἀντίθετες.

2. Ἄν τά ἐνδεχόμενα A καί B εἶναι ἀνεξάρτητα τότε τά ζεύγη (A,B'), (A',B), (A',B') εἶναι ζεύγη ἀνεξάρτητων ἐνδεχομένων. Ὁμοίως ἄν τά ἐνδεχόμενα A,B,Γ εἶναι ἀνεξάρτητα τό ἴδιο ἰσχύει καί γιά τίς τριάδες ἐνδεχομένων (A',B,Γ), (A,B',Γ), (A,B,Γ'), (A',B',Γ), (A',B,Γ'), (A,B',Γ'), (A',B',Γ') κ.ο.κ.

3. Ἡ ἀνά ζεύγη ἀνεξαρτησία δέν συνεπάγεται καί ὁλική ἀνεξαρτησία.

<u>Ἀντιπαράδειγμα</u> <u>2.22</u>:

Ἔστω ὅτι ρίχνουμε δύο ζάρια. Ἔστω ἐπίσης τά ἐνδεχόμενα

A_1 = {ἀποτέλεσμα πρώτου ζαριοῦ ἄρτιο}

A_2 = {ἀποτέλεσμα δεύτερου ζαριοῦ ἄρτιο}

A_3 = {ἄθροισμα ἀποτελεσμάτων ἄρτιος}.

Τότε

$$P(A_1) = P(A_2) = P(A_3) = 1/2$$

$$P(A_1A_2) = P(A_1A_3) = P(A_2A_3) = 1/4$$

$$P(A_1A_2A_3) = 1/4 \neq P(A_1)P(A_2)P(A_3) = 1/8 \quad .$$

Βλέπουμε δηλαδή ὅτι τά ἐνδεχόμενα εἶναι ἀνεξάρτητα ἀνά δύο ἀλλά ὄχι καί ὁλικῶς ἀνεξάρτητα.

2. 8. Ἐπαναλαμβανόμενα καὶ ἀνεξάρτητα πειράματα

Στό ἐδάφιο 2.1 ἀναφέραμε ὅτι ὑπάρχουν σύνθετα πειράματα πού ἀποτελοῦνται ἀπό ἄλλα ἐπί μέρους πειράματα. Ἄλλες φορές ἔχουμε ἕνα πείραμα πού ἐπαναλαμβάνεται δύο ἤ περισσότερες φορές. Ἔστω Π_1 καί Π_2 δύο ἐπί μέρους πειράματα μέ δειγματικούς χώρους S_1 καί S_2 καί ἐνδεχόμενα A_1 καί A_2 ἀντίστοιχα. Τό σύνθετο πείραμα συμβολίζεται μέ $\Pi=\Pi_1\times\Pi_2$ καί ὁ δειγματικός του χῶρος μέ $S=S_1\times S_2$. Ἐνδεχόμενα τοῦ νέου δειγματικοῦ χώρου εἶναι τά $A=A_1\times A_2$. Μέ ἁπλά λόγια $A_1\times A_2$ σημαίνει A_1A_2, δηλαδή πραγματοποίηση τοῦ A_1 καί τοῦ A_2. Ὑπάρχουν βέβαια καί ἄλλα ἐνδεχόμενα τοῦ S πού ἔχουν διαφορετική μορφή. Ἄν τά πειράματα Π_1 καί Π_2 εἶναι τά ἴδια, ἔστω ἴσα μέ τό Π^* καί πραγματοποιοῦνται διαδοχικά τότε λέμε ὅτι τό Π^* εἶναι ἐ π α ν α λ α μ β α ν ό μ ε ν ο. Ἄν

$$P(A_1 \times A_2) = P(A_1 A_2) = P(A_1)P(A_2)$$

γιά κ ά θ ε ἐ ν δ ε χ ό μ ε ν ο A_1 τ ο ῦ Π_1 κ α ί κ ά θ ε ἐ ν δ ε χ ό μ ε ν ο A_2 τ ο ῦ Π_2 τότε τά πειράματα Π_1 καί Π_2 λέγονται ἀ ν ε ξ ά ρ τ η τ α. Ἄν $\Pi_1 = \Pi^*$ καί $\Pi_2 = \Pi^*$ τότε λέγομε ὅτι τό πείραμα Π^* εἶναι ἀ ν ε ξ ά ρ τ η τ α ἐ π α ν α λ α μ β α ν ό μ ε ν ο. Γιά ἐπαναλαμβανόμενα πειράματα ὁ συμβολισμός $A_1 \times A_2$ σημαίνει ὅτι πρῶτα πραγματοποιεῖται τό A_1 καί κατόπιν τό A_2. Οἱ παραπάνω ἔννοιες γενικεύονται εὔκολα καί γιά περισσότερα ἀπό δύο πειράματα. Ἔτσι k πειράματα $\Pi_1, \Pi_2, \ldots, \Pi_k$ εἶναι ἀνεξάρτητα ὅταν κάθε σύνολο ἀπό k ἐνδεχόμενα, ἕνα ἀπό κάθε πείραμα, ἀποτελεῖ σύνολο ἀνεξάρτητων ἐνδεχομένων.

Ὅταν τά πειράματα εἶναι ἀνεξάρτητα ἡ πραγματοποίηση ἑνός ἐνδεχομένου τοῦ ἑνός πειράματος δέν ἐπηρεάζει τήν πραγματοποίηση τῶν ἀποτελεσμάτων τῶν ἄλλων πειραμάτων. Παραδείγματα ἀνεξάρτητων πειραμάτων εἶναι τό ρίψιμο ἑνός ζαριοῦ δύο ἤ περισσότερες φορές, ἡ δειγματοληψία n ἀντικειμένων μέ ἐπανάθεση ἀπό ἕνα πεπερασμένο πληθυσμό N ἀντικειμένων, ἡ ἐκπομπή σωματιδίων ἀπό δύο ἀνεξάρτητες πηγές κλπ.

Ὅπως καί μέ τά ἐνδεχόμενα ὁ ἔλεγχος τῆς ἀνεξαρτησίας πειραμάτων δέν εἶναι τόσο εὔκολος. Συχνά ἀποφαινόμαστε ὅτι τά πειράματα εἶναι ἀνεξάρτητα ἄν ἡ διαδικασία ἐκτελέσεως τοῦ ἑνός εἶναι τελείως ἀνεξάρτητη ἀπό τή διαδικασία ἐκτελέσεως τῶν ἄλλων.(

ΑΣΚΗΣΕΙΣ

2.1. Ἕνα κλουβί περιέχει πέντε ὅμοια κουνέλια, μερικά ἀπό τά ὁποῖα ἔχουν ἐμβολιαστεῖ γιά κάποια νόσο.
α) Ἐάν ἕνα κουνέλι εἶναι ἐμβολιασμένο νά βρεθεῖ ἡ πιθανότητα ἐκλογῆς του σέ δεῖγμα μεγέθους ἕνα.
β) Ἐάν δύο κουνέλια εἶναι ἐμβολιασμένα νά βρεθεῖ ἡ πιθανότητα ἐκλογῆς τους σέ δεῖγμα μεγέθους τρία.
γ) Ἐάν δύο κουνέλια ἐμβολιάσθησαν, ἕνα τήν Τετάρτη καί ἕνα τήν Πέμπτη, νά βρεθεῖ ἡ πιθανότητα ἐκλογῆς τους μέ τή σειρά ἐμβολιασμοῦ σέ δεῖγμα μεγέθους δύο.
Ὅλα τά ἀνωτέρω δείγματα λαμβάνονται χωρίς ἐπανάθεση.

2.2. Ἕξη ζάρια ρίχνονται. Ποία εἶναι ἡ πιθανότητα νά ἐμφανιστοῦν ὅλα τά ἀποτελέσματα *1, 2, 3, 4, 5, 6*;

2.3. Δώδεκα γεωμετρικά σχήματα δίδονται σέ ἕνα πίθηκο, τρία σχήματος τετραγώνου, τρία ὀρθογωνίου, τρία κύκλου καί τρία τριγώνου. Ὁ πίθηκος ἐκλέγει τρία κάθε εἴδους στή σειρά, π.χ. τρία τρίγωνα, τρία τετράγωνα, κλπ. Ἔχει ὁ πίθηκος τήν ἱκανότητα νά συσχετίζει ὅμοια σχήματα; Νά ὑπολογιστεῖ ἡ πιθανότητα τοῦ παραπάνω ἐνδεχομένου.

2.4. Ρίχνονται τρία ζάρια. Ποιά εἶναι ἡ πιθανότητα τό ἄθροισμα τῶν ἀποτελεσμάτων (πλευρές) νά εἶναι *12*; Ποιά εἶναι ἡ πιθανότητα νά ἐμφανιστεῖ τό ἕνα *(1)* τουλάχιστον μιά φορά.

2.5. Ἕνα κουτί περιέχει *2* ἄσπρες, *3* κόκκινες καί *5* μαῦρες μπάλες. Ἕνα δεύτερο κουτί περιέχει 1 ἄσπρη, 6 κόκκινες καί 3 μαῦρες μπάλες. Μιά μπάλα ἐκλέγεται τυχαῖα ἀπό κάθε κουτί. Νά βρεθεῖ ἡ πιθανότητα καί οἱ δυό μπάλες πού ἐκλέχτηκαν νά εἶναι διαφορετικοῦ χρώματος.

2.6. Πέντε χαρτιά ἐκλέγονται ἀπό μία συνηθισμένη τράπουλα τῶν 52 χαρτιῶν. Ἔστω τά ἐνδεχόμενα A = {καί τά πέντε χαρτιά εἶναι σπαθιά}, B ={τά πέντε χαρτιά εἶναι ἄσσος, ρήγας, ντάμα, βαλές, δέκα ὅλα τοῦ ἴδιου χρώματος} , C = {τό ἐνδεχόμενο B χωρίς τόν περιορισμό τοῦ χρώματος}. Νά βρεθοῦν οἱ πιθανότητες $P(A)$, $P(B)$, $P(C)$ καί $P(A+B)$.

2.7. Ἕνα κουτί περιέχει 10 λευκά, 4 μαῦρα καί 2 κόκκινα σφαιρίδια. Ἐάν ληφθοῦν χωρίς ἐπανάθεση δύο σφαιρίδια νά ὑπολογισθεῖ ἡ πιθανότητα τό πολύ ἕνα ἀπό τά σφαιρίδια νά εἶναι λευκό.

2.8. Δύο ἐνδεχόμενα A_1 A_2, εἶναι ἀνεξάρτητα μέ πιθανότητα 0,1 τό κάθε ἕνα. Νά βρεθεῖ ἡ $P(A_1 \cup A_2)$.

2.9. Ξέροντας ὅτι τά ἐνδεχόμενα A,B,C εἶναι ἀνά δύο ἀσυμβίβαστα, ἐξετᾶστε ποιές ἀπό τίς παρακάτω πιθανότητες εἶναι σωστές.

$i)$ $P(A) = 0,4$, $P(B) = 0,4$, $P(A \cup B) = 0,3$.

$ii)$ $P(A) = 0,6$, $P(A \cap B') = 0,5$.

2.10. Ἄν A,B εἶναι δύο ἐνδεχόμενα τοῦ δειγματικοῦ χώρου S μέ $A \subset B$ δεῖξτε ὅτι

$$P(A) \leq P(B) \quad \text{καί} \quad P(B-A) = P(B)-P(A).$$

2.11. Έστω A, B δύο ενδεχόμενα σέ ένα δειγματικό χῶρο S. Νά δειχθοῦν οἱ ἀκόλουθες σχέσεις

i) $P(A \cap B) \geq P(A) + P(B) - 1$

ii) Ἐάν $P(A/B) > P(A)$ τότε $P(B/A) > P(B)$.

2.12. Δίδεται ὅτι $P(A) = 0{,}48$, $P(B) = 0{,}37$ καί $P(AB) = 0{,}13$. Νά βρεθοῦν (*i*) $P(A \cup B)$, (*ii*) $P(A' \cup B')$, (*iii*) $P(AB')$, (*iv*) $P(A'B')$.

2.13. Έστω τό ἐνδεχόμενο A καί B μέ $P(A) = 0{,}25$, $P(B/A) = 0{,}5$ καί $P(A/B) = 0{,}25$. Νά βρεθοῦν (α) $P(A/B')$ καί (β) $P(A'/B')$. Εἶναι τά A, B ἀνεξάρτητα;

2.14. Ἐάν τά ἐνδεχόμενα A, B, Γ εἶναι ἀνεξάρτητα καί $P(A) = 1/2$, $P(B) = 1/3$ καί $P(\Gamma) = 1/4$, νά βρεθοῦν *i)* $P(A' \cup B)$, *ii)* $P(A \cup B' \cup \Gamma)$.

2.15. Γιά τυχόντα ἐνδεχόμενα A καί B δεῖξτε ὅτι

$$P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C).$$

Δίδεται ὅτι $P(C) > 0$.

2.16. Δίδονται τά ἐνδεχόμενα A, B, Γ καί οἱ πιθανότητες $P(A) = P(B) = P(\Gamma) = 1/3$, $P(AB) = P(A\Gamma) = P(B\Gamma) = 1/4$ καί $P(AB\Gamma) = 1/5$. Νά βρεθοῦν οἱ πιθανότητες

(*i*) Νά συμβεῖ τοὐλάχιστον ἕνα ἀπό τά A, B, Γ.

(*ii*) Νά συμβοῦν τοὐλάχιστον δύο ἀπό τά A, B, Γ.

(*iii*) Νά συμβεῖ ἀκριβῶς ἕνα ἀπό τά A, B, Γ.

(*iv*) Νά συμβοῦν ἀκριβῶς δύο ἀπό τά A, B, Γ.

2.17. Νά ἐξετάσετε ἄν οἱ ἀκόλουθες προτάσεις εἶναι σωστές ή όχι

(*i*) 'Εάν $P(B|A') = P(B|A)$, τότε τά ένδεχόμενα A καί B είναι άνεξάρτητα.

(*ii*) 'Εάν $a = P(A)$ καί $b = P(B)$, τότε $P(A|B) \geq (a+b-1)/b$.

2.18. "Ας ύποθέσουμε ότι A,B είναι δύο άνεξάρτητα ένδεχόμενα. Νά άποδειχθεί ότι τά A καί B' είναι έπίσης άνεξάρτητα.

2.19. Μία διαφημιστική έταιρεία παρατηρεί ότι περίπου *1* στούς *50* ύποψήφιους άγοραστές ένός έμπορεύματος βλέπει μία διαφήμιση τοῦ έμπορεύματος σέ κάποιο περιοδικό καί *1* στούς *5* βλέπει μιά όμοια διαφήμιση στήν τηλεόραση. Ένας στούς τρείς άγοράζει τό προϊόν έάν έχουν δεί τήν διαφήμιση καί *1* στούς *10* έάν δέν τήν έχουν δεί. 'Ο Γιάννης είναι ένας ύποψήφιος άγοραστής. Ποιά είναι ή πιθανότητα νά άγοράσει τό έμπόρευμα;

2.20. Τρείς παίκτες A,B,Γ στή σειρά ρίχνουν ένα ζάρι ό ένας κατόπιν τοῦ άλλου. 'Ο A κερδίζει, όταν έπιτύχει *5* ή *6*. 'Ο B κερδίζει, όταν έπιτύχει άρτιο άριθμό καί ό Γ έπίσης κερδίζει, όταν έπιτύχει περιττό άριθμό. Τό παιγνίδι σταματᾶ εύθύς ώς ένας άπό τούς τρείς παίκτες κερδίσει. Νά βρεθεί ή πιθανότητα ό A τελικῶς νά κερδίσει.

2.21. Ένα πείραμα συνίσταται στήν ταυτόχρονη ρίψη τριῶν ζαριῶν. Νά βρεθοῦν:

i) Ἡ πιθανότητα τό ἄθροισμα τῶν ἀποτελεσμάτων νά εἶναι δώδεκα.

ii) Ἡ πιθανότητα τουλάχιστον ἕνας ἄσσος νά ἐμφανιστεῖ.

2.22. Παίρνουμε στήν τύχη *10* ἄτομα. Ποιά ἡ πιθανότητα νά ὑπάρχουν τουλάχιστον δύο πού νά ἔχουν γεννηθεῖ τόν ἴδιο μῆνα;

2.23. Τό κουτί *A* περιέχει *5* μαῦρα καί *4* ἄσπρα σφαιρίδια. Τό κουτί *B* περιέχει *2* μαῦρα καί *6* ἄσπρα σφαιρίδια. Δύο σφαιρίδια μεταφέρονται ἀπό τό *A* στό *B* καί στήν συνέχεια ἕνα σφαιρίδιο ἐξάγεται ἀπό τό *B*. Ποία ἡ πιθανότητα τό ἐξαχθέν σφαιρίδιο νά εἶναι λευκό;

2.24. Δοχεῖο *A* περιέχει *5* μαῦρες καί *6* ἄσπρες σφαῖρες καί δοχεῖο *B* περιέχει *8* μαῦρες σφαῖρες. Δύο σφαῖρες μεταφέρονται ἀπό τό *B* δοχεῖο στό *A* καί κατόπιν μία σφαίρα ἐκλέγεται ἀπό τό *A*.

i) Ποία ἡ πιθανότητα ἡ σφαίρα αὐτή νά εἶναι ἄσπρη;

ii) Δεδομένου ὅτι ἡ σφαίρα αὐτή εἶναι ἄσπρη ποία ἡ πιθανότητα δύο ἄσπρες σφαῖρες νά ἔχουν μεταφερθεῖ ἀπό τό δοχεῖο *B* στό *A*;

2.25. Ἕνα δοχεῖο περιέχει *10* μπάλες ἀριθμημένες ἀπό τό *1* ἕως τό *10*. Τέσσερες μπάλες ἐκλέγονται χωρίς ἐπανάθεση καί ἔστω x ὁ δεύτερος μικρότερος ἀπό τούς τέσσερες ἀριθμούς πού ἐκλέχτηκαν. Νά βρεθεῖ ἡ πιθανότητα τό x νά εἶναι *3*.

2.26. Ένα άτομο έχει πιθανότητες $1/6$, $1/2$, $1/3$ νά άποκτήσει $0,1,2$ παιδιά άντιστοίχως. Οί ίδιες πιθανότητες ίσχύουν γιά κάθε ένα άπό τά παιδιά του.
i) Ποιά ή πιθανότητα τό άτομο αύτό νά μήν άποκτήσει κανένα έγγόνι;
ii) Δεδομένου ότι τό άτομο αύτό άπέκτησε ένα έγγόνι ποιά ή πιθανότητα νά έχει δύο παιδιά;

2.27. Ένας παίκτης έκλέγει τυχαίως ένα άπό δύο νομίσματα *A* καί *B*. Τό *A* έχει πιθανότητα $3/4$ νά φέρει κορώνα καί τό *B* έχει πιθανότητα κορώνας ίση πρός $1/4$. Ό παίκτης ρίχνει τό έκλεγέν νόμισμα δύο φορές. Ποιά είναι ή πιθανότητα νά φέρει (*i*) δύο κορώνες, (*ii*) μία κορώνα;

2.28. Τόν Φεβρουάριο ή πιθανότητα χιονοπτώσεως μία όποιαδήποτε μέρα είναι $0,2$. Ή πιθανότητα ή θερμοκρασία νά πέσει κάτω άπό -5^oC είναι $0,5$. Άν ή θερμοκρασία πέσει κάτω άπό -5^oC μία όποιαδήποτε μέρα ή πιθανότητα χιονοπτώσεως είναι $0,35$. Μία μέρα τού Φεβρουαρίου έπεσε χιόνι. Ποία ή πιθανότητα ή θερμοκρασία νά έπεσε κάτω άπό -5^oC τή μέρα αύτή;

2.29. Δύο άτομα, άνεξάρτητα τό ένα άπό τό άλλο, διάλεγουν άπό ένα διψήφιο άριθμό. Έστω *A* τό ένδεχόμενο ότι καί τά δύο άτομα διαλέγουν τό 17 καί *B* τό ένδεχόμενο ότι καί τά δύο άτομα διαλέγουν τόν ίδιο άριθμό. Ποίο άπό τά ένδεχόμενα *A* καί *B* έχει μεγαλύτερη πιθανότητα; Δικαιολογείστε τήν άπάντηση σας διαισθητικά καί ύπολογίζοντας τά $P(A)$ καί $P(B)$.

2.30. Ένας χιμπατζής παίζει μέ μία γραφομηχανή-παιγνίδι- πού έχει πλήκτρα μέ τά γράμματα A, Γ, E, I, K, O, Π, T. Εάν ό χιμπατζής χτυπήση 4 πλήκτρα στήν τύχη ποία ή πιθανότητα νά χτυπήση τό ίδιο γράμμα τέσσερες φορές;

2.31. Σέ ένα μεγάλο πληθυσμό τό ποσοστό τών άτόμων πού έχουν μαύρα μαλλιά καί τό ποσοστό τών άτόμων πού έχουν καστανά μαλλιά είναι 1/3 καί 1/3 άντίστοιχα. Παρατηρείται ότι 30% τών άνθρώπων μέ μαύρα μαλλιά καί 50% τών άνθρώπων μέ καστανά μαλλιά έχουν γαλανά μάτια. Έπίσης τό 40% τών άνθρώπων πού δέν έχουν ούτε μαύρα ούτε καστανά μαλλιά έχουν γαλανά μάτια. Ένα άτομο έκλέγεται τυχαία άπό τόν παραπάνω πληθυσμό καί βλέπουμε ότι έχει γαλανά μάτια. Ποία ή πιθανότητα νά έχει μαύρα μαλλιά;

2.32. Ή κυρία X δακτυλογραφεί τό 40% τών γραμμάτων ένός γραφείου καί ή κυρία Ψ τό 60%. Άπό τήν πείρα γνωρίζουμε ότι ή κυρία X δακτυλογραφεί ένα γράμμα χωρίς λάθη 95% τών φορών καί ή κυρία Ψ 90% τών φορών. Ένα γράμμα βρέθηκε νά έχει λάθος. Ποιά είναι ή πιθανότητα τό γράμμα νά γράφτηκε άπό τήν κυρία X.

2.33. Είναι γνωστό άπό τήν πείρα ότι μία μικροβιολογική έξέταση αίματος γιά μιά όρισμένη άρρώστια μάς δίνει λάθος θετική άντίδραση περίπου 10% τών φορών καί σωστή θετική άντίδραση 80% τών φορών. Ή παραπάνω έξέταση γίνεται σέ όλους τούς μαθητές ένός σχολείου γιά τούς όποίους είναι γνωστό ότι 2% πάσχουν άπό αύτή τήν άρρώστεια. Ύπολογίστε τίς πιθανότητες *i)* θετικῆς ἀντίδρασης

ii) ὁ μαθητής πού ἐξετάζεται νά εἶναι ἀσθενής ὅταν ἡ ἀντίδραση εἶναι θετική.

2.34. Ἕνας ἀσθενής εἶναι γνωστό νά ἔχει ἀκριβῶς μία ἀπό τίς δύο ἀρρώστιες B_1 ἤ B_2. Ὑπάρχουν δύο δυνατά σύνολα συμπτωμάτων

Γ={πυρετός,σπυριά} καί Δ={λαιμός, πονοκέφαλος}. Ἔχει βρεθεῖ ὅτι $P(\Gamma|B_1) = 0,6$, $P(\Delta|B_1) = 0,4$, $P(\Gamma|B_2) = 0,7$ καί $P(\Delta|B_2) = 0,3$. Ἐπί πλέον ἡ ἀρρώστια B_1 ἐμφανίζεται δύο φορές πιό συχνά ἀπό τήν ἀρρώστια B_2 καί ποτέ δέν ἐμφανίζονται μαζί. Ἄν ὁ ἀσθενής, πού ἔχει ἐκλεγεῖ στήν τύχη ἀπό τόν πληθυσμό, ἔχει πυρετό καί σπυριά, νά βρεθεῖ ἡ πιθανότητα νά ἔχει τήν ἀρρώστια B_1.

2.35. Ὁ Μιχάλης ἀγοράζει *10* ἀπό *1000* λαχεῖα σέ μία λαχειοφόρο. Ὑπάρχουν *5* βραβεῖα καί κατά τήν κλήρωση *5* λαχεῖα ἐκλέγονται τυχαῖα καί χωρίς ἐπανατοποθέτηση. Νά βρεθεῖ ἡ πιθανότητα ὁ Μιχάλης νά κερδίσει τουλάχιστον ἕνα βραβεῖο.

2.36. Φοιτητές ψήφισαν σέ δυό τμήματα T_1 καί T_2 τούς ὑποψήφιους K_1, K_2. Στή μέτρηση βρέθηκαν: Τμῆμα T_1: *20* ὑπέρ K_1 καί *40* ὑπέρ K_2. Τμῆμα T_2: *30* ὑπέρ K_1 καί *30* ὑπέρ K_2. Διαλέγουμε στήν τύχη ἕνα τμῆμα καί παίρνουμε τυχαῖα *5* ψήφους, χωρίς ἐπανατοποθέτηση, ἀπ'τούς ὁποίους οἱ *3* εἶναι ὑπέρ τοῦ K_1 καί οἱ *2* ὑπέρ τοῦ K_2. Ποιά ἡ πιθανότητα νά ἔχουμε διαλέξει τό τμῆμα T_2;

2.37. Σέ ἕνα συρτάρι ὑπάρχουν ἀνακατεμένα 150 γραπτά, 40 τῆς τάξης A, 50 τῆς τάξης B, 60 τῆς τάξης Γ. Ἀπό αὐτά 15% τῆς τάξης A, 20% τῆς τάξης B καί 10% τῆς τάξης Γ εἶναι βαθμολογημένα κάτω τῆς βάσεως. Παίρνουμε στήν τύχη ἕνα γραπτό. Ποιά ἡ πιθανότητα τό γραπτό αὐτό νά ἔχει βαθμολογία κάτω ἀπό τήν βάση;

2.38. Τό ἐγραστήριο Φυσικῆς ἑνός Πανεπιστημίου προμηθεύεται μετρητές τάσης ἠλεκτρικοῦ ρεύματος ἀπό δύο ἀποθῆκες A_1, A_2 ἠλεκτρικῶν εἰδῶν. Ἡ πιθανότητα νά προμηθευθεῖ ἀπό τήν ἀποθήκη A_1 εἶναι ἴση μέ τήν πιθανότητα νά προμηθευθεῖ ἀπό τήν ἀποθήκη A_2. Εἶναι ἀκόμη γνωστό ὅτι τό 5% τῶν μετρητῶν τάσης τῆς ἀποθήκης A_1 καί τό 20% τῆς ἀποθήκης A_2 εἶναι ἐλαττωματικοί. Ποιά εἶναι ἡ πιθανότητα ἕνας ὁποιοσδήποτε μετρητής τάσης ὁ ὁποῖος θά φθάσει στό ἐν λόγω ἐργαστήριο νά μήν εἶναι ἐλαττωματικός;

2.39. Σέ μιά κληρωτίδα ὑπάρχουν οἱ ἀριθμοί 221, 232, 422, 798. Ἕνας ἀπό αὐτούς ἐξάγεται τυχαία. Ἔστω A, B, Γ τά ἐνδεχόμενα ἐξαγωγῆς τοῦ ἀριθμοῦ τοῦ ὁποίου τό πρῶτο, τό δεύτερο καί τό τρίτο ψηφίο εἶναι 2 ἀντίστοιχα. Νά ἐξετασθεῖ ἄν τά ἐνδεχόμενα A, B, Γ εἶναι ἀνεξάρτητα.

2.40. Οἱ οἰκογένειες μίας κοινότητας χωρίζονται σέ τέσσερες ὁμάδες εἰσοδήματος O_1, O_2, O_3, O_4 οἱ ὁποῖες περιλαμβάνουν 100, 200, 300 καί 400 οἰκογένειες ἀντιστοίχως. Ὁ παρακάτω πίνακας δίνει τά ποσοστά

οἰκογενειῶν ἀνά ὁμάδα πού ἔχουν: κανένα παιδί, ἕ-
να παιδί, δύο παιδιά καί τρία ἤ περισσότερα παι-
διά.

'Ομάδα εἰσοδήματος	'Αριθμός παιδιῶν ἀνά οἰκογένεια			
	0	1	2	≥ 3
0_1	40%	30%	20%	10%
0_2	10%	40%	30%	20%
0_3	20%	10%	40%	30%
0_4	30%	20%	10%	40%

Μία οἰκογένεια ἐκλέγεται τυχαία ἀπό τήν κοινότητα καί εὑρίσκεται νά μήν ἔχει παιδιά. Νά βρεθεῖ ἡ πιθανότητα ἡ οἰκογένεια αὐτή νά προέρχεται ἀπό τή δεύτερη ὁμάδα (0_2) εἰσοδήματος.

2.41. Δίνονται τέσσερα κάλπικα νομίσματα τέτοια ὥστε ἡ πιθανότητα κορώνας γιά τό ν-στό νόμισμα νά εἶναι $\nu/4$, $\nu = 1,2,3,4$.

(α) Νά βρεθεῖ ἡ πιθανότητα κορώνας γιά ἕνα νόμισμα πού ἐκλέγεται στήν τύχη ἀπό τά τέσσερα κάλπικα νομίσματα.

(β) 'Εκλέγεται ἕνα νόμισμα ἀπό τά τέσσερα, ρίχνεται καί δείχνει κορώνα. Ποία ἡ πιθανότητα νά εἶναι τό δεύτερο νόμισμα;

2.42. 'Η πρωϊνή, ἀπογευματινή καί νυχτερινή βάρδια ἐργατῶν σέ ἕνα ἐργοστάσιο ἐρευνῶνται ὀπτικά στήν πύλη ἐξόδου ἀπό τό προσωπικό ἀσφαλείας γιά τόν περιορισμό τῶν ἀπωλειῶν λόγω κλοπῆς. Οἱ βάρδιες αὐτές ἀπασχολοῦν *400*, *400* καί *100* ἐργάτες ἀντίστοιχα μέ ποσοστά κλοπῆς ἀνά ἄτομο ἴσα πρός *0,4%* , *0,8%* καί *1,2%* ἀντίστοιχα. Ἄν κάποια μέρα λείψει ἕνα ἠλεκτρικό τρυπάνι ποιά ἡ πιθανότητα ὅτι ἡ κλοπή ἔγινε κατά τή νυκτερινή βάρδια;

2.43. Ὁ Μιχάλης γράφεται ὡς πρωτοετής σέ κάποιο Πανεπιστήμιο. Ἡ πιθανότητα νά πετύχει ὑποτροφία εἶναι *0,30*. Ἄν πάρει ὑποτροφία ἡ πιθανότητα νά πάρει πτυχίο εἶναι *0,85* καί ἄν δέν πάρει ὑποτροφία ἡ πιθανότητα αὐτή εἶναι μόνο *0,45* . Νά βρεθεῖ ἡ πιθανότητα ὁ Μιχάλης νά πάρει πτυχίο.

2.44. Μία παλαιά συσκευή τηλεοράσεως χαλάει *60%* τῶν φορῶν ἀπό βλάβη τῆς λυχνίας της, *20%* ἀπό διακύμανση τῆς τάσης τοῦ ρεύματος, *10%* ἀπό βλάβη τῆς λυχνίας καί διακύμανση τῆς τάσης. Ἐπίσης χαλάει καί ἀπό ἄλλες αἰτίες. Νά βρεθεῖ (*i*) ἡ πιθανότητα ἡ τηλεόραση νά χαλάσει <u>μόνο</u> ἀπό βλάβη τῆς λυχνίας, (*ii*) ἡ πιθανότητα νά χαλάσει <u>μόνο</u> ἀπό τή διακύμανση τῆς τάσης τοῦ ρεύματος καί (*iii*) ἡ πιθανότητα νά χαλάσει <u>μόνο</u> ἀπό βλάβη τῆς λυχνίας ἤ <u>μόνο</u> ἀπό διακύμανση τῆς τάσης.

2.45. Στή Στατιστική Μηχανική r μή διακεκριμένα σωματίδια θεωροῦνται ὅτι κατανέμονται σέ $n(>r)$ μικρές περιοχές τοῦ χώρου τῶν φάσεων. Κατά τήν ὑπόθεση τῶν *Fermi-Dirac* εἶναι ἀδύνατον δύο ἤ περισσότερα σωματίδια νά καταλάβουν τήν ἴδια περιοχή. Ὑποθέτοντας ὅτι ὅλες οἱ κατανομές τῶν σωματιδίων εἶναι ἰσοπίθανες νά βρεθεῖ ἡ πιθανότητα μίας κατανομῆς (ἡ διάταξη δέν παίζει ρόλο).

2.46. Ἕνα κιβώτιο μέ 15 λυχνίες περιέχει 5 ἐλαττωματικές. Οἱ λυχνίες ἐκλέγονται ἀνά μία ἀπό τό κιβώτιο στήν τύχη καί χωρίς ἐπανάθεση καί ἐξετάζονται. Νά βρεθεῖ ἡ πιθανότητα ἡ δέκατη λυχνία πού ἐκλέγεται νά εἶναι ἡ πέμπτη ἐλαττωματική.

2.47. Ἄν $A_1, A_2, \ldots$ εἶναι ἀνά δύο ἀσυμβίβαστα ἐνδεχόμενα καί $A = \bigcup_{i=1}^{\infty} A_i$ τότε νά δειχθεῖ ὅτι γιά κάθε ἐνδεχόμενο B ἰσχύει ἡ σχέση

$$P(AB) = \sum_{i=1}^{\infty} P(A_i)P(B|A_i) .$$

2.48. Ἔστω A, B, C τρία ἐνδεχόμενα καί ἔστω ἐπί πλέον ὅτι τό ἐνδεχόμενο C συνεπάγεται τό ἐνδεχόμενο AB δηλ. $C \subset AB$. Νά δειχθεῖ ὅτι $P(C) \leq P(A') + P(B')$

ΚΕΦΑΛΑΙΟ 3

τυχαιες μεταβλητες-κατανομες πιθανοτητας

Δύο θεμελιώδεις έννοιες τῆς θεωρίας Πιθανοτήτων καί τῆς Στατιστικῆς εἶναι οἱ έννοιες τῆς τυχαίας μεταβλητῆς καί τῆς κατανομῆς πιθανότητας. Οἱ έννοιες αὐτές εἰσάγονται καί ἀναπτύσσονται στό Κεφάλαιο αὐτό. Οἱ κατανομές πιθανότητας περιγράφονται μέ τίς ἀθροιστικές συναρτήσεις κατανομῆς, τίς συναρτήσεις πιθανότητας καί τίς συναρτήσεις πυκνότητας πιθανότητας. Καί οἱ τρεῖς αὐτές συναρτήσεις παίζουν πρωταρχικό ρόλο στή θεωρία καί μεθοδολογία πού παρουσιάζεται στά ἑπόμενα κεφάλαια.

3.1 Τυχαῖες μεταβλητές

Εἶναι φανερό ὅτι οἱ έννοιες τοῦ δειγματικοῦ χώρου καί τοῦ ἐνδεχομένου καθώς καί οἱ νόμοι τῶν πιθανοτήτων πού παρουσιάστηκαν στό προηγούμενο κεφάλαιο μαζί μέ τό συμβολισμό τους εἶναι περιορισμένης ἐκτάσεως καί δέν ἐπιτρέπουν τήν πλήρη καί ἐπωφελή ἐκμετάλλευση τοῦ μαθηματικοῦ δυναμικοῦ. Τά ἀποτελέσματα ἑνός πειράματος σπάνια περιγράφονται μέ μία ποσότητα ἤ μέ σύμβολα σάν τά A, B, ἤ E_1, E_2, K, Γ, $\{\cdot\}$, $\{\cdot\cdot\}$, κλπ. Συχνά ἀναφέρονται σέ μία ἤ περισσότερες ποσότητες οἱ ὁποῖες στή γλώσσα τῶν φυσικῶν ἐπιστημῶν καλοῦνται μεταβλητές.Γιά παράδειγμα,ἕνα ἐρωτηματολόγιο στό ὁποῖο ἀπαντᾶ ἕνα τυχαῖα ἐπιλεγόμενο ἄτομο περιλαμβάνει δέκα ἐρωτήσεις ἡ κάθε μία μέ 5 δυνατές ἀπαντήσεις, ἡ ἐτήσια παραγωγή ἑνός ἀγροῦ σέ καλαμπόκι ἐξαρτᾶται ἀπό τήν ποικιλία τοῦ καλαμποκιοῦ, τήν ποσότητα τοῦ λιπάσματος, τό κλίμα καί ἄλλους ἐλεγχόμενους ἤ μή παράγοντες. Ἡ πιό συστηματική, αὐστηρή καί ἐκτεταμένη μελέτη τῆς Θεωρίας τῶν Πιθανοτήτων καί τῆς Στατιστικῆς γίνεται μέ τή βοήθεια τῆς ἔννοιας τῆς τυχαίας μεταβλητῆς. Ἄς δώσουμε τώρα δύο παραδείγματα.

Παραδείγματα:

3.1: Ἄς ὑποθέσουμε ὅτι ἕνα ἄτομο δίνει ἐξετάσεις γιά τήν κατάληψη θέσεως σέ μία τράπεζα. Οἱ ἐξετάσεις καί τά ἀποτελέσματα μποροῦν νά θεωρηθοῦν σάν πείραμα τύχης. Τά ἀποτελέσματα τοῦ πειράματος εἶναι ἐπιτυχία,πού συμβολίζεται μέ τό E, καί ἀποτυχία,πού συμβολίζεται μέ τό A. Ἡ ἴδια περιγραφή μπορεῖ νά γίνει καί μέ τή χρήση μίας μεταβλητῆς X ἡ ὁποία παίρνει τίς (κωδικές) τιμές 0 ὅταν οἱ ἐξετάσεις καταλήγουν σέ ἀποτυχία καί 1 ὅταν καταλήγουν σέ ἐπιτυχία. Ἔτσι τά ἐνδεχόμενα A καί E γράφονται ἰσοδύναμα

$$A = \{X=0\} \quad \text{καί} \quad E = \{X=1\}.$$

Ἀπό τούς περιγραφικούς ὅρους ἀποτυχία καί ἐπιτυχία μεταφερόμαστε στούς ἀριθμούς $0,1$ πού ἐκφράζουν ἀκριβῶς τό ἴδιο πρᾶγμα. Ἔτσι ἐπιτυγχάνεται μία ἀπεικόνιση τῶν ἀποτελεσμάτων τοῦ πειράματος στό σύνολο $\{0,1\}$. Ἡ ἀπεικόνιση γίνεται μέ τή βοήθεια τῆς μεταβλητῆς X καί ἀντί νά μιλᾶμε γιά $P(A)$ ἤ $P(E)$ μποροῦμε νά μιλᾶμε ἰσοδύναμα γιά $P(X=0)$ καί $P(X=1)$ ἀντίστοιχα. Τό X εἶναι μία τυχαία μεταβλητή.

3.2: Ἀς θεωρήσουμε τό Παράδειγμα 2.4 τῆς βελόνας τῆς μαγνητικῆς πυξίδας. Τό πείραμα συνίσταται σέ ἕνα στριφογύρισμα τῆς βελόνας. Τά ἀποτελέσματα τοῦ τυχαίου πειράματος εἶναι οἱ (φυσικές, γεωμετρικές) γωνίες θ πού σχηματίζονται μέ τή σταθερή ἀκτίνα OA. Αὐτά μποροῦμε νά τά περιγράψουμε μέ μία μεταβλητή X πού παίρνει τιμές ἀπό 0 ἕως 2π ἀκτίνια. Κάθε τιμή τῆς X ἀντιπροσωπεύει καί μία γωνία πού μπορεῖ νά σχηματιστεῖ. Ἔτσι ἄν A εἶναι τό ἐνδεχόμενο ἡ γωνία πού σχηματίζεται εἶναι ἡ θ μποροῦμε νά γράψουμε

$$A = \{\text{γωνία } \theta\} = \{X=x\} ,$$

ὅπου x εἶναι τό μέγεθος τῆς γωνίας θ. Τό X εἶναι μία τυχαία μεταβλητή.

Τά δύο αὐτά παραδείγματα μᾶς δείχνουν ὅτι τά ἀποτελέσματα ἑνός πειράματος μποροῦν πάντοτε νά ἀντιστοιχηθοῦν μέ πραγματικούς ἀριθμούς καί αὐτό μᾶς ὁδηγεῖ στόν ὁρισμό τῆς ἔννοιας τῆς τυχαίας μεταβλητῆς.

Ὁρισμός 3.1: Τ υ χ α ί α μ ε τ α β λ η τ ή (τ.μ) εἶναι μία μονοσήμαντη συνάρτηση μέ πεδίο ὁρισμοῦ ἕνα δειγματικό χῶρο S καί τιμές ἕνα ὑποσύνολο τῆς εὐθείας τῶν πραγματικῶν ἀριθμῶν.

Οἱ τυχαῖες μεταβλητές λέγονται καί σ τ ο χ α σ τ ι-κ έ ς μ ε τ α β λ η τ έ ς ἤ μ ε τ α β λ η τ έ ς τ ύ-χ η ς καί συμβολίζονται συνήθως μέ κεφαλαῖα γράμματα X, Y, Z κλπ. ἤ X_1, X_2, Y_1 κλπ. Μέ μικρά γράμματα x, y, z, x_1, y_1 κλπ. συμβολίζονται οἱ τιμές τῶν τυχαίων μεταβλητῶν. Τά ἐνδεχόμενα γράφονται $\{X=x\}$, $\{X\leq x\}$ κλπ.

Τά ἑπόμενα παραδείγματα ἐμπεδώνουν τήν ἔννοια τῆς τ.μ. καί προετοιμάζουν τό ἔδαφος γιά τίς κατανομές πιθανότητας.

<u>Παραδείγματα:</u>

3.3: Ἄς θεωρήσουμε τό Παράδειγμα 2.2 μέ τό ζάρι. Ἡ ἀντιστοιχία ὁρίζει μία τυχαία μεταβλητή X μέ πεδίο ὁρισμοῦ
Ἡ ἀντιστοιχία

τό S καί πεδίο τιμῶν τό σύνολο $\{1,2,\ldots,6\}$. Τό ἐνδεχόμενο E_3 συμβολίζεται μέ $\{X=3\}$. Εἶναι φανερό ὅτι ἡ τ.μ. X παίρνει τίς διάφορες τιμές της x μέ κάποια πιθανότητα. Συγκεκριμένα

$$P(X=x) = \frac{1}{6} \quad , \quad x = 1,2,\ldots,6. \qquad (3.1)$$

3.4: Ἄς θεωρήσουμε τό Παράδειγμα 2.5 ὅπου ἕνα νόμισμα ρίχνεται δύο φορές. Ἰσοδύναμα μποροῦμε νά ρίξουμε δύο νομίσματα μία φορά. Μία τυχαία μεταβλητή μπορεῖ νά οριστεῖ μέ τήν ἀκόλουθη ἀπεικόνιση.

$$KK \to 1, \quad K\Gamma \to 2, \quad \Gamma K \to 3, \quad \Gamma\Gamma \to 4 .$$

Ἔτσι ὅταν γράφουμε $X=3$ ἐννοοῦμε ὅτι τό πρῶτο νόμισμα ἔφερε Γ καί τό δεύτερο K, δηλ. $\{X=3\}=\Gamma K$. Θά μπορούσαμε βέβαια νά χρησιμοποιήσουμε καί ἄλλους ἀριθμούς. Ἐπί πλέον ἔχουμε

$$P(X=x) = 1/4, \qquad x = 1,2,3,4. \qquad (3.2)$$

Στό ἴδιο παράδειγμα μία ἄλλη τ.μ. Y ορίζεται μέ τήν ἀντιστοιχία

$$KK \to 2, \quad K\Gamma \to 1, \quad \Gamma K \to 1, \quad \Gamma\Gamma \to 0 .$$

Ἔτσι ἔχουμε

$$\{Y=2\} = KK, \quad \{Y=1\} = \{K\Gamma, \Gamma K\}, \quad \{Y=0\} = \Gamma\Gamma$$

καί

$$P(Y=2) = P(KK) = 1/4$$

$$P(Y=1) = P(K\Gamma) + P(\Gamma K) = 1/4+1/2 = 1/2 \qquad (3.3)$$

$$P(Y=0) = P(\Gamma\Gamma) = 1/4.$$

Ἡ τ.μ. Y μετρᾶ τόν ἀριθμό τῶν πλευρῶν K πού μποροῦν νά ἐμφανιστοῦν στή ρίψη ἑνός νομίσματος δύο φορές.

3.5: Ἄς θεωρήσουμε ἕνα ἄτομο τό ὁποῖο παίρνει μέρος σέ ἐξετάσεις γιά τήν ἀπόκτηση διπλώματος ὁδηγήσεως αὐτοκινήτου. Τά δύο δυνατά ἀποτελέσματα εἶναι ἀποτυχία, A, καί ἐπιτυχία, E. Ἄν ἀποτύχει τό ἄτομο ἐπανέρχεται πάλι στίς ἐξετάσεις κ.ο.κ. Τό πείραμα σταματᾶ ὅταν ὁ ὑποψήφιος πετύχει στίς ἐξετάσεις. Θεωρητικά ὁ ὑποψήφιος μπορεῖ νά προσπαθήσει μία, δύο, τρεῖς,..., ἄπειρες φορές. Ἔτσι τά ἀποτελέσματα (δειγματικός χῶρος) τοῦ πειράματος εἶναι

$$S = \{E, AE, AAE, AAAE, \dots\}.$$

Μία τ.μ. X ὁρίζεται ὡς ἑξῆς:

$$\{X=1\} = E, \quad \{X=2\} = AE, \quad \{X=3\} = AAE \quad \text{κ.ο.κ.}$$

Δηλαδή ἡ τυχαία μεταβλητή X μᾶς δίνει τόν ἀριθμό τῶν προσπαθειῶν τοῦ ὑποψήφιου ὁδηγοῦ μέχρις ὅτου ὁ ὑποψήφιος ἐπιτύχει καί ἀποκτήσει τό δίπλωμα. Τό σύνολο τῶν τιμῶν τῆς X εἶναι ἀριθμήσιμο.

Ἔστω τώρα ὅτι σέ κάθε προσπάθεια ἡ πιθανότητα ἐπιτυχίας εἶναι p καί ἀποτυχίας q, ὅπου $p+q=1$. Ἔστω ἐπί πλέον ὅτι οἱ προσπάθειες εἶναι ἀνεξάρτητες μεταξύ τους. Τά ἐνδεχόμενα A καί E στά ἐπαναλαμβανόμενα πειράματα τῶν ἐξετάσεων-προσπαθειῶν εἶναι ἀνεξάρτητα. Θά εὕρουμε τήν πιθανότητα ὁ ὑποψήφιος νά πετύχει στήν πρώτη, δεύτερη, τρίτη, κλπ. προσπάθεια. Μέ ἄλλα λόγια

$$P(X=1) = P(E) = p$$

$$P(X=2) = P(AE) = P(A)P(E) = qp$$

$$P(X=3) = P(AAE) = P(A)P(A)P(E) = q^2p$$

καί γενικεύοντας

$$P(X=x) = P(\underbrace{AA\dots A}_{x-1}E) = q^{x-1}p, \quad x=1,2,3,\dots \tag{3.4}$$

Τό πεδίο τιμῶν μίας τ.μ. δέν εἶναι κατ'ἀνάγκην ἕνα ὑποσύνολο τῆς εὐθείας τῶν πραγματικῶν ἀριθμῶν. Μπορεῖ νά εἶναι καί ὑποσύνολο τοῦ διδιάστατου χώρου ἤ τοῦ χώρου τῶν k διαστάσεων. Στήν περίπτωση αὐτή μιλᾶμε γιά τυχαῖα διανύσματα, μιά ἔννοια πού θά ὁρίσουμε στό Κεφάλαιο 6.

Ἀπό θεωρητικῆς πλευρᾶς τυχαία μεταβλητή εἶναι μία συνάρτηση $X(s)$ μέ πεδίο ὁρισμοῦ τό S καί πεδίο τιμῶν τούς πραγματικούς ἀριθμούς. Γιά νά ἀποδάσει ἡ ἀντιστοιχία αὐτή θά πρέπει οἱ ἀντίστροφες εἰκόνες συνόλων (ὑποσυνόλων) τῶν πραγματικῶν ἀριθμῶν νά εἶναι ἐνδεχόμενα. Ἔτσι ἄν $A \subset R^1$ τότε θά πρέπει τό $X^{-1}(A) = \{s \in S : X(s) \in A\}$ νά εἶναι ἕνα ἐνδεχόμενο. Μπορεῖ νά ἀποδειχθεῖ ὅτι γιά νά εἶναι ἡ συνάρτηση $X(s)$ τ.μ. ἀρκεῖ τό σύνολο $X^{-1}(-\infty, x]$ νά εἶναι ἕνα ἐνδεχόμενο στό S γιά κάθε $x \in R^1$. Ἔτσι οἱ πιθανότητες μεταφέρονται ἀπό τά ἐνδεχόμενα στά σύνολα πραγματικῶν ἀριθμῶν:

$$P(A) = P[X^{-1}(A)], \quad A \subset R^1 \tag{3.5}$$

Ὑπάρχουν συναρτήσεις $X(s)$ πού δέν εἶναι τ.μ. Αὐτές ἔχουν μόνο θεωρητική σημασία καί δέν θά μᾶς ἀπασχολήσουν στό βιβλίο αὐτό.

Μιλήσαμε προηγουμένως γιά ἀριθμήσιμους καί γιά συνεχεῖς δειγματικούς χώρους. Μία ἀνάλογη ταξινόμηση ὑπάρχει καί μεταξύ τῶν τ.μ. καί τῶν κατανομῶν αὐτῶν.

3. 2 Διακριτὲς τυχαῖες μεταβλητὲς καὶ κατανομὲς πιθανότητας

'Ορισμός 3.2: Διακριτή τυχαία μεταβλητή. Ἔστω X μία τυχαία μεταβλητή. Ἐάν τό σύνολο τῶν τιμῶν τῆς X εἶναι τό πολύ ἀριθμήσιμο (δηλ. πεπερασμένο ἤ ἀπείρως ἀριθμήσιμο) τότε ἡ τ.μ. λέγεται διακριτή ἤ ἀπαριθμητή ἤ καί ἀσυνεχής.

Οἱ τυχαῖες μεταβλητές τῶν Παραδειγμάτων 3.1, 3.3, 3.4, 3.5 εἶναι διακριτές. Ἄλλα παραδείγματα διακριτῶν τ.μ. εἶναι τά ἀκόλουθα: ὁ ἀριθμός τῶν τηλεφωνικῶν κλήσεων πού φθάνουν σέ ἕνα τηλεφωνικό κέντρο μία ὁρισμένη ὥρα, ὁ ἀριθμός τῶν ἀτόμων πού περιμένουν σέ μία στάση λεωφορείου ἤ σέ μία οὐρά, ὁ ἀριθμός τῶν ἐλαττωμάτων κάποιας ὕλης ἤ προϊόντος π.χ. ἐλαστικά αὐτοκινήτων, ἠλεκτρικοί λαμπτῆρες, τόπια ὑφασμάτων κλπ., ὁ ἀριθμός τῶν βλημάτων πού πετυχαίνουν ἕνα στόχο σέ μία σειρά βολῶν, ὁ ἀριθμός τῶν ἀτυχημάτων πού συμβαίνουν σέ μία ὡρισμένη περιοχή γιά ὥρισμένο χρονικό διάστημα. Γενικά οἱ διακριτές τ.μ. παριστάνουν "ἀριθμούς μετρήσεων" καί, στίς περιπτώσεις αὐτές, οἱ τιμές τους εἶναι οἱ φυσικοί ἀριθμοί $0,1,2,\ldots$

Ἀπό τά προηγούμενα παραδείγματα εἶναι φανερό ὅτι σέ κάθε τιμή μίας διακριτῆς τ.μ. ἀντιστοιχεῖ καί μία πιθανότητα. Ἡ ἀντιστοίχηση αὐτή ὁρίζει μία συνάρτηση ἡ ὁποία ὁρίζεται ὡς ἀκολούθως:

'Ορισμός 3.3: Συνάρτηση Πιθανότητας. Ἔστω X μία διακριτή τ.μ. Ἡ συνάρτηση

$$p_X(x) = P(X=x)$$

μέ πεδίο ὁρισμοῦ τίς τιμές τῆς X καί πεδίο τιμῶν τίς πιθανότητες τῶν τιμῶν αὐτῶν λέγεται σ υ ν ά ρ τ η σ η π ι θ α ν ό τ η τ α ς (σ.π.) ἤ καί κ α τ α ν ο μ ή π ι θ α ν ό τ η τ α ς τῆς τ.μ. X.

Στά Παράδειγματα 3.3 καί 3.5 οἱ συναρτήσεις

$$p_X(x) = P(X=x) = 1/6 \quad , \qquad x = 1,2,3,4,5,6$$

καί

$$p_X(x) = P(X=x) = q^{x-1}p, \qquad x = 1,2,3, \ldots$$

εἶναι συναρτήσεις πιθανότητας. Στό Παράδειγμα 3.4 ἡ σ.π. τῆς τ.μ. Y εἶναι

$$p_Y(y) = \begin{cases} 1/4 \ , & y = 0,2 \\ 1/2 \ , & y = 1 \ . \end{cases}$$

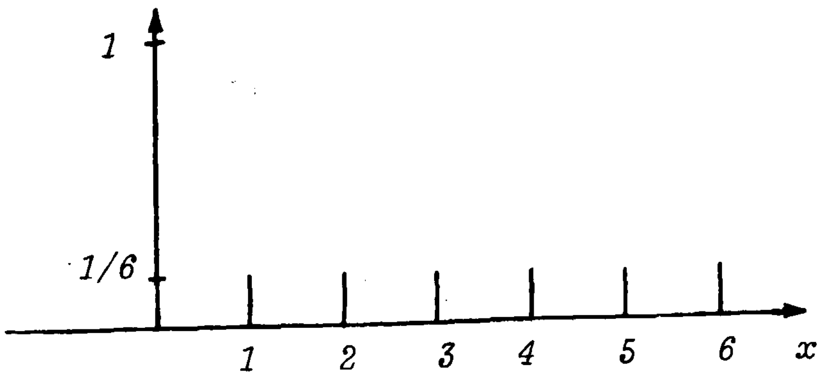
'Από τόν ὁρισμό καί τά ἀξιώματα τῶν πιθανοτήτων εἶναι φανερό ὅτι κάθε σ.π. ἔχει τίς ἑξῆς ἰδιότητες:

1. $p_X(x) \geq 0$ γιά κάθε x.	
2. $\sum_x p_X(x) = 1$	(3.6)
3. $P(X \epsilon B) = \sum_{x \epsilon B} p_X(x)$, B ὑποσύνολο πεδίου τιμῶν τοῦ X.	

Οἱ ἰδιότητες 1 καί 2 δίνουν τίς ἀπαραίτητες συνθῆκες γιά τόν καθορισμό μίας σ.π. Μία τυχοῦσα συνάρτηση $p(x)$ εἶναι σ.π. μίας τ.μ. X ἄν ἱκανοποιεῖ τίς ἰδιότητες 1 καί 2.

'Η συνάρτηση πιθανότητας πολλές φορές ὀνομάζεται καί σ υ ν ά ρ τ η σ η μ ά ζ α ς γιατί μοιράζει κατά κάποιο τρόπο μία μάζα ἴση μέ 1, τήν πιθανότητα τοῦ δειγματικοῦ χώρου, στίς διάφορες τιμές τῆς τ.μ., τά διάφορα δειγματικά σημεῖα.

Συχνά ἀντί τῆς μαθηματικῆς μορφῆς τῆς $p_X(x)$ δίνεται ἡ γραφική της παράσταση. Ἔτσι τό γράφημα στό Σχῆμα 3.1



Σχῆμα 3.1 Συνάρτηση πιθανότητας

δίνει τήν κατανομή πιθανότητας τῆς τ.μ. X τοῦ Παραδείγτος 3.3.

Ἄν D εἶναι τό σύνολο τῶν τιμῶν x μίας τ.μ. X τότε ἡ σ.π. μπορεῖ νά ὁριστεῖ καί γ ι ά κ ά θ ε $x \in R$ ὡς ἑξῆς

$$p_X(x) = \begin{cases} P(X=x), & \text{ἐάν } x \in D \\ 0 & , \text{ἐάν } x \notin D. \end{cases} \qquad (3.7)$$

Μέ αὐτό τόν τρόπο ἐπεκτείνεται ἡ ἔννοια τῆς τ.μ. X σ'ὅλο τό R δίνοντας πιθανότητα μηδέν στίς τιμές πού δέν ἀνήκουν στό D. Ἡ ἐπέκταση αὐτή, ἄν καί δέν ἀλλάζει τίποτε στήν οὐσία, ἐπιτρέπει τήν ὁμοιόμορφη παρουσίαση τῆς θεωρίας.

Ἡ κατανομή πιθανότητας μίας διακριτῆς τ.μ. X μπορεῖ νά ὁριστεῖ ἰσοδύναμα καί μέ τή βοήθεια τῆς ἀθροιστικῆς συνάρτησης κατανομῆς ἤ τῶν γραφημάτων αὐτῆς.

'Ορισμός 3.4: 'Α θ ρ ο ι σ τ ι κ ή σ ύ ν ά ρ τ η σ η κ α τ α ν ο μ ῆ ς. Ἔστω μία τ.μ. Ἡ συνάρτηση

$$F_X(x) = P(X \leq x), \quad x \in R \qquad (3.8)$$

λέγεται ἀ θ ρ ο ι σ τ ι κ ή σ υ ν ά ρ τ η σ η κ α τ α ν ο μ ῆ ς (α.σ.κ.) τῆς X ἤ ἁπλά σ υ ν ά ρ τ η σ η κ α τ α ν ο μ ῆ ς (σ.κ.) τῆς X.

Ἡ α.σ.κ. συμβολίζεται μέ τό $F_X(x)$ ἤ ἁπλά τό $F(x)$. Ἡ $F(x)$ δίνει τήν πιθανότητα ἡ τ.μ. X νά πάρει ὁποιαδήποτε τιμή μικρότερη ἤ ἴση μέ τήν x. Κατά συνέπεια

$$F_X(x) = \sum_{y \leq x} P(X \leq y) = \sum_{y \leq x} p_X(y), \text{ γιά κάθε } x \in R. \quad (3.9)$$

Οἱ συναρτήσεις $F_X(x)$ καί $p_X(x)$ εἶναι ἰσοδύναμες μέ τήν ἔννοια ὅτι ἐάν ξέρουμε τήν μία μποροῦμε τότε νά προσδιορίσουμε τήν ἄλλη. Πράγματι ἰσχύει

$$P(X=x) = p_X(x) = F_X(x) - F_X(x-1) . \qquad (3.10)$$

Μία τ.μ. εἶναι καθορισμένη ὅταν ξέρουμε τίς τιμές πού παίρνει καί τίς πιθανότητες τῶν τιμῶν αὐτῶν. Δηλαδή ὅταν γνωρίζομε τήν σ.π. ἤ ἰσοδύναμα τήν α.σ.κ.

<u>Παραδείγματα:</u>

3.6: Ἔστω ἡ τ.μ. Y τοῦ Παραδείγματος 3.4. Ἡ σ.π. τῆς Y εἶναι

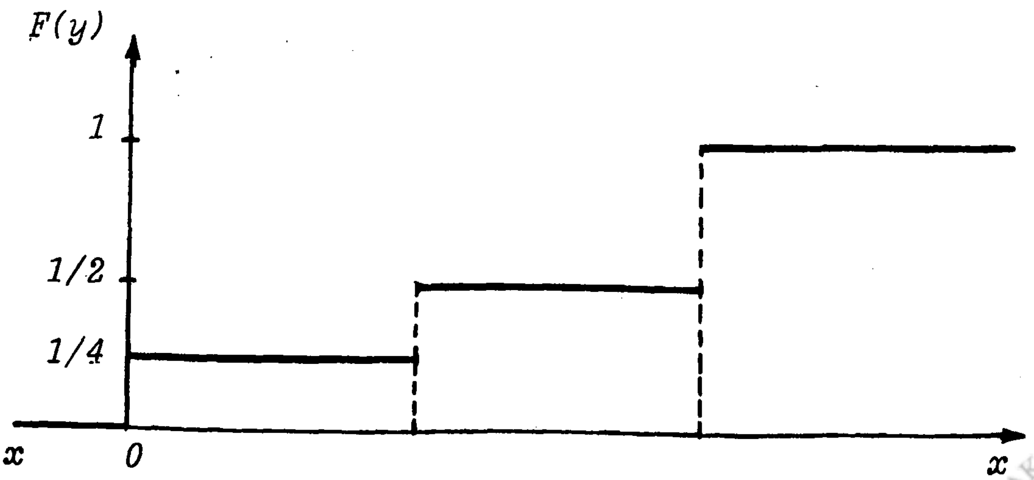
$$p_Y(y)=\begin{cases} 1/4\ , & y=0,\ 2 \\ 1/2\ , & y=1 \\ 0 & \text{ἀλλοῦ}\ . \end{cases}$$

Ἡ α.σ.κ. τῆς Y εὑρίσκεται ὡς ἑξῆς: ἐάν $y<0$ τότε $F_Y(y) = P(Y\leq y) = 0$ διότι $P(Y=y) = p_Y(y) = 0$ ὅταν $y<0$. Ἐάν $0\leq y<1$ τότε $F_Y(y) = P(Y\leq y) = P(Y=0) = p_Y(0) = 1/4$. Ἐάν $1\leq y<2$ τότε $F_Y(y) = P(Y\leq y) = P(Y=0)+P(Y=1) = p_Y(0)+p_Y(1) = 1/4+1/2 = 3/4$. Ἐάν $2\leq y<\infty$ τότε $F_Y(y) = P(Y\leq y) = P(Y=0)+P(Y=1)+P(Y=2) = 1/4+1/2+1/4 = 1$.

Ἔτσι

$$F(y) = \begin{cases} 0 & ,\ y<0 \\ 1/4 & ,\ 0\leq y<1 \\ 3/4 & ,\ 1\leq y<2 \\ 1 & ,\ 2\leq y<\infty\ . \end{cases}$$

Ἡ γραφική παράσταση τῆς $F_Y(y)$ δίνεται στό Σχῆμα 3.2.



Σχῆμα 3.2 Ἀθροιστική συνάρτηση διακριτῆς τυχαίας μεταβλητῆς

Βλέπουμε ὅτι ἡ $F(y)$ εἶναι μία κλιμακωτή, μή φθίνουσα, συνεχής ἀπό δεξιά συνάρτηση.

Ἄν $B = \{0,2\}$ τότε

$$P(Y \varepsilon B) = \sum_{y \varepsilon B} p_Y(y) = p_Y(0)+p_Y(2) = 1/2 .$$

3.7: Ἄς θεωρήσουμε τήν τ.μ. τοῦ Παραδείγματος 3.5. Ἡ σ.π. εἶναι

$$p_X(x) = P(X=x) = \begin{cases} q^{x-1}p, & x = 1,2,3,\ldots \\ 0 & \text{, γιά ὅλες τίς ἄλλες τιμές τοῦ } x . \end{cases}$$

Ἔτσι ἄν $x > 1$ τότε $F_X(x) = 0$.

$$F_X(x) = \sum_{y \leq x} p_X(y) = p+qp+q^2p+\ldots+q^{x-1}p = p\left(\frac{q^x-1}{q-1}\right) .$$

Ἄρα

$$F_X(x) = \begin{cases} 0 & , \quad x<1 \\ p\,\dfrac{q^x-1}{q-1} & , \quad x \geq 1 . \end{cases}$$

Οἱ ἀκόλουθες πιθανότητες ὑπολογίζονται εὔκολα μέ τή βοήθεια τῆς α.σ.κ.

$$\boxed{\begin{aligned} P(a<X\leq b) &= F_X(b)-F_X(a) \\ P(a\leq X\leq b) &= F_X(b)-F_X(a-1) \\ P(a<X<b) &= F_X(b-1)-F_X(a) \\ P(a\leq X<b) &= F_X(b-1)-F_X(a-1). \end{aligned}} \qquad (3.11)$$

3. 3 Συνεχεῖς τυχαῖες μεταβλητὲς καὶ κατανομὲς πιθανότητας

Οἱ συνεχεῖς τυχαῖες μεταβλητές ἀντιστοιχοῦν σέ συνεχεῖς δειγματικούς χώρους. Τά πεδία τιμῶν τους εἶναι ὑπεραριθμήσιμα, συνήθως διαστήματα πεπερασμένα ἤ ἄπειρα τῆς εὐθείας τῶν πραγματικῶν ἀριθμῶν ἤ ὁλόκληρη ἡ εὐθεία. Συχνά τά πεδία τιμῶν τῶν συνεχῶν τ.μ.ταυτίζονται μέ τόν δειγματικό χῶρο. Αὐτό γίνεται ὅταν τά ἀποτελέσματα τοῦ τυχαίου πειράματος παρίστανται μέ σημεῖα τῆς εὐθείας γραμμῆς καί οἱ τιμές τῆς τ.μ. εἶναι ἀριθμοί κατάλληλα ὁρισμένοι στά σημεῖα αὐτά μέ τή βοήθεια κάποιου κανόνα. Εἶναι φανερό ὅτι ἄν ὁ δειγματικός χῶρος S εἶναι ἀριθμήσιμος δέν μπορεῖ νά προκύψουν συνεχεῖς τ.μ.

Ὁ ὁρισμός τῆς κατανομῆς πιθανότητας μίας συνεχοῦς τ.μ. παρουσιάζει κάποια δυσκολία δεδομένου ὅτι θά πρέπει νά ὁρίσουμε τήν πιθανότητα γιά κάθε ἐνδεχόμενο. Ὑπενθυμίζουμε ὅτι ἕνα ὑποσύνολο ἑνός συνεχοῦς δειγματικοῦ χώρου δέν εἶναι κατ'ἀνάγκην ἐνδεχόμενο διότι μπορεῖ νά μήν ἐκφράζεται σάν ἕνωση ἤ τομή ἀριθμήσιμου πλήθους ἐνδεχόμενων. Τό ἴδιο ἰσχύει καί γιά τά ὑποσύνολα τοῦ R^1 σέ σχέση μέ τά διαστήματα. Τό σύνολο A τῶν ἐνδεχόμενων εἶναι "κατά τι" μικρότερο τοῦ συνόλου ὅλων τῶν ὑποσυνόλων τοῦ S. Γιά παράδειγμα συνόλου πού δέν ἐκφράζεται ὡς ἕνωση ἤ τομή ἀριθμήσιμου πλήθους διαστημάτων ὁ ἀναγνώστης παραπέμπεται στή θεωρία Μέτρου.

Ἄς ἐπανέλθουμε τώρα στό Παράδειγμα 2.3 μέ τή βελόνη τῆς μαγνητικῆς πυξίδας. Ἀδιαφορώντας γιά τήν μαθηματική αὐστηρότητα τοῦ συλλογισμοῦ σκεφτόμαστε ὅτι ἀφοῦ στήν περιφέρεια τῆς πυξίδας ὑπάρχουν ἄπειρες (ὑπεραριθμήσιμες) ἐνδείξεις ἡ πιθανότητα ἡ ἐκτέλεση τοῦ πειράματος νά ὁδηγήσει σέ μία συγκεκριμένη τιμή x τῆς X εἶναι $P(X=x) = 1/\infty = 0$. Βλέπουμε δηλαδή ὅτι ἐνῶ ἡ τ.μ. X μπορεῖ νά πάρει ὁποιαδήποτε τιμή x στό διάστημα $[0,2\pi]$, $P(X=x) = 0$. Βέβαια δέν ἔχουμε ἀντινομία διότι στήν πραγματικότητα δέν παρατηροῦμε μία γωνία x ἀλλά ἕνα διάστημα $x+\Delta x$ λόγω σφαλμάτων στή μέτρηση. Οἱ σκέψεις αὐτές μᾶς ὁδηγοῦν στόν ἑξῆς ὁρισμό.

Ὁρισμός 3.5: Σ υ ν ε χ ή ς τ υ χ α ί α μ ε τ α β λ η τ ή. Ἔστω X μία τ.μ. μέ πεδίο ὁρισμοῦ τόν δειγματικό χῶρο S καί πεδίο τιμῶν ἕνα πεπερασμένο ἤ ἄπειρο διάστημα τῆς εὐθείας τῶν πραγματικῶν ἀριθμῶν (ἤ ὁλόκληρη τήν εὐθεία). Ἐάν

$$P(X=x) = 0 \text{ γιά κάθε τιμή } x \text{ τῆς } X$$

τότε ἡ τ.μ. X λέγεται σ υ ν ε χ ή ς.

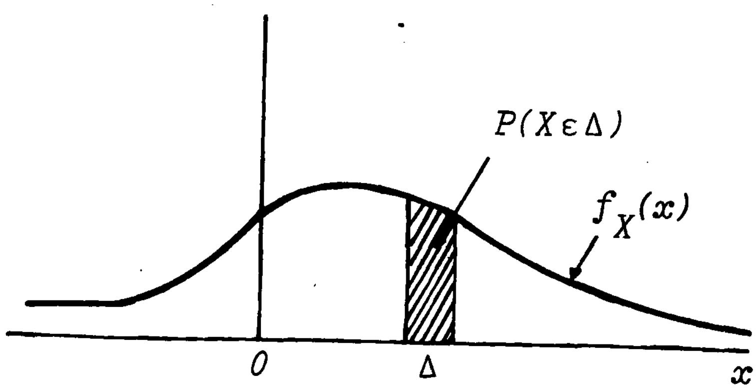
Οἱ συνεχεῖς τ.μ. συνήθως μετροῦν ποσότητες ὅπως βάρος, μῆκος, χρόνος,ὄγκος,ἀπόδοση κλπ.

Ὑπάρχουν δύο τρόποι μέ τούς ὁποίους μποροῦμε νά περιγράψουμε τίς πιθανότητες καί τίς κατανομές συνεχῶν τ.μ.

Α. Ἐφ'ὅσον $P(X=x) = 0$ τότε ἀρκεῖ νά βροῦμε ἕνα τρόπο νά ὁρίσουμε τίς πιθανότητες γιά τά διαστήματα. Αὐτό γίνεται μέ τή βοήθεια μίας μή ἀρνητικῆς συναρτήσεως $f_X(x)$ τέτοιας ὥστε γιά κάθε διάστημα Δ

$$P(X\varepsilon\Delta) = P\{x: x\varepsilon\Delta\} = \int_{\Delta} f_X(x)dx \,. \qquad (3.12)$$

Μέ ἄλλα λόγια $P(X\varepsilon\Delta)$ εἶναι τό ἐμβαδόν τοῦ χώρου πού ὁρίζεται ἀπό τό διάστημα Δ στόν ἄξονα τῶν x καί τήν καμπύλη $f_X(x)$ ὅπως φαίνεται στό Σχῆμα 3.3.



Σχῆμα 3.3 Πιθανότητα διαστήματος συνεχοῦς τυχαίας μεταβλητῆς

Στό παράδειγμα τῆς πυξίδας ἡ συνάρτηση $f_X(x)$ εἶναι

$$f_X(x) = \frac{1}{2\pi}\,, \quad 0 \leq x \leq 2\pi$$

καί έτσι άν $[a,b]$ εἶναι κάποιο ὑποδιάστατο τοῦ $[0,2\pi]$

$$P(a<X\leq b) = \frac{b-a}{2\pi} .$$

Ἔτσι ὁρίζουμε τίς πιθανότητες στά διαστήματα καί αὐτό γίνεται σύμφωνα μέ τά Ἀξιώματα 1 καί 2 τῆς πιθανότητας. Χρησιμοποιώντας καί τό Ἀξίωμα 3 μποροῦμε νά ὁρίσουμε τίς πιθανότητες σέ ὑποσύνολα πού εἶναι ἀριθμήσιμες ἑνώσεις ἤ τομές διαστημάτων, δηλαδή ὅλα τά ἐνδεχόμενα.

Ὁρισμός 3.6: Συνάρτηση πυκνότητας πιθανότητας. Ἔστω μία συνεχής τ.μ. X Ἐάν ὑπάρχει μία μή ἀρνητική συνάρτηση $f_X(x)$ μέ πεδίο ὁρισμοῦ τό πεδίο τιμῶν τῆς X τέτοια ὥστε

$$P(X\varepsilon\Delta) = \int_{\Delta} f_X(x)dx$$

γιά κάθε διάστημα Δ τοῦ πεδίου τιμῶν τῆς X, τότε ἡ $f_X(x)$ λέγεται συνάρτηση πυκνότητας πιθανότητας (σ.π.π) ἤ ἁπλῶς κατανομή πιθανότητας.

Ἄν δέν ὑπάρχει κίνδυνος συγχύσεως θά γράφουμε $f(x)$ ἀντί $f_X(x)$. Τό πεδίο ὁρισμοῦ τῆς $f(x)$ μπορεῖ νά ἐπεκταθεῖ ἄν ὁρίσουμε $f(x) = 0$ γιά κάθε τιμή x τῆς X πού βρίσκεται ἐκτός τοῦ πεδίου ὁρισμοῦ της.

Ἰδιότητες τῆς $f_X(x)$:

$$\boxed{\begin{array}{ll} 1. & f_X(x) \geq 0 \quad \text{γιά κάθε} \quad x\varepsilon R \\ 2. & \int_{-\infty}^{+\infty} f(x)dx = 1 \end{array}} \qquad (3.13)$$

Οἱ ἰδιότητες αὐτές προκύπτουν εὔκολα ἀπό τόν ὁρισμό.

Οἱ ἰδιότητες *1* καί *2* δίνουν τίς ἀπαραίτητες συνθῆκες γιά τόν καθορισμό μίας σ.π.π. Μία τυχοῦσα συνάρτηση $f(x)$ εἶναι σ.π.π. μίας τ.μ. X, ἄν ἱκανοποιεῖ τίς ἰδιότητες *1* καί *2*.

Ἄν $\Delta = [x, x+dx]$ τότε ἰσχύει ἡ προσεγγιστική σχέση

$$P(x<X\leq x+dx) \approx f(x)dx\,, \tag{3.14}$$

ἡ δέ ποσότητα $f(x)dx$ λέγεται στοιχεῖο ἤ διαφορικό πιθανότητας.

Ἡ σ.π.π. δίνεται συνήθως μέ κάποιο μαθηματικό τύπο ἀλλά ὑπάρχουν καί περιπτώσεις ὅπου δίνεται μόνο ἡ γραφική της παράσταση. Εἶναι φανερό ὅτι οἱ δύο αὐτοί τρόποι εἶναι ἰσοδύναμοι.

Β. Ἀντί νά ὁρίσουμε τήν πιθανότητα γιά κάθε διάστημα τῆς εὐθείας γραμμῆς μποροῦμε νά διαλέξουμε τά ἰδιαίτερα διαστήματα τοῦ τύπου $(-\infty, x]$ καί νά ὁρίσουμε τήν ἀθροιστική συνάρτηση κατανομῆς $F_X(x)$ τῆς X μέ τόν ἴδιο ἀκριβῶς τρόπο ὅπως στή διακριτή περίπτωση δηλ.

$$F_X(x) = P(X\leq x) \quad \text{γιά κάθε} \quad x \in R.$$

Ἐάν ὑπάρχει ἡ σ.π.π. $f_X(x)$ τότε ἔχουμε

$$F_X(x) = \int_{-\infty}^{x} f(y)dy \quad \text{γιά κάθε} \ x \in R \tag{3.15}$$

καί ἀπό τό θεμελιῶδες θεώρημα τοῦ ἀπειροστικοῦ λογισμοῦ ἔχουμε

$$f(x) = F'(x) \quad \text{γιά κάθε} \quad x\,. \tag{3.16}$$

Ἄ ρ α ἡ $F_X(x)$ ε ἶ ν α ι σ υ ν ε χ ή ς γιά κάθε $x\epsilon R$. Χρησιμοποιώντας τήν ἔννοια τοῦ στοιχείου ἤ διαφορικοῦ πιθανότητας ἔχουμε

$$f(x) = \lim_{dx\to 0} \frac{P\{x<X\leq x+dx\}}{dx} = \lim_{dx\to 0} \frac{F(x+dx)-F(x)}{dx} = F'(x).$$

Ἡ σχέση αὐτή εἶναι βασική διότι μᾶς ἐπιτρέπει κάτω ἀπό ὁρισμένες προϋποθέσεις νά εὑρίσκουμε τήν μία συνάρτηση ἀπό τήν ἄλλη. Π.χ. ἄν $F_X(x) = 1-e^{-3x}$ τότε $f_X(x)=3e^{-3x}$. Ἄν γνωρίζουμε τήν συνάρτηση $F_X(x)$ μίας συνεχοῦς τ.μ. X τότε μποροῦμε νά ὑπολογίσουμε πιθανότητες διαστημάτων ὡς ἑξῆς

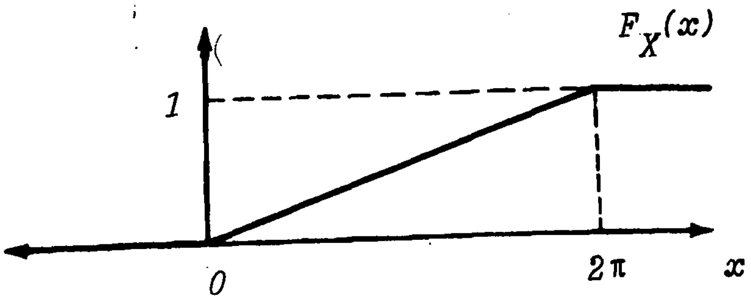
$$\left.\begin{array}{l} P(a<X\leq b) \\ P(a<X<b) \\ P(a\leq X<b) \\ P(a\leq X\leq b) \end{array}\right\} = F(b)-F(a) \,. \qquad (3.17)$$

Στό παράδειγμα τῆς πυξίδας χρησιμοποιώντας βασικές ἔννοιες μποροῦμε εὔκολα νά δοῦμε ὅτι:

$$F_X(x) = P(X\leq x) = \begin{cases} 0 & \text{ἄν } x\leq 0 \\ \dfrac{x}{2\pi} & \text{ἄν } 0<x\leq 2\pi \\ 1 & \text{ἄν } x>2\pi. \end{cases} \qquad (3.18)$$

Τό ἴδιο ἀποτέλεσμα βρίσκουμε ἄν χρησιμοποιήσουμε τήν σχέση (3.15).

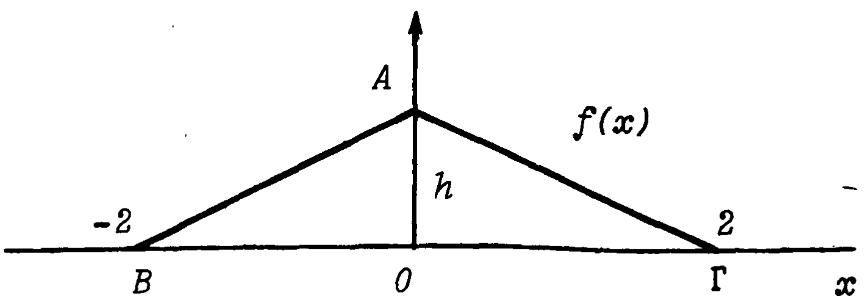
Ἡ γραφική παράσταση τῆς (3.18) εἶναι:



Σχῆμα 3.4 'Αθροιστική συνάρτηση ὁμοιόμορφης κατανομῆς

Παράδειγμα 3.8:

Ἔστω $f(x)$ μία συνάρτηση μέ γράφημα ὅπως στό Σχῆμα 3.5.



Σχῆμα 3.5 Συνάρτηση πυκνότητας πιθανότητας

Νά βρεθοῦν (α) τό h ὥστε ἡ $f(x)$ νά ἀποτελεῖ σ.π.π. μίας τ.μ. X μέ τιμές στό διάστημα $[-2,2]$, (β) ἡ α.σ.κ. τῆς τ.μ. X καί (γ) ἡ $P(|X|<1)$.

α) Θά πρέπει νά ἔχουμε

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

Μέ ἄλλα λόγια τό ἐμβαδό τοῦ τριγώνου $AB\Gamma$ θά πρέπει νά ἰσοῦται μέ τή μονάδα. Ἔτσι ἐμβαδό ΑΒΓ = $4h/2 = 1$ ἤ $h = 1/2$.

β) Γιά νά βροῦμε τήν α.σ.κ. χρειαζόμαστε τήν σ.π.π. Ἡ γραμμή BA ἔχει ἐξίσωση

$$y = \frac{1}{4}(x+2)$$

καί ἡ γραμμή ΑΓ ἔχει ἐξίσωση

$$y = -\frac{1}{4}(x-2)\ .$$

Ἔτσι ἔχουμε

$$f(x) = \begin{cases} 0 & \text{ἄν} \quad x \leq -2 \\ \frac{1}{4}(x+2) & \text{ἄν} \quad -2<x\leq 0 \\ -\frac{1}{4}(x-2) & \text{ἄν} \quad 0<x\leq 2 \\ 0 & \text{ἄν} \quad x>2. \end{cases}$$

Ἐπειδή $F(x) = \int_{-\infty}^{x} f(y)dy$ θά ἔχουμε

$$F(x) = \begin{cases} \int_{-\infty}^{x} 0\cdot dy = 0 & ,\ x\leq -2 \\ \int_{-2}^{x} \frac{1}{4}(y+2)dy = \frac{1}{8}(x+2)^2 & ,\ -2<x\leq 0 \\ \int_{-2}^{0} \frac{1}{4}(x+2)dx + \int_{0}^{x} -\frac{1}{4}(y-2)dy = 1-\frac{1}{8}(x-2)^2, & 0<x\leq 2 \\ 1 & ,\ 2<x<\infty\ . \end{cases}$$

Γενικές Παρατηρήσεις

1. Οἱ συνεχεῖς τ.μ. ὅπως ορίστηκαν πιό πάνω λέγονται ἀκριβέστερα ἀ π ο λ ύ τ ω ς σ υ ν ε χ ε ῖ ς.

2. Ἡ ἔννοια τῆς σ.π.π. εἶναι ἀνάλογη μέ τήν ἔννοια τῆς πυκνότητας μάζας στή Φυσική.

3. Τό γεγονός ὅτι $P(X=x) = 0$ γιά μία συνεχή τ.μ. δέν σημαίνει ὅτι στήν πράξη τό ἐνδεχόμενο $\{X=x\}$ δέν μπορεῖ νά συμβεῖ. Συμβαίνει μέ τήν ἔννοια ὅτι παρατηροῦμε μέν τό $\{X=x\}$ στήν πραγματικότητα ὅμως ἔχουμε παρατηρήσει τό ἐνδεχόμενο $\{x<X<x+dx\}$, ὅπου τό dx ὀφείλεται σέ διάφορα σφάλματα.

4. Πολλές φορές ἀντί γιά τούς ὅρους σ.π.π. ἤ σ.π. χρησιμοποιοῦμε τόν ὅρο κ α τ α ν ο μ ή π ι θ α ν ό τ η τ α ς ἤ ἁπλῶς κ α τ α ν ο μ ή.

5. Γιά κάθε τ.μ. ὑπάρχει πάντα ἡ α.σ.κ.

6. Μία τ.μ. εἶναι γνωστή μόνο ὅταν γνωρίζουμε τήν κατανομή της.

7. Ἡ α.σ.κ. μιάς τ.μ. δέν εἶναι κατ'ἀνάγκην πάντα συνεχής. Στήν διακριτή περίπτωση εἶναι ἀσυνεχής μέ ἀριθμήσιμο ἀριθμό ἀσυνεχειῶν (πηδημάτων). Στή σ υ ν ε χ ή π ε ρ ί π τ ω σ η ἡ α.σ.κ. ε ἶ ν α ι π ά ν τ α σ υ ν ε χ ή ς.

8. Ὑπάρχουν κατανομές τῶν ὁποίων ἡ α.σ.κ. δέν εἶναι οὔτε κλιμακωτή μέ ἀριθμήσιμο πλῆθος ἀσυνεχειῶν οὔτε συνεχής. Ἔτσι οἱ ἀντίστοιχες τ.μ. δέν εἶναι οὔτε διακριτές οὔτε συνεχεῖς. Τό Παράδειγμα 3.9 στό ἑπό-

μενο ἐδάφιο μᾶς δίνει μία τέτοια τ.μ.

9. Ἡ α.σ.κ. εἶναι πάντα παραγωγίσιμη παντοῦ ἐκτός ἀπό ἀριθμήσιμο πλῆθος σημείων. Στή συνεχή περίπτωση ὑπάρχουν α.σ.κ. τῶν ὁποίων ἡ παράγωγος εἶναι ἴση μέ τό μηδέν σχεδόν παντοῦ. Τέτοιες κατανομές λέγονται ἰδιάζουσες.

3.4 Ἰδιότητες ἀθροιστικῶν συναρτήσεων κατανομῆς

Θεώρημα 3.1: Ἰδιότητες α.σ.κ.

1.	$0 \leq F(x) \leq 1$ γιά κάθε $x \in R$.
2.	$F(x) \uparrow$, δηλαδή μή φθίνουσα: ἄν $x_1 < x_2$ τότε $F(x_1) \leq F(x_2)$.
3.	$F(-\infty) = \lim\limits_{x \to -\infty} F_X(x) = 0$ καί $F(+\infty) = \lim\limits_{x \to +\infty} F_X(x) = 1.$
4.	$F(x)$ συνεχής ἀπό τά δεξιά γιά κάθε $x \in R$.

Ἀπόδειξη:

1. Προφανής.

2. Ἐπειδή $\{-\infty < X \leq x_1\} \subset \{-\infty < X \leq x_2\}$ τό ἀποτέλεσμα προκύπτει ἀπό τό θεώρημα 2.5.

3. Αὐστηρή ἀπόδειξη τῆς ἰδιότητας 3 ἀπαιτεῖ γνώσεις θεωρίας Μέτρου καί ξεφεύγει ἀπό τό σκοπό τοῦ βιβλίου.

Διαισθητικά βλέπουμε ὅτι τό ἐνδεχόμενο $\{X\leq-\infty\}$ εἶναι ἀδύνατο καί τό ἐνδεχόμενο $\{X\leq+\infty\}$ εἶναι βέβαιο. Ἔτσι $F(-\infty) = 0$ καί $F(+\infty) = 1$.

4. Αὐστηρή ἀπόδειξη τῆς 4 ξεφεύγει ἀπό τό σκοπό τοῦ βιβλίου. Ἡ δεξιά συνέχεια ὀφείλεται στό γεγονός ὅτι ἡ $F(x)$ ὁρίζεται ὡς $P(X\leq x)$. ▼

Θεώρημα 3.2:

<u>Ἔστω</u> X <u>μία</u> <u>τ.μ.</u> <u>μέ</u> <u>α.σ.κ.</u> $F(x)$. <u>Ἰσχύουν</u> <u>οἱ</u> <u>ἀκόλουθες</u> <u>σχέσεις</u>:

(a) $P(X>x) = 1-F(x)$.

(b) $P(a<X\leq b) = F(b)-F(a)$.

(c) $P(X<x) = \lim\limits_{y\to x-} F(y) \equiv F(x-)$.

(d) $P(X=x) = P(X\leq x)-P(X<x)$

$= F(x)-F(x-)$.

(e) $P(a<X<b) = F(b-)-F(a)$.

Ἀπόδειξη: Ἡ (a) εἶναι προφανής. Γιά τήν (b) παρατηροῦμε ὅτι

$$\{-\infty<X\leq b\} = \{-\infty<X\leq a\}\cup\{a<X\leq b\}.$$

Ἔτσι ἄν πάρουμε τίς πιθανότητες καί στά δύο μέλη ἔχουμε

$$F(b) = F(a)+P(a<X\leq b)$$

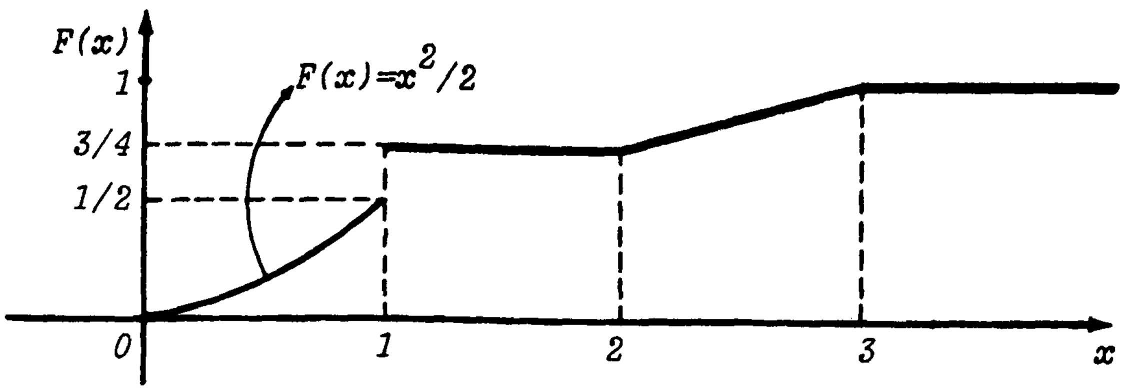
ἤ

$$P(a<X\leq b) = F(b)-F(a).$$

Ἡ ἀπόδειξη τῶν (c) καί (d) ξεφεύγουν ἀπό τό σκοπό τοῦ βιβλίου. Σημειώνουμε ὅτι $F(x-)$ εἶναι τό ἀριστερό ὅριο τῆς $F(\cdot)$ στό σημεῖο x. Δηλαδή $F(x-) = \lim_{h\to 0} F(x-h)$, $h>0$. Ἡ (e) ἀποδεικνύεται εὔκολα μέ τίς σχέσεις (b) καί (c). ▼

Παράδειγμα 3.9:

Ἔστω X τυχαία μεταβλητή μέ ἀθροιστική συνάρτηση κατανομῆς $F(x)$ ὅπως δίδεται στό Σχῆμα 3.6.



Σχῆμα 3.6

Νά ὑπολογιστοῦν τά ἀκόλουθα

a. $P(X=1/2)$ d. $P(X<1)$

b. $P(X=1)$ e. $P(X>2)$

c. $P(X\leq 1)$ f. $P(1/2<X<5/2)$.

Ἀπό τό σχῆμα μποροῦμε νά βροῦμε ὅτι ἡ ἐξίσωση τῆς α.σ.κ. τῆς τυχαίας μεταβλητῆς εἶναι

$$F(x) = \begin{cases} 0 & , \quad x<0 \\ x^2/2 & , \quad 0\leqq x<1 \\ 3/4 & , \quad 1\leqq x\leqq 2 \\ (x+1)/4 & , \quad 2\leqq x\leqq 3 \\ 1 & , \quad x\geqq 3 \;. \end{cases}$$

Γιά νά ὑπολογίσουμε τά a καί b χρησιμοποιοῦμε τήν ἰδιότητα (d) τοῦ θεωρήματος 3.2.

a. $$P(X=1/2) = 0 \;,$$

διότι $F(x)$ συνεχής στό σημεῖο $x = 1/2$.

b. $$P(X=1) = F(1)- \lim_{h\to 0+} F(1-h) = 3/4-1/2 = 1/4$$

διότι

$$\lim_{h\to 0+} F(1-h) = \lim_{h\to 0} \frac{(1-h)^2}{2} = \lim_{h\to 0+} \frac{(1-2h+h^2)}{2} = 1/2 \,.$$

c. $$P(X\leqq 1) = F(1) = 3/4 \,.$$

d. $$P(X<1) = P(X\leqq 1)-P(X=1)$$

$$= F(1)-1/4 = 3/4-1/4 = 1/2$$

ἤ

$$P(X<1) = F(1) = 1/2 \,.$$

$e.\quad P(X>2) = 1-P(X\leq 2)$

$$= 1-F(2) = 1-3/4 = 1/4\ .$$

$f.\quad P(1/2<X<5/2) = P\ (1/2<X\leq 5/2)-P(X=5/2)$

$$= F(5/2)-F(1/2)-P(X=5/2)$$

$$= (3\ 5)/4-1/8-0 = 3/4.$$

Ἡ τ.μ. τοῦ παραδείγματος αὐτοῦ δέν εἶναι οὔτε συνεχής, (διότι $P(X=1) = 1/4$) οὔτε ἀπαριθμητή (διότι τό σύνολο τῶν τιμῶν της δέν εἶναι ἀριθμήσιμο).

3.5 Συναρτήσεις τυχαίων μεταβλητῶν

Συχνά ἀντί γιά τή τυχαία μεταβλητή X (πού συνδέεται μέ κάποιο δειγματικό χῶρο S) χρησιμοποιοῦμε μία πραγματική συνάρτηση $h(X)$ τῆς X. Αὐτό γίνεται εἴτε γιά λόγους εὐκολίας εἴτε διότι ἡ τ.μ. X εἶναι ἀδύνατο ἤ δύσκολο νά παρατηρηθεῖ εἴτε διότι ὑπάρχουν τεχνικοί (μαθηματικοί) λόγοι πού ἐπιβάλουν τή χρήση τῆς $h(X)$. Εἶναι φανερό ὅτι ἡ $h(X)$ εἶναι καί αὐτή τυχαία μεταβλητή. Ἡ εὕρεση τῆς κατανομῆς τῆς $h(X)$ εἶναι ἕνα ἐνδιαφέρον πρόβλημα πού θά μελετήσουμε στό Κεφάλαιο 6. Οἱ τιμές τῆς $h(X)$ θά συμβολίζονται μέ $h(x)$. Γενικά (μετρήσιμες) συναρτήσεις τυχαίων μεταβλητῶν εἶναι τυχαῖες μεταβλητές. Ὑπάρχουν βέβαια συναρτήσεις τῆς X πού δέν εἶναι τυχαῖες μεταβλητές. Αὐτές ἔχουν μόνο θεωρητική σημασία καί δέν θά μᾶς ἀπασχολήσουν στό βιβλίο αὐτό.

Παραδείγματα συναρτήσεων τυχαίων μεταβλητῶν εἶναι τά ἀκόλουθα:ὅταν ἐνδιαφερόμαστε γιά τίς σχετικές καί ὄχι τίς ἀπόλυτες μεταβολές μεταξύ τῶν τιμῶν μίας τυχαίας μεταβλητῆς συχνά χρησιμοποιοῦμε τή λογαριθμική κλίμακα. Μέ ἄλλα λόγια μελετᾶμε τήν τυχαία μεταβλητή lnX ὅπου X εἶναι ἡ ἀρχική μεταβλητή. Συχνά γιά λόγους εὐκολίας ἤ ἄλλους τεχνικούς λόγους οἱ μετρήσεις $X_1, \ldots, X_n$ κωδικοποιοῦνται ἀλλάζοντας τή θέση τους ἤ τήν κλίμακα τους ἤ καί τά δύο. Ἀ λ λ α γ ή τ ῆ ς θ έ σ ε ω ς τῶν μετρήσεων σημαίνει ἐφαρμογή τοῦ μετασχηματισμοῦ $h(X) = X+b$, ὅπου b μία σταθερή ποσότης. Ἀ λ λ α γ ή τ ῆ ς κ λ ί μ α κ ο ς τῶν μετρήσεων σημαίνει ἐφαρμογή τοῦ μετασχηματισμοῦ $h(X) = aX$, ὅπου a σταθερή ποσότης. Ἀ λ λ α γ ή τ ῆ ς θ έ σ ε ω ς καί τ ῆ ς κ λ ί μ α κ ο ς τῶν μετρήσεων σημαίνει ἐφαρμογή τοῦ μετασχηματισμοῦ $h(X) = aX+b$. Οἱ μετασχηματισμοί αὐτοί εἶναι συναρτήσεις τῆς τυχαίας μεταβλητῆς X. Ἡ ἀλλαγή ἀπό καρτεσιανές συντεταγμένες σέ πολικές περικλείει συναρτήσεις τυχαίων μεταβλητῶν.

Α Σ Κ Η Σ Ε Ι Σ

3.1. Νά δειχθεῖ ὅτι ἡ συνάρτηση

$$P(X=x) = \begin{cases} 2^{x-1}/3^x , & x=1,2,\ldots \\ 0 , & \text{ἀλλοῦ} \end{cases}$$

ἀποτελεῖ μία σ.π. διακριτῆς τ.μ.

3.2. Ἔστω X τ.μ. μέ σ.π.

$$p(x) = c/x^2 , \quad x=1,2,\ldots$$

Νά βρεθεῖ ἡ τιμή τῆς σταθερῆς c.

3.3. Νά βρεθεῖ ἡ συνάρτηση πιθανότητας (σ.π.) ἀπό τήν ἀκόλουθη α.σ.κ.

$$F(x) = \begin{cases} 0 , & -\infty<x<1 \\ 1/4 , & 1\leq x<4 \\ 3/4 , & 4\leq x<6 \\ 1 , & x\geq 6 \end{cases} .$$

3.4. Ἕνα ραδιόφωνο περιέχει 6 τρανζίστορς, δύο ἀπό τά ὁποῖα εἶναι ἐλαττωματικά. Τρία τρανζίστορς ἐκλέγονται στήν τύχη καί ἐλέγχονται. Ἄν Y εἶναι ὁ ἀριθμός τῶν ἐλαττωματικῶν τρανζίστορς νά βρεθεῖ ἡ κατανομή πιθανότητας τοῦ Y.

3.5. Νά ἐλεγχθεῖ ἄν ἡ συνάρτηση

$$F(x) = \begin{cases} 1-e^{-5x^2} , & x>0 \\ 0 , & x\leq 0 \end{cases}$$

μπορεῖ νά παριστᾶ τήν α.σ.κ. μίας συνεχοῦς τ.μ.
"Αν ναί, νά βρεθεῖ ἡ σ.π.π. αὐτῆς.

3.6. "Εστω X μία τ.μ. μέ σ.π.π.

$$f(x) = \begin{cases} 1/x^2 , & 1<x<\infty \\ 0 , & \text{ἀλλοῦ} . \end{cases}$$

"Αν $A_1 = \{x:1<x<2\}$ καί $A_2 = \{x:4<x<5\}$

νά βρεθοῦν

(i) $P(A_1 U A_2)$ καί (ii) $P(A_1 A_2)$.

3.7. "Εστω X μία συνεχής τ.μ. μέ σ.π.π.

$$f(x) = \begin{cases} kx , & 0\leq x\leq 1/2 \\ k(1-x) , & 1/2\leq x<1 \\ 0 , & \text{ἀλλοῦ} . \end{cases}$$

(i) Νά βρεθεῖ ἡ τιμή τῆς σταθερῆς k.

(ii) "Αν A εἶναι τό ἐνδεχόμενο $\{X<1/2\}$ καί B τό ἐνδεχόμενο $\{1/4<X<3/4\}$ νά βρεθεῖ ἡ πιθανότητα τοῦ A καί ἡ πιθανότητα τοῦ B.

(iii) Εἶναι τά A καί B ἀνεξάρτητα ἐνδεχόμενα;

3.8. "Εστω X τ.μ. μέ κατανομή

$$f(x) = \begin{cases} 0,2 , & -1<x\leq 0 \\ 0,2+cx , & 0<x<1 \\ 0 , & \text{ἀλλοῦ} . \end{cases}$$

(i) Νά ὑπολογιστεῖ ἡ σταθερά c.

(ii) Νά ὑπολογιστεῖ ἡ πιθανότητα $P(-0,5\leq X<0,5)$.

3.9. Ἔστω X τ.μ. μέ α.σ.κ.

$$F(x) = \begin{cases} c(1-1/x^4) & , \quad x \geq 1 \\ 0 & , \quad x<1 . \end{cases}$$

(*i*) Νά βρεθεῖ ἡ τιμή τῆς σταθερᾶς c.

(*ii*) Νά βρεθεῖ ἡ σ.π.π. τῆς τ.μ. X.

3.10. Μπορεῖ ἡ συνάρτηση

$$f(x) = \begin{cases} x-2 , & 0<x<2 \\ 0 , & \text{ἀλλοῦ} \end{cases}$$

νά εἶναι μία συνάρτηση πυκνότητας πιθανότητας;

3.11. Νά βρεθεῖ ἡ ἀθροιστική συνάρτηση κατανομῆς (α.σ.κ.) ἀπό τήν ἀκόλουθη συνάρτηση πυκνότητας πιθανότητας (σ.π.π.).

$$f(x) = \begin{cases} xe^{-x} & , \quad x>0 \\ 0 & , \quad x \leq 0 . \end{cases}$$

3.12. Δίδεται τ.μ. X μέ κατανομή

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

Ἄν $A_1 = \{x: 1<x \leq 4\}$, $A_2 = \{x: 2<x \leq 5\}$ νά ὑπολογιστοῦν οἱ πιθανότητες $P(A_1+A_2)$, $P(A_1A_2)$, $P(A_1)$ καί $P(A_2)$.

3.13. Ἡ τ.μ. ἔχει σ.π.π.

$$f(x) = \begin{cases} 1-|x| & , \quad |x|<1 \\ 0 & , \quad \text{ἀλλοῦ} . \end{cases}$$

Νά βρεθεῖ ἡ α.σ.κ. $F(x)$ τῆς X καί νά ὑπολογιστεῖ ἡ πιθανότητα $P(|X|<1/2)$.

3.14. Ἔστω X τ.μ. μέ σ.π.π.

$$f(x) = \begin{cases} 12(1-x)^2/11 & , \quad -1/2\leq x\leq 1/2 \\ 0 & , \quad \text{ἄλλοῦ} . \end{cases}$$

Νά βρεθεῖ ἡ α.σ.κ. τῆς X.

3.15. Νά βρεθοῦν οἱ σ.π.π. ἀπό τίς ἀκόλουθες α.σ.κ.

(i) $$F(x) = \begin{cases} 0 & , \quad x\leq 0 \\ x^2/2 & , \quad 0<x\leq 1 \\ -1/8+(3x-x^2/2)/4 & , \quad 1<x\leq 3 \\ 1 & , \quad x>3 \end{cases}$$

(ii) $$F(x) = \begin{cases} 0 & \quad x\leq 0 \\ x(3-2\sqrt{2})+2(\ 2-1), & \quad 0<x<1 \\ 1 & , \quad x\geq 1 . \end{cases}$$

3.16. Ὁ χρόνος ἀναμονῆς (σέ λεπτά) σέ στάση λεωφορείου ἔχει συνάρτηση κατανομῆς

$$F(x) = \begin{cases} 0 & , \quad x\leq 0 \\ x/10 & , \quad 0\leq x\leq 5 \\ 1/2 & , \quad 5\leq x\leq 10 \\ x/20 & , \quad 10\leq x\leq 20 \\ 1 & , \quad 20\leq x<\infty . \end{cases}$$

(*i*) Παραστήσατε γραφικώς τήν $F(x)$.

(*ii*) Εἶναι ἡ κατανομή συνεχής;

(*iii*) Ποιά ἡ πιθανότητα νά περιμένει κανείς α) περισσότερο ἀπό πέντε λεπτά, β) λιγώτερο ἀπό δέκα λεπτά καί γ) περισσότερο ἀπό δεκαπέντε λεπτά δεδομένου ὅτι ἤδη περιμένει ἐπί πέντε λεπτά;

3.17. Ἕνας πωλητής βαρέων μηχανημάτων μπορεῖ νά ἐπικοινωνήσει μέ ἕνα ἤ δύο ὑποψήφιους πελάτες τήν ἡμέρα μέ πιθανότητα πωλήσεως *1/3* καί *2/3* ἀντίστοιχα. Κάθε ἐπικοινωνία ἔχει ὡς ἀποτέλεσμα μηδέν πώληση ἤ μία πώληση *500000* δραχμῶν μέ πιθανότητες *9/10* καί *1/10* ἀντίστοιχα. Νά δοθεῖ ἡ κατανομή πιθανότητας τῶν πωλήσεων τῆς ἡμέρας.

3.18. Ἡ διάρκεια σέ λεπτά τῶν ὑπεραστικῶν τηλεφωνικῶν συνδιαλέξεων ἀκολουθεῖ μία κατανομή μέ α.σ.κ.

$$F(x) = \begin{cases} 1-e^{-x/5} & , \; x>0 \\ 0 & , \; x\leq 0 \, . \end{cases}$$

Νά βρεθεῖ ἡ πιθανότητα ἡ διάρκεια μίας ὑπεραστικῆς συνδιαλέξεως (*i*) νά ὑπερβεῖ τά *5* λεπτά καί (*ii*) νά εἶναι μικρότερη τῶν *6* λεπτῶν δεδομένου ὅτι ἔχει ἤδη διαρκέσει *3* λεπτά.

ΚΕΦΑΛΑΙΟ 4

ειδικες κατανομες

Στό Κεφάλαιο αὐτό παρουσιάζουμε διάφορες ειδικές κατανομές καί ἐφαρμογές αὐτῶν. Οἱ κατανομές αὐτές ἀποτελοῦν πρότυπα (μοντέλα) πιθανοτήτων. Μέ τή βοήθεια τους μποροῦμε νά ὑπολογίζουμε τίς πιθανότητες διαφόρων ἐνδεχομένων χωρίς νά χρειάζεται νά ἀνατρέχουμε κάθε φορά στίς βασικές ἀρχές τῆς θεωρίας τῶν Πιθανοτήτων. Πρῶτα παρουσιάζουμε τίς εἰδικές διακριτές κατανομές καί στή συνέχεια τίς εἰδικές συνεχεῖς κατανομές.

Α΄ ΕΙΔΙΚΕΣ ΔΙΑΚΡΙΤΕΣ ΚΑΤΑΝΟΜΕΣ

4.1 Διωνυμικὸ πείραμα - Δοκιμὲς Bernoulli

Πολλά ἀπό τά τυχαῖα πειράματα ἔχουν τό χαρακτηριστικό ὅτι ἀποτελοῦνται ἀπό μία πεπερασμένη ἤ ἄπειρη σειρά δοκιμῶν (προσπαθειῶν) ἤ ἐπαναλήψεων ἑνός μερικοῦ πειράματος τοῦ ὁποίου τά ἀποτελέσματα εἶναι μόνον δύο: ἕνα ἐνδεχόμενο ἤ τό ἀντίθετο αὐτοῦ. Γιά εὐκολία ἄς ὀνομάσουμε τά ἀποτελέσματα αὐτά ἐπιτυχία καί ἀποτ υ χ ί α. Θά συμβολίζουμε τήν έπιτυχία μέ τό E καί τήν άποτυχία μέ τό A. Στή σειρά αύτή τῶν δοκιμῶν ένδιαφερόμαστε νά μελετήσουμε τόν άριθμό τῶν έπιτυχιῶν καί ίδιαίτερα νά βροῦμε τή κατανομή πιθανότητας τοῦ άριθμοῦ αύτοῦ.

Ἄς άναφέρουμε μερικά παραδείγματα: σέ μία σφυγμομέτρηση τῆς κοινῆς γνώμης ένδιαφερόμαστε νά μελετήσουμε τήν κατανομή τοῦ άριθμοῦ τῶν άτόμων πού ύποστηρίζουν τή θέση ένός πολιτικοῦ κόμματος μεταξύ έκείνων πού λαμβάνουν μέρος στή σφυγμομέτρηση. Οἱ δοκιμές τοῦ πειράματος εἶναι ή έξέταση τῶν άτόμων πού λαμβάνουν μέρος στή σφυγμομέτρηση. Ἐπιτυχία εἶναι ή ύποστήριξη άπό ἕνα ἄτομο τῆς θέσης τοῦ πολιτικοῦ κόμματος καί άποτυχία τό άντίθετο. Ἐδῶ ὁ πληθυσμός θεωρεῖται ἄπειρος.

Σέ μία μελέτη γιά τόν ήλεκτροφωτισμό μίας άγροτικῆς περιοχῆς ένδιαφερόμαστε νά βροῦμε τό ποσοστό τῶν άγροτικῶν σπιτιῶν πού ἔχουν ήλεκτρικό ρεῦμα. Τριακόσια σπίτια έκλέγονται στήν τύχη κ α ί μ έ έ π α ν ά θ ε σ η καί τό κάθε ἕνα έξετάζεται ἄν ἔχει ήλεκτρικό ρεῦμα. Τά τριακόσια σπίτια άποτελοῦν τόν άριθμό δοκιμῶν τοῦ πειράματος. Ἐπιτυχία εἶναι τό νά ἔχει ἕνα σπίτι ήλεκτρικό ρεῦμα καί άποτυχία τό νά μήν ἔχει. Ποιά ή πιθανότητα ὁ άριθμός τῶν σπιτιῶν πού διαλέγουμε στό πείραμα καί πού ἔχουν ρεῦμα νά εἶναι μεγαλύτερος τοῦ όκτώ; Μία άπάντηση στό έρώτημα αύτό θά δοθεῖ πιό κάτω.

Χάριν εύκολίας θά συμβολίζουμε μέ n τόν άριθμό τῶν δοκιμῶν τοῦ πειράματος. Ἔστω δέ X ὁ άριθμός τῶν έπιτυχιῶν σέ n δοκιμές τοῦ πειράματος. Σκοπός μας εἶναι νά βροῦμε τή κατανομή πιθανότητας τῆς διακριτῆς τ.μ. X. Εἶναι φανερό ὅτι γιά νά πετύχουμε τό σκοπό αύτό πρέπει νά κάνουμε ὡρισμένες ὑποθέσεις - ἐξιδανικεύσεις. Ἔτσι δεχόμαστε ὅτι ἰσχύουν τά ἑξῆς:

(*i*)	Ἡ πιθανότητα ἐπιτυχίας ἤ ἀποτυχίας παραμένει σταθερή σέ ὅλες τίς δοκιμές τοῦ πειράματος.
(*ii*)	Οἱ δοκιμές τοῦ πειράματος εἶναι στατιστικῶς ἀνεξάρτητες.

Θά συμβολίζουμε τήν πιθανότητα ἐπιτυχίας μέ τό p, $0 \leq p \leq 1$, δηλαδή $P(E) = p$. Ἡ πιθανότητα ἀποτυχίας θά συμβολίζεται μέ τό q, δηλαδή $P(A) = q$. Ἔτσι $p+q = 1$.

Ὁρισμός 4.1: Διωνυμικό πείραμα.

Ἕνα πείραμα ἀποτελούμενο ἀπό δοκιμές ἡ κάθε μία ἀπό τίς ὁποῖες ἔχει μόνο δύο δυνατά ἀποτελέσματα E ἤ A καί γιά τό ὁποῖο ἰσχύουν οἱ παραπάνω ὑποθέσεις (*i*) καί (*ii*) λέγεται διωνυμικό πείραμα καί κάθε ἐπί μέρους δοκιμή λέγεται δοκιμή τοῦ *Bernoulli*.

Ὁ δειγματικός χῶρος n δοκιμῶν τοῦ *Bernoulli* περιλαμβάνει 2^n σημεῖα ἤ σειρές ἀπό n σύμβολα E καί A. Κάθε τέτοιο σημεῖο ἀντιπροσωπεύει ἕνα δυνατό ἀποτέλεσμα τοῦ διωνυμικοῦ πειράματος.

4. 2 Διωνυμική κατανομή

Ἄς θεωρήσουμε ἕνα διωνυμικό πείραμα στό ὁποῖο ἐκτελοῦμε n δοκιμές τοῦ *Bernoulli*, ὅπου n δοθείς ἀριθμός. Ἔστω X ἡ τ.μ. τῆς ὁποίας οἱ τιμές x παριστοῦν τόν ἀριθμό τῶν ἐπιτυχιῶν πού θά σημειωθοῦν στήν ἀκολουθία τῶν n δοκιμῶν. Εἶναι φανερό ὅτι ἡ X εἶναι μία διακριτή τ.μ. μέ τιμές $0,1,2,\ldots,n$.

Ἄς εὕρουμε τώρα τήν κατανομή πιθανότητας τῆς X δηλαδή τή συνάρτηση $p_X(x) = P(X=x)$. Ἔστω ὅτι τό x εἶναι σταθερό. Τά ἀποτελέσματα τοῦ διωνυμικοῦ πειράματος εἶναι τῆς μορφῆς:

$$EEAAEA \ldots E \;\; \text{ἤ} \;\; AAEA \ldots AE, \text{ κλπ.}, \qquad (4.1)$$

ὅπου ὁ συνολικός ἀριθμός τῶν E καί A εἶναι n. Ὅταν $X=x$ τά ἀποτελέσματα περιέχουν x E καί $n-x$ A. Τότε, λόγω τῶν συνθηκῶν (i) καί (ii) τοῦ διωνυμικοῦ πειράματος, ἡ πιθανότητα κάθε ἑνός ἀποτελέσματος μέ x E καί $n-x$ A εἶναι $p^x(1-p)^{n-x}$. Ἀπομένει νά ὑπολογίσουμε τόν ἀριθμό τῶν ἀποτελεσμάτων τῆς μορφῆς (4.1) ὅταν $X=x$. Ὁ ἀριθμός αὐτός εἶναι ὁ ἀριθμός τῶν συνδυασμῶν τῶν n θέσεων ἀνά x στίς ὁποῖες x θέσεις ἐμφανίζεται τό E, δεδομένου ὅτι ἡ σειρά μέ τήν ὁποία γίνονται οἱ ἐπιτυχίες E καί οἱ ἀποτυχίες A δέν παίζει κανένα ρόλο. Ὁ ἀριθμός αὐτός ἰσοῦται ἀκόμη μέ τόν ἀριθμό τῶν μεταθέσεων n πραγμάτων ἀπό τά ὁποῖα x εἶναι ἴδια, (E), καί $n-x$ εἶναι ἐπίσης ἴδια ἀλλά ἄλλου τύπου, (A), δηλαδή $n!/[x!(n-x)!]$. Συνεπῶς, ἔχουμε

$$\begin{aligned} p_X(x) &= P\{X=x\} = P\{EEAE\ldots EA \;\; \text{ἤ} \;\; AAEA\ldots AE \;\; \text{ἤ} \;\; \ldots\} \\ &= p^x(1-p)^{n-x}+p^x(1-p)^{n-x}+\ldots = \binom{n}{x}p^x q^{n-x}. \end{aligned}$$

Ἔτσι

$$p_X(x) = \binom{n}{x} p^x q^{n-x} \;, \; x=0,1,2,\ldots \qquad (4.2)$$

Ἡ παραπάνω σχέση δίνει τή λεγόμενη δ ι ω ν υ μ ι κ ή κ α τ α ν ο μ ή. Βάσει τοῦ θεωρήματος τοῦ Νεύτωνα (1.1) ἔχουμε

$$\sum_{x=0}^{n} p_X(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (p+q)^n = 1 .$$

Ἄρα ἡ $p_X(x)$ δίνει μία ἀληθή σ.π.

Μία τ.μ. X ἡ ὁποία ἔχει τήν κατανομή *(4.2)* θά λέγεται δ ι ω ν υ μ ι κ ή τ.μ. Γιά συντομία αὐτό θά συμβολίζεται μέ

$$X \sim B(n,p) .$$

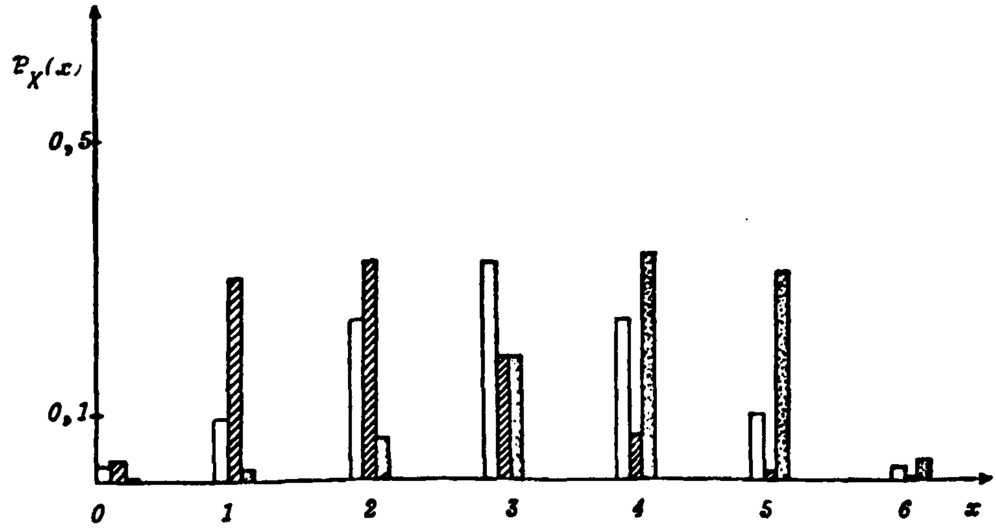
Τά n,p λέγονται καί π α ρ ά μ ε τ ρ ο ι τῆς διωνυμικῆς κατανομῆς. Ἐάν $n = 1$ τότε ἡ X λέγεται καί τ.μ. τ ο ῦ *B e r n o u l l i* καί ἡ κατανομή αὐτῆς κατανομή τοῦ *Bernoulli*.

Ἡ α.σ.κ. τῆς X δίδεται ἀπό τή σχέση

$$F_X(x) = \begin{cases} 0 & , \quad x<0 \\ \sum_{y=0}^{[x]} \binom{n}{y} p^y q^{n-y} & , \quad 0 \le x \le n \\ 1 & , \quad x>n \end{cases}$$

ὅπου $[x]$ παριστᾶ τό ἀκέραιο μέρος τοῦ θετικοῦ ἀριθμοῦ x.

Τό Σχῆμα 4.1 δίνει γραφικές παραστάσεις τῆς 4.2 γιά $(n = 6,\ p = 0,5)$, $(n = 6,\ p = 0,3)$ καί $(n = 6,\ p = 0,7)$. Παρατηροῦμε ὅτι γιά $p = 0,5$ ἡ γραφική παράσταση εἶναι συμμετρική. Γιά $p = 0,3$ ἡ γραφική παράσταση ἀνεβαίνει ἀπότομα καί μετά κατεβαίνει πιό ὁμαλα ἐνῶ ἀντίθετη εἶναι ἡ συμπεριφορά της γιά $p = 0,7$. Ἡ παρατήρηση αὐτή εἶναι γενικώτερη καί ἰσχύει γιά κάθε $p < 0,5$ καί $p > 0,5$ ἀντίστοιχα.



Κατανομή ▯ : $n = 6$, $p = 0,5$

Κατανομή ▨ : $n = 6$, $p = 0,3$

Κατανομή ▦ : $n = 6$, $p = 0,7$

Σχῆμα 4.1: Διωνυμικές κατανομές

Ὁ ὑπολογισμός τῶν διωνυμικῶν πιθανοτήτων βάσει τῆς *(4.2)* εἶναι σχετικά ἁπλός γιά μικρές τιμές τοῦ n $(n \leq 20)$. Γιά μεγαλύτερες τιμές τοῦ n ἡ ὑπολογιστική προσπάθεια δέν εἶναι ἀσήμαντη καί γίνεται ὑπερβολική γιά μεγάλες τιμές τοῦ n. Ὁ ἀναγωγικός τύπος

$$p_X(x) = \frac{(n-x+1)p}{xq}\, p_X(x-1) \quad , \quad x = 1,2,\ldots,n ,$$

εἶναι χρήσιμος καίτοι συνήθως γιά μεγάλες τιμές τοῦ n

προτιμοῦμε τόν κατά προσέγγιση ὑπολογισμό τοῦ $p_X(x)$ μέ τήν κατανομή τοῦ *Poisson* ἤ τήν κανονική κατανομή ὅπως θά δοῦμε πιό κάτω.

Πίνακες δίνοντες τίς διωνυμικές πιθανότητες γιά διάφορες τιμές τοῦ n καί p χρησιμοποιοῦνται εὐρύτατα στή Στατιστική. Ἕνας τέτοιος πίνακας εἶναι ὁ Πίνακας Ι τοῦ Παραρτήματος. Ἐπίσης ὑπάρχουν καί πίνακες δίνοντες τήν ἀθροιστική διωνυμική συνάρτηση κατανομῆς.

Ἡ ἑπόμενη σχέση εἶναι χρήσιμη καί συνδέει μία διωνυμική τ.μ. μέ ἐπί μέρους τ.μ. τοῦ *Bernoulli*. Ἔστω X ὁ ἀριθμός τῶν ἐπιτυχιῶν ἑνός διωνυμικοῦ πειράματος μέ n δοκιμές καί $P(E) = p$. Τότε $X \sim B(n,p)$. Ἔστω ἐπίσης οἱ τ.μ.

$$X_i = \begin{cases} 1 & \text{ἐάν τό ἀποτέλεσμα τῆς } i \text{ δοκιμῆς εἶναι } E. \\ 0 & \text{ἐάν τό ἀποτέλεσμα τῆς } i \text{ δοκιμῆς εἶναι } A. \end{cases}$$

Τότε

$$X = X_1 + X_2 + \dots + X_n. \qquad (4.4)$$

Μέ ἄλλα λόγια **τό ἄθροισμα n ἀνεξάρτητων τ.μ. τοῦ *Bernoulli* μέ παράμετρο p εἶναι μία διωνυμική κατανομή μέ παραμέτρους n καί p.**

Ἐπίσης ἔχουμε

$$P\{x \text{ ἐπιτυχίες } E \text{ σέ } n \text{ δοκιμές}\}$$

$$= P\{n-x \text{ ἀποτυχίες } A \text{ σέ } n \text{ δοκιμές}\}$$

ἤ

$$\binom{n}{x} p^x q^{n-x} = \binom{n}{n-x} q^{n-x} p^x.$$

Ἡ σχέση αὐτή εἶναι χρήσιμη γιά τόν ὑπολογισμό διωνυμικῶν πιθανοτήτων ἀπό διωνυμικούς πίνακες στούς ὁποίους τό p περιορίζεται ἀπό 0 μέχρι $1/2$.

Ἐφαρμογές: Τό πιό γνωστό καί ἁπλό παράδειγμα δοκιμῶν τοῦ *Bernoulli* δίνεται ἀπό διαδοχικές ρίψεις ἑνός ἀληθινοῦ ἤ συμμετρικοῦ νομίσματος. Ἐδῶ $p = q = 1/2$. Ἐάν τό νόμισμα δέν εἶναι συμμετρικό πάλι ὑποθέτουμε ὅτι οἱ διαδοχικές ρίψεις εἶναι ἀνεξάρτητες ὥστε νά ἔχουμε τό πρότυπο τῶν δοκιμῶν τοῦ *Bernoulli* στίς ὁποῖες ἡ πιθανότητα p ἐπιτυχίας μπορεῖ νά πάρει μία αὐθαίρετη τιμή. Ὁμοίως ἡ διαδοχική ρίψη ἑνός ἀληθινοῦ ζαριοῦ καί ἡ διάκριση μεταξύ τῶν ἐνδεχομένων ἄσσος δηλαδή (E) καί μή ἄσσος (A) ὁδηγεῖ σέ δοκιμές τοῦ *Bernoulli* μέ $p = 1/6$ ἐνῶ ἡ διάκριση μεταξύ περιττοῦ ἤ ἄρτιου ἀποτελέσματος ὁδηγεῖ σέ διωνυμικό πείραμα μέ $p = 1/2$. Ἄλλη ἐφαρμογή προσφέρεται στόν Βιομηχανικό Ἔλεγχο Ποιότητας ὅπου ἀντικείμενα-εἴδη (π.χ. βίδες, λαμπτῆρες, κουζίνες κλπ.) παράγονται μαζικῶς καί κατά τόν ἔλεγχο ποιότητας ταξινομοῦνται ὡς E στήν περίπτωση πού ἐκπληρώνουν τίς προδιαγραφές καί ὡς A στήν περίπτωση πού δέν ἐκπληρώνουν τίς προδιαγραφές. Δοκιμές τοῦ *Bernoulli* ἔχουμε καί στίς δειγματοληψίες μέ ἐπανάθεση ἀπό πεπερασμένους πληθυσμούς.

Τό σχῆμα τῶν δοκιμῶν τοῦ *Bernoulli* εἶναι ἕνα θεωρητικό πρότυπο (μοντέλο) πού μόνο ἡ ἐμπειρία μπορεῖ νά δείξει ἐάν εἶναι κατάλληλο γιά τήν περιγραφή ἑνός συγκεκριμένου πειράματος. Ἔτσι ἡ γνώση μας ὅτι οἱ διαδοχικές ρίψεις νομίσματος ἀκολουθοῦν τό πρότυπο τῶν δοκιμῶν τοῦ

Bernoulli πηγάζει ἀπό ἐμπειριακή μαρτυρία. Ὁ ἁπλός ἄνθρωπος πιστεύει ὅτι ἐάν φέρουμε στή σειρά *17* κορῶνες, τά γράμματα γίνονται κατόπιν πιό πιθανά. Αὐτό δέν σημαίνει κατ'ἀνάγκη ὅτι τό νόμισμα εἶναι ἐλαττωματικό. Εἶναι πολύ πιθανό νά μήν ἔχουμε ἀνεξαρτησία τῶν δοκιμῶν.

Στίς δειγματοληψίες, τό βιομηχανικό ποιοτικό ἔλεγχο κλπ. τό σχῆμα τῶν δοκιμῶν τοῦ *Bernoulli* δίνει ἕνα ἰδανικό πρότυπο ἄσχετα καί ἄν ποτέ δέν ἐπιτυγχάνεται ἑκατό τοῖς ἑκατό. Πολλές φορές μέ τήν πάροδο τοῦ χρόνου ἡ $P(E) = p$ δέν παραμένει σταθερή διότι π.χ. ἐπέρχονται ἀλλοιώσεις στό προϊόν τῆς βιομηχανικῆς γραμμῆς. Ἀπό πλευρᾶς ποιοτικοῦ ἐλέγχου εἶναι ἐπιθυμητό ἡ διαδικασία παραγωγῆς νά ἀκολουθεῖ τό σχῆμα τοῦ *Bernoulli*. Ὁ συνεχής ἔλεγχος ἀποσκοπεῖ στό νά ἀνακαλύψει βασικές ἀλλαγές ἀπό τό ἰδανικό σχῆμα καί νά χρησιμοποιήσει αὐτές σάν ἔνδειξη ἐπερχόμενης ἀνωμαλίας.

<u>Παραδείγματα:</u>

4.1: Εἶναι γνωστό ὅτι κατά μέσο ὅρο *30%* τῶν ἀγροτικῶν σπιτιῶν μίας περιοχῆς διαθέτουν ἠλεκτρικό ρεῦμα. Ποιά ἡ πιθανότητα σέ μία δειγματοληψία *20* σπιτιῶν

α) Τό πολύ τρία ἀπό αὐτά νά διαθέτουν ἠλεκτρικό,

β) Τουλάχιστον *8* ἀπό αὐτά νά διαθέτουν ἠλεκτρικό,

γ) Περισσότερα ἀπό τρία καί λιγώτερα ἀπό ὀκτώ νά διαθέτουν ἠλεκτρικό;

Ἐδῶ ἔχουμε διωνυμικό πείραμα μέ $n = 20$ καί $p = 0,3$. Συνεπῶς

α) $$P(X \leq 3) = \sum_{x=0}^{3} \binom{20}{x} (0,3)^x (0,7)^{20-x}$$

$$= 0,0008+0,0068+0,0278+0,0716 = 0,1070 .$$

β) $P(X \geq 8) = 1-P(X<8) = 1-P(X \leq 7)$

$$= 1-[P(X \leq 3)+P(X=4)+P(X=5)+P(X=6)+P(X=7)]$$

$$= 1-[0,1070+0,1304+0,1789+0,1916+0,1643]$$

$$= 0,2278.$$

γ) $P(3 \leq X<8) = P(X=3)+P(X=4)+P(X=5)+P(X=6)+P(X=7)$

$$= 0,0716+0,1304+0,1789+0,1916+0,1643$$

$$= 0,7368 .$$

4.2: Ἄς ὑποθέσουμε ὅτι δέκα ἐργάτες ἑνός ἐργοστασίου χρησιμοποιοῦν παροχές ἠλεκτρικοῦ ρεύματος σέ διάφορες φάσεις τῆς δουλειᾶς τους. Τό πρόβλημα πού ἀπασχολεῖ τό ἐργοστάσιο εἶναι πόσες παροχές χρειάζονται γιά νά ἐξυπηρετηθοῦν οἱ ἀνάγκες τῶν ἐργατῶν καλύτερα, δηλαδή κατά τρόπο ὥστε νά μήν περιμένει πολύ καιρό ὁ ἕνας ἐργάτης γιά νά χρησιμοποιήσει τήν παροχή. Ἐξυπακούεται ὅτι λόγω ὑψηλοῦ κόστους τό ἐργοστάσιο θέλει νά ἀποφύγει τήν ἐγκατάσταση δέκα παροχῶν.

Σάν μιά χονδρική προσέγγιση μποροῦμε νά ὑποθέσουμε ὅτι γιά κάθε χρονική στιγμή κάθε ἐργάτης ἔχει τήν ἴδια πιθανότητα p νά χρειασθεῖ μιά παροχή. Ἐάν οἱ ἐργάτες ἐργάζονται ὁ ἕνας ἀνεξάρτητα ἀπό τόν ἄλλο τότε ὁ ἀριθμός X τῶν ἐργατῶν πού μπορεῖ νά χρειαστεῖ παροχές τήν ἴδια στιγμή ἔχει τήν κατανομή $B(10,p)$. Τό ἐργοστάσιο γνωρίζει ἀκόμη ὅτι κατά μέσο ὅρο ὁ κάθε ἐργάτης χρησιμοποιεῖ τήν παροχή του *12* λεπτά τήν ὥρα. Αὐτό δίνει τήν τιμή $p = 12/60 = 1/5$ γιά τήν παράμετρο p. Ἄς δοῦμε τώρα τήν πιθανότητα ὥστε ἑπτά ἤ περισσότεροι ἐργάτες νά χρειαστοῦν παροχή τήν ἴδια στιγμή:

$$P(X \geq 7) = \sum_{x=7}^{10} \binom{10}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{10-x}$$

$$= 0{,}0008+0{,}0001+0{,}0000+0{,}0000$$

$$= 0{,}0009 \; .$$

Αὐτό σημαίνει ὅτι ἄν τό ἐργοστάσιο ἐγκαταστήσει ἕξη μόνο παροχές ἡ πιθανότητα γιά ὑπερζήτηση εἶναι

$$0{,}0009 = \frac{1}{10000/9} = 1/1111{,}11 \; ,$$

δηλαδή γιά ἕνα λεπτό περίπου στά *1111* λεπτά ἤ ἕνα λεπτό στίς *18,5* περίπου ὧρες.

4. 3 Ὑπεργεωμετρικὴ κατανομὴ

Ἡ κατανομή αὐτή ἐμφανίζεται ὅταν κάνουμε δειγματοληψία χωρίς ἐπανάθεση ἀπό ἕνα πεπερασμένο πληθυσμό. Ἄς ὑποθέσουμε ὅτι ὁ πληθυσμός ἀποτελεῖται ἀπό N στοιχεῖα κάθε ἕνα ἀπό τά ὁποῖα εἶναι δυνατόν νά ἀνήκει σέ μία ἀπό δύο κατηγορίες C καί C'. Ἔστω Np ὁ ἀριθμός τῶν στοιχείων πού ἀνήκουν στή κατηγορία C καί Nq ὁ ἀριθμός ἐκείνων πού ἀνήκουν στή κατηγορία C', ὅπου p καί q, $0 \leq p$, $q \leq 1$, εἶναι τά ποσοστά τῶν στοιχείων τῶν κατηγοριῶν C καί C' ἀντίστοιχα. Προφανῶς

$Np+Nq = N$ καί $p+q = 1$. Ἔστω τώρα ὅτι n στοιχεῖα ἐκλέγονται στήν τύχη ἀπό τόν πληθυσμό χωρίς ἐπανάθεση. Ἔστω X ἡ τ.μ. πού παριστάνει τόν ἀριθμό τῶν στοιχείων τῆς κατηγορίας C πού περιλαμβάνονται στό δεῖγμα τῶν n στοιχείων. Προφανῶς ἡ X εἶναι μία διακριτή τ.μ.

Δεδομένου ὅτι δέν ἐνδιαφερόμαστε γιά τή σειρά μέ τήν ὁποία ἐμφανίζονται τά στοιχεῖα ὁ ἀριθμός τῶν δυνατῶν δειγμάτων εἶναι $\binom{N}{n}$. Ἀπό αὐτά, ὁ ἀριθμός τῶν δειγμάτων πού περιέχουν x στοιχεῖα τῆς κατηγορίας C καί $n-x$ τῆς κατηγορίας C' εἶναι, $\binom{Np}{x}\binom{Nq}{n-x}$ βάσει τοῦ Κανόνα $m.n.$ (θεώρημα 1.1). Ἄρα

$$P\{X=x\} = \binom{Np}{x}\binom{Nq}{n-x}/\binom{N}{n} \; . \qquad (4.5)$$

Ἡ (4.5) δίνει τή σ.π. $p_X(x)$ τῆς τ.μ. X. Τό ὅτι ἡ (4.5) ἀποτελεῖ μία σ.π. ἀποδεικνύεται εὔκολα ἀπό τίς ἰδιότητες

$$\sum_{r=0}^{k} \binom{m}{r}\binom{n}{k-r} = \binom{m+n}{k}$$

καί

$$\binom{n}{r} = 0 \quad \text{γιά } n \text{ καί } r \text{ θετικά καί ἀκέραια καί } r>n.$$

τῶν διωνυμικῶν συντελεστῶν.

Ἡ κατανομή (4.5) λέγεται ὑ π ε ρ γ ε ω μ ε τ ρ ι κ ή καί κατά τόν ἴδιο τρόπο ἡ X λέγεται ὑ π ε ρ γ ε ω μ ε τ ρ ι κ ή τ.μ. καί συμβολίζεται μέ

$$X \sim Hg(N,n,p) \; .$$

Δοθέντων τῶν N καί n, τό p, $0 \leq p \leq 1$ ἀποτελεῖ τήν π α ρ ά μ ε τ ρ ο τῆς κατανομῆς.

Εἶναι εὔκολο νά δεῖ κανείς ὅτι οἱ τιμές πού μπορεῖ νά πάρει μία ὑπεργεωμετρική τ.μ. X μέ θετική πιθανότητα δέν εἶναι $0,1,\ldots,n$ ἀλλά ἐξαρτῶνται ἀπό τή σχέση πού ὑπάρχει μεταξύ τῶν n,Np καί Nq. Οἱ διωνυμικοί συντελεστές $\binom{n}{r}$ εἶναι μή μηδενικοί ὅταν τό r εἶναι ἀκέραιος καί $0 \leqq r \leqq n$. Ἄρα ἡ $p_X(x)$ εἶναι θετική ὅταν

$$x \leqq Np,\ n-x \leqq Nq,\ x \geqq 0 \text{ καί } n-x \geqq 0.$$

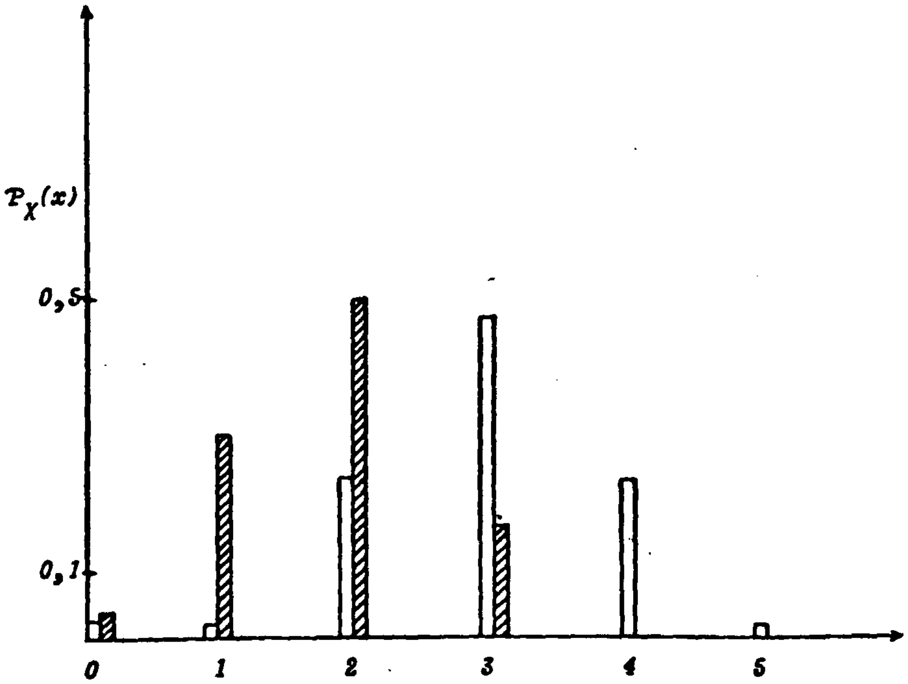
Ἀπό αὐτές τίς ἀνισότητες βλέπουμε ὅτι ἡ μέγιστη δυνατή τιμή τοῦ x εἶναι τό $min(n,Np)$ καί ἡ ἐλάχιστη δυνατή εἶναι τό $max(n-Nq,0)$. Ἄρα τό πεδίο τῶν τιμῶν μίας ὑπεργεωμετρικῆς τ.μ. εἶναι

$$max(0,n-Nq),\ldots,min(n,Np).$$

Ἡ ἀθροιστική συνάρτηση κατανομῆς τῆς ὑπεργεωμετρικῆς τ.μ. δίνεται ἀπό τή σχέση

$$F_X(x) = \begin{cases} 0 & , x < max(n-Nq,0) \\ \sum\limits_{y=max(n-Nq,0)}^{[x]} \binom{Np}{n}\binom{Nq}{n-y}/\binom{N}{n}, & max(0,n-Np) \leqq x \leqq min(n,Nq) \\ 1 & , x > min(n,Np). \end{cases} \tag{4.6}$$

Γιά τήν διευκόλυνση τῶν ὑπολογισμῶν ὑπάρχουν πίνακες δίνοντες τή σ.π. ἤ τήν α.σ.κ. τῆς ὑπεργεωμετρικῆς κατανομῆς γιά μικρές τιμές τοῦ $N(\leq 100)$. Συνήθως οἱ πίνακες αὐτοί χρησιμοποιοῦνται στή Στατιστική γιά τήν εὕρεση ἀκριβῶν διαστημάτων ἐμπιστοσύνης τῆς παραμέτρου p. Γιά τούς ὑπολογισμούς μποροῦμε νά χρησιμοποιοῦμε καί τόν ἐπαναλη-



Κατανομή ▯ : $N = 10$, $Np = 5$, $n = 6$

Κατανομή ▨ : $N = 10$, $Np = 3$, $n = 6$

Σχῆμα 4.2: Ὑπεργεωμετρικές Κατανομές

πτικό τύπο

$$p_X(x+1) = \frac{(n-x)(Np-x)}{(x+1)(Nq-n+x+1)} p_X(x) \quad ,$$

ξεκινῶντας ἀπό τήν τιμή τοῦ $p_X(0)$.

Τό Σχῆμα 4.2 δίνει τήν γραφική παράσταση τῆς *(4.5)* γιά διάφορες τιμές τῶν παραμέτρων της.

Ἄς θεωρήσουμε τώρα τή σ.π. τῆς ὑπεργεωμετρικῆς κατανομῆς ὡς μία ἀκολουθία συναρτήσεων τοῦ x καί πάρουμε τό ὅριο καθώς τό $N \to \infty$ μέ τά n καί p σταθερά. Τότε εἶναι εὔκολο νά δείξουμε ὅτι

$$\lim \binom{Np}{x}\binom{Nq}{n-x}/\binom{N}{x} = \binom{n}{x}p^x q^{n-x} \; .$$

Μέ ἄλλα λόγια γιά μεγάλες τιμές τοῦ N ἡ ὑπεργεωμετρική κατανομή προσεγγίζει τή διωνυμική. Πράγματι συχνά χρησιμοποιοῦμε τή διωνυμική κατανομή μέ τό ἴδιο p γιά νά προσεγγίσουμε τίς ὑπεργεωμετρικές πιθανότητες ἐ φ ' ὅ σ ο ν τ ό n δ έ ν ὑ π ε ρ β α ί ν ε ι τ ό 5% τ ο ῦ N.

Οἱ ἐ φ α ρ μ ο γ έ ς τῆς ὑπεργεωμετρικῆς κατανομῆς εἶναι πολύπλευρες. Στό βιομηχανικό ποιοτικό ἔλεγχο σύνολα ἀπό N ἀντικείμενα ὑπόκεινται στόν ἔλεγχο δειγματοληψίας. Κάθε ἐλεγχόμενο ἀντικείμενο χαρακτηρίζεται ὡς ἐλαττωματικό ἤ μή. Ἡ δειγματοληψία γίνεται συνήθως (καί κατά ρεαλιστικό τρόπο) χωρίς ἐπανάθεση. Τά ἐλαττωματικά ἀντικείμενα ἀνήκουν στή κατηγορία C καί τά μή ἐλαττωματικά στή κατηγορία C'. Τότε τό ὑπεργεωμετρικό πρότυπο εἶναι ἐφαρμόσιμο. Ὁ ἀριθμός Np τῶν ἐλαττωματικῶν ἀντικειμένων εἶναι φυσικά ἄγνωστος. Ἕνα δεῖγμα μεγέθους n λαμβάνεται καί ὁ ἀριθμός ἐλαττωματικῶν ἀντικειμένων προσδιορίζεται. Τό ὑπεργεωμετρικό πρότυπο μᾶς ἐπιτρέπει νά βγάλουμε συμπεράσματα γιά τό πιθανό μέγεθος τοῦ Np χρησιμοποιῶντας μεθόδους τῆς Στατιστικῆς.

Χρησιμοποιῶντας τό ὑπεργεωμετρικό πρότυπο πιθανότητας καί μεθόδους τῆς Στατιστικῆς μποροῦμε ἐπίσης νά ἐκτιμήσουμε τό μέγεθος ἑνός πληθυσμοῦ ζώων ἐφαρμόζοντες διπλή δειγματοληψία χωρίς ἐπανάθεση. Ἡ κεντρική ἰδέα ἔχει ὡς ἑξῆς: ἄς ὑποθέσουμε γιά παράδειγμα ὅτι θέλουμε νά ἐκτιμήσουμε τόν ἄγνωστο ἀριθμό N τῶν ψαριῶν στή Λίμνη τῶν Ἰωαννίνων. Γιά τό σκοπό αὐτό ψαρεύουμε (κατά κάποιο τρόπο) Np ψάρια, τά σημαδεύουμε μέ μία ἀνεξίτηλο κόκκινη κηλίδα καί τά ἀφήνουμε καί πάλι ἐλεύθερα. Μετά ἀπό μία ἤ δύο μέρες ξαναφαρεύουμε n ψάρια καί μετροῦμε τόν ἀριθμό x τῶν ψαριῶν πού ἔχουν τήν κόκκινη κηλίδα. Τό ὑπεργεωμετρικό πρότυπο εἶναι τώρα ἐφαρμόσιμο καί χρησιμοποιῶντας μεθόδους τῆς Στατιστικῆς πού θά μάθουμε παρακάτω μποροῦμε νά ἐκτιμήσουμε τό N.

Παράδειγμα 4.3:

Ἔστω ὅτι σέ μία τάξη ἀπό *20* φοιτητές τό *40%* εὐνοεῖ ἕνα νέο ἐσωτερικό κανονισμό. Ἐάν διαλέξουμε στή τύχη καί χωρίς ἐπανάθεση *6* φοιτητές, α) ποία ἡ πιθανότητα δύο ἀκριβῶς ἀπό αὐτούς νά εὐνοοῦν τόν ἐσωτερικό κανονισμό καί β) ποία ἡ πιθανότητα τουλάχιστον δύο ἀπό αὐτούς νά εὐνοοῦν τόν ἐσωτερικό κανονισμό;

Ἐπειδή ξέρουμε ὅτι τό *40%* τῶν φοιτητῶν τῆς τάξης εὐνοεῖ τόν ἐσωτερικό κανονισμό καί *60%* δέν τόν εὐνοεῖ οἱ *20* φοιτητές ἀποτελοῦν δύο ὁμάδες μέ $Np = 20\times0{,}4 = 8$ ἄτομα ἡ πρώτη καί $Nq = 20\times0{,}6 = 12$ ἄτομα ἡ δεύτερη. Ἔτσι ἡ κατανομή τοῦ ἀριθμοῦ X τῶν φοιτητῶν πού εὐνοοῦν τόν κανονισμό στό δεῖγμα τῶν *6* ἀτόμων εἶναι

$$p_X(x) = \frac{\binom{8}{x}\binom{12}{6-x}}{\binom{20}{6}}, \quad x = 0,1,2,\ldots,6 .$$

Ἀπό αὐτή τή σχέση ἔχουμε

α) $$P(X=2) = \frac{\binom{8}{2}\binom{12}{4}}{\binom{20}{6}} = 0{,}3576$$

β) P (τουλάχιστον δύο εὐνοοῦν τόν κανονισμό)

$= 1-P$(τό πολύ ἕνας εὐνοεῖ τόν κανονισμό)

$$= 1-P(X=0)-P(X=1) = 1- \frac{\binom{8}{2}\binom{12}{6}}{\binom{20}{6}} - \frac{\binom{8}{1}\binom{12}{5}}{\binom{20}{6}} = 0,8127 \; .$$

4. 4 Γεωμετρική κατανομή

Ἡ κατανομή αὐτή εἶναι ἡ κατανομή τῆς τυχαίας μεταβλητῆς τοῦ Παραδείγματος 3.5. Ἀναφέρεται στό διωνυμικό πείραμα ἤ τίς δοκιμές τοῦ *Bernoulli* ὅπου ἡ τ.μ. X τώρα ἀντιπροσωπεύει τόν ἀριθμό τῶν δοκιμῶν πού ἐκτελοῦνται μέχρις ὅτου ἐμφανιστεῖ ἡ πρώτη ἐπιτυχία E. Μέ ἄλλα λόγια ἡ γεωμετρική κατανομή εἶναι ἡ κατανομή τοῦ ἀριθμοῦ τῶν "δοκιμῶν ἀναμονῆς" ἤ "χρόνου ἀναμονῆς" μέχρις ὅτου ἐμφανιστεῖ ἡ πρώτη ἐπιτυχία σέ ἕνα διωνυμικό πείραμα.

Ἀπό τά ἐκτιθέμενα στό Παράδειγμα 3.5 εἶναι εὔκολο νά δοῦμε ὅτι ἡ σ.π. τῆς γεωμετρικῆς κατανομῆς δίνεται ἀπό τόν τύπο

$$p_X(x) = pq^{x-1} \; , \qquad x = 1,2,3,\ldots \; . \qquad (4.7)$$

Τό p, $0 \leq p \leq 1$, εἶναι ἡ πιθανότητα ἐπιτυχίας τῶν δοκιμῶν τοῦ *Bernoulli* καί ἀποτελεῖ τήν παράμετρο τῆς κατανομῆς.

Μία τ.μ. X μέ κατανομή τήν (4.7) λέγεται γεωμετρική ἤ τ.μ. τοῦ *Pascal* καί συμβολίζεται μέ

$$X \sim Geo(p) \; .$$

Τό ὄνομα ἀπορρέει ἀπό τό γεγονός ὅτι γιά διαδοχικές τιμές τοῦ x οἱ πιθανότητες (4.7) ἀποτελοῦν μία φθίνουσα γεωμετρική πρόοδο. Εἶναι δέ φανερό ὅτι

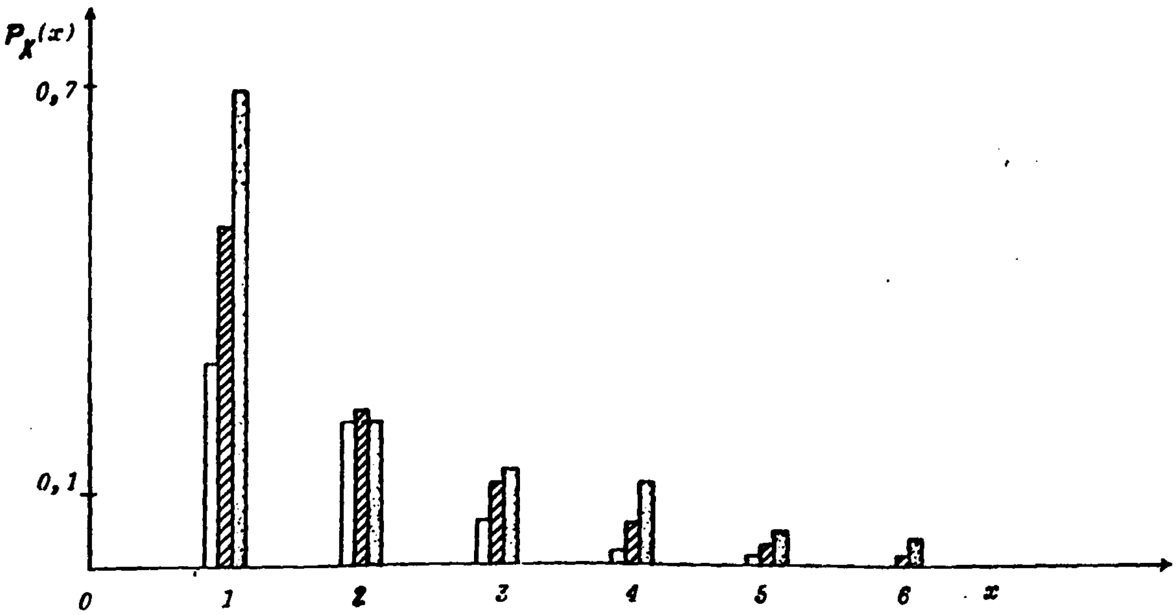
$$\sum_{x=1}^{\infty} pq^{x-1} = p/(1-q) = 1$$

γεγονός πού μαζί μέ τό μή ἀρνητικό τῶν τιμῶν τῆς (4.7) ἀποδεικνύει ὅτι ἡ ἀνωτέρω $p_X(x)$ ἀποτελεῖ μία σ.π.

Ἡ α.σ.κ. τῆς γεωμετρικῆς τ.μ. εἶναι

$$F(x) = \begin{cases} 0 & , \quad x < 1 \\ \sum_{y=1}^{[x]} pq^{y-1} & , \quad x \geq 1. \end{cases} \tag{4.8}$$

Τό Σχῆμα 4.3 δίνει τίς γραφικές παραστάσεις τῆς (4.7) γιά $p = 0,3$, $p = 0,5$, $p = 0,7$.

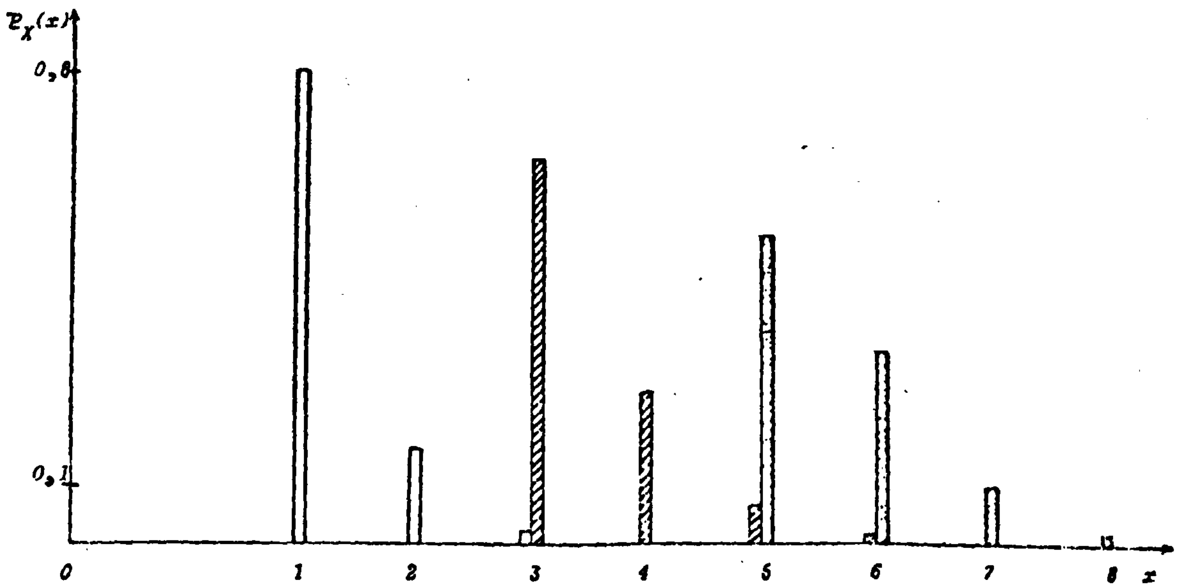


Κατανομή ▯ : $p = 0,3$,

Κατανομή ▨ : $p = 0,5$,

Κατανομή ▩ $p = 0,7$.

Σχῆμα 4.3: Γεωμετρικές κατανομές



Κατανομή : $p = 0,9$, $k = 2$

Κατανομή : $p = 0,9$, $k = 4$

Κατανομή : $p = 0,9$, $k = 6$

Σχῆμα 4.4: Ἀρνητικές διωνυμικές κατανομές

Οἱ ἐ φ α ρ μ ο γ έ ς τῆς γεωμετρικῆς κατανομῆς εἶναι εὐρύτατες. Σέ κάθε φαινόμενο πού μπορεῖ νά παρασταθεῖ ἤ προσεγγιστεῖ μέ μία σειρά δοκιμῶν τοῦ *Bernoulli*, ἡ γεωμετρική κατανομή ἀποτελεῖ τό μοντέλο πιθανοτήτων πού μπορεῖ νά υἱοθετηθεῖ γιά τό "χρόνο" ἀναμονῆς. Ἀναφέρουμε μερικές τυπικές περιπτώσεις: Σέ ἕνα παιχνίδι "κορώνα - γράμματα" ὅπου ἐνδιαφερόμαστε γιά τό χρόνο ἀναμονῆς μέχρις ὅτου ἐμφανιστεῖ ἡ πρώτη κορώνα. Ἕνας γιατρός ἐνδιαφέρεται γιά τήν πιθανότητα ὁ πέμπτος ἄρρωστος πού θά τόν ἐπισκεφτεῖ στό γραφεῖο του μία μέρα νά πάσχει ἀπό κάποιο ἰό γρίππης. Ὁ βιομήχανος ἐνδιαφέρεται γιά τή πιθανότητα τό ὄγδοο κατασκευαζόμενο ἀντικείμενο νά εἶναι ἐλαττωματικό κ.ο.κ.

Παράδειγμα 4.4:

"Ας ἀναφερθοῦμε στό Παράδειγμα 3.5 καί ἄς ὑποθέσουμε ὅτι ἡ πιθανότητα ἕνας ὑποψήφιος ὁδηγός νά πάρει τό δίπλωμα ὁδηγήσεως σέ μία ὁποιαδήποτε προσπάθεια εἶναι 60% (πιθανόν ὄχι τόσο ρεαλιστικό ποσοστό). Τότε ἡ πιθανότητα ὁ ὑποψήφιος νά πάρει τό δίπλωμα του στή πέμπτη προσπάθεια εἶναι

$$P(X=5) = 0{,}6(1-0{,}6)^4 \approx 0{,}015 .$$

4.5 Ἀρνητική διωνυμική κατανομή ἤ κατανομή τοῦ Pascal

Ἡ κατανομή αὐτή ἀποτελεῖ γενίκευση τῆς γεωμετρικῆς κατανομῆς. Θεωροῦμε καί πάλι μία σειρά δοκιμῶν τοῦ *Bernoulli* καί ἀντί νά ἐνδιαφερόμαστε γιά τόν ἀριθμό τῶν "δοκιμῶν ἀναμονῆς" μέχρι τήν πρώτη ἐπιτυχία ἐνδιαφερόμαστε τώρα γιά τόν ἀριθμό X τῶν "δοκιμῶν ἀναμονῆς" ("χρόνου ἀναμονῆς") μέχρις ὅτου ἐμφανιστεῖ ἡ k ἐπιτυχία ($k \geq 1$, ἀκέραιος). Εἶναι φανερό ὅτι οἱ δυνατές τιμές x τῆς τ.μ. X εἶναι

$$k, k+1, k+2, \ldots$$

"Εστω x σταθερό ($x \geq k$). Τότε τό ἐνδεχόμενο $\{X=x\}$ σημαίνει ὅτι $k-1$ ἐπιτυχίες ἔγιναν στίς $x-1$ πρῶτες δοκιμές καί ἡ τελευταία δοκιμή, δηλαδή ἡ δοκιμή x τάξης, κατέληξε σέ ἐπιτυχία. Φυσικά οἱ $k-1$ ἐπιτυχίες στίς $x-1$ πρῶτες δοκιμές ἀκολουθοῦν τή διωνυμική κατανομή. "Αρα βάσει τῆς ἀνεξαρτησίας θά ἔχουμε

$P(X=x) = P\{k-1$ E στίς πρῶτες $x-1$ δοκιμές$\}$ $P\{E$ στή x δοκιμή$\}$

$$= \binom{x-1}{k-1} p^{k-1} q^{x-1-(k-1)} p \quad .$$

Ἔτσι ἡ σ.π. τῆς **ἀρνητικῆς διωνυμικῆς κατανομῆς** δίνεται ἀπό τόν τύπο

$$p_X(x) = \binom{x-1}{k-1} p^k q^{x-k}, \qquad x = k,\ k+1, \ldots \qquad (4.9)$$

Τά k καί p, $0 \leq p \leq 1$, εἶναι καί πάλι οἱ **παράμετροι** τῆς κατανομῆς. Μία τ.μ. X μέ κατανομή τήν *(4.9)* λέγεται **ἀρνητική διωνυμική** ἤ τ.μ. τοῦ ***Pascal*** καί συμβολίζεται μέ

$$X \sim NB(k,p).$$

Θά ἀποδείξουμε τώρα ὅτι ἡ συνάρτηση $p_X(x)$ τῆς *(4.9)* εἶναι πράγματι μία σ.π. Παράλληλα θά δικαιολογήσουμε καί τό ὄνομα ἀρνητική διωνυμική. Στηριζόμαστε στήν ἐπέκταση τοῦ ὁρισμοῦ τοῦ διωνυμικοῦ συντελεστῆ $\binom{x}{r}$ γιά κάθε πραγματικό ἀριθμό x καί κάθε ἀκέραιο r [βλέπε §1.4 σχέση *(1.4)*] καί στόν διωνυμικό τύπο τοῦ Νεύτωνα *(1.5)*. Ἔχουμε

$$\sum_{x=k}^{\infty} p_X(x) = \sum_{x=k}^{\infty} \binom{x-1}{k-1} p^k q^{x-k} = p^k \sum_{x=k}^{\infty} \binom{x-1}{x-k} q^{x-k}$$

$$= p^k \{1+\binom{k}{1}q+\binom{k+1}{2}q^2+\binom{k+2}{3}q^3+ \ldots\} \qquad (4.10)$$

$$= p^k (1-q)^{-k} = p^k p^{-k} = 1 \quad ,$$

διότι ἡ σχέση *(1.4)* μᾶς δίνει

$$\binom{-k}{y}(-1) = \binom{y+k-1}{y} \qquad (4.11)$$

καί ἔτσι ὁ διωνυμικός τύπος τοῦ Νεύτωνα *(1.5)* γίνεται

$$(1-q)^{-k} = \sum_{y=0}^{\infty} \binom{-k}{y}(-q)^{y}$$

$$= \sum_{y=0}^{\infty} \binom{y+k-1}{y} q^{y}$$

$$= 1+\binom{k}{1}q+\binom{k+1}{2}q^{2}+\binom{k+2}{3}q^{3} + \ldots$$

'Επειδή δέ $p_X(x) \geq 0$ ἡ σχέση *(4.10)* ἀποδεικνύει ὅτι ἡ συνάρτηση $p_X(x)$ εἶναι πράγματι μία σ.π.

Τό ὄνομα "ἀρνητική διωνυμική" κατανομή ἀπορρέει ἀπό τό γεγονός ὅτι βάσει τῆς σχέσης *(4.10)* οἱ τιμές τῆς $p_X(x)$ γιά $x=k,\ k+1,\ k+2,\ldots$ εἶναι διαδοχικοί ὅροι τοῦ "ἀρνητικοῦ" διωνυμικοῦ ἀναπτύγματος τοῦ $p^{k}(1-q)^{-k}$. 'Επίσης καί ἀπό τό γεγονός ὅτι ἡ $p_X(x)$ μέ τή βοήθεια τῆς *(4.11)* μπορεῖ νά γραφεῖ ἰσοδύναμα

$$p_X(x) = \binom{-k}{x-k} p^{k}(-q)^{x-k} \quad , \ x = k,\ k+1,\ldots \qquad (4.12)$$

'Η α.σ.κ. τῆς ἀρνητικῆς διωνυμικῆς τ.μ. εἶναι:

$$F(x) = \begin{cases} 0 & , \ x < k \\ \sum_{n=k}^{[x]} \binom{y-1}{k-1} p^{k} q^{x-k} & , \ x \geq k \ . \end{cases} \qquad (4.13)$$

Τό Σχῆμα 4.4 στή σελίδα 135 δίνει γραφικές παραστάσεις τῆς

(4.9) γιά $p = 0,9$ καί $k = 2,4,6$.

Εἶναι εὔκολο τώρα νά δειχθεῖ ὅτι ἡ ἀρνητική διωνυμική κατανομή εἶναι ἕνα ἄθροισμα k ἀνεξάρτητων γεωμετρικῶν κατανομῶν. Ἄς θεωρήσουμε καί πάλι τίς δοκιμές τοῦ *Benoulli*. Ἔστω Y_1 ὁ ἀριθμός τῶν δοκιμῶν μέχρις ὅτου ἐμφανιστεῖ ἡ πρώτη ἐπιτυχία, Y_2 ὁ ἀριθμός τῶν δοκιμῶν ἀπό τήν πρώτη ἐπιτυχία μέχρις ὅτου ἐμφανιστεῖ ἡ δεύτερη ἐπιτυχία κ.ο.κ. καί τελικά ἔστω Y_k ὁ ἀριθμός τῶν δοκιμῶν ἀπό τήν $k-1$ ἐπιτυχία μέχρι τήν k ἐπιτυχία. Τότε ἐάν $X \sim NB(k,p)$ βάσει τῶν ὁρισμῶν ἔχουμε

$$X = Y_1+Y_2+\ldots+Y_k\,. \qquad (4.14)$$

Ὁ ὑπολογισμός τῶν τιμῶν τῶν πιθανοτήτων μίας ἀρνητικῆς διωνυμικῆς κατανομῆς μπορεῖ νά γίνει εὔκολα μέ τή βοήθεια τῆς διωνυμικῆς κατανομῆς καί τῆς σχέσης

$$p_X(x) = \frac{k}{x}\,p_Y(k) \qquad (4.15)$$

ὅπου $X \sim NB(k,p)$ καί $Y \sim B(x,p)$.

Γιά $k=1$ ἡ ἀρνητική διωνυμική κατανομή γίνεται ἡ γεωμετρική κατανομή. Οἱ ἐφαρμογές αὐτῆς εἶναι ἀνάλογες μέ τίς ἐφαρμογές τῆς γεωμετρικῆς κατανομῆς.

<u>Παράδειγμα</u> <u>4.5</u>:

Ἔστω ὅτι ἡ πιθανότητα ἕνα νήπιο ἐκτιθέμενο σέ μία μεταδοτική νόσο νά προσβληθεῖ ἀπό αὐτή εἶναι 20%. Τότε ή πιθανότητα τό δωδέκατο νήπιο πού έκτίθεται στή νόσο νά εἶναι τό τρίτο πού θά προσβληθεῖ ἀπό αὐτή εὑρίσκεται ὑπολογίζοντας τό $p_X(12)$ ὅπου $X \sim NB(3,\ 0,2)$, δηλαδή

$$p_X(12) = \binom{11}{2} 0,2^3 (1-0,2)^9 \approx 0,059 \ .$$

4.6 Κατανομή τοῦ Poisson

Ἄς θεωρήσουμε μία ραδιενεργό οὐσία ἡ ὁποία μέ τή πάροδο τοῦ χρόνου ἐκπέμπει σωματίδια ἑνός ὡρισμένου τύπου. Θέτουμε τήν οὐσία αὐτή ὑπό παρακολούθηση γιά ἕνα μοναδιαῖο χρονικό διάστημα π.χ. ἕνα λεπτό, μία ὥρα, κλπ. καί ἐνδιαφερόμαστε γιά τόν ἀριθμό X τῶν σωματιδίων πού ἐκπέμπονται στό διάστημα αὐτό. Ὁ ἀριθμός X εἶναι προφανῶς μία τ.μ. Θέλουμε νά εὕρουμε τήν κατανομή τῆς X.

Εἶναι φανερό ὅτι οἱ δυνατές τιμές τῆς τ.μ. X εἶναι $0,1,2,\ldots,\infty$ (τοὐλάχιστον θεωρητικά). Γιά νά εὕρουμε τή κατανομή τῆς X θά χρειαστεῖ νά κάνουμε ὡρισμένες ὑποθέσεις οἱ ὁποῖες, φυσικά, θά προδικάζουν τήν ὀρθότητα τοῦ μαθηματικοῦ μας συλλογισμοῦ: κατ'ἀρχήν δεχόμαστε ὅτι οἱ συνθῆκες τοῦ πειράματος παραμένουν σταθερές μέ τή πάροδο τοῦ χρόνου. Στή συνέχεια ὑποδιαιροῦμε τό μοναδιαῖο χρονικό διάστημα σέ ἕνα μεγάλο ἀριθμό n μικρῶν ἴσων ὑποδιαστημάτων κατά τέτοιο τρόπο ὥστε τό πολύ ἕ ν α σ ω μ α τ ί δ ι ο νά δύναται νά ἐκπεμφθεῖ σέ κάθε ὑποδιάστημα μέ θετική πιθανότητα. Ἔστω p ἡ πιθανότητα αὐτή καί ἔστω ἐπί πλέον ὅτι εἶναι σταθερή γιά κάθε ὑποδιάστημα. Εἶναι λογικό νά δεχτοῦμε ὅτι ὅσο μικρότερο εἶναι τό ὑποδιάστημα τόσο μικρότερο εἶναι τό p. Ἔστω ἐπίσης ὅτι ἡ ἐκπομπή τῶν σωματιδίων εἶναι ἀνεξάρτητη ἀπό ὑποδιάστημα σέ ὑποδιάστημα. Τότε γιά λόγους πρακτικούς μποροῦμε νά ποῦμε ὅτι ἡ ἐκπομπή τῶν σωματιδίων μοιάζει μέ μία σειρά δοκιμῶν τοῦ *Bernoulli* μέ ἐπιτυχία E τήν ἐκπομπή ἑνός σωματιδίου καί σταθερό τό p. Ἔτσι μία κατά προσέγγιση τιμή τῆς $P(X=x)$ εἶναι ἡ $\binom{n}{x}p^x q^{n-x}$.

Ἔστω τώρα ὅτι ὁ ἀριθμός n τῶν ὑποδιαστημάτων τείνει στό ἄπειρο. Καθώς τό n τείνει στό ἄπειρο τό p τείνει στό μηδέν. Ἄς δεχτοῦμε λοιπόν ὅτι ἡ ποσότητα $\lambda = np$ παραμένει σταθερή. Τότε

$$\begin{aligned} P(X=x) &= \lim_{n\to\infty} \binom{n}{x} p^x q^{n-x} \\ &= \lim_{n\to\infty} \frac{n(n-1)\dots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \\ &= \frac{\lambda^x}{x!} \lim_{n\to\infty} \frac{n(n-1)\dots(n-x+1)}{n\,n\,\dots\,n} \left(1-\frac{\lambda}{n}\right)^{n-x} \\ &= \frac{\lambda^x}{x!} e^{-\lambda}, \end{aligned}$$

διότι

$$\lim_{n\to\infty} (1-\lambda/n)^{-x} = 1 \quad \text{καί} \quad \lim_{n\to\infty} (1-\lambda/n)^{n} = e^{-\lambda} .$$

Ἡ παραπάνω σύγκλιση εἶναι ὁμοιόμορφη ὡς πρός x.

Ἡ συνάρτηση

$$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0,1,2,\dots, \quad \text{καί} \quad \lambda > 0,$$

είναι πράγματι μία σ.π. διότι είναι μή άρνητική καί έπί πλέον

$$\sum_{x=0}^{\infty} p_X(x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1.$$

Ἡ *(4.16)* δίνει τήν κατανομή τῆς τ.μ. X τοῦ άνωτέρω παραδείγματος καί άποτελεῖ τή περίφημη κατανομή τοῦ *Poisson* μέ παράμετρο λ. Ὁ *S.Poisson* (1781-1840) ἦτο διάσημος Γάλλος μαθηματικός, πού άνεκάλυψε τήν παραπάνω κατανομή ὡς τό ὅριο τῆς διωνυμικῆς κατανομῆς καθώς τό $n\to\infty$ μέ $\lambda = np$ σταθερό. Ἀργότερα ή κατανομή αὐτή μελετήθηκε μέσα στό γενικώτερο πλαίσιο τῆς θεωρίας τῶν Στοχαστικῶν Διαδικασιῶν. Ἡ φυσική σημασία τῆς παραμέτρου λ θά καταφανεῖ στό Κεφάλαιο 5. Ἡ τ.μ. X πού ἔχει τήν κατανομή (4.16) λέγεται τ.μ. τοῦ *Poisson* μέ παράμετρο λ καί συμβολίζεται μέ

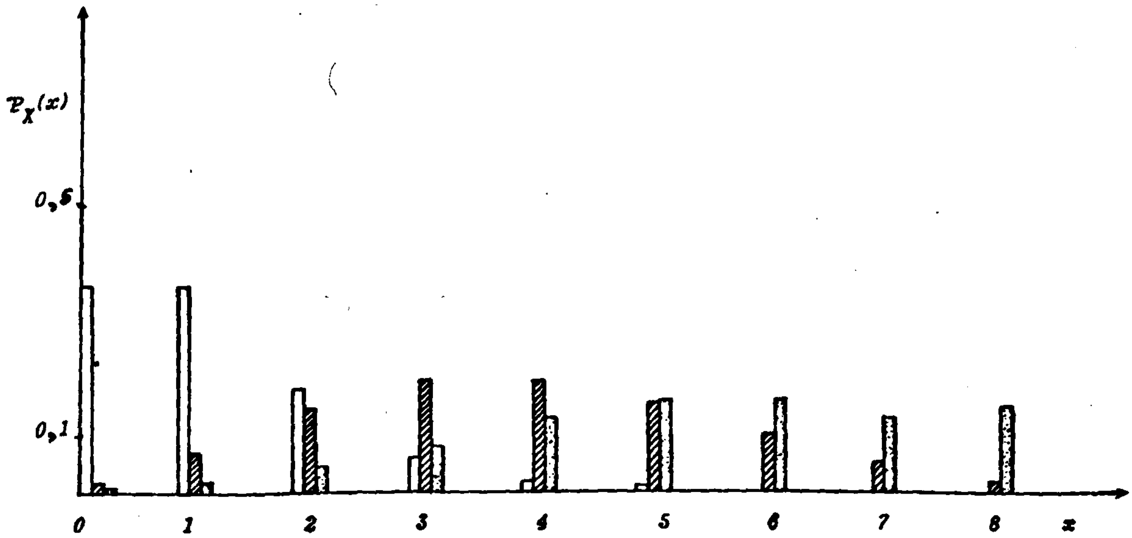
$$X \sim P(\lambda) .$$

Ἡ α.σ.κ. τῆς (4.16) δίνεται άπό τήν

$$F(x) = \begin{cases} 0 & \text{ἐάν } x < 0 \\ \sum_{y=0}^{[x]} \dfrac{e^{-\lambda}\lambda^y}{y!} & \text{ἐάν } x \geq 0 . \end{cases}$$

Ἀντιπροσωπευτικές γραφικές παραστάσεις τῆς σ.π. τῆς κατανομῆς τοῦ *Poisson* μέ $\lambda = 1{,}0$, $\lambda = 4{,}0$ καί $\lambda = 6{,}0$ δίνονται στό Σχῆμα 4.5.

Γιά διευκόλυνση τῶν ὑπολογισμῶν άλλά καί γιά λόγους



Κατανομή ▯ : λ = $1,0$

Κατανομή ▨ : λ = $4,0$

Κατανομή ▯ : λ = $6,0$

Σχῆμα 4.5: Κατανομές τοῦ *Poisson*

στατιστικῆς ἐπαγωγῆς ἡ βιβλιογραφία προσφέρει πίνακες τῆς κατανομῆς τοῦ *Poisson* οἱ ὁποῖοι δίνουν τίς τιμές τῆς σ.π. (4.16) γιά διάφορες τιμές τοῦ λ. Ἕνας τέτοιος πίνακας εἶναι καί ὁ Πίνακας II τοῦ Παραρτήματος. Ἄλλοι πίνακες δίνουν τίς τιμές τῆς α.σ.κ. (4.17) γιά διάφορες τιμές τοῦ λ.

Συχνά ἡ κατανομή τοῦ *Poisson* χρησιμοποιεῖται γιά τόν κατά προσέγγιση ὑπολογισμό τῶν πιθανοτήτων τῆς διωνυμικῆς κατανομῆς. Ὁ Πίνακας 4.1 πού παραθέτουμε συγκρίνει τίς τιμές τῆς διωνυμικῆς κατανομῆς μέ $n = 20$ καί $p = 0,1$

μέ τίς τιμές τῆς κατανομῆς τοῦ *Poisson* μέ $\lambda = np =$
$= 20 \times 0 \times 0,1 = 2$.

	Πιθανότητες				
x	Διωνυμικές	*Poisson*	x	Διωνυμικές	*Poisson*
0	0,1216	0,1353	7	0,0020	0,0034
1	0,2702	0,2707	8	0,0004	0,0009
2	0,2852	0,2707	9	0,0001	0,0002
3	0,1901	0,1804	10	0,0000	0,0000
4	0,0898	0,0902	11	0,0000	0,0000
5	0,0319	0,0361	12	0,0000	0,0000
6	0,0089	0,0120			

Πίνακας 4.1 Σύγκριση τῆς διωνυμικῆς κατανομῆς $B(n = 20,\ p = 0,1)$ μέ τήν κατανομή τοῦ *Poisson* $P(\lambda = 2)$

Ὁ πίνακας δείχνει ὅτι ἡ προσέγγιση εἶναι ἱκανοποιητική. Ἡ προσέγγιση εἶναι ἰδιαίτερα ἱκανοποιητική ὅταν τό n εἶναι μεγάλο καί τό p μικρό. Κατά κανόνα ἡ κατανομή τοῦ *Poisson* δίνει ἱκανοποιητική προσέγγιση στίς διωνυμικές πιθανότητες ὅταν $n \geq 20$ καί $p \leq 0,05$ ἤ ὅταν $n \geq 100$ καί $p \leq 0,1$(ἤ $\lambda \leq 10$).

Ἡ κατανομή τοῦ *Poisson* εἶναι μία ἀπό τίς πιό βασικές κατανομές τῆς θεωρίας Πιθανοτήτων καί τῆς Στατιστικῆς. Ἔχει μεγάλη γενικότητα καί τεράστιο εὖρος ἐφαρμογῶν. Παρόμοιες κατανομές εἶναι ἡ διωνυμική καί ἡ κανονική πού θά μελετήσουμε πιό κάτω.

Ἡ κατανομή τοῦ *Poisson* ἐ φ α ρ μ ό ζ ε τ α ι ὅταν ἔχουμε μία σειρά τ υ χ α ί ω ν ἐ ν δ ε χ ο μ έ ν ω ν τ ά ὁ π ο ῖ α γ ί ν ο ν τ α ι μ έ τ ή π ά ρ ο δ ο τ ο ῦ χ ρ ό ν ο υ καί γιά τά ὁποῖα ἰσχύουν οἱ ὑποθέσεις πού ἀναφέραμε στήν ἀρχή τοῦ παρόντος ἐδαφίου. Συγκεκριμένα εἶναι ἀπαραίτητο οἱ συνθῆκες τοῦ πειράματος ἤ φαινομένου νά παραμένουν σταθερές μέ τή πάροδο τοῦ χρόνου καί τά χρονικά διαστήματα τά ὁποῖα δέν ἐπικαλύπτονται νά εἶναι στατιστικῶς ἀνεξάρτητα μέ τήν ἔννοια ὅτι ἡ πληροφορία σχετικά μέ τόν ἀριθμό τῶν τυχαίων ἐνδεχομένων σέ ἕνα χρονικό διάστημα νά μήν ἀποκαλύπτει τίποτε γιά τό ἄλλο διάστημα. Ἐπίσης θά πρέπει ἡ πιθανότητα νά συμβεῖ ἕνα τυχαῖο γεγονός σέ ἕνα μικρό διάστημα νά εἶναι ἀνάλογος τοῦ μήκους τοῦ διαστήματος καί ἡ πιθανότητα νά συμβοῦν δύο ἤ περισσότερα τυχαῖα ἐνδεχόμενα σέ ἕνα πολύ μικρό διάστημα νά εἶναι ἀμελητέα. Ἡ τελευταία αὐτή ὑπόθεση συμφωνεῖ μέ τή διαισθητική εἰκόνα πού ἔχουμε σχετικά μέ μ ε μ ο ν ο μ έ ν α τ υ χ α ῖ α ἐ ν δ ε χ ό μ ε ν α. Ἀποκλείει ὅμως τήν περίπτωση τά ἐνδεχόμενα μας νά ἐμφανίζονται ἀνά ζεύγη. Γιά παράδειγμα ἀναφέρουμε τή περίπτωση τῶν τροχαίων ἀτυχημάτων. Ἡ πιθανότητα νά συμβοῦν δύο ἀτυχήματα σέ ἕνα μικρό χρονικό διάστημα εἶναι ἀμελητέα σέ σύγκριση μέ τήν πιθανότητα νά συμβεῖ ἕνα ἀτύχημα. Ἕνα ἀτύχημα ὅμως ἐνδέχεται νά περιλαμβάνει δύο αὐτοκίνητα καί ἔτσι ἐάν μέ τυχαῖο ἐνδεχόμενο ἐννοοῦμε "καταστροφή ἑνός αὐτοκινήτου" τότε τά τυχαῖα ἐνδεχόμενα ἐνδέχεται νά ἐμφανιστοῦν ἀνά ζεύγη καί οἱ βασικές ὑποθέσεις δέν ἰσχύουν.

Διαδικασία τοῦ Poisson

Ἄς ἐπανέλθουμε τώρα στό ἀρχικό παράδειγμα τῆς ραδιενεργοῦ οὐσίας καί θεωρήσουμε ἀντί τοῦ μοναδιαίου χρονικοῦ διαστήματος ἕνα οἱοδήποτε χρονικό διάστημα μήκους t. Ἔστω $X(t)$ ὁ ἀριθμός τῶν σωματιδίων πού ἐκπέμπονται στό διάστημα $(0,t)$. Ζητᾶμε νά προσδιορίσουμε τή κατανομή $p(x,t)$ τῆς τ.μ. $X(t)$ γιά σταθερό t.

Σκεπτόμενοι μέ τόν ἴδιο τρόπο ὅπως καί προηγομένως ὑποδιαιροῦμε τό διάστημα $(0,t)$ σέ $n = t/\Delta t$ ὑποδιαστήματα μήκους Δt καί δεχόμαστε τίς ἴδιες ὑποθέσεις. Ἰδιαίτερα δεχόμαστε ὅτι ἔχουμε πάλι δοκιμές τοῦ *Bernoulli* μέ πιθανότητα ἐπιτυχίας

$$p = P\{\text{μία } E \text{ σέ ἕνα οἱοδήποτε ὑποδιάστημα μήκους } \Delta t\} = \lambda\Delta t$$

Τότε παίρνοντας τό ὅριο τῆς διωνυμικῆς κατανομῆς μέ $\Delta t \to 0$ (ἤ $n \to \infty$) εὑρίσκουμε κατά τόν ἴδιο τρόπο τήν κατανομή τοῦ *Poisson* μέ παράμετρο λt.

$$p(x,t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} \quad , \qquad x = 0,1,2,\ldots \qquad (4.18)$$

Ἔστω τώρα ὅτι τό t μεταβάλλεται καί ὅτι γιά κάθε t οἱ ἴδιες ἀρχικές ὑποθέσεις ἰσχύουν. Τότε ἡ (4.18) δίνει τή σ.π. τῆς τ.μ. $X(t)$ γιά κάθε t.

Ἡ οἰκογένεια τῶν τ.μ. (ἤ ἡ τυχαία συνάρτηση) $X(t)$ μέ $t > 0$ καί σ.π. τήν $p(x,t)$ τῆς (4.18) λέγεται **στοχαστική διαδικασία τοῦ *Poisson*** ἤ ἁπλῶς **διαδικασία τοῦ *Poisson***. Οἱ ὑποθέσεις πού ἔχουμε δεχτεῖ εἶναι γνωστές καί ὡς ἀ ξ ι ώ μ α τ α τ ῆ ς δ ι α δ ι κ α σ ί α ς τ ο ῦ *P o i s s o n* καί μποροῦν νά διατυπωθοῦν ὡς ἑξῆς:

'<u>Αξίωμα</u> <u>1</u>:

$$P\{1\ E \text{ στό διάστημα } (t, t+\Delta t)\} = \lambda \Delta t \quad \text{γιά κάθε } t$$

(ἰδιότητα στατικότητας)

'<u>Αξίωμα</u> <u>2</u>:

$$P\{2 \text{ ἤ περισσότερες } E \text{ στό διάστημα} (t, t+\Delta t)\} = \text{ἀμελητέα}$$

'<u>Αξίωμα</u> <u>3</u>:

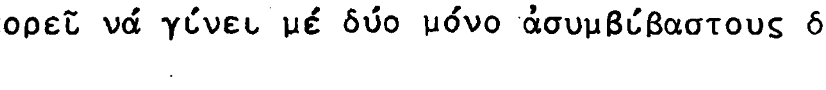
$$P\{n_1 E \text{ στό διάστημα } (t_1, t_1+\Delta t_1) \text{ καί } n_2\ E \text{ στό διάστημα } (t_2, t_2+\Delta t_2),\ \ t_1 < t_2\}$$

$$= P\{n_1\ E \text{ στό } (t_1, t_1+\Delta t_1)\} P\{n_2\ E \text{ στό } (t_2, t_2+\Delta t_2)\}$$

(ἰδιότητα ἀνεξαρτησίας)

Μέ βάση τά παραπάνω ἀξιώματα μποροῦμε νά ξαναβροῦμε τήν $p(x,t)$ τῆς (4.18) μέ τρόπο διάφορετικό ἀπό τόν προηγούμενο. Ὁ τρόπος αὐτός χρησιμοποιεῖται συχνά στή Θεωρία τῶν Στοχαστικῶν Διαδικασιῶν. Ἔστω ὅτι x τυχαῖα ἐνδεχόμενα παρατηροῦνται στό διάστημα $(0, t+\Delta t)$. Ἡ ἀντίστοιχη πιθανότητα εἶναι

$$p(x, t+\Delta t).$$



Βάσει τοῦ δεύτερου ἀξιώματος τό παραπάνω σύνθετο ἐνδεχόμενο μπορεῖ νά γίνει μέ δύο μόνο ἀσυμβίβαστους δυνατούς τρόπους: x τυχαῖα ἐνδεχόμενα παρατηροῦνται στό διάστημα $(0,t)$ καί κανένα στό $(t,t+\Delta t)$ ἤ $x-1$ τυχαῖα ἐνδεχόμενα στό διάστημα $(0,t)$ καί 1 στό $(t,t+\Delta t)$. Κατά συνέπεια ἔχουμε:

$$\begin{aligned} p(x,t+\Delta t) &= P[X(t+\Delta t) = x] \\ &= P\{x \text{ } E \text{ στό } (0,t) \text{ καί } 0 \text{ } E \text{ στό } (t,t+\Delta t)\} \\ &\quad + P\{x-1 \text{ } E \text{ στό } (0,t) \text{ καί } 1 \text{ } E \text{ στό } (t,t+\Delta t)\} \\ &= P\{x \text{ } E \text{ στό } (0,t)\}P\{0 \text{ } E \text{ στό } (t,t+\Delta t)\} \\ &\quad + P\{x-1 \text{ } E \text{ στό } (0,t)\}P\{1 \text{ } E \text{ στό } (t,t+\Delta t)\} \\ &\qquad\qquad \text{(λόγω ἀνεξαρτησίας)} \\ &= p(x,t)(1-\lambda\Delta t)+p(x-1,t)\lambda\Delta t \end{aligned}$$

διότι

$$P\{0 \text{ } E \text{ στό } (t,t+\Delta t)\} = 1-\lambda\Delta t$$

καί

$$P\{1 \text{ } E \text{ στό } (t,t+\Delta t)\} = \lambda\Delta t.$$

Μετά ἀπό πράξεις βρίσκουμε

$$\frac{p(x,t+\Delta t)-p(x,t)}{\Delta t} = -\lambda p(1,t)+p(x-1,t)$$

ἤ παίρνοντες τό ὅριο γιά $\Delta t \to 0$

$$\frac{\partial p(x,t)}{\partial t} = -\lambda p(x,t)+p(x-1,t) \cdot$$

Ἡ παραπάνω διαφορική ἐξίσωση ἰσχύει γιά $x=0,1,2,\ldots$

καί ἐπιλύεται ὡς ἀκολούθως:

Γιά $x = 0$ ἔχουμε τήν ὁ μ ο γ ε ν ή δ ι α φ ο ρ ι κ ή ἐ ξ ί σ ω σ η

$$\frac{\partial p(0,t)}{\partial t} = -\lambda p(0,t)$$

δεδομένου ὅτι $p(-1,t) = 0$ γιά κάθε t. Μέ τή βοήθεια τῆς ἀρχικῆς συνθήκης $p(0,0) = 1$ βρίσκουμε τή λύση

$$p(0,t) = e^{-\lambda t} .$$

Γιά $x = 1$ παίρνουμε τή γ ρ α μ μ ι κ ή δ ι α φ ο ρ ι κ ή ἐ ξ ί σ ω σ η

$$\frac{\partial p(1,t)}{\partial t} + \lambda p(1,t) = e^{-\lambda t}$$

μέ λύση τήν

$$p(1,t) = \lambda t e^{-\lambda t}$$

δεδομένου ὅτι $p(1,0) = 0$.

Γιά $x = 2$ προκύπτει ἡ γραμμική διαφορική ἐξίσωση

$$\frac{\partial p(2,t)}{\partial t} + \lambda p(2,t) = \lambda^2 t e^{-\lambda t}/2 .$$

Τελικά μέ τή χρήση τῆς ἐπαγωγῆς βρίσκουμε

$$p(x,t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} , \quad x = 0,1,2,\ldots , \quad t > 0, \quad \lambda > 0 \blacktriangledown$$

Τυχαῖα ἐνδεχόμενα τά ὁποῖα ἀκολουθοῦν τίς ὑποθέσεις τῆς κατανομῆς τοῦ *Poisson* λέγονται συχνά "σ π ά ν ι α ἐ ν δ ε χ ό μ ε ν α". Ὁ ὅρος "σπάνια" ἐνδεχόμενα δέν είναι πετυχημένος καί είναι άσχετος μέ τή χρήση ή τό θεμελιώδη ρόλο τής κατανομής τού *Poisson*. 'Εκτός άπό τήν περίπτωση τών ραδιενεργών διασπάσεων καί τών τροχαίων άτυχημάτων ή κατανομή τού *Poisson* έφαρμόζεται καί στίς τηλεφωνικές κλήσεις πού φτάνουν σ'ένα τηλεφωνικό κέντρο κατά τή διάρκεια μίας χρονικής περιόδου. Στή Βιολογία ό άριθμός $X(k)$ τών κυττάρων μέ k άκριβώς χρωματοσωμικές άλλαγές άκολουθεί τήν κατανομή τού *Poisson*. Χρωματοσωμικές άλλαγές είναι οί διαδικασίες πού διενεργούν σέ όργανικά κύτταρα οί άκτίνες X. 'Ο άριθμός τών αύτοκινήτων (ή πελατών) πού φτάνουν σέ ένα σταθμό διοδίων (σέ ένα *super market*) σέ μία χρονική περίοδο άκολουθεί τήν κατανομή τού *Poisson*.

Οί διάφορες έφαρμογές τής κατανομής (διαδικασίας) τού *Poisson* έχουν έπηρεάσει καί τήν όρολογία της. Συχνά τά τυχαία ένδεχόμενα άναφέρονται καί ώς "ά φ ί ξ ε ι ς ἐ ν δ ε χ ο μ έ ν ω ν" καί τό λ όρίζεται ώς ό "ρ υ θ μ ό ς τ ώ ν ά φ ί ξ ε ω ν".

<u>Παράδειγμα 4.6:</u>

Τό τηλεφωνικό κέντρο ένός Σταθμού Α' Βοηθειών δέχεται τηλεφωνήματα έκτάκτου άνάγκης πού άκολουθούν τή κατανομή τού *Poisson* μέ ρυθμό άφίξεων *2* τηλεφωνήματα άνά *30* λεπτά κατά μέσον όρο. Νά βρεθεί ή πιθανότητα τό τηλεφωνικό κέντρο νά δεχτεί τουλάχιστον *3* τέτοια τηλεφωνήματα άπό τίς *12* τό μεσημέρι μέχρι τής *2* μ.μ.

Έστω X ό άριθμός τών τηλεφωνημάτων έκτάκτου άνάγκης άπό *12-2* μ.μ. Τότε ή τ.μ. X άκολουθεί κατανομή τού

Poisson. Ἄν ὡς μονάδα τοῦ χρόνου ληφθεῖ ἡ ὥρα, τότε $\lambda = 4$ καί συνεπῶς $\lambda t = 4\times 2 = 8$.

Ἡ πιθανότητα πού ζητᾶμε εἶναι:

$$\begin{aligned} P(X \geq 3) &= 1-P(X \leq 2) \\ &= 1-P(X=0)-P(X=1)-P(X=2) \\ &= 1-\frac{e^{-8}8^0}{0!} - \frac{e^{-8}8^1}{1!} - \frac{e^{-8}8^2}{2!} \\ &\approx 0{,}9862 \ . \end{aligned}$$

Ἡ κατανομή (ἤ ἡ διαδικασία τοῦ *Poisson*) μᾶς δίνει τή κατανομή τυχαίων ἐνδεχομένων ἤ σημείων κατά μῆκος τοῦ ἄξονα τοῦ χρόνου. Ἡ ἴδια κατανομή μπορεῖ νά ἐφαρμοστεῖ καί γιά νά περιγράψουμε τή κατανομή τυχαίων σημείων στό ἐπίπεδο ἤ τό χῶρο τῶν τριῶν διαστάσεων καί γενικώτερα τό χῶρο τῶν k διαστάσεων. Ἀντί γιά διαστήματα μήκους t ἔχουμε περιοχές μέ ἐμβαδό ἤ ὄγκο t καί ἡ βασική ὑπόθεση εἶναι ὅτι ἡ πιθανότητα νά βροῦμε x σημεῖα σέ μία ὁποιαδήποτε περιοχή ἐξαρτᾶται μόνο ἀπό τό ἐμβαδόν ἤ τόν ὄγκο τῆς περιοχῆς καί ὄχι ἀπό τό σχῆμα της. Οἱ ἴδιες ὑποθέσεις ἰσχύουν ὅπως καί πρίν:

(1) ἐάν τό t εἶναι μικρό, ἡ πιθανότητα νά βροῦμε περισσότερα ἀπό ἕνα σημεῖα τῆς περιοχῆς ἐμβαδοῦ ἤ ὄγκου t εἶναι μικρή σέ σύγκριση μέ τό t.

(2) μή ἐπικαλυπτόμενες περιοχές εἶναι στατιστικῶς ἀνεξάρτητες.

'Ακολουθώντας τήν ἴδια μέθοδο βρίσκουμε καί πάλι τήν κατανομή τοῦ *Poisson* μέ παράμετρο λt. 'Αστέρες στό διάστημα, σταφίδες σέ ἕνα γλυκό, σπόροι ζιζανίων ἀνεμειγμένοι μέ σπόρους χλόης, ἐλαττώματα σέ ὑλικά, βακτήρια σέ ἕνα ὑγρό ἤ ἕνα δίσκο μικροσκοπίου κατανέμονται σύμφωνα μέ τό μοντέλο πιθανότητας τοῦ *Poisson*.

Παραδείγματα:

4.7: 'Ο μέσος ὅρος τῶν κηλίδων πού ἐμφανίζονται σέ κάποιο τύπο ταινίας εἶναι μία ἀνά *2000* μέτρα.Μέ τήν ὑπόθεση ὅτι ὁ ἀριθμός τῶν κηλίδων ἀκολουθεῖ τή κατανομή τοῦ *Poisson*, νά βρεθεῖ ἡ πιθανότητα σέ *5000* μέτρα ταινίας α) νά μήν ὑπάρχουν κηλίδες β) νά ὑπάρχουν τό πολύ δύο κηλίδες καί γ) νά ὑπάρχουν ἀκριβῶς δύο κηλίδες.

Ὅταν σάν μονάδα μήκους ληφθεῖ τό μέτρο τότε ὁ μέσος ὅρος λ τῶν κηλίδων ἀνά μέτρο εἶναι $\lambda = 1/2000 = 0,0005$. Γιά τά *5000* μέτρα ἔχουμε $\lambda t = 0,0005 \times 5000 = 2,5$ καί ἔτσι

$$P(X=x) = \frac{2,5^x e^{-2,5}}{x!}, \quad x = 0,1,2,\ldots$$

α) $$P(X=0) = \frac{2,5^0 e^{-2,5}}{0!} \simeq 0,0821$$

β) $$P(X \leq 2) = P(X=0)+P(X=1)+P(X=2)$$

$$= \frac{e^{-2,5}2,5^0}{0!} + \frac{e^{-2,5}2,5^1}{1!} + \frac{e^{-2,5}2,5^2}{2!}$$

$$\simeq 0,5438\ .$$

γ) P(τουλάχιστο δύο κηλίδες) $= P(X \geq 2) = 1-P(X \leq 1)$

$$= 1- \frac{e^{-2,5}2,5^0}{0!} - \frac{e^{-2,5}2,5^1}{1!}$$

$$\approx 0,7125 .$$

4.8: Στόν Πίνακα 4.2 δίνουμε τά ἀποτελέσματα τῶν φημισμένων πειραμάτων τῶν *Rutherford* καί *Geiger*, οἱ ὁποῖοι παρετήρησαν τούς ἀριθμούς τῶν σωματιδίων a πού ἐκπέμπονται ἀπό μία ραδιενεργό οὐσία σέ $n = 2608$ χρονικές πειρόδους τῶν $7,5$ δευτερολέπτων ἡ κάθε μία. Στόν πίναqκα αὐτό f_x παριστάνει τόν ἀριθμό τῶν περιόδων (συχνοτήτα) στίς ὁποῖες ὁ ἀριθμός τῶν ἐκπεμπομένων σωματιδίων ἦταν ἀκριβῶς x. Ὁ μέσος ἀριθμός λ τῶν ἐκπεμπομένων σωματιδίων σέ μία χρονική περίοδο τῶν $7,5$ δευτερολέπτων εἶναι.

$$\lambda = \frac{\sum_{x=0}^{10} f_x x}{n} = \frac{10086}{2608} = 3,87 .$$

Ἄν ἡ κατανομή τοῦ ἀριθμοῦ τῶν ἐκπεμπομένων σωματιδίων εἶναι *Poisson* τότε

$$P(X=x) = e^{-3,87} \frac{3,87^x}{x!} .$$

Οἱ *Rutherford* καί *Geiger* ὑπελόγισαν τίς πιθανότητες $P(X=x)$ ἐκπομπῆς x σωματιδίων σέ μία χρονική περίοδο καί στή συνέχεια τίς ἀναμενόμενες συχνότητες $nP(X=x)$ πού δίνονται στήν τρίτη στήλη τοῦ Πίνακα 4.2. Βλέπουμε ὅτι οἱ θεωρητικές τιμές (συχνότητες) $nP(X=x)$ συμπίπτουν

x	f_x	$nP(X=x)$
0	57	54,399
1	203	210,523
2	383	407,361
3	525	525,496
4	532	508,418
5	408	393,515
6	273	253,817
7	139	140,325
8	45	67,882
9	27	29,189
$x\geq 10$	16	17,075

Πίνακας 4.2 Πείραμα *Rutherford - Geiger*

κατά πολύ μέ τίς συχνότητες f_x πού πραγματοποιήθηκαν (δεύτερη στήλη τοῦ πίνακα). Ἔτσι μποροῦμε νά ποῦμε ὅτι τό μοντέλο τοῦ *Poisson* περιγράφει ἱκανοποιητικά τό φαινόμενο τῆς ἐκπομπῆς σωματιδίων α ἀπό μία ραδιενεργό οὐσία. Βέβαια πρέπει νά ἐκτιμήσουμε τό μέγεθος τῶν διακυμάνσεων πού ὀφείλεται στήν τύχη γιά νά κρίνουμε τήν καταλληλότητα τῆς προσαρμογῆς *(goodness of fit)*. Ὁ στατιστικός κρίνει τήν καταλληλότητα τῆς προσαρμογῆς μέ τό τέστ X^2 πού θά μελετήσουμε σέ ἑπόμενο κεφάλαιο.

Β΄ ΕΙΔΙΚΕΣ ΣΥΝΕΧΕΙΣ ΚΑΤΑΝΟΜΕΣ

4. 7 Ὁμοιόμορφη κατανομή

Ἡ ἁπλούστερη ἀπό τίς συνεχεῖς κατανομές, ἡ ὁμοιόμορφη κατανομή ἔχει παρουσιαστεῖ προηγουμένως στό πλαίσιο τοῦ παραδείγματος τῆς πυξίδας (βλ.§3.3).Εἶναι ἡ κατανομή πού ἐκφράζει τό ἰσοπίθανο (ὁμοιόμορφο) τῶν ἀποτελεσμάτων ἑνός πειράματος. Ἐμφανίζεται τόσο στή διακριτή ὅσο καί τή συνεχή περίπτωση. Στό ἐδάφιο αὐτό θά παρουσιάσουμε τή συνεχή ὁμοιόμορφο κατανομή.

Ἔστω X μία τ.μ. μέ σ.π.π.

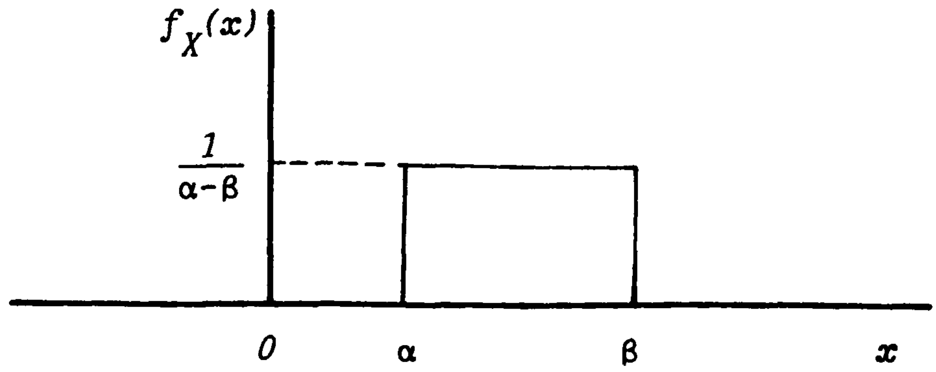
$$f_X(x) = \begin{cases} \dfrac{1}{\beta-\alpha}\,, & \alpha \leqq x \leqq \beta \\ 0\,, & \text{ἀλλοῦ}\,, \end{cases} \tag{4.19}$$

ὅπου α καί β εἶναι σταθεροί πραγματικοί ἀριθμοί. Τότε ἡ X λέγεται ὁμοιόμορφη τ.μ. στό διάστημα (α,β) καί συμβολίζεται μέ

$$X \sim U(\alpha,\beta).$$

Ἡ κατανομή τῆς X λέγεται ὁμοιόμορφη στό διάστημα (α,β). Καταχρηστικά λέμε ὅτι ἡ τ.μ. παίρνει τιμές στό διάστημα (α,β). Ἐάν τά σημεῖα α καί β εἶναι ἄγνωστα (μεταβλητά) τότε τά α καί β εἶναι οἱ παράμετροι τῆς κατανομῆς.

Ἡ γραφική παράσταση τῆς συναρτήσεως *(4.19)* δίδεται στό Σχῆμα 4.6.



Σχῆμα 4.6 Ὁμοιόμορφη κατανομή

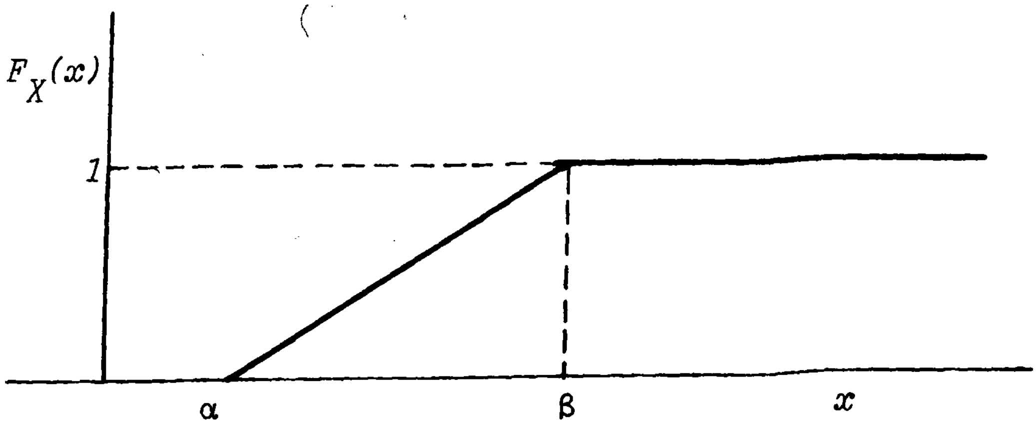
Εἶναι φανερό ὅτι ἡ παραπάνω συνάρτηση εἶναι μία γνήσια σ.π.π. Τό ἐμβαδόν τοῦ ὀρθογωνίου τοῦ Σχήματος 4.6 ἰσοῦται μέ τή μονάδα. Ἐξ αἰτίας τοῦ σχήματος αὐτοῦ ἡ ὁμοιόμορφη κατανομή λέγεται καί ὀ ρ θ ο γ ώ ν ι α κατανομή.

Ἡ α.σ.κ. τῆς ὁμοιόμορφης τ.μ. εἶναι

$$F_X(x) = \begin{cases} 0 \quad , & x \leq \alpha \\ \dfrac{x-\alpha}{\beta-\alpha} \quad , & \alpha \leq x \leq \beta \\ 1 \quad , & x \geq \beta \end{cases} \qquad (4.20)$$

καί ἡ γραφική της παράσταση δίνεται στό Σχῆμα 4.7.

Ἡ χαρακτηριστική ἰδιότητα τῆς ὁμοιόμορφης κατανομῆς εἶναι ὅτι ἴσα ὑποδιαστήματα τοῦ (α,β) ἔχουν τήν ἴδια πιθανότητα. Οἱ ἐφαρμογές τῆς $U(\alpha,\beta)$ στηρίζονται σ'αὐτή ἀκριβῶς τήν ἰδιότητα.Παράλληλα ἡ ὁμοιόμορφη κατανομή ἔχει καί μία ἄλλη σημαντική ἰδιότητα πού θά συναντήσουμε ὅταν μελετήσουμε τό θέμα τῶν κατανομῶν μετασχηματισμῶν τυχαίων



Σχῆμα 4.7 Ἀθροιστική συνάρτηση
Ὁμοιόμορφης κατανομῆς.

μεταβλητῶν. Χρησιμοποιεῖται εὐρύτατα γιά τήν ἐπεξήγηση διαφόρων πτυχῶν τῆς στατιστικῆς θεωρίας.

4.8 Ἐκθετική κατανομή

Ἄς θεωρήσουμε μία διαδικασία τοῦ *Poisson* $X(t)$ μέ παράμετρο λ. Τά τυχαῖα ἐνδεχόμενα τῆς διαδικασίας αὐτῆς παρατηροῦνται μέ τή πάροδο τοῦ χρόνου. Ἔστω T ὁ χρόνος ἀ ν α μ ο ν ῆ ς ἀπό τήν ἀρχή τῶν μετρήσεων μέχρι τήν ἐμφάνιση (ἄφιξη) τοῦ πρώτου τυχαίου ἐνδεχομένου. Εἶναι φανερό ὅτι τό T εἶναι μία συνεχής τ.μ. Γιά νά εὕρουμε τήν κατανομή πιθανότητας τῆς τ.μ. T σκεπτόμαστε ὅτι ὁ χρόνος ἀναμονῆς T εἶναι μεγαλύτερος μίας τιμῆς t, $t \geq 0$, ὅταν δέν ἔχουμε καμμία ἄφιξη τυχαίου ἐνδεχομένου στό χρονικό διάστημα $(0,t)$. Μέ ἄλλα λόγια

$$\{T > t\} = \{X(t) = 0\}.$$

Ἄρα

$$P(T>t) = P(X(t) = 0) = e^{-\lambda t} ,$$

δεδομένου ὅτι ἡ σ.π.π. τῆς $X(t)$ εἶναι $e^{-\lambda t}(\lambda t)^x/x!$. Συνεπῶς

$$P(T\leq t) = 1-P(T>t) = 1-e^{-\lambda t} ,$$

καί ἡ α.σ.κ. τῆς T εἶναι

$$F_T(t) = \begin{cases} 0 & , \quad t \leq 0 \\ 1-e^{-\lambda t} & , \quad t > 0 . \end{cases} \qquad (4.21)$$

Παραγωγίζοντες τήν $F_T(t)$ εὑρίσκουμε τή σ.π.π. τῆς T

$$f_T(t) = \begin{cases} 0 & , \ t \leq 0 \\ \lambda e^{-\lambda t} & , \ t > 0 \end{cases} , \ \lambda > 0 . \qquad (4.22)$$

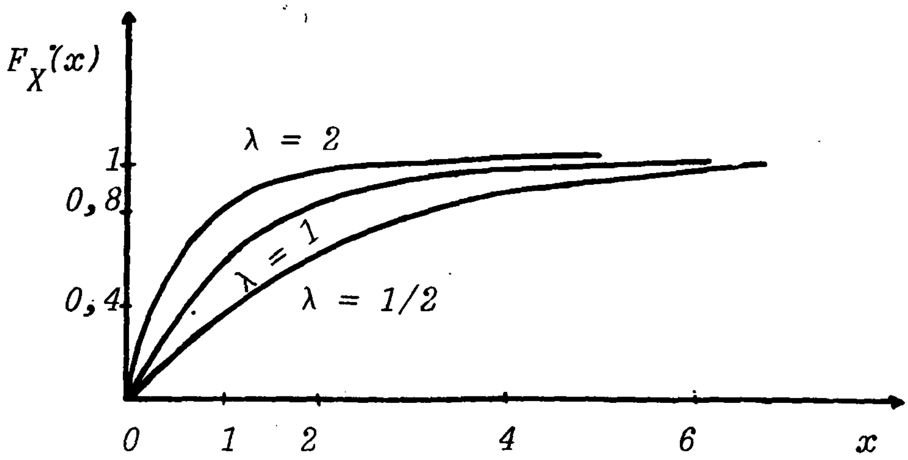
Ἀπό τόν ὁρισμό τῆς διαδικασίας τοῦ *Poisson* εἶναι φανερό ὅτι ἡ κατανομή (4.22) δίνει τήν κατανομή τοῦ χρόνου ἀναμονῆς μεταξύ δύο οἱωνδήποτε ἀφίξεων τυχαίων ἐνδεχομένων.

Μία τυχαία μεταβλητή X μέ σ.π.π. τήν (4.22) καί α.σ.κ. τήν (4.21) (ὅπου θέτουμε x ἀντί t) λέγεται ἐ κ θ ε τ ι κ ή (ἤ ἀρνητική ἐκθετική) τ.μ. μέ π α ρ ά μ ε τ ρ ο λ καί συμβολίζεται μέ

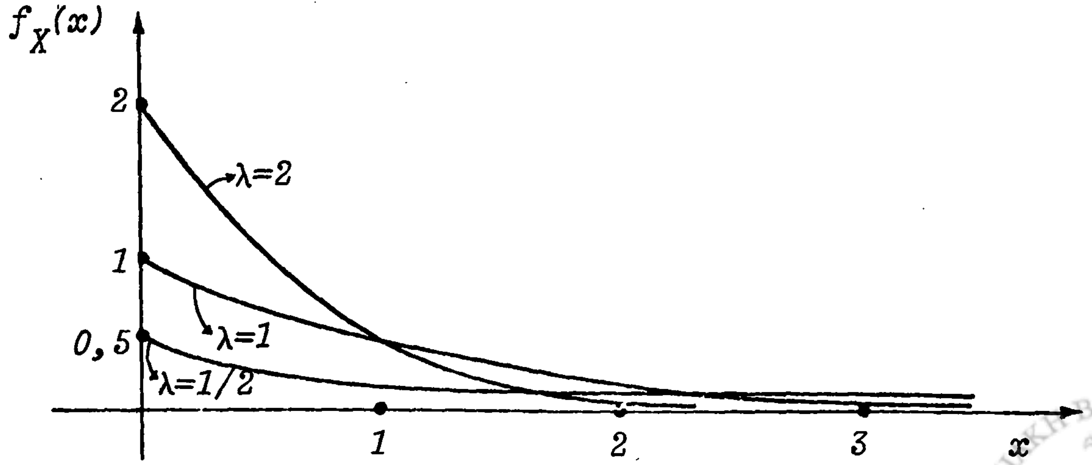
$$X \sim Ek\theta(\lambda).$$

Ἡ ἀντίστοιχη κατανομή λέγεται ἐκθετική. Εἶναι εὔκολο νά δειχθεῖ ὅτι τό ὁλοκλήρωμα τῆς (4.22) ἀπό $-\infty$ ἕως $+\infty$ ἰσοῦται μέ τήν μονάδα. Ὁ ὑπολογισμός πιθανοτήτων διαστημάτων βάσει τῆς ἐκθετικῆς κατανομῆς γίνεται μέ τή βοήθεια τῆς α.σ.κ. (4.21) ἤ μέ ἀπ'εὐθείας ὁλοκλήρωση τῆς σ.π.π. (4.22).

Ἡ γραφική παράσταση τῆς α.σ.κ. τῆς ἐκθετικῆς τ.μ. δίνεται στό Σχῆμα 4.8 καί τῆς σ.π.π. στό Σχῆμα 4.9.



Σχῆμα 4.8: Ἀθροιστικές συναρτήσεις ἐκθετικῶν κατανομῶν



Σχῆμα 4.9: Συναρτήσεις πυκνότητας πιθανότητας ἐκθετικῶν κατανομῶν

Παραδείγματα:

4.10: Έστω ότι οί πέστροφες πού πιάνει ένας ψαράς στόν ποταμό Λοΰρο ακολουθούν τή διαδικασία τοϋ *Poisson* μέ $\lambda = 1,8$ καί μονάδα χρόνου τήν ώρα. Νά βρεθεί ή πιθανότητα μετά ένα "ψάρεμα" ό ψαράς νά περιμένει λιγότερο άπό *20* λεπτά μέχρι τό νέο "ψάρεμα".

Ό χρόνος αναμονής μεταξύ δύο "ψαρεμάτων" ακολουθεί εκθετική κατανομή μέ παράμετρο $\lambda = 1,8$. Τά *20* λεπτά είναι *1/3* τής ώρας. Έτσι ή πιθανότητα πού ζητάμε είναι:

$$P(T\leq 1/3) = \int_0^{1/3} 1,8e^{-1,8t}dt = 1-e^{-1,8/3} = 0,4511.$$

Ό χρόνος ή ή διάρκεια ζωής ηλεκτρικών ή ηλεκτρονικών συσκευών ή άλλων μηχανημάτων (ψυγεία, πλυντήρια κλπ.) συνήθως ακολουθεί τήν εκθετική κατανομή. Τό ίδιο ισχύει καί γιά πολλές τυχαίες μεταβλητές εμπορικοοικονομικής φύσεως όπως π.χ. οί ημερήσιες πωλήσεις, ημερησία κατανάλωση ενός είδους κλπ.

4.11: Μετά τήν πρόσληψη ενός νέου υπαλλήλου ή αύξηση τών ημερήσιων πωλήσεων ενός καταστήματος ακολουθεί μία εκθετική κατανομή μέ $\lambda = 1/2$. Εάν οί ημερήσιες πωλήσεις είναι ανεξάρτητες μεταξύ τους καί διαλέξουμε τρεις μέρες στήν τύχη ποιά ή πιθανότητα (*i*) καί τίς τρεις μέρες ή αύξηση νά είναι μεγαλύτερη άπό 10 μονάδες καί (*ii*) ή αύξηση νά είναι μεγαλύτερη άπό *8* μονάδες τουλάχιστον σέ δύο άπό τίς τρεις μέρες.

(*i*) 'Η πιθανότητα ἡ αὔξηση στίς πωλήσεις μία ὁποιαδήποτε τυχαία ἡμέρα νά εἶναι μεγαλύτερη ἀπό 10 μονάδες εἶναι

$$P(X \geq 10) = \int_{10}^{\infty} \tfrac{1}{2}\, e^{-x/2} dx = e^{-5} .$$

Λόγω τῆς ἀνεξαρτησίας, γιά τίς τρεῖς ἡμέρες ἔχουμε διωνυμικό πείραμα μέ $p = e^{-5}$. Ἄρα ἡ πιθανότητα ἡ ἴδια αὔξηση νά παρατηρηθεῖ καί τίς 3 ἡμέρες εἶναι

$$\binom{3}{3}(e^{-5})^3(1-e^{-5})^0 = e^{-15} = 3{,}06\times 10^{-7} .$$

(*ii*) 'Η πιθανότητα ἡ αὔξηση νά εἶναι μεγαλύτερη ἀπό 8 μονάδες τήν ἡμέρα εἶναι

$$P(X \geq 8) = \int_{8}^{\infty} \tfrac{1}{2}\, e^{-x/2} dx = e^{-4} .$$

Βάσει τῆς διωνυμικῆς κατανομῆς ἡ πιθανότητα τοῦ β΄ μέρους εἶναι

$$\binom{3}{2}(e^{-4})^2(1-e^{-4})+\binom{3}{3}(e^{-4})^3(1-e^{-4})^0 = 3e^{-8}-2e^{-12} = 9{,}94\times 10^{-4}.$$

'Ι δ ι ό τ η τ α τ ῆ ς ἀ μ ν η σ ί α ς τ ῆ ς ἐ κ θ ε τ ι κ ῆ ς κ α τ α ν ο μ ῆ ς. Κλείνουμε τό ἐδάφιο αὐτό μέ μία χαρακτηριστική ἰδιότητα τῆς διαδικασίας τοῦ *Poisson* καί τῆς ἐκθετικῆς κατανομῆς πού εἶναι γνωστή ὡς ἰδιότητα τῆς ἀ μ ν η σ ί α ς.

Ἔστω ὅτι θέλουμε νά βροῦμε τήν κατανομή τοῦ χρόνου T ἀναμονῆς μεταξύ δύο ἀφίξεων δοθέντος ὅτι ὁ χρόνος αὐτός εἶναι μεγαλύτερος δεδομένου χρόνου t_0. Μέ ἄλλα λόγια ζητᾶμε τήν κατανομή τῆς τ.μ. $T|T>t_0$. Γιά $t>t_0$ ἔχουμε

$$P(T\leq t|T>t_0) = \frac{P(T\leq t, T>t_0)}{P(T>t_0)} = \frac{P(t_0<T\leq t)}{1-P(T\leq t_0)}$$

$$= \frac{F_T(t)-F_T(t_0)}{1-F_T(t_0)} = \frac{1-e^{-\lambda t}-(1-e^{-\lambda t_0})}{1-(1-e^{-\lambda t_0})}$$

$$= 1-e^{-\lambda(t-t_0)} .$$

Παραγωγίζοντες ὡς πρός t βρίσκομε τή σ.π.π. τῆς $T|T>t_0$:

$$f_{T|T>t_0}(t) = \begin{cases} 0 & , \quad t \leq t_0 \\ \lambda e^{-\lambda(t-t_0)} & , \quad t > t_0. \end{cases} \tag{4.23}$$

Βλέπουμε δηλαδή ὅτι ὁ ἐ π ι π ρ ό σ θ ε τ ο ς χρόνος ἀναμονῆς ἔχει τήν ἴ δ ι α κατανομή μέ τόν ἀ ρ χ ι κ ό χ ρ ό ν ο ἀναμονῆς, δηλαδή ἐκθετική μέ παράμετρο λ. Εἶναι εὔκολο τώρα νά δικαιολογήσουμε τήν ἀκόλουθη σχέση

$$P(T>t+t_0|T>t_0) = P(T>t) . \tag{4.24}$$

Μέ ἄλλα λόγια ἡ πιθανότητα ἀναμονή πού συνεχίζεται σέ χρόνο t_0 νά συνεχιστεῖ πέρα ἀπό τό χρόνο $t+t_0$ εἶναι ἀνεξάρτητη τῆς προηγούμενης διάρκειας τῆς ἀναμονῆς. Αὐτό λέγεται ἀ μ ν η σ ί α τ ῆ ς ἐ κ θ ε τ ι κ ῆ ς κ α τ α ν ο μ ῆ ς. Ἡ διαδικασία τοῦ *Poisson* κάθε στιγμή ξεχνᾶ τό παρελθόν καί ἀρχίζει ἐκ νέου.

Ἡ ἐ κ θ ε τ ι κ ή κ α τ α ν ο μ ή ε ἶ ν α ι ἡ μ ό ν η ἀ π ό τ ί ς σ υ ν ε χ ε ῖ ς κ α τ α ν ο μ έ ς μ έ τ ή ν ἰ δ ι ό τ η τ α τ ῆ ς ἀ μ ν η σ ί α ς (4.24). Ἡ μ ό ν η ἀ π ό τ ί ς δ ι α κ ρ ι τ έ ς κατ α ν ο μ έ ς π ο ύ ἔ χ ε ι τ ή ν ἴ δ ι α ἰ δ ι ό - τ η τ α ε ἶ ν α ι ἡ γ ε ω μ ε τ ρ ι κ ή κ α τ α ν ο - μ ή. 'Εάν ὁ χρόνος ζωῆς ἑνός ἀτόμου ἤ ἑνός ἐξαρτήματος μηχανῆς ἀκολουθεῖ τήν ἐκθετική κατανομή (ἤ γεωμετρική στή περίπτωση πού ὁ χρόνος μετριέται μέ διακριτό τρόπο) τότε "γ έ ρ α σ μ α" δέν γίνεται· ἐφ'ὅσον τό ἄτομο ζεῖ, ἔχει τήν ἴδια πιθανότητα νά ἀποσυντεθεῖ τήν ἑπόμενη στιγμή. 'Αντίστροφα, ἐάν εἶναι γνωστό ὅτι ἕνα φαινόμενο ἔχει τήν ἰδιότητα τῆς ἀμνησίας ἤ "ἔλλειψη γεράσματος" τότε ἡ κατανομή τῆς διάρκειας του πρέπει νά εἶναι γεωμετρική ἤ ἐκθετική.

Παραδείγματα:

4.12: Ἔστω ὅτι ὁ χρόνος ζωῆς X ἑνός ἐξαρτήματος μίας συσκευῆς ἀκολουθεῖ ἐκθετική κατανομή μέ παράμετρο $\lambda = 7$. Ἔστω ἐπί πλέον ὅτι τό ἐξάρτημα ἔχει "ζήσει" x_0 ὧρες. Νά βρεθεῖ ἡ πιθανότητα τό ἐξάρτημα νά διαρκέσει τουλάχιστον x ἐπί πλέον ὧρες.

'Η πιθανότητα πού ζητᾶμε εἶναι

$$P(X>x_0+x \mid X>x_0) = \frac{P(X>x_0+x, X>x_0)}{P(X>x_0)}$$

$$= \frac{P(X>x_0+x)}{P(X\leqq x_0)}$$

$$= \frac{1-P(X\leqq x_0+x)}{1-P(X\leqq x_0)}$$

Βάσει δέ τῆς (4.21) ἔχομε

$$P(X>x_0+x|X>x_0) = \frac{e^{-7(x_0+x)}}{e^{-7x_0}} = e^{-7x} .$$

Παρατηροῦμε ὅτι ἡ πιθανότητα τό ἐξάρτημα νά διαρκέσει τουλάχιστον x ἀρχικές ὧρες εἶναι ἐπίσης e^{-7x}. Τό ἐξάρτημα δέν "κουράζεται" ἤ "φθείρεται".

4.13: Ἔστω ὅτι οἱ ἀφίξεις λεωφορείων σέ μία στάση ἀκολουθοῦν διαδικασία τοῦ *Poisson* μέ $\lambda = 6$ ἀφίξεως ἀνά ὧρα ἤ $0,1$ ἀνά λεπτό. Νά βρεθεῖ ἡ πιθανότητα ὁ Δημήτρης νά περιμένει στή στάση λιγότερο ἀπό ἐπί πλέον λεπτά δοθέντος ὅτι ἔχει ἤδη περιμένει 15 λεπτά.

Βάσει τῆς (4.23) ἡ πιθανότητα πού ζητᾶμε εἶναι

$$P(T\leq 20|T>15) = \int_{15}^{20} \lambda e^{-\lambda(t-15)}dt$$

$$= \int_{0}^{5} \lambda e^{-\lambda t}dt = 1-e^{-0,1\times 5} \approx 0,39 .$$

4.9 Οἱ μαθηματικὲς συναρτήσεις Γάμμα καὶ Βῆτα καὶ ἄλλες χρήσιμες μαθηματικὲς γνώσεις

Οἱ συναρτήσεις καί τά θεωρήματα πού παρουσιάζονται στό ἐδάφιο αὐτό ὁδηγοῦν σέ δύο ἀξιόλογες κατανομές πιθανότητας καί διευκολύνουν τό λογιστικό μέρος τῆς θεωρίας Πιθανοτήτων καί τῆς Στατιστικῆς. Ἀρχίζουμε μέ τόν ὁρισμό τῶν συναρτήσεων Γάμμα καί Βῆτα.

Ὁρισμοί:

4.2: Τό γενικευμένο ὁλοκλήρωμα

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx \quad , \quad \alpha > 0 \qquad (4.24)$$

λέγεται σ υ ν ά ρ τ η σ η Γ ά μ μ α καί ὁρίζεται γιά κάθε $\alpha > 0$.

4.3: Τό ὁλοκλήρωμα

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx \quad , \quad \alpha > 0, \quad \beta > 0 \quad (4.25)$$

λέγεται σ υ ν ά ρ τ η σ η Β ή τ α καί ὁρίζεται γιά κάθε $\alpha > 0$, $\beta > 0$.

Μέ ἐξαίρεση ὁρισμένες ἁπλές περιπτώσεις, τά ὁλοκληρώματα αὐτά δέν ὑπολογίζονται μέ κλειστούς τύπους. Ἐκτεταμένοι πίνακες τιμῶν τῶν συναρτήσεων Γάμμα καί Βήτα ὑπάρχουν στή Μαθηματική Βιβλιογραφία.

Τά θεωρήματα πού ἀκολουθοῦν εἶναι βασικῆς σημασίας. Ἡ ἀπόδειξη τους παραλείπεται ὅταν εἶναι προφανής.

Θεώρημα 4.1:

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1), \quad \alpha > 1. \qquad (4.26)$$

Ἀπόδειξη: Ὁλοκληρώνοντες κατά παράγοντες ἔχομε

$$\begin{aligned}\Gamma(\alpha) &= -\int_0^\infty x^{\alpha-1}de^{-x} = -\{x^{\alpha-1}e^{-x}\Big|_0^\infty - (\alpha-1)\int_0^\infty e^{-x}x^{\alpha-2}dx\} \\ &= -\{0-0-(\alpha-1)\int_0^\infty x^{\alpha-2}e^{-x}dx\} \\ &= (\alpha-1)\Gamma(\alpha-1) \;\blacktriangledown\end{aligned}$$

Θεώρημα 4.2:

$$\Gamma(1) = 1 . \qquad (4.27)$$

Πόρισμα 4.1: Ἐάν n ἀκέραιος θετικός ἀριθμός τότε

$$\Gamma(n) = (n-1)! \qquad (4.28)$$

Ἀπόδειξη: Μέ ἐπαγωγή βάσει τῶν θεωρημάτων 4.1 καί 4.2 ▼

Θεώρημα 4.3:

<u>Σχέση μεταξύ Βήτα καί Γάμμα συναρτήσεων</u>

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} . \qquad (4.29)$$

Ἀπόδειξη: Θέτομε $x = sin^2\theta$ μέ $dx = 2sin\theta cos\theta d\theta$ καί ὁ μετασχηματισμός αὐτός μᾶς δίνει

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$$

$$=2\int_0^{\pi/2} sin^{2\alpha-1}\theta cos^{2\beta-1}\theta d\theta \qquad (4.30)$$

Ἐφαρμόζοντες τούς μετασχηματισμούς

$$x = y^2,\ dx = 2ydy$$

καί

$$x = z^2, \quad dx = 2zdz$$

στά ολοκληρώματα Γ(α) καί Γ(β) άντίστοιχα βρίσκουμε

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx = 2\int_0^\infty y^{2\alpha-1} e^{-y^2} dy$$

$$\Gamma(\beta) = \int_0^\infty x^{\beta-1} e^{-x} dx = 2\int_0^\infty z^{2\beta-1} e^{-z^2} dz .$$

Έτσι πολλαπλασιάζοντες Γ(α) μέ Γ(β) έχομε

$$\Gamma(\alpha)\Gamma(\beta) = 4\int_0^\infty \int_0^\infty y^{2\alpha-1} z^{2\beta-1} e^{-(y^2+z^2)} dydz$$

καί άλλάζοντες τίς συντεταγμένες (z,y) σέ πολικές (r,θ) βάσει τῶν τύπων

$$\begin{aligned} z &= r\cos\theta \\ y &= r\sin\theta \end{aligned} \quad \text{μέ ἰακωβιανή} \quad |J| = r$$

έχομε

$$\Gamma(\alpha)\Gamma(\beta) = 4\int_0^\infty \int_0^{\pi/2} (r\sin\theta)^{2\alpha-1} (r\cos\theta)^{2\beta-1} e^{-r^2} r d\theta dz \tag{4.31}$$

$$= \{2\int_0^\infty e^{-r^2} r^{2\alpha+2\beta-1} dr\}\{2\int_0^{\pi/2} (\sin\theta)^{2\alpha-1} (\cos\theta)^{2\beta-1} \theta d\theta\}.$$

Τέλος έφαρμόζομε τόν μετασχηματισμό $x = r^2$, $dx = 2rdr$ στό ολοκλήρωμα Γ(α+β)

$$\Gamma(\alpha+\beta) = 2\int_0^\infty r^{2(\alpha+\beta)-1} e^{-r^2} dr \tag{4.32}$$

'Αντικαθιστῶντες τίς σχέσεις (4.30), (4.31) καί (4.32) στήν (4.29) ἔχομε τό ἀποτέλεσμα ▼

Θ ε ώ ρ η μ α 4.4:

$$I = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2}dx = \sqrt{2\pi} \quad . \tag{4.33}$$

'Απόδειξη: 'Υπάρχουν διάφοροι τρόποι ἀποδείξεως. 'Εδῶ χρησιμοποιοῦμε διπλά ὁλοκληρώματα:

$$I^2 = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2}dx \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2}dy$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-(x^2+y^2)/2}dxdy \; .$$

'Αλλάζοντες τά (x,y) σέ πολικές συντεταγμένες (r,θ) ὅπως καί στήν ἀπόδειξη τοῦ Θεωρήματος 4.3 βρίσκουμε μέ κατ' εὐθεῖαν ὁλοκλήρωση

$$I^2 = \int_0^{2\pi}\int_0^{\infty} re^{-\frac{1}{2}r^2}drd\theta$$
$$= \int_0^{2\pi} d\theta \int_0^{\infty} re^{-\frac{1}{2}r^2}dr = 2\pi\times 1 \quad .$$

"Αρα $I = \sqrt{2\pi}$ ▼

Θ ε ώ ρ η μ α 4.5:

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi} \; . \tag{4.34}$$

Ἀπόδειξη: Ἐφαρμόζοντες στό ὁλοκλήρωμα $\Gamma(\frac{1}{2})$ τό μετασχηματισμό $x = \frac{1}{2}y^2$ καί χρησιμοποιοῦντες τό θεώρημα 4.4 ἔχομε

$$\Gamma(\tfrac{1}{2}) = \int_0^\infty x^{-\frac{1}{2}}e^{-x}dx = \int_0^\infty \frac{y^{-1}}{2^{-\frac{1}{2}}}\, e^{-\frac{1}{2}y^2}\, y\,dy$$

$$= \sqrt{2}\int_0^\infty e^{-\frac{1}{2}y^2}dy = \sqrt{2}\,\tfrac{1}{2}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2}dy$$

$$= \frac{\sqrt{2}}{2}\sqrt{2\pi} = \sqrt{\pi}\ .$$

Στό προτελευταῖο βῆμα χρησιμοποιήσαμε τήν ἰδιότητα τῶν ἀρτίων συναρτήσεων $\varphi(x) = \varphi(-x)$:

$$\int_{-\alpha}^{\alpha} \varphi(x)dx = 2\int_0^{\alpha} \varphi(x)dx\ . \ \blacktriangledown$$

Γιά περιττές συναρτήσεις $\varphi(x) = -\varphi(-x)$ ἰσχύει

$$\int_{-\alpha}^{\alpha} \varphi(x)dx = 0$$

Πόρισμα 4.2: Ἐάν n ἀκέραιος θετικός ἀριθμός τότε

$$\Gamma(n+\tfrac{1}{2}) = \frac{1.3.5\ldots(2n-1)}{2^n}\sqrt{\pi}\ . \qquad (4.35)$$

Πόρισμα 4.3:

$$B(\alpha,\beta) = B(\beta,\alpha). \qquad (4.36)$$

4.10 Κατανομή Γάμμα

῾Η κατανομή Γάμμα ἀποτελεῖ γενίκευση τῆς ἐκθετικῆς καί χρησιμοποιεῖται γιά τήν περιγραφή διαφόρων φυσικῶν φαινομένων. Μπορεῖ νά ἀποδειχθεῖ ὅτι ὁ χρόνος ἀναμονῆς μέχρι τῆς ἐμφανίσεως τοῦ α στή σειρά (α φυσικός ἀριθμός) τυχαίου ἐνδεχομένου σέ μία διαδικασία τοῦ *Poisson* μέ παράμετρο $\lambda = 1/\beta$ ἀκολουθεῖ τή κατανομή Γάμμα.

῾Η σ.π.π. τῆς κατανομῆς αὐτῆς εἶναι:

$$f(x) = \begin{cases} \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)}\, x^{\alpha-1} e^{-x/\beta} , & x > 0 \\ 0 , & x \leqq 0 , \end{cases} \qquad (4.37)$$

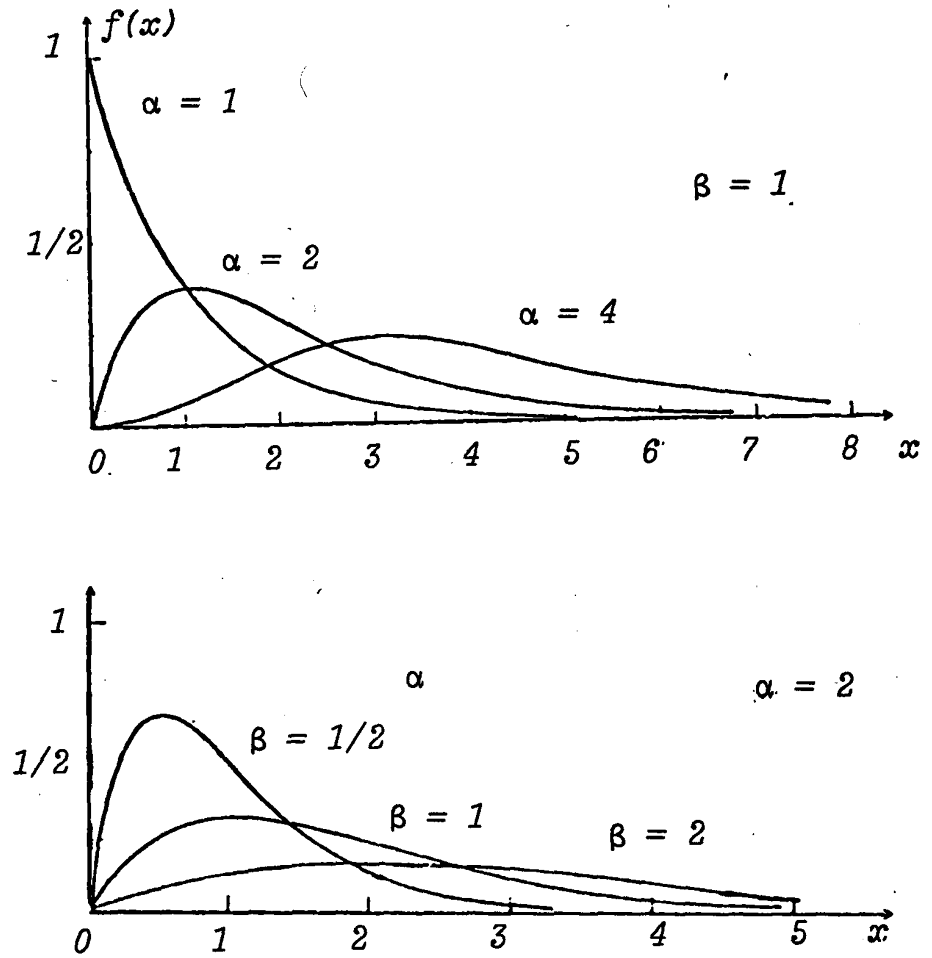
ὅπου $\alpha > 0$, $\beta > 0$ εἶναι οἱ παράμετροι τῆς κατανομῆς. Χρησιμοποιοῦντες τόν ὁρισμό τῆς συναρτήσεως Γάμμα εἶναι εὔκολο νά δειχθεῖ ὅτι ἡ συνάρτηση *(4.37)* εἶναι μία γνήσια σ.π.π. ὡς μή ἀρνητική μέ ὁλοκλήρωμα ἴσο πρός τήν μονάδα. Μία τ.μ. X μέ σ.π.π. τήν *(4.37)* λέγεται Γάμμα τ.μ. καί συμβολίζεται μέ

$$X \sim G(\alpha, \beta).$$

῾Η γραφική παράσταση τῆς Γάμμα σ.π.π. δίνεται στό Σχῆμα 4.10 γιά διάφορες τιμές τῶν παραμέτρων α καί β.

᾿Εάν $\alpha = 1$ καί $\beta = 1/\lambda$ ἡ κατανομή Γάμμα γίνεται ἡ ἐκθετική κατανομή μέ παράμετρο λ.

Γιά $\alpha = \nu/2$ καί $\beta = 2$ ἡ κατανομή Γάμμα γίνεται μιά ἄλλη εἰδική κατανομή, γνωστή ὡς χί τετράγω-



Σχῆμα 4.10: Κατανομές Γάμμα

νομέν βαθμούς ἐλευθερίας χ_ν^2 ἡ ὁποία θά μελετηθεῖ στό Κεφάλαιο 8. 'Ιδιαίτερα ἡ σ.π.π. τῆς κατανομῆς χ_ν^2 εἶναι

$$f(x) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)}\, x^{(\nu/2)-1} e^{-x/2}, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{4.38}$$

Συχνά γράφουμε

$$\chi^2_\nu = G(\alpha=\nu/2, \beta=2) \quad . \qquad (4.39)$$

'Εάν έξαιρέσει κανείς μερικές άπλές περιπτώσεις, π.χ. $\alpha = 2$ καί όποιοδήποτε β, ή α.σ.κ. τῆς Γάμμα κατανομῆς δέν μπορεῖ νά ύπολογιστεῖ μέ τύπο κλειστῆς μορφῆς. 'Εάν τό α εἶναι φυσικός άριθμός τό όλοκλήρωμα

$$F_X(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x y^{\alpha-1} e^{-y/\beta} dy$$

ύπολογίζεται έπακριβῶς· διαφορετικά χρειάζεται άριθμητική όλοκλήρωση. Γιά τούς λόγους αύτούς ό ύπολογισμός πιθανοτήτων βάσει τῆς κατανομῆς Γάμμα παρουσιάζει δυσκολίες καί συχνά γίνεται χρήση πινάκων τῆς συναρτήσεως Γάμμα ἤ, σέ ειδικές περιπτώσεις, τῆς κατανομῆς χ^2_ν.

Τό έπόμενο θεώρημα δίνεται χωρίς άπόδειξη. Μπορεῖ εὔκολα ν'άποδειχθεῖ μέ διαδοχικές όλοκληρώσεις κατά παράγοντας ἤ βάσει τοῦ όρισμοῦ τῆς Γάμμα τυχαίας μεταβλητῆς ὡς τοῦ χρόνου άναμονῆς γιά τήν έμφάνιση τοῦ α στή σειρά τυχαίου ένδεχομένου μίας διαδικασίας τοῦ *Poisson*.

Θ ε ώ ρ η μ α 4.6:

<u>'Εάν</u> $X \sim G(\alpha, \beta)$ <u>καί</u> α <u>φυσικός</u> <u>άριθμός</u> <u>τότε</u> <u>ή</u> <u>α.σ.κ.</u> <u>τῆς</u> X <u>δίνεται</u> <u>άπό</u> <u>τόν</u> <u>τύπο</u>

$$F_X(x) = 1 - \sum_{j=0}^{\alpha-1} \frac{e^{-x/\beta}(x/\beta)^j}{j!} \quad . \qquad (4.40)$$

Γιά $\beta = 1$ ή $F_X(x)$ λέγεται μ ή π λ ή ρ η ς σ υ-ν ά ρ τ η σ η Γ ά μ μ α. Ἐκτεταμένοι πίνακες δίνοντες τίς τιμές τῆς συναρτήσεως αὐτῆς ὑπάρχουν στή βιβλιογραφία [*Pearson, K. (1922). Tables of Incomplete Gamma Functions, Cambridge University Press*].

Μέχρι τώρα ἔχουμε δεῖ δύο εἴδη "χ ρ ό ν ο υ ἀ ν α μ ο ν ῆ ς": διακριτός καί συνεχής. Ὁ διακριτός χρόνος ἀναμονῆς ἐμφανίζεται μέ τή γεωμετρική καί τήν ἀρνητική διωνυμική κατανομή καί ὁ συνεχής μέ τήν ἐκθετική καί τήν κατανομή Γάμμα. Ἡ ἐκθετική κατανομή ἀντιστοιχεῖ στήν γεωμετρική καί ἡ Γάμμα στήν ἀρνητική διωνυμική κατανομή.

Παράδειγμα 4.14:

Ἡ περισσότερη ἀπό χίλια λίτρα ἡμερήσια κατανάλωση βενζίνης, μετρούμενη σέ χιλιάδες λίτρα, ἑνός πρατηρίου βενζίνης μπορεῖ νά θεωρηθεῖ ὅτι ἔχει κατανομή Γάμμα μέ $\alpha = 3$ καί $\beta = 2$. Ἐάν οἱ δεξαμενές τοῦ πρατηρίου περιέχουν *5000* λίτρα ποιά ἡ πιθανότητα τό πρατήριο νά μείνει χωρίς βενζίνη μία μέρα.

Ἐάν Y εἶναι ἡ ἡμερήσια κατανάλωση βενζίνης τοῦ πρατηρίου σέ χιλιάδες λίτρα τότε $X = Y-1 \sim G(3,2)$. Ἡ σ.π.π. τῆς X εἶναι

$$f_X(x) = \begin{cases} \dfrac{1}{2^3 \times 2!}\, x^2 e^{-x/2}, & x > 0 \\ 0, & x \leq 0 . \end{cases}$$

Τό πρατήριο θά μείνει χωρίς βενζίνη ἐάν $Y \geq 5$. Ἄρα ἡ πιθανότητα πού ζητᾶμε εἶναι

$$P(Y \geq 5) = P(X+1 \geq 5) = P(X \geq 4)$$

$$= \int_4^\infty \frac{1}{2^3 \times 2!} x^2 e^{-x/2} dx \quad (\text{θέτομε} \quad z = x/2)$$

$$= \int_2^\infty \tfrac{1}{2} z^2 e^{-z} dz = 5e^{-2} = 0{,}6767 \; .$$

4.11 Κατανομή Βήτα

Όταν ή τυχαία μεταβλητή παίρνει τιμές μεταξύ 0 καί 1 (π.χ. συνεχῆ ποσοστά) συχνά χρησιμοποιοῦμε τή Βήτα κατανομή. Ἡ ὁμοιόμορφη κατανομή στό διάστημα $(0,1)$ εἶναι είδική περίπτωση τῆς κατανομῆς πού μελετᾶμε στό ἐδάφιο αὐτό.

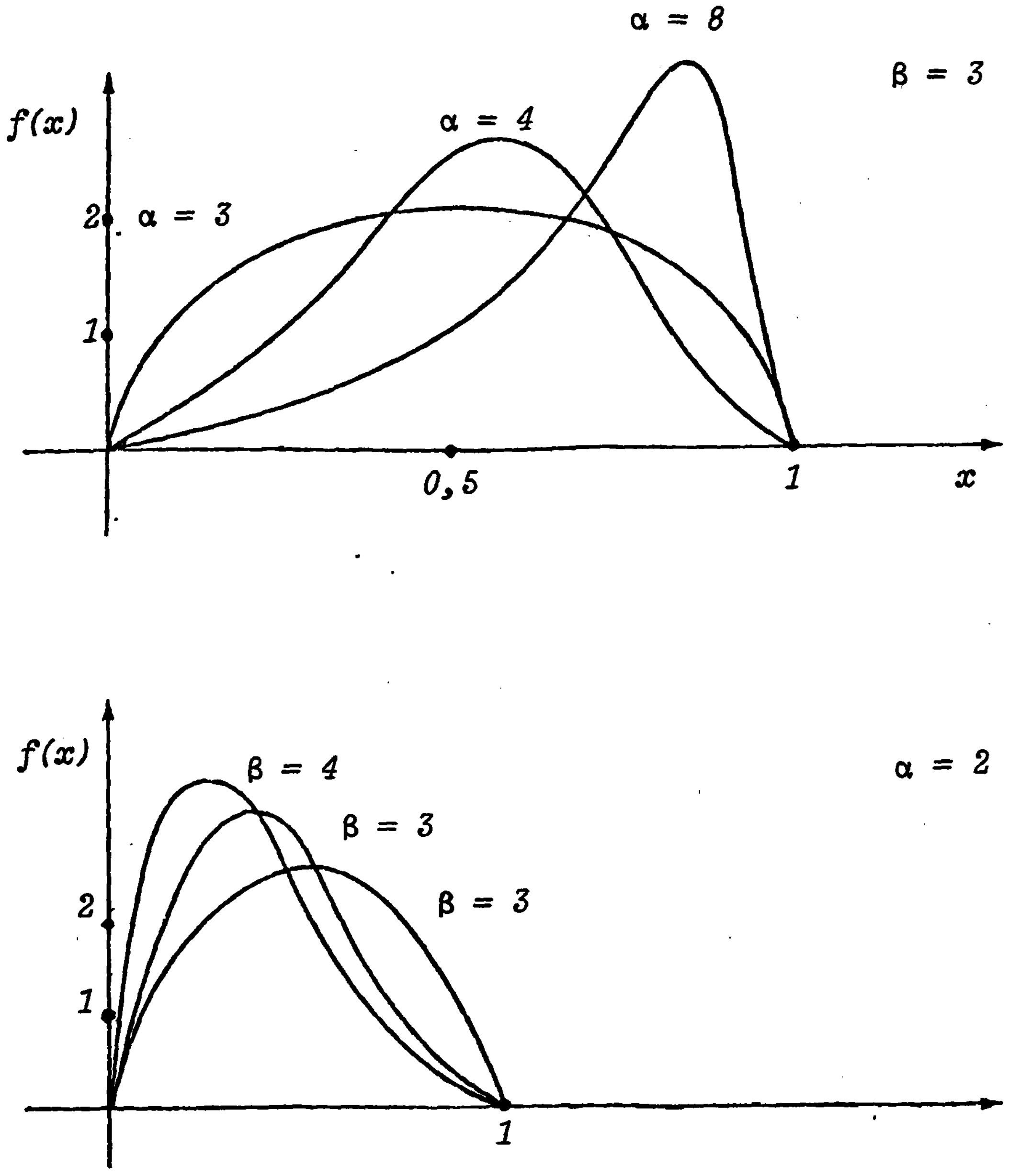
Ἡ σ.π.π. τῆς **Βήτα κατανομῆς** εἶναι

$$f_X(x) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & 0<x<1 \\ 0 & , \text{ἀλλοῦ} , \end{cases} \tag{4.41}$$

ὅπου $\alpha > 0$, $\beta > 0$ εἶναι οἱ **παράμετροι** τῆς κατανομῆς. Ἡ ἰσότητα στό πρῶτο σκέλος τῆς (4.41) στηρίζεται στό θεώρημα 4.3 τῆς §4.9. Βάσει τοῦ ὁρισμοῦ τῆς συναρτήσεως Βήτα εἶναι εὔκολο νά δειχθεῖ ὅτι ἡ (4.41) εἶναι μία γνήσια σ.π.π. Μία τ.μ. X μέ σ.π.π. τήν (4.41) λέγεται Βήτα τ.μ. καί συμβολίζεται μέ

$$X \sim Be(\alpha,\beta)$$

Ἡ γραφική παράσταση τῆς Βήτα σ.π.π. δίδεται στό Σχῆμα 4.11.



Σχῆμα 4.11: Κατανομές Βήτα

Βλέπουμε ὅτι, ὅταν $\alpha < \beta$ ἡ σ.π.π. εἶναι λοξή πρός τά δεξιά καί ὅταν $\beta < \alpha$ εἶναι λοξή πρός τ'ἀριστερά. Ὅταν

$\alpha = \beta$ ἡ κατανομή εἶναι συμμετρική. Ἡ γραφική παράσταση τῆς $Be(\alpha,\beta)$ ἔχει σχῆμα U ὅταν $\alpha < 1$ καί $\beta < 1$ καί σχῆμα J ὅταν $\alpha < 1$ ἤ $\beta < 1$. Ὅταν $\alpha = \beta = 1$ ἡ κατανομή γίνεται $U(0,1)$.

Ἡ α.σ.κ. τῆς Βήτα κατανομῆς εἶναι

$$F_X(x) = \begin{cases} 0 & , \quad x \leq 0 \\ \displaystyle\int_0^x \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}\, dy & , \quad 0 < x < 1 \\ 1 & , \quad x \geq 1 \ . \end{cases} \tag{4.42}$$

Ὅπως καί στή περίπτωση τῆς Γάμμα κατανομῆς ἔτσι καί ἐδῶ τό ὁλοκλήρωμα τῆς *(4.42)* δέν ὑπολογίζεται μέ κλειστό τύπο ἐκτός ἐάν τά α καί β ἔχουν ὁρισμένες ἀκέραιες τιμές. Ἔτσι ὁ ὑπολογισμός τῶν πιθανοτήτων τῆς Βήτα κατανομῆς παρουσιάζει δυσκολίες. Τό ὁλοκλήρωμα τῆς (4.42) λέγεται μή πλήρης συνάρτηση Βήτα καί ἐκτεταμένοι πίνακες δίνοντες τίς τιμές του γιά διάφορα x,α καί β ὑπάρχουν στή βιβλιογραφία.

<u>Παράδειγμα 4.15</u>:

Τό ἡμερήσιο ποσοστό τῶν ἀρρένων πού δέχονται πρῶτες βοήθειες σ'ἕνα νοσοκομεῖο μπορεῖ νά θεωρηθεῖ ὡς τυχαία μεταβλητή μέ κατανομή $Be(\alpha=6,\ \beta=4)$. Νά βρεθεῖ ἡ πιθανότητα μία μέρα τό ποσοστό αὐτό νά κυμαίνεται μεταξύ *60%* καί *80%*.

Ἡ πιθανότητα πού ζητᾶμε εἶναι

$$P(0,6<X<0,8) = \int_{0,6}^{0,8} \frac{\Gamma(10)}{\Gamma(4)\Gamma(6)} x^5(1-x)^3 dx$$

$$= 0,4032.$$

4. 12 Κανονική κατανομή ή κατανομή του Gauss

Ἡ κανονική κατανομή εἶναι ἡ πιό βασική καί περισσότερο γνωστή κατανομή τῆς θεωρίας Πιθανοτήτων καί τῆς Στατιστικῆς. Ἀποτελεῖ τή θεμελιώδη κατανομή πάνω στήν ὁποία στηρίζεται ὁλόκληρο τό ἐπιστημονικό οἰκοδόμημα τῶν κλάδων αὐτῶν. Ἔχει μοναδικές μαθηματικές, πιθανοθεωρητικές καί στατιστικές ἰδιότητες καί ὑπό ὡρισμένες συνθῆκες μπορεῖ νά ἐφαρμοστεῖ σέ κάθε ἐπιστημονικό πρόβλημα.

Ἀνακαλύφθηκε τό 1733 ἀπό τόν *Abraham De Moivre* ὡς τό ὅριο τῆς διωνυμικῆς κατανομῆς. Τό 1774 ὁ *Pierre Laplace* ἐμελέτησε τίς μαθηματικές της ἰδιότητες. Παρ'ὅλα αὐτά ἀποδίδεται στόν *Gauss* ὁ ὁποῖος τήν ἀναφέρει γιά πρώτη φορά σέ μία ἐργασία του τό 1809. Μελετήθηκε ἐκτεταμένα τό 18ο καί 19ο αἰώνα ἀπό ἐπιστήμονες οἱ ὁποῖοι παρατήρησαν ὅτι τά μοντέλα (κατανομές) τῶν σφαλμάτων τῶν μετρήσεων τους μποροῦσαν νά προσεγγισθοῦν ἱκανοποιητικά ἀπό μία συνεχή καμπύλη. Ἡ καμπύλη αὐτή ἐλέγετο "κανονική καμπύλη τῶν σφαλμάτων" καί ἀποδίδετο στούς νόμους τῆς τύχης. Πρός τιμήν τοῦ *Gauss* ἀναφέρεται ὡς κατανομή τοῦ *Gauss*. Σήμερα ὁ ὅρος κανονική κατανομή εἶναι ὁ πιό διαδεδομένος.

Διακρίνουμε δύο περιπτώσεις: (α) τήν τυπική (ἤ τυποποιημένη) κανονική κατανομή μέ σ.π.π.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad -\infty < x < \infty \tag{4.43}$$

καί (β) τήν (γενική) κανονική κατανομ ή με σ.π.π.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad \begin{array}{l} -\infty < x < \infty \\ -\infty < \mu < \infty \\ \sigma > 0, \end{array} \quad (4.44)$$

όπου μ καί σ^2 ἤ (μ καί σ) εἶναι οἱ παράμετροι τῆς κατανομῆς. Προφανῶς οἱ συναρτήσεις *(4.43)* καί *(4.44)* εἶναι μή ἀρνητικές. Τό θεώρημα 4.4 τῆς §4.9 ἀποδεικνύει ὅτι τό ὁλοκλήρωμα τῆς *(4.43)* ἀπό $-\infty$ ἕως $+\infty$ ἰσοῦται μέ τή μονάδα καί [μέ τή βοήθεια τοῦ μετασχηματισμοῦ $y = (x-\mu)/\sigma$] τό ἴδιο ἰσχύει γιά τή συνάρτηση *(4.44)*. Ἔτσι καί οἱ δύο προηγούμενες συναρτήσεις εἶναι σ.π.π.

Μία τ.μ. X μέ σ.π.π. τή *(4.43)* λέγεται τ υ π ι κ ή κ α ν ο ν ι κ ή τ.μ. ἤ μ ε τ α β λ η τ ή Z καί συμβολίζεται μέ

$$X \sim N(0,1).$$

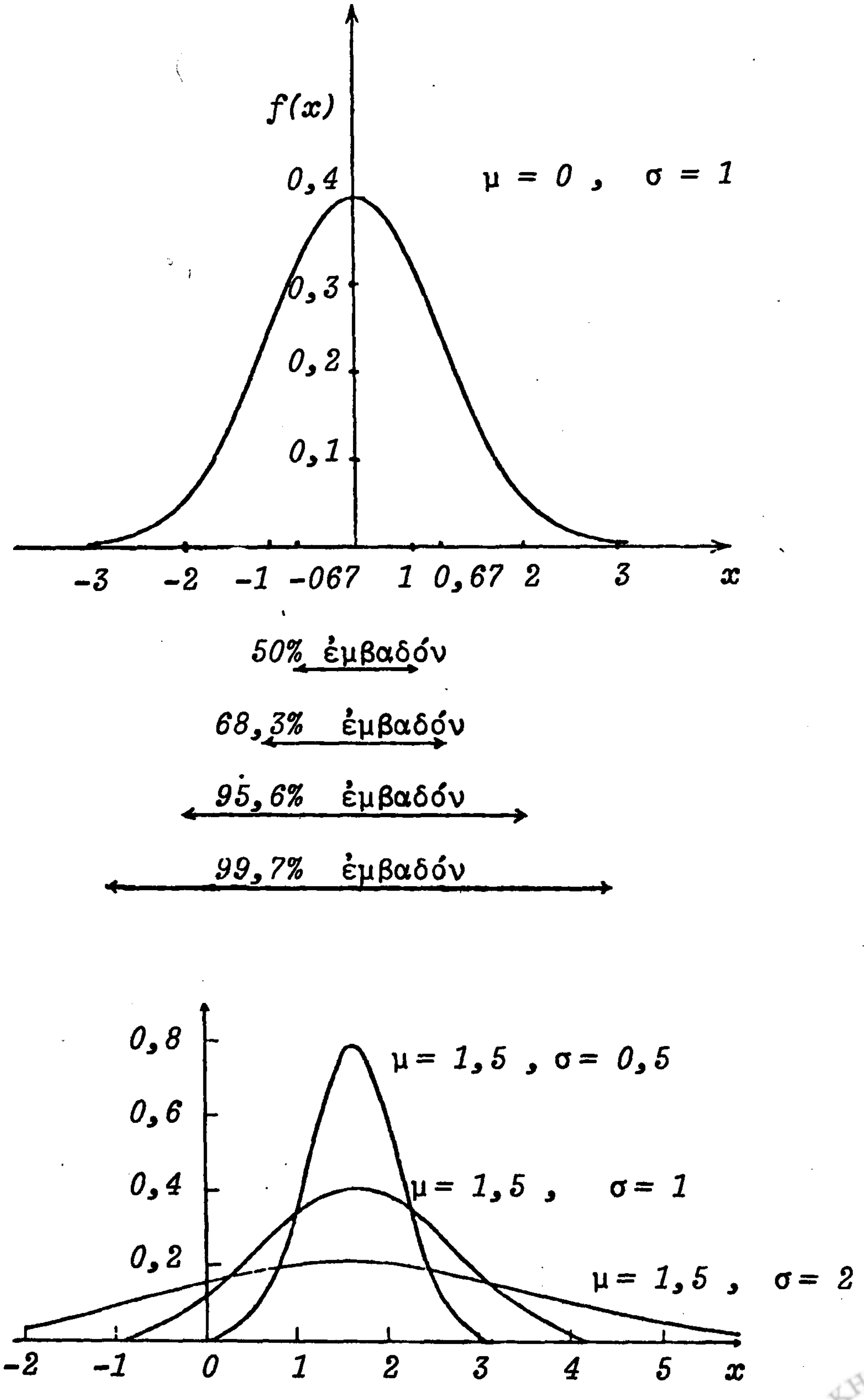
Μία τ.μ. X μέ σ.π.π.τή *(4.44)* λέγεται κ α ν ο ν ι κ ή τ. μ. καί συμβολίζεται μέ

$$X \sim N(\mu,\sigma^2).$$

Ἡ τυπική κατανομή προκύπτει ἀπό τή γενική θέτοντες $\mu = 0$ καί $\sigma^2 = 1$.

Οἱ γραφικές παραστάσεις τῶν συναρτήσεων *(4.43)* καί *(4.44)* δίνονται στό Σχῆμα 4.12.

Ἡ σημασία τῶν παραμέτρων μ καί σ^2 θά καταφανεῖ στό ἑπόμενο κεφάλαιο. Τό μ ἀντιπροσωπεύει τή μ έ σ η τ ι μ ή καί τό σ^2 τ ή δ ι α κ ύ μ α ν σ η τῆς κατανομῆς. Ἡ σ.π.π. τῆς κανονικῆς κατανομῆς εἶναι σ υ μ μ ε τ ρ ι κ ή γύρω ἀπό τό σημεῖο $x = \mu$, δηλαδή



Σχῆμα 4.12: Κανονικές κατανομές

$$f(\mu-x) = f(\mu+x) \qquad \text{γιά κάθε} \quad x \in R$$

καί τό μ έ γ ι σ τ ο της παρατηρεῖται στό σημεῖο μ μέ τιμή $1/(\sigma\sqrt{2\pi})$. 'Από αὐτό καί ἀπό τό γεγονός ὅτι τό ἐμβαδόν μεταξύ τῆς καμπύλης τῆς $f(x)$ καί τοῦ ἄξονα τῶν x ἰσοῦται μέ τή μονάδα συμπεραίνουμε ὅτι ὅσο μεγαλύτερο εἶναι τό σ τόσο περισσσότερο "ἁπλωμένη" εἶναι ἡ $f(x)$ καί ἀντίστροφα. 'Η συνάρτηση $f(x)$ ἔχει σ η μ ε ῖ α ἀ ν α κ ά μ ψ ε ω ς τά $x = \mu-\sigma$ καί $x = \mu+\sigma$.

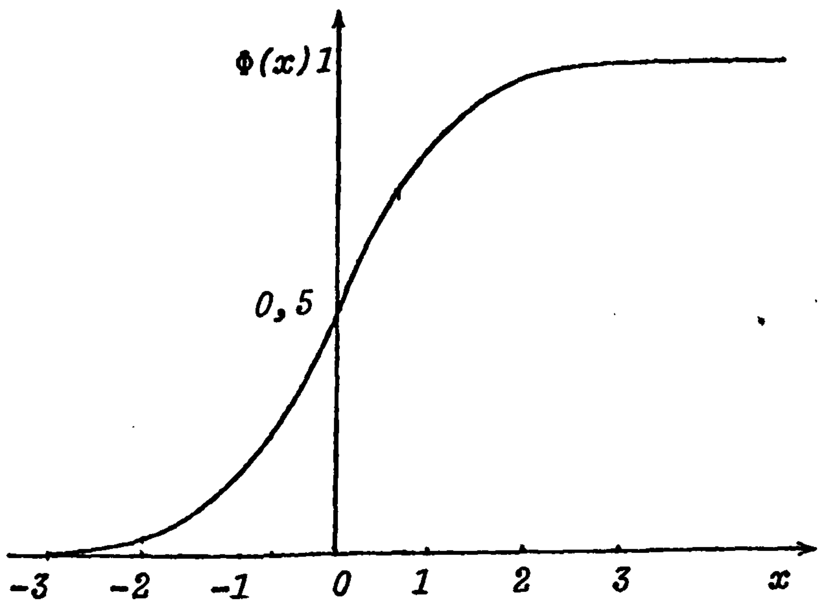
'Η α.σ.κ. μίας κανονικῆς τ.μ. μέ παραμέτρους μ καί σ^2 εἶναι

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \,, \quad -\infty < x < \infty. \tag{4.45}$$

'Ιδιαίτερα ἡ α.σ.κ. τῆς τυπικῆς κανονικῆς κατανομῆς συμβολίζεται μέ $\Phi(x)$, δηλαδή

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt, \quad -\infty < x < \infty \,. \tag{4.46}$$

'Η γραφική παράσταση τῆς $\Phi(x)$ δίνεται στό Σχῆμα 4.13.



Σχῆμα 4.13: 'Αθροιστική συνάρτηση τυπικ. κανον.κατανομῆς

Θεώρημα 4.7:

(α) Ἐάν $X \sim N(0,1)$ τότε $Y = \sigma X + \mu \sim N(\mu, \sigma^2)$

(β) Ἐάν $X \sim N(\mu, \sigma^2)$ τότε $Y = (X-\mu)/\sigma \sim N(0,1)$.

Ἀπόδειξη: θά ἀποδείξουμε μόνο τή δεύτερη πρόταση μέ τή μέθοδο τῆς α.σ.κ. Ἡ πρώτη πρόταση ἀποδεικνύεται ἀνάλογα. Ἔχομε

$$F_Y(y) = P(Y \leq y) = P\left(\frac{X-\mu}{\sigma} \leq y\right)$$

$$= P(X \leq \sigma y + \mu) = F_X(\sigma y + \mu).$$

Ἀλλά ἡ $F_X(x)$ δίνεται ἀπό τήν (4.45). Ἔτσι ἀλλάζοντες μεταβλητές ἔχομε

$$F_Y(y) = \int_{-\infty}^{\sigma y+\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

$$= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt = \Phi(y).$$

Ἡ α.σ.κ. τῆς τ.μ. Y εἶναι ἡ $\Phi(y)$ ἡ ὁποία ὁρίζει μονοσήμαντα τήν κατανομή. Ἄρα $Y \sim N(0,1)$. ▼

Τό ὁλοκλήρωμα μέ τό ὁποῖο ὁρίζεται ἡ $\Phi(x)$ καί κατά συνέπεια καί τό ὁλοκλήρωμα τῆς (4.45) δέν μποροῦν νά ὑπολογιστοῦν μέ κλειστό τύπο παρά μόνον ἀριθμητικά. Ἐκτεταμένοι πίνακες τῆς $f(x)$ καί τῆς $\Phi(x)$ ὑπάρχουν στήν βιβλιογραφία. Οἱ πίνακες τῆς $\Phi(x)$ χρησιμοποιοῦνται γιά τόν ὑπολογισμό πιθανοτήτων τῆς κανονικῆς κατανομῆς. Οἱ πίνακες αυτοί είτε δίνουν τήν $\Phi(x)$ είτε μόνο τό ολοκλήρωμα

$$\int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} t^2} dt \qquad (4.47)$$

γιά θετικές τιμές τοῦ x. Λόγω τῆς συμμετρίας τῆς κατανομῆς, ἀπό τίς τιμές αὐτές μποροῦμε νά ὑπολογίσουμε τή $\Phi(x)$ γιά κάθε x. 'Ισχύουν οἱ ἀκόλουθες σχέσεις

$$\Phi(x) = 1-\Phi(-x) \quad \text{γιά κάθε } x.$$

$$\Phi(x) = \begin{cases} \frac{1}{2} + \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} t^2} dt & \text{γιά } x \geqq 0 \\ \frac{1}{2} - \int_0^{-x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} t^2} dt & \text{γιά } x \leqq 0 . \end{cases}$$

Γιά τίς ἀνάγκες τοῦ παρόντος βοηθήματος θά χρησιμοποιήσουμε πίνακες πού δίνουν τίς τιμές τοῦ ὁλοκληρώματος (4.47) [βλέπε Πίνακα III τοῦ Παραρτήματος].

Δίνουμε τώρα ἁπλά ἀριθμητικά παραδείγματα ὑπολογισμοῦ πιθανοτήτων:

(i) Ἔστω $X \sim N(0,1)$. Τότε

$$\begin{aligned} P(-1{,}2<X<2{,}22) &= P(-1{,}2<X<0)+P(0<X<2{,}22) \\ &= P(0<X<1{,}2)+P(0<X<2{,}22) \\ &= 0{,}3849+0{,}4864 = 0{,}8713. \end{aligned}$$

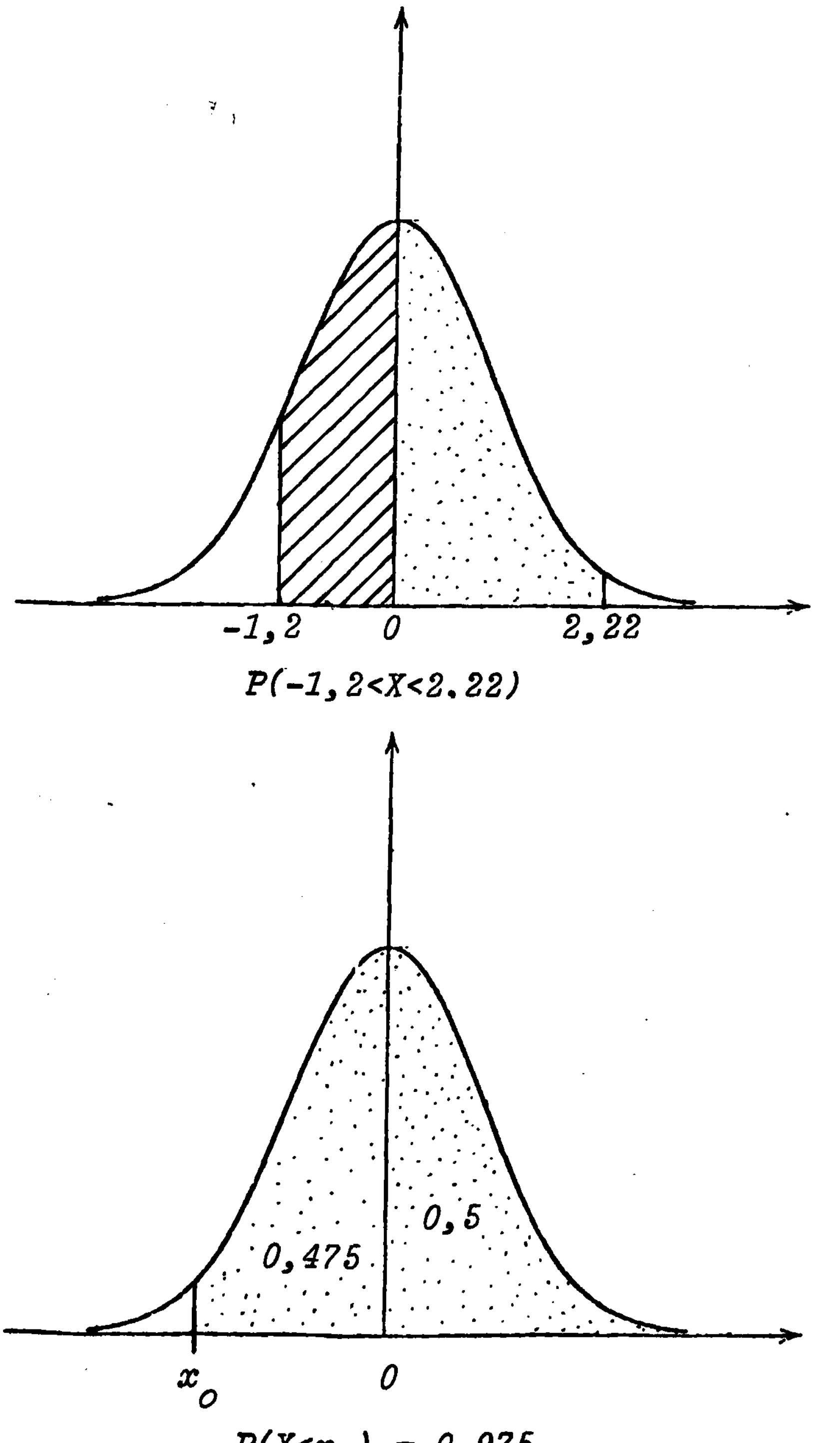
'Εάν θέλουμε νά βροῦμε τό σημεῖο x_0 τέτοιο ὥστε

$$P(X>x_0) = 0{,}975 ,$$

τό x_0 πρέπει νά εἶναι ἀρνητικό. 'Ισοδύναμα ἔχομε $P(0<X<-x_0) = 0,475$ καί ἀπό τόν Πίνακα III βρίσκουμε

$$-x_0 = 1,96 \quad \text{ἤ} \quad x_0 = -1,96.$$

Οἱ ὑπολογισμοί αὐτοί γραφικά φαίνονται στό Σχῆμα 4.13.



Σχῆμα 4.13: Ὑπολογισμοί πιθανοτήτων κανον. κατανομῆς

(ii) Έστω τώρα $X \sim N(2,4)$, δηλαδη $\mu = 2$ καί $\sigma^2 = 4$. Τότε μέ τή βοήθεια τοῦ θεωρήματος 4.7(β) ἔχομε

$$P(2,5<X<4,7) = P\left(\frac{2,5-2}{2} < \frac{x-\mu}{\sigma} < \frac{4,7-2}{2}\right)$$

$$= P(0,25<Z<1,35),$$

ὅπου $Z = (X-\mu)/\sigma \sim N(0,1)$. Ἔτσι

$$P(2,5<X<4,7) = P(0<Z<1,35)-P(0<Z<0,25)$$

$$= 0,4115-0,0987 = 0,3128 .$$

Γιά νά βροῦμε τό x_0 τέτοιο ὥστε $P(X<x_0) = 0,8$ ἔχομε τίς ἀκόλουθες ἰσοδύναμες σχέσεις

$$P\left(\frac{x-\mu}{\sigma} < \frac{x_0-2}{2}\right) = 0,8$$

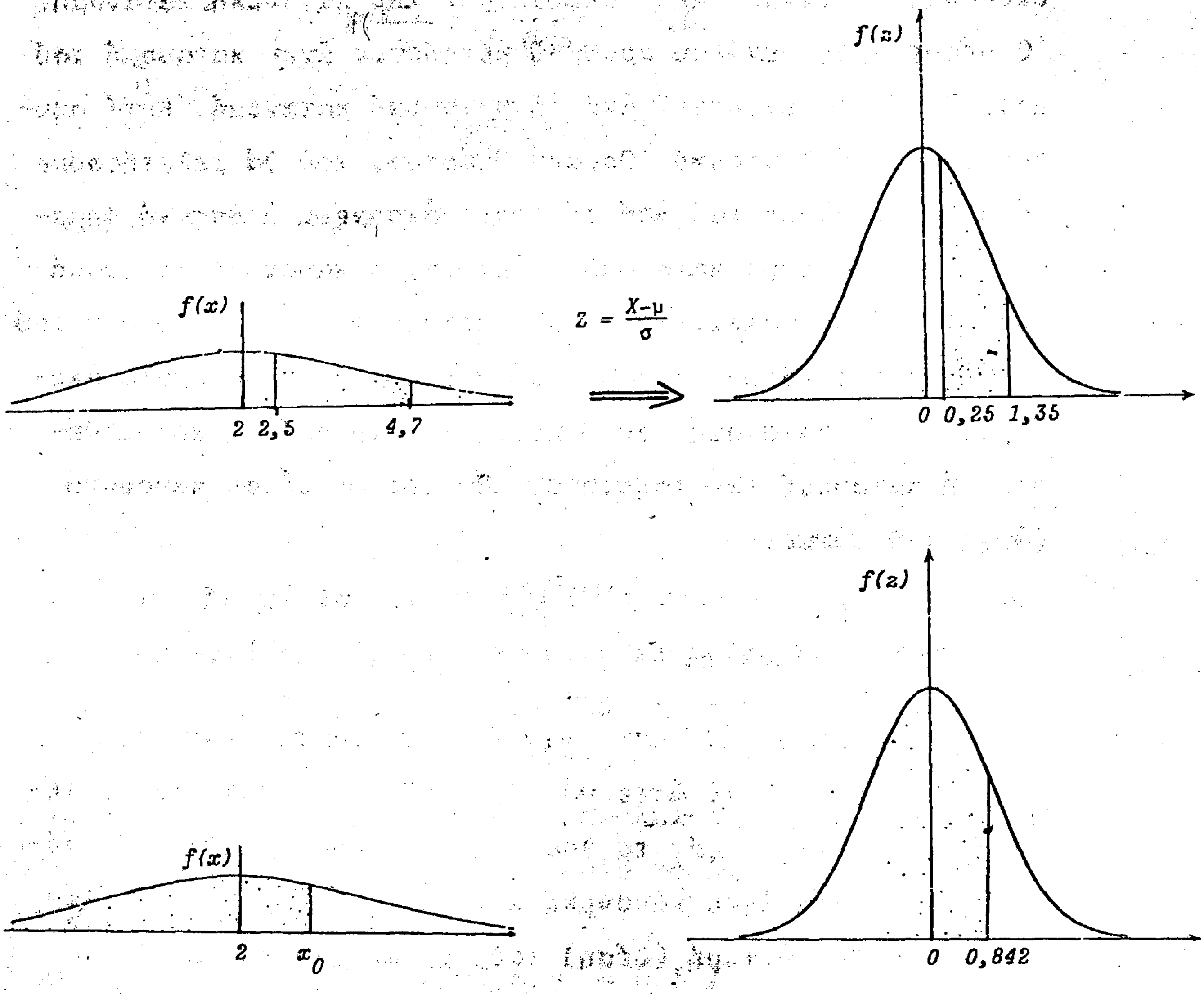
$$P\left(0<Z < \frac{x_0-2}{2}\right) = 0,3 .$$

Ἀπό τόν Πίνακα III παίρνουμε

$$\frac{x_0-2}{2} = 0,842 \quad \text{ἤ} \quad x_0 = 3,684 .$$

Τό Σχῆμα 4.14 παριστάνει γραφικά τούς παραπάνω ὑπολογισμούς.

Οἱ πράξεις $Z = (X-\mu)/\sigma$ λέγονται τ υ π ο π ο ί η σ η τῆς τ.μ. X.



Σχῆμα 4.14 Ὑπολογισμοί πιθανοτήτων κανονικῆς κατανομῆς.

'Η κανονική κατανομή δίνει τό κατάλληλο μοντέλο πιθανοτήτων γιά τή περιγραφή πολλῶν τυχαίων φαινομένων. Ποσότητες ὅπως π.χ. τό μῆκος, τό βάρος, ὁ ὄγκος, τό ὕψος, ἡ ἐπίδοση, οἱ βαθμοί κλπ. ἀκολουθοῦν τήν κανονική κατανομή. 'Ο μέσος ὅρος μεγάλου ἀριθμοῦ μετρήσεων ἔχει κατανομή πού μπορεῖ νά προσεγγιστεῖ ἀπό τή κανονική κατανομή. Αὐτό συνδέεται μέ τό Κεντρικό 'Οριακό Θεώρημα, πού θά μελετήσουμε σέ ἄλλο κεφάλαιο καί ἀπό τό ὁποῖο ἀπορρέει ἡ βασική σημασία τῆς κανονικῆς κατανομῆς. 'Επίσης ἡ κανονική κατανομή εἶναι τό ὅριο πολλῶν κατανομῶν πιθανότητας (ὑπό ἔννοια πού θά δοῦμε πιό κάτω). Τυχαῖα (μή συστηματικά) σφάλματα ἀκολουθοῦν τήν κατανομή τοῦ *Gauss*. 'Υπό ὡρισμένες προϋποθέσεις ἡ κατανομή τῶν ταχυτήτων τῶν μορίων εἶναι κανονική (ἀρχή τοῦ *Maxwell*).

Παραδείγματα:

4.16: Μία αὐτόματη μηχανή πωλήσεως πορτοκαλάδας μπορεῖ νά ρυθμιστεῖ ὥστε νά "βγάλει" κατά μέσο ὅρο μ λίτρα ἀνά κύπελλο. 'Εάν τό ποσό τῆς πορτοκαλάδας πού "βγαίνει" κάθε φορά ἔχει κανονική κατανομή μέ $\sigma = 0,008$ λίτρα νά βρεθεῖ ἡ τιμή (θέση) τοῦ μ ὥστε κύπελλα τῶν $0,230$ λίτρων νά ξεχυλίζουν μόνο μία στίς ἑκατό φορές.

"Εστω X ἡ τ.μ. πού μετρᾶ τό ποσό τῆς πορτοκαλάδας. Τότε

$$X \sim N(\mu, \sigma^2 = 0,008^2) .$$

Γιά νά ξεχυλίσει τό κύπελλο θά πρέπει νά συμβεῖ τό ἐνδεχόμενο $\{X > 0,230\}$. "Ετσι ἔχομε

$$P(X > 0,230) = 0,01$$

ή

$$P\left(\frac{X-\mu}{\sigma} > \frac{0,23-\mu}{0,008}\right) = 0,01$$

ή

$$P\left(Z > \frac{0,23-\mu}{0,008}\right) = 0,01 .$$

'Από τούς Πίνακες III έχομε

$$\frac{0,23-\mu}{0,008} = 2,329 .$$

"Αρα

$$\mu = 0,2281.$$

4.17: Οἱ βαθμοί σέ ἕνα μάθημα ὑποτίθεται ὅτι ἀκολουθοῦν κατά προσέγγιση κανονική κατανομή μέ $\mu = 6,5$, $\sigma = 0,5$ καί ἄριστα τό δέκα. 'Εάν 3 μαθητές ἐκλεγοῦν στήν τύχη νά βρεθεῖ ἡ πιθανότητα ἀκριβῶς δύο ἀπό τούς τρεῖς μαθητές νά ἔχουν βαθμό μεγαλύτερο ἀπό 7.

"Εστω X ἡ τ.μ. πού παριστᾶ τό βαθμό στό μάθημα. Τότε

$$X \sim N(\mu=6,5, \sigma^2 = 0,5^2).$$

"Εστω p ἡ πιθανότητα ὁ μαθητής πού ἐκλέγεται στήν τύχη νά ἔχει βαθμό μεγαλύτερο ἀπό 7. Τότε μέ τυποποίηση καί τή βοήθεια τοῦ Πίνακα III ἔχομε

$$p = P(X>7) = P\left(\frac{X-\mu}{\sigma} > \frac{7-6,5}{0,5}\right)$$

$$= P(Z>1) = 0,1586.$$

Ἡ πιθανότητα αὐτή παραμένει ἡ ἴδια γιά κάθε μαθητή τῆς τάξεως. Ὑποθέτουμε δηλαδή ὅτι τό διωνυμικό πρότυπο μπορεῖ νά υἱοθετηθεῖ. Ἄρα ἡ πιθανότητα πού ζητᾶμε εἶναι

$$\binom{3}{2}p^2(1-p) = \binom{3}{2}(0,1586)^2(0,8414) = 0,0635.$$

4. 13 Ἄλλες συνεχεῖς κατανομές

Στό ἐδάφιο αὐτό δίνουμε ἕνα κατάλογο ἄλλων χρήσιμων συνεχῶν κατανομῶν

α) Κατανομή τοῦ *Cauchy* μέ σ.π.π.

$$f(x) = \frac{1}{\pi}\frac{\alpha}{\alpha^2+(x-\mu)^2}, \quad -\infty<x<\infty, \quad \begin{matrix}-\infty<\mu<\infty\\ \alpha>0\end{matrix} . \qquad (4.48)$$

β) Κατανομή τοῦ *Laplace* ἤ διπλή ἐκθετική μέ σ.π.π.

$$f(x) = \frac{1}{2\sigma}e^{-\left|\frac{x-\mu}{\sigma}\right|}, \quad -\infty<x<\infty \quad \begin{matrix}-\infty<\mu<\infty\\ \sigma>0\end{matrix} \qquad (4.49)$$

Χρησιμοποιεῖται σέ μηχανολογικά προβλήματα.

γ) Λογαριθμοκανονική κατανομή μέ σ.π.π.

$$f(x) = \begin{cases} \dfrac{1}{x\beta\sqrt{2\pi}}e^{-\frac{(lnx-ln\alpha)^2}{2\beta^2}} & , \; x > 0 \\ 0 & , \; x \le 0 \end{cases} \quad \alpha,\beta>0 \qquad (4.50)$$

Ἐάν X ἔχει τήν λογαριθμοκανονική κατανομή τότε ἡ τ.μ. lnX ἔχει τήν κανονική κατανομή. Ἔχει χρησιμοποιηθεῖ σέ προβλήματα Φυσικῆς, Βιομηχανίας, Οἰκονομικῶν Ἐπιστημῶν, Κοινωνιολογίας, Βιολογίας καί Ἀνθρωπομετρίας.

δ) Κατανομή τοῦ *Weibull* μέ σ.π.π.

$$f(x) = \begin{cases} \frac{\beta}{\alpha}(x-\gamma)^{\beta-1} e^{-\frac{(x-\gamma)^\beta}{\alpha}} & , x > \gamma \\ & \qquad \alpha, \beta, \gamma > 0 \\ 0 & , x \leqq \gamma . \end{cases} \tag{4.51}$$

Ἐφαρμόζεται σέ πολλά φυσικά φαινόμενα. Περιγράφει ἱκανοποιητικά τή διάρκεια ζωῆς διαφόρων ἀντικειμένων ὅπως συσκευῶν, ἐξαρτημάτων κλπ.

ε) Κατανομή τοῦ *Maxwell* μέ σ.π.π.

$$f(x) = \begin{cases} \frac{2}{\pi} x^2 e^{-\frac{1}{2} x^2} & , x > 0 \\ 0 & , x \leqq 0 . \end{cases} \tag{4.52}$$

Ἐφαρμόζεται σέ προβλήματα τῆς Φυσικῆς (ταχύτης μορίων).

στ) Κατανομή τοῦ *Rayleigh* μέ σ.π.π.

$$f(x) = \begin{cases} x e^{-\frac{1}{2} x^2} & , \quad x \geqq 0 \\ 0 & , \quad x \leqq 0 . \end{cases} \tag{4.53}$$

Ὁμοίως ἐφαρμόζεται σέ προβλήματα τῆς Φυσικῆς.

ζ) Κατανομή ἄκρων τιμῶν ἤ τοῦ *Gumbel* μέ σ.π.π.

$$f(x) = e^{-x-e^{-x}} \quad , \quad -\infty < x < \infty . \qquad (4.54)$$

Ἐφαρμόζεται στήν ὑδρολογία, μοντέλα ἀποτυχίας, μετεωρολογία κλπ.

η) Κατανομή τοῦ *Pareto* μέ σ.π.π.

$$f(x)=\begin{cases} \alpha k^{\alpha}(x^{\alpha+1})^{-1} , & x \geq k, \\ & \alpha \geq 0,\ k \geq 0 \\ 0 , & x < k . \end{cases} \qquad (4.55)$$

Ἐφαρμόζεται σέ οἰκομομετρικά καί ἠλεκτρομηχανολογικά προβλήματα.

ΑΣΚΗΣΕΙΣ

4.1. Ἔστω $p(x,N,n,p)$ ἡ σ.π. μιᾶς ὑπεργεωμετρικῆς τ.μ. μέ παραμέτρους (N,n,p). Νά δειχθεῖ ὅτι

$$\lim_{N\to\infty} p(x,N,n,p) = \binom{n}{x}p^x(1-p)^{n-x} .$$

4.2. Τέσσερα ἄτομα παίρνουν ὁ κάθε ἕνας ἀπό ἕνα ὁμοιόμορφο δοχεῖο. Κάθε δοχεῖο περιέχει 10 ἄσπρες καί 40 μαῦρες μπάλες. Κάθε ἄτομο ἐκλέγει χωρίς ἐπανάθεση καί στήν τύχη 8 μπάλες ἀπό τό δοχεῖο του. Νά βρεθεῖ ἡ πιθανότητα 3 ἀπό τά παραπάνω ἄτομα νά τραβήξουν τό κάθε ἕνα τουλάχιστον 2 ἄσπρες μπάλες.

4.3. Ἕνα ζάρι ρίχνεται 20 φορές. Νά βρεθεῖ ἡ πιθανότητα νά πάρουμε λιγώτερο ἀπό 4 τριάρια.

4.4. Σέ μία τάξη ὑπάρχουν 20 ἀγόρια καί 15 κορίτσια. Παίρνουμε στήν τύχη καί χωρίς ἐπανατοποθέτηση 4 μαθητές. (*i*) Ποιά ἡ πιθανότητα νά πάρουμε 3 ἀγόρια καί 1 κορίτσι; (*ii*) Ἄν ἐπαναλάβουμε τό πείραμα 3 φορές ποία ἡ πιθανότητα νά πάρουμε δύο φορές 3 ἀγόρια καί 1 κορίτσι;

4.5. Ἔστω ὅτι ὁ χρόνος ζωῆς σέ ὧρες μιᾶς λυχνίας ραδιοφώνου ἔχει πυκνότητα $f(x) = 100/x^2$ ὅταν $x > 100$ καί 0 ὅταν $x \leq 100$. Νά βρεθεῖ ἡ πιθανότητα α) καμμία ἀπό τρεῖς τέτοιες λυχνίες πού λειτουργοῦν ἀνεξάρτητα σέ ἕνα ραδιόφωνο νά μήν ἀντικατασταθεῖ κατά τή διάρκεια τῶν πρώτων 150 ὡρῶν λειτουργίας καί β) καί οἱ τρεῖς ἀρχικές λυχνίες νά ἀντικατασταθοῦν κατά τή διάρκεια τῶν πρώτων *150* ὡρῶν.

4.6. Ἕνα ἐργοστάσιο ἔχει *4* ὅμοιες γεννήτριες ἠλεκτρικοῦ ρεύματος οἱ ὁποῖες λειτουργοῦν ἡ μία ἀνεξάρτητα ἀπό τίς ἄλλες. Ἔστω ὅτι ἡ καλή ἤ ἡ κακή λειτουργία μίας γεννήτριας κατά τή διάρκεια μίας περιόδου μίας ὥρας εἶναι ἀνεξάρτητη ἀπό τή κατάσταση τῆς γεννήτριας ὁποιαδήποτε ἄλλη στιγμή. Ἔστω ὅτι ἡ πιθανότητα κακῆς λειτουργίας μίας γεννήτριας κατά τή διάρκεια μίας περιόδου μίας ὥρας εἶναι *0,02*. Νά βρεθεῖ ἡ πιθανότητα τουλάχιστον δύο γεννήτριες νά ἐπιζήσουν δύο ὧρες.

4.7. Νά βρεθεῖ ἡ πιθανότητα νά πάρουμε ἄθροισμα ἀποτελεσμάτων μικρότερο τοῦ *5* ὅταν ἕνα τέλειο ζάρι ρίχνεται *3* φορές.

4.8. Ἕνας μηχανισμός ἐκκινήσεως πού χρησιμοποιεῖται σ'ἕνα διαστημόπλοιο ἔχει μεγάλη ἀξιοπιστία καί φημίζεται νά ἀρχίζει μέ πιθανότητα *0,99999*. Ποιά ἡ πιθανότητα μίας τουλάχιστον ἀποτυχίας στίς ἑπόμενες *10.000* ἐκκινήσεις;

4.9 Ἡ βιβλιογραφία ἀναφέρει ποσοστό θνησιμότητας μίας ἀσθένειας *30%*. Τρία περιστατικά τῆς ἀσθένειας αὐτῆς εἰσάγονται σ'ἕνα Νοσοκομεῖο καί οἱ ἀσθενεῖς δέν ἐπιζοῦν. Εἶναι πιθανόν νά ἔχει αὐξηθεῖ τό ποσοστό θνησιμότητας;

4.10. Δύο παῖχτες παίζουν "κορώνα-γράμματα" μέ ἕνα τέλειο νόμισμα. Σέ κάθε "κορώνα" ὁ πρῶτος κερδίζει μία δραχμή. Οἱ παῖχτες ξεκινοῦν τό παιχνίδι μέ *6* δραχμές ὁ καθένας. (*i*) Ποιά ἡ πιθανότητα μετά *6* ρίψεις τοῦ νομίσματος καί οἱ δύο παῖχτες νά ἔχουν τά ἴδια χρήματα; (*ii*) Ποιά ἡ πιθανότητα ὁ ἕνας παίχτης νά κερδίσει ὅλα τά χρήματα στήν 6^{η} ρίψη τοῦ νομίσματος;

4.11. 'Από *100* ἀνθρώπους σέ κάποιο χωριό, οἱ *50* πάντα λένε τήν ἀλήθεια, οἱ *30* πάντα ψεύδονται καί οἱ *20* πάντα ἀρνοῦνται ν'ἀπαντήσουν. Ἕνα δεῖγμα *30* ἀνθρώπων ἐκλέγεται. (*i*) 'Εάν ἡ δειγματοληψία γίνεται μ έ ἐ π α ν ά θ ε σ η, νά βρεθεῖ ἡ πιθανότητα τό δεῖγμα νά περιέχει *10* ἀνθρώπους κάθε κατηγορίας. (*ii*) 'Εάν ἡ δειγματοληψία γίνεται χ ω ρ ί ς ἐ π α ν ά θ ε σ η νά βρεθεῖ ἡ πιθανότητα τό δεῖγμα νά περιέχει ἀκριβῶς *12* ψεῦτες.

4.12. Σέ μία αἴθουσα ἐξετάσεων μέ *50* φοιτητές οἱ *10* ἀντιγράφουν. (*i*) Ποιά ἡ πιθανότητα σέ *5* φοιτητές, πού ἐκλέγονται στήν τύχη,νά μήν ὑπάρχει κανένας πού ἀντιγράφει; (*ii*) Ἄν ἐπαναλάβουμε τό παραπάνω πείραμα τρεῖς φορές ποιά ἡ πιθανότητα δύο φορές νά μή βροῦμε ἀντιγραφέα;

4.13. Ἔστω ὅτι ἀγνοώντας τίς δίδυμες , τρίδυμες κλπ. γεννήσεις οἱ ἁπλές γεννήσεις ἀνθρώπου ἔχουν πιθανότητα *0,5* νά φέρουν ἀγόρι. 'Υποθέτοντας ὅτι οἱ διάφορες γεννήσεις εἶναι ἀνεξάρτητες ὡς πρός τό φύλο, νά βρεθεῖ ἡ πιθανότητα μία οἰκογένεια *6* παιδιῶν νά ἔχει (*i*) ἀκριβῶς *4* ἀγόρια, (*ii*) τουλάχιστον *1* ἀγόρι (*iii*) τό πολύ ἕνα ἀγόρι.

4.14. Εἶναι γνωστό ὅτι ἕνα μέρος ἑνός μεγάλου πληθυσμοῦ ἀνθρώπων πάσχει ἀπό μία ἀσθένεια Α. Ἐπίσης εἶναι γνωστό ὅτι ἄν ἐκλέξουμε *10* ἄτομα ἀπ'τόν πληθυσμό αὐτό, ἡ πιθανότητα νά βροῦμε *5* ἄτομα ἄρρωστα εἶναι διπλάσια τῆς πιθανότητας νά βροῦμε *4* ἄτομα ἄρρωστα. Ποιά ἡ πιθανότητα νά βροῦμε *2* ἄρρωστα ἄτομα ὅταν ἐκλέγουμε *8*;

4.15. Ἔστω ὅτι οἱ τέσσερες μηχανές ἐπιβατικοῦ ἀεροσκάφους ἔχουν τοποθετηθεῖ κατά τέτοιο τρόπο ὥστε νά λειτουργοῦν ἀνεξάρτητα ἡ μία ἀπό τήν ἄλλη καί ὅτι ἡ πιθανότητα βλάβης μίας μηχανῆς, κατά τήν διάρκεια τῆς πτήσεως εἶναι *0,01*. Νά βρεθεῖ ἡ πιθανότητα σέ μία δεδομένη πτήση:(α) Νά μήν παρατηρηθεῖ καμμία βλάβη καί (β) νά μήν παρατηρηθεῖ πάνω ἀπό μία βλάβη.

4.16. Ἕνας ἀντιπρόσωπος βαρέων μηχανημάτων πωλεῖ τά μηχανήματα καί τά ἐξαρτήματα τους μόνο μέ ἀντιπρόσωπους πωλήσεων οἱ ὁποῖοι ἐπισκέπτονται τούς ὑποψήφιους πελάτες. Ἡ πιθανότητα πωλήσεως ἑνός ἐξαρτήματος εἶναι *0,3* καί κάθε πελάτης μπορεῖ νά ἀγοράσει ἕνα μόνο ἐξάρτημα. Ἄν ὁ ἀντιπρόσωπος ἔχει ἀπόθεμα *3* ἐξαρτήματα νά βρεθεῖ ἡ πιθανότητα νά χρειαστοῦν λιγώτερες ἀπό *5* ἐπισκέψεις σέ ὑποψήφιους πελάτες γιά νά ἐξαντληθεῖ τό ἀπόθεμα. Μεταξύ ὑποψήφιων πελατῶν δέν ὑπάρχει καμμία σχέση.

4.17. Ἕνας πύραυλος ἔχει *70%* πιθανότητα πετυχημένης βολῆς. Νά βρεθεῖ ἡ πιθανότητα νά χρειαστοῦν τουλάχιστον *10* ἐκτοξεύσεις γιά *7* πετυχημένες βολές.

4.18. Ἕνα παιχνίδι μπιλλιάρδου ἀποτελεῖται ἀπό κτυπήματα καί ὁ κανονισμός τοῦ παιχνιδιοῦ καθορίζει ὅτι ἕνας παίκτης μπορεῖ νά παίζει συνέχεια ἕως ὅτου χάσει κτύπημα. Ἄς ὑποθέσουμε ὅτι ὁ Γιῶργος κατά κανόνα χάνει *20%* ἀπό τά κτυπήματά του καί ὅτι τά κτυπήματα του εἶναι ἀνεξάρτητα. Νά βρεθεῖ ἡ πιθανότητα ὁ Γιῶργος νά τελειώσει τή σειρά του (*i*) σέ *5* ἀκριβῶς κτυπήματα καί (*ii*) τό πολύ σέ *5* κτυπήματα.

4.19. Σέ ἕναν ἔλεγχο ποιότητας προϊόντων (μέ ἐπανατοποθέτηση) ἡ πιθανότητα νά ἐμφανιστεῖ *5* φορές ἐλαττωματικό ἀντικείμενο σέ *10* δειγματοληψίες εἶναι διπλάσια τῆς πιθανότητας νά ἐμφανιστεῖ ἐλαττωματικό ἀντικείμενο *4* φορές στόν ἴδιο ἀριθμό δειγματοληψιῶν. (*i*) Ποιά ἡ πιθανότητα νά ἐμφανιστεῖ ἐλαττωματικό ἀντικείμενο τουλάχιστον *1* φορά σέ *10* δειγματοληψίες; (*ii*) Ποιά ἡ πιθανότητα τό πρῶτο ἐλαττωματικό ἀντικείμενο νά ἐμφανιστεῖ στήν τέταρτη δειγματοληψία; (*iii*) Ποιά ἡ πιθανότητα τό τρίτο ἐλλατωματικό νά ἐμφανιστεῖ στήν 6η δειγματοληψία;

4.20. Ἕνας νεαρός προσπαθεῖ νά πετύχει ἕνα στόχο μέ πέτρες. Ἡ πιθανότητα ὅτι θά πετύχει τό στόχο σέ μιά ὁποιαδήποτε προσπάθεια εἶναι *0,8*. Ποιά εἶναι ἡ πιθανότητα νά χρειαστοῦν τουλάχιστον *9* δοκιμές γιά νά ἔχει *7* ἐπιτυχίες;

4.21. Ἕνα καινούργιο ἐμβόλιο ἔχει πιθανότητα *70%* ἀνοσοποιήσεως κουνελιῶν ἀπό μία ἀσθένεια. Ἕνας μεγάλος πληθυσμός κουνελιῶν ἐμβολιάζεται καί ὡρισμένα κουνέλια διαλέγονται διαδοχικά γιά ἐργαστηριακή παρακολούθηση.

(α) Ποιά ή πιθανότητα νά βροῦμε ἀκριβῶς *3* ἀνοσοποιημένα κουνέλια ὅταν διαλέξουμε *10*; Πόσα ἀπό τά *10* κουνέλια ἐλπίζουμε νά εἶναι ἀνοσοποιημένα; (β) Ποιά ἡ πιθανότητα νά χρειαστεῖ νά διαλέξουμε *4* κουνέλια μέχρι νά βροῦμε τό πρῶτο ἀνοσοποιημένο;

4.22. Ὁ ἀριθμός τῶν τηλεφωνικῶν κλήσεων πού παίρνει ὁ χειριστής ἑνός τηλεφωνικοῦ κέντρου μεταξύ *9.00* καί *9.10΄* τό πρωί ἀκολουθεῖ κατανομή *Poisson* μέ μέση τιμή *3*. Νά βρεθοῦν (*i*) ἡ πιθανότητα κατά τήν αὐριανή ἡμέρα ὁ τηλεφωνητής νά μήν λάβει καμμιά κλήση στό διάστημα *9.00* καί *9.10΄*. (*ii*) Ἡ πιθανότητα κατά τίς ἑπόμενες τρεῖς ἡμέρες ὁ τηλεφωνητής νά λάβει σ υ ν ο λ ι κ ά μιά κλήση στό ἴδιο διάστημα.

4.23. Ἀπό μία ραδιενεργό οὐσία ἐκπέμπονται κατά μέσο ὅρο *3* σωματίδια α ἀνά λεπτό, σύμφωνα μέ τήν κατανομή τοῦ *Poisson*.(*i*) Ποιά ἡ πιθανότητα σέ *2* λεπτά νά ἐκπεμφθοῦν *4* σωματίδια. (*ii*) Δίνεται ὅτι ὁ χρόνος μεταξύ δύο διαδοχικῶν ἐκπομπῶν ἀκολουθεῖ τήν ἐκθετική κατανομή μέ παράμετρο $\lambda=3$. Νά βρεθεῖ ἡ πιθανότητα ὁ χρόνος μεταξύ δύο διαδοχικῶν ἐκπομπῶν νά εἶναι τό πολύ *2* λεπτά.

4.24. Οἱ ἀφίξεις τηλεφωνικῶν κλήσεων σέ ἕνα Σταθμό Πρώτων Βοηθειῶν ἀκολουθοῦν κατανομή *Poisson* μέ $\lambda=3$ κλήσεις τήν ὥρα. Νά βρεθεῖ ἡ πιθανότητα (*i*) σέ διάστημα *2* ὡρῶν νά ἔχουμε τουλάχιστον *4* κλήσεις καί (*ii*) ὁ χρόνος πού μεσολαβεῖ μεταξύ τῆς 4ης καί 5ης κλήσεως νά εἶναι μικρότερος τῶν *10* λεπτῶν.

4.25. Ἔστω ὅτι ὁ ἀριθμός τῶν ἐλαττωμάτων ἀνά cm^2 σέ φύλλα κοντραπλακέ ἀκολουθεῖ τήν κατανομή *Poisson* μέ ἕνα ἐλάττωμα ἀνά $50\ cm^2$ κατά μέσο ὅρο. Νά βρεθεῖ ἡ πιθανότητα ἕνα φύλλο διαστάσεων $5cm \times 5cm$ νά ἔχει τό πολύ ἕνα ἐλάττωμα.

4.26. Σ'ἕνα πολυσέλιδο κείμενο βρίσκουμε ὅτι μόνο *13,5%* τῶν σελίδων δέν ἔχουν τυπογραφικά λάθη. Ἄν ὁ ἀριθμός τῶν λαθῶν ἀνά σελίδα εἶναι τυχαία μεταβλητή μέ κατανομή τοῦ *Poisson* νά βρεθεῖ τό ποσοστό τῶν σελίδων πού ἔχουν ἀκριβῶς ἕνα λάθος.

4.27. Οἱ ἀφίξεις πελατῶν σέ ἕνα φαρμακεῖο ἀκολουθοῦν τήν κατανομή τοῦ *Poisson* μέ ρυθμό ἀφίξεων *3* πελάτες ἀνά *2* ὧρες κατά μέσο ὅρο. (*i*) Νά βρεθεῖ ἡ πιθανότητα τό φαρμακεῖο νά δεχθεῖ *8* ἕως *10* πελάτες ἀπό τίς *9* π.μ. μέχρι τίς *3* μ.μ. (*ii*) Ἔστω ὅτι ὁ χρόνος ἐξυπηρέτησης κάθε πελάτη εἶναι σταθερός καί ἴσος μέ *15* λεπτά. Ποιά ἡ πιθανότητα τό φαρμακεῖο νά μείνει χωρίς πελάτη περισσότερο ἀπό μισή ὥρα.

4.28. Ὁ ἀριθμός τῶν πελατῶν πού φθάνουν σέ ἕνα κατάστημα, ἀκολουθώντας μία διαδικασία τοῦ *Poisson*, εἶναι κατά μέσο ὅρο *20* τήν ὥρα. Ποιά ἡ πιθανότητα ὁ χρόνος μεταξύ δύο διαδοχικῶν ἀφίξεων νά εἶναι *i*) μικρότερος τῶν *3* λεπτῶν. *ii*) Μεγαλύτερος τῶν *4* λεπτῶν. *iii*) Ποιά ἡ πιθανότητα σέ *30*'τῆς ὥρας νά ἔλθουν *15* πελάτες.

4.29. Ὑποθέτουμε ὅτι ὁ ἀριθμός τῶν τυπογραφικῶν λαθῶν ἀνά σελίδα ἀκολουθεῖ τήν κατανομή *Poisson*. Ἕνα βιβλίο μέ *200* σελίδες ἔχει *40* τυπογραφικά λάθη. Ἄν διαλέξουμε *10* σελίδες ἀπό τό παραπάνω βιβλίο τυχαίως, νά βρεθεῖ ἡ πιθανότητα νά μήν ἔχουν τυπογραφικά λάθη.

4.30. Ἕνας ἠλεκτρονικός μετρητής βακτηριδίων μετράει κατά μέσο ὅρο 5 βακτηρίδια σέ ἕνα cm^3 ἑνός ὑγροῦ. Ὑποθέτοντας ὅτι ὁ ἀριθμός τῶν μετρουμένων βακτηριδίων σέ ἕνα cm^3 ἀκολουθεῖ τήν κατανομή τοῦ *Poisson* νά βρεθεῖ ἡ πιθανότητα ὁ μετρητής νά μετρήσει (*i*) τό πολύ 4 βακτηρίδια σέ 1 cm^3. (*ii*) τουλάχοστον 3 βακτηρίδια.

4.31. Εἶναι γνωστό ὅτι οἱ ἀφίξεις πελατῶν σέ ἕνα ἰατρεῖο ἀκολουθοῦν τήν κατανομή *Poisson* μέ μέση τιμή 6 πελάτες τήν ὥρα. Δεδομένου ὅτι ἡ αἴθουσα ἀναμονῆς τοῦ ἰατρείου ἔχει 4 καθίσματα καί ὅτι ὁ χρόνος ἐξετάσεως κάθε ἀσθενοῦς εἶναι σταθερός καί ἴσος μέ 40', ποιά ἡ πιθανότητα κατά τή στιγμή πού τελειώνει ἡ ἐξέταση τοῦ πρώτου ἀσθενοῦς νά ὑπάρχουν τουλάχιστον δύο ὄρθιοι πελάτες στήν αἴθουσα ἀναμονῆς;

4.32. Ὁ ἀριθμός τῶν ἀκτίνων γάμμα πού ἐκπέμπονται ἀνά δευτερόλεπτο ἀπό μιά ραδιενεργό οὐσία εἶναι μιά τυχαία μεταβλητή μέ κατανομή *Poisson* μέ παράμετρο $\lambda=2$. Ἄν ἕνα ὄργανο πού μετρᾶ τόν ἀριθμό τῶν ἀκτίνων γάμμα, οἱ ὁποῖες ἐκπέμπονται, σταματᾶ νά δουλεύει ὅταν οἱ ἀκτῖνες εἶναι περισσότερες ἀπό 4, ποιά ἡ πιθανότητα νά σταματήσει νά δουλεύει τό παραπάνω ὄργανο;

4.33. Σέ μιά κλινική ὑπολογίζεται ὅτι κατά μέσον ὅρο 3 ἀσθενεῖς τήν ἡμέρα (24 ὧρες) καταφθάνουν καί ζητάνε κρεββάτι. Οἱ ἀφίξεις ἀσθενῶν ἀνά μέρα ἀκολουθοῦν διαδικασία *Poisson*. Ποιά ἡ πιθανότητα ὁ χρόνος μεταξύ τῶν ἀφίξεων τοῦ πρώτου καί τοῦ δεύτερου ἀσθενῆ νά εἶναι μικρότερος τῶν 4 ὡρῶν.

4.34. Μέ τή βοήθεια ἕνός τυχαίου μηχανισμοῦ πού ἀκολουθεῖ τήν ὁμοιόμορφη κατανομή στό διάστημα $(0,1)$ ἐκλέγεται ἕνας ἀριθμός ἀπό τό διάστημα αὐτό. Νά βρεθεῖ ἡ πιθανότητα *i*) τό δεύτερο δεκαδικό ψηφίο τοῦ ἀριθμοῦ νά εἶναι τό *2* καί *ii*) τό δεύτερο δεκαδικό ψηφίο τῆς τετραγωνικῆς ρίζας τοῦ ἀριθμοῦ νά εἶναι τό *5*.

4.35. 'Ο χρόνος ζωῆς τῶν κατοίκων μίας χώρας εἶναι ἐκθετική τυχαία μεταβλητή μέ $\lambda=1/50$ (σέ ἔτη). Νά βρεθεῖ τό ποσοστό τῶν κατοίκων πού ζοῦν πάνω ἀπό *70* χρόνια.

4.36. 'Υποθέτουμε ὅτι ἡ διάρκεια τῆς χιονόπτωσης σέ ἕνα χωριό τῆς Β. 'Ελλάδας εἶναι μία τυχαία μεταβλητή T μέ συνάρτηση πυκνότητας πιθανότητας

$$f_T(t) = \begin{cases} \frac{1}{5}\, e^{-t/5} & , \quad 0 < t < \infty \\ 0 & , \quad \text{ἀλλοῦ} \;, \end{cases}$$

ὅπου τό t μετρᾶται σέ λεπτά. Ἄν ἡ διάρκεια τῆς χιονόπτωσης κατά μία μέρα, ἔχει φθάσει ἤδη στά *8* λεπτά, ποιά εἶναι ἡ πιθανότητα νά διαρκέσει ἄλλα *6* λεπτά ἀκόμα;

4.37. 'Η διάρκεια σέ λεπτά τῶν ὑπεραστικῶν τηλεφωνικῶν συνδιαλέξεων ἀκολουθεῖ τήν ἐκθετική κατανομή μέ ἀθροιστική συνάρτηση κατανομῆς

$$F(x) = 1-e^{-x/5} \;, \quad x \geq 0.$$

Νά βρεθεῖ ἡ πιθανότητα ἡ διάρκεια μίας ὑπεραστικῆς συνδιαλέξεως (i) νά ὑπερβεῖ τά 5 λεπτά (ii) νά εἶναι μικρότερη τῶν 6 λεπτῶν δεδομένου ὅτι ἦταν μεγαλύτερη τῶν 3 λεπτῶν.

4.38. Εἶναι γνωστό ὅτι ὁ χρόνος ζωῆς τοῦ ἰοῦ τῆς γρίππης μέσα στόν ὀργανισμό ἑνός ἀτόμου τό ὁποῖο ὑφίσταται ἀντιγριπική θεραπεία ἀκολουθεῖ τήν ἐκθετική κατανομή μέ μέση τιμή 3 μέρες. (i) Ποιά ἡ πιθανότητα ἀπό 10 ἄτομα πού ἔχουν προσβληθεῖ ἀπό τόν ἰό τά 3 νά γίνουν καλά σέ χρονικό διάστημα ἀπό 2 ἕως 4 μέρες; (ii) Ποιά ἡ πιθανότητα ἕνας ἀσθενής νά γίνει καλά συνολικά σέ λιγότερο ἀπό 5 μέρες δεδομένου ὅτι ἔχει 2 μέρες ἄρρωστος.

4.39. Ἕνας γιατρός μέ πολύ μεγάλο ἀριθμό πελατῶν θέλει νά κανονίσει τά ραντεβοῦ τῶν ἀσθενῶν του ὥστε ὁ χρόνος ἀναμονῆς τους στόν προθάλαμο τοῦ ἰατρείου νά μήν εἶναι μεγάλος. Δοθέντος ὅτι ὁ χρόνος ἐξετάσεως κάθε πελάτη ἀκολουθεῖ τήν ἐκθετική κατανομή μέ μέση τιμή 30 λεπτά καί ὅτι ὁ γιατρός κανονίζει τά ραντεβοῦ του κάθε 40 λεπτά, ποιά ἡ πιθανότητα ὁ δεύτερος πελάτης κάθε μέρας νά περιμένει στόν προθάλαμο μεταξύ 5 καί 10 λεπτά .Δίδεται ὅτι οἱ ἀσθενεῖς φθάνουν στό ἰατρεῖο ἀκριβῶς τήν καθορισμένη ὥρα τοῦ ραντεβοῦ.

4.40. Ὁ χρόνος ζωῆς τῶν ἰνδικῶν χοιριδίων τά ὁποῖα ἐμβολιάζονται μέ ἕνα μολυσματικό ἰό ἀκολουθεῖ τήν ἐκθετική κατανομή μέ μέση τιμή 1 χρόνο. Νά βρεθεῖ ἡ πιθανότητα (i) ἕνα χοιρίδιο τό ὁποῖο ἐμβολιάζεται μέ τόν ἰό νά ζήσει μεταξύ 8 καί 10 μῆνες καί (ii) τό χοιρίδιο νά ζήσει λιγότερο άπό 9 μῆνες δεδομένου ὅτι ἔζησε περισσότερο ἀπό 5.

4.41. 'Ο χρόνος ζωῆς τῶν βρεφῶν τά ὁποῖα γεννιοῦνται μέ μία ἀνίατη νόσο A ἀκολουθεῖ τήν ἐκθετική κατανομή μέ μέση τιμή 1 χρόνο. (i) Νά βρεθεῖ ἡ πιθανότητα ἕνα βρέφος πού γεννιέται μέ τήν ἀσθένεια A νά ζήσει μεταξύ 8 καί 10 μῆνες. (ii) Νά βρεθεῖ ἡ πιθανότητα τό βρέφος νά ζήσει λιγότερο ἀπό 9 μῆνες δεδομένου ὅτι ἔζησε περισσότερο ἀπό 5 μῆνες. (iii) Ποιά ἡ πιθανότητα ἀπό 10 τέτοια ἀσυσχέτιστα βρέφη ἀκριβῶς τρία νά ζήσουν 12 μῆνες;

4.42. 'Ο χρόνος ζωῆς τῶν λαμπτήρων χειρουργικοῦ θαλάμου ἀκολουθεῖ τήν ἐκθετική κατανομή μέ $\lambda=0,2$ (σέ ὧρες). Γιά τήν ἐπιτυχή διεκπεραίωση μίας ἐγχειρήσεως διάρκειας 10 ὡρῶν χρειάζεται ἡ λειτουργία τουλάχιστον 2 ἀπό τούς 4 ὑπάρχοντες λαμπτῆρες. Ποιά ἡ πιθανότητα ἐπιτυχοῦς διεκπαιρεώσεως τῆς ἐγχειρήσεως;

4.43. 'Ο χρόνος ζωῆς ἑνός πλυντηρίου ἔχει κατά προσέγγιση κανονική κατανομή μέ μέση τιμή $\mu=3,1$ ἔτη καί τυπική ἀπόκλιση $\sigma=1,2$ ἔτη. Ἄν τό πλυντήριο ἔχει ἐγγύηση ἑνός ἔτους νά βρεθεῖ τό ποσοστό τῶν πλυντηρίων πού θά χρειαστεῖ νά ἀντικατασταθοῦν.

4.44. 'Ο χρόνος πού χρειάζεται γιά νά δοθοῦν οἱ ἀπαντήσεις σέ ἕνα γραπτό τέστ ἀκολουθεῖ κανονική κατανομή μέ μέση τιμή $\mu=70$ min καί τυπική ἀπόκλιση $\sigma=12$ min. Πότε πρέπει νά τελειώσει τό διαγώνισμα ἄν θέλουμε νά δώσουμε ἀρκετό χρόνο ὥστε τό 90% τῶν ὑποψηφίων νά προλάβουν νά ὁλοκληρώσουν τό τέστ;

4.45. Μία ὑπηρεσία χρησιμοποιεῖ πολλούς ἠλεκτρικούς λαμπτῆρες οἱ ὁποῖοι εἶναι ἀναμμένοι συνέχεια μέρα καί νύχτα. Ὁ χρόνος ζωῆς κάθε λαμπτήρα θεωρεῖται ὅτι ἀκολουθεῖ κανονική κατανομή μέ μέση τιμή $\mu=60$ μέρες καί τυπική ἀπόκλιση $\sigma=20$ μέρες. Τήν 1η Ἰανουαρίου ἡ ὑπηρεσία τοποθετεῖ 10.000 νέους λαμπτῆρες. Πόσοι λαμπτῆρες ἀναμένεται νά χρειάζονται ἀντικατάσταση μέχρι τήν 1η Φεβρουαρίου τοῦ ἴδιου ἔτους;

4.46. Ὁ δείκτης εὐφυΐας γιά κάποιο πληθυσμό ἔχει κανονική κατανομή μέ μέση τιμή $\mu=100$ καί τυπική ἀπόκλιση $\sigma=15$. Ποιό ποσοστό τοῦ πληθυσμοῦ ἔχει δείκτη εὐφυΐας μικρότερο τοῦ 90 καί ποιό ποσοστό μεγαλύτερο τοῦ 125;

4.47. Μία βιομηχανία κατασκευάζει παιδικές τροφές τῶν ὁποίων τό βάρος ἀκολουθεῖ τήν κανονική κατανομή μέ μέση τιμή 900gr καί τυπική ἀπόκλιση 50gr. Σέ ἕνα ἀγορανομικό ἔλεγχο ὁ ἐλεγκτής διαλέγει στήν τύχη 40 κουτιά ἀπό τήν παραγωγή μίας μέρας καί τά ζυγίζει ὅλα μαζί. Ποιά εἶναι ἡ πιθανότητα τό συνολικό βάρος τῶν 40 κουτιῶν νά εἶναι μικρότερο τῶν 35.500gr;

4.48. Ἡ ποσότητα τοῦ πετρελαίου πού καίει ἕνας καυστήρας σέ 24 ὧρες εἶναι τυχαία μεταβλητή μέ κανονική κατανομή, $\mu=200$ λίτρα, $\sigma=20$. Ἄν ὑπάρχουν ἀκόμη 210 λίτρα στό ντεπόζιτο ποιά ἡ πιθανότητα νά διαρκέσουν γιά 24 ὧρες;

ΚΕΦΑΛΑΙΟ 5

μαθηματικη ελπιδα-χαρακτηριστικα κατανομων και τυχαιων μεταβλητων-ροπογεννητριες

Στό Κεφάλαιο αὐτό παρουσιάζουμε τήν ἔννοια τῆς μαθηματικῆς ἐλπίδας μίας τυχαίας μεταβλητῆς. Παράλληλα ἀναπτύσσουμε τίς ἔννοιες τῆς μέσης τιμῆς, διακύμανσης καί ροπῆς μίας τ.μ. καί μελετᾶμε τίς ἰδιότητες τους. Στό τέλος περιγράφουμε διάφορα ἀριθμητικά ἤ ποσοτικά χαρακτηριστικά τῶν κατανομῶν καί τῶν τυχαίων μεταβλητῶν. Ἀπό αὐτά τό πιό σημαντικό εἶναι ἡ ροπογεννήτρια συνάρτηση ἡ ὁποία ἔχει ἀξιόλογες ἰδιότητες καί μᾶς προσφέρει ἕνα πολύτιμο ἐργαλεῖο γιά τή μελέτη τῆς θεωρίας τῶν κατανομῶν πιθανότητας.

5.1 Μαθηματική ἐλπίδα

Ἡ ἔννοια τῆς ἀναμενόμενης τιμῆς ἤ μαθηματικῆς ἐλπίδας μίας τυχαίας μεταβλητῆς ὁδηγεῖ σέ ἕνα ἀπό τά σημαντικώτερα χαρακτηριστικά της γνωρίσματα. Συνδέεται μέ τήν ἔννοια τῆς μέσης τιμῆς μίας σειρᾶς μετρήσεων καί ἀπό μία ἄποψη εἶναι ἡ ἐπικρατέστερη τιμή τῆς τ.μ. X.

Ἔστω μία διακριτή τυχαία μεταβλητή X ἡ ὁποία παίρνει τιμές $x_1, x_2, \ldots, x_k$ μέ πιθανότητες $p(x_1), p(x_2), \ldots, p(x_k)$ ἀντίστοιχα ὅπου $p(x)$ εἶναι ἡ συνάρτηση πιθανότητας τῆς X. Ἐάν τό τυχαῖο πείραμα πού συνδέεται μέ τήν X ἐπαναληφθεῖ ἀνεξάρτητα ἕνα μεγάλο ἀριθμό φορῶν, θά περιμέναμε οἱ διάφορες τιμές τῆς X νά ἐμφανιστοῦν σέ ποσοστά φορῶν πού θά εἶναι κατά προσέγγιση ἴσα πρός τίς πιθανότητές τους. Ἐάν δέ ἐπί πλέον γιά κάθε τιμή x_i τῆς X ὑπάρχει καί κέρδος x_i δραχμῶν τότε τό ἀναμενόμενο κέρδος ἀπό κάθε τιμή x_i θά εἶναι $x_i p(x_i)$ καί τό συνολικό ἀναμενόμενο κέρδος $\sum_{i=1}^{k} x_i p(x_i)$.

Γιά παράδειγμα, ἔστω ὅτι ρίχνουμε ἕνα τέλειο νόμισμα Ἐάν κερδίζουμε 1 δραχμή ὅταν ἐμφανίζεται ἡ πλευρά K καί χάνουμε 1 δραχμή (δηλ. κερδίζουμε -1 δραχμή) ὅταν ἐμφανίζεται ἡ πλευρά Γ, τότε σέ μία ρίψη τοῦ νομίσματος περιμένουμε νά κερδίσουμε $1\times(1/2)$ δραχμές ἀπό τήν πλευρά K καί $(-1)\times(1/2)$ ἀπό τήν πλευρά Γ, δηλαδή, συνολικά

$$1\times(1/2) + (-1)\times(1/2) = 0$$

κέρδος. Ἐάν ἐπίσης ρίξουμε ἕνα τέλειο ζάρι 6 φορές καί σέ κάθε ἀποτέλεσμα i κερδίζουμε i δραχμές, $i=1,2,3,\ldots,6$, τότε θά περιμένουμε νά ἐμφανιστεῖ μία φορά 1, δηλαδή νά κερδίσουμε 1 δραχμή, μία φορά 2, δηλαδή νά κερδίσουμε 2 δραχμές κ.ο.κ. Τό ἀναμενόμενο κέρδος μας στίς 6 ρίψεις θά εἶναι

$$1+2+3+4+5+6 = 6\times\frac{1}{6}+12\times\frac{1}{6}+18\times\frac{1}{6}+24\times\frac{1}{6}+30\times\frac{1}{6}+36\times\frac{1}{6} = 21 .$$

Οἱ ἰδέες αὐτές μᾶς ὁδηγοῦν στόν ἀκόλουθο ὁρισμό. Τό ἀναμενόμενο κέρδος μας στή μία ρίψη θά εἶναι $21/6$ δηλαδή

$$\frac{21}{6} = 1\times\frac{1}{6} +2\times\frac{2}{6}+\ldots+6\times\frac{6}{6} = \Sigma\, xp(x) .$$

Ὁρισμός 5.1: Ἀ ν α μ ε ν ό μ ε ν η τ ι μ ή ἤ μ α θ η μ α τ ι κ ή ἐ λ π ί δ α τ υ χ α ί α ς μ ε τ α β λ η τ ῆ ς. Ἔστω X μία τυχαία μεταβλητή, τότε ἡ ἀ ν α μ ε ν ο μ ε ν η τ ι μ ή ἤ μ α θ η μ α τ ι κ ή ἐ λ π ί δ α τῆς X συμβολίζεται μέ $E(X)$ καί ὁρίζεται μέ τή σχέση

$$E(X) = \begin{cases} \sum\limits_{x} xp(x) , & \text{ἐάν ἡ τ.μ. } X \text{ εἶναι ἀπαριθμητή} \\ & \text{μέ κατανομή } p(x) . \\ \int\limits_{-\infty}^{+\infty} xf(x)dx, & \text{ἐάν ἡ τ.μ. } X \text{ εἶναι συνεχής} \\ & \text{μέ κατανομή } f(x). \end{cases} \qquad (5.1)$$

Ἡ ἀναμενόμενη τιμή τῆς X συμβολίζεται καί μέ EX, μ_x ἤ ἁπλᾶ μ καί λέγεται συνήθως μέση τιμή τῆς τ.μ. X.

Ὁ παραπάνω ὁρισμός ἰσχύει ὑπό τήν προϋπόθεση ὅτι

τό ἄθροισμα $\sum\limits_{x} xp(x)$ ἤ τό ὁλοκλήρωμα $\int\limits_{-\infty}^{+\infty} xf(x)dx$

συγκλίνουν ἀπόλυτα, δηλαδή

$$\sum_{x}|x|p(x) < \infty \quad \text{ἤ} \quad \int_{-\infty}^{+\infty} |x|f(x)dx < \infty . \qquad (5.2)$$

Τότε λέμε ὅτι ὑπάρχει ἡ μέση τιμή EX τῆς τυχαίας μεταβλητῆς X. Αὐτό εἶναι ἀπαραίτητο διότι, ἄν π.χ. $\Sigma|x|p(x) = \infty$ τότε τό ἄθροισμα $\sum_x xp(x)$ θά παρίστανε διαφορετικούς ἀριθμούς ἀνάλογα μέ τόν τρόπο ἀθροίσεως(ἀναδιατάσσοντας τούς ὅρους του θά ἄλλαζε τιμή). Γιά παράδειγμα ἔστω X τ.μ. μέ τιμές $(-2)^x$ καί πιθανότητα $1/2^x$ γιά κάθε τιμή, $x = 1,2,\ldots$ Τότε

$$\sum_{x=1}^{\infty} |x|p(x) = \sum_{x=1}^{\infty} |(-2)^x|(1/2^x) = \sum_{x=1}^{\infty} 1 = \infty$$

καί

$$EX = (-2)\frac{1}{2} +(-2)^2 \frac{1}{2^2} +(-2)^3 \frac{1}{2^3} + \ldots$$

$$= -1+1-1+1- \ldots$$

πού κυμαίνεται μεταξύ -1 καί 0.

Παραδείγματα:

5.1: Χίλιοι λαχνοί τῶν 10 δραχμῶν ὁ κάθε ἕνας πωλοῦνται σέ μιά λαχειοφόρο ἀγορά τῆς ὁποίας τό μοναδικό βραβεῖο εἶναι ἕνα ραδιόφωνο ἀξίας 3000 δραχμῶν. Ἐάν ὁ Γιῶργος ἀγοράσει δυό λαχεῖα ποιό εἶναι τό ἀναμενόμενο κέρδος του;

Ἔστω X ἡ τ.μ. πού ἐκφράζει τό κέρδος τοῦ Γιώργου. Ἐάν ὁ Γιῶργος κερδίσει, τό κέρδος του θά εἶναι $x = 3000 -(2\times 10) = 2980$ δραχμές. Ἐάν χάσει, τό κέρδος του θά εἶναι $x = -20$ δραχμές. Οἱ ἀντίστοιχες πιθανότητες εἶναι

$2/1000 = 0{,}002$ καί $1-0{,}002 = 0{,}998$. Ἄρα τό ἀναμενόμενο κέρδος τοῦ Γιώργου θά εἶναι

$$E(X) = \sum_x xp(x) = (2980\times 0{,}002)+(-20)\times(0{,}998) = -14 ,$$

δηλαδή ἀπώλεια 14 δραχμές.

5.2: Νά βρεθεῖ ὁ ἀναμενόμενος χρόνος ἀπορροφήσεως ἑνός φαρμάκου ὅταν ἡ κατανομή τοῦ χρόνου αὐτοῦ T εἶναι

$$f(t) = \begin{cases} \dfrac{1}{10} , & 5 \leq t \leq 15 \\ 0 , & \text{ἀλλοῦ}. \end{cases}$$

Ἔχουμε

$$E(T) = \int_{-\infty}^{+\infty} t\, f(t)dt = \int_5^{15} t\, \frac{1}{10}\, dt = \frac{1}{10}\, \frac{t^2}{2}\Big|_5^{15} = 10$$

λεπτά τῆς ὥρας, πρᾶγμα πού περιμέναμε λόγω τοῦ ὅτι ἡ κατανομή τῆς τ.μ. T εἶναι ὁμοιόμορφη.

5.3: Ἔστω X τυχαία μεταβλητή μέ σ.π.π.

$$f(x) = \begin{cases} 1/x^2 , & 1 < x < \infty \\ 0 , & \text{ἀλλοῦ}. \end{cases}$$

Ἡ τ.μ. X δέν ἔχει μέση τιμή. Πράγματι

$$= EX = \int_1^{+\infty} x \frac{1}{x^2}\, dx$$

$$= \lim_{t\to+\infty} \int_1^t \frac{dx}{x} = \lim_{t\to+\infty} [\ln t - \ln 1] = \infty .$$

Ὅπως ἀναφέραμε στό ἐδάφιο 3.5 συναρτήσεις τυχαίων μεταβλητῶν εἶναι ἐπίσης τυχαῖες μεταβλητές. Ἀνάλογα ὁρίζεται καί ἡ μαθηματική ἐλπίδα μίας συναρτήσεως $h(X)$ μίας τυχαίας μεταβλητῆς X.

Ὁρισμός 5.2: Ἀ ν α μ ε ν ό μ ε ν η τ ι μ ή ἤ μ α θ η μ α τ ι κ ή ἐ λ π ί δ α τ ῆ ς $h(X)$. Ἔστω $h(X)$ μία συνάρτηση μίας τ.μ. X. Ἡ ἀ ν α μ ε ν ό μ ε ν η τ ι μ ή ἤ μ α θ η μ α τ ι κ ή ἐ λ π ί δ α $E[h(X)]$ τῆς $h(X)$ ὁρίζεται μέ τή σχέση

$$E[h(X)] = \begin{cases} \sum_x h(x)p_X(x) & \text{ἄν ἡ τ.μ. } X \text{ εἶναι διακριτή μέ σ.π. } p_X(x) \\ \int_{-\infty}^{+\infty} h(x)f_X(x)dx & \text{ἄν ἡ τ.μ. } X \text{ εἶναι συνεχής μέ σ.π.π. } f_X(x) . \end{cases} \qquad (5.3)$$

Ἡ ἀναμενόμενη τιμή τῆς $h(X)$ συμβολίζεται καί μέ $Eh(X)$ καί ὑπάρχει ὑπό τόν ὅρο ὅτι τό ἄθροισμα ἤ τό ὁλοκλήρωμα στήν (5.3) συγκλίνουν ἀπόλυτα.

5. 2 'Ιδιότητες τῆς μαθηματικῆς ἐλπίδας

Οἱ βασικές ἰδιότητες τῆς μαθηματικῆς ἐλπίδας μίας τυχαίας μεταβλητῆς δίνονται στό ἀκόλουθο θεώρημα.

Θεώρημα 5.1:

"Εστω X τυχαία μεταβλητή, $a, b, c, c_1, c_2, \ldots, c_n$ σταθερές ποσότητες καί $h(X)$, $h_1(X), \ldots, h_n(X)$ συναρτήσεις τῆς X. Τότε

1. $$E(c) = c\ .$$

2. $$E(aX+b) = aE(X)+b$$

καί γενικά

$$E[ah(X)+b] = aE[h(X)]+b\ . \tag{5.4}$$

3. $$E[h_1(X)+h_2(X)] = E[h_1(X)]+E[h_2(X)]$$

καί γενικά

$$E\left[\sum_{i=1}^{k} c_i h_i(X)\right] = \sum_{i=1}^{k} c_i E[h_i(X)]. \tag{5.5}$$

4. "Αν $X \geqq 0$ τότε $E(X) \geqq 0$

καί γενικά

ἄν $h_1(X) \geqq h_2(X)$ τότε $E[h_1(X)] \geqq E[h_2(X)]$.

5. $$|E[h(X)]| \leqq E|h(X)|.$$

6. <u>Ἄν ἡ</u> $E(X^n)$ <u>ὑπάρχει τότε ὑπάρχει ἡ</u> $E(X^m)$ <u>γιά κάθε</u> $m = 1,2,\ldots,n-1$.

Ἀπόδειξη: Οἱ παραπάνω ἰδιότητες εἶναι ἁπλές συνέπειες τῶν ὁρισμῶν τῶν ἰδιοτήτων πεπερασμένων ἀθροισμάτων ἤ ὁλοκληρωμάτων. Ἔτσι γιά τήν ἰδιότητα 3 ἔχουμε στή συνεχή περίπτωση.

$$
\begin{aligned}
E[h_1(X)+h_2(X)] &= \int_{-\infty}^{+\infty} [h_1(x)+h_2(x)]f_X(x)dx \\
&= \int_{\infty}^{\infty} h_1(x)f_X(x)dx+\int_{-\infty}^{+\infty} h_2(x)f_X(x)dx \\
&= E[h_1(X)]+E[h_2(X)].
\end{aligned}
$$

Γιά τή γενική ἰδιότητα 4 ἔχουμε

$$h_1(X) \geq h_2(X) \Longrightarrow h_1(X)-h_2(X) \geq 0$$

καί ἀπό τό πρῶτο μέρος τῆς 4 ἕπεται

$$E[h_1(X)-h_2(X)] \geq 0 \qquad Eh_1(X) \geq Eh_2(X) .$$

Στήν ἰδιότητα 6, ἡ ὕπαρξη τῆς $E(X^n)$ σημαίνει ὅτι

$$E|X|^n < \infty.$$

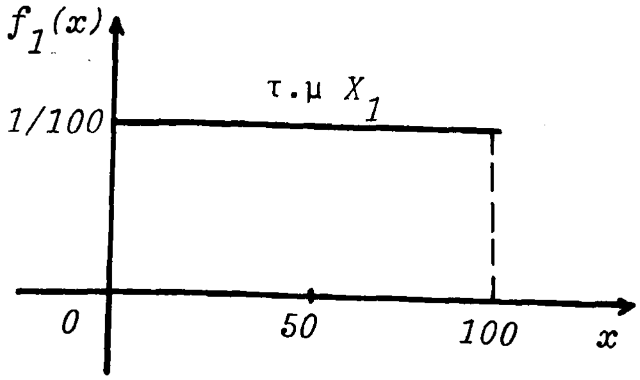
Ἐπί πλέον ἔχουμε

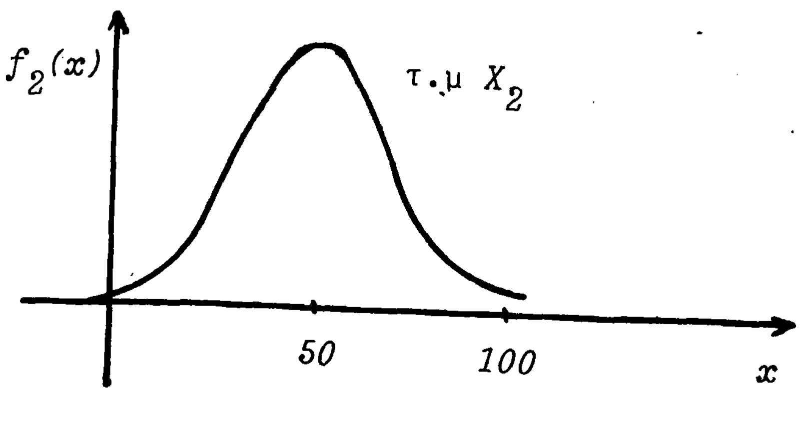
$$|X|^m \leq 1+|X|^n$$

γιά κάθε X καί κάθε $m = 1,2,\ldots,n-1$. Ἔτσι ἐφαρμόζοντες τήν ἰδιότητα 4 παίρνουμε τό ἀποτέλεσμα. ▼

5. 3 Διακύμανση (διασπορὰ) - τυπικὴ ἀπόκλιση

Ἡ μέση τιμή μίας τυχαίας μεταβλητῆς εἶναι ἕνα χρήσιμο χαρακτηριστικό γνώρισμα αὐτῆς. Μᾶς δίνει τή θέση ἤ τό "κέντρο βάρους" τῆς κατανομῆς καί ἔτσι εἶναι γνωστή καί ὡς μ έ τ ρ ο θ έ σ ε ω ς α ὐ τ ῆ ς. Παρά ταῦτα οἱ πληροφορίες πού μᾶς δίνει εἶναι περιορισμένης μορφῆς ὅπως φαίνεται ἀπό τό παρακάτω παράδειγμα δύο τυχαίων μεταβλητῶν X_1 καί X_2 μέ κατανομές $f_1(x)$ καί $f_2(x)$ οἱ ὁποῖες ἔχουν τό ἴδιο πεδίο τιμῶν ἀλλά εἶναι τελείως διαφορετικῆς μορφῆς.





Γιά τή μεταβλητή X_1 ὅλες οἱ τιμές μεταξύ 0 καί 100 εἶναι ἰσοπίθανες, ἐνῶ γιά τό X_2 οἱ τιμές κοντά στό 0 ἤ 100 εἶναι ἀρκετά ἀπίθανο νά πραγματοποιηθοῦν. Ἡ διαφορά μετξύ μίας τιμῆς τῆς X_1 καί τῆς μέσης τιμῆς $\mu = 50$ εἶναι πιθανό νά εἶναι μεγάλη, ἐνῶ ἡ ἴδια διαφορά γιά τό X_2 εἶναι ἀπίθανο νά εἶναι μεγάλη. Ἔτσι ὁδηγούμαστε στό νά ρωτήσουμε κατά πόσο ἀναμένεται μία τυχαία μεταβλητή νά ἀποκλίνει ἀπό τή μέση της τιμή.

Μέ ἄλλα λόγια χρειαζόμαστε ἕνα μέτρο τῆς μεταβλητότητας ἤ διασπορᾶς τῶν τιμῶν τῆς τ.μ. Ἐπειδή οἱ ἀποκλίσεις ἀπό τή μέση τιμή μπορεῖ νά εἶναι θετικές ἤ ἀρνητικές χρησιμοποιοῦμε τό τετράγωνο τῶν ἀποκλίσεων. Ἡ διακύμανση (διασπορά) μίας τυχαίας μεταβλητῆς ὁρίζεται μέ τή βοήθεια τῆς μαθηματικῆς ἐλπίδας.

Ὁρισμός 5.3: Διακύμανση (διασπορά) τυχαίας μεταβλητῆς. Ἔστω X μία τυχαία μεταβλητή. Ἡ διακύμανση ἤ διασπορά *(variance)* τῆς X συμβολίζεται μέ $Var(X)$ ἤ σ_X^2 ἤ ἁπλῶς σ^2 καί ὁρίζεται μέ τή σχέση

$$Var(X) = E(X-\mu)^2 = \begin{cases} \sum\limits_x (x-\mu)^2 p(x)\,, & \text{ἐάν } X \text{ ἀπαριθμητή μέ κατανομή } p(x) \\ \int\limits_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx, & \text{ἐάν } X \text{ συνεχής μέ κατανομή } f(x) \end{cases} \qquad (5.6)$$

ὅπου $\mu = EX$.

Ὁρισμός 5.4: Τυπική ἀπόκλιση τυχαίας μεταβλητῆς. Ἡ τυπική ἀπό-κ λ ι σ η σ_X , ἢ ἁπλῶς σ, μιᾶς τ.μ. X ὁρίζεται ὡς ἡ θετική τετραγωνική ρίζα τῆς διακυμάνσεως αὐτῆς, δηλαδή

$$(\quad \sigma_X = \sqrt{Var(X)} \; .$$

Ἡ διακύμανση καί ἡ τυπική ἀπόκλιση μιᾶς συναρτήσεως $h(X)$ μιᾶς τυχαίας μεταβλητῆς X ὁρίζονται ἀνάλογα:

$$Var[h(X)] = E[h(X) - Eh(X)]^2$$
$$\sigma_{h(X)} = \sqrt{Varh(X)} \; . \qquad (5.7)$$

Ἀντί τῶν παραπάνω συμβολισμῶν χρησιμοποιοῦμε καί τούς ἀκολούθους ἁπλούστερους: $Var\ h(X)$ ἢ $\sigma^2_{h(X)}$.

Ἡ διακύμανση σ^2 εἶναι πάντα μή ἀρνητική καί μετρᾶται μέ τό τετράγωνο τῶν μονάδων μετρήσεως τῆς X, ἐνῶ ἡ τυπική ἀπόκλιση μέ τίς ἴδιες μονάδες ὅπως καί ἡ X. Εἶναι ἐπίσης φανερό ὅτι ἄν ἡ διακύμανση σ^2 μιᾶς τυχαίας μεταβλητῆς X ἰσοῦται μέ τό μηδέν τότε ἡ X ἐκφυλίζεται σέ μία σταθερή τιμή, τή μέση τιμή, μέ πιθανότητα ἴση μέ τή μονάδα.

Οἱ βασικές ἰδιότητες τῆς διακύμανσης δίνονται στό παρακάτω θεώρημα. Ἀπό αὐτές ἡ πρώτη εἶναι πολύ χρήσιμη διότι διευκολύνει τόν ὑπολογισμό τῆς διακύμανσης μιᾶς τυχαίας μεταβλητῆς.

Θ ε ώ ρ η μ α 5.2:

Ἔστω X τυχαία μεταβλητή, a,b,c σταθερές καί $h(X)$ μία συνάρτηση τοῦ X. Τότε

1. $$Var(X) = E(X^2)-(EX)^2$$

καί γενικά

$$Var[h(X)] = E[h(X)]^2-[Eh(X)]^2. \qquad (5.8)$$

2. $$Var\ (c) = 0$$

3. $$Var(aX+b) = a^2 Var(X)$$

<u>καί γενικά</u>

$$Var[ah(X)+b] = a^2 Var\ h(X). \qquad (5.9)$$

'Απόδειξη: 'Αποδεικνύουμε τίς είδικές μόνο περιπτώσεις. Οἱ γενικές ἀποδεικνύονται ἀνάλογα. Γιά τήν ἰδιότητα *1* ἔχουμε

$$\begin{aligned} Var\ X &= E(X-\mu)^2 = E(X^2-2\mu X+\mu^2) \\ &= EX^2-E(2\mu X)+E(\mu^2) \\ &= EX^2-2\mu^2+\mu^2 = EX^2-\mu^2 \\ &= EX^2-(EX)^2 . \end{aligned}$$

Χρησιμοποιήσαμε τίς ἰδιότητες *1* καί *2* τοῦ θεωρήματος 5.1.

'Η ἰδιότητα *2* εἶναι προφανής διότι ἄλλωστε σταθερές ποσότητες δέν ἔχουν μεταβλητότητα.

Γιά τήν ἰδιότητα *3* ἔχουμε

$$\begin{aligned} Var\ (aX+b) &= E[aX+b-E(aX+b)]^2 \\ &= E(aX+b-aEX-Eb) \\ &= E[a(X-Ex)]^2 = a^2E(X-EX)^2 \\ &= a^2 Var\ X\ . \ \blacktriangledown \end{aligned}$$

<u>Παραδείγματα:</u>

5.4: Νά βρεθεῖ ἡ διακύμανση σ^2 καί ἡ τυπική ἀπόκλιση σ τῆς τυχαίας μεταβλητῆς T τοῦ Παραδείγματος 5.2.

Ἔχουμε

$$\sigma^2 = ET^2-(ET)^2 ,$$

καί

$$ET^2 = \int_{-\infty}^{+\infty} t^2 f(t)dt = \int_5^{15} t^2 \frac{1}{10}\, dt$$

$$= \frac{1}{10}\frac{t^3}{3}\Big|_5^{15} = \frac{325}{3} .$$

Ἔτσι

$$\sigma^2 = \frac{325}{3} - 10^2 = \frac{25}{3} \approx 8{,}33$$

καί

$$\sigma = \sqrt{\frac{25}{3}} \approx 2{,}9 .$$

5.5: Ἔστω X συνεχής τ.μ. μέ σ.π.π.

$$f(x) = \begin{cases} \frac{1}{2}(x+1) & , \quad -1 < x < 1 \\ 0 & , \quad \text{ἀλλοῦ} . \end{cases}$$

Ἡ μέση τιμή της εἶναι

$$\mu = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-1}^{1} x(\frac{x+1}{2})dx = 1/3 .$$

Ἡ διακύμανση θά βρεθεῖ ἀπό τόν τύπο

$$Var(X) = E(X^2)-(EX)^2 = E(X^2)-\mu^2 .$$

Ἔχουμε

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_{-1}^{1} x^2 (\frac{x+1}{2})dx = 3/9 .$$

Ἔτσι

$$Var(X) = \frac{3}{9} - \frac{1}{9} = \frac{2}{9} .$$

Ἡ διακύμανση μετρᾶ τήν μεταβλητότητα μίας τυχαίας μεταβλητῆς. Ἐάν εἶναι μηδέν τότε ἡ τυχαία μεταβλητή εἶναι σταθερά. Ὅσο πιό μικρή εἶναι ἡ διακύμανση τόσο πιό μικρή εἶναι καί ἡ μεταβλητότητα τῆς X. Τό ἑπόμενο θεώρημα ἐκτός ἀπό τή γενική μαθηματική του σημασία μᾶς δίνει καί ἕνα χαρακτηριστικό τρόπο γιά νά ἐπεξηγήσουμε τίς δύο αὐτές ἔννοιες.

Θεώρημα 5.3: (Θεώρημα ἤ ἀνισότητα τοῦ Chebyshev).

Ἔστω X μία τυχαία μεταβλητή μέ μέση τιμή μ καί πεπερασμένη διακύμανση σ^2. Τότε γιά κάθε $\varepsilon > 0$

$$P\{|X-\mu| \geq \varepsilon\} \leq \frac{Var(X)}{\varepsilon^2} . \qquad (5.10)$$

Ἀπόδειξη: Ἀπό τόν ὁρισμό τῆς διακύμανσης ἔχουμε

$$Var(X) = E(X-\mu)^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx$$

ὅπου $f(x)$ εἶναι ἡ σ.π.π. τῆς X. Ἄν περιορίσουμε τό διάστημα ὁλοκληρώσεως σ'ἐκεῖνα τά x γιά τά ὁποῖα $|x-\mu|>\varepsilon$ ἔχουμε:

$$Var(X) \geq \int_{\{x:|x-\mu|\geq\varepsilon\}} (x-\mu)^2 f(x)dx$$

$$\geq \int_{\{x:|x-\mu|\geq\varepsilon\}} \varepsilon^2 f(x)dx=\varepsilon^2 P(|X-\mu|\geq\varepsilon)$$

διότι

$$\int_{\{x:|x-\mu|\geq\varepsilon\}} f(x)dx = P(|X-\mu|\geq\varepsilon)$$

σύμφωνα μέ τόν ὁρισμό τῆς σ.π.π.

'Ανάλογα ἀποδεικνύεται τό θεώρημα ὅταν X εἶναι διακριτή. ▼

Ἄν στήν ἀνισότητα (5.10) θέσουμε $\varepsilon = k\sigma$, ὅπου $k > 0$ παίρνουμε τήν ἀνισότητα

$$P(|X-\mu| \geq k\sigma) \leq 1/k^2. \tag{5.11}$$

Μέ ἄλλα λόγια ἐπιτυγχάνουμε ἕνα ἄνω φράγμα τῆς πιθανότητας $P(|X-\mu| \geq k\sigma)$. 'Ομοίως ἔχουμε

$$P(|X-\mu| \leq k\sigma) \geq 1-1/k^2$$

ἤ (5.12)

$$P(\mu-k\sigma \leq X \leq \mu+k\sigma) \geq 1-1/k^2 .$$

'Η ἀνισότητα αὐτή ἑρμηνεύεται ὡς ἀκολούθως: τουλάχιστον $1-(1/k^2)$ τῶν τιμῶν τῆς τ.μ. X βρίσκεται ἐντός k τυπικῶν ἀποκλίσεων ἀπό τή μέση τιμή τῆς X. Ἐπίσης δίνει ἕνα κάτω φράγμα τῆς πιθανότητας $P(|X-\mu| \leq k\sigma)$. Ἰδιαίτερα ἄν $k = 2$

$$P\{\mu-2\sigma<X<\mu+2\sigma\} \geq 3/4 = 75\%$$

καί ἄν $k = 3$

$$P\{\mu-3\sigma<X<\mu+3\sigma\} \geq 8/9 = 88,9\% \ .$$

Ὁ ρόλος τῆς διακυμάνσεως ἤ τῆς τυπικῆς ἀποκλίσεως ὡς μέτρα μεταβλητότητας μιᾶς τυχαίας μεταβλητῆς εἶναι πιά σαφής.

Ἡ ἀνισότητα τοῦ *Chebyshev* εἶναι μερική περίπτωση τῆς ἀκόλουθης γενικώτερης πρότασης γνωστῆς ὡς ἀ ν ι σ ό τ η τ α ς τ ο ῦ *M a r k o v*: ἄν $h(X)$ εἶναι μία συνάρτηση τῆς τυχαίας μεταβλητῆς X μέ $h(X) \geq 0$, $E[h(X)] < \infty$ καί $c > 0$ μία σταθερά τότε

$$P[h(X) \geq c] \leq \frac{E[h(X)]}{c} \ . \qquad (5.13)$$

Ἡ ἀπόδειξη τῆς (5.13) εἶναι παρόμοια μέ ἐκείνη τοῦ θεωρήματος τοῦ *Chebyshev*. Ἄν $h(X) = (X-\mu)^2$ προκύπτει ἡ ἀνισότητα τοῦ *Chebyshev*.

Ἡ ἀνισότητα τοῦ *Chebyshev* εἶναι ἕνα σημαντικό ἀποτέλεσμα τῆς Μαθηματικῆς Στατιστικῆς. Συνήθως γιά νά ὑπολογίσουμε τήν πιθανότητα ἑνός ἐνδεχόμενου πού περιγράφεται μέ τή βοήθεια μιᾶς τυχαίας μεταβλητῆς X χρειαζόμαστε τήν κατανομή τῆς X. Ἐάν ὅμως ἡ κατανομή τῆς X δέν εἶναι γνωστή ἀλλά γνωρίζουμε τήν μέση τιμή καί τήν τυπική της ἀπόκλιση, ἡ ἀνισότητα τοῦ *Chebyshev* μᾶς ἐπιτρέπει νά ὑπολογίζουμε φράγματα, πού δέν ἐξαρτῶνται ἀπό τήν κατανομή τῆς X, γιά τίς πιθανότητες ἐνδεχομένων τῆς τ.μ. X τῆς μορφῆς $(\mu-k\sigma,\mu+k\sigma)$.

Παραδείγματα:

5.6: Ἔστω X μία τυχαία μεταβλητή μέ κατανομή

$$f(x)=\begin{cases}\dfrac{1}{2\sqrt{3}}, & -\sqrt{3}<x<\sqrt{3}\\ 0, & \text{ἀλλοῦ}.\end{cases}$$

Εὔκολα μποροῦμε νά δοῦμε ὅτι $\mu=0$ καί $\sigma^2=1$. Ἄς πάρουμε $k=3/2$ καί ἄς ὑπολογίσουμε μέ ἀκρίβεια τήν πιθανότητα

$$P(|X-\mu|\geq k\sigma)=P(|X|\geq 3/2)$$

$$=1-\int_{-3/2}^{3/2}\frac{1}{2\sqrt{3}}\,dx=1-\frac{\sqrt{3}}{2}\approx 0{,}134.$$

Μέ τήν ἀνισότητα τοῦ *Chebyshev* ἔχουμε

$$P(|X-\mu|\geq k\sigma)\leq\frac{1}{k^2}=\frac{4}{9}\approx 0{,}47.$$

Ἡ τιμή αὐτή εἶναι ἕνα ἄνω φράγμα τῆς πιθανότητας $P(|X|\geq 3/2)$. Ἡ ἀκριβής πιθανότητα $0{,}134$ εἶναι βέβαια κατά πολύ μικρότερη ἀπό τό ἄνω φράγμα $0{,}47$ πού δίνει ἡ ἀνισότητα τοῦ *Chebyshev*. Αὐτό δέν πρέπει νά μᾶς ξενίζει ἄν λάβουμε ὑπ'ὄψη τήν γενικότητα τῆς ἀνισότητας τοῦ *Chebyshev*. Ἄν θέλουμε βελτίωση τῆς ἀνισότητας πρέπει νά κάνουμε ὑποθέσεις γιά τήν κατανομή τῆς X.

5.7: Ἔστω X διακριτή τυχαία μεταβλητή μέ

$$P(X=-1) = P(X=1) = \frac{1}{8}$$

$$P(X=0) = \frac{6}{8}.$$

Ἐδῶ $\mu = 0$ καί $\sigma^2 = 1/4$. Ἄν $k = 2$ τότε $1/k^2 = 1/4$ καί

$$P(|X-\mu| \geq k\sigma) = P(|X| \geq 1) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

Ἔτσι ἡ πιθανότητα $P(|X-\mu| \geq k)$ "πιάνει" τό ἄνω φράγμα $1/k^2 = 1/4$ πού προκύπτει ἀπό τήν ἀνισότητα τοῦ *Chebyshev*. Αὐτό ὀφείλεται στήν κατανομή πού ὑποθέσαμε γιά τό X.

5.4 Ροπές

Ἡ μέση τιμή καί ἡ διακύμανση μίας τυχαίας μεταβλητῆς δέν εἶναι τά μόνα ἀριθμητικά μέτρα πού περιγράφουν χαρακτηριστικά γνωρίσματα τῶν κατανομῶν. Ὑπάρχουν καί ἄλλα ὅπως οἱ ροπές, τά μέτρα λοξότητας καί κυρτώσεως μίας κατανομῆς, ἡ διάμεσος, ἡ κορυφή, τά ποσοστιαῖα σημεῖα κλπ. Στό ἐδάφιο αὐτό περιγράφουμε τίς διάφορες ροπές. Οἱ ροπές μίας τυχαίας μεταβλητῆς κατατάσσονται στίς ἀκόλουθες τέσσερες κατηγορίες.

α) Ἁ π λ έ ς ρ ο π έ ς ἤ ρ ο π έ ς π ε ρ ί τ ό μ η δ έ ν.

Ἡ ποσότητα

$$\mu_k = E(X^k) = \begin{cases} \sum_x x^k p(x), & \text{ἄν } X \text{ διακριτή τ.μ.} \\ \int_{-\infty}^{+\infty} x^k f(x)dx, & \text{ἄν } X \text{ συνεχής τ.μ.} \end{cases} \quad (5.14)$$

ὅπου k θετικός ἀκέραιος ἀριθμός λέγεται (ἀπλή) ροπή k τάξεως περί τό μηδέν τῆς τ.μ. X. Εἰδικώτερα βλέπουμε ὅτι

$$\mu_1 = \mu$$

δηλαδή ἡ ροπή πρώτης τάξεως περί τό μηδέν εἶναι ἡ ἴδια ἡ μέση τιμή τῆς τ.μ. 'Ανάλογα ὁρίζονται οἱ ροπές περί τό μηδέν μίας συναρτήσεως $h(X)$ τῆς τ.μ. X.

β) Κ ε ν τ ρ ι κ έ ς ρ ο π έ ς ἤ ρ ο π έ ς π ε ρ ί τ ό μ.

Ἄν μ εἶναι ἡ μέση τιμή μίας τ.μ. X τότε ἡ ποσότητα

$$\lambda_k = E(X-\mu)^k = \begin{cases} \sum\limits_x (x-\mu)^k p(x), & \text{ἄν } X \text{ διακριτή τ.μ.} \\ \int_{-\infty}^{+\infty} (x-\mu)^k f(x)dx, & \text{ἄν } X \text{ συνεχής τ.μ.,} \end{cases} \qquad (5.15)$$

ὅπου k θετικός ἀκέραιος ἀριθμός, λέγεται κ ε ν τ ρ ι κ ή ρ ο π ή ἤ ρ ο π ή π ε ρ ί τ ό μ k τ ά ξ ε ω ς τῆς τ.μ. X. Εἰδικώτερα βλέπουμε ὅτι

$$\lambda_1 = 0$$

γιά κάθε κατανομή καί συνεπῶς τό μέτρο αὐτό εἶναι ἄχρηστο ὡς ἀριθμητικό χαρακτηριστικό τῶν κατανομῶν. 'Επίσης

$$\lambda_2 = \sigma^2$$

δηλαδή κεντρική ροπή δευτέρας τάξεως εἶναι ἡ διακύμανση τῆς X.

Οἱ κεντρικές ροπές μίας τυχαίας μεταβλητῆς συνδέονται μέ τίς ἀπλές ροπές μέ τίς ἀκόλουθες σχέσεις

$$\boxed{\begin{aligned} \lambda_2 &= \mu_2-\mu_1^2 \\ \lambda_3 &= \mu_3-3\mu_2\mu_1+2\mu_1^3 \\ \lambda_4 &= \mu_4-4\mu_3\mu_1+6\mu_2\mu_1^2-3\mu_1^4 \end{aligned}} \qquad (5.16)$$

καί γενικώτερα

$$\lambda_k = \mu_k-\binom{k}{1}\mu_{k-1}\mu_1+\binom{k}{2}\mu_{k-2}\mu_1^2-\ldots+(-1)^k\binom{k}{k}\mu_1^k \qquad (5.17)$$

Οἱ σχέσεις αὐτές μποροῦν νά ἀποδειχτοῦν ἄν ἀναπτύξουμε τίς δυνάμεις $(x-\mu)^k$ στήν (5.15) καί λάβουμε ὑπ'ὄψη μας τήν (5.14). Ἀνάλογα ὁρίζονται οἱ κεντρικές ροπές μίας συνάρτησης $h(X)$ τῆς τ.μ. X.

γ) Τ υ π ι κ έ ς ἤ κ α ν ο ν ι κ ο π ο ι η μ έ ν ες ρ ο π έ ς.

Ἡ ποσότητα

$$\alpha_k = E\left(\frac{X-\mu}{\sigma}\right)^k \quad , \quad k = 1, 2, 3, \ldots \qquad (5.18)$$

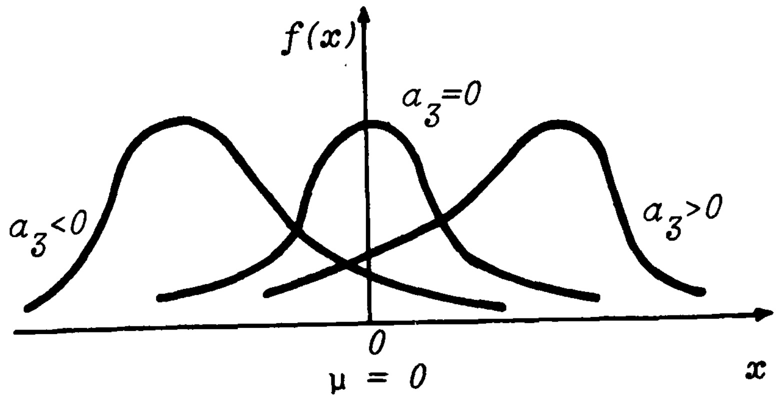
ὅπου μ ἡ μέση τιμή καί σ^2 ἡ διακύμανση τῆς X, ὀνομάζεται τ υ π ι κ ή ἤ κ α ν ο ν ι κ ο π ο ι η μ έ ν η ρ ο π ή k τ ά ξ ε ω ς τ ῆ ς X.

Εἶναι εὔκολο νά δοῦμε ὅτι

$$\alpha_1 = 0 \quad \text{καί} \quad \alpha_2 = 1$$

γιά κάθε τ.μ. X καί συνεπῶς οἱ ποσότητες αὐτές εἶναι ἄχρηστες ὡς ἀριθμητικά χαρακτηριστικά τῶν κατανομῶν.

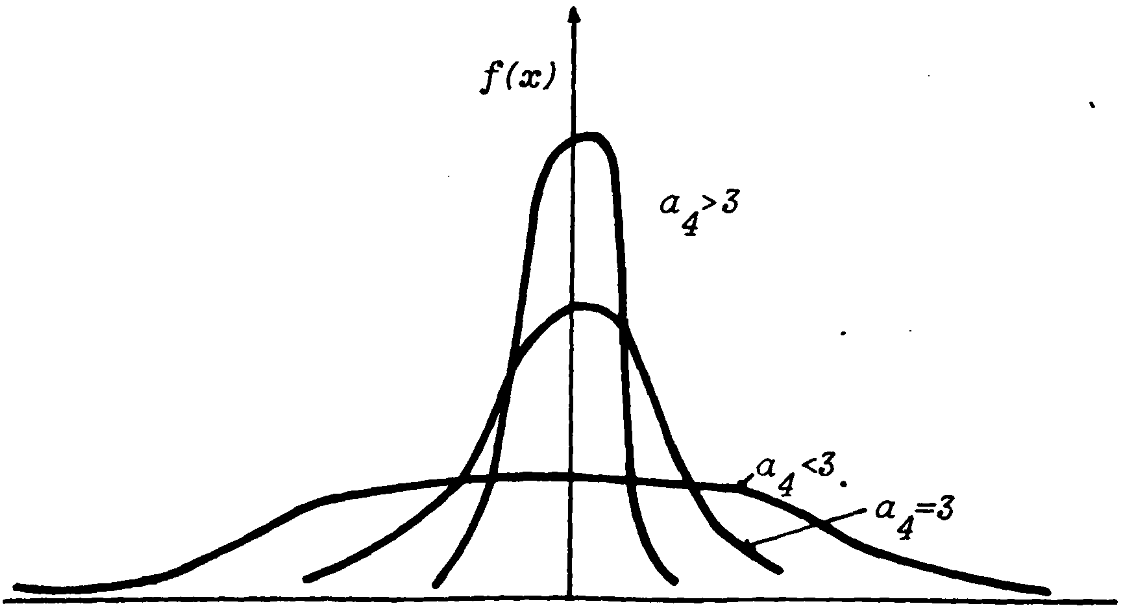
'Ιδιαίτερο ἐνδιαφέρον παρουσιάζουν οἱ ροπές α_3 καί α_4. Ἡ ροπή α_3 ὀνομάζεται καί μ έ τ ρ ο ἤ σ υ ν τ ε λ ε σ τ ή ς λ ο ξ ό τ η τ α ς (ἀσυμμετρίας) τῆς κατανομῆς γιατί μετρᾶ τήν λοξότητα (συμμετρία ἤ ἀσυμμετρία) τῆς καμπύλης τῆς σ.π.π. $f(x)$ ἤ τῆς σ.π. $p(x)$ τῆς τ.μ. X. Ἄν ἡ κατανομή



Σχῆμα 5.1 Λοξότης

εἶναι συμμετρική περί τό μ τότε $\alpha_3 = 0$. Τό ἀντίστροφο δέν ἰσχύει (βλ. Ἄσκηση 5.50). Ἄν $\alpha_3 > 0$ τότε ἡ καμπύλη εἶναι λοξή πρός τά δεξιά ἐνῶ ἄν $\alpha_3 < 0$ ἡ καμπύλη εἶναι λοξή πρός τά ἀριστερά (βλέπε Σχῆμα 5.1). Ἡ ροπή α_4 ὀνομάζεται καί μ έ τ ρ ο ἤ σ υ ν τ ε λ ε σ τ ή ς κ υ ρ τ ώ σ ε ω ς τῆς κατανομῆς. Μεγάλες τιμές τοῦ α_4 σημαίνουν καμπύλη πολύ κυρτή, μικρές τιμές τοῦ α_4 σημαίνουν καμπύλη μέ μικρή κυρτότητα.

Τό Σχῆμα 5.2 δίνει τρεῖς ἀντιπροσωπευτικές καμπύλες



Σχῆμα 5.2 Κύρτωσις

συναρτήσεων πυκνότητας πιθανότητας μέ διάφορα α_4 γιά μία τ.μ. X. Ἡ καμπύλη μέ $\alpha_4 = 3$ ἀντιστοιχεῖ σέ τ.μ. X μέ κατανομή $N(\mu,\sigma^2)$.

Τόσο τό α_3 ὅσο καί τό α_4 εἶναι ἀναλλοίωτα σέ γραμμικούς μετασχηματισμούς.

δ) Π α ρ α γ ο ν τ ι κ έ ς ρ ο π έ ς.

Ο ἱ π α ρ α γ ο ν τ ι κ έ ς ρ ο π έ ς k τ ά - ξ ε ω ς συμβολίζονται μέ $\mu_{[k]}$ καί ὁρίζονται ὡς ἡ ἀναμενόμενη τιμή τοῦ γινομένου $X(X-1)\ldots(X-k+1)$ δηλαδή

$$\mu_{[k]} = E[X(X-1)\ldots(X-k+1)], \qquad (5.19)$$

ὅπου k θετικός καί ἀκέραιος.

5.5 Ροπὲς εἰδικῶν κατανομῶν

Στό ἐδάφιο αὐτό δίνουμε τίς βασικές ροπές (μέση τιμή, διακύμανση, α_3 καί α_4) διαφόρων εἰδικῶν κατανομῶν.

α) Διωνυμική κατανομή $B(n,p)$.

$$\boxed{\mu = n\,p \ , \quad \sigma^2 = n\,p\,q} \qquad (5.20)$$

'Απόδειξη:

$$\mu = E(X) = \sum_{x=0}^{n} x\binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$= n\,p \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)![(n-1)-(x-1)]} p^{x-1} q^{n-1-(x-1)}$$

(θέτουμε $y = x-1$)

$$= n\,p \sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)} p^y q^{n-1-y}$$

$$= n\,p \sum_{y=0}^{n-1} \binom{n-1}{y} p^y q^{n-1-y}$$

$$= n\,p(p+q)^{n-1} = n\,p\ .$$

Γιά τήν εὕρεση τῆς σ^2 θά χρησιμοποιήσουμε τόν τύπο

$$\sigma^2 = E(X^2)-(EX)^2\ .$$

Κατά συνέπεια μᾶς χρειάζεται ἡ $E(X^2)$. 'Αντί νά ὑπολογίσουμε τήν ροπή αὐτή κατ'εὐθεῖαν θά βροῦμε πρῶτα τήν παραγοντική ροπή δευτέρας τάξεως $\mu_{[2]}$ πού εἶναι πιό εὔκολο. 'Από τήν $\mu_{[2]}$ θά πάρουμε τήν $E(X^2)$.

$$\mu_{[2]} = E[X(X-1)]$$

$$= \sum_{x=0}^{n} x(x-1)\,\frac{n!}{x!(n-x)!}\,p^{x-2}q^{n-x}$$

$$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!}\,p^{x-2}q^{n-x}$$

γιατί $x(x-1) = 0$ ὅταν $x = 0$ ἤ 1. Ἐργαζόμενοι ὅπως καί προηγουμένως, δηλαδή θέτοντας $y = x-2$, ἔχουμε

$$\mu_{[2]} = n(n-1)p^2 \sum_{y=0}^{n-2} \binom{n-2}{y} p^y q^{n-2-y}$$

$$= n(n-1)p^2(p+q)^{n-2}$$

$$= n(n-1)p^2 .$$

Ἀλλά

$$\mu_{[2]} = E[X(X-1)] = E[X^2-X]$$

$$= E(X^2)-E(X) = E(X^2)-n\,p$$

καί ἔτσι

$$E(X^2) = n(n-1)p^2+np.$$

Ἑπομένως

$$\sigma^2 = n(n-1)p^2+np-(np)^2$$

$$= np[(n-1)p+1-np] = np(1-p)$$

$$= n\,p\,q \,.\blacktriangledown$$

Δίνουμε χωρίς απόδειξη τίς ροπές α_3 καί α_4.

$$\boxed{\alpha_3 = \frac{(q-p)}{\sqrt{n\,p\,q}} \quad \text{καί} \quad \alpha_4 = 3+\frac{1-6pq}{npq}} \tag{5.21}$$

β) Ὑ π ε ρ γ ε ω μ ε τ ρ ι κ ή κ α τ α ν ο μ ή $Hg(N,n,p)$

$$\boxed{\mu = n\,p \quad , \quad \sigma^2 = n\,p\,q\,\frac{N-n}{N-1}} \tag{5.22}$$

$$\boxed{\begin{aligned}
\alpha_3 &= \frac{(q-p)\sqrt{N-1}\,(N-2n)}{(N-2)\sqrt{(Npq(N-n)}} \\[2ex]
\alpha_4 &= \frac{\dfrac{(Np-1)(Np-2)(Np-3)(n-1)(n-2)(n-3)}{(N-1)(N-2)(N-3)}}{npq^2\,\dfrac{(N-n)^2}{(N-1)^2}} \\[2ex]
&\quad - \frac{4np\,\dfrac{(Np-1)(Np-2)(n-1)(n-2)}{(N-1)(N-2)}}{npq^2\,\dfrac{(N-n)^2}{(N-1)^2}} \\[2ex]
&\quad + \frac{6n^2p^2\,\dfrac{(Np-1)(n-1)}{N-1} - 3n^3p^3}{npq^2\,\dfrac{(N-n)^2}{(N-1)^2}}
\end{aligned}} \tag{5.23}$$

Ἡ ἀπόδειξη τῶν τύπων αὐτῶν δίνεται ὡς ἄσκηση.

γ) Γ ε ω μ ε τ ρ ι κ ή κ α τ α ν ο μ ή $Geo(p)$

$$\boxed{\mu = 1/p \quad , \quad \sigma^2 = \frac{1-p}{p^2} = \frac{q}{p^2}} \tag{5.24}$$

'Απόδειξη: "Εχουμε

$$\mu = \sum_{x=1}^{\infty} xpq^{x-1} = p \sum_{x=1}^{\infty} xq^{x-1} .$$

Εἶναι γνωστό ὅτι ἡ σειρά $\sum_{n=0}^{\infty} \alpha^n$ μέ $|\alpha| < 1$ συγκλίνει ὁμοιόμορφα καί τό ἄθροισμα της εἶναι (φθίνουσα γεωμετρική πρόοδος)

$$1+\alpha+\alpha^2+\ldots+\alpha^n+\ldots = \frac{1}{1-\alpha} .$$

"Αν παραγωγίσουμε τά μέλη αὐτῆς τῆς σχέσεως ὡς πρός α δύο φορές ἔχουμε

$$1+2\alpha+3\alpha^2+\ldots+n\alpha^{n-1}+\ldots = \frac{1}{(1-\alpha)^2}$$

$$2\cdot 1+3\cdot 2\alpha+\ldots+n(n-1)\alpha^{n-2}+\ldots = \frac{2}{(1-\alpha)^3} .$$

Βασιζόμενοι στήν πρώτη ταυτότητα ἔχουμε

$$\mu = p\frac{1}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

Βασιζόμενοι στήν δεύτερη ταυτότητα ἔχουμε

$$\mu_{[2]} = E[X(X-1)] = \sum_{x=1}^{\infty} x(x-1)pq^{x-1}$$

$$= pq \sum_{x=2}^{\infty} x(x-1)q^{x-2}$$

$$= pq \frac{2}{(1-q)^3} = \frac{2q}{p^2}$$

'Από ἐδῶ συμπεραίνουμε ὅτι

$$E(X^2) = \frac{2q}{p^2} + \frac{1}{p}$$

Ἔτσι ἔχουμε

$$\sigma^2 = E(X^2)-(EX)^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1}{p^2}(2q+p-1) = \frac{q}{p^2} . ▼$$

Οἱ ροπές α_3 καί α_4 δίνονται χωρίς ἀπόδειξη

$$\boxed{\alpha_3 = \frac{1+q}{\sqrt{q}} , \quad \alpha_4 = 9+(p^2/q) .} \qquad (5.25)$$

δ) Ἀρνητική Διωνυμική $NB(k,p)$.

$$\boxed{\mu = \frac{k}{p} , \quad \sigma^2 = \frac{kq}{p^2} .} \qquad (5.26)$$

$$\boxed{\alpha_3 = \frac{1+q}{\sqrt{kq}} , \quad \alpha_4 = \frac{p^2+6q}{kq} .} \qquad (5.27)$$

Ἡ ἀπόδειξη τῶν τύπων αὐτῶν δίνεται ὡς ἄσκηση.

ε) *P o i s s o n* $P(\lambda)$

$$\boxed{\mu = \lambda , \quad \sigma^2 = \lambda .} \qquad (5.28)$$

'Απόδειξη: "Εχουμε

$$\mu = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-1)!}$$

$$=\lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda,$$

$$\mu_{[2]} = \sum_{x=0}^{\infty} \frac{x(x-1)e^{-\lambda}\lambda^x}{x!}$$

$$= \lambda^2 e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} = \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2 \quad ,$$

καί

$$E[X(X-1)] = E(X^2)-E(X) = \lambda^2 .$$

"Ετσι

$$E(X^2) = \lambda^2+\lambda .$$

Συνεπῶς

$$\sigma^2 = \lambda^2+\lambda-\lambda^2 = \lambda . \blacktriangledown$$

Δίνουμε χωρίς ἀπόδειξη τίς ροπές α_3 καί α_4 .

$$\boxed{\alpha_3 = \frac{1}{\sqrt{\lambda}} \quad , \quad \alpha_4 = 3+\frac{1}{\lambda} .} \qquad (5.29)$$

στ) 'Ο μ ο ι ό μ ο ρ φ η κ α τ α ν ο μ ή $U(\alpha,\beta)$

$$\boxed{\mu = \frac{\alpha+\beta}{2} \quad , \quad \sigma^2 = \frac{(\beta-\alpha)^2}{12} .} \qquad (5.30)$$

'Απόδειξη:

$$\mu = \int_{-\infty}^{+\infty} xf(x)dx = \int_{\alpha}^{\beta} x \frac{1}{\beta-\alpha}\, dx = \frac{x^2}{2(\beta-\alpha)}\Big|_{\alpha}^{\beta} = \frac{\beta^2-\alpha^2}{2(\beta-\alpha)} = \frac{\alpha+\beta}{2}$$

$$E(X^2) = \int_{\alpha}^{\beta} x^2 \frac{1}{\beta-\alpha}\, dx = \frac{x^3}{3(\beta-\alpha)}\Big|_{\alpha}^{\beta} = \frac{\beta^3-\alpha^3}{3(\beta-\alpha)} = \frac{\alpha^2+\alpha\beta+\beta^2}{3}$$

$$\sigma^2 = \frac{\alpha^2+\alpha\beta+\beta^2}{3} - \frac{(\alpha+\beta)^2}{4} = \frac{(\alpha-\beta)^2}{12} \ .\blacktriangledown$$

Δίνουμε χωρίς απόδειξη τίς ροπές α_3 καί α_4 .

$$\boxed{\alpha_3 = 0 \quad , \quad \alpha_4 = \frac{9}{5} \ .} \tag{5.31}$$

ζ) Ἐκθετική κατανομή $Ek\theta(\lambda)$.

$$\boxed{\mu = \frac{1}{\lambda} \quad , \quad \sigma^2 = \frac{1}{\lambda^2} \ .} \tag{5.32}$$

'Απόδειξη:

$$\mu = \int_0^{\infty} x\lambda e^{-\lambda x} dx = -\int_0^{\infty} x de^{-\lambda x} = -xe^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\, dx$$

$$= \lim_{t\to\infty}(-te^{-\lambda t}) - \frac{1}{\lambda} e^{-\lambda x}\Big|_0^{\infty} = 0 + \frac{1}{\lambda} = \frac{1}{\lambda} \ .$$

$$E(X^2) = \int_0^{\infty} x^2\, e^{-\lambda x} dx = -\int_0^{\infty} x^2 de^{-\lambda x} = \int_0^{\infty} e^{-\lambda x} 2x dx$$

$$= -2\int_0^{\infty} \frac{1}{\lambda}\, xd(e^{-\lambda x}) = \frac{2}{\lambda} \int_0^{\infty} e^{-\lambda x} dx = \frac{2}{\lambda^2} .$$

Συνεπῶς

$$\sigma^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} .$$

Ἡ εὕρεση τῶν ροῶν α_3 καί α_4 δίνεται ὡς ἄσκηση.

$$\boxed{\alpha_3 = 2 \quad , \quad \alpha_4 = 9 \ .} \tag{5.33}$$

η) Γ ά μ μ α κ α τ α ν ο μ ή $G(\alpha,\beta)$.

$$\boxed{\mu = \alpha\beta \quad , \quad \sigma^2 = \alpha\beta^2 \ .} \tag{5.34}$$

Ἀπόδειξη:

$$\mu = \int_0^\infty x \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)} dx = \int_0^\infty \frac{x^\alpha e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)} dx .$$

Ἄν κάνουμε τήν ἀντικατάσταση $y = \frac{x}{\beta}$ ἔχουμε

$$\mu = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty \beta^\alpha y^\alpha e^{-y} \beta dy = \frac{\beta\Gamma(\alpha+1)}{\Gamma(\alpha)} = \frac{\beta\alpha\Gamma(\alpha)}{\Gamma(\alpha)} = \alpha\beta .$$

Ὁμοίως βρίσκουμε ὅτι

$$E(X^2) = \beta^2(\alpha+1)\alpha .$$

Συνεπῶς

$$\sigma^2 = \alpha(\alpha+1)\beta^2 - (\alpha\beta)^2 = \alpha\beta^2 . \blacktriangledown$$

Ἡ εὕρεση τῶν ροπῶν α_3 καί α_4 δίνεται ὡς ἄσκηση.

$$\boxed{\alpha_3 = \frac{2\sqrt{\alpha}}{\alpha} \quad , \quad \alpha_4 = 3+\frac{6}{\alpha}} \tag{5.35}$$

Οἱ k τάξεως ροπές περί τό μηδέν τῆς X εἶναι

$$\boxed{\mu_k = EX^k = \frac{\beta^k\Gamma(\alpha+k)}{\Gamma(\alpha)} \quad , \; k = 1,2,3,\ldots} \tag{5.36}$$

Ἀπόδειξη:

$$\mu_k = EX^k = \int_0^\infty \frac{x^k}{\beta^\alpha\Gamma(\alpha)}\, x^{\alpha-1}e^{-x/\beta}dx$$

$$= \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty x^{\alpha+k-1}e^{-x/\beta}dx$$

$$= \frac{\beta^{\alpha+k}\Gamma(\alpha+k)}{\beta^\alpha\Gamma(\alpha)} = \frac{\beta^k\Gamma(\alpha+k)}{\Gamma(\alpha)} \quad . \; \blacktriangledown$$

Οἱ κεντρικές ροπές λ_k τῆς X βρίσκονται τώρα εὔκολα μέ τή βοήθεια τῶν σχέσεων (5.16) καί (5.17).

θ) Β ή τ α κ α τ α ν ο μ ή $Be(\alpha,\beta)$.

$$\boxed{\mu = \frac{\alpha}{\alpha+\beta} \quad , \quad \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}} \tag{5.36a}$$

'Απόδειξη:

$$\mu = \int_0^1 \frac{x x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}\, dx$$

$$= \int_0^1 \frac{x^{\alpha+1-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}\, dx = \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} .$$

"Αν λάβουμε ὑπ'ὄψη τήν σχέση $B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ καί ἀναπτύξουμε τίς συναρτήσεις φθάνουμε στό ἀποτέλεσμα. 'Εργαζόμενοι παρόμοια μποροῦμε νά ἀποδείξουμε τόν τύπο γιά τήν διακύμανση.

Οἱ ροπές α_3 καί α_4 δίνονται χωρίς ἀπόδειξη.

$$\boxed{\begin{aligned} \alpha_3 &= \frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}} , \\ \alpha_4 &= \frac{3(\alpha+\beta+1)[2(\alpha^2-\alpha\beta+\beta^2)+\alpha\beta(\alpha+\beta)]}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)} . \end{aligned}} \tag{5.37}$$

ι) Κ α ν ο ν ι κ ή κ α τ α ν ο μ ή

1. Τυπική κατανομή κατανομή $N(0,1)$.

$$\boxed{\mu = 0 , \quad \sigma^2 = 1 , \quad \alpha_3 = 0 , \quad \alpha_4 = 3} \tag{5.38}$$

'Απόδειξη:

$$\mu = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx .$$

'Επειδή ή όλοκληρωτέα συνάρτηση εἶναι περιττή δηλαδή $f(-x) = -f(x)$ τό παραπάνω όλοκλήρωμα εἶναι μηδέν.

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \; .$$

'Επειδή ή όλοκληρωτέα συνάρτηση εἶναι ἄρτια δηλαδή $f(-x) = f(x)$ ἔχουμε

$$E(X^2) = 2\int_{0}^{\infty} \frac{x^2}{\sqrt{2\pi}} e^{-x^2/2} dx \; .$$

"Αν σ'αὐτό τό όλοκλήρωμα κάνουμε τόν μετασχηματισμό $y = x^2/2$ λαμβάνουμε

$$E(X^2) = 2 \int_{0}^{\infty} \frac{2ye^{-y}}{2\pi} \; \frac{dy}{2\, y^{\frac{1}{2}}}$$

$$= \frac{2}{\sqrt{\pi}} \int_{0}^{\infty} y^{\frac{1}{2}} e^{-y} dy = \frac{2}{\sqrt{\pi}} \quad (3/2)$$

$$= \frac{2}{\sqrt{\pi}} \frac{1}{2} \Gamma(1/2) = \frac{\sqrt{\pi}}{\sqrt{\pi}} = 1 \; .$$

"Ετσι

$$\sigma^2 = EX^2 - \mu^2 = 1 \; .$$

$$\alpha_3 = \int_{-\infty}^{+\infty} \left(\frac{x-\mu}{\sigma}\right)^3 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_{-\infty}^{+\infty} \frac{x^3}{\sqrt{2\pi}} e^{-x^2/2} dx = 0$$

ἐπειδή ή όλοκληρωτέα συνάρτηση εἶναι περιττή.

$$\alpha_4 = E\left(\frac{X-\mu}{\sigma}\right)^4 = E(X^4) \; .$$

Ἐργαζόμενοι ὅπως στήν περίπτωση τοῦ ὑπολογισμοῦ τοῦ $E(X^2)$ λαμβάνουμε τό ἀποτέλεσμα $E(X^4) = 3$. ▼

2. Κανονική κατανομή $N(\mu,\sigma^2)$

$$\boxed{\begin{array}{ll} \text{Μέση τιμή} = \mu \;, & \alpha_3 = 0 \\ \text{Διακύμανση} = \sigma^2, & \alpha_4 = 3 \end{array}} \qquad (5.39)$$

Ἡ ἀπόδειξη εἶναι ἀνάλογη μέ τίς προηγούμενες ἀποδείξεις γιά τήν τυπική κανονική κατανομή καί παραλείπεται.

Οἱ k τάξεως κεντρικές ροπές τῆς X εἶναι

$$\boxed{\lambda_k = E(X-\mu)^k = \begin{cases} 0 & , \; k = 1,3,5,\ldots \\ \sigma^k(k-1)(k-3)\ldots3\cdot1 & , \; k = 2,4,6,\ldots \end{cases}} \qquad (5.40)$$

Ἀπόδειξη: Ἔχουμε

$$\lambda_k = E(X-\mu)^k = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)^k e^{-(x-\mu)^2/2\sigma^2} dx$$

$$= \frac{\sigma^k}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^k e^{-y^2/2} dy \qquad \text{(ἀλλαγή μεταβλητῆς } y = (x-\mu)/\sigma)$$

Ἄν τό k εἶναι περιττός ἀριθμός ἡ τελευταία ὁλοκληρωτέα συνάρτηση εἶναι περιττή. Ἔτσι $\lambda_k = 0$ γιά $k = 1, 3, 5, \ldots$ Ἄν τό k εἶναι ἄρτιος κάνουμε ἀλλαγή μεταβλητῆς $y^2 = 2z$ καί παίρνουμε

$$\lambda_k = \sqrt{\frac{2}{\pi}}\,\sigma^k 2^{(k-1)/2} \int_0^\infty z^{(k-1)/2} e^{-z} dz$$

$$= \sigma^k 2^{(k-1)/2}\Gamma((k+1)/2)$$

$$= \sigma^k (k-1)(k-3)\ldots 3.1 \quad . \blacktriangledown$$

5. 6 Κορυφή — Ποσοστιαία σημεία — Διάμεσος

Ἐκτός ἀπό τά ἀριθμητικά χαρακτηριστικά πού ἀναφερθηκαν μέχρι τώρα, ὑπάρχουν καί ἄλλες ποσότητες πού περιγράφουν τή θέση μίας κατανομῆς.

Ὁρισμός 5.5: Κορυφή *(mode)*. Ἔστω X μία τυχαία μεταβλητή. Κάθε σημεῖο k γιά τό ὁποῖο ἰσχύει ἡ σχέση

$$p_X(k) = \max_x p_X(x) \quad \text{ἄν } X \text{ διακριτή τ.μ.}$$

ἤ $\qquad\qquad (5.41)$

$$f_X(k) = \max f_X(x) \quad \text{ἄν } X \text{ συνεχής τ.μ.}$$

λέγεται κορυφή ἤ πιό ἐπικρατούσα τιμή *(mode)* τῆς τυχαίας μεταβλητῆς (ἤ καταχρηστικά τῆς κατανομῆς).

Κορυφή μίας τυχαίας μεταβλητῆς εἶναι ἕνα σημεῖο μέ τή μεγαλύτερη συχνότητα (πιθανότητα). Ἀπό μαθηματικῆς πλευρᾶς κορυφή εἶναι τό σημεῖο ἤ τά σημεῖα στά ὁποῖα ἐμφανίζεται τό μέγιστο τῆς $p(x)$ ἤ $f(x)$. Δεδομένου ὅτι οἱ $p(x)$ καί $f(x)$ εἶναι συναρτήσεις χωρίς ἰδιαίτερες μαθηματικές ἰδιότητες (ἐκτός τοῦ ὅτι εἶναι κατανομές πιθανότητας), ἡ κορυφή μίας τ.μ. μπορεῖ νά μήν ὑπάρχει ἤ μία κατανομή μπορεῖ νά ἔχει μία, δύο ἤ καί περισσότερες κορυφές.

Ἡ εὕρεση τῶν κορυφῶν μίας τυχαίας μεταβλητῆς εἶναι πρόβλημα εὑρέσεως τοῦ μέγιστου μίας συναρτήσεως. Τό πρόβλημα αὐτό μπορεῖ νά λυθεῖ εἴτε μέ ἁπλές μαθηματικές ἀρχές εἴτε μέ μεθόδους ἀπειροστικοῦ λογισμοῦ. Αὐτό γίνεται στά ἑπόμενα θεωρήματα πού ἀφοροῦν τή διωνυμική κατανομή, τήν κατανομή τοῦ *Poisson* καί τήν κανονική κατανομή.

Θεώρημα 5.4:

Ἔστω X τυχαία μεταβλητή μέ διωνυμική κατανομή $B(n,p)$ καί $\nu = (n+1)p$. Ἐάν τό ν εἶναι ἀκέραιος ἡ κατανομή ἔχει δύο κορυφές στά σημεῖα $\nu-1$ καί ν, ἄλλως ἔχει μία κορυφή στό σημεῖο $[\nu]$, ὅπου $[\nu]$ συμβολίζει τό ἀκέραιο μέρος τοῦ ν.

Ἀπόδειξη: Ἔστω $p(x) = n!p^x q^{n-x}/[x!(n-x)!]$, $x = 0,1,\ldots,n,$ ἡ σ.π. τῆς διωνυμικῆς κατανομῆς. Ἔχουμε

$$\frac{p(x)}{p(x-1)} = \frac{n-x+1}{x}\ \frac{p}{q}\ . \qquad (5.42)$$

Ἔτσι

ἄν $\dfrac{n-x+1}{x}\ \dfrac{p}{q} > 1$ ἔπεται $p(x-1) < p(x)$

ἤ

ἄν $x < (n+1)p$ ἔπεται $p(x-1) < p(x)$

Μέ ἄλλα λόγια ἡ $p(x)$ εἶναι γνησίως αὔξουσα γιά κάθε x τέτοιο ὥστε $x < (n+1)p = \nu$ καί γνησίως φθίνουσα γιά κάθε x τέτοιο ὥστε $x > (n+1)p = \nu$.

Ἄν τώρα τό ν εἶναι ἀκέραιος ἡ $p(x)$ εἶναι γνησίως αὔξουσα γιά κάθε x τέτοιο ὥστε $x \leq \nu-1$ καί γνησίως φθίνουσα γιά κάθε x τέτοιο ὥστε $x \geq \nu+1$. Ἔτσι ὑποψήφια μέγιστα εἶναι τά σημεῖα $\nu-1$, ν καί $\nu+1$. Ἔχουμε ὅμως

$$\frac{p(\nu+1)}{p(\nu)} = \frac{n-\nu-1+1}{\nu+1} \quad \frac{p}{q} = \frac{np-\nu p}{(\nu+1)q}$$
$$= \frac{\nu-p-\nu p}{(\nu+1)q} < \frac{\nu+1-p-\nu p}{(\nu+1)q} = 1 \tag{5.43}$$

Ἔτσι $p(\nu) > p(\nu+1)$. Ἐπί πλέον ἡ (5.42) γιά $x = \nu$ δίνει $p(\nu-1) = p(\nu)$. Ἄρα ἡ $p(x)$ ἔχει ἕνα μέγιστο στά δύο σημεῖα $\nu-1$ καί ν.

Ἄν τό ν δέν εἶναι ἀκέραιος ἡ $p(x)$ εἶναι γνησίως αὔξουσα γιά κάθε $x \leq [\nu]$ καί γνησίως φθίνουσα γιά κάθε $x \geq [\nu]+1$. Ἔτσι ὑποψήφια μέγιστα εἶναι τά σημεῖα $[\nu]$ καί $[\nu]+1$. Μέ βάση τή σχέση (5.43) μποροῦμε νά δοῦμε ὅτι $p([\nu]+1) < p([\nu])$. Ἄρα ἡ $p(x)$ ἔχει μέγιστο στό σημεῖο $x = [\nu]$. ▼

Θεώρημα 5.5:

Ἔστω X τυχαία μεταβλητή μέ κατανομή Poisson $P(\lambda)$. Ἐάν τό λ εἶναι ἀκέραιος τότε ἡ κατανομή ἔχει κορυφές στά σημεῖα $\lambda-1$ καί λ. Ἐάν τό λ δέν εἶναι ἀκέραιος τότε ἡ κατανομή ἔχει μία κορυφή στό σημεῖο $[\lambda]$, ὅπου $[\lambda]$ συμβολίζει τό ἀκέραιο μέρος τοῦ λ.

'Απόδειξη: Ἄν $p(x)$ ἡ σ.π.π. τῆς κατανομῆς *Poisson* $P(\lambda)$ ἔχουμε

$$\frac{p(x)}{p(x-1)} = \frac{\lambda}{x} .$$

'Από αὐτή τή σχέση συμπεραίνουμε ὅτι ἡ $p(x)$ εἶναι γνησίως αὔξουσα ὅταν $x < \lambda$ καί γνησίως φθίνουσα ὅταν $x > \lambda$. 'Εάν τό λ εἶναι ἀκέραιος τότε $p(\lambda) = p(\lambda-1)$ καί κατά συνέπεια

$$p(\lambda) = p(\lambda-1) = \max_x p(x) .$$

'Η κατανομή ἔχει δύο κορυφές στά σημεῖα $\lambda-1$ καί λ.

'Εάν τό λ δέν εἶναι ἀκέραιος τότε ἡ $p(x)$ εἶναι γνησίως αὔξουσα ὅταν $x \leq [\lambda]$ καί γνησίως φθίνουσα ὅταν $x \geq [\lambda]+1$. 'Επί πλέον

$$\frac{p([\lambda]+1)}{p([\lambda])} = \frac{\lambda}{[\lambda]+1} < 1$$

ἤ

$$p([\lambda]+1) < p([\lambda]) .$$

Συνεπῶς τό σημεῖο $[\lambda]$ εἶναι κορυφή τῆς κατανομῆς. ▼

Θ ε ώ ρ η μ α 5.6:

<u>'Η κορυφή μιᾶς κανονικῆς τυχαίας μεταβλητῆς</u>, $X \sim N(\mu,\sigma^2)$, <u>εἶναι τό σημεῖο</u> $X = \mu$.

'Απόδειξη: 'Η σ.π.π. τῆς X εἶναι

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} .$$

Παραγωγίζοντες τήν $f(x)$ έχουμε

$$f'(x) = -\frac{1}{\sqrt{2\pi\sigma^2}}\frac{(x-\mu)}{\sigma^2}e^{-(x-\mu)^2/2\sigma^2}$$

Έτσι $f'(x) = 0$ όταν $x = \mu$ ενώ $f''(\mu) < 0$. Συνεπώς τό σημεΐο $X = \mu$ μεγιστοποιεΐ τήν $f(x)$.

Στό ίδιο αποτέλεσμα φτάνουμε σύντομα άν παρατηρήσουμε ότι ό εκθέτης στήν έκφραση τής $f(x)$ γίνεται μέγιστος όταν $x = \mu$. ▼

Μία άλλη κατηγορία χαρακτηριστικών σημείων μίας κατανομής είναι τά ποσοστιαΐα σημεΐα.

Ορισμός 5.6: Ποσοστιαΐα σημεΐα.
Έστω X μία τυχαία μεταβλητή μέ α.σ.κ. $F(x)$. Κάθε σημεΐο x_p, $0 \leqq p \leqq 1$, γιά τό όποΐο ισχύει

$$P(X \leqq x_p) = F(x_p) = p \qquad \text{άν ή } X \text{ είναι συνεχής.}$$

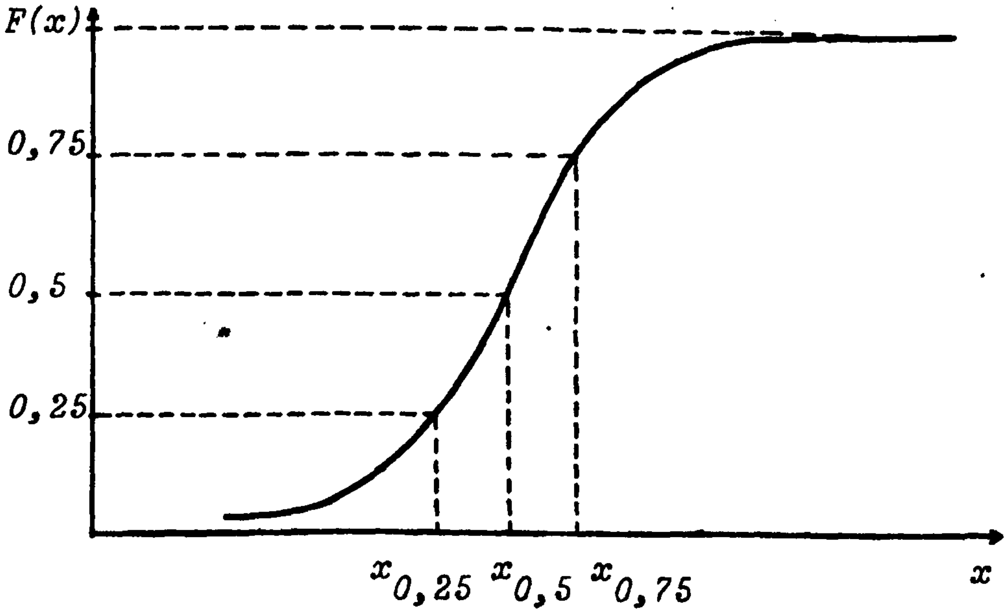
ή (5.44)

$$P(X \leqq x_p) \geqq p \quad \text{καί} \quad P(X \geqq x_p) \geqq 1-p \quad \text{άν ή } X \text{ είναι διακριτή.}$$

λέγεται p - ποσοσταΐο σημεΐο (*quantile*) τής τ.μ. X ή τής κατανομής αυτής.

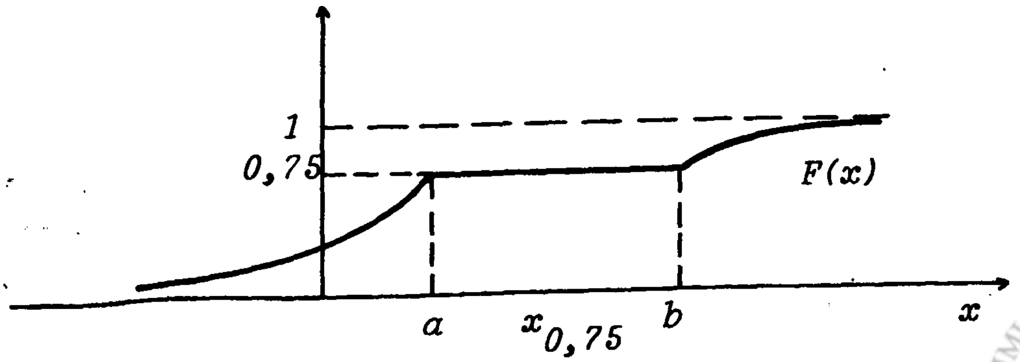
Γραφικά τά ποσοστιαΐα σημεΐα παρίστανται εύκολα μέ τή βοήθεια τής α.σ.κ.(βλέπε Σχήμα 5.3).

Όταν ή $F(x)$ είναι γνησίως αύξουσα τά x_p ποσοστιαΐα σημεΐα ορίζονται μονοσήμαντα γιά κάθε p. Άν ή $F(x)$ δέν είναι γνησίως αύξουσα όπως π.χ. στό Σχήμα 5.4 τότε τά x_p ποσοστιαΐα σημεΐα είναι δυνατόν νά μήν προσδιορίζονται



Σχῆμα 5.3 Ποσοστιαῖα σημεῖα

μονοσήμαντα γιά κάθε p. Ἄν $p = 0,75$ ἀπό τό σχῆμα φαίνεται ὅτι κάθε σημεῖο y τοῦ διαστήματος $[a,b]$ ἔχει τήν ἰδιότητα $F(y) = 0,75$. Σ'αὐτή τήν περίπτωση διαλέγουμε συνήθως ὡς $x_{0,75}$ ποσοστιαῖο σημεῖο τό σημεῖο $(a+b)/2$. Ὅταν



Σχῆμα 5.4

ή X είναι ἀσυνεχής δέν είναι πάντα δυνατή ή εὕρεση σημείων x_p τέτοιων ὥστε $F(x_p) = p$. Ἔτσι τροποποιεῖται ὁ ὁρισμός αὐτός καί προκύπτει τό σκέλος τοῦ Ὁρισμοῦ 5.6 γιά τή διακριτή τ.μ. Τό σκέλος αὐτό θά μποροῦσε νά χρησιμοποιηθεῖ καί γιά τήν περίπτωση συνεχοῦς τ.μ.

Ἄν $p = 0{,}01,\ 0{,}02, \ldots, 0{,}99$, τά σημεῖα $x_{0,01}, x_{0,02}, \ldots, x_{0,99}$ λέγονται καί πρῶτο, δεύτερο,..., ἐνενηκοστό ἐνατο ἑκατοστιαῖα σημεῖα ἤ ἑκατοστημόρια *(percentiles)*. Ἄν $p = 0{,}5$ τό σημεῖο λέγεται καί διάμεσος τῆς κατανομῆς ἤ τῆς τ.μ. καί συνήθως συμβολίζεται μέ τό m.

Τά σημεῖα $x_{0,25}$, $x_{0,50}$, $x_{0,75}$ λέγονται σημεῖα τοῦ πρώτου, δευτέρου καί τρίτου τεταρτημορίου. Χωρίζουν τήν κατανομή σέ τέσσερα ἴσα μέρη πού λέγονται τεταρτημόρια, πρῶτο, δεύτερο, τρίτο καί τέταρτο ἀντίστοιχα. Ἡ διαφορά $x_{0,75} - x_{0,25}$ λέγεται ἐνδοτεταρτημοριακό πλάτος *(interquantile range)*

Παραδείγματα:

5.8: Ἔστω X τ.μ. μέ ὁμοιόμορφο $U(0,1)$ κατανομή καί $p = 2/3$. Ἐπειδή $F(x) = x,\ 0 < x < 1$, τό $x_{2/3}$ ποσοστιαῖο σημεῖο εἶναι $F(x_{2/3}) = 2/3$ ἄρα $x_{2/3} = 2/3$. Ἐδῶ τά ποσοστιαῖα σημεῖα εἶναι μονοσήμαντα ὁρισμένα.

5.9: Νά βρεθοῦν τά x_p ποσοστιαῖα σημεῖα γιά μία τ.μ. X μέ κατανομή $Ek\theta(\lambda)$.

Ἕνα τέτοιο σημεῖο θά πρέπει νά ἱκανοποιεῖ τή σχέση

$$P(X \leq x_p) = F(x_p) = p\ ,$$

ὅπου

$$F(x_p) = \int_0^{x_p} \lambda e^{-\lambda y} dy = 1-e^{-\lambda x_p} \quad .$$

Συνεπῶς

$$1-e^{-\lambda x_p} = p \quad \text{ἤ} \quad x_p = -\frac{1}{\lambda} \ln(1-p).$$

Καί πάλι τά ποσοστιαῖα σημεῖα εἶναι μονοσήμαντα ὁρισμένα.

5.10: Ἔστω X συνεχής τ.μ. μέ κατανομή

$$f(x) = \begin{cases} \frac{1}{2}, & 0 \le x \le 1 \\ 1, & 2,5 < x < 3 \\ 0, & \text{ἀλλοῦ.} \end{cases}$$

Τότε ἐπειδή γιά κάθε m, $1 \le m \le 2,5$ ἔχουμε

$$P(X\le m) = \int_0^m f(x)dx = \int_0^1 \frac{1}{2}\, dx = \frac{1}{2}$$

κάθε σημεῖο τοῦ διαστήματος $[1,\ 2,5]$ μπορεῖ νά ληφθεῖ ὡς διάμεσος τῆς κατανομῆς. Ἡ διάμεσος δέν εἶναι μονοσήμαντα ὁρισμένη. Ἐπίσης εἶναι φανερό ὅτι κάθε σημεῖο τοῦ διαστήματος $[2,5,\ 3]$ μπορεῖ νά ληφθεῖ ὡς κορυφή τῆς κατανομῆς.

5.11: Ἄν X εἶναι διακριτή τ.μ. μέ

$$P(X=1) = 0,1 \ , \ P(X=2) = 0,3 \ , \ P(X=3) = 0,2 \ , \ P(X=4) = 0,4$$

τότε ἐπειδή $P(X\le 3) = 0,6>1/2$ καί $P(X\ge 3) = 0,6>1/2$ τό σημεῖο $X = 3$ εἶναι ἡ μοναδική διάμεσος τῆς κατανομῆς. Ἡ κορυφή αὐτῆς τῆς κατανομῆς εἶναι τό σημεῖο $X = 4$.

5.12: Ἔστω X διακριτή τ.μ. μέ σ.π.

$$p(x) = \begin{cases} \frac{1}{4} \ , & x = 1,2,3,4 \\ 0 \ , & \text{ἀλλοῦ} \ . \end{cases}$$

Εἶναι εὔκολο νά δεῖ κανείς ὅτι ὅλα τά σημεῖα τοῦ διαστήματος $[2,3]$ ἱκανοποιοῦν τόν ὁρισμό τῆς διαμέσου. Π.χ. $m = 2$ διότι $P(X \leq 2) = 1/2 \geq 1/2$ καί $P(X \geq 2) = 3/4 \geq 1/2$.

5.13: Ἔστω X ἡ τ.μ. πού παριστάνει τήν ποσότητα τοῦ περιεχομένου φιαλῶν πορτοκαλάδας καί ἔστω ἐπί πλέον ὅτι $X \sim N(\mu=1$ λίτρο, $\sigma = 0,05$ λίτρα). Νά βρεθεῖ ἡ ποσότητα x τέτοια ὥστε 20% τῶν φιαλῶν νά ἔχουν ποσότητα μικρότερη τοῦ x.

Ἐδῶ τό ζητούμενο εἶναι τό σημεῖο $x_{0,2}$. Ἔχουμε

$$P(X \leq x_{0,2}) = 0,2$$

ἤ

$$P\left(\frac{X-\mu}{\sigma} \leq \frac{x_{0,2} - 1}{0,05}\right) = 0,2$$

καί ἀπό τόν Πίνακα III παίρνουμε

$$\frac{x_{0,2} - 1}{0,05} = -2,055$$

ἤ

$$x_{0,2} = 1+0,05\times(-2,055) = 0,89725 \text{ λίτρα.}$$

Γενικά γιά τήν κανονική κατανομή ἰσχύει ἡ σχέση

$$x_p = \mu + \sigma\Phi^{-1}(p). \qquad (5.45)$$

5. 7 Ταξινόμηση τῶν χαρακτηριστικῶν κατανομῶν

Ἔχοντας συμπληρώσει τήν μελέτη τῶν χαρακτηριστικῶν τῶν κατανομῶν δίνουμε μία ταξινόμηση αὐτῶν σχετικά μέ τήν πληροφορία πού μᾶς δίνει τό κάθε ἕνα ὅσον ἀφορᾶ τήν μορφή τῆς κατανομῆς.

α) Μέτρα θέσεως τῆς κατανομῆς:

Μέση τιμή, διάμεσος, κορυφή ποσοστιαῖα σημεῖα $(x_{0,25}+x_{0,75})/2$

β) Μέτρα διασπορᾶς τῆς κατανομῆς.

διακύμανση, τυπική ἀπόκλιση συντελεστής μεταβλητότητας = σ/μ ἐνδοτεταρτημοριακό πλάτος $x_p - x_{1-p}$, $1/2 < p < 1$

γ) Μέτρα λοξότητας τῆς κατανομῆς:

συντελεστής α_3

δ) Μέτρα κυρτώσεως τῆς κατανομῆς:

$$\boxed{\text{συντελεστής } \alpha_4}$$

5.8 Γεννήτριες συναρτήσεις

Στά προηγούμενα ἐδάφια ὁρίσαμε τίς διάφορες ροπές μίας τ.μ. X ἤ μίας συναρτήσεως αὐτῆς $h(X)$. Ὑπάρχουν συναρτήσεις, οἱ λεγόμενες γ ε ν ν ή τ ρ ι ε ς συναρτήσεις, οἱ ὁποῖες μποροῦν νά χρησιμοποιηθοῦν γιά τήν εὕρεση τῶν ροπῶν καί οἱ ὁποῖες εἶναι χαρακτηριστικές γιά τήν κάθε κατανομή. Οἱ γεννήτριες αὐτές συναρτήσεις διακρίνονται σέ τρεῖς κατηγορίες: ρ ο π ο γ ε ν ν ή τ ρ ι ε ς, π α ρ α γ ο ν τ ι κ έ ς ρ ο π ο γ ε ν ν ή τ ρ ι ε ς καί χ α ρ α κ τ η ρ ι σ τ ι κ έ ς σ υ ν α ρ τ ή σ ε ι ς. Ἐδῶ, γιά λόγους μαθηματικοῦ ὑπόβαθρου θά περιοριστοῦμε στή μελέτη μόνον τῶν δύο πρώτων κατηγοριῶν καί ἰδιαίτερα τῆς πρώτης παρ'ὅλο πού ἡ τρίτη κατηγορία εἶναι ἡ πιό ἐνδιαφέρουσα.

5.8.1 Ροπογεννήτριες συναρτήσεις.

Ὁρισμός 5.7: Ἔστω X μία τ.μ. Ἡ ρ ο π ο γ ε ν ν ή τ ρ ι α σ υ ν ά ρ τ η σ η τῆς X συμβολίζεται μέ $m_X(t)$ ἤ ἁπλῶς $m(t)$ καί ὁρίζεται ὡς ἀκολούθως

$$m_X(t) = E(e^{tX}) = \begin{cases} \sum\limits_x e^{tx}p(x), & \text{ὅταν } X \text{ εἶναι διακριτή} \\ \int\limits_{-\infty}^{+\infty} e^{tx}f(x)dx, & \text{ὅταν } X \text{ εἶναι συνεχής}, \end{cases} \qquad (5.46)$$

ὅπου $p(x)$ εἶναι ἡ σ.π. τῆς X, $f(x)$ ἡ σ.π.π. τῆς

X καί t εἶναι παράμετρος μεταβαλλόμενη στά διαστήματα γιά τά ὁποῖα τό ἄθροισμα ἤ τό ὁλοκλήρωμα συγκλίνουν ἀπόλυτα δηλαδή εἶναι $<\infty$.

Ἡ $m_X(t)$ ὑπάρχει πάντοτε στό σημεῖο $t = 0$, ὅπως ἄμεσα προκύπτει ἀπό τόν ὁρισμό της. Μάλιστα δέ ἰσχύει

$$m_X(0) = 1$$

γιά κάθε τ.μ. X . Ὑπάρχουν ὅμως περιπτώσεις πού ἡ $m_X(t)$ δέν ὑπάρχει γιά καμμία ἄλλη τιμή τοῦ t ἐκτός τῆς $t = 0$ (βλέπε Ἄσκηση 5.46). Ἄν ἡ $m_X(t)$ ὑπάρχει σέ μία περιοχή τοῦ $t = 0$ τότε ἔχει παραγώγους κάθε τάξεως στήν περιοχή αὐτή. Σέ ὅτι ἀκολουθεῖ θά δεχθοῦμε ὅτι ὅλες οἱ ὑπό συζήτηση ροπογεννήτριες συναρτήσεις ὑπάρχουν τουλάχιστο σέ ἕνα διάστημα $(-c, c)$, $c > 0$.

Θά δώσουμε τώρα τίς ροπογεννήτριες τῶν πιό γνωστῶν ἀπό τίς εἰδικές κατανομές πού συναντήσαμε μέχρι τώρα.

α) Διωνυμική $B(n,p)$.

$$\boxed{m(t) = (pe^t+q)^n \quad \text{γιά κάθε} \quad t \in R.} \qquad (5.47)$$

Ἀπόδειξη:

$$m(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx}\binom{n}{x}p^x q^{n-x}$$

$$= \sum_{x=0}^{n}\binom{n}{x}(pe^t)^x q^{n-x} = (pe^t+q)^n. \blacktriangledown$$

β) Γεωμετρική $Geo\ (p)$.

$$\boxed{m(t) = \frac{pe^t}{1-qe^t} \quad \text{γιά κάθε} \quad t < -ln\ p\ .} \qquad (5.48)$$

'Απόδειξη:

$$m(t) = \sum_{x=1}^{\infty} e^{tx} pq^{x-1} = \sum_{x=1}^{\infty} (qe^t)^x$$

$$= \frac{pqe^t}{q} \sum_{x=1}^{\infty} (qe^t)^{x-1}$$

καί ἄν $qe^t < 1$ δηλαδή $t < -ln\ p$

$$m(t) = \frac{p}{q} \frac{qe^t}{1-qe^t} \ . \ ▼$$

γ) 'Α ρ ν η τ ι κ ή Δ ι ω ν υ μ ι κ ή $NB(k,p)$.

$$\boxed{m(t) = \frac{p^k e^{tk}}{(1-qe^t)^k}, \quad t < -ln\ q\ .} \qquad (5.49)$$

'Απόδειξη: 'Από τόν ὁρισμό καί τήν σχέση (4.9) ἔχουμε

$$m(t) = \sum_{x=k}^{\infty} e^{tx} \binom{x-1}{k-1} p^k q^{x-k} = \sum_{x=k}^{\infty} \binom{x-1}{k-1} p^k (e^t q)^{x-k} e^{tk}$$

$$= p^x e^{tk} \sum_{x=k}^{\infty} \binom{x-1}{k-1} (e^t q)^{x-k}$$

$$= p^x e^{tk} \frac{1}{(1-e^t q)^k} \quad \text{ἄν} \quad t < -ln\ q\ . \ ▼$$

δ) *Poisson* $P(\lambda)$.

$$\boxed{m(t) = e^{\lambda(e^{t}-1)} \quad \text{γιά κάθε} \quad t \in R\ .} \qquad (5.50)$$

Ἀπόδειξη:

$$m(t) = \sum_{x=0}^{\infty} e^{tx}\frac{\lambda^{x}e^{-\lambda}}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^{t})^{x}}{x!}$$

$$= e^{-\lambda}e^{\lambda^{t}}\ . \blacktriangledown$$

ε) Ὁμοιόμορφη $U(\alpha,\beta)$.

$$\boxed{m(t) = \frac{e^{\beta t}-e^{\alpha t}}{t(\beta-\alpha)} \quad \text{γιά κάθε} \quad t \in R.} \qquad (5.51)$$

Μπορεῖ εὔκολα νά ἀποδειχθεῖ μέ ἁπλή ὁλοκλήρωση.

ζ) Ἐκθετική $Ek\theta(\lambda)$.

$$\boxed{m(t) = \frac{\lambda}{\lambda-t}, \quad t < \lambda\ .} \qquad (5.52)$$

Ἀπόδειξη:

$$m(t) = \int_{0}^{\infty} e^{tx}\lambda e^{-\lambda x}dx = \lambda\int_{0}^{\infty} e^{-(\lambda-t)x}dx$$

$$= \frac{\lambda}{\lambda-t}\int_{0}^{\infty} e^{-y}dy \qquad (y=(\lambda-t)x, \lambda-t>0)$$

$$= -\frac{\lambda}{\lambda-t}e^{-y}\Big|_{0}^{\infty} = \frac{\lambda}{\lambda-t},$$

ὑπό τόν ὅρο ὅτι $\lambda - t > 0$. ▼

η) Γ ά μ μ α $G(\alpha,\beta)$.

$$\boxed{m(t) = (1-\beta t)^{-\alpha}, \quad t < 1/\beta .} \qquad (5.53)$$

'Απόδειξη:

$$m(t) = \int_0^\infty e^{tx} \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} dx = \int_0^\infty \frac{x^{\alpha-1} e^{-x((1/\beta)-t)}}{\beta^\alpha \Gamma(\alpha)} dx$$

$$= \int_0^\infty x^{\alpha-1} \frac{e^{-\frac{x}{\beta/(1-\beta t)}}}{\beta^\alpha \Gamma(\alpha)} dx$$

καί ἄν $\frac{1}{\beta} - t > 0$ δηλαδή $t < \frac{1}{\beta}$ ἔχουμε

$$m(t) = \frac{\left(\frac{\beta}{1-\beta t}\right)^\alpha \Gamma(\alpha)}{\beta^\alpha \Gamma(\alpha)} = (1-\beta t)^{-\alpha}. \quad ▼$$

ϑ) Κ α ν ο ν ι κ ή $N(\mu,\sigma^2)$.

$$\boxed{m(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}, \quad t \in R .} \qquad (5.54)$$

'Απόδειξη:

$$m(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2-2tx\sigma^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{[x-(\mu+t\sigma^2)]^2-(2t\mu+t^2\sigma^2)\sigma^2}{2\sigma^2}} dx$$

$$= e^{\mu t+\frac{1}{2}\sigma^2 t^2} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{[x-(\mu+t\sigma)]^2}{2\sigma^2}} dx$$

$$= e^{\mu t+\frac{1}{2}\sigma^2 t^2} .$$

Τό ὁλοκλήρωμα

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{[x-(\mu+t\sigma)]^2}{2\sigma^2}} dx = 1$$

διότι ἡ ὁλοκληρωτέα συνάρτηση εἶναι σ.π.π. κανονικῆς κατανομῆς μέ μέση τιμή $\mu+t\sigma$ καί διακύμανση σ^2. ▼

'Ο παραπάνω τρόπος ὁλοκληρώσεως λέγεται " σ υ μ π λ ή ρ ω σ η τ ο ῦ τ ε τ ρ α γ ώ ν ο υ"

"Αν στήν ἔκφραση (5.46) εἶναι δυνατή ἡ παραγώγιση ὡς πρός t ἐντός τῶν συμβόλων ἀθροίσεως καί ὁλοκληρώσεως τότε έχουμε

$$\frac{d^r m_X(t)}{dt^r} = \left\{ \begin{array}{l} \sum_x x^r e^{tx} p(x) \\ \\ \int_{-\infty}^{+\infty} x^r e^{tx} f(x) dx \end{array} \right\} = E(x^r e^{tX})$$

δηλαδή

$$\frac{d^r m_X(t)}{dt^r} = E(x^r e^{tX}) \quad , \quad r = 1,2,\ldots$$

καί εάν $t = 0$ έχουμε

$$\left.\frac{d^r m_X(t)}{dt^r}\right|_{t=0} = \lim_{t \to 0} \frac{d^r m_X(t)}{dt^r} = E(X^r) = \mu_r . \quad (5.55)$$

Έτσι βλέπουμε ότι ἡ π α ρ ά γ ω γ ο ς r-τ ά ξ ε ω ς τ ῆ ς $m_X(t)$ σ τ ό σ η μ ε ῖ ο $t = 0$ ἰ σ ο ῦ τ α ι, κ ά τ ω ἀ π ό ὡ ρ ι σ μ έ ν ε ς σ υ ν θ ῆ κ ε ς, μ έ τ ή ν π ε ρ ί τ ό μ η δ έ ν ρ ο π ή τ ά ξ ε ω ς r τ ῆ ς τ.μ. X. Σέ αὐτό τό συμπέρασμα βασίζεται καί τό ὄνομα πού δώσαμε σ'αὐτές τίς συναρτήσεις.

Ἡ παραπάνω ἰδιότητα εἶναι πολύ σημαντική διότι μᾶς ἐπιτρέπει τόν ὑπολογισμό τῶν ροπῶν μίας τ.μ. μέ τή βοήθεια τῆς ροπογεννήτριας συναρτήσεως της.

Παράδειγμα 5.14:

Ἔστω X μία *Poisson* τ.μ. μέ παράμετρο λ. Μέ τή βοήθεια τῆς ροπογεννήτριας συναρτήσεως νά βρεθοῦν ἡ μέση τιμή καί ἡ διακύμανση τῆς X.

Ἔχουμε

$$m_X(t) = e^{\lambda(e^t-1)} \quad , \quad t \in R ,$$

$$m_X'(t) = \lambda e^t e^{\lambda(e^t-1)}$$

καί

$$m''(t) = \lambda e^t e^{\lambda(e^t-1)} + \lambda^2 e^{2t} e^{\lambda(e^t-1)} .$$

Γιά $t = 0$ ἔχουμε

$$m'(0) = \lambda = \mu_1 = \mu = E(X)$$

$$m''(0) = \lambda+\lambda^2 = \mu_2 = E(X^2).$$

Ἔτσι $\mu = \lambda$ καί $\sigma^2 = \lambda+\lambda^2-\lambda^2 = \lambda$.

Ἐάν στήν *(5.46)* ἀντικαταστήσουμε τό e^{tx} μέ τό ἀνάπτυγμα του

$$e^{tx} = 1+ \frac{tx}{1!} + \frac{(tx)^2}{2!} + \dots$$

τότε, κάτω ἀπό ὡρισμένες συνθῆκες [ἀρκεῖ νά ὑπάρχει ἡ $m_X(t)$ γιά κάθε t σέ κάποιο διάστημα $(-c,c)$], παίρνουμε

$$m_X(t) = 1+ \mu_1 \frac{t}{1!} + \mu_2 \frac{t^2}{2!} + \dots$$

ἤ

$$m_X(t) = 1 + \sum_{r=1}^{\infty} \mu_r \frac{t^r}{r!} \quad . \qquad (5.56)$$

Βλέπουμε δηλαδή ὅτι οἱ συντελεστές τοῦ $t^r/r!$ στό ἀνάπτυγμα τῆς $m_X(t)$ σέ σειρά *Mac Laurin* εἶναι οἱ ροπές τάξεως r περί τό μηδέν.

Μερικές φορές ἡ ἐφαρμογή τῆς μεθόδου τῆς παραγωγίσεως τῆς ροπογεννήτριας [βλέπε *(5.55)*] γιά τήν εὕρεση τῶν ροπῶν μίας κατανομῆς ὁδηγεῖ σέ ἀπροσδιοριστίες ὅταν $t = 0$. Γιά νά ἀποφύγουμε τίς ἀπροσδιοριστίες αὐτές καί τόν κανόνα τοῦ *L'Hospital*, χρησιμοποιοῦμε τό ἀνάπτυγμα *(5.56)*. Γιά παράδειγμα ἄς πάρουμε τήν ὁμοιόμορφη κατανομή $U(\alpha,\beta)$. Ἡ *(5.51)* μᾶς δίνει

$$m_X(t) = \frac{e^{\beta t} - e^{\alpha t}}{(\beta-\alpha)t} \quad , \quad t \in R \; .$$

Χρησιμοποιοῦντες τά σέ σειρά ἀναπτύγματα τῶν $e^{\beta t}$ καί $e^{\alpha t}$ παίρνουμε

$$m\,(t) = \frac{1}{(\beta-\alpha)t}\left\{1+ \frac{\beta t}{1!} + \frac{(\beta t)^2}{2!} +\ldots-\left[1+ \frac{\alpha t}{1!} + \frac{(\alpha t)^2}{2!} +\ldots\right]\right\}$$

$$= \frac{1}{(\alpha-\beta)t}\left[(\beta-\alpha)t+ \frac{(\beta^2-\alpha^2)t^2}{2!} + \frac{(\beta^3-\alpha^3)t^3}{3!} + \ldots\right]$$

$$= 1+ \frac{\beta+\alpha}{2}\, t+ \frac{\beta^2+\alpha\beta+\alpha^2}{3}\, \frac{t^2}{2!} + \ldots$$

Ἔτσι $\mu_1 = (\alpha+\beta)/2$, $\mu_2 = (\alpha^2+\alpha\beta+\beta^2)/3$ κλπ.

Τά ἀποτελέσμτα αὐτά συμφωνοῦν μέ τήν (5.30).

Παρατηρήσεις:

1. Εἶναι δυνατό μία τ.μ. X νά μήν ἔχει καμμία ροπή ἀλλά νά ἔχει ροπογεννήτρια συνάρτηση. Βλέπε Ἄσκηση 5.47.

2. Εἶναι δυνατό μία τ.μ. X νά ἔχει ὅλες ἤ μερικές ροπές καί νά μήν ἔχει ροπογεννήτρια συνάρτηση (ἐκτός ἀπό τό σημεῖο $t=0$). Βλέπε Ἄσκηση 5.49.

3. Εἶναι δυνατό μία τ.μ. X νά ἔχει ὅλες ἤ μερι- ἀπό τίς ροπές της, νά ἔχει ροπογεννήτρια συνάρτηση καί ἐν τούτοις ἡ ροπογεννήτρια συνάρτηση νά μήν παράγει τίς ροπές. Βλέπε Ἄσκηση 5.48.

Ὑπάρχουν παραδείγματα πού τεκμηριώνουν τίς παραπάνω παρατηρήσεις.

Τό ἑπόμενο θεώρημα ἀφορᾶ τήν ροπογεννήτρια μίας τ.μ. πού εἶναι γραμμικός μετασχηματισμός μίας ἄλλης τ.μ.

Θεώρημα 5.7:

Ἄν $m_X(t)$ εἶναι ἡ ροπογεννήτρια μίας τ.μ. X καί $Y = \alpha X+b$ εἶναι μία ἄλλη τ.μ. τότε

$$m_Y(t) = m_{\alpha X+b}(t) = e^{bt} m_X(\alpha t) . \qquad (5.57)$$

'Απόδειξη:

$$m_Y(t) = E[e^{(\alpha X+b)t}] = E[e^{\alpha tX}e^{bt}] = e^{bt}E[e^{\alpha tX}]$$

$$= e^{bt}m_X(\alpha t). \blacktriangledown$$

'Από τά προηγούμενα εἶναι φανερό ὅτι ἄν ἔχουμε δύο τ.μ. X, Y πού ἔχουν τίς ἴδιες κατανομές τότε $m_X(t) = m_Y(t)$ γιά ὅλες τίς τιμές τοῦ t. Μέ ἄλλα λόγια ἡ κατανομή ὁρίζει μονοσήμαντα τήν ροπογεννήτρια συνάρτηση. Τό ἐρώτημα πού τίθεται εἶναι ἄν ἡ ροπογεννήτρια συνάρτηση ὁρίζει μονοσήμαντα τήν κατανομή. Ἄν οἱ ροπογεννήτριες συναρτήσεις δύο τ.μ. συμπίπτουν τί συμβαίνει μέ τίς ἀντίστοιχες κατανομές; Ἡ ἀπάντηση εἶναι ὅτι καί οἱ ἀντίστοιχες κατανομές ταυτίζονται ὅπως φαίνεται ἀπό τό ἀκόλουθο θεώρημα. Θά δώσουμε μόνο τήν διατύπωση τοῦ θεωρήματος. Ἡ ἀπόδειξη του ξεφεύγει ἀπό τό σκοπό τοῦ βιβλίου.

Θεώρημα 5.8: (Θεώρημα μονοσήμαντου τῶν ροπογεννητριῶν).

Ἄν οἱ ροπογεννήτριες συναρτήσεις δύο τ.μ. X, Y συμπίπτουν δηλαδή ἄν $m_X(t) = m_Y(t)$ γιά κάθε t τότε καί οἱ ἀντίστοιχες κατανομές συμπίπτουν.

Τό παραπάνω θεώρημα εἶναι θεμελιῶδες. Μᾶς ἐπιτρέπει τόν προσδιορισμό τῶν κατανομῶν τυχαίων μεταβλητῶν ὅταν εἶναι γνωστές οἱ ροπογεννήτριες τους καί μποροῦν νά ταυτιστοῦν μέ τίς ροπογεννήτριες γνωστῶν κατανομῶν. Τά παραδείγματα πού ἀκολουθοῦν δείχνουν πῶς ἐφαρμόζεται τό Θεώρημα 5.8.

Παραδείγματα:

5.15: Ἄν ἡ ροπογεννήτρια μιᾶς τ.μ. εἶναι

$$m(t) = \frac{1}{10} e^{t} + \frac{2}{10} e^{2t} + \frac{3}{10} e^{3t} + \frac{4}{10} e^{4t}, \quad t \in R,$$

νά βρεθεῖ ἡ κατανομή (σ.π.) τῆς X.

Ἐπειδή

$$m(t) = \sum_{x} e^{tx} p(x)$$

ἔχουμε

$$\frac{1}{10} e^{t} + \frac{2}{10} e^{2t} + \frac{3}{10} e^{3t} + \frac{4}{10} e^{4t} \equiv p(\alpha_1) e^{\alpha_1 t} + p(\alpha_2) e^{\alpha_2 t} + \ldots$$

γιά κάθε $t \in R$, ὅπου $\alpha_1, \alpha_2, \ldots$ εἶναι οἱ τιμές τῆς X. Ἐπειδή ἡ σχέση αὐτή εἶναι μία ταυτότητα ὡς πρός t· τό δεύτερο μέλος θά πρέπει νά ἔχει τούς ἴδιους ἀκριβῶς ὅρους μέ τό πρῶτο. Συνεπῶς

$$\frac{1}{10} e^{t} = p(\alpha_1) e^{\alpha_1 t}, \quad \frac{2}{10} e^{2t} = p(\alpha_2) e^{\alpha_2 t}, \ldots$$

Ἔτσι ἔχουμε

$$p(\alpha_1) = p(1) = \frac{1}{10}, \quad p(\alpha_2) = p(2) = \frac{2}{10},$$

$$p(\alpha_3) = p(3) = \frac{3}{10}, \quad p(\alpha_4) = p(4) = \frac{4}{10}.$$

Ἄρα

$$p(x) = \begin{cases} \frac{x}{10}, & x = 1,2,3,4 \\ 0, & \text{ἀλλοῦ}. \end{cases}$$

5.16: Ἂν

$$m(t) = \frac{1}{(1-t/5)^2} \quad , \quad t < 5$$

εἶναι ἡ ροπογεννήτρια μιᾶς τ.μ. X τότε ἡ κατανομή της βρίσκεται ὡς ἑξῆς. Συγκρίνοντας τήν ἔκφραση αὐτή μέ τήν ἔκφραση (5.53) πού δίνει τήν ροπογεννήτρια τῆς Γάμμα κατανομῆς ἔχουμε

$$\frac{1}{(1-\beta t)^\alpha} = \frac{1}{(1-t/5)^2} \ .$$

Ἔτσι συμπεραίνουμε ὅτι ἡ δοθεῖσα συνάρτηση εἶναι ἡ ροπογεννήτρια Γάμμα κατανομῆς μέ παράμετρους $\alpha = 2$ καί $\beta = 1/5$ καί σύμφωνα μέ τό θεώρημα μονοσήμαντου τῶν ροπογεννητριῶν ἡ κατανομή τῆς X θά εἶναι

$$f(x) = \frac{1}{(1/5)^2\Gamma(2)}\, x e^{-\frac{x}{1/5}} = 25xe^{-5x} \quad , \ x > 0 \ .$$

5.17: Ἂν ἡ ροπογεννήτρια μιᾶς τ.μ. X εἶναι

$$m(t) = e^{t^2/2}$$

τότε οἱ ροπές περί τό μηδέν αὐτῆς μποροῦν νά βρεθοῦν ὡς ἑξῆς. Ἀναπτύσσουμε τήν $m(t)$ σέ σειρά *Mac Laurin* καί ἔχουμε

$$e^{t^2/2} = 1 + \frac{t^2}{2}\frac{1}{1!} + \left(\frac{t^2}{2}\right)^2\frac{1}{2!} + \ldots + \left(\frac{t^2}{2}\right)^k\frac{1}{k!} + \ldots$$

$$= 1 + \frac{1}{2!}t^2 + \frac{3\ 1}{4!}t^4 + \ldots + \frac{(2k-1)(2k-3)\ldots 3\ 1}{(2k)!}t^{2k} + \ldots$$

καί σύμφωνα μέ τήν σχέση *(5.56)* ἔχουμε

$$E(X^{2k}) = (2k-1)(2k-3)\ldots 3\times 1 = \frac{(2k)!}{2^k k!}$$

5.8.2 Πιθανογεννήτριες ἤ Παραγοντικές Ροπογεννήτριες Συναρτήσεις

Ὁρισμός 5.8: Ἡ πιθανογεννήτρια συνάρτηση ἤ παραγοντική ροπογεννήτρια συνάρτηση μιᾶς τ.μ. X συμβολίζεται μέ $g_X(t)$ ἤ ἁπλῶς $g(t)$ καί ὁρίζεται ὡς ἀκολούθως

$$g_X(t) = E(t^X) = \begin{cases} \sum\limits_x t^x p(x) & \text{ἄν ἡ } X \text{ εἶναι διακριτή} \\ \int_{-\infty}^{+\infty} t^x f(x)dx & \text{ἄν ἡ } X \text{ εἶναι συνεχής.} \end{cases} \qquad (5.58)$$

Γιά τήν ὕπαρξη τῆς $g_X(t)$ ἰσχύουν τά ἴδια σχόλια πού ἔγιναν γιά τήν $m_X(t)$. Σημειώνουμε μόνον ὅτι γιά $|t| < 1$ ἡ $g_X(t)$ ὑπάρχει πάντοτε. Ἄν ἰσχύουν οἱ συνθῆκες πού ἐπιτρέπουν παραγώγιση κάτω ἀπό τό σύμβολο ὁλοκληρώσεως καί μέσα στό σύμβολο τῆς ἀθροίσεως τότε ἔχουμε

$$\frac{d^r g_X(t)}{dt^r} = \begin{cases} \sum\limits_x x(x-1)\ldots(x-r+1)t^{x-r}p(x) \\ \int_{-\infty}^{+\infty} x(x-1)\ldots(x-r+1)t^{x-r}f(x)dx \end{cases}$$

$$= E[X(X-1)\ldots(X-r+1)t^{X-r}]$$

Ἄν σ'αὐτή τήν ἰσότητα θέσουμε $t = 1$ ἔχουμε

$$\left.\frac{d^r g_X(t)}{dt^r}\right|_{t=1} = E[X(X-1)\ldots(X-r+1)] = \mu_{[r]} \tag{5.59}$$

Ἔτσι βλέπουμε ὅτι ἀπό τήν συνάρτηση $g_X(t)$ μέ παραγώγιση στό σημεῖο $t = 1$ μποροῦμε νά πάρουμε τίς παραγοντικές ροπές.

Ἄν ἡ τ.μ. παίρνει τίς τιμές $0,1,2,3,\ldots$ τότε

$$g_X(t) = E(t^X) = p_0+p_1t+p_2t^2+\ldots \tag{5.60}$$

Ἔτσι ὁ συντελεστής τοῦ t^r τοῦ ἀναπτύγματος τῆς $g_X(t)$ σέ δυνάμεις εἶναι ἡ πιθανότητα $P(X=r)$.

Ἄς δοῦμε τώρα ἕνα παράδειγμα. Ἡ πιθανογεννήτρια συνάρτηση τῆς κατανομῆς τοῦ *Poisson* $P(\lambda)$ εἶναι

$$g_X(t) = e^{\lambda(t-1)}$$

Ἐπιπλέον

$$g_X'(t) = \lambda e^{\lambda(t-1)} \quad \text{καί} \quad g''(t) = \lambda^2 e^{\lambda(t-1)}$$

Ἔτσι

$$g_X'(1) = \lambda = \mu_{[1]} = \lambda_1 = \mu_1$$

$$g_X''(1) = \lambda^2 = \mu_{[2]} = E[X(X-1)]$$

ὅπως ἔχουμε ἤδη δεῖ στά προηγούμενα.

Ἀκόμη ἀναπτύσσοντας τήν $g_X(t)$ σέ σειρά ἔχουμε

$$g_X(t) = e^{-\lambda}\left(1+ \frac{\lambda t}{1!} + \frac{(\lambda t)^2}{2!} +\ldots+ \frac{(\lambda t)^r}{r!} + \ldots\right)$$

$$= e^{-\lambda}+ \frac{e^{-\lambda}\lambda^1}{1!}\, t+ \frac{e^{-\lambda}\lambda^2}{2!}\, t^2+\ldots+ \frac{e^{-\lambda}\lambda^r}{r!}\, t^r+\ldots\ .$$

Βλέπουμε δηλαδή ὅτι οἱ συντελεστές τοῦ t^r εἶναι ἀκριβῶς οἱ πιθανότητες $P(X=r)$.

Οἱ δύο συναρτήσεις $m_X(t)$ καί $g_X(t)$ συνδέονται μέ τή σχέση

$$g_X(t) = m_X(\log t), \quad t > 0\ . \qquad (5.61)$$

Οἱ σχέσεις αὐτές προκύπτουν εὔκολα ἀπό τίς (5.46) καί (5.58).

ΑΣΚΗΣΕΙΣ

5.1. Ἔστω X τ.μ. μέ σ.π.

$$p(x) = \frac{\alpha^x}{(1+\alpha)^{x+1}} \qquad x = 0,\ 1,\ 2,\ \ldots\ ,$$

ὅπου α θετική σταθερά. Νά δειχθεῖ ὅτι ἡ μέση τιμή τῆς X εἶναι α καί ἡ διακύμανση $\alpha(\alpha+1)$.

5.2. Δίνεται X τ.μ. μέ διωνυμική κατανομή $B(n,p)$. Νά βρεθοῦν οἱ τυποποιημένες ροπές α_3 καί α_4 τῆς X.

5.3. Νά ἀποδειχθοῦν οἱ σχέσεις

$$\lambda_3 = \mu_3-3\mu_2\mu_1+2\mu_1^3$$

$$\lambda_4 = \mu_4-4\mu_3\mu_1+6\mu_2\mu_1^2-3\mu_1^4$$

$$\lambda_k = \mu_k-\binom{k}{1}\mu_{k-1}\mu_1+\binom{k}{2}\mu_{k-2}\mu_1^2-\ldots+(-1)^k\binom{k}{k}\mu_1^k$$

ὅπου μ_k καί λ_k εἶναι οἱ ροπές k τάξεως μίας τ.μ. X περί τό μηδέν καί τή μέση τιμή μ ἀντίστοιχα.

5.4. Δίνεται X τ.μ. μέ ὑπεργεωμετρική κατανομή $Hg(N,n,p)$. Νά βρεθοῦν οἱ ροπές μ, σ^2, α_3 καί α_4 τῆς X.

5.5. Δίνεται X τ.μ. μέ γεωμετρική κατανομή $Geo(p)$. Νά βρεθοῦν οἱ ροπές α_3 καί α_4 τῆς X.

5.6. Δίνεται X τ.μ. μέ ἀρνητική διωνυμική κατανομή $NB(k,p)$. Νά βρεθοῦν οἱ ροπές μ, σ^3, α_3 καί α_4 τῆς X.

5.7. Δίνεται X τ.μ. μέ κατανομή τοῦ *Poisson* $P(\lambda)$. Νά βρεθοῦν οἱ ροπές α_3 καί α_4 τῆς X.

5.8. Δίνεται X τ.μ. μέ ὁμοιόμορφη κατανομή $U(\alpha,\beta)$. Νά βρεθοῦν οἱ ροπές α_3 καί α_4 τῆς X.

5.9. Δίνεται X τ.μ. μέ Γάμμα κατανομή $G(\alpha,\beta)$. Νά βρεθοῦν οἱ ροπές σ^2, α_3 καί α_4.

5.10. Δίνεται X τ.μ. μέ Βήτα κατανομή $Be(\alpha,\beta)$. Νά βρεθοῦν οἱ ροπές σ^2, α_3 καί α_4.

5.11. Δίνεται X τ.μ. μέ κανονική κατανομή $N(\mu,\sigma^2)$. Νά βρεθοῦν οἱ ροπές μ, σ^2, α_3 καί α_4.

5.12. Δίνεται ἡ κατανομή

x	0	1	2	3	4
$P(X=x)$	0,1	0,2	0,4	0,2	0,1

Νά βρεθοῦν α) ἡ δεύτερη παραγοντική ροπή, β) ὁ συντελεστής λοξότητας α_3, γ) ὁ συντελεστής κυρτώσεως α_4 καί δ) ἡ ροπογεννήτρια συνάρτηση.

5.13. Ἔστω X τ.μ. μέ σ.π.π.

$$f(x) = \begin{cases} 12(1-x^2)/11, & -\tfrac{1}{2} \leq x \leq \tfrac{1}{2} \\ 0 & , \text{ἀλλοῦ} \end{cases}$$

α) Νά βρεθοῦν ἡ $E(X)$ καί ἡ $E(|X|)$ καί β) νά ὑπολογιστεῖ ἡ $P(|X| \leq 1/4)$.

5.14. Ὁ χρόνος πέψεως, μετρούμενος σέ ὧρες, μιᾶς μονάδας τροφῆς εἶναι μία τυχαία μεταβλητή τῆς ὁποίας ἡ συνάρτηση πυκνότητας πιθανότητας εἶναι

$$f(x) = \begin{cases} 4xe^{-2x} & , \quad x > 0 \\ 0 & , \quad \text{ἀλλοῦ.} \end{cases}$$

(α) Ποιά εἶναι ἡ πιθανότητα μία μονάδα τροφῆς νά μήν ἔχει πλήρως χονευθεῖ μετά ἀπό μία ὥρα; (β)Ποιός χρόνος χρειάζεται κατά μέσο ὅρο γιά τή πέψη μίας μονάδας τροφῆς;

5.15. Δίνεται ἡ συνάρτηση

$$f(x) = \begin{cases} kx(1-x) & , \quad 0 \leq x \leq 1 \\ 0 & , \quad \text{ὁπουδήποτε ἀλλοῦ.} \end{cases}$$

Νά βρεθεῖ ἡ σταθερά k ὥστε ἡ $f(x)$ νά εἶναι κατανομή μίας τ.μ. X καί στή συνέχεια νά ὑπολογιστοῦν ἡ $E(X)$ καί $Var(X)$.

5.16. Ἄν $E(X) = 1$ καί $E(X^2) = 4$ νά βρεθοῦν ἡ μέση τιμή καί ἡ διακύμανση τῆς τ.μ. $Y = 2X-3$.

5.17. Νά ἀποδειχτεῖ ὅτι

$$[E(X-\alpha)]^2 \leq E(X-\alpha)^2$$

γιά κάθε α σταθερό.

5.18. Ἔστω X τυχαία μεταβλητή μέ κατανομή χ_4^2. Νά βρεθοῦν ἡ $E(X)$ καί ἡ $Var(X)$.

5.19. Ἔστω X διωνυμική τυχαία μεταβλητή μέ $n = 20$ καί $p = 1/3$. Νά δοθοῦν ἡ κορυφή (-φές) τῆς κατανομῆς αὐτῆς.

5.20. Νά βρεθεῖ ἡ μέση τιμή καί ἡ διακύμανση τῆς κατανομῆς μέ (ἀθροιστική) συνάρτηση κατανομῆς:

$$F(x) = \begin{cases} 0 & , \quad x < 0 \\ x/8 & , \quad 0 \leq x < 2 \\ x^2/16 & , \quad 2 \leq x < 4 \\ 1 & , \quad 4 \leq x \end{cases} .$$

5.21. Μία ἠλεκτρονική συσκευή ἔχει χρόνο ζωῆς X ὁ ὁποῖος μετρούμενος σέ κατάλληλες μονάδες χρόνου εἶναι μία τ.μ. μέ σ.π.π.

$$f(x) = \begin{cases} 1/(1+x)^2 & , \quad x > 0 \\ 0 & , \quad \text{ἀλλοῦ}. \end{cases}$$

Ἄν τό κόστος κατασκευῆς μίας συσκευῆς εἶναι 80 δραχμές καί ὁ κατασκευαστής τήν πωλεῖ πρός 400 δραχμές καί ἐγγυᾶται πλήρη ἐπιστροφή χρημάτων ὅταν $X \leq 1$ νά βρεθεῖ τό ἀναμενόμενο κέρδος τοῦ κατασκευαστῆ ἀνά συσκευή.

5.22. Ἕνα τέλειο νόμισμα ρίχνεται μέχρις ὅτου ἐμφανιστεῖ ἡ πρώτη κορώνα. Νά βρεθεῖ ὁ ἀναμενόμενος ἀριθμός ρίψεων τοῦ νομίσματος.

5.23. Μία τ.μ. X ἔχει $E(X) = 1$, $Var(X) = 1$, συντελεστή λοξότητας $\alpha_3 = 2$, συντελεστής κυρτώσεως $\alpha_4 = 9$ καί

$E(X^k)=k!$ γιά $k \geq 5$. Νά βρεθεῖ ἡ ροπογεννήτρια συνάρτηση τῆς X καί νά ἀναγνωριστεῖ ἡ κατανομή.

5.24. Τό μέγεθος τῶν ἡμερήσιων πωλήσεων ἑνός ἐμπορεύματος καί οἱ ἀντίστοιχες πιθανότητες ἔχουν ὡς ἀκολούθως γιά κάποιο *supermarket*:

Μέγεθος πωλήσεων (σέ ἑκατοντάδες μονάδες)	1	2	3	4
πιθανότητες	1/4	1/4	1/4	1/4

Νά βρεθεῖ ὁ ἀριθμός τῶν μονάδων τοῦ ἐμπορεύματος πού πρέπει νά ἀποθηκεύσει ὁ ἔμπορος σέ ἡμερήσια βάση γιά νά μεγιστοποιήσει τό ἀναμενόμενο κέρδος του ὅταν γιά κάθε μονάδα ἐμπορεύματος ἔχει κέρδος *20* δραχμές ἄν πωληθεῖ καί ζημιά *40* δραχμές ἄν δέν πωληθεῖ.

5.25. Ἔστω X τ.μ. μέ διωνυμική κατανομή μέ μέση τιμή *12* καί διακύμανση *8*. (*i*) Νά βρεθεῖ τό n καί (*ii*) νά ὑπολογιστεῖ ἡ $E(X^2+2X)$.

5.26. Ἕνας σπουδαστής παίρνει μέρος σέ μία ἐξέταση ἡ ὁποία ἀποτελεῖται ἀπό ἑκατό ἐρωτήσεις. Κάθε ἐρώτηση ἔχει τέσσερες ἀπαντήσεις ἀπό τίς ὁποῖες μία μόνο εἶναι σωστή. Ὁ σπουδαστής ἀπάντησε σωστά σέ *90* ἐρωτήσεις, ὅταν διαπίστωσε ὅτι τοῦ μένουν ἀκόμη *5* λεπτά. Γιά τό λόγο αὐτό ἀπαντᾶ στίς ὑπόλοιπες ἐρωτήσεις στήν τύχη. Νά βρεθεῖ ὁ ἀναμενόμενος ἀριθμός τῶν σωστῶν ἀπαντήσεων τοῦ σπουδαστή καί ἡ τυπική ἀπόκλιση τοῦ ἀριθμοῦ αὐτῆς. Ἄν κάθε ἐρώτηση βαθμολογεῖται μέ μία μονάδα καί γιά κάθε λανθασμένη ἀπάντηση ἀφαιρεῖται τό 1/3 τῆς μονάδας νά βρεθεῖ ὁ ἀναμενόμενος βαθμός τοῦ σπουδαστῆ καί ἡ διακύμανση του.

5.27. Γιά νά ἀπορροφηθεῖ ἕνα φάρμακο χρειάζονται τουλάχιστον 5 *min* ἀλλά ποτέ περισσότερο ἀπό 15 *min*. Ἄν X εἶναι ὁ χρόνος ἀπορροφήσης τοῦ φαρμάκου καί ἄν ὅλες οἱ χρονικές στιγμές μεταξύ 5 καί 15 εἶναι ἰσοπίθανες, νά ὑπολογιστοῦν ἡ μέση τιμή καί ἡ διακύμανση τοῦ X.

5.28. Μία θεραπεία ἀνοσοποιεῖ 80% τῶν κουνελιῶν ἐναντίον κάποιας ἀσθένειας. Ἕνα δεῖγμα 50 κουνελιῶν ἐξετάζεται. Νά βρεθεῖ ὁ ἀναμενόμενος ἀριθμός τῶν ἀνοσοποιημένων κουνελιῶν καί ἡ τυπική ἀπόκλιση.

5.29. Ἔστω X διακριτή τ.μ. μέ κατανομή $P(X=\alpha) = P(X=-\alpha) = 1/8$, $P(X=0) = 3/4$ καί $P(X=x) = 0$, $x \neq \alpha, 0$. Νά ὑπολογιστεῖ ἡ $P(|X| \geq 2\sigma)$, ὅπου σ ἡ τύπική ἀπόκλιση τῆς X καί κατόπιν νά συγκριθεῖ ἡ πιθανότητα αὐτή μέ τό ἀντίστοιχο φράγμα τῆς ἀνισότητας τοῦ *Chebyshev*.

5.30. Ἡ ἀπαριθμητή τυχαία μεταβλητή X ἔχει συνάρτηση πυκνότητας τήν

$$p(x) = \begin{cases} \dfrac{k^2-1}{k^2} & , x = 0 \\ \dfrac{1}{2k^2} & , |x| = k \\ 0 & , \text{ἀλλοῦ} \end{cases}$$

ὅπου $k \geq 1$. Νά ὑπολογιστοῦν ἡ μέση τιμή καί ἡ διακύμανση τῆς X. Νά βρεθεῖ ἡ $P(|X-\mu|<k\sigma)$ καί νά συγκριθεῖ μέ τό κατώτερο πέρας πού δίνει ἡ ἀνισότητα *Chebyshev*.

5.31. Ἔστω X τ.μ. μέ σ.π.π.

$$f(x) = \begin{cases} \dfrac{1}{2\sqrt{3}} & , -\sqrt{3} < x < \sqrt{3} \\ 0 & , \text{ἀλλοῦ} . \end{cases}$$

Νά βρεθεῖ ἡ $P(|X| \geq 3/2)$ καί νά συγκριθεῖ ἡ πιθανότητα αὐτή μέ τό ἄνω φράγμα τῆς ἀνισότητας τοῦ *Chebyshev*.

5.32. Ἔστω ὅτι ὁ ἀριθμός τῶν ἀεροπλάνων πού φθάνουν σέ κάποιο ἀεροδρόμιο σέ μία ὁποιαδήποτε εἰκοσάλεπτο περίοδο ἀκολουθεῖ τήν κατανομή τοῦ *Poisson* μέ μέση τιμή 100. Χρησιμοποιώντας τήν ἀνισότητα τοῦ *Chebyshev* νά βρεθεῖ ἕνα κάτω φράγμα στήν πιθανότητα ὁ ἀριθμός τῶν παραπάνω ἀεροπλάνων νά εἶναι μεταξύ 70 καί 1300.

5.33. Δίνεται X τ.μ. μέ σ.π.π.

$$f(x) = \begin{cases} 12x(1-x)^2 & , \quad 0 < x < 1 \\ 0 & , \quad \text{ἀλλοῦ}. \end{cases}$$

(*i*) Νά βρεθεῖ ἡ κορυφή τῆς κατανομῆς αὐτῆς.
(*ii*) Νά βρεθεῖ ἡ μέση τιμή καί ἡ διακύμανση τῆς X . (*iii*) Νά βρεθεῖ ἡ μέση τιμή καί ἡ διακύμανση

τῆς $Y = 1/X$.

5.34. Νά βρεθεῖ ἡ ροπογεννήτρια συνάρτηση τῆς κατανομῆς

$$f(x) = \begin{cases} \dfrac{x^3 e^{-x^2/2}}{96} & , \quad x > 0 \\ 0 & , \quad \text{ἀλλοῦ}. \end{cases}$$

Ἐπίσης νά βρεθοῦν ἡ μέση τιμή καί ἡ διακύμανση τῆς κατανομῆς αὐτῆς.

5.35. Ἐάν ἡ ροπογεννήτρια συνάρτηση τῆς τυχαίας μεταβλητῆς X εἶναι

$$M_X(t) = \left(\frac{3}{4} + \frac{1}{4}\, e^t\right)^{16}$$

νά ὑπολογιστεῖ ἡ πιθανότητα $P(\mu-2\sigma < X < \mu+2\sigma)$.

5.36. Νά βρεθεῖ ἡ κατανομή γιά τήν ὁποία ἡ ροπογεννήτρια συνάρτηση εἶναι $m(t) = (1+2e^t)^4/81$. Ὁμοίως ἄν $m(t) = exp(2t+t^2)$.

5.37. Ἄν ἡ τ.μ. X ἔχει ροπογεννήτρια συνάρτηση τήν $exp(e^t-1)$ νά βρεθεῖ ἡ $E(X)$.

5.38. Δίνονται οἱ ἀκόλουθες ροπογεννήτριες συναρτήσεις

(i) $m_X(t) = (0,5\, e^t+0,5)^5$,

(ii) $m_X(t) = (e^t+1)^5/32$,

(iii) $m_X(t) = 5(5-t)^{-1}$,

(iv) $m_X(t) = (1-t/5)^{-1}$,

(v) $m_X(t) = (1-2t)^{-4}$,

(vi) $m_X(t) = 16^{-1}(\frac{1}{2} - t)^{-4}$

(vii) $m_X(t) = e^{t+t^2}$

(viii) $m_X(t) = e^{t^2}$

(ix) $m_X(t) = exp(3e^t-1)$.

Γιά κάθε περίπτωση νά άναγνωριστεῖ ἡ άντίστοιχη κατανομή καί νά βρεθοῦν ἡ μέση τιμή καί ἡ διακύμανση αὐτῆς.

5.39. Ἔστω ὅτι ἡ τ.μ. X ἔχει κανονική κατανομή $N(\mu,\sigma^2)$. Νά δεχτεῖ ὅτι

$$E(|X-\mu|) = \sigma\sqrt{2/\pi} .$$

5.40. Ἄν μ καί m εἶναι ἡ μέση τιμή καί ἡ διάμεσος τῆς τ.μ. X άντίστοιχα νά δεχτεῖ ὅτι ἡ $E(|X-\alpha|)$ ἐλαχιστοποιεῖται ὅταν $\alpha = m$ ἐνῶ ἡ $E(X-\alpha)^2$ ἐλαχιστοποιεῖται ὅταν $\alpha = \mu$.

5.41. Ἔστω X τ.μ. μέ $P(X\leq 0) = 0$ καί $E(X) = \mu<\infty$. Νά Νά δεχτεῖ ὅτι $P(X\geq 2\mu) \leq 1/2$.

5.42. Ἄν ἡ τ.μ. X ἔχει $E(X) = 3$ καί $E(X)^2= 13$ νά βρεθεῖ ἕνα κάτω φράγμα τῆς πιθανότητας $P(-2<X<8)$.

5.43. Νά δεχθεῖ ὅτι ὁ 'Ορισμός 5.7 τοῦ ποσοστιαίου σημείου εἶναι ἰσοδύναμος μέ κάθε ἕναν ἀπό τούς ἀπόλουθους ὁρισμούς.

(i) Ἔστω X τ.μ. καί $0<p<1$. Κάθε σημεῖο x_p γιά τό ὁποῖο ἰσχύει

$$P(X \leq x_p) \geq p \quad \text{καί} \quad P(X \geq x_p) \geq 1-p$$

λέγεται p-ποσοστιαῖο σημεῖο τῆς X.

(ii) Ἔστω X τ.μ. καί $0<p<1$. Κάθε σημεῖο x_p γιά τό ὁποῖο ἰσχύει

$$F(x_p-0) = P(X<x_p) \leq p \leq P(X \leq x_p)$$

λέγεται p-ποσοστιαῖο σημεῖο τῆς X.

5.44. Νά δεχτεῖ ὅτι οἱ κορυφές τῆς γεωμετρικῆς καί τῆς ἐκθετικῆς κατανομῆς εἶναι τά σημεῖα $k = 1$ καί $k = 0$ ἀντίστοιχα.

5.45. Δίνεται ἡ κατανομή

$$f(x) = \begin{cases} 1/x^2 & , \quad x \geq 1 \\ 0 & , \quad \text{ἀλλοῦ} . \end{cases}$$

Νά δεχτεῖ ὅτι ἡ ῥοπογεννήτρια συνάρτηση $m(t)$ τῆς κατανομῆς αὐτῆς δέν ὑπάρχει ἐκτός ἄν $t = 0$.

5.46. Ἔστω X διακριτή τ.μ. μέ τιμές $x = 1,2,\ldots$ καί σ.π.π. $p(x) = 1/[x(x+1)]$, $x = 1,2,\ldots$. Νά δεχτεῖ ὅτι ἡ X δέν ἔχει καμμία ῥοπή καί ὅτι, ἀντίθετα, ἡ ῥοπογεννήτρια συνάρτηση τῆς X ὑπάρχει καί εἶναι

$$m(t) = \begin{cases} 1 & , \quad t = 0 \\ 1+(e^{-t}-1)\ln(1-e^{t}) & , \quad t < 0 \\ \text{δέν ὑπάρχει} & , \quad t > 0. \end{cases}$$

5.47. Ἔστω X διακριτή τ.μ. μέ τιμές $x = 1, 2, 2^2, 2^3, \ldots$
.. καί σ.π. $P(X=2^n) = e^{-1}/n!$, $n = 0,1,2,\ldots$
(α) Νά βρεθοῦν οἱ ροπές (περί τό μηδέν) τῆς X.
(β) Νά δεχτεῖ ὅτι ἡ ροπογεννήτρια συνάρτηση τῆς X ὑπάρχει γιά $t \leq 0$ ἀλλά δέν ἔχει παραγώγους στό $t = 0$. Ἔτσι ἡ ροπογεννήτρια συνάρτηση δέν μπορεῖ νά μᾶς δώσει τίς ροπές.

5.48 Ἔστω X διακριτή τ.μ. μέ τιμές $\pm 2^n$ καί

$P(X=2^n) = P(X=-2^n) = 1/(2en!)$, $n = 0,1,2,\ldots$

Νά βρεθοῦν οἱ ροπές (περί τό μηδέν) τῆς X καί νά δεχτεῖ ὅτι ἡ ροπογεννήτρια συνάρτηση τῆς X ὑπάρχει μόνο γιά $t = 0$.

5.49. Ἔστω ἡ σ.π.π.

$$f(x) = \frac{1}{48} e^{-|x|^{1/4}} [1-k \sin|x|^{1/4}] , \quad x \in R ,$$

$k = -1$ ἄν $x < 0$ καί $k = 1$ ἄν $x \geq 0$. Νά δεχθεῖ ὅτι ἡ $f(x)$ εἶναι μή συμμετρική ἀλλά $\alpha_3 = 0$.

5.50. Νά δεχτεῖ ὅτι ἡ κορυφή τῆς ὑπεργεωμετρικῆς κατανομῆς εἶναι τό σημεῖο

$$k = [\frac{(n+1)(Np+1)}{N+2}] .$$

5.51. Νά δεχτεῖ ὅτι ἡ κορυφή τῆς ἀρνητικῆς διωνυμικῆς κατανομῆς εἶναι τό σημεῖο

$$k = 1+ [\frac{n-1}{p}].$$

5.53. Νά δεχτεῖ ἡ "ἔλλειψη μνήμης" στή γεωμετρική κατανομή, δηλαδή

$$P(X>m+n|X>m) = P(X>n), \quad m,n = 1,2,\ldots$$

ΚΕΦΑΛΑΙΟ 6

πολυδιαστατες κατανομες. ανεξαρτησια-αλλαγη μεταβλητων

Στό Κεφάλαιο αὐτό παρουσιάζουμε μερικές χρήσιμες ἔννοιες ἀπό τίς πολυδιάστατες τυχαῖες μεταβλητές καί κατανομές. Κατόπιν παρουσιάζουμε τήν ἔννοια τῆς ἀνεξαρτησίας τυχαίων μεταβλητῶν. Καί στή συνέχεια ἀναπτύσσουμε τό θέμα τῆς ἀλλαγῆς μεταβλητῶν γιά μονοδιάστατες τυχαῖες μεταβλητές. Τά ἀποτελέσματα αὐτά, μερικά ἀπό τά ὁποῖα δίδονται χωρίς ἀποδείξεις, ὁλοκληρώνουν τήν Εἰσαγωγή στή Θεωρία Πιθανοτήτων καί Τυχαίων Μεταβλητῶν καί ἀνοίγουν τό δρόμο γιά τή Στατιστική.

6.1 Πολυδιάστατες τυχαῖες μεταβλητὲς καὶ κατανομὲς

Ἡ ἔννοια τῆς πολυδιάστατης τυχαίας μεταβλητῆς ἀποτελεῖ γενίκευση τῆς ἔννοιας τῆς μονοδιάστατης τυχαίας μεταβλητῆς πού παρουσιάστηκε στό Κεφάλαιο 3. Τό ἴδιο ἰσχύει καί γιά τήν ἔννοια τῆς πολυδιάστατης κατανομῆς.

Ἡ μονοδιάστατη τυχαία μεταβλητή ὁρίστηκε σάν μία μονοσήμαντη συνάρτηση μέ πεδίο ὁρισμοῦ τό δειγματικό χῶρο S καί πεδίο τιμῶν ἕνα ὑποσύνολο τῆς εὐθείας τῶν πραγματικῶν ἀριθμῶν. Ὅταν τό πεδίο τιμῶν τῆς συναρτήσεως εἶναι ἕνα ὑποσύνολο τοῦ Εὐκλείδιου χώρου R^n τῶν n διαστάσεων τότε ἡ παραπάνω συνάρτηση ὁρίζει μία n-διάστατη τυχαία μεταβλητή. Οἱ μεταβλητές αὐτές συμβολίζονται μέ $(X_1,X_2,\ldots,X_n)$ καί διακρίνονται σέ συνεχεῖς καί διακριτές. Οἱ τυχαῖες μεταβλητές $X_1,X_2,\ldots,X_n$ εἶναι οἱ συνιστῶσες τῆς πολυδιάστατης τυχαίας μεταβλητῆς καί λέγονται καί π ε ρ ι θ ω ρ ι α κ έ ς τυχαῖες μεταβλητές.

Ἡ κατανομή τῆς πολυδιάστατης τυχαίας μεταβλητῆς ἐκφράζει τόν τρόπο μέ τόν ὁποῖο ἡ μοναδιαία μάζα πιθανότητας κατανέμεται στό χῶρο R^n. Δίδεται εἴτε μέ τήν ἀπό κ ο ι ν ο ῦ σ υ ν ά ρ τ η σ η π ι θ α ν ό τ η τ α ς (σ.π.)

$$p_{X_1,\ldots,X_n}(x_1,\ldots,x_n)=P(X_1=x_1,\ldots,X_n=x_n) \qquad (6.1)$$

ὅταν ἡ τυχαία μεταβλητή εἶναι διακριτή ἤ τήν ἀπό κ ο ι ν ο ῦ σ υ ν ά ρ τ η σ η π υ κ ν ό τ η τ α ς π ι θ α ν ό τ η τ α ς (σ.π.π.)

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) \qquad (6.2)$$

ὅταν ἡ μεταβλητή εἶναι συνεχής εἴτε μέ τήν ἀπό κ ο ι ν ο ῦ ἀ θ ρ ο ι σ τ ι κ ή σ υ ν ά ρ τ η σ η κ α τ α ν ο μ ῆ ς (α.σ.κ.)

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)=P(X_1\le x_1,\ldots,X_n\le x_n). \qquad (6.3)$$

Παραδείγματα:

6.1: Ἔστω ὅτι ρίχνουμε δύο ζάρια. Τά 36 σημεῖα τοῦ δειγματικοῦ χώρου τοῦ πειράματος αὐτοῦ εἶναι

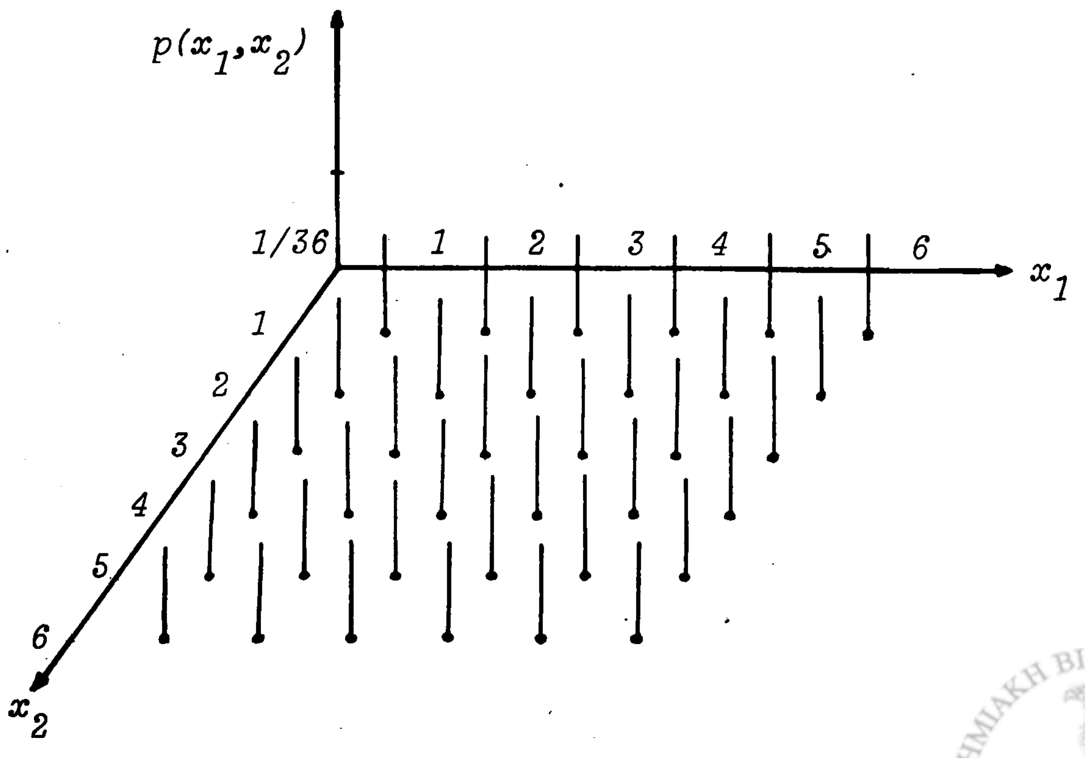
$$S = \{(\cdot,\cdot),(\cdot,\cdot\cdot),\ldots,(\cdot,\vdots\vdots),\ldots,(\vdots\vdots,\vdots\vdots)\}$$

Ἔστω X_1 ἡ τ.μ. πού παριστάνει τά ἀποτελέσματα τοῦ ἑνός ζαριοῦ καί X_2 ἡ τ.μ. μέ τά ἀποτελέσματα τοῦ ἄλλου ζαριοῦ. Τότε (X_1,X_2) εἶναι μία διδιάστατη τυχαία μεταβλητή μέ πεδίο ὁρισμοῦ τό S καί τιμές τά σημεῖα (i,j), $i,j=1,2,\ldots,6$.

Ἡ ἀπό κοινοῦ συνάρτηση πιθανότητας (σ.π.) τῶν (X_1,X_2) εἶναι

$$p_{X_1,X_2}(x_1,x_2)=\frac{1}{36}\,,\quad x_1,x_2=1,2,\ldots,6$$

καί ἔχει γραφική παράσταση



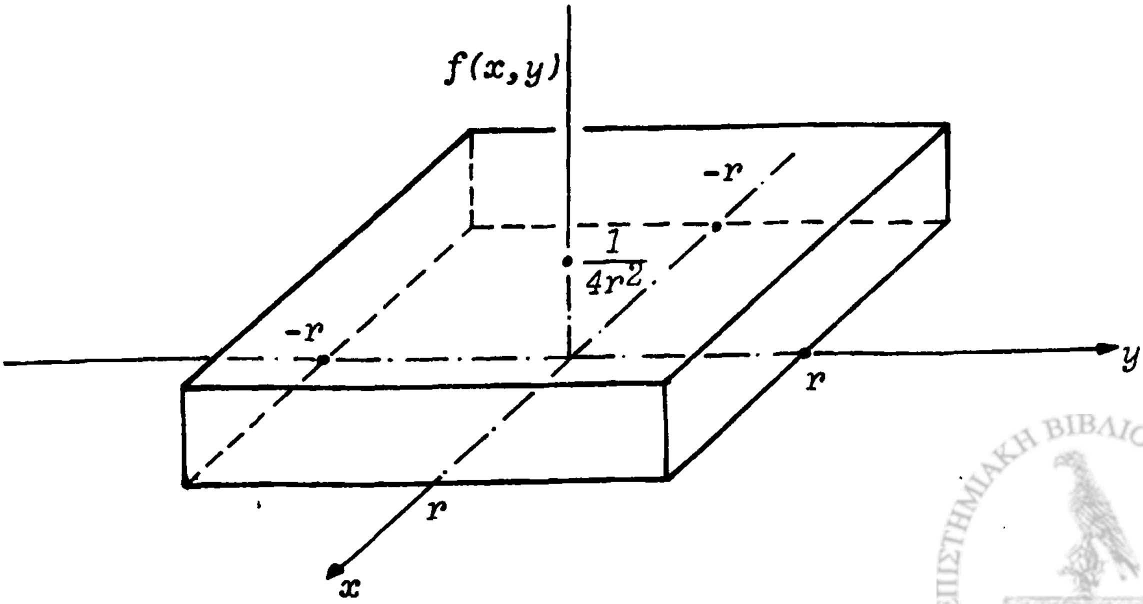
'Αθροίζοντες τήν $p_{X_1,X_2}(x_1,x_2)$ ὡς πρός x_2 καί x_1 ἀντιστοίχως· εὑρίσκομε $p_{X_1}(x_1)=1/6$, $x_1=1,\ldots,6$ καί $p_{X_2}(x_2)=1/6$, $x_2=1,\ldots,6$.

6.2: Ἄς θεωρήσουμε τή βελόνα τῆς μαγνητικῆς πυξίδας τοῦ Παραδείγματος 3.2. Ἔστω ὅτι τό μῆκος αὐτῆς εἶναι r καί (X,Y) οἱ συνιστῶσες τῆς βελόνας ὡς πρός ἕνα σύστημα ἀξόνων x,y μέ ἀρχή τό κέντρο τῆς πυξίδας. Καθώς περιστρέφεται ἡ βελόνα, τά X,Y παίρνουν τυχαῖες τιμές στό διάστημα $[-r,r]$. Προφανῶς (X,Y) εἶναι μία διδιάστατη τυχαία μεταβλητή μέ τιμές στό ὑποσύνολο $[-r,r]\times[-r,r]$ τοῦ ἐπιπέδου.

Ἡ ἀπό κοινοῦ συνάρτηση πυκνότητας πιθανότητας (σ.π.π.) τῶν (X,Y) εἶναι

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{4r^2} & , \ -r<x<r, \ -r<y<r \\ 0 & , \ \text{ἀλλοῦ} \end{cases}$$

καί ἔχει γραφική παράσταση



Πιό ρεαλιστικά παραδείγματα παίρνουμε ὅταν κάνουμε δειγματοληψία ἀπό πεπερασμένους ἤ ἄπειρους πληθυσμούς. Ἔτσι ὅταν παίρνουμε ἕνα δεῖγμα ἀπό μαθητές ἑνός γυμνασίου μπορεῖ νά ἐνδιαφερόμαστε γιά τήν ἐπίδοση X_1 στά μαθηματικά, τό βάρος X_2, τόν ἀριθμό X_3 τῶν ἡμερῶν ἀπουσίας, κ.ο.κ τοῦ κάθε μαθητοῦ. Οἱ τυχαῖες μεταβλητές X_1, X_2, X_3 ἀποτελοῦν ἀπό κοινοῦ μία τρισδιάστατη τυχαία μεταβλητή.

Ἡ μελέτη τῶν πολυδιάστατων κατανομῶν ἀκολουθεῖ τήν ἴδια κατεύθυνση ὅπως καί ἡ μελέτη τῶν μονοδιάστατων κατανομῶν. Ἔτσι μέ τή χρήση τοῦ ἀπειροστικοῦ λογισμοῦ πολλῶν μεταβλητῶν μποροῦν νά ὑπολογιστοῦν οἱ πιθανότητες διαφόρων πολυδιάστατων ἐνδεχομένων καί νά θεμελιωθοῦν οἱ ἰδιότητες τῶν ἀθροιστικῶν συναρτήσεων κατανομῶν.

Ὑπάρχουν ἐπίσης μοντέλα εἰδικῶν πολυδιάστατων κατανομῶν καί χαρακτηριστικά τῶν κατανομῶν ἀνάλογα μέ τή μαθηματική ἐλπίδα, τή μέση τιμή, τή διακύμανση, τίς ροπές, τίς ροπογεννήτριες, τίς κορυφές κ.λ.π.Τά θέματα αὐτά ξεφεύγουν ἀπό τούς στόχους τοῦ παρόντος βιβλίου.

6. 2 'Ανεξάρτητες τυχαῖες μεταβλητὲς

Οἱ ἀνεξάρτητες τυχαῖες μεταβλητές παίζουν πρωτεύοντα ρόλο στή Θεωρία Πιθανοτήτων καί τή Στατιστική. Ἀποτελοῦν τήν ἁπλούστερη ὑπόθεση στή πιθανοθεωρητική καί στατιστική μελέτη ἑνός φαινομένου.

Ἡ ἔννοια τῆς ἀνεξαρτησίας ἀναφέρεται σέ δύο ἤ περισσότερες μεταβλητές καί ἐξετάζει τή σχέση μεταξύ τῆς ἀπό κοινοῦ κατανομῆς τους καί τῶν ἐπί μέρους (περιθωριακῶν) κατανομῶν τῶν τυχαίων μεταβλητῶν. Ἐπειδή οἱ τυχαῖες μεταβλητές περιγράφουν διάφορα ἐνδεχόμενα τοῦ δειγματικοῦ χώρου S, ἡ ἔννοια τῆς ἀνεξαρτησίας τυχαίων μεταβλητῶν εἶναι φυσικό νά ἀποτελεῖ γενίκευση τῆς ἔννοιας τῶν στοχαστικῶς ἀνεξαρτήτων ἐνδεχομένων πού μελετήσαμε στό Δεύτερο Κεφάλαιο. Ἔτσι ἔχουμε τόν ἀκόλουθο.

Ὁρισμός 6.1: Οἱ τυχαῖες μεταβλητές $X_1, X_2, \ldots, X_n$ λέγονται στοχαστικά ἀνεξάρτητες ἤ ἁπλά ἀνεξάρτητες ἄν γιά κάθε συλλογή ἀπό (Borel) ὑποσύνολα $B_1, B_2, B_3, \ldots, B_n$ τοῦ R^1 ἰσχύει ἡ σχέση

$$P(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = P(X_1 \in B_1) P(X_2 \in B_2) \ldots P(X_n \in B_n). \tag{6.4}$$

Τυχαῖες μεταβλητές πού δέν εἶναι ἀνεξάρτητες λέγονται καί ἐξηρτημένες.

Ὁ παραπάνω ὁρισμός συνεπάγεται ὅτι ἄν n τυχαῖες μεταβλητές εἶναι ἀνεξάρτητες τότε εἶναι καί ἀνεξάρτητες ἀνά δύο. Τό ἀντίστροφο δέν εἶναι σωστό. Ὑπενθυμίζουμε ὅτι τό ἴδιο ἰσχύει καί γιά τά ἐνδεχόμενα: n ἀνά δύο ἀνεξάρτητα ἐνδεχόμενα δέν εἶναι καί ὁλικά ἀνεξάρτητα.

Διαισθητικά δύο ἤ περισσότερες τυχαῖες μεταβλητές εἶναι ἀνεξάρτητες ἄν ἡ μία δέν ἐπηρεάζει τίς ἄλλες. Αὐτό συμβαίνει ὅταν κάθε μία ἀπό τίς τυχαῖες μεταβλητές προκύπτει ἀπό ἕνα τυχαῖο πείραμα καί τά πειράματα εἶναι ἀνεξάρτητα. Ἔτσι οἱ ἀριθμοί τῶν τηλεφωνικῶν κλήσεων πού φθάνουν σ'ἕνα τηλεφωνικό κέντρο ἀπό τίς 8 ὥς τίς 9 τό πρωί σέ δύο διαδοχικές ήμέρες μποροῦν νά θεωρηθοῦν ὡς ἀνεξάρτητες τυχαῖες μεταβλητές.

Ὁ ἔλεγχος τῆς ἀνεξαρτησίας τυχαίων μεταβλητῶν μέ βάση τόν παραπάνω ὁρισμό δέν εἶναι εὔκολος. Ὑπάρχουν ὁρισμένα κριτήρια δύο ἀπό τά ὁποῖα κάνουν τό πρόβλημα εὔκολο.

Θεώρημα 6.1: (Κριτήρια ἀνεξαρτησίας τυχαίων μεταβλητῶν):

Οἱ τυχαῖες μεταβλητές $X_1, X_2, \ldots, X_n$ εἶναι ἀνεξάρτητες ἄν καί μόνον ἄν γιά κάθε $x_1, \ldots, x_n$ ἰσχύει μία ἀπό τίς σχέσεις

1) $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)=F_{X_1}(x_1)\ldots F_{X_n}(x_n)$ (6.5)

2) $p_{X_1,\ldots,X_n}(x_1,\ldots,x_n)=p_{X_1}(x_1)\ldots p_{X_n}(x_n)$

ὅταν οἱ τ.μ. $X_1, X_2, \ldots, X_n$ εἶναι διακριτές

ἤ (6.6)

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

ὅταν οἱ τ.μ. $X_1, X_2, \ldots, X_n$ εἶναι συνεχεῖς,

ὅπου F_X, P_X καί f_X συμβολίζουν τήν ἀθροιστική συνάρτηση κατανομῆς (α.σ.κ.), τή συνάρτηση πιθανότητας (σ.π.) καί τή συνάρτηση πυκνότητας πιθανότητας (σ.π.π.) ἀντίστοιχα τῆς τυχαίας μεταβλητῆς X.

Ἡ Μαθηματική Στατιστική καί ὁ Ἀπειροστικός Λογισμός μᾶς ἐπιτρέπουν τόν προσδιορισμό τῶν περιθωριακῶν κατανομῶν μιάς πολυδιάστατης κατανομῆς. Τό ἀντίστροφο πρόβλημα δέν εἶναι πάντοτε δυνατό ἐκτός ἐάν οἱ περιθωριακές τυχαῖες μεταβλητές εἶναι ἀνεξάρτητες.

Οἱ συνέπειες τῆς ἀνεξαρτησίας τυχαίων μεταβλητῶν εἶναι πολλές καί σημαντικές. Ἀναφέρουμε χωρίς ἀπόδειξη δύο θεωρήματα, τό δεύτερο ἀπό τά ὁποῖα ἔχει ἰδιαίτερη σημασία.

Θεώρημα 6.2:

Ἔστω $X_1,\ldots,X_n$ n ἀνεξάρτητες τυχαῖες μεταβλητές καί $y_i=h_i(x_i)$, $i=1,2,\ldots,n$, n συναρτήσεις (μετασχηματισμοί) τῶν x_i. Τότε οἱ τυχαῖες μεταβλητές $Y_1=h_1(X_1),\ldots,$ $Y_n=h_n(X_n)$ εἶναι ἀνεξάρτητες.

Θεώρημα 6.3:

1. Ἔστω X,Y δύο ἀνεξάρτητες τυχαῖες μεταβλητές μέ πεπερασμένες μέσες τιμές. Τότε

$$EXY = (EX)(EY) \;, \qquad (6.7)$$

ὅπου EXY εἶναι ἡ ἀναμενόμενη τιμή τῆς τυχαίας μεταβλητῆς XY.

2. Ἔστω $X_1,\ldots,X_n$ n ἀνεξάρτητες τυχαῖες μετατές καί $y_i=h_i(x_i)$, $i=1,2,\ldots,n$, n συναρτήσεις τῶν x_i. Ἔστω ἐπί πλέον ὅτι $Eh_i(X_i)<\infty$, $i=1,2,\ldots,n$. Τότε

$$E[h_1(X_1)\ldots h_n(X_n)]=[Eh_1(X_1)]\ldots[Eh_n(X_n)]. \qquad (6.8)$$

6. 3 Ἀλλαγὴ μεταβλητῶν: κατανομὴ συναρτήσεως τυχαίας μεταβλητῆς

Στό ἐδάφιο αὐτό μελετᾶμε τόν τρόπο μέ τόν ὁποῖο βρίσκουμε τήν κατανομή μίας συναρτήσεως μίας τυχαίας μεταβλητῆς. Ἔστω ὅτι δίδεται μία τυχαία μεταβλητή X καί ἡ κατανομή αὐτῆς. Ἔστω ἐπίσης $Y=h(X)$ μία δοθεῖσα συνάρτηση τοῦ X. Ἐνδιαφερόμαστε νά εὕρουμε τήν κατανομή τῆς τυχαίας μεταβλητῆς Y. Τό πρόβλημα αὐτό μαζί μέ τή γενίκευση του γιά πολυδιάστατες τυχαῖες μεταβλητές εἶναι ἀπό τά θεμελιώδη προβλήματα τῆς Μαθηματικῆς Στατιστικῆς δεδομένου ὅτι ἡ Στατιστική Συμπερασματολογία στηρίζεται στή γνώση τῆς κατανομῆς μετασχηματιζομένων τυχαίων μεταβλητῶν. Εἶναι ἐπίσης γνωστό καί σάν "ἀ λ λ α γ ή μ ε τ α β λ η τ ῶ ν".

Ἀπό θεωρητικῆς πλευρᾶς ἡ λύση τοῦ προβλήματος εἶναι εὔκολη διότι γιά σταθερό y ἡ ἀθροιστική συνάρτηση κατανομῆς τῆς τυχαίας μεταβλητῆς Y εἶναι

$$F_Y(y)=P(Y\leq y)=P[h(X)\leq y]$$

καί ἡ τελευταία πιθανότητα εἶναι πιθανότητα ἐνδεχομένου πού περιγράφεται μέ τήν τυχαία μεταβλητή X. Θεωρητικά αὐτή ἡ πιθανότητα εὑρίσκεται ὁλοκληρώνοντας ἤ ἀθροίζοντας τήν πυκνότητα τοῦ X πάνω στή περιοχή πού καθορίζεται ἀπό τό ἐνδεχόμενο $\{h(x)\leq y\}$. Ἡ δυσκολία τοῦ προβλήματος ἔγκειται στό ὅτι κατά κανόνα ἡ πιθανότητα αὐτή δέν ὑπολογίζεται εὔκολα.

Ἀπό τή θεωρία πού ἔχει παρουσιαστεῖ μέχρι τώρα βλέπουμε πώς τρεῖς μέθοδοι προσφέρονται γιά τή λύση τοῦ προβλήματος: (1) ἡ μέθοδος τῶν ἀθροιστικῶν συναρτήσεων κατανομῆς πού ἀναφέρθηκε πιό πάνω καί ἐπεξηγεῖται στό (Α), (2) ἡ μέθοδος τῆς συναρτήσεως πυκνότητας πιθανότητας ἤ τοῦ μετασχηματισμοῦ πού ἀναπτύσσεται στό (Β) καί (3) ἡ μέθοδος τῆς ροπογέννητριας συναρτήσεως πού ἀναπτύσσεται στή (Γ). Ἡ ὅλη μεθοδολογία ἀποτελεῖ μέρος τῆς Θεωρίας Κατανομῶν.

(Α) Μέθοδος τῆς ἀθροιστικῆς συναρτήσεως κατανομῆς. Ἡ μέθοδος αὐτή εἶναι δυναμική καί εὔχρηστη. Ἀποτελεῖ ἐφαρμογή τῆς παραπάνω ἰδέας τῆς ἀθροιστικῆς συναρτήσεως κατανομῆς.

Ἡ ἀθροιστική συνάρτηση κατανομῆς τῆς τυχαίας μεταβλητῆς Y θά βρεθεῖ ἀπό τόν τύπο:

$$F_Y(y)=P(Y\leq y)=P[h(X)\leq y]=\begin{cases}\displaystyle\int_{\{h(x)\leq y\}} f(x)dx & \text{ἐάν ἡ } X \text{ εἶναι συνεχής}\\ \displaystyle\sum_{\{x:h(x)\leq y\}} p_X(x) & \text{ἐάν ἡ } X \text{ εἶναι διακριτή.}\end{cases} \tag{6.9}$$

Τά παρακάτω παραδείγματα δίνουν τόν τρόπο ἐφαρμογῆς τῆς μεθόδου.

<u>Παραδείγματα:</u>

6.3: Ἔστω $X\sim N(0,1)$ καί $Y=X^2$. Τότε ἡ μέθοδος τῆς ἀθροιστικῆς συναρτήσεως κατανομῆς γιά τήν Y δίνει

$$F_Y(y)=P(Y\leq y)=P(X^2\leq y)=P(-\sqrt{y}\leq X\leq\sqrt{y})=2P(0\leq X\leq\sqrt{y})$$

$$=2\int_0^{\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx=\frac{2}{\sqrt{2\pi}}\int_0^y\frac{1}{2\sqrt{t}}e^{-t/2}dt=$$

$$=\int_0^y\frac{t^{1-1/2}e^{-t/2}}{\Gamma(1/2)\;2^{1/2}}dt\;,\qquad y>0\;.$$

Ἔτσι $Y\sim G(\alpha=1/2,\;\beta=2)$.

6.4: Ἔστω X συνεχής τυχαία μεταβλητή μέ ἀθροιστική συνάρτηση κατανομῆς $F_X(x)$ καί συνάρτηση πυκνότητας πιθανότητας $f_X(x)$. Ἔστω ἐπίσης $Y=aX+b$, $a<0$. Τότε ἡ α.σ.κ. τῆς Y εἶναι

$$F_Y(y)=P(Y\leq y)=P(aX+b\leq y)=P(X\geq(y-b)/a)=$$

$$=1-P(X\leq(y-b)/a)=1-F_X((y-b)/a).$$

Παραγωγίζοντες ὡς πρός y ἔχουμε

$$F_Y(y)=-\frac{1}{a}f_X((y-b)/a)$$

ἡ ὁποία εἶναι ἡ συνάρτηση πυκνότητας πιθανότητας τῆς Y.

(Β) Μ έ θ ο δ ο ς τ ο ῦ μ ε τ α σ χ η μ α τ ι σ μ ο ῦ. Ἡ μέθοδος αὐτή μᾶς ἐπιτρέπει νά εὑρίσκουμε τή συνάρτηση πιθανότητας ἤ τή συνάρτηση πυκνότητας πιθανότητας μετασχηματιζομένων τυχαίων μεταβλητῶν. Ἡ οὐσία τῆς μεθόδου ἀφορᾶ συνεχεῖς τυχαῖες μεταβλητές. Εἶναι ἀρκετά γενική καίτοι ἡ ἐφαρμογή της ἐνδέχεται νά παρουσιάζει τεχνικές δυσκολίες.

Στήν ἀρχή μελετᾶμε τήν περίπτωση μίας διακριτῆς τυχαίας μεταβλητῆς X μετασχηματιζομένης σέ μία ἄλλη μονοδιάστατη τυχαία μεταβλητή Y βάσει τῆς συναρτήσεως $y=h(x)$. Ἡ νέα τυχαία μεταβλητή εἶναι $Y=h(X)$. Ἔστω ὅτι τά σημεῖα πιθανότητας (μάζης) τῆς X εἶναι $x_1,x_2,\ldots$ Τότε ἡ Y εἶναι ἐπίσης διακριτή τυχαία μεταβλητή μέ τιμές τά σημεῖα $y_1,y_2,\ldots$ ὅπου $y_j=h(x_i)$ ἐνδεχομένως γιά περισσότερα τοῦ ἑνός x_i. Ἡ συνάρτηση πιθανότητας τῆς Y εὑρίσκεται βάσει τῶν νόμων τῶν πιθανοτήτων

$$P_Y(y)=P(Y=y_j)=P[h(X)=y_j]=\sum_{\{x_i:h(x_i)=y_i\}} P(X=x_i)$$

ἤ

$$P_Y(y)=\sum_{\{x:h(x)=y\}} p(x) \qquad (6.10)$$

<u>Παραδείγματα</u>:

6.5: Ἔστω X διακριτή τυχαία μεταβλητή μέ τιμές -3, -2, -1, 0, 1, 2, 3 καί ἡ κάθε μία μέ πιθανότητα $1/7$. Ἔστω ἐπίσης $Y=X^2$. Τότε ἡ Y παίρνει τιμές 0, 1, 4, 9 μέ πιθανότητες $P(Y=0)=P(X=0)=1/7$, $P(Y=1)=P(X=-1)+$ $+P(X=1)=2/7$, $P(Y=4)=2/7$ καί $P(Y=9)=2/7$.

6.6: Ἔστω $X\sim P(\lambda)$ καί $Y=X^2+X-1$. Οἱ τιμές τῆς Y εἶναι -1, 1, 5, 11, $19,\ldots$ Ἡ $y=x^2+x-1$ ἤ $x^2+x-1-y=0$ μᾶς δίδει $x=-1+\sqrt{5+4y}$. Ἡ ἄλλη ρίζα εἶναι ἀρνητική καί ἀπορρίπτεται γιατί $x\geq 0$.

$$p_Y(y)=p_X(x)=\frac{e^{-\lambda}\lambda^x}{x!}=\frac{e^{-\lambda}\lambda^{-1+\sqrt{5+4y}}}{(-1+\sqrt{5+4y})!}\;.$$

6.3 Ἀλλαγή μεταβλητῶν: κατανομή συναρτήσεως τυχαίας μεταβλητῆς

Ἔστω τώρα ὅτι X εἶναι συνεχής τυχαία μεταβλητή. Ἡ ἐφαρμογή τῆς μεθόδου τῆς ἀθροιστικῆς συναρτήσεως κατανομῆς καί ὑπο ὡρισμένες συνθῆκες μᾶς ὁδηγεῖ στό ἑπόμενο θεώρημα.

Θ ε ώ ρ η μ α 6.4:

Ἔστω X μία συνεχής τυχαία μεταβλητή μέ συνάρτηση πυκνότητας πιθανότητας $f_X(x)$. Θέτομε $S=\{x:f_X(x)>0\}$. Ὑποθέτομε ὅτι

(i) $y=h(x)$ εἶναι ἕνας ἀμφιμονοσήμαντος (ἕνα-πρός-ἕνα) μετασχηματισμός (συνάρτηση) πού ἀπεικονίζει τό σύνολο S σ'ἕνα σύνολο T τῶν y

(ii) ἡ ἀντίστροφη συνάρτηση $x=h^{-1}(y)$ εἶναι παραγωγίσιμη καί ἡ παράγωγος της συνεχής καί μή μηδενική γιά κάθε $y\in T$.

Τότε ἡ τυχαία μεταβλητή $Y=h(X)$ εἶναι συνεχής μέ συνάρτηση πυκνότητας πιθανότητας

$$f_Y(y)=\begin{cases} f_X[h^{-1}(y)]\left|\dfrac{dh^{-1}(y)}{dy}\right| , & y\in T \\ 0 , & \text{ἀλλοῦ} , \end{cases} \qquad (6.11)$$

ὅπου $|.|$, σημαίνει τήν ἀπόλυτο τιμή τῆς συναρτήσεως.

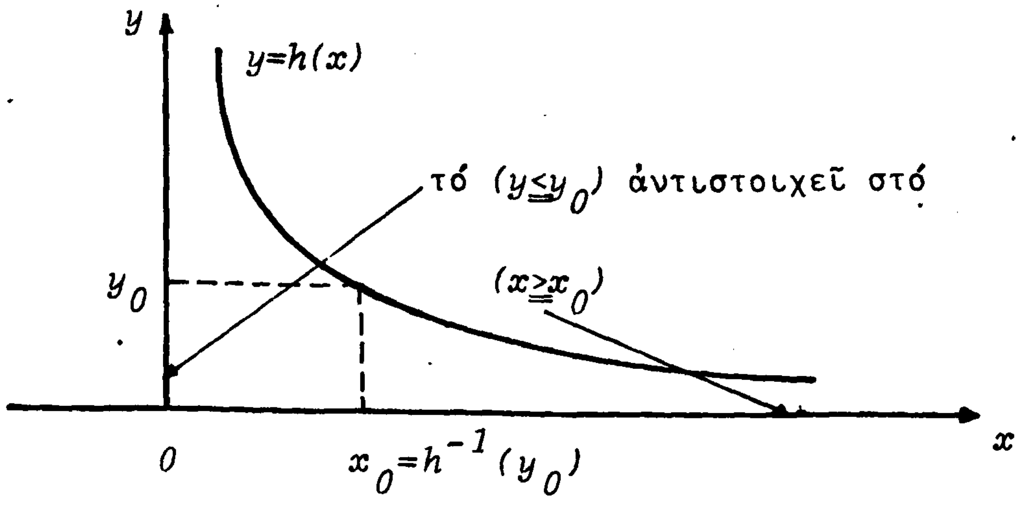
Τό παραπάνω θεώρημα δίδει συνθῆκες τίς ὁποῖες πρέπει νά πληροῖ ἡ $h(x)$ γιά νά ἐξασφαλίζεται ἡ συνέχεια τῆς Y ὅταν ἡ X εἶναι συνεχής. Ἀποτελεῖ βασικό θεώρημα τοῦ Ἀπειροστικοῦ Λογισμοῦ ἐπί τῆς ἀλλαγῆς μεταβλητῶν στό ὁρισμένο ὁλοκλήρωμα δεδομένου ὅτι ἐάν $B\subset T$, καί A ἡ εἰκόνα τοῦ B ἀπό τή συνάρτηση h^{-1} ἔχομε

$$P(Y\in B)=P[h(X)\in B]=P(X\in A)= \int_A f_X(x)dx$$

Δίνουμε μία σκιαγραφία τῆς ἀποδείξεως του.

Ἀπόδειξη: Ἄς θεωρήσουμε τήν περίπτωση πού τό S εἶναι ἕνα διάστημα. Ἔστω ὅτι ἡ $h(x)$ εἶναι γνησίως αὔξουσα συνάρτηση πάνω στό S δηλαδή $h'(x)>0$, πρᾶγμα τό ὁποῖο ἰσχύει τότε καί μόνο τότε ἐάν $dh^{-1}(y)/dy>0$ πάνω στό T. Γιά $y\in T$, $F_Y(y) = P(Y\leq y) = P[h(X)\leq y]$ $= P\ [X\leq h^{-1}(y)] = F_X[h^{-1}(y)]$ καί κατά συνέπεια $f_Y(y) = f_X(x)dx/dy = f_X\ h^{-1}(y)/dy$.

Ἡ περίπτωση ἡ $h(x)$ νά εἶναι γνησίως φθίνουσα δίδεται ὡς ἄσκηση στόν ἀναγνώστη. Τό παρακάτω σχῆμα ἐπεξηγεῖ τή σχέση μεταξύ τῶν ἐνδεχομένων τῶν ἀθροιστικῶν συναρτήσεων κατανομῶν στή περίπτωση αὐτή. ▼



Χάριν εὐκολίας ἡ (6.11) συνήθως γράφεται ὡς

$$f_Y(y)=f_X(x)\ \left|\frac{dx}{dy}\right|\ ,\quad dy/dx\neq 0 \qquad (6.12)$$

<u>Παραδείγματα:</u>

6.7: Ἔστω $X\sim B(a,b)$ καί $Y=-\ell nX$. Ζητεῖται ἡ κατανομή τῆς Y. Ἔχομε $S=\{x:f_X(x)>0\}=(0,1)$. Ἡ συνάρτηση $y=-\ell nx$ εἶναι ἕνας ἀμφιμονοσήμαντος μετασχηματισμός τοῦ S ἐπί τοῦ $T=(0,+\infty)$ καί

$$x=h^{-1}(y)=e^{-y},\qquad dh^{-1}(y)/dy=-e^{-y}.$$

Ἄρα ἡ (6.12) μᾶς δίδει

$$f_Y(y)=\frac{1}{B(a,\beta)}\,x^{a-1}(1-x)^{\beta-1}\left|\frac{dx}{dy}\right|=\frac{1}{B(a,\beta)}\,(e^{-y})^{a-1}.$$

$$\cdot(1-e^{-y})^{\beta-1}e^{-y}=\frac{1}{B(a,\beta)}\,e^{-ay}(1-e^{-y})^{\beta-1}\ ,\quad y>0\ .$$

6.8: Ἔστω $X\sim Ek\theta(\lambda)$ καί $Y=\sqrt{X}$. Ζητεῖται ἡ κατανομή τῆς Y. Ἔχομε $S=(0,\infty)$, $T=(0,\infty)$ καί ἡ συνάρτηση $y=\sqrt{x}$ ἀμφιμονοσήμαντος ἀπό S ἐπί T. $x=y^2$ καί $dx/dy=2y$. Ἄρα ἡ (6.12) δίδει

$$f_Y(y)=\lambda e^{\lambda x}2y=2\lambda ye^{-\lambda y^2},\quad y>0\qquad (\text{κατανομή Weibull}).$$

Τό ἑπόμενο θεώρημα καλύπτει τή περίπτωση πού ὁ μετασχηματισμός $y=h(x)$ δέν εἶναι ἐξ ὁλοκλήρου ἀμφιμονοσήμαντος ἀλλά μόνο κομματιαστά. Οἱ λεπτομέρειες δίδονται στήν ἐκφώνηση.

Θ ε ώ ρ η μ α 6.5:

Ἔστω X μία συνεχής τυχαία μεταβλητή μέ συνάρτηση πυκνότητας πιθανότητας $f_X(x)$ καί $y=h(x)$ ἕνας μετασχηματισμός. Θέτομε $S=\{x:f_X(x)>0\}$ καί $T=\{y:y=h(x)\}$, $x\in S\}$. Ὑποθέτομε ὅτι

(i) ὑπάρχει μία διαμέριση $S_1,\ldots,S_k$ τοῦ S (τά σύνολα S_i εἶναι ξένα μεταξύ τους) τέτοια ὥστε γιά κάθε S_i ἡ συνάρτηση $h(x)$ εἶναι μία ἀμφιμονοσήμαντη ἀπεικόνιση τοῦ S_i στό T. Ἔστω $y=h_i(x)$ ὁ περιορισμός τῆς συναρτήσεως h στό S_i καί $T_i=\{y:y=h_i(x),\ x\in S_i\}$, $i=1,\ldots,k$. $T=\bigcup_{i=1}^{k}T_i$ ἀλλά τά T_i δέν εἶναι κατ'ἀνάγκην ξένα μεταξύ τους.

(ii) γιά κάθε i ἡ ἀντίστροφη συνάρτηση $x=h_i^{-1}(y)$ εἶναι παραγωγίσιμη καί ἡ παράγωγος της συνεχής καί μή μηδενική γιά κάθε $y\in T_i$.

Τότε ἡ τυχαία μεταβλητή $Y=h(X)$ εἶναι συνεχής μέ συνάρτηση πυκνότητας πιθανότητας

$$f_Y(y)=\begin{cases}\Sigma f_X\ h_i^{-1}(y)\ \left|\dfrac{dh_i^{-1}(y)}{dy}\right| & ,\ y\in T\\[2ex] 0 & ,\ \text{ἀλλοῦ}\end{cases}\qquad(6.13)$$

ὅ π ο υ τ ό ἄ θ ρ ο ι σ μ α ε ἶ ν α ι γ ι ά ἐ κ ε ῖ ν α τ ά i, γ ι ά τ ά ὁ π ο ῖ α $h_i(x)=y$, $i=1,\ldots,k$.

Ἡ ἀπόδειξη τοῦ θεωρήματος εἶναι ἀνάλογος μέ ἐκείνη τοῦ θεωρήματος 6.4 καί παραλείπεται.

Ἡ οὐσία τοῦ θεωρήματος εἶναι ὅτι γιά νά εὕρωμε τήν κατανομή τῆς $Y=h(X)$ ἐργαζόμεθα βάσει τοῦ θεωρήματος 6.4. σέ κάθε ζεῦγος (S_i,T_i) καί ἐάν ἕνα y ἀνήκει σέ δύο ἤ περισσότερα T_i ἀθροίζομε τίς ἐπί μέρους συναρτήσεις πυκνοτήτων πιθανοτήτων.

Μία ἁπλή περίπτωση πού οἱ ὑποθέσεις τοῦ θεωρήματος 6.5 ἰσχύουν εἶναι ἡ ἀκόλουθη: $S=[a,b]$, ἡ $h(x)$ ἔχει πρώτη παράγωγο $h'(x)$ συνεχῆ γιά κάθε $x\in[a,b]$ καί $h'(x)$ μηδενίζεται στά σημεῖα $a_1<a_2<a_3<\ldots<a_{k-1}$.

Παραδείγματα:

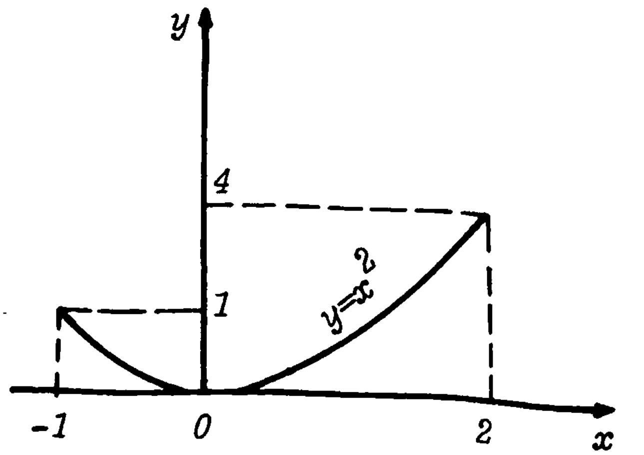
6.9: Ἔστω X συνεχής τυχαία μεταβλητή μέ συνάρτηση πυκνότητας πιθανότητας $f_X(x) = e^{-|x|}/2$, $-\infty<x<\infty$ καί $Y=X^2$. Προφανῶς ὁ μετασχηματισμός $y=x^2$, $-\infty<x<\infty$ δέν εἶναι ἀμφιμονοσήμαντος. Ἔστω $S_1=\{x:x<0\}$ καί $S_2=\{x:x\geq 0\}$. Προφανῶς ἡ $y=x^2$ εἶναι ἀμφιμονοσήμαντη γιά κάθε S_i, $i=1,2$ καί οἱ ὑποθέσεις τοῦ θεωρήματος 6.5 ἱκανοποιοῦνται. Ἄρα

$$f_Y(y)=f_X(-\sqrt{y})\,\frac{1}{2\sqrt{y}} + f_X(\sqrt{y})\,\frac{1}{2\sqrt{y}} = \frac{1}{2\sqrt{y}}\,e^{-\sqrt{y}}, \quad y>0 .$$

6.10: Ἔστω X συνεχής τυχαία μεταβλητή μέ συνάρτηση πυκνότητας πιθανότητας

$$f_X(x) = \begin{cases} \frac{2}{9}\,(x+1) & , \quad -1<x<2 \\ 0 & , \quad \text{ἀλλοῦ} \end{cases}$$

καί $Y=X^2$ (Βλέπε Σχῆμα)



Ἔστω $S_1=(-1,0)$ καί $S_2=[0,2)$. Ἡ $y=x^2$ εἶναι ἀμφιμονοσήμαντη γιά τά S_1 καί S_2. $T_1=(0,1)$ καί $T_2=[0,4)$. $h_1^{-1}(x)=-\sqrt{y}$, $h_2^{-1}=\sqrt{y}$. Ἄν $0<y<1$ τότε $y \in T_1, T_2$ καί θά πρέπει νά προσθέσουμε τίς ἀντίστοιχες σ.π.π., ἐνῶ ἄν $y \in [1,4)$ τότε $y \in T_2$ καί δέν ἔχομε νά προσθέσουμε. Συνεπῶς

$$f_Y(y)=\begin{cases} \frac{2}{9}(-\sqrt{y}+1)\frac{1}{2\sqrt{y}}+\frac{2}{9}(\sqrt{y}+1)\frac{1}{2\sqrt{y}} \\ \\ \frac{2}{9}(\sqrt{y}+1)\ \frac{1}{2\sqrt{y}} \end{cases} = \begin{cases} \frac{2}{9\sqrt{y}} & , \quad 0<y<1 \\ \\ (\sqrt{y}+1)\frac{1}{9\sqrt{y}} & , \quad 1\leq y<4. \end{cases}$$

Τό ἑπόμενο θεώρημα εἶναι ἀξιοσημείωτο διότι μᾶς δίδει ἕνα τρόπο νά παραγάγουμε τίς τιμές διαφόρων συνεχῶν τυχαίων μεταβλητῶν.

Θεώρημα 6.6:

Ἄν ἡ α.σ.κ, $F_X(x)$, μίας συνεχοῦς τυχαίας μεταβλητῆς εἶναι μονοτόνως αὔξουσα, τότε ἡ τυχαία μεταβλητή $U=F_X(X)$ ἔχει ὁμοιόμορφο κατανομή στό διάστημα $(0,1)$. Ἀντίστροφα, ἐάν $V\sim U(0,1)$ τότε $X=F_X^{-1}(V)$ ἔχει ἀθροι-

<u>στική συνάρτηση κατανομῆς τήν</u> $F_X(x)$.

'Απόδειξη: 'Επειδή $0 \leq F_X(x) \leq 1$ ἕπεται ὅτι $0 \leq U \leq 1$. 'Η α.σ.κ τῆς U εἶναι

$$F_V(u) = P[F_X(X) \leq u] = \begin{cases} 0 & , u<0 \\ P[X \leq F^{-1}(u)] = F_X(F_X^{-1}(u)) = u, & 0 \leq u \leq 1 \\ 1 & , u>1 \end{cases}$$

Παραγωγίζοντες ὡς πρός u ἔχουμε $f_V(u) = \begin{cases} 1, & 0 \leq u \leq 1 \\ 0, & \text{ἀλλοῦ} \end{cases}$.

'Αντίστροφα : $P(X \leq x) = P[F_X^{-1}(V) \leq x] = P[V \leq F_X(x)] = F_X(x)$. ▼

"Εστω u μία τιμή τῆς τυχαίας μεταβλητῆς $U \sim U(0,1)$ καί $F_X(x)$ ἡ ἀθροιστική συνάρτηση κατανομῆς μίας συνεχοῦς τυχαίας μεταβλητῆς X. 'Η $F^{-1}(u)$ ἀποτελεῖ μία τιμή τῆς X .

(Γ) Μέθοδος τῆς ροπογεννήτριας συναρτήσεως. 'Η τελευταία αὐτή μέθοδος εὑρέσεως τῶν κατανομῶν συναρτήσεων τυχαίων μεταβλητῶν εἶναι ἰδιαίτερα χρήσιμη στή πολυδιάστατη περίπτωση. Στηρίζεται στήν ἰδιότητα τοῦ μονοσήμαντου τῆς ροπογεννητρίας.

'Η ἰδέα τῆς μεθόδου εἶναι ἡ ἑξῆς: "Εστω X τυχαία μεταβλητή καί $Y=h(X)$ ὁ μετασχηματισμός αὐτῆς. 'Η ροπογεννήτρια συνάρτηση τῆς Y εἶναι

$$m_Y(t) = Ee^{tY} = Ee^{th(X)} = \begin{cases} \sum\limits_{e} e^{th(x)} p_X(x) & \text{ἤ} \\ \int\limits_{-\infty}^{\infty} e^{th(x)} f_X(x) dx \ . \end{cases}$$

Ἐάν ἡ ἄθροιση ἤ ἡ ὁλοκλήρωση ἐκτελεστοῦν καί ἡ συνάρτηση πού προκύπτει ἀναγνωριστεῖ τότε βρίσκουμε τήν κατανομή τῆς Y. Ἡ μέθοδος θά ἐφαρμοστεῖ πιό κάτω πολλές φορές γιά τήν εὕρεση τῆς κατανομῆς ἀθροισμάτων ἀνεξάρτητων τυχαίων μεταβλητῶν.

<u>Παράδειγμα</u> <u>6.11</u>:

Ἔστω $X \sim N(0,1)$ καί $Y=X^2$. Τότε

$$m_Y(t) = \int_{-\infty}^{\infty} e^{tx^2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-1/2\ x^2(1-2t)} dx =$$

$$= \frac{1}{(1-2t)^{1/2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = (1-2t)^{-1/2}$$

Ἄρα $\qquad Y \sim G(a=\frac{1}{2},\ \beta=2) = X_1^2$.

6. 4 Ἀθροίσματα ἀνεξάρτητων τυχαίων μεταβλητῶν

Στό ἐδάφιο αὐτό μελετᾶμε ἀθροίσματα καί γραμμικούς συνδιασμούς ἀνεξάρτήτων τυχαίων μεταβλητῶν. Ἔστω X_1, $\dots, X_n$ n ἀνεξάρτητες τυχαῖες μεταβλητές, Y τό ἄθροισμα αὐτῶν καί Z ἕνας γραμμικός συνδυασμός μέ σταθερούς συντελεστές a_i. Μέ ἄλλα λόγια

$$Y=X_1+X_2+\ldots+X_n=\sum_{i=1}^{n} X_i$$

καί

$$Z=a_1X_1+a_2X_2+\ldots+a_nX_n=\sum_{i=1}^{n} a_iX_i\ .$$

θέλουμε νά μελετήσουμε τήν κατανομή καί τίς ροπές τῶν Y καί Z.

Ἡ εὕρεση τῆς κατανομῆς τῶν Y καί Z μπορεῖ νά θεωρηθεῖ σάν πολυδιάστατο πρόβλημα ἀλλαγῆς μεταβλητῶν. Ἡ μέθοδος ὅμως τῆς ροπογεννήτριας μᾶς δίνει ἕνα ἁπλό καί χρήσιμο θεώρημα.

Θ ε ώ ρ η μ α 6.7:

Ἔστω $X_1,\ldots,X_n$ n ἀνεξάρτητες τυχαῖες μεταβλητές γιά τίς ὁποῖες οἱ ροπογεννήτριες συναρτήσεις $\{m_{X_i}(t)\}$, $i=1,2,\ldots,n$, ὑπάρχουν γιά ἕνα κοινό διάστημα Δ τοῦ t.

1. Ἡ ροπογεννήτρια συνάρτηση τοῦ ἀθροίσματος $Y=\sum_{i=1}^{n} X_i$ εἶναι τό γινόμενο τῶν ροπογεννήτριων συναρτήσεων τῶν X_i. Μέ ἄλλα λόγια

$$m_{X_1+X_2+\ldots+X_n}(t)=m_{X_1}(t)\ldots m_{X_n}(t) \quad \forall t\in\Delta. \qquad (6.14)$$

2. Ἄν οἱ τυχαῖες μεταβλητές $\{X_i\}$ εἶναι ἰσόνομες, δηλαδή ἔχουν τήν ἴδια κατανομή μέ τήν τυχαία μεταβλητή X, τότε

$$m_{X_1+X_2+\ldots+X_n}(t)=[m_X(t)]^n . \qquad (6.15)$$

'Απόδειξη: Μέ βάση τό θεώρημα 6.6 καί τούς ὁρισμούς ἔχουμε

$$m_{X_1+X_2+\ldots+X_n}= Ee^{(X_1+X_2+\ldots+X_n)t} =Ee^{X_1t+X_2t+\ldots+X_nt}$$

$$=E[(e^{X_1t})(e^{X_2t})\ldots(e^{X_nt})]=(Ee^{tX_1})(Ee^{tX_2})(Ee^{tX_n})$$

$$= m_{X_1}(t)m_{X_2}(t)\ldots m_{X_n}(t) .$$

Τό δεύτερο μέρος τοῦ θεωρήματος ἀπορρέει εὔκολα ἀπό τό πρῶτο. ▼

Τό παραπάνω θεώρημα μᾶς δίνει τή δυνατότητα νά βρίσκουμε τήν κατανομή ἀθροισμάτων ἀνεξάρτητων τυχαίων μεταβλητῶν γιά πολλές περιπτώσεις. Στή συνέχεια δίνουμε μία σειρά ἐφαρμογῶν τοῦ θεωρήματος 6.7. 'Από αὐτές ἀποδεικνύουμε ἐνδεικτικά μόνο τήν πρώτη καί τήν τελευταία. Οἱ ἄλλες ἀποδεικνύονται μέ ἀνάλογο τρόπο.

Προτάσεις:

6.1: Ἔστω $X_1,\ldots,X_k$ k ἀνεξάρτητες τυχαῖες μεταβλητές μέ $X_i \sim B(n_i,p)$(διωνυμική), $i=1,2,\ldots,k$. Τότε ἡ τυχαία μεταβλητή $Y=\sum_{i=1}^{k} X_i$ εἶναι $B(\sum_{i=1}^{k} n_i,p)$.

'Απόδειξη: Ἡ ροπογεννήτρια τῆς διωνυμικῆς κατανομῆς εἶναι $(pe^t+q)^n$. Ἔτσι ἐφαρμόζοντας τό θεώρημα 6.7

ἔχουμε

$$m_Y(t)=(pe^t+q)^{n_1}\ldots(pe^t+q)^{n_k}=(pe^t+q)^{\sum_{i=1}^{k} n_i}$$

Τό μονοσήμαντο τῆς ροπογεννήτριας μᾶς δίνει τό ἀποτέλεσμα. ▼

6.2: Ἔστω $X_1,\ldots,X_n$ n ἀνεξάρτητες *Poisson* τυχαῖες μεταβλητές μέ παραμέτρους $\lambda_1,\ldots,\lambda_n$ ἀντιστοίχως καί $Y=\sum_{i=1}^{n} X_i$.

Τότε $Y\sim Poisson\ (\lambda_1+\ldots+\lambda_n)$.

6.3: Ἔστω $\{X_i\}$, $i=1,2,\ldots,n$, n ἀνεξάρτητες ἀρνητικές διωνυμικές τυχαῖες μεταβλητές μέ παραμέτρους (k_i,p), $i=1,2,\ldots,n$ ἀντιστοίχως καί $Y=\sum_{i=1}^{n} X_i$. Τότε ἡ κατανομή τῆς τυχαίας μεταβλητῆς Y εἶναι ἐπίσης ἀρνητική διωνυμική μέ παραμέτρους $(\sum_{i=1}^{n} k_i,p)$.

6.4: Ἔστω $X_1,\ldots,X_n$ n ἀνεξάρτητες τυχαῖες μεταβλητές μέ κατανομή Γάμμα καί παραμέτρους (a_1,β), (a_2,β), $\ldots,(a_n,\beta)$ ἀντιστοίχως. Ἔστω ἐπίσης $Y=\sum_{i=1}^{n} X_i$. Τότε ἡ τυχαία μεταβλητή Y ἔχει ἐπίσης τήν κατανομή Γάμμα μέ παραμέτρους $(\sum_{i=1}^{n} a_i,\beta)$.

6.5: Ἔστω $X_i\sim N(\mu_i,\sigma_i^2)$, $i=1,\ldots,n$, n ἀνεξάρτητες κανονικές τυχαῖες μεταβλητές καί $Y=\sum_{i=1}^{n} X_i$. Ἡ κατανομή τῆς Y εἶναι ἐπίσης κανονική μέ μέση τιμή $\mu=\sum_{i=1}^{n}\mu_i$ καί διακύμανση $\sigma^2=\sum_{i=1}^{n}\sigma_i^2$.

Ἀπόδειξη: Ἡ ροπογεννήτρια συνάρτηση τῆς κανονικῆς κατανομῆς μέ παραμέτρους (μ_i,σ_i^2) εἶναι

$$m_{X_i}(t)=e^{\mu_i t+\frac{1}{2}\sigma_i^2 t^2}$$

Ἔτσι ἔχουμε

$$m_Y(t)=m_{X_1}(t)\ldots m_{X_n}(t)=$$

$$=e^{\mu_1 t+\frac{1}{2}\sigma_1^2 t^2}\ldots e^{\mu_n t+\frac{1}{2}\sigma_n^2 t^2}=$$

$$=e^{(\mu_1+\ldots+\mu_n)t+\frac{1}{2}(\sigma_1^2+\ldots+\sigma_n^2)t^2}\;.$$

Τό ἀποτέλεσμα προκύπτει ἀπό τή μονοσήμαντο ἰδιότητα τῆς ροπογεννήτριας συναρτήσεως. ▼

Ἡ ἑπομένη πρόταση ἀποτελεῖ γενίκευση τῆς Πρότασης (6.5) καί ἀποδεικνύεται μέ τόν ἴδιο συλλογισμό. Ἡ ἀπόδειξη ἀφήνεται σάν ἄσκηση στόν ἀναγνώστη. Λόγω τῆς σπουδαιότητας της δίνεται σάν

Θεώρημα 6.8:

Ἔστω $X_i \sim N(\mu_i,\sigma_i^2)$, $i=1,2,\ldots,n$, n ἀνεξάρτητες κανονικές τυχαῖες μεταβλητές καί

$$Z=a_1X_1+a_2+X_2+\ldots+a_nX_n$$

ἕνας γραμμικός συνδυασμός τῶν X_i μέ σταθερούς συντελεστές a_i . Τότε

$$Z \sim N(\Sigma a_i \mu_i\ ,\ \Sigma a_i^2 \sigma_i^2)\ . \qquad (6.16)$$

<u>Ἐφαρμογή:</u> Ἔστω ὅτι τό σφάλμα πού γίνεται στή μέτρηση τοῦ μήκους ἑνός δρόμου ἀκολουθεῖ κανονική κατανομή μέ μέση τιμή *10 cm* καί τυπική ἀπόκλιση *2 cm*. Τό μῆκος τοῦ δρόμου μετριέται *3* φορές. Νά βρεθεῖ ἡ πιθανότητα τό ὁλικό σφάλμα νά εἶναι μικρότερο τῶν *22 cm*.

Ἄν X_i, $i=1,2,3$ εἶναι τό σφάλμα σέ κάθε μέτρηση τό ὁλικό σφάλμα θά εἶναι $X_1+X_2+X_3$. Εἶναι προφανές ὅτι τά X_i εἶναι ἀνεξάρτητες ἰσόνομες κανονικές τυχαῖες μεταβλητές μέ παραμέτρους $\mu_i=10cm$ καί $\sigma_i^2=4cm^2$. Ἔτσι $X_1+X_2+X_3 \sim N(\mu=30,\ \sigma^2=12)$. Μέ τή βοήθεια τοῦ πίνακα τῆς τυπικῆς κανονικῆς κατανομῆς ἡ πιθανότητα πού ζητᾶμε βρίσκεται ὅτι εἶναι

$$P(X_1+X_2+X_3 \leqq 22)=P\left(\frac{\Sigma X_i-\mu}{\sigma} \leqq \frac{22-30}{\sqrt{12}}\right)$$

$$= P\left(Z \leqq \frac{-2}{3.46}\right)=P(Z \leqq -2.31)=0,0104=\ 1,04\%\ .$$

Τά παραπάνω παραδείγματα χαρακτηρίζονται ἀπό τό γεγονός ὅτι τά ἀθροίσματα ἀνεξάρτητων τυχαίων μεταβλητῶν ἑνός τύπου ἔχουν κατανομή τοῦ ἰδίου τύπου. Τέτοιες κατανομές λέγονται ἀ ν α π α ρ α γ ω γ ι κ έ ς. Ὅλες οἱ κατανομές δέν εἶναι ἀναπαραγωγικές. Ἐπί παραδείγματι ἡ γεωμετρική καί ἡ ἐκθετική κατανομή δέν εἶναι ἀναπαραγωγικές. Τό ἄθροισμα n ἀνεξάρτητων καί ἰσόνομων γεωμετρικῶν τυχαίων μεταβλητῶν μέ παράμετρο p εἶναι ἀρνητική διωνυμική τυχαία μεταβλητή μέ παραμέτρους

(n,p) καί τό ἄθροισμα n ἀνεξάρτητων καί ἰσόνομων ἐκθετικῶν τυχαίων μεταβλητῶν μέ παράμετρο λ εἶναι Γάμμα τυχαία μεταβλητή μέ παραμέτρους $(n,1/\lambda)$.

Ἡ μέση τιμή καί ἡ διακύμανση ἀθροίσματος ἀνεξάρτητων τυχαίων μεταβλητῶν εἶναι τά ἀθροίσματα τῶν ἀντιστοίχων μέσων τιμῶν καί διακυμάνσεων τῶν τυχαίων μεταβλητῶν. Τό ἀποτέλεσμα αὐτό ἀπορρέει ἀπό τό ἀκόλουθο γενικώτερο θεώρημα πού ἀφορᾶ γραμμικούς συνδυασμούς τυχαίων μεταβλητῶν.

Θεώρημα 6.9:

Ἔστω $X_1,X_2,\dots,X_n$ n ἀνεξάρτητες τυχαῖες μεταβλητές καί

$$Z=a_1X_1+a_2X_2+\dots+a_nX_n$$

ἕνας γραμμικός συνδυασμός τῶν X_i μέ σταθερούς συντελεστές a_i, $i=1,2,\dots,n$. Τότε

$$EZ=\sum_{i=1}^{n} a_i EX_i \qquad (6.17)$$

καί

$$Var(Z)=\sum_{i=1}^{n} a_i^2\, Var(X_i)\ . \qquad (6.18)$$

Ἀπόδειξη: Τό πρῶτο μέρος τοῦ θεωρήματος ἀπορρέει εὔκολα ἀπό τίς ἰδιότητες τοῦ ἀθροίσματος καί τοῦ ὁλοκληρώματος. Θά ἀποδείξουμε τό δεύτερο μέρος γιά τήν περίπτωση $n = 2$

$$
\begin{aligned}
Var(Z) &= EZ^2-(EZ)^2 = \\
&= E(a_1X_1+a_2X_2)^2 - [E(a_1X_1+a_2X_2)]^2 \\
&= E(a_1^2X_1^2+a_2^2X_2^2+2a_1a_2X_1X_2) -[a_1^2(EX_1)^2+a_2^2(EX_2)^2 \\
&\quad +2a_1a_2(EX_1)(EX_2)]=a_1^2[EX_1^2-(EX_1)^2]+a_2^2[EX_2^2 \\
&\quad -(EX_2)^2]+2a_1a_2[EX_1X_2-(EX_1)(EX_2)] \\
&=a_1^2Var\ X_1+a_2^2Var\ X_2 ,
\end{aligned}
$$

διότι $EX_1X_2=(EX_1)(EX_2)$ λόγω τῆς ἀνεξαρτησίας τῶν X_1 καί X_2.▼

6. 5 Κατανομή τῶν max Xi καί min Xi

Στό ἐδάφιο αὐτό ἐξετάζουμε δύο ἰδιάζουσες καί ἐνδιαφέρουσες συναρτήσεις n ἀνεξάρτητων τυχαίων μεταβλητῶν $\{X_i\}$. τήν $Y=max\ X_i=max\ (X_1,X_2,\ldots,X_n)$ καί τήν $Z=minX_i = min\ (X_i,X_2,\ldots,X_n)$. Οἱ συναρτήσεις αὐτές ἔχουν τήν ἀκόλουθη ἔννοια: εἶναι γνωστό ὅτι οἱ τυχαῖες μεταβλητές X_i εἶναι πραγματικές συναρτήσεις μέ πεδίο ὁρισμοῦ τόν δειγματικό χῶρο S. Ἔτσι γιά κάθε $s\in S$ ἔχουμε τούς πραγματικούς ἀριθμούς $X_1(s)$, $X_2(s),\ldots,X_n(s)$. $Y(s)$ εἶναι τό μέγιστο τῶν $X_1(s)$, $X_2(s),\ldots,X_n(s)$ καί $Z(s)$ εἶναι τό ἐλάχιστο αὐτῶν.

<u>(Α) Κατανομή τοῦ $max\ X_i$</u>: Ἔστω $F_{X_i}(x)$ ἡ ἀθροιστική συνάρτηση κατανομῆς τῆς τυχαίας μεταβλητῆς

X_i, $i=1,2,\ldots,n$.
Τότε ἔχουμε

$$F_Y(y)=P(Y\leq y)=P(\max X_i\leq y)$$

$$=P(X_1\leq y, X_2\leq y,\ldots,X_n\leq y)$$

καί λόγω τῆς ἀνεξαρτησίας τῶν $\{X_i\}$ (βλέπε σχέση 6.9)

$$F_Y(y)=P(X_1\leq y)P(X_2\leq y)\ldots P(X_n\leq y).$$

Ἄρα

$$F_{\max X_i}(y)=F_{X_1}(y)F_{X_2}(y)\ldots(F_{X_n}(y), \quad y\in R^1 . \qquad (6.19)$$

Ἄν οἱ τυχαῖες μεταβλητές X_i εἶναι συνεχεῖς μέ συνάρτηση πυκνότητας πιθανότητας $f_{X_i}(x)$ τότε ἡ συνάρτηση πυκνότητας πιθανότητας $f_Y(y)$ τῆς Y εἶναι

$$f_{\max X_i}(y)=\sum_{i=1}^{n}\frac{F_{X_1}(y)\ldots F_{X_n}(y)}{F_{X_i}(y)}f_{X_i}(y), \quad y\in R^1 . \qquad (6.20)$$

Πόρισμα 6.1: Ἄν οἱ τυχαῖες μεταβλητές X_i, $i=1,2,\ldots,n$ εἶναι ἀνεξάρτητες καί ἰσόνομες μέ τήν τυχαία μεταβλητή X τότε

$$F_{\max X_i}(y)=[F_X(y)]^n , \quad y\in R^1 . \qquad (6.21)$$

καί ἄν ἐπί πλέον οἱ X_i εἶναι συνεχεῖς, τότε

$$f_{\max X_i}(y)=n[F_X(y)]^{n-1}f_X(y), \quad y\in R^1 . \qquad (6.22)$$

(Β) Κ α τ α ν ο μ ή τ ο ῦ $min\ X_i$: Χρησιμοποιῶντας τόν ἴδιο συμβολισμό ἔχουμε

$$F_Z(z)=P(Z\leqq z)=P(\min X_i\leqq z)=1-P(\min X_i>Z)$$

$$=1-P(X_1>z,X_2>z,\ldots,X_n>z)$$

καί λόγω τῆς ἀνεξαρτησίας τῶν $\{X_i\}$

$$F_Z(z)=1-P(X_1>z)P(X_2>z)\ldots P(X_n>z)$$

$$=1-[1-P(X_1\leqq z)]\ldots[1-P(X_n\leqq z)].$$

Ἄρα

$$F_{\min X_i}(z)=1-[1-F_{X_1}(z)]\ldots[1-F_{X_n}(z)],\quad z\in R^1 \qquad (6.23)$$

Ἄν πάλι οἱ τυχαῖες μεταβλητές X_i εἶναι συνεχεῖς τότε ἡ συνάρτηση πυκνότητας πιθανότητας τοῦ $\min X_i$ εἶναι

$$f_{\min X_i}(z)=\sum_{i=1}^{n}\frac{[1-F_{X_1}(z)]\ldots[1-F_{X_n}(z)]}{[1-F_{X_i}(z)]}f_{X_i}(z),\quad z\in R^1 . \qquad (6.24)$$

Πόρισμα 6.2: Ἄν οἱ τυχαῖες μεταβλητές X_i, $i=1,2,\ldots,n$ εἶναι ἀνεξάρτητες καί ἰσόνομες μέ τήν τυχαία μεταβλητή X τότε

$$F_{\min X_i}(z)=1-[1-F_X(z)]^n,\quad z\in R^1 \qquad (6.25)$$

καί ἄν ἐπί πλέον οἱ X_i εἶναι συνεχεῖς, τότε

$$f_{min\ X_i}(z)=n[1-F_X(z)]^{n-1} f_X(z),\quad z\in R^1 \ . \qquad (6.26)$$

Παραδείγματα:

6.12: Ἔστω ὅτι ὁ χρόνος ζωῆς μιᾶς ἠλεκτρικῆς λάμπας ἔχει ἐκθετική κατανομή μέ μέση τιμή *1000* ὧρες. Ἑκατό *(100)* τέτοιες λάμπες ἐγκαθίστανται ταυτόχρονα σέ μία φωτεινή διαφήμηση. Νά βρεθοῦν ἡ κατανομή καί ἡ μέση τιμή τοῦ χρόνου ζωῆς τῆς λάμπας πού θά καεῖ πρώτη.

Ἔστω X_i ὁ χρόνος ζωῆς τῆς i λάμπας, $i=1,2,\ldots,100$. Τότε $Z=\min X_i$ εἶναι ὁ χρόνος ζωῆς τῆς λάμπας πού θά καεῖ πρώτη. (Ὑποθέτουμε ὅτι οἱ τυχαῖες μεταβλητές X_i εἶναι ἀνεξάρτητες καί ἰσόνομες).

$$f_X(x)=\begin{cases}\frac{1}{1000}e^{-x/1000}, & x\geq 0\\ 0, & \text{ἀλλοῦ}\end{cases}$$

$$F_X(x)=\begin{cases}1-e^{-x/1000}, & x\geq 0\\ 0, & \text{ἀλλοῦ}\ .\end{cases}$$

Ἄρα

$$f_{min\ X_i}(z)=\begin{cases}100[1-(1-e^{-z/1000})]^{99}\frac{1}{1000}e^{-z/1000}, & z>0\\ 0, & \text{ἀλλοῦ}\end{cases}$$

$$= \begin{cases} \dfrac{100}{1000} e^{-\frac{100z}{1000}} & , \ z \geq 0 \\ 0 & , \ \text{ἀλλοῦ} \ . \end{cases}$$

Ἔτσι ἡ κατανομή τοῦ χρόνου ζωῆς τῆς λάμπας πού θά καεῖ πρώτη εἶναι ἐκθετική μέ $\lambda = 1/10$ καί ὁ μέσος χρόνος ζωῆς αὐτῆς εἶναι *10* ὧρες.

6.13: Ἔστω $X_1, X_2, \ldots, X_n$ n ἀνεξάρτητες καί ἰσόνομες τυχαῖες μεταβλητές μέ κατανομή τήν ὁμοιόμορφη πάνω στό διάστημα $(0, \theta)$, ὅπου θ εἶναι ἕνας σταθερός πραγματικός ἀριθμός. Ἡ ἀθροιστική συνάρτηση τῆς ὁμοιόμορφης κατανομῆς εἶναι

$$F_X(x) = \begin{cases} 0 & , \ x \leq 0 \\ \dfrac{x}{\theta} & , \ 0 \leq x \leq \theta \\ 1 & , \ x \geq \theta \ . \end{cases}$$

Ἄρα οἱ *(6.21)* καί *(6.25)* δίδουν:

$$F_{\max X_i}(y) = \begin{cases} 0 & , \ y \leq 0 \\ y^{n}/\theta^{n} & , \ 0 \leq y \leq \theta \\ 1 & , \ y \geq \theta \ . \end{cases}$$

$$F_{\min X_i}(z) = \begin{cases} 0 & , \ z \leq 0 \\ 1-[1-z/\theta]^{n} & , \ 0 \leq z \leq \theta \\ 1 & , \ z \geq \theta \ . \end{cases}$$

Ἡ συνάρτηση πυκνότητας πιθανότητας τῆς $U(0,\theta)$ εἶναι

$$f_X(x)=\begin{cases} 1/\theta\ , & 0\leq x\leq\theta \\ 0\ , & \text{ἀλλοῦ}\ . \end{cases}$$

Ἄρα οἱ (6.22) καί (6.26) δίνουν

$$f_{\max X_i}(y)=\begin{cases} n(\frac{y}{\theta})^{n-1}\frac{1}{\theta}=\frac{ny^{n-1}}{\theta^n}\ , & 0\leq y\leq\theta \\ 0\ , & \text{ἀλλοῦ} \end{cases}$$

καί

$$f_{\min X_i}(z)=\begin{cases} n[1-z/\theta]^{n-1}\frac{1}{\theta}\ , & 0\leq z\leq\theta \\ 0\ , & \text{ἀλλοῦ}\ . \end{cases}$$

ΑΣΚΗΣΕΙΣ

6.1. Ἔστω ὅτι οἱ τ.μ. X καί Y ἔχουν διακριτή ἀπό κοινοῦ κατανομή μέ σ.π.

$$p(x,y) = \begin{cases} c|x+y|, & \begin{array}{l} x=-2,-1,0,1,2 \\ y=-2,-1,0,1,2 \end{array} \\ 0 & \text{ἀλλοῦ} \end{cases} .$$

Νά βρεθοῦν α) $P(X=1)$, β) $P(X=0,\ Y=-2)$ καί γ) $P(|X-Y|\leq 1)$.

6.2. Ἄν X καί Y εἶναι ἀναξάρτητες τ.μ. μέ τήν ἴδια γεωμετρική κατανομή $p(x)=q^{x-1}p$, $x=1,2,\ldots$ καί $p+q=1$, νά βρεθεῖ ἡ πιθανότητα $P(X=Y)$.

6.3. Ἄν X συνεχής τυχαία μεταβλητή μέ ἀ.σ.κ. $F(x)$ καί σ.π.π. $f(x)$ νά βρεθεῖ ἡ ἀ.σ.κ. τῆς $Y=|X|+4$ συναρτήσει τῆς $F(x)$ καί ἡ σ.π.π. τῆς Y συναρτήσει τῆς $f(x)$.

6.4. Ἔστω τ.μ. μέ σ.π.π. $f(x)=(1+x)/2$ ἄν $-1<x<1$ καί $f(x)=0$ ἀλλοῦ. Νά βρεθοῦν ἡ ἀ.σ.κ. καί σ.π.π. τῆς $Y=X^2$.

6.5. Ἀποδείξατε μέ τή μέθοδο τῆς ροπογεννήτριας συνάρτησης ὅτι ἄν $X\sim N(10,16)$ τότε $5X-2\sim N(48,400)$.

6.6. Ἔστω X τ.μ. τέτοια ὥστε ἡ τ.μ. lnX εἶναι $N(0,1)$. Νά βρεθεῖ ἡ πυκνότητα τῆς X.

6.7. Νά δειχθεῖ ὅτι ἡ κατανομή τῆς τ.μ. $X=(\sum_{i=1}^{n} X_i)/n$, ὅπου $X_1,\ldots,X_n$ εἶναι ἀνεξάρτητες Γάμμα μεταβλητές μέ παραμέτρους a καί β, εἶναι ἐπίσης Γάμμα μέ παραμέτρους na καί $β/n$

6.8. Ἔστω $X \sim Gamma\ (a=2,\ β=2)$ καί $F(x)$ ἡ ἁ.σ.κ. τῆς X. Νά βρεθοῦν a) $P[1/4\ F(X)]1/2$, b) $P[F(X)]2,5$ c) $E[F(X)]^3$.

6.9. Μέ τή βοήθεια ἑνός τυχαίου μηχανισμοῦ ἐκλέγεται ὁμοιόμορφα στήν τύχη ἕνας ἀριθμός ἀπό τό διάστημα $(0,1)$. Νά βρεθοῦν οἱ πιθανότητες (i) τό δεύτερο δεκαδικό ψηφίο τοῦ ἀριθμοῦ νά εἶναι τό ψηφίο 2 καί (ii) τό δεύτερο δεκαδικό ψηφίο τῆς τετραγωνικῆς ρίζας τοῦ ἀριθμοῦ νά εἶναι τό 5.

6.10. Στά παιδιά μίας ἡλικίας τό ὕψος X καί τό βάρος Y συνδέονται μέ τή σχέση $Y=2X+4$. Ἔστω ὅτι τό βάρος ἔχει κανονική κατανομή μέ μέση τιμή 100 καί τυπική ἀπόκλιση 5. Ποία ἡ πιθανότητα τό ὕψος ἑνός παιδιοῦ τῆς ἰδίας ἡλικίας νά εἶναι μεταξύ 30 καί 48;

6.11. Ἄν ἡ τ.μ. X ἔχει ὁμοιόμορφο κατανομή στό διάστημα $[-1,1]$ νά βρεθεῖ ἡ σ.π.π. τῆς $Y=X^3$.

6.12. Ἄν ἡ σ.π.π. τῆς τ.μ. X εἶναι $f(x)=x^2/9$ ἄν $0<x<3$ καί $f(x)=0$ ἀλλοῦ, νά βρεθεῖ ἡ κατανομή τῆς τ.μ. $Y=X^3$.

6.13. Ἄν X τ.μ. μέ σ.π.π. $f(x)=12x^2(1-x)$ ἄν $0<x<1$ καί $f(x)=0$ ἀλλοῦ νά βρεθεῖ ἡ σ.π.π. τῆς $Y=1/X$.

6.14. Ἔστω X τ.μ. μέ σ.π.π. $f(x)=320x^3/(1+2x)^6$, ὅταν $x>0$ καί $f(x)=0$ ἀλλοῦ. Νά βρεθεῖ καί ἀναγνωριστεῖ ἡ κατανομή τῆς τ.μ. $Y=2X/(1+2X)$.

6.15. Ὁ Νῖκος πηγαίνει σπίτι του ἀπό τό γραφεῖο πρῶτα μέ τόν ὑπόγειο σιδηρόδρομο (*metro*) καί κατόπιν μέ ἀστικό λεωφορεῖο. Ἔστω ὅτι ὁ χρόνος ἀναμονῆς γιά τό *metro* σέ ὥρα αἰχμῆς εἶναι ἐκθετική τ.μ. μέ μέση τιμή 2 *min*. Ἔστω ἐπίσης ὅτι ὁ χρόνος ἀναμονῆς γιά τό λεωφορεῖο σέ ὥρα αἰχμῆς εἶναι ἐκθετική τ.μ. μέ μέση τιμή 3 *min*. Νά βρεθεῖ ἡ πιθανότητα ὁ συνολικός χρόνος ἀναμονῆς τοῦ Νίκου νά εἶναι λιγώτερος ἀπό 6 *min*.

6.16. Νά δειχθεῖ ὅτι ἄν X καί Y εἶναι ἀνεξάρτητες *Poisson* μέ παραμέτρους λ καί μ ἀντίστοιχα, τότε $X+Y$ εἶναι *Poisson* τ.μ. μέ παράμετρο λ+μ.

6.17. Ἄν οἱ τ.μ. X καί Y εἶναι ἀνεξάρτητες κανονικές $N(\mu_x=25,\ \sigma_x^2=4)$, $N(\mu_y=35,\ \sigma_y^2=16)$ ἀντίστοιχα καί $Z=3X-2Y$, νά βρεθοῦν ἡ $P(-2<Z<19)$.

6.18. Ἄν $X_1,X_2,\ldots X_n$ εἶναι n ἀνεξάρτητες καί ἰσοδύναμες ἐκθετικές τ.μ. μέ παράμετρο λ, νά βρεθοῦν οἱ κατανομές τῶν τ.μ. $\max X_i$ καί $\min X_i$.

6.19. Τά ἄτομα πού χρησιμοποιοῦν τό ἀσανσέρ μίας πολυκατοικίας θεωροῦνται ὅτι προέρχονται ἀπό κάποιο κανονικό πληθυσμό μέ μέση τιμή 60Kg καί τυπική ἀπόκλιση 10Kg. Ποία ἡ πιθανότητα 4 ἄτομα νά ἔχουν συνολικά βάρος πού δέν θά ξεπερνᾶ τά $280Kg$, τό ὅριο φορτίου τοῦ ἀσανσέρ;

6.20. Νά ἀποδειχθεῖ ἡ σχέση

$$F_X(x) + F_Y(y) - 1 \leq F_{X,Y}(x,y) \leq \sqrt{F_X(x)F_Y(y)}$$

γιά κάθε x,y ὅπου $F_{X,Y}(x,y)$, $F_X(x)$, $F_Y(y)$ εἶναι οἱ ἀθροιστικές συναρτήσεις κατανομῆς τῶν τ.μ. (X,Y), X καί Y ἀντίστοιχα.

6.21. Ἄν X καί Y εἶναι ἀνεξάρτητες τ.μ. μέ διωνυμικές κατανομές $B(n=3,\ p=1/3)$ καί $B(n=2,\ p=1/2)$ ἀντίστοιχα, νά βρεθεῖ ἡ πιθανότητα $P(X=Y)$.

6.22. Ἕξη χαρτιά ἐκλέγονται χωρίς ἐπανατοποθέτηση ἀπό μία κοινή τράπουλα τῶν 52 χαρτιῶν. Νά βρεθεῖ ἡ ἀπό κοινοῦ κατανομή (σ.π.) τοῦ ἀριθμοῦ τῶν ἄσσων X καί τοῦ ἀριθμοῦ τῶν "βαλέ" Y.

6.23. Ἄν ἡ τ.μ. X ἔχει ὁμοιόμορφο κατανομή στό διάστημα $(-\pi/2\ ,\ \pi/2)$, νά βρεθεῖ ἡ κατανομή τῆς τ.μ. $Y=\text{ἐφ}\,X$.

6.24. Ἄν ἡ τ.μ. X ἔχει κανονική κατανομή $N(\mu,\sigma^2)$ νά βρεθεῖ ἡ κατανομή, ἡ μέση τιμή καί ἡ διακύμανση τῆς τ.μ. $Y=e^X$.

6.25. Ἄν ἡ τ.μ. X ἔχει ὁμοιόμορφο κατανομή στό διάστημα $(0,1)$ νά βρεθεῖ ἡ κατανομή τῆς τ.μ. $Y=1/X$.

6.26. Ἄν ἡ τ.μ. X ἔχει Βῆτα κατανομή $B(\alpha,\beta)$ νά βρεθεῖ ἡ κατανομή τῆς τ.μ. $Y=1-X$.

6.27. Ἄν ἡ τ.μ. X ἔχει ἐκθετική κατανομή 'Εκθ $(\lambda=1)$ νά βρεθεῖ ἡ κατανομή τῆς τ.μ. $Y=X/(1+X)$.

6.28. Ἄν ἡ τ.μ. X ἔχει κατανομή

$$f_X(x)=\frac{1}{\pi(1+x^2)}, \quad -\infty < x < +\infty$$

νά βρεθεῖ ἡ κατανομή τῆς τ.μ. $Y=1/X$.

6.29. Ἄν ἡ τ.μ. X ἔχει κατανομή

$$f_X(x) = \begin{cases} \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}} & , x>0 \\ 0 & , \text{ἀλλοῦ} \end{cases}$$

νά βρεθεῖ ἡ κατανομή τῆς τ.μ. $Y = \alpha X^{\beta}$.

6.30. Ἄν ἡ τ.μ. X ἔχει ὁμοιόμορφο κατανομή στό διάστημα $(0,1)$ νά βρεθεῖ ἡ κατανομή τῆς τ.μ. $Y=3X+1$.

6.31. Ἄν X_1, X_2 ἀνεξάρτητες τυχαῖες μεταβλητές ἡ κάθε μία μέ ἐκθετική κατανομή 'Εκθ$(\lambda=1)$ νά βρεθεῖ ἡ κατανομή τῆς τ.μ. $Y=(X_1+X_2)/2$.

6.32. Νά δειχθεῖ ὅτι ἄν οἱ τ.μ. $X_1, X_2, \ldots, X_n$ εἶναι ἀνεξάρτητες ἀρνητικές διωνυμικές κατανομές μέ παραμέτρους (k_i, p), $i = 1, \ldots, n$, ἀντίστοιχα τότε ἡ κατανομή τῆς τ.μ. $Y = X_1 + \ldots + X_n$ εἶναι ἐπίσης ἀρνητική διωνυμική μέ παραμέτρους $(\sum_{i=1}^{n} k_i, p)$.

6.33. Ἄν ἡ διακριτή τυχαία μεταβλητή X ἔχει σ.π. $p(x) = (1/2)^x$, $x = 1, 2, \ldots$, νά βρεθεῖ ἡ κατανομή τῆς τ.μ. $Y = X^3$.

6.34. Ἄν X_1, X_2 εἶναι ἀνεξάρτητες τ.μ. μέ κανονικές κατανομές $N(\mu_1 = 6, \sigma_1^2 = 1)$ καί $N(\mu_2 = 7, \sigma_2^2 = 1)$ ἀντίστοιχα, νά βρεθεῖ ἡ πιθανότητα $P(X_1 > X_2)$.

6.35. Ἔστω X_1, X_2 ἀνεξάρτητες τ.μ. μέ διωνυμικές κατανομές $B(n_1, p_1 = 1/2)$ καί $B(n_2, p_2 = 1/2)$, ἀντίστοιχα. Νά δειχθεῖ ὅτι ἡ κατανομή τῆς τ.μ. $Y = X_1 - X_2 + n_2$ εἶναι διωνυμική μέ παραμέτρους $n = n_1 + n_2$, $p = 1/2$.

Μερος II: ΣΤΑΤΙΣΤΙΚΗ

ΚΕΦΑΛΑΙΟ 7

τυχαια δειγματα - συχνοτητες - γραφικες παραστασεις - αριθμητικα χαρακτηριστικα

7.1 Γενικότητες

Στατιστική εἶναι ὁ κλάδος ὁ ὁποῖος ἀσχολεῖται μέ τήν σχεδίαση πειραμάτων καί μεθόδων δειγματοληψιῶν, τή συλλογή καί ἀνάλυση ἀριθμητικῶν δεδομένων (μετρήσεων) καί τήν ἐξαγωγή συμπερασμάτων γιά ἕνα σύνολο βάσει τῶν πληροφοριῶν πού περιέχονται σ'ἕνα δεῖγμα ἀπό τό σύνολο αὐτό. Τό θεωρητικό ὑπόβαθρο τῆς Στατιστικῆς εἶναι ἡ θεωρία τῶν Πιθανοτήτων.

Ἡ Στατιστική μαζί μέ τήν θεωρία Πιθανοτήτων χρησιμοποιεῖται σέ ὅλους σχεδόν τούς κλάδους τῆς ἀνθρώπινης δραστηριότητας καί γνώσης ὅπως π.χ.οἱ ἐπιχειρήσεις, ἡ διοίκηση, ἡ ἐκπαίδευση, ἡ ψυχολογία, ἡ γεωπονική, τά οἰκονομικά,ἡ βιολογία, ἡ γενετική, ἡ λογοτεχνία κ.λ.π. Μέ τήν πάροδο τοῦ χρόνου καί τήν ἐξάπλωση τῶν ἠλεκτρονικῶν ὑπολογιστῶν ἡ ἐφαρμογή καί ἡ ἀνάπτυξη τῶν στατιστικῶν μεθόδων γίνεται ὁλοένα καί μεγαλύτερη.

Ἰδιαίτεροι κλάδοι ἔχουν δημιουργηθεῖ ἀπό τήν χρησιμοποίηση τῶν στατιστικῶν ἐργαλείων καί ἐννοιῶν στήν

Οἰκονομία, Βιολογία καί 'Ιατρική Ψυχολογία κ.λ.π. ὅπως ἡ Οἰκονομετρία, ἡ Βιομετρία, ἡ Ψυχομετρία κ.λ.π.

Οἱ μέθοδοι τῆς Στατιστικῆς χρησιμοποιοῦνται κατά παράδοση γιά περιγραφικούς σκοπούς, γιά σύμπτυξη καί περιληπτική παρουσίαση τῶν ἀριθμητικῶν δεδομένων. 'Η Περιγραφική Στατιστική ἀσχολεῖται μέ τήν πινακοποίηση τῶν δεδομένων, τήν παράσταση τους ὑπό μορφή γραφημάτων ἤ εἰκόνων καί τόν ὑπολογισμό περιγραφικῶν μέτρων. Τά θέματα αὐτά καλύπτονται στά ἑπόμενα ἐδάφια.

'Η σύγχρονη Στατιστική ἀσχολεῖται κυρίως μέ τή λεγόμενη Στατιστική Συμπερασματολογία. 'Εφ'ὅσον δέν χρησιμοποιοῦμε τά μέτρα πού ὑπολογίζουμε γιά νά κάνουμε γενικεύσεις, τότε ἁπλῶς περιγράφουμε ὅτι παρατηροῦμε. 'Αλλά εὐθύς ὡς κάνουμε μιά ἐπαγωγική γενίκευση τότε ἀφήνουμε τήν περιγραφή καί μπαίνουμε στό χῶρο τῆς συμπερασματολογίας (ἐπαγωγῆς). Ἕνας δημοσιογράφος, γιά παράδειγμα, κάνει σφυγμομέτρηση κοινῆς γνώμης σέ μία πόλη, καί ἐρωτᾶ ἑκατό ἀνθρώπους ἄν ὑποστηρίζουν μία ἀλλαγή σέ κάποιο νόμο. 'Εβδομήντα πέντε, ἀπό αὐτούς, ἀποκρίνονται ὅτι ὑποστηρίζουν τήν ἀλλαγή. Ἄν ὁ δημοσιογράφος γράφει ὅτι ἑβδομήντα πέντε (75) ἀπό ἑκατό (100) ἄτομα, πού ἐρωτήθηκαν, ὑποστηρίζουν τήν ἀλλαγή, τότε ἁπλῶς, αὐτός, περιγράφει ὅτι παρατήρησε. Ἄν ὅμως γράφει ὅτι 75% τῶν κατοίκων τῆς πόλης ὑποστηρίζουν τήν ἀλλαγή, τότε συμπεραίνει ὅτι, τό ποσοστό τῶν κατοίκων τῆς πόλης, πού ὑποστηρίζουν τήν ἀλλαγή, εἶναι τό ἴδιο μέ ἐκεῖνο τοῦ μικροῦ δείγματος του. Ξεκινώντας ἀπό τήν πληροφορία πού τοῦ δίνει τό δεῖγμα διατυπώνει μία πρόταση γιά ένα μεγαλύτερο σύνολο άπό τό όποιο πάρθηκε τό δεΐγμα. Στήν πραγματικότητα είναι ίσως δύσκολο νά μή γενικεύσει έστω καί ύποσυνείδετα.

Ἡ παραπάνω στατιστική συμπερασματολογία έχει ίσχύ, άν σωστές στατιστικές άρχές έφαρμόστηκαν στήν έπιλογή τοῦ δείγματος καί τοῦ έκτιμητοῦ. Ἄν, παραδείγματος χάριν, τό δεΐγμα τοῦ δημοσιογράφου ἦταν τυχαΐο (λεπτομέρειες γιά τυχαΐα δείγματα δίνονται στή §7.2), τότε όρθῶς μπορεΐ νά ύποθέσει ότι τό άληθινό ποσοστό γιά όλη τήν κοινότητα πού ύποστηρίζει τήν άλλαγή πρέπει νά εὐρίσκεται γύρω στό 75%. Μήπως είναι 65% ή 85%; Είναι δύσκολο νά κρίνει. Ἡ άπάντηση χρειάζεται θεωρία Πιθανοτήτων. Μέ τή βοήθειά της μποροῦμε νά κατασκευάσουμε ένα άξιόπιστο διάστημα ποσοστῶν έκείνων πού ύποστηρίζουν τήν άλλαγή στό νόμο.

Ὅταν τά άτομα ή ύλικά πού διερευνῶνται έχουν μεγάλη μεταβλητότητα ή όταν ό έλεγχόμενος πειραματισμός είναι άδύνατος, (π.χ. σφυγμομέτρηση κοινῆς γνώμης,έρευνες άγορᾶς, κοινωνιολογία, οίκονομία κ.λ.π.) ή στατιστική σχεδίαση τῶν δειγματοληπτικῶν έρευνῶν παρέχει στούς έρευνητές μεθόδους χωρίς τίς όποιες θά ἦταν άδύνατον νά βγάλουν όποιοδήποτε συμπέρασμα.

Τά άριθμητικά δεδομένα πού έπεξεργάζεται ή Στατιστική προέρχονται άπό μία ή περισσότερες μεταβλητές οί όποιες άναφέρονται σέ διάφορα φαινόμενα ή πειράματα (φυσικά, κοινωνικά, ίατρικά κ.λ.π.).

Ἰδιαίτερα οί μεταβλητές τῆς Στατιστικῆς είναι τυχαΐες μεταβλητές. Τά φαινόμενα ή πειράματα κυβερνῶνται άπό τούς νόμους τῆς τύχης καί οί τιμές τῶν μετα -

βλητῶν εἶναι ἀποτέλεσμα παραγόντων τύχης. Συχνά τά ἀριθμητικά δεδομένα ἤ οἱ τιμές τῶν μεταβλητῶν λέγονται καί π α ρ α τ η ρ ή σ ε ι ς ἤ ἁπλῶς μ ε τ ρ ή σ ε ι ς, ἰδιαίτερα ὅταν προκύπτουν ἀπό διαδικασίες μετρήσεων ἤ παρατηρήσεων. Στήν ἀρχική τους φάση οἱ μετρήσεις ἤ παρατηρήσεις δέν εἶναι κατ'ἀνάγκην ἀριθμητικοῦ χαρακτῆρος. Μπορεῖ νά εἶναι σύμβολα, εἰκόνες, καμπύλες κ.λ.π. Εὔκολα ὅμως μέ κατάλληλες συναρτήσεις (μετασχηματισμούς)μετατρέπονται σέ ἀριθμητικά δεδομένα. Τά δεδομένα αὐτά συμβολίζονται μέ:

$x_1, x_2, \ldots, x_n$	ὅταν προέρχονται ἀπό τήν μεταβλητή x
$y_1, y_2, \ldots, y_n$	ὅταν προέρχονται ἀπό τήν μεταβλητή y
$(x_1,y_1), \ldots, (x_n,y_n)$	ὅταν προέρχονται ἀπό τήν μεταβλητή (x,y)

x_{ij}, $i=1,2,\ldots,k$, $j=1,2,\ldots,n_i$ μέ $n_1+n_2+\ldots+n_k=n$ κ.λ.π.

n παριστάνει τό πλῆθος ἤ ἀριθμό τῶν μετρήσεων.

Οἱ μεταβλητές διακρίνονται σέ δύο γενικές κατηγορίες: π ο σ ο τ ι κ έ ς καί π ο ι ο τ ι κ έ ς μεταβλητές. Οἱ πρῶτες εἶναι μεταβλητές πού μποροῦν νά μετρηθοῦν π.χ. τό μῆκος, τό ὕψος, τό πλάτος, τό βάρος, ἡ ἡλικία, ὁ χρόνος μεταξύ διαδοχικῶν ἐμφανίσεων ἑνός ἐνδεχομένου, ὁ ἀριθμός τηλεφωνικῶν κλήσεων σέ μία ὥρα κ.λ.π. 'Αντίθετα οἱ ποιοτικές μεταβλητές δέν μποροῦν νά μετρηθοῦν καί ἀναφέρονται σέ καταστάσεις, κατηγορίες,ἐπίπεδα, εἴδη, παρουσία ἤ ἀπουσία ἑνός ἤ περισσοτέρων γνωρισμάτων κ.λ.π. Γιά παράδειγμα τό φύλο τῶν ἀνθρώπων τό χρῶμα, ἡ οἰκονομικοκοινωνική κατάσταση (φτωχός, πλούσιος κ.λ.π.), ἡ ἐπιτυχία, ἡ ἀποτυχία κ.λ.π. .

'Ανάλογη εἶναι καί ἡ διάκριση τῶν μεταβλητῶν σέ συνεχεῖς καί διακριτές ἤ ἀπαριθμητές μεταβλητές πού ἔγινε στήν ἀρχή τῶν μαθημάτων αὐτῶν.

7. 2 Πληθυσμός - Τυχαῖο δεῖγμα - Κατανομή

Οἱ πρῶτες βασικές ἔννοιες τῆς Στατιστικῆς εἶναι ὁ πληθυσμός καί τό δεῖγμα. 'Η ἔννοια τοῦ πληθυσμοῦ ταυτίζεται μέ τήν ὁμώνυμη ἔννοια στή θεωρία Πιθανοτήτων. Τό σύνολο τῶν τιμῶν μίας τυχαίας μεταβλητῆς (μονοδιάστατης ἤ πολυδιάστατης) εἶναι ἕνας (θεωρητικός) πληθυσμός. Ἔτσι μερικές φορές ὁ πληθυσμός συμβολίζεται ὅπως καί οἱ τυχαῖες μεταβλητές καί λέμε ἔστω ὁ πληθυσμός X,Y,Z κ.λ.π.

Τό μέγεθος τοῦ πληθυσμοῦ μπορεῖ νά εἶναι πεπερασμένο, ἄπειρο ἤ ὑπεράριθμο. Τό μέγεθος ἐπίσης μπορεῖ νά αὐξάνεται ἤ ἐλαττοῦται ἀνάλογα μέ τό ἰδιαίτερο ἐνδιαφέρον μας. Ἄν, ἐπί παραδείγματι, ἐνδιαφερόμαστε γιά τήν γενική βαθμολογία τῶν εἰσαγωγικῶν ἐξετάσεων στά 'Ανώτατα 'Εκπαιδευτικά 'Ιδρύματα μίας χρονιᾶς, ὁ πληθυσμός μας ἀποτελεῖται ἀπό ὅλους τούς ὑποψήφιους πού ἔδωσαν ἐξετάσεις καί ἰδιαίτερα ἀπό ὅλες τίς γενικές βαθμολογίες τῆς χρονιᾶς ἐκείνης. Ἄν ἐνδιαφερόμαστε γιά τή γενική βαθμολογία τοῦ Πολυτεχνικοῦ-Φυσικομαθηματικοῦ Κύκλου τότε ἔχουμε ἕνα διαφορετικό (μικρότερο) πληθυσμό. Ἔτσι οἱ πληθυσμοί στήν Στατιστική διαμορφώνονται ἤ καθορίζονται ἀνάλογα μέ τή σφαίρα τοῦ ἐνδιαφέροντος μας.

Ὁ πληθυσμός ἀποτελεῖ τό ἄγνωστο μέρος τῆς Στατιστικῆς. Σκοπός τῆς Στατιστικῆς εἶναι ἡ ἐξαγωγή συμπερασμάτων γιά τόν πληθυσμό βάσει τοῦ δείγματος.

Τό δεῖγμα ὁρίζεται ἁπλῶς σάν ἕνα μέρος τοῦ πληθυσμοῦ. Ἀναφερόμενοι στό παραπάνω παράδειγμα οἱ πρωτοετεῖς φοιτητές τῆς Ἰατρικῆς Σχολῆς τοῦ Παν/μίου Ἰωαννίνων ἀποτελοῦν ἕνα δεῖγμα καί οἱ γενικές βαθμολογίες τους στίς εἰσαγωγικές ἐξετάσεις ἀποτελοῦν ἕνα δεῖγμα γενικῶν βαθμολογιῶν. Οἱ τιμές τῶν μεταβλητῶν πού προκύπτουν ἀπό τά μέλη τοῦ δείγματος λέγονται καί τιμές τοῦ δείγματος. Ὑπάρχουν διάφορα εἴδη δειγμάτων. Βασικός στόχος εἶναι τό δεῖγμα νά ἀντιπροσωπεύει τόν πληθυσμό. Στήν ἀρχή θά περιοριστοῦμε στό τυχαῖο δεῖγμα.

Τυχαῖο δεῖγμα εἶναι τό δεῖγμα πού ἐκλέγεται κατά τέτοιο τρόπο ὥστε ὅλα τά μέλη τοῦ πληθυσμοῦ νά ἔχουν ἴση καί ἀνεξάρτητο πιθανότητα νά συμπεριληφθοῦν στό δεῖγμα.

Εἶναι προφανές ὅτι οἱ τιμές τοῦ τυχαίου δείγματος εἶναι τιμές τῆς τυχαίας μεταβλητῆς (ἤ τῶν τυχαίων μεταβλητῶν) τοῦ πληθυσμοῦ. Ἔτσι οἱ γενικές βαθμολογίες τοῦ δείγματος εἶναι τιμές τῆς τυχαίας μεταβλητῆς "γενική βαθμολογία εἰσαγωγικῶν ἐξετάσεων". Ἄν δέ φανταστοῦμε ὅτι ἡ ὅλη δειγματοληψία μπορεῖ νά ἐπαναληφθεῖ νοερῶς ἕνα μεγάλο (ἄπειρο) ἀριθμό φορῶν, βλέπομε ὅτι οἱ τιμές τοῦ δείγματος μεταβάλλονται σέ ὅλο τό φάσμα τῶν τιμῶν τῆς τυχαίας μεταβλητῆς. Μέ τίς νοερές αὐτές ἐπαναλήψεις κάθε τιμή τοῦ δείγματος γίνεται τυχαία μεταβλητή ἰσόνομη μέ τήν X καί ἀνεξάρτητη ἀπό τίς ἄλλες. Ἔτσι ἔχουμε τόν ἀκόλουθο ὁρισμό:

Ὁρισμός 7.1: Τυχαΐο δεΐγμα μεγέθους n ἀπό ἕνα πληθυσμό X ἤ μία κατανομή εἶναι n ἀνεξάρτητες καί ἰσόνομες τυχαΐες μεταβλητές $X_1, X_2, \ldots, X_n$.

Ἀπό τά παραπάνω βλέπουμε ὅτι οἱ τιμές ἑνός τυχαίου δείγματος παίζουν ἕνα διπλό ρόλο: ἀφ'ἑνός εἶναι συγκεκριμμένες ἀριθμητικές τιμές, ἀφ'ἑτέρου εἶναι καί τυχαΐες μεταβλητές μέ τήν ἔννοια τῶν νοερῶν ἐπαναλήψεων πού ἀναφέραμε πιό πάνω. Ἔτσι τό τυχαΐο δεΐγμα ἄλλοτε θά συμβολίζεται μέ $x_1, x_2, \ldots, x_n$ καί ἄλλοτε μέ $X_1, X_2, \ldots, X_n$.

Ὑπάρχουν πολλοί τρόποι μέ τούς ὁποίους μποροῦμε νά ἐκλέξουμε ἕνα τυχαΐο δεΐγμα. Ὁ πλέον συνηθισμένος τρόπος εἶναι μέ τή βοήθεια π ι ν ά κ ω ν τ υ χ α ί ω ν ἀ ρ ι θ μ ῶ ν. Δεΐγμα πίνακος τυχαίων ἀριθμῶν παρατίθεται πιό κάτω.

99 60 50 50 60	12 48 08 01 88	60 51 73 38 54	47 41 67 43 84	21 35 12 90 43
88 61 29 18 05	31 29 56 94 33	58 88 68 42 10	08 11 15 96 06	15 51 48 39 96
19 28 81 63 24	30 96 40 11 59	36 16 01 02 60	36 16 26 83 53	41 09 50 85 01
53 61 62 34 47	04 37 74 97 09	77 36 92 80 45	99 26 28 24 24	54 63 43 63 54
40 03 44 30 11	42 25 70 19 79	90 12 36 16 80	28 70 24 86 07	76 17 01 50 80
65 15 18 13 54	05 13 69 91 51	84 57 52 89 88	12 52 03 39 71	19 48 20 94 16
95 79 58 84 86	00 04 73 69 94	89 12 93 84 29	72 62 79 66 98	65 17 54 69 56
75 26 86 16 42	65 03 22 43 68	87 68 70 09 18	92 85 94 60 32	97 44 95 82 72
92 45 48 29 84	56 60 50 64 07	71 46 35 31 52	21 80 61 25 30	31 99 58 07 04
43 00 97 26 90	99 85 55 75 16	09 55 34 16 16	94 32 12 12 07	32 90 97 62 47

Πίνακας 7.1 Τυχαΐοι Ἀριθμοί

Οἱ πίνακες αὐτοί κατασκευάζονται ἔτσι ὥστε κάθε ἀριθμός (μονοψήφιος, διψήφιος κ.λ.π.) νά ἔχει τήν ἴδια πυκνότητα (1/10, 1/100 κ.λ.π. ἀντιστοίχα) νά ἐμφανιστεῖ στόν πίνακα.

Ἡ τυχαία δειγματοληψία γίνεται ὡς ἑξῆς: Παίρνουμε (ἤ κατασκευάζουμε) ἕνα κατάλογο (frame) τοῦ πληθυσμοῦ καί ἀπαριθμοῦμε τά στοιχεῖα αὐτοῦ. Ἔτσι κάθε στοιχεῖο ἀντιστοιχεῖ σ'ἕνα μονοψήφιο ἤ διψήφιο κ.λ.π. ἀριθμό. Ξεκινώντας ἀπό ἕνα τυχαῖο σημεῖο τοῦ πίνακα τυχαίων ἀριθμῶν παίρνουμε ἕνα ἀριθμό μονοψήφιο, διψήφιο κ.λ.π. ἀνάλογα μέ τό ἄν τά στοιχεῖα τοῦ πληθυσμοῦ εἶναι μονοψήφια ἤ διψήφια κ.λ.π. Ἄν ὁ ἀριθμός αὐτός ἀντιστοιχεῖ σέ στοιχεῖα τοῦ πληθυσμοῦ, τό στοιχεῖο αὐτό ἐκλέγεται στό δεῖγμα. Ἄν ὄχι ὁ ἀριθμός αὐτός ἀπορρίπτεται καί παίρνουμε τόν γειτονικό ἀριθμό. Προχωροῦμε ἀπό γειτονικό σέ γειτονικό ἀριθμό γιά τήν ἐκλογή τῶν μελῶν τοῦ δείγματος.

Ἀντί γιά πίνακες τυχαίων ἀριθμῶν μποροῦμε νά χρησιμοποιήσουμε καί ἕνα τηλεφωνικό κατάλογο. Ἔχει ἀποδειχθεῖ ὅτι τά κεντρικά ψηφία ἑπταψήφιων ἀριθμῶν τηλεφώνου συμπεριφέρονται σάν τυχαῖοι ἀριθμοί.

Τυχαῖα δείγματα δέν ἐμφανίζονται πάντοτε στήν πράξη. Ἄλλοστε οἱ πιθανότητες ἐκλογῆς τῶν στοιχείων τοῦ πληθυσμοῦ εἶναι ἄνισες καί ἄλλοτε ἡ ἀνεξαρτησία δέν ἰσχύει. Ἡ Στατιστική στηρίζεται σέ δείγματα πιθανότητας, δηλαδή δείγματα πού ἔχουν ὡρισμένη πιθανότητα νά ἐκλεγοῦν. Ἀργότερα θά ὀνομάσουμε τήν πιθανότητα αὐτή πιθανοφάνεια (likelihood).

Ὁ πληθυσμός ἀποτελεῖ τό ἄγνωστο μέρος τῆς Στατιστικῆς. Τό δεῖγμα ἀποτελεῖ τό γνωστό μέρος. Ἡ Στατιστική εἶναι ἑπαγωγική ἐπιστήμη, ἡ ὁποία προσπαθεῖ νά βγάλει συμπεράσματα ἀπό τό ἐπί μέρους, γνωστό τμῆμα τοῦ πληθυσμοῦ, τό δεῖγμα, γιά τό καθολικό μέρος, ὁλόκληρο τόν πληθυσμό.

Μέ κάθε μέτρηση ὁ νοῦς τοῦ ἀνθρώπου μπορεῖ νά συσχετίσει ἕνα μέτρο κανονικότητος τῆς μετρήσεως (ἤ ἐλλείψεως αὐτῆς) καί πιό συγκεκριμμένα ἕνα μέτρο πιθανότητος πραγματοποιήσεως τῆς μετρήσεως. Αὐτό ὁδηγεῖ στήν ἔννοια τῆς κ α τ α ν ο μ ῆ ς ἡ ὁποία κατά κάποιο τρόπο δείχνει τήν θέση τῆς μετρήσεως ὡς πρός τίς ἄλλες μετρήσεις γνωστές ἤ ἄγνωστες. Ἔτσι ἔχουμε δύο κατανομές: τήν κ α τ α ν ο μ ή τ ο ῦ π λ η θ υ σ μ ο ῦ (τῆς τ.μ.X) καί τήν κ α τ α ν ο μ ή τοῦ δ ε ί γ μ α τ ο ς. Ἡ κατανομή τοῦ πληθυσμοῦ εἶναι γνωστή ἀπό τήν Θεωρία Πιθανοτήτων. Ἡ κατανομή τοῦ δείγματος στηρίζεται στήν ἔννοια τῆς σχετικῆς συχνότητος καί ἀναπτύσσεται πιό κάτω.

7. 3 Περιγραφή καί σύμπτυξη τῶν ἀριθμητικῶν

Ἔστω $x_1, x_2, \ldots, x_n$ ἕνα δεῖγμα ἀπό κάποιο πληθυσμό X. Οἱ τιμές x_i δέν εἶναι κατ'ἀνάγκη διάφορες μεταξύ τους. Ὁ ἀριθμός τῶν φορῶν πού ἡ τιμή x_i ἐμφανίζεται στό δεῖγμα λέγεται σ υ χ ν ό τ η ς τῆς τιμῆς καί συμβολίζεται μέ f_i.

Σ χ ε τ ι κ ή σ υ χ ν ό τ η ς τῆς τιμῆς x_i εἶναι ὁ ἀριθμός f_i/n. Ἄν k εἶναι ὁ ἀριθμός τῶν διακεκριμμένων τιμῶν τοῦ δείγματος τότε

$$\sum_{i=1}^{k} f_i = n \quad \text{καί} \quad \sum_{i=1}^{k} \frac{f_i}{n} = 1 .$$

Ἡ **ἀθροιστική συχνότης** F_i τῆς τιμῆς x_i ὁρίζεται ὡς ἑξῆς:

$$F_i = \begin{cases} \text{ἄθροισμα τῶν συχνοτήτων } f_i \text{ ὅλων τῶν} \\ \text{τιμῶν } x_j \text{ μέ τήν ἰδιότητα } x_j \leq x_i . \end{cases}$$

Εἶναι προφανές ὅτι $x_i < x_j$ ἄν καί μόνο ἄν $F_i < F_j$.

Ἡ συχνότης, σχετική συχνότης κ.λ.π. δέν ἀναφέρονται μόνο σέ μονοδιάστατες μετρήσεις x_i ἀλλά καί σέ πολυδιάστατες (x_i,y_i), (x_i,y_i,z_i) κ.λ.π.

<u>Παράδειγμα 7.1:</u>

Ὕψη 750 ἀνδρῶν (μέ προσέγγιση τοῦ πλησιέστερου ἑκατοστοῦ).

Ὕψος (cm) X_i :	155	156	157	158	159	160	161
Συχνότης f_i :	3	4	5	5	7	7	9
Ὕψος (cm) X_i :	162	163	164	165	166	167	168
Συχνότης f_i :	10	11	13	20	24	27	30
Ὕψος (cm) X_i :	169	170	171	172	173	174	175
Συχνότης f_i :	34	35	36	37	39	42	45
Ὕψος (cm) X_i :	176	177	178	179	180	181	182
Συχνότης f_i :	40	36	31	30	29	28	26

Ὕψος (cm) X_i :	183	184	185	186	187	188	189
Συχνότης f_i :	25	23	17	1	5	4	2

Παράδειγμα 7.2:

Ἀριθμός ἐργατικῶν ἀτυχημάτων ἑνός πυριτιδοποιείου σέ περίοδο 5 ἑβδομάδων.

Ἀριθμός ἀτυχημάτων X_i :	0	1	2	3	4	5
Συχνότης f_i :	447	132	42	21	3	2

Παράδειγμα 7.3:

Βάρος καί διάμετρος πορτοκαλιῶν

Βάρος	Διάμετρος (προσέγγιση μισοῦ cm)									
(Προσέγγιση 2 gr)	23.0	23.5	24.0	24.5	25.0	25.5	26.0	26.5	27.0	Σύνολο
190		1								1
192			1							1
194	1	2	4	3	2	1				13
196	1	1	6	7	5		1			21
198		2	8	12	13	5	3	1		44
200	1	3	7	13	20	14	5	2		65
202		1	2	9	11	18	10	4	1	56
204			1	1	1	2	3	2	1	11
206					2	4	5	1	1	13
208						1	2	2		5
210							2			2
Σύνολο	3	10	29	45	54	45	31	12	3	232

7.4 Περιγραφή καί σύμπτυξη τῶν ἀριθμητικῶν δεδομένων.

Ἡ ἀνάγκη περιγραφῆς ἤ περιληπτικῆς παρουσίασης τῶν μετρήσεων εἶναι προφανής. Ὁ νοῦς τοῦ ἀνθρώπου μπορεῖ νά συγκρατήσει μόνο ἕνα ὡρισμένο ἀριθμό στοιχείων, παραστάσεων ή γνώσεων. Μία από τίς λειτουργίες τοῦ ἐγκεφάλου εἶναι ἡ συστηματική ταξινόμηση τῶν γνώσεων. Σήμερα μέ τήν πρόοδο τῆς τεχνολογίας ὁ ἀριθμός τῶν ἀριθμητικῶν δεδομένων τά ὁποῖα συλλέγονται γιά ἕνα ἐρευνητικό σκοπό μπορεῖ νά εἶναι τεράστιος. Γίνεται λοιπόν ἀναγκαία ἡ σύμπτυξη ή περιληπτική παρουσίαση τῶν ἀριθμητικῶν δεδομένων. Τά ἀριθμητικά δεδομένα μποροῦν νά συμπτυχθοῦν μέ τρεῖς τρόπους: α) Σ τ α τ ι σ τ ι κ ο ί π ί ν α κ ε ς (Π ί ν α κ ε ς σ υ χ ν ο τ ή τ ω ν), β) Γ ρ α φ ι κ έ ς π α ρ α σ τ ά σ ε ι ς (ἱ σ τ ο γ ρ ά μ μ α τ α) καί γ) Ἀ ρ ι θ μ η τ ι κ ά μ ε γ έ θ η. Ἱστορικά οἱ κλάδοι τῆς Στατιστικῆς καί τῆς Ἀναλύσεως τῶν Μετρήσεων ξεκίνησαν μέ τούς δύο πρώτους τρόπους. Οἱ σύγχρονες θεωρίες χρησιμοποιοῦν τόν τρίτο τρόπο σχεδόν ἀποκλειστικά. Παρά τό γεγονός αὐτό καί τῶν τριῶν τρόπων εἶναι ἐπιβεβλημένη ἡ χρήση γιατί ὁ κάθε ἕνας παρουσιάζει πλεονεκτήματα καί ἀποκαλύπτει πτυχές τοῦ πληθυσμοῦ οἱ ὁποῖες ἐνδεχομένως νά ἀποκρύπτονται ἀπό τούς ἄλλους.

7.4.1 Π ί ν α κ ε ς σ υ χ ν ο τ ή τ ω ν

Ὑπάρχουν διάφοροι πίνακες συχνοτήτων. Ἀπό αὐτούς ἀναφέρουμε τούς ἑξῆς:

α) Ἁπλός (μή διατεταγμένος) πίνακας συχνοτήτων: Οἱ ἀρχικές μετρήσεις $x_1, x_2, \ldots, x_n$ μποροῦν νά συμπτυχθοῦν μέ τήν βοήθεια τῶν συχνοτήτων ὡς ἑξῆς:

Μέτρηση	x_1	x_2	...	x_k
Συχνότης	f_1	f_2	...	f_k ,

ὅπου ὅλες οἱ τιμές $x_1, x_2, \ldots, x_k$ εἶναι διάφορες μεταξύ τους καί k ὁ ἀριθμός τῶν μετρήσεων αὐτῶν.

β) Διατεταγμένος πίνακας συχνοτήτων: Ἐάν τά x_1, $x_2, \ldots, x_k$ διαταχθοῦν κατά σειρά μεγέθους προκύπτει ὁ πίνακας:

Μέτρηση	Συχνότης	Ἀθροιστική συχνότης
$x_{(1)}$	f_1	$F_1=f_1$
$x_{(2)}$	f_2	$F_2=f_1+f_2$
. . . .	. . . .	. . . .
$x_{(i)}$	f_i	$F_i=f_1+\ldots+f_i$
. . . .	. . . .	. . . .
$x_{(k)}$	f_k	$F_k=n$
Σύνολο	n	

Πίνακας 7.2 Διατεταγμένος πίνακας συχνοτήτων ὅπου
$x_{(1)} < x_{(2)} < \ldots < x_{(k)}$

Προφανῶς καμμία ἀπώλεια μετρήσεων δέν γίνεται μέ τούς δύο παραπάνω τρόπους. Παρά ταῦτα ἐάν ὁ ἀριθμός k εἶναι μεγάλος τέτοιες συμπτύξεις εἶναι μικρῆς σημασίας, διότι εἶναι ἰσοδύναμοι μέ τά ἀρχικά δεδομένα.

Τά παραδείγματα τοῦ ἐδαφίου 7.3 εἶναι διατεταγμένοι πίνακες συχνοτήτων.

γ) Ὁμαδοποιημένες μετρήσεις-Πίνακες συχνοτήτων: Μία ἀπό τίς πλέον συνήθεις μεθόδους συμπτύξεως μετρήσεων εἶναι ἡ ὁμαδοποίηση αὐτῶν. Ὁ ἀριθμός k τῶν ομάδων, συνήθως *5* ώς *30*, εκλέγεται αυθαίρετα καί τό μήκος κάθε κατηγορίας ή ομάδος καθορίζεται από τό *k* καί τό εὖρος (range), δηλ. τήν διαφορά μεταξύ μεγίστης καί ελαχίστης μετρήσεως. Καταβάλλεται φροντίδα ὥστε τά ὅρια τῶν ομάδων νά εἶναι διάφορα τῶν μετρήσεων. Αὐτό εξασφαλίζεται καθορίζοντας τά ὅρια τῶν ομάδων μέ περισσότερα σημαντικά ψηφία από εκεῖνα τῶν μετρήσεων. Ἔτσι ἄν οἱ μετρήσεις εἶναι *1,2*, *1,8*, *10,6*, *4,9* κ.λ.π. δηλαδή ἔχουν ἀκρίβεια πρώτου δεκαδικοῦ ψηφίου (γίνεται αὐτομάτως στρογγύλεμα εἰς τό πλησιέστερο πρῶτο δεκαδικό ψηφίο π.χ. τό *1,23* καταχωρεῖται ὡς *1,2* καί *1,27* ὡς *1,3*) τότε τά ὅρια τῶν ομάδων μ π ο ρ ε ῖ νά εἶναι *0,25-0,75*, *0,75 -1,25*, *1,25-1,75* κ.λ.π. Συνήθως τά μήκη τῶν ομάδων εκλέγονται ἴ σ α μποροῦν ὅμως νά εἶναι καί ἄνισα. Κάθε ομάδα ἀντιπροσωπεύεται από τό μέσο σημεῖο της πού λέγεται σ η μ ε ῖ ο ἤ τ ι μ ή (class mark) τῆς ομάδος, τήν συχνότητα τῆς ομάδος καί τήν ἀθροιστική της συχνότητα. Σ υ χ ν ό τ η ς ὁ μ ά δ ο ς εἶναι ὁ ἀριθμός τῶν μετρήσεων πού πέφτουν στήν ομάδα. Προφανῶς ή σύμπτυξη αὐτή συνεπάγεται απώλεια μετρήσεων. Τέτοια ομοδοποίηση ὅμως δέν παύει νά ἔχει πρακτική χρησιμότητα, δεδομένου ὅτι ἀποτελεῖ προϋπόθεση κατασκευῆς ἱστογραμμάτων καί πολυγώνων, τά οποῖα κατά προσέγγιση απεικονίζουν τήν κατανομή τοῦ πληθυσμοῦ. Ὁ συνήθης πίνακας ομαδοποιημένων μετρήσεων εἶναι τῆς μορφῆς:

Ὁμάδα	Ὅρια	Τιμή ὁμάδος	Συχνότης	Ἀθρ.συχνότης
1	L_1U_1	X_1	f_1	$F_1 = f_1$
2	L_2U_2	X_2	f_2	$F_2 = f_1 + f_2$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
i	L_iU_i	X_i	f_i	$F_i = f_1 + f_2 + \ldots + f_i$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
k	L_kU_k	X_k	f_k	n
Σύνολο			n	

Πίνακας 7.3 Ὁμαδοποιημένες μετρήσεις.

ὅπου L_i, U_i εἶναι τό κατώτερο καί τό ἀνώτερο ὅριο ἀντίστοιχα τῆς ὁμάδας i καί X_i τό σημεῖο ἤ ἡ τιμή αὐτῆς. Γιά λόγους συνεχείας παίρνουμε $U_i = L_{i+1}$, $i = 1, 2, \ldots, k-1$.

Ὁ πίνακας τῶν ὁμαδοποιημένων μετρήσεων μᾶς δίνει τήν λεγόμενη κατανομή συχνοτήτων τῶν μετρήσεων.

Μερικές φορές τά δεδομένα περιέχουν ἀκρότατες μετρήσεις, δηλαδή μετρήσεις πού εἶναι ἤ πολύ μικρότερες ἤ πολύ μεγαλύτερες ἀπό τίς ἄλλες. Τότε ἄν κάνουμε ὁμαδοποίηση μέ τόν συνήθη τρόπο θά προκύψουν ὁμάδες μέ μηδενική συχνότητα. Τό πρόβλημα αὐτό ἀποφεύγεται μέ δύο τρόπους: εἴτε διατηρώντας τόν ἴδιο ἀριθμό ὁμάδων καί χρησιμοποιοῦντες "ἀνοικτά", δηλαδή ἀπείρου μήκους διαστήματα, γιά τήν πρώτη ἤ τήν τελευταία ὁμάδα, εἴτε μεταβάλλοντες τόν ἀριθμό τῶν ὁμάδων. Ἔτσι ἔχουμε μέν καλύτερη εἰκόνα τῆς κατανομῆς τῶν μετρήσεων χάνουμε ὅμως τίς πραγματικές τιμές τῶν ἀκρότατων μετρήσεων. Ἐπί πλέον προκύπτουν δυσκολίες στόν ὑπολογισμό τῶν ἀριθμητικῶν χαρακτηριστικῶν τοῦ δείγματος.

<u>Παράδειγμα</u> <u>7.4:</u>

Ἑξήντα ἄτομα λαμβάνουν μέρος σέ μία κλινική δοκιμή. Οἱ μετρήσεις χοληστερίνης καί αἵματος μέ προσέγγιση μονάδος ἔδωσαν τά ἀκόλουθα ἀποτελέσματα:

239	212	249	227	218	310	281	330	226	233
223	161	195	233	249	284	284	174	170	256
169	299	210	301	199	258	258	195	227	244
355	234	195	196	354	282	282	286	286	176
195	163	297	211	228	309	309	225	223	195
248	284	173	256	169	209	209	200	258	284

Ὁ διατεταγμένος πίνακας συχνοτήτων εἶναι τῆς μορφῆς:

x_i	f_i
161	1
163	2
169	2
. . . .	. . . .
355	1
Σύνολο	60

Μεγαλύτερη σύμπτυξη τῶν δεδομένων ἐπιτυγχάνεται μέ ὁμαδοποίηση. Ἔτσι διαλέγοντας ὁμαδοποίηση μέ *8* ὁμάδες καί παρατηρώντας ὅτι ἡ μεγίστη μεῖον ἡ ἐλαχίστη τιμή εἶναι *355-161 = 194*, διαιροῦμε τό *194/8 = 24,25*. Ἄρα παίρνουμε γιά μῆκος τῆς κάθε ὁμάδος *25* μονάδες μέ ὅρια ὁμάδων τά *160,5-185,5*, *185,5-210,5* κ.λ.π.

Οἱ ὁμάδες, τά ὅρια τους, οἱ τιμές τῶν ὁμάδων, τό μέτρημα τῶν συχνοτήτων καί οἱ συχνότητες παρουσιάζονται στόν παρακάτω πίνακα.

Ὁμάδα	Ὅρια	Τιμή Ὁμάδος	Συχνότητες		Ἀθροιστ. συχνότ.	Σχετικές συχνότητ.
1	160,5-185,5	173	卌 |||	= 8	8	0,1333
2	185,5-210,5	198	卌 卌 |	=11	19	0,1833
3	210,5-235,5	223	卌 卌 |||	=13	32	0,2167
4	235,5-260,5	248	卌 卌 |	=10	42	0,1667
5	260,5-285,5	273	卌 |	= 6	48	0,1000
6	285,5-310,5	298	卌 ||	= 7	55	0,1167
7	310,5-335,5	323	|||	= 3	58	0,0500
8	335,5-360,5	348	||	= 2	60	0,0333
		Σύνολο		60		1,000

Πίνακας 7.4 Ὁμαδοποιημένες μετρήσεις χοληστερίνης.

Ἄν μία ἀπό τίς παραπάνω μετρήσεις ἦταν *450* τότε θά προέκυπταν ὁμάδες μηδενικῆς συχνότητας. Στήν περίπτωση αὐτή ἡ τελευταία ὁμάδα θά μποροῦσε νά γίνει "ἀνοικτή" δηλαδή *335.5* καί πάνω.

δ) Άλλοι στατιστικοί πίνακες: Γενικά πίνακες πού περιέχουν διάφορα άριθμητικά στοιχεῖα λέγονται στατιστικοί πίνακες.

Παραδείγματα: (1) πίνακες μέ τό έμβαδόν ἤ τό μῆκος τῶν ἀκτῶν διαφόρων γεωγραφικῶν περιοχῶν μιᾶς χώρας.

(2) πίνακες μέ τήν ἀνά μήνα μεγίστη, ἐλαχίστη καί μέση θερμοκρασία τοῦ ἀέρος σέ διάφορες πόλεις.

(3) πίνακες μέ τόν ἀνά μήνα ἀριθμό ἡμερῶν βροχῆς γιά διάφορες πόλεις.

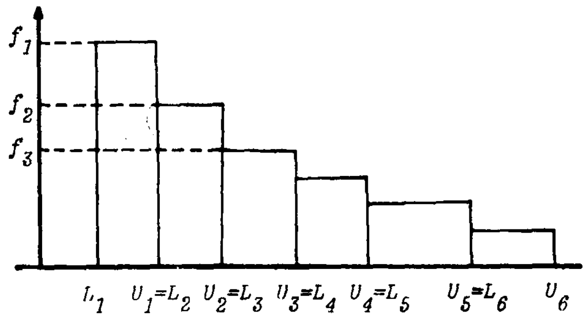
(4) πίνακες θνησιμότητος ἀνά ἡλικία καί αἰτία θανάτου γιά ἕνα ὁρισμένο ἔτος.

(5) πίνακες ἀνεργίας κατά ἡλικία, φύλλο, περιοχή κ.λ.π.

Ὁ ἀναγνώστης παραπέμπεται γιά περισσότερη ἐμπειρία στίς ἐπετηρίδες καί στατιστικές ἐκδόσεις τῆς Ἐθνικῆς Στατιστικῆς Ὑπηρεσίας τῆς Ἑλλάδος.

7.4.2 Γραφικές μέθοδοι συμπτύξεως

α) Ἱστογράμματα (Histogram): Τό ἱστόγραμμα εἶναι ὁ πλέον γνωστός γραφικός τρόπος συμπτύξεως καί παρουσιάσεως τῶν μετρήσεων. Τό διάγραμμα αὐτό γίνεται σ'ἕνα σύστημα ἀξόνων x καί y καί ἀποτελεῖται ἀπό ἕνα σύνολο ὀρθογωνίων τοποθετημένων στόν ἄξονα τῶν X, ὅπως φαίνεται στό παρακάτω σχῆμα. Ἡ βάση τῶν ὀρθογωνίων παριστάνει τό μῆκος τῶν ὁμάδων τῶν μετρήσεων καί τό ὕψος τήν συχνότητα τῶν ὁμάδων (ἤ ποσότητα ἀνάλογο τῆς συχνότητας).

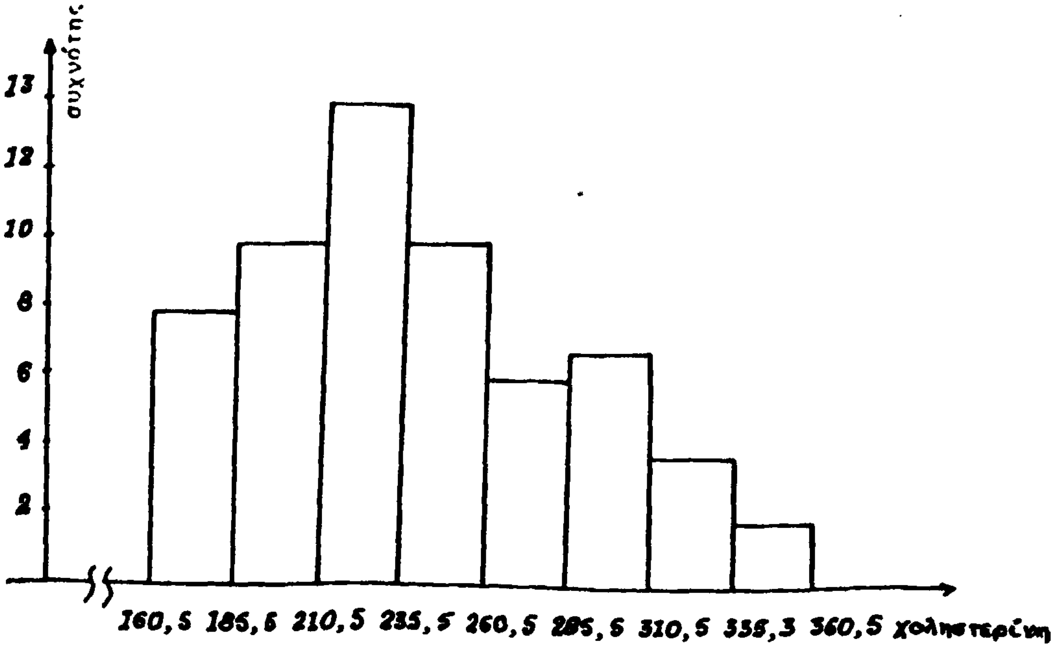


Σχῆμα 7.1 'Ιστογράμματα

Τό βασικό κριτήριο γιά τήν κατασκευή ἱστογράμματος εἶναι ὅτι τό ποσοστό τοῦ ἱστογραμματικοῦ ἐμβαδοῦ κάθε ὁμάδος πρέπει νά εἶναι ἴσο πρός τήν σχετική συχνότητα τῆς ὁμάδος. Ἔτσι ὁ σχεδιασμός τοῦ ἱστογράμματος γίνεται θέτοντας στόν ἄξονα τῶν y τίς τιμές f_i/d_i, $i=1,2,\dots,k$, ὅπου $d_i=U_i-L_i$. 'Εάν οἱ ὁμάδες εἶναι ἴσες, δηλαδή ἔχουν τό ἴδιο μῆκος ($d_1=d_2=\dots=d_k$) τότε τό d μπορεῖ νά ἀπορροφηθεῖ στήν κλίμακα τοῦ ἄξονα τῶν y. Ἔτσι οἱ τιμές στό y θά παριστάνουν τίς ἁπλές συχνότητες f_i. 'Εάν στόν ἄξονα τῶν y θέσουμε τίς τιμές f_i/nd_i τότε τό συνολικό ἐμβαδόν τοῦ ἱστογράμματος θά ἰσοῦται μέ τήν μονάδα.

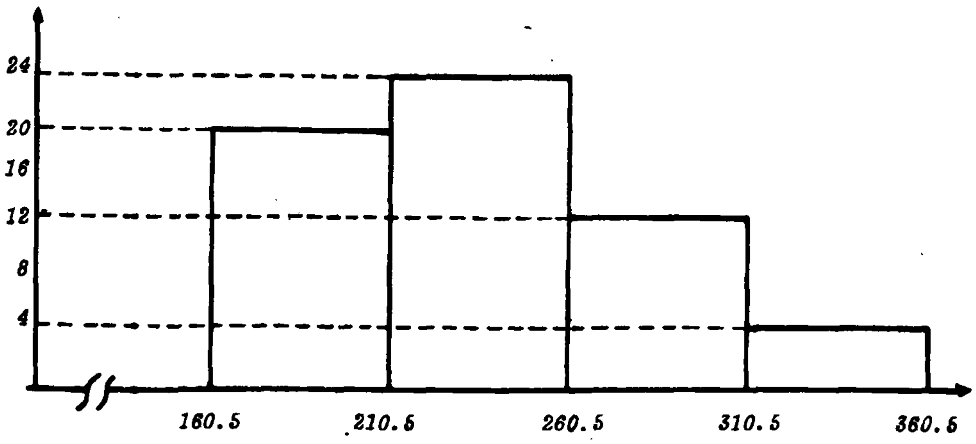
Τό ἱστόγραμμα κατασκευάζεται βάσει τοῦ πίνακος συχνοτήτων ὁμαδοποιημένων μετρήσεων καί μέ τή συσχέτιση ἐμβαδοῦ καί συχνότητος (ἤ καί σχετικῆς συχνότητος) ἐπιτυγχάνεται μία γραφική καί όπτική άξιολόγηση τῆς κατανομῆς τῶν μετρήσεων. Ἄν ἡ πρώτη ἤ ἡ τελευταία ὁμάδα εἶναι "ἀνοικτή" παραλείπεται στήν κατασκευή τοῦ ἱστογράμματος.

Τό ἱστόγραμμα τῶν δεδομένων τοῦ Παραδείγματος 7.4 (βλ. Πίνακα 7.4) εἶναι:



Σχῆμα 7.4 Ἱστόγραμμα χοληστερίνης μέ μῆκος ὁμάδων 25

Όταν τό μῆκος τῶν ὁμάδων εἶναι *50* καί *12* τότε προκύπτουν τά ἀκόλουθα ἱστογράμματα.

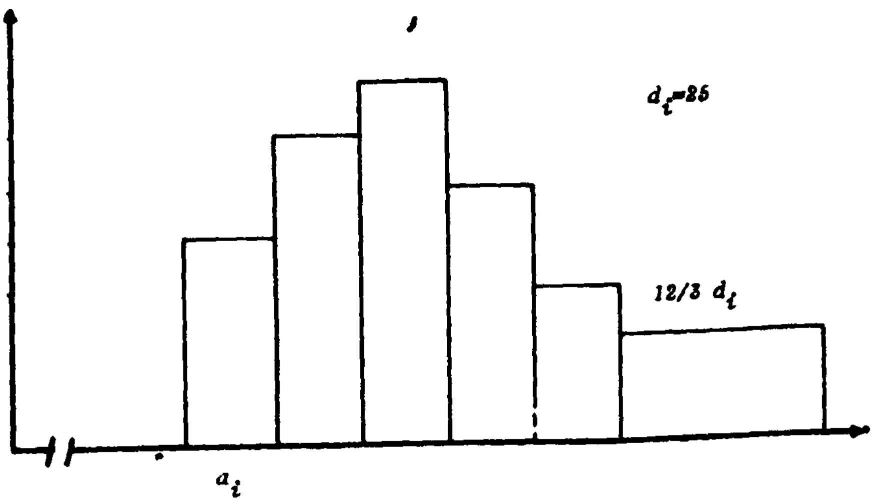


Σχῆμα 7.4 Ἱστόγραμμα χοληστερίνης μέ μῆκος ὁμάδων *50*

Σχῆμα 7.5 Ἱστόγραμμα χοληστερίνης μέ μῆκος ὁμάδων *12*

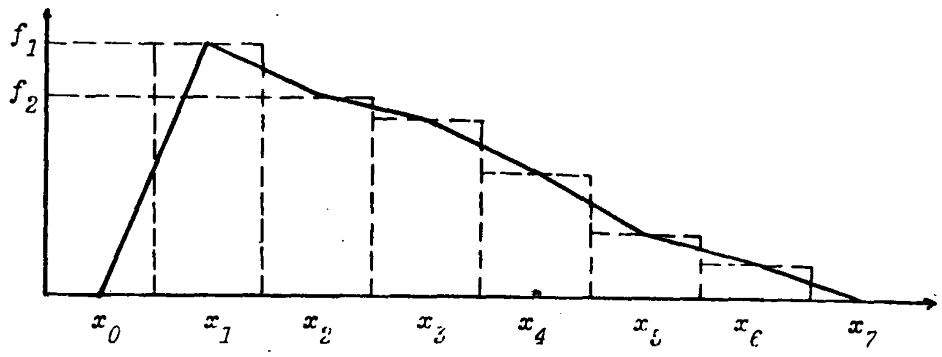
Τό μῆκος τῶν ὁμάδων ἐπηρρεάζει τήν εἰκόνα τοῦ ἱστογράμματος. Ὅσο πιό μεγάλο εἶναι τό μῆκος τῶν ὁμάδων τόσο πιό ἀσαφές καί ἀκαθόριστο εἶναι τό σχῆμα τῆς κατανομῆς τῶν μετρήσεων (βλ. σχῆμα 7.4). Ἄν οἱ ὁμάδες εἶναι πολλές καί ἔχουν μικρό μῆκος τό ἱστόγραμμα παρουσιάζει ἀνωμαλίες πού ἀντανακλοῦν τήν μεταβλητικότητα τῆς δειγματοληψίας (βλ.σχῆμα 7.5)

Ἄν ἐνοποιήσουμε τίς τρεῖς τελευταῖες ὁμάδες τοῦ πίνακα 7.4 προκύπτει τό ἱστόγραμμα μέ ἄνισες ὁμάδες τοῦ Σχήματος 7.6.



Σχῆμα 7.6 'Ιστόγραμμα χοληστερίνης μέ ἄνισες ὁμάδες

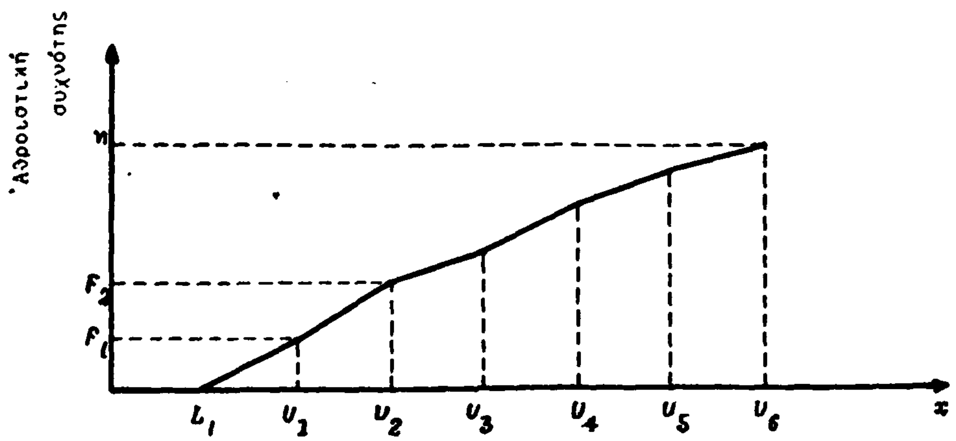
β) Πολύγωνο συχνοτήτων: Τό πολύγωνο συχνοτήτων ἀποτελεῖ δεύτερο τρόπο γραφικῆς συμπτύξεως καί παρουσιάσεως τῶν μετρήσεων. Κατασκευάζεται ἑνώνοντας μέ εὐθεῖες γραμμές τά μέσα τῶν ἄνω πλευρῶν τῶν ὀρθογωνίων τοῦ ἱστογράμματος, ὅπως φαίνεται στό Σχῆμα 7.7.



Σχῆμα 7.7 Πολύγωνο συχνοτήτων

Ἀπό πλευρᾶς χρησιμότητος τό πολύγωνο συχνοτήτων εἶναι ἰσοδύναμο μέ τό ἱστόγραμμα.

γ) Πολύγωνο ἀθροιστικῶν συχνοτήτων: Τό πολύγωνο αὐτό κατασκευάζεται χρησιμοποιοῦντες τίς ἀθροιστικές συχνότητες καί ἑνώνοντας μέ εὐθεῖες γραμμές τά ἀκόλουθα σημεῖα τοῦ (x,y) ἐπιπέδου: $(L_1,0)$, (U_1,F_1), (U_2,F_2), $\ldots,(U_k,n)$ ἤ ἰσοδύναμα $(L_1,0)$, $(U_1,F_1/n)$, $(U_2,F_2/n),\ldots,$ $(U_k,1)$. Χρησιμοποιοῦντες τό ἱστόγραμμα τοῦ Σχήματος 7.1 καί τό $1/n$ τῆς κλίμακος ἔχουμε:



Σχῆμα 7.8 Πολύγωνο ἀθροιστικῶν συχνοτήτων.

Ὅπως καί στήν περίπτωση πινάκων συχνότητας ἔτσι καί τώρα ὁ κάθε ἕνας ἀπό τούς γραφικούς τρόπους συμπτύξεως παρουσιάζει ἰδιαίτερα πλεονεκτήματα. Συνεπῶς εἶναι ἀδύνατη ἡ ὑπόδειξη τοῦ πλέον ἐνδεδειγμένου τρόπου γραφικῆς παρουσιάσεως τῶν μετρήσεων. Ὁ σκοπός καί ἡ προβλεπόμενη χρήση τῆς γραφικῆς παραστάσεως ἀποτελοῦν τό καλύτερο κριτήριο ἐπιλογῆς. Ἐάν ἐπί παραδείγματι σκοπός τῆς συμπτύξεως εἶναι ἡ εὕρεση καί ἀξιολόγηση τῆς συγκριτικῆς θέσεως μελλοντικῶν μετρήσεων τότε τό πολύγωνο ἀθροιστικῶν συναρτήσεων. εἶναι ἡ πλέον ἐνδεδειγμένη γραφική σύμπτυξη.

Ἐκτός ἀπό τούς παραπάνω γραφικούς τρόπους παρουσιάσεως τῶν δεδομένων ὑπάρχουν καί ἄλλοι, ὅπως τά ραβδογράμματα, ἁπλά ἤ σύνθετα, τά είδοδιαγράμματα, τά κυκλικά διαγράμματα, οἱ κύλινδροι κ.λ.π. Ὅλα αὐτά ἀποτελοῦν παραλλαγές τῶν παραπάνω τριῶν γραφικῶν τρόπων. Ὁ ἀναγνώστης παραπέμπεται στή σχετική βιβλιογραφία.

7.4.3 Ἀριθμητικά μεγέθη-Χαρακτηριστικά τῶν μετρήσεων.

Οἱ ἀριθμητικές μέθοδοι περιγραφῆς καί συμπτύξεως τῶν μετρήσεων χρησιμοποιοῦν ἀριθμητικές ποσότητες ἤ χαρακτηριστικά τά ὁποῖα κατά κάποιο τρόπο περιγράφουν τίς μετρήσεις. Ὁ βασικός σκοπός εἶναι ἡ ἐξακρίβωση τῆς κατανομῆς τοῦ πληθυσμοῦ καί τά χαρακτηριστικά αὐτά ἀποσκοποῦν στήν ἀπόδοση λεπτομερειῶν τῆς κατανομῆς ὅπως ἡ θέση, ἡ ἔκταση (εὖρος), ἡ συμμετρία, ἡ λοξότης, ἡ κύρτωσις , ἡ αἰχμή κ.λ.π. τῆς κατανομῆς. Διακρίνουμε τρεῖς κατηγορίες περιγραφικῶν μέτρων:

α) **Μέτρα θέσεως:** Συχνά τά ἀριθμητικά δεδομένα παρουσιάζουν τό χαρακτηριστικό γνώρισμα νά συγκεντρώνονται γύρω ἀπό κάποια "κεντρική" τιμή. Αὐτό τό φαινόμενο εἶναι γνωστό σάν κεντρική τάση τῶν μετρήσεων. Ἡ κεντρική τιμή καθορίζει τήν θέση τοῦ κέντρου ἤ μέσου τῆς κατανομῆς τῶν μετρήσεων καί ἀνάλογα μέ τό τί ἐννοοῦμε μέ τόν ὅρο κέντρο προκύπτουν τά διάφορα μέτρα θέσεως. Τά μέτρα αὐτά εἶναι ἀντίστοιχα τῶν μέτρων θέσεως τῆς (θεωρητικῆς) κατανομῆς τοῦ πληθυσμοῦ καί εἶναι ἡ μέση τιμή, ἡ διάμεσος, ἡ κορυφή κ.λ.π.

Ὁ ἀ ρ ι θ μ η τ ι κ ό ς μ έ σ ο ς ὅ ρ ο ς τῶν μετρήσεων ἤ ἡ δ ε ι γ μ α τ ι κ ή μ έ σ η τ ι μ ή ἤ ἡ μ έ σ η τ ι μ ή τ ο ῦ δ ε ί γ μ α τ ο ς $X_1,\ldots,X_n$ συμβολίζεται μέ $\bar{X}$ καί ὁρίζεται ἀπό τή σχέση

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} .$$

Μετρεῖται μέ τίς ἴδιες μονάδες ὅπως καί οἱ ἀρχικές μετρήσεις (cm,gr, δραχμές κ.λ.π.).

Στήν περίπτωση μετρήσεων μέ συχνότητες ἤ ὁμαδοποιημένων μετρήσεων ἡ μέση τιμή δίδεται ἀπό τήν σχέση

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i X_i}{n} ,$$

ὅπου X_i ἡ τιμή τῆς ὁμάδος i καί f_i ἡ συχνότητα της.

Γιά τά δεδομένα τοῦ Παραδείγματος 7.4 ἔχουμε

$$\overline{X} = \frac{239+212+\ldots+284}{60} = \frac{14427}{60} = 240,45 .$$

Ἄν ὅμως χρησιμοποιήσουμε τήν ὁμαδοποίηση τοῦ Πίνακα 7.4 ἔχουμε

$$\overline{X} = \frac{173\times8+198\times11+\ldots+348\times2}{60} = \frac{1384+2178+\ldots+696}{60} =$$

$$= \frac{14430}{60} = 238,85 .$$

Ἡ διαφορά ἀπό τήν προηγούμενη μέση τιμή ὀφείλεται σέ ἀπώλεια πληροφορίας λόγω ὁμαδοποιήσεως.

Ἡ μέση τιμή τοῦ δείγματος μέ βάρη $w_1,\ldots,w_n$ δίδεται ἀπό τή σχέση

$$\overline{X}_w = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i} \quad . \qquad (7.5)$$

Τά βάρη ἐκφράζουν τήν σημασία πού ἀποδίδουμε σέ κάθε τιμή.

Ἄν ἔχουμε τρία δείγματα μέ μεγέθη n_1, n_2 καί n_3 καί μέσες τιμές $\overline{X}_1$, $\overline{X}_2$, $\overline{X}_3$ ἀντίστοιχα, ἡ μέση τιμή $\overline{X}$ ὅλων τῶν μετρήσεων εἶναι ἡ μέση τιμή τῶν $\overline{X}_1$, $\overline{X}_2$, $\overline{X}_3$ μέ βάρη n_1, n_2, n_3, δηλαδή

$$\overline{X} = \overline{X}_w = \frac{n_1\overline{X}_1 + n_2\overline{X}_2 + n_3\overline{X}_3}{n_1+n_2+n_3} \quad .$$

Ἡ τιμή αὐτή εἶναι διαφορετική ἀπό τήν $(\overline{X}_1+\overline{X}_2+\overline{X}_3)/3$ ἐκτός ἐάν $\overline{X}_1=\overline{X}_2=\overline{X}_3$ ἤ $n_1=n_2=n_3$.

Ἡ **διάμεσος τῶν μετρήσεων** ἤ ἡ **δειγματική διάμεσος** ἤ ἡ **διάμεσος τοῦ δείγματος** εἶναι ἡ μέση μέτρηση στήν κατά σειρά μεγέθους διάταξή τους.

Ἔτσι ἄν $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ εἶναι ἕνα διατεταγμένο δεῖγμα καί n περιττό μέ διακεκριμμένη τή μέτρηση $X_{((n+1)/2)}$ (δηλαδή μέ συχνότητα ἴση μέ ἕνα) τότε ἡ διάμεσος τοῦ δείγματος εἶναι ἡ $X_{((n+1)/2)}$ μέτρηση. Ἄν τό n εἶναι ἄρτιο καί οἱ μετρήσεις $X_{(n/2)}$, $X_{(n/2+1)}$ ἔχουν συχνότητα ἕνα, τότε ἡ δειγματική διάμεσος εἶναι ὁποιοδήποτε σημεῖο μεταξύ τῶν μετρήσεων $X_{(n/2)}$ καί $X_{(n/2+1)}$.

Συνήθως παίρνουμε τό σημεῖο

$$M = \tfrac{1}{2}\,[X_{(n/2)} + X_{(n/2+1)}] \;. \qquad (7.6)$$

Ἄν οἱ παραπάνω προϋποθέσεις δέν πληροῦνται τότε ἡ διάμεσος δέν ὑπάρχει. Π.χ. γιά τά δεδομένα τοῦ Παραδείγματος 7.2 διάμεσος δέν ὑπάρχει.

Ἄν οἱ μετρήσεις εἶναι ὁμαδοποιημένες (βλ. Πίνακα 7.3) τότε ὡς διάμεσος τοῦ δείγματος λαμβάνεται τό σημεῖο M τοῦ ἄξονα τῶν x μέ τήν ἰδιότητα ἡ ἀπό τό M παράλληλος πρός τόν ἄξονα τῶν y νά χωρίζει τό ἐμβαδόν τοῦ ἱστογράμματος σέ δύο ἴσα μέρη. Ὁ ὁρισμός αὐτός τῆς διαμέσου ὑποθέτει ὅτι οἱ μετρήσεις εἶναι ὁμοιόμορφα κατανεμημένες σέ κάθε ὁμάδα. Μέ τόν συμβολισμό τῆς §7.1.4γ καί ὑπό τήν προϋπόθεση ὅτι τά μήκη τῶν ὁμάδων εἶναι ἴσα πρός d, ἡ διάμεσος M εὑρίσκεται ἐντός τῆς ὁμάδος i γιά τήν ὁποία

$$F_{i-1} < n/2 \leq F_i\,, \quad i=1,2,\ldots,k, \quad F_0 = 0$$

καί

$$dF_{i-1} + (M-L_i)f_i = 1/2\;dn \quad \text{ἤ}$$

$$M = L_i + \frac{d}{f_i}(n/2 - F_{i-1}) \;.$$

Γιά τά ὁμαδοποιημένα δεδομένα τοῦ Πίνακα 7.4 ἡ διάμεσος M εὑρίσκεται στήν $i=3^{\eta}$ ὁμάδα ($F_2=19<60/2<F_3 = 32$). Ἔτσι:

$$M = L_3 + \frac{d}{f_3}\,(n/2-F_2) = 210{,}5 + \frac{25}{13}\,(\frac{60}{2} - 19) = 231{,}66 \;.$$

Ἡ κ ο ρ υ φ ή τῶν μετρήσεων ἤ τοῦ δείγματος ἤ ἡ δ ε ι γ μ α τ ι κ ή κ ο ρ υ φ ή εἶναι, ὅπως καί στήν περίπτωση τοῦ πληθυσμοῦ, ἡ τιμή μέ τήν μεγαλύτερη συχνότητα. Ἔτσι στό Παράδειγμα 7.1 τό ὕψος *175cm* εἶναι ἡ κορυφή τοῦ δείγματος. Γιά τά ὁμαδοποιημένα δεδομένα τοῦ Πίνακα 7.4 ἡ 3η ὁμάδα ἔχει τήν μεγαλύτερη συχνότητα καί ὡς κορυφή λαμβάνεται ἡ τιμή τῆς ὁμάδας αὐτῆς *223mgr*. Εἶναι φανερό ὅτι οἱ κατανομές τῶν μετρήσεων μπορεῖ νά ἔχουν μία μόνο κορυφή ἤ καμμία κορυφή ἤ νά εἶναι πολυκόρυφες.

Τό μ έ σ ο ε ὖ ρ ο ς (midrange) τῶν μετρήσεων εἶναι

$$1/2\ (\min X_i + \max X_i).$$

Εἶναι καί αὐτό ἄλλο ἕνα μέτρο θέσεως τῆς κατανομῆς τῶν μετρήσεων, πού προφανῶς ἐπηρεάζεται ἀπό τίς ἄκρες τιμές τοῦ δείγματος.

Ἀπό τά διάφορα μέτρα θέσεως ποιό εἶναι τό καλύτερο ἤ προτιμώτερο; Ἡ ἀπάντηση στό ἐρώτημα αὐτό ἐξαρτᾶται ἀπό τόν σκοπό καί τόν τρόπο χρησιμοποιήσεως τῶν μέτρων αὐτῶν. Ἔτσι χρειάζεται νά ξέρουμε τά πλεονεκτήματα καί τά μειονεκτήματα τῶν διαφόρων μέτρων.

Ἄν ἡ κατανομή τῶν μετρήσεων εἶναι συμμετρική καί μονοκόρυφη, τότε ἡ μέση τιμή $\overline{X}$, ἡ διάμεσος M καί ἡ κορυφή K συμπίπτουν. Ὅσο ἡ κατανομή γίνεται πιό ἀσυμμετρική τόσο οἱ διαφορές μεταξύ τῶν μέτρων αὐτῶν γίνονται μεγαλύτερες.

Ἡ μέση τιμή $\overline{X}$ εἶναι εὐαίσθητη στίς ἀκρότατες τιμές τοῦ δείγματος ἐνῶ τό ἀντίθετο συμβαίνει γιά τή διάμεσο M καί κορυφή K. Ἄν ἡ κατανομή τῶν μετρήσεων εἶναι λοξή τότε ἡ διάμεσος καί ἡ κορυφή ἀποτελοῦν καλύτερα περιγραφικά μέτρα τῆς θέσεως τῆς κατανομῆς τῶν μετρήσεων.

Ἡ μέση τιμή $\overline{X}$ ὑπερέχει τῶν ἄλλων μέτρων θέσεως ὅσον ἀφορᾶ τίς μαθηματικές καί στατιστικές ἰδιότητες. Ἔχει μεγαλύτερη εὐστάθεια (μικρότερη μεταβλητότητα) καί ἐπιδέχεται θεωρητική ἐπεξεργασία πλουσιώτερη καί εὐκολότερη σέ σχέση μέ τά ἄλλα μέτρα. Τά τελευταῖα αὐτά πλεονεκτήματα κάνουν τήν μέση τιμή μοναδική σχεδόν ἐκλογή γιά ὅλους ἐκτός ἀπό τούς καθαρά περιγραφικούς σκοπούς.

β) **Μέτρα μεταβλητότητας (διασπορᾶς ἤ διακυμάνσεως) τῶν μετρήσεων:** Οἱ ἀριθμητικές τιμές τῶν στατιστικῶν δεδομένων εἶναι κατά κανόνα διαφορετικές μεταξύ τους. Παρουσιάζουν δηλαδή τήν λεγομένη διασπορά ἤ διακύμανση ἄλλες φορές σέ μικρό καί ἄλλες φορές σέ μεγάλο βαθμό. Ἡ μεταβλητότητα αὐτή μετρεῖται μέ τά λεγόμενα μέτρα διασπορᾶς ἤ διακυμάνσεως. Τά μέτρα αὐτά μαζί μέ τά μέτρα θέσεως δίνουν μιά πληρέστερη περιγραφή τῶν μετρήσεων, δεδομένου ὅτι μπορεῖ νά ἔχουμε ὁμάδες δεδομένων μέ τήν ἴδια μέση τιμή ἤ διάμεσο ἤ κορυφή ἀλλά τελείως διαφορετικές διακυμάνσεις.

Τό πιό ἁπλό μέτρο διασπορᾶς τῶν μετρήσεων εἶναι τό δειγματικό εὖρος (range) ἤ τό εὖρος τοῦ δείγματος τό ὁποῖο ὁρίζεται ἀπό τήν σχέση

$$R = \max X_i - \min X_i \quad . \tag{7.8}$$

Παρά τήν εὐκολία ὑπολογισμοῦ, τό εὖρος δέν εἶναι καί τόσο ἱκανοποιητικό μέτρο μεταβλητότητας διότι: (i) στηρίζεται μόνο στίς δύο ἄκρες μετρήσεις καί ἀγνοεῖ τίς ἄλλες, (ii) καθώς ὁ ἀριθμός τῶν μετρήσεων μεγαλώνει τό εὖρος τείνει νά γίνει μεγαλύτερο (ἔτσι μπορεῖ νά χρησιμοποιηθεῖ γιά τήν σύγκριση τῆς μεταβλητότητας δύο ὁμάδων δεδομένων μόνο ὅταν οἱ ὁμάδες περιέχουν τόν ἴδιο ἀριθμό μετρήσεων) καί (iii) παρουσιάζει τήν μικρότερη εὐστάθεια (μεγαλύτερη μεταβλητότητα) σέ ἐπαναληπτικές δειγματοληψίες σέ σχέση μέ τά ἄλλα μέτρα διασπορᾶς ἐκτός τῶν περιπτώσεων πού τό μέγεθος τοῦ δείγματος εἶναι πολύ μικρό. Λόγω τῆς εὐκολίας ὑπολογισμοῦ καί τῆς εὐστάθειας του σέ μικρά δείγματα, τό εὖρος χρησιμοποιεῖται συχνά στόν στατιστικό ποιοτικό ἔλεγχο ὅπου τέσσερεις ἤ πέντε μετρήσεις συνήθως ἀρκοῦν.

Ἡ διακυμανση (ἤ διασπορά) τοῦ δείγματος ἤ δειγματική διασπορά δίδεται ἀπό τήν σχέση

$$S'^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 . \qquad (7.9)$$

Ἀναπτύσσοντας τό τετράγωνο καί χρησιμοποιοῦντες ἰδιότητες τοῦ ἀθροίσματος Σ ἔχουμε

$$\Sigma(X_i-\overline{X})^2=\Sigma(X_i^2-2\overline{X}X_i+\overline{X}^2)=\Sigma X_i^2-2\overline{X}\Sigma X_i+n\overline{X}^2$$

$$=\Sigma X_i^2-2\left(\frac{\Sigma X_i}{n}\right)\Sigma X_i+n\left(\frac{\Sigma X_i}{n}\right)^2=\Sigma X_i^2-\frac{(\Sigma X_i)^2}{n} .$$

Ἔτσι

$$S'^2 = \frac{1}{n}\left\{\Sigma X_i^2-\frac{(\Sigma X_i)^2}{n}\right\} . \qquad (7.10)$$

Γιά τόν ὑπολογισμό τοῦ S'^2 ἀρκεῖ νά ὑπολογίσουμε τό ἄθροισμα ΣX_i καί τό ἄθροισμα τετραγώνων ΣX_i^2 τῶν μετρήσεων. Ὁ τύπος (7.10) λέγεται τύπος μηχανῆς.

Γιά λόγους πού θά φανοῦν ὅταν μελετήσουμε τήν ἐκτίμηση τῶν στατιστικῶν παραμέτρων, ὡς διακύμανση τοῦ δείγματος συχνά παίρνεται ἡ ποσότης

$$S^2 = \frac{1}{n-1}\,\Sigma(X_i-\bar{X})^2 \tag{7.11}$$

ἡ ὁποία λέγεται καί *μέση τετραγωνική ἀπόκλιση*. Ὁ τύπος μηχανῆς γιά τήν ἀπόκλιση S^2 εἶναι

$$S^2 = \frac{1}{n-1}\left\{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}\right\}. \tag{7.12}$$

Οἱ διακυμάνσεις S'^2 καί S^2 ἐκφράζουν τήν ἴδια στατιστική ἔννοια:τήν μεταβλητότητα τοῦ δείγματος. Ἐπίσης ἡ διαφορά μεταξύ τῶν τιμῶν τῶν S'^2 καί S^2 εἶναι ἀσήμαντη, ἰδίως ὅταν τό μέγεθος τοῦ δείγματος εἶναι μεγάλο. Δέν παύουν ὅμως νά ἔχουν διαφορετικές στατιστικές ἰδιότητες ὅπως θά δοῦμε πιό κάτω (τό S'^2 τείνει νά ὑποεκτιμᾶ τή διακύμανση σ^2).

Ἡ διακύμανση στηρίζεται στίς ποσότητες $X_1-\bar{X},\ldots,X_n-\bar{X}$ πού λέγονται *ἀποκλίσεις ἀπό τή μέση τιμή*. Ὁ μέσος ὅρος τῶν ἀποκλίσεων αὐτῶν δέν προσφέρεται γιά μέτρο διακυμάνσεως διότι

$$\sum_{i=1}^{n}(X_i-\bar{X}) = \Sigma X_i - n\bar{X} = 0 \tag{7.13}$$

γιά ὁποιοδήποτε σύνολο μετρήσεων $X_1,\ldots,X_n$.

Ἀντί τοῦ μέσου ὅρου τῶν ἀποκλίσεων μποροῦμε νά χρησιμοποιήσουμε τόν μέσο ὅρο τῶν ἀπολύτων ἀποκλίσεων

$$\frac{1}{n} \sum_{i=1}^{n} |X_i - \bar{X}| \; . \qquad (7.14)$$

Τό μέτρο αὐτό λέγεται μέση ἀπόλυτη ἀπόκλιση. Παρουσιάζει περιορισμένο ἐνδιαφέρον λόγω τῶν τεχνικῶν δυσκολιῶν τῆς ἀπολύτου τιμῆς.

Ἡ ποσότης

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 \qquad (7.15)$$

λέγεται ἄθροισμα τετραγώνων τῶν ἀποκλίσεων ἤ λόγω ἁπλότητας ἄθροισμα τετραγώνων καί εἶναι ἴση μέ τό μηδέν ἐάν καί μόνον ἐάν οἱ τιμές τῶν μετρήσεων εἶναι ἴσες μεταξύ τους. Ὁμοίως τό ἄθροισμα τετραγώνων παίρνει μεγάλες τιμές ἐάν καί μόνον ἐάν οἱ τιμές τῶν μετρήσεων διαφέρουν (ἔχουν μεγάλη διασπορά) μεταξύ τους. Τό ἴδιο ἰσχύει καί γιά τίς ποσότητες S'^2 καί S^2.

Τό ἄθροισμα τετραγώνων συνδέεται μέ ἕναν ἀριθμό, ὁ ὁποῖος εἶναι γνωστός σάν βαθμός ἐλευθερίας τοῦ ἀθροίσματος. Ὁ ἀριθμός αὐτός ὁρίζεται ὡς ὁ ἀριθμός τῶν τετραγώνων μεῖον τόν ἀριθμό τῶν ἀνεξαρτήτων γραμμικῶν περιορισμῶν πού ὑπάρχουν ἐπί τῶν ὅρων τοῦ ἀθροίσματος. Ἔτσι τό ἄθροισμα τετραγώνων (7.15) ἔχει $n-1$ βαθμούς ἐλευθερίας διότι ὑπάρχει μόνον ἕνας γραμμικός περιορισμός, ὁ (7.13) μέ τήν βοήθεια τοῦ ὁποίου μπορεῖ νά προσδιορισθεῖ ἡ n-οστή ἀπόκλιση ὅταν δίδονται οἱ ἄλλες $n-1$ ἀποκλίσεις. Οἱ βαθμοί ἐλευθερίας συνοδεύουν καί τήν ποσότητα S^2.

Ἡ τ υ π ι κ ή ἀ π ό κ λ ι σ η τ ο ῦ δ ε ι γ μα-τ ο ς ἤ δ ε ι γ μ α τ ι κ ή τ υ π ι κ ή ἀ π ό κ λ ι σ η εἶναι ἡ τετραγωνική ρίζα τῆς διακυμάνσεως δηλαδή

$$S' = \sqrt{S'^2} \quad \text{ἤ} \quad S = \sqrt{S^2}\,. \tag{7.16}$$

Ἡ τυπική ἀπόκλιση ἐκφράζει τήν μεταβλητότητα τῶν μετρήσεων κατά περισσότερο ἄμεσο τρόπο δεδομένου ὅτι μετρεῖται μέ τίς ἴδιες μονάδες μετρήσεως ὅπως καί τά ἀρχικά δεδομένα σέ ἀντίθεση μέ τήν διακύμανση πού μετρεῖται μέ τό τετράγωνο τῶν μονάδων.

Ὡς σ υ ν τ ε λ ε σ τ ή ς μ ε τ α β λ η τ ό τ η τ ο ς τῶν μετρήσεων ὁρίζεται ἡ ποσότης

$$C.V. = \frac{S}{\bar{X}} \quad \text{ἤ} \quad C.V. = \frac{S}{\bar{X}} \times 100\%$$

(*C.V.= coefficient of variation*), ἡ ὁποία εἶναι καθαρός ἀριθμός καί ἐκφράζει τήν μεταβλητότητα τῶν μετρήσεων ἀπαλλαγμένη ἀπό τήν ἐπίδραση τῆς μέσης τιμῆς.

Ἡ χρησιμότητα τοῦ συντελεστοῦ μεταβλητότητας φαίνεται ἀπό τό ἀκόλουθο παράδειγμα: Ἕνα σύστημα ἤ μέθοδος κατασκευάζει σωλῆνες (π.χ. κάλυκες) μίας ὁρισμένης διαμέτρου. Ἔστω ὅτι ἡ τυπική ἀπόκλιση τῆς διαμέτρου εἶναι *0,05 cm*. Πότε τό σύστημα (μέθοδος) ἔχει περισσότερη ἀκρίβεια ὅταν ἡ μέση διάμετρος εἶναι *10cm* ἤ $\frac{1}{2}$ *cm*;

Ἄς ὑπολογίσουμε τούς συντελεστές μεταβλητότητας:

μέ μέση διάμετρο *10cm* $C.V. = \frac{0,05}{10} \cdot 100\% = 0,5\%$

μέ μέση διάμετρο $0.5cm$ $C.V. = \frac{0.05}{0.5} \cdot 100\% = 10\%$

Ἄρα ὅταν ἡ μέση διάμετρος εἶναι $10cm$ ὁ συντελεστής μεταβλητότητας εἶναι μικρότερος καί τό σύστημα ἔχει μεγαλύτερη ἀκρίβεια.

Παράδειγμα 7.5:

Γιά τά δεδομένα τοῦ Παραδείγματος 7.4 ἔχουμε:

$$\Sigma X_i = 14{,}427 \qquad \Sigma X_i^2 = 3615497$$

$$S'^2 = \frac{1}{60}\left\{ 3615497 - \frac{(14{,}427)^2}{60} \right\} = 2442{,}08$$

$$S' = 49{,}42$$

$$S^2 = \frac{1}{59}\left\{ 3615497 - \frac{(14{,}427)^2}{60} \right\} = 2483{,}42$$

$$S = 49{,}83 \; .$$

Εἶναι φανερό ὅτι ἡ δειγματική διακύμανση καί ἡ τυπική ἀπόκλιση S'^2 καί S' ἀντιστοιχοῦν ἐννοιολογικά μέ τήν διακύμανση καί τήν τυπική ἀπόκλιση σ^2 καί σ τοῦ πληθυσμοῦ. Ἡ στατιστική σημασία τῆς διακυμάνσεως καί κατά συνέπεια τῆς τυπικῆς ἀποκλίσεως καταφαίνεται ἀπό τό Θεώρημα τοῦ *Chebyshev* (βλ. Κεφάλαιο V): γιά ὁποιοδήποτε τυχαῖο δεῖγμα $x_1, \ldots, x_n$ τοὐλάχιστον $(1-1/k^2)$ τῶν μετρήσεων $(k=1,2,\ldots)$ θά εὑρίσκεται σέ ἀπόσταση k τυπικῶν ἀποκλίσεων ἀπό τήν μέση τιμή, δηλαδή ἐντός τοῦ διαστήματος $[(\bar{X}-ks,\ \bar{X}+ks)]$.

Στήν περίπτωση μετρήσεων μέ συχνότητα ἤ ὁμαδοποιημένων μετρήσεων ἡ διακύμανση δίδεται ἀπό τίς σχέσεις

$$S^2=\frac{1}{n-1}\sum_{i=1}^{k} f_i(X_i-\bar{X})^2=\frac{1}{n-1}\left\{\sum_{i=1}^{k} f_iX_i^2-\frac{(\Sigma f_iX_i)^2}{n}\right\}.$$

Παράδειγμα 7.6:

Γιά τά ὁμαδοποιημένα δεδομένα τοῦ Παραδείγματος 7.4 ἔχουμε:

$$\Sigma f_iX_i=14427 \qquad \Sigma f_iX_i^2=3556190$$

$$S^2=\frac{1}{59}\{3556190-3468972{,}15\}=1478{,}27$$

$$S=38{,}45\ .$$

γ) Ροπές τοῦ δείγματος: οἱ ροπές τοῦ δείγματος ὁρίζονται ὅπως οἱ ροπές τοῦ πληθυσμοῦ καί ἔχουν τήν ἴδια ἐννοιολογική σημασία. Ἔτσι ἔχουμε:

(1) δειγματική ροπή k τάξεως πέρί τό μηδέν:

$$m_k=\frac{1}{n}\sum_{i=1}^{n} X_i^k, \qquad k=1,2,\ldots$$

(2) δειγματική κεντρική ροπή k τάξεως:

$$v_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k, \qquad k=1,2,\ldots$$

(3) δειγματική τυπική ἤ κανονικοποιημένη ροπή k τάξεως:

$$a_k=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{S}\right)^k,\ k=1,2,\ldots$$

Ἀπό τίς παραπάνω ροπές πιό χρήσιμες εἶναι ἡ $m_1=\bar{X}$,

$v_2=S'^2$, a_3 ὁ δειγματικός συντελεστής λοξότητας τῆς κατανομῆς τῶν μετρήσεων καί a_4 ὁ δειγματικός συντελεστής κυρτότητος τῆς κατανομῆς. Ἡ ἑρμηνεία τῶν a_3 καί a_4 εἶναι ἡ ἴδια μέ τήν ἑρμηνεία τῶν α_3 καί α_4 ἀντίστοιχα (βλ. § 5.4).

ΑΣΚΗΣΕΙΣ

7.1. Γιά κάθε ὁμάδα μετρήσεων νά βρεθοῦν (α) ἡ μέση τιμή, (β) ἡ διάμεσος, (γ) ἡ κορυφή, (δ) τό εὖρος (ε) ἡ διακύμανση S^2 καί (στ) ἡ τυπική ἀπόκλιση S.

(i) 1, 3, 3, 5, 6, 6

(ii) -4, 2, 0, -6, -4, 6, 2, 0, 4

(iii) 16, 22, 2, 8, 6, 20, 14, 24 .

7.2. Νά δειχθοῦν οἱ σχέσεις

$$v_2 = m_2 - m_1^2$$

$$v_3 = m_3 - 3m_1 m_2 + 2m_1^3$$

$$v_4 = m_4 - 4m_1 m_3 + 6m_1^2 m_2 - 3m_1^4 .$$

7.3. Δίδονται οἱ χρόνοι ζωῆς 60 λυχνιῶν ἑνός ὁρισμένου τύπου. Ὁ χρόνος ζωῆς μετρεῖται σέ ὧρες μέ ἀκρίβεια μετρήσεων τήν πλησιέστερη ὥρα.

365	510	381	354	301	497
321	250	445	462	373	442
331	364	501	286	451	444
364	382	460	404	342	473
271	388	337	438	552	351
342	301	265	309	448	573
455	373	457	399	395	273
332	554	308	443	461	412

270	324	251	299	605	341
539	370	306	541	398	366

(i) Νά γίνει ὁμαδοποίηση τῶν δεδομένων σέ δέκα ὁμάδες.

(ii) Νά κατασκευαστεῖ τό ἱστόγραμμα καί τό πολύγωνο ἀθροιστικῶν συχνοτήτων.

(iii) Γιά τίς ὁμαδοποιημένες μετρήσεις νά βρεθεῖ ἡ μέση τιμή, ἡ διάμεσος καί ἡ τυπική ἀπόκλιση S.

(iv) Νά βρεθοῦν τά τεταρτημόρια καί τό 10το ἑκατοστιαῖο σημεῖο.

7.4. Νά δειχθεῖ ὅτι τό ἄθροισμα τῶν τετραγώνων τῶν ἀποκλίσεων n μετρήσεων ἀπό μία τιμή a γίνεται ἐλάχιστο ἄν καί μόνο ἄν $a=\bar{X}$.

7.5. Νά δειχθεῖ ὅτι

$$E(m_k) = \mu_k$$

7.6. Νά δειχθεῖ ὅτι

(i) $$S^2 = \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} (X_i - X_j)^2$$

(ii) $$Var(S^2) = \frac{1}{n} \left(\lambda_4 - \frac{n-3}{n-1} \sigma^4\right)$$

ὅπου $$\lambda_k = E(X-\mu)^k.$$

ΚΕΦΑΛΑΙΟ 8

στατιστικα-δειγματικες κατανομες

8.1 Δειγματοληψία καί Συμπερασματολογία -Δειγματικές κατανομές

'Η εὕρεση τῶν ἀγνώστων ποσοτικῶν ἤ ποιοτικῶν στοιχείων ἑνός πληθυσμοῦ γίνεται μέ τή βοήθεια δειγμάτων πού ἐκλέγονται ἀπ'αὐτόν. Ἔστω ὅτι στίς δημοτικές ἐκλογές ὁ πληθυσμός μίας πόλης εἶναι διχασμένος σέ ἐκείνους πού ὑποστηρίζουν τόν ὑποψήφιο A καί ἐκείνους πού ὑποστηρίζουν τόν ὑποψήφιο B. Μιά σφυγμομέτρηση τῆς κοινῆς γνώμης ἐπιδιώκει νά προβλέψει τό ἀποτέλεσμα τῆς ἐκλογῆς. Ἔτσι ἐκλέγεται ἕνα τυχαῖο δεῖγμα 100 ψηφοφόρων (συνήθως χωρίς ἐπανατοποθέτηση) καί ἔστω X ὁ ἀριθμός ἐκείνων πού ὑποστηρίζουν τόν ὑποψήφιο A. Ἔστω ἐπίσης p τό ποσοστό τῶν ψηφοφόρων τῆς πόλης πού ὑποστηρίζουν τόν A. Ἄν τό p ἦταν γνωστό δέν θά ὑπῆρχε ἀνάγκη προβλέψεως ἤ δειγματοληψίας. 'Η σφυγμομέτρηση προσπαθεῖ νά "μαντέψει" τό p βάσει τοῦ X. Ἄν π.χ. $X=70$ ξέρουμε, καί χωρίς τή χρήση τῆς Στατιστικῆς 'Ανάλυσης, ὅτι τό p θά εἶναι γύρω στό 70%. Εἶναι ἀπίθανο τό p νά εἶναι

25% καί νά ἔχουμε ἕνα $X=70$. Ἐπίσης ξέρουμε ὅτι ἄν τό μέγεθος τοῦ δείγματος ἦταν 1000 ἀντί 100 θά μπορούσαμε νά κάνουμε πιό ἀκριβέστερη "ἐκτίμηση" τοῦ p. Σκοπός μας εἶναι νά πάμε πιό κάτω ἀπό αὐτές τίς ἁπλές καί ἀρχικές ἰδέες καί νά καταλάβουμε λεπτομερέστερα πῶς γίνεται ἡ "ἐκτίμηση" τοῦ p καί τί εἶδος ἀκρίβειας μποροῦμε νά ἔχουμε.

Τά ἄγνωστα στοιχεῖα τοῦ πληθυσμοῦ συνήθως ἐκφράζονται συναρτήσει ἀριθμῶν πού λέγονται (στατιστικές) π α ρ ά μ ε τ ρ ο ι. Π.χ. τό p τοῦ παραπάνω δείγματος, ὁ μέσος χρόνος μεταξύ δύο διαδοχικῶν ἀφίξεων κλήσεων σ'ἕνα τηλεφωνικό κέντρο, τό μ τῆς κανονικῆς κατανομῆς, ἤ $P(X>1)$ κ.λ.π. Τά συμπεράσματα γιά μιά παράμετρο ἐξάγονται μέ τή βοήθεια τῶν σ τ α τ ι σ τ ι κ ῶ ν (ἤ σ τ α τ ι σ τ ι κ ῶ ν σ υ ν α ρ τ ή σ ε ω ν), τά ὁποῖα εἶναι ἀριθμοί πού μποροῦν νά ὑπολογιστοῦν ἀπό τό δεῖγμα. Π.χ. τό X τοῦ παραπάνω παραδείγματος, τό $\overline{X}$, S^2 κ.λ.π.

Ὁρισμός 8.1:

Μία συνάρτηση τῶν μεταβλητῶν $X_1,X_2,\ldots,X_n$ ἑνός τυχαίου δείγματος πού δέν περιέχει ἄγνωστες παραμέτρους λέγεται σ τ α τ ι σ τ ι κ ό ἤ σ τ α τ ι σ τ ι κ ή σ υ ν ά ρ τ η σ η (σ.σ.).

Τά στατιστικά συμβολίζονται συνήθως μέ $T=T(X_1,\ldots,X_n)$, $S=S(X_1,\ldots,X_n)$ κ.λ.π.

Ἀπό τά προηγούμενα εἶναι φανερό ὅτι τόσον οἱ τιμές x_i ἑνός τυχαίου δείγματος ὅσο καί τά στατιστικά T παίζουν ἕνα διπλό ρόλο: ἀφ'ἑνός μέν εἶναι γνωστοί ἀριθμοί (ἔχουν γνωστές ἀριθμητικές τιμές), ἀφ'ἑτέρου εἶναι καί τυχαῖες μεταβλητές. Στό προηγούμενο παράδειγμα $X=70$. Τό X ἔχει τήν τιμή 70 ἀλλά εἶναι καί μία διωνυμική τ.μ. μέ παραμέτρους (n,p). Οἱ στατιστικές συναρτήσεις ὡς συναρτήσεις τυχαίων μεταβλητῶν εἶναι ἐπίσης τυχαῖες μεταβλητές. Παίρνουν τίς διάφορες τιμές τους ἔπειτα ἀπό ν ο ε ρ έ ς ἐ π α ν α λ ή ψ ε ι ς τ ο ῦ π ε ι ρ ά μ α τ ο ς πού γεννᾶ τίς τιμές τοῦ τυχαίου δείγματος. Συνήθως θά συμβολίζουμε τίς ἀριθμητικές τιμές τοῦ τ. δείγματος ἤ τῶν στατιστικῶν μέ μικρά γράμματα $x_1,x_2,\ldots$ ἤ $y_1,y_2,\ldots,$ $t,\bar{x},s^2$ κ.λ.π. (πρῶτος ρόλος) καί τίς τυχαῖες μεταβλητές μέ κεφαλαῖα γράμματα $X_1,X_2,$ $\ldots,$ $Y_1,Y_2,\ldots,$ $T,\bar{X},S^2$ κ.λ.π. (δεύτερος ρόλος). Οἱ κατανομές πιθανότητας τῶν στατιστικῶν λέγονται δ ε ι γ μ α τ ι κ έ ς κ α τ α ν ο μ έ ς. Ἔτσι ὁ $\bar{X}$, τό S^2, S κ.λ.π. ἔχουν ἀπό μία δειγματική κατανομή τό κάθε ἕνα. Τά στατιστικά συμπεράσματα ἐξάγονται χρησιμοποιῶντας τίς κατανομές αὐτές.

Ἄν ἐνδιαφερόμαστε γιά τό μέσο βάρος μ (σέ κιλά) τοῦ πληθυσμοῦ ψαριῶν στή Λίμνη Ἰωαννίνων τό μ εἶναι μία ἄγνωστη παράμετρος καί προφανῶς θά μείνει γιά πάντα ἄγνωστη. Γιά νά "ἐκτιμήσουμε" τό μ παίρνουμε ἕνα δεῖγμα ἀπό ψάρια καί ὑπολογίζουμε τό μέσο βάρος τους $\bar{x}$. Μέ βάση τό $\bar{x}$ βγάζουμε συμπέρασμα γιά τό μ. Χωρίς στατιστική θεωρία ξέρουμε ὅτι τό $\bar{x}$ "ἐκτιμᾶ" κάπως τό μ, ἀλλά γιά νά μάθουμε πῶς νά βγάζουμε συμπεράσματα καί πόση ἀκρίβεια θά ἔχουν τά συμπεράσματα αὐτά, χρειαζόμαστε νά ξέρουμε τήν κατανομή τοῦ $\bar{X}$, τή δειγματική κατανομή τῆς τ.μ. $\bar{X}$.

Εἶναι φανερό ὅτι ἡ εὕρεση τῶν δειγματικῶν κατανομῶν στηρίζεται στή μεθοδολογία τῆς ἀλλαγῆς μεταβλητῶν. Πρίν προχωρήσουμε στή μελέτη τῶν δειγματικῶν κατανομῶν θά δώσουμε μερικά γενικά ἀποτελέσματα πού ἀφοροῦν τίς μέσες τιμές καί διακυμάνσεις τῶν χαρακτηριστικῶν τοῦ δείγματος καί τά ὁποῖα ἰσχύουν χωρίς συγκεκριμένη ὑπόθεση γιά τό εἶδος τῆς κατανομῆς τοῦ πληθυσμοῦ.

8. 2 Μέσες τιμὲς καὶ διακυμάνσεις $\overline{X}$ καὶ S^2

Ἐφ'ὅσον τά στατιστικά $\overline{X}$, S^2 κ.λ.π. εἶναι τυχαῖες μεταβλητές εἶναι σωστό νά μιλᾶμε γιά τίς μέσες τιμές, διακυμάνσεις τους κ.λ.π.

Θεώρημα 8.1:

Ἔστω $X_1,\ldots,X_n$ ἕνα τυχαῖο δεῖγμα ἀπό ἕνα πληθυσμό μέ μέση τιμή μ καί διακύμανση σ^2. Τότε

(i) $$E(\overline{X}) = \mu \tag{8.1}$$

(ii) $$Var(\overline{X}) = \frac{\sigma^2}{n} \tag{8.2}$$

(iii) $$E(S^2) = \sigma^2 . \tag{8.3}$$

Ἀπόδειξη:

(i) $$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} E(X_i)\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu .$$

$$(ii)\ Var(\overline{X}) = Var(\frac{1}{n}\Sigma X_i) = \frac{1}{n^2}\sum_{i=1}^{n} VarX_i = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}.$$

$$(iii)\ E(S^2) = E[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2]$$

$$= \frac{1}{n-1}E[\sum_{i=1}^{n}(X_i-\mu+\mu-\overline{X})^2]$$

$$= \frac{1}{n-1}E[\sum_{i=1}^{n}(X_i-\mu)^2+\sum_{i=1}^{n}(\overline{X}-\mu)^2-2\sum_{i=1}^{n}(\overline{X}-\mu)(X_i-\mu)]$$

$$= \frac{1}{n-1}E[\sum_{i=1}^{n}(X_i-\mu)^2+n(\overline{X}-\mu)^2-2(\overline{X}-\mu)(n\overline{X}-n\mu)]$$

$$= \frac{1}{n-1}E[\sum_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2]$$

$$= \frac{1}{n-1}[\sum_{i=1}^{n}E(X_i-\mu)^2-nE(\overline{X}-\mu)^2].$$

Βάσει τοῦ *(ii)* ἔχουμε $E(\overline{X}-\mu)^2 = Var(\overline{X})=\sigma^2/n$.
Ἔτσι

$$E(S^2) = \frac{1}{n-1}[\sum_{i=1}^{n}\sigma^2-n\ \sigma^2/n] = \sigma^2.\ \blacktriangledown$$

Πόρισμα 8.1: Ἔστω $X_{11},X_{12},\ldots,X_{1n}$, ἕνα τυχαῖο δεῖγμα μεγέθους n_1 ἀπό ἕνα πληθυσμό μέ μέση τιμή μ_1 καί διακύμανση σ_1^2 καί $X_{21},X_{22},\ldots,X_{2n}$, ἕνα τυχαῖο δεῖγμα μεγέθους n_2 ἀπό ἕνα ἄλλο ἀνεξάρτητο πληθυσμό μέ μέση τιμή μ_2 καί διχύμανση σ_2^2 . Τότε

$$(i) \qquad E(\overline{X}_1-\overline{X}_2) = \mu_1-\mu_2 \qquad (8.4)$$

$$(ii) \qquad Var(\overline{X}_1-\overline{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} , \qquad (8.5)$$

ὅπου $\overline{X}_1 = (\sum_{i=1}^{n_1} X_{1j})/n_1$ καί $\overline{X}_2 = (\sum_{j=1}^{n_2} X_{2j})/n_2$.

8. 3 Κατανομὲς ποὺ ἀπορρέουν ἀπὸ τὴν κανονικὴ

Οἱ μονοδιάστατες κατανομές πού δίδονται στό ἐδάφιο αὐτό ἀπορρέουν ἀπό τήν κανονική κατανομή καί παίζουν ὅπως θά δοῦμε πρωταρχικό ρόλο στή στατιστική συμπερασματολογία. Τό ἴδιο ἰσχύει καί γιά τά ἀντίστροφα ἑκατοστιαῖα σημεῖα αὐτῶν.

Ὁρισμός 8.2 κατανομή X^2

Ἔστω $X_1,\ldots,X_\nu$ ν ἀνεξάρτητες καί ἰσόνομες τ.μ. μέ τυπική κανονική κατανομή. Ἡ κατανομή τῆς τ.μ. $Y = X_1^2+\ldots+X_\nu^2$ εἶναι ἡ X_ν^2 , χί- τ ε τ ρ ά γ ω ν ο μ έ ν β α θ μ ο ύ ς ἐ λ ε υ θ ε ρ ί α ς (β.ε.). Συνολικά

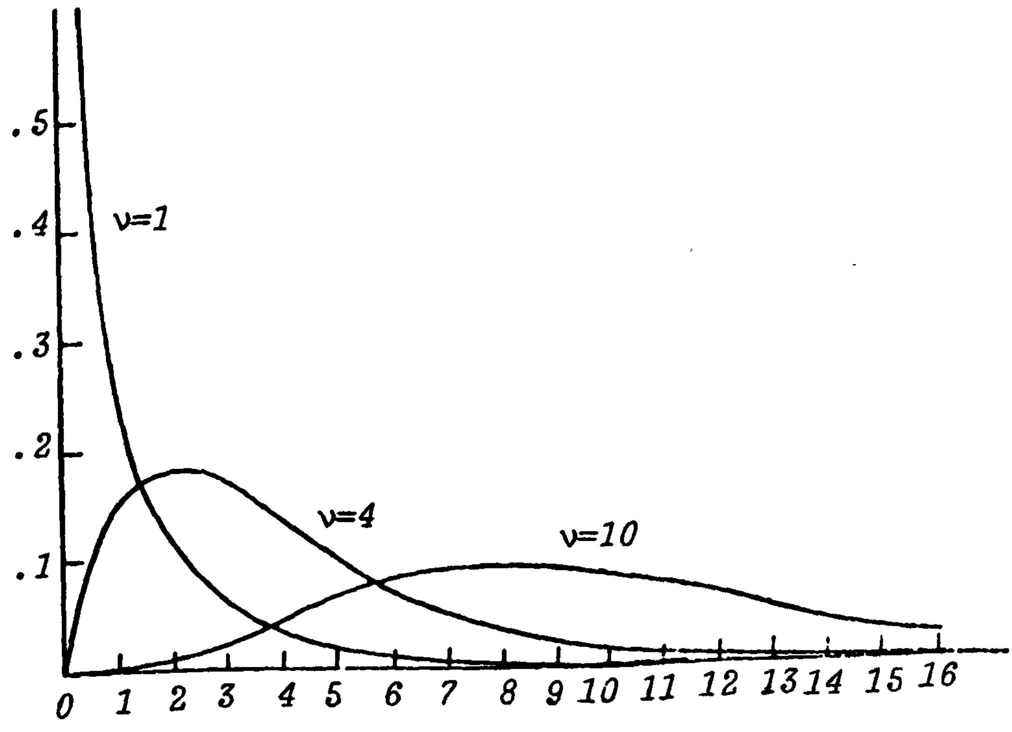
$$X_\nu^2 = \Sigma N(0,1)^2. \qquad (8.6)$$

Ἡ χί τετράγωνο κατανομή μέ ν βαθμούς ἐλευθερίας παρουσιάστηκε καί μελετήθηκε στό ἐδάφιο 4.10 ὡς εἰδική περίπτωση τῆς Γάμμα κατανομῆς μέ α=ν/2 καί β=2.

'Ο 'Ορισμός *8.2* ἀποτελεῖ ἕνα ἰσοδύναμο τρόπο παρουσιάσεως τῆς κατανομῆς X^2_ν. Πράγματι μέ τή θεωρία τοῦ ἐδαφίου *6.4* μποροῦμε ν' ἀποδείξουμε ὅτι ἡ τ.μ. Y τοῦ 'Ορισμοῦ *8.2* ἔχει κατανομή $G(\alpha=\nu/2,\ \beta=2)$.

'Η X^2_ν κατανομή εἶναι λοξή πρός τά δεξιά καί ἡ πυκνότητα της ἐξαρτᾶται μόνο ἀπό τούς βαθμούς ἐλευθερίας ν. Μάλιστα δέ καθώς οἱ βαθμοί ἐλευθερίας αὐξάνουν ἡ λοξότης ἐλαττοῦται.

Τό Σχῆμα *8.1* μᾶς δίνει τρεῖς X^2 κατανομές μέ *1,4* καί *10* βαθμούς ἐλευθερίας.



Σχῆμα 8.1 X^2 κατανομές μέ *1,4* καί *10* βαθμούς ἐλευθερίας.

Μπορεῖ νά ἀποδειχθεῖ ὅτι ὅταν $\nu \to \infty$ ἡ X^2 κατανομή προσεγγίζεται ἀπό τήν κανονική κατανομή.

Οἱ συνήθεις πίνακες τῆς X^2 -κατανομῆς (βλέπε Πίνακα IV στό Παράρτημα) διάφορους β.ε. $\nu \leq 30$ ἀ ν τ ί σ τ ρ ο φ α (ἀπό πάνω) ἑκατοστιαῖα σημεῖα $X^2_{\alpha,\nu}$ τά ὁποῖα ὁρίζονται ἀπό τή σχέση

$$P(X^2 \geq X^2_{\alpha,\nu}) = \alpha \quad , \qquad (8.7)$$

π.χ. $X^2_{0,05,20} = 31,40$. Ὅταν τό ν εἶναι μεγαλύτερο τοῦ 30 οἱ πιθανότητες τῆς κατανομῆς X^2 συνήθως προσεγγίζονται μέ τίς πιθανότητες τῆς κανονικῆς κατανομῆς (βλ.Κεντρικό Ὁριακό Θεώρημα § 8.5).

Ἐπίσης ἔχουμε

$$\boxed{E(X^2_\nu) = \nu \quad \text{καί} \quad Var(X^2_\nu) = 2\nu \,.} \qquad (8.8)$$

Ὁρισμός 8.3 Κατανομή t

Ἔστω $X \sim N(0,1)$, $Y \sim X^2_\nu$ καί X,Y ἀνεξάρτητες τ.μ. Ἡ κατανομή τῆς τ.μ. $Z = \dfrac{X}{\sqrt{Y/\nu}}$ λέγεται t_ν (τί μ έ ν β α θ μ ο ύ ς ἐ λ ε υ θ ε ρ ί α ς). Συμβολικά

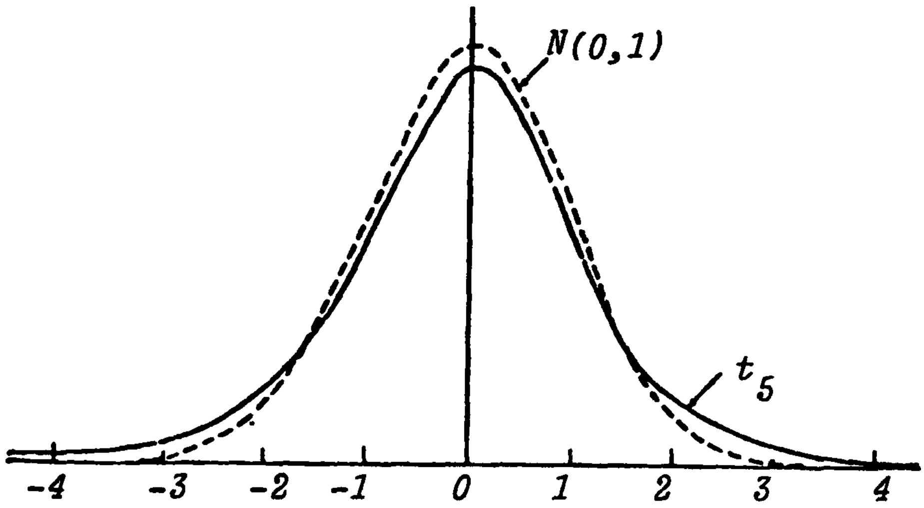
$$t_\nu = \frac{N(0,1)}{\sqrt{X^2{}_\nu/\nu}} \qquad (8.9)$$

Ἡ κατανομή t_ν εἶναι γνωστή καί ὡς ἡ κατανομή τοῦ *S t u d e n t*, ψευδώνυμο τοῦ *W.S. Gosset* ὁ ὁποῖος πρῶτος τήν ἀνακάλυψε τό 1908.

Μπορεῖ ν'ἀποδειχθεῖ ὅτι ἡ σ.π.π. τῆς t κατανομῆς δίνεται από τή σχέση

$$f_{t_\nu}(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)\sqrt{\pi\nu}}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty<t<+\infty .$$

Ἡ t_ν κατανομή εἶναι συμμετρική περί τόν ἄξονα.



Σχῆμα 8.2 Κατανομές t_5 καί $t_\infty = N(0,1)$

$x = 0$ καί ἔχει μέση τιμή καί διακύμανση:

$$E(t_\nu) = 0 , \quad Var(t_\nu) = \frac{\nu}{\nu-2} \cdot \qquad (8.10)$$

Οἱ οὐρές τῆς κατανομῆς t ἔχουν περισσότερη πιθανότητα ἀπό τίς ἀντίστοιχες οὐρές τῆς τυπικῆς κανονικῆς κατανομῆς ἡ ὁποία εἶναι ψηλότερη στό μέσον. Ὅσο μεγαλώνουν οἱ βαθμοί ἐλευθερίας τόσο περισσότερο ἡ κατανομή προσεγγίζει τήν τυπική κανονική. Στήν πραγματικότητα

$$t_{\infty} = N(0,1). \tag{8.11}$$

Μέ ἀνάλογο τρόπο ὁρίζονται τά ἀντίστροφα ἑκατοστιαῖα σημεῖα $t_{\alpha,\nu}$ τῆς κατανομῆς t_ν, τά ὁποῖα δίνονται στούς στατιστικούς πίνακες γιά $t_{\alpha,\nu} > 0$ καί διάφορες τιμές τῶν βαθμῶν ἐλευθερίας ν (Πίνακας V στό Παράρτημα).

$$P(t_\nu \geqq t_{,\nu}) = \alpha \; . \tag{8.12}$$

Λόγω συμμετρίας ἔχουμε

$$P(t_\nu \leqq -t_{\alpha,\nu}) = P(t_\nu \geqq t_{\alpha,\nu}) = \alpha$$

$$t_{1-\alpha,\nu} = -t_{\alpha,\nu}$$

π.χ. $\quad t_{0,01,15} = 2{,}602, \quad t_{0,05,\infty} = 1{,}645 \; .$

Ὁρισμός 8.4 Κατανομή F

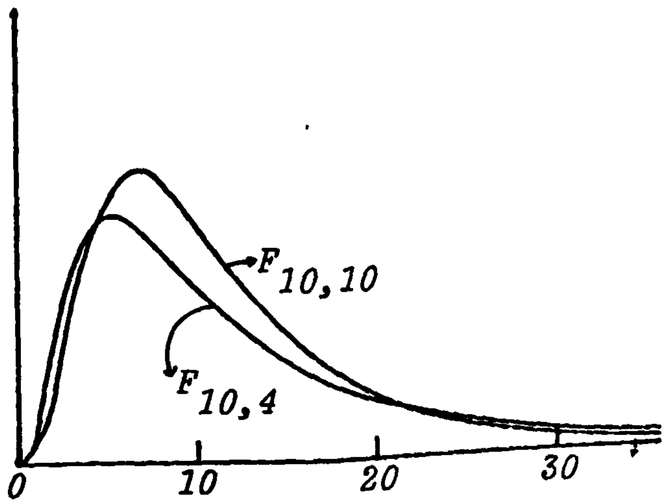
Ἔστω $X_1 \sim X^2_{\nu_1}$, $X_2 \sim X^2_{\nu_2}$ καί X_1, X_2 ἀνεξάρτητες τ.μ. Ἡ κατανομή τῆς τ.μ. $\dfrac{X_1/\nu_1}{X_2/\nu_2}$ λέγεται F_{ν_1,ν_2} (F μέ ν_1,ν_2 β α θ μ ο ύ ς ἐ λ ε υ θ ε ρ ί α ς). Συμβολικά

$$F_{\nu_1,\nu_2} = \frac{X^2_{\nu_1}/\nu_1}{X^2_{\nu_2}/\nu_2} \; . \tag{8.13}$$

Μπορεῖ ν'ἀποδειχθεῖ ὅτι ἡ σ.π.π. τῆς F κατανομῆς δίνεται ἀπό τή σχέση

$$f_{F_{\nu_1,\nu_2}}(x) = \begin{cases} \dfrac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \dfrac{x^{\frac{\nu_1}{2}-1}}{\left(1+\frac{\nu_1 x}{\nu_2}\right)^{\frac{\nu_1+\nu_2}{2}}} & , 0<x<\infty \\ 0 & , \text{ἀλλοῦ} \end{cases}.$$

Ἡ κατανομή F εἶναι ἀσυμμετρική. Δύο κατανομές F δίδονται στό Σχῆμα 8.3.



Σχῆμα 8.3 Κατανομές $F_{10,4}$ καί $F_{10,10}$

Ἡ μέση τιμή καί ἡ διακύμανση τῆς F_{ν_1,ν_2} εἶναι

$$E(F_{\nu_1,\nu_2}) = \frac{\nu_2}{\nu_2-2} \quad , \quad \nu_2 > 2 \qquad (8.14a)$$

$$Var(F_{\nu_1,\nu_2}) = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)} \quad , \quad \nu_2 > 4 \quad . \tag{8.14b}$$

Τά ἀντίστροφα (ἀπό πάνω) ἑκατοστιαῖα σημεῖα τῆς F κατανομῆς ὁρίζονται ἀπό τή σχέση (βλ. Σχῆμα 8.3)

$$P(F_{\nu_1,\nu_2} \geqq F_{\alpha,\nu_1,\nu_2}) = \alpha \tag{8.15}$$

καί δίνονται σέ στατιστικούς πίνακες.(Βλέπε Πίνακα VI στό Παράρτημα γιά $\alpha = 0,05$ καί $\alpha = 0,01$). Ὅταν $\alpha > 0,5$ χρησιμοποιοῦμε τή σχέση

$$F_{1-\alpha,\nu_1,\nu_2} = \frac{1}{F_{\alpha,\nu_2,\nu_1}} \quad . \tag{8.16}$$

Ἀπό τούς προηγούμενους ὁρισμούς εὔκολα βλέπουμε ὅτι ἰσχύει ἡ σχέση

$$t_\nu^2 = F_{1,\nu} \quad . \tag{8.17}$$

Ὁρισμός 8.5: Ἔστω $Z \sim N(0,1)$. Τά ἀντίστροφα (ἀπό πάνω) ἑκατοστιαῖα σημεῖα τῆς τυπικῆς κανονικῆς κατανομῆς ὁρίζονται ἀπό τή σχέση

$$P(Z \geqq z_\alpha) = \alpha \quad . \tag{8.18}$$

Ἰσοδύναμα ἔχουμε

$$\Phi(z_\alpha) = 1-\alpha \quad \text{καί} \quad z_{1-\alpha} = -z_\alpha \;.$$

Γιά α=5%, 2,5% καί 0,5% ἔχουμε

$$z_{0,05} = 1,645$$

$$z_{0,025} = 1,96 \qquad (8.19)$$

$$z_{0,005} = 2,58 \;.$$

8. 4 Δειγματοληψία ἀπὸ κανονικοὺς πληθυσμοὺς

Ἔστω ὅτι οἱ πληθυσμοί ἀπό τούς ὁποίους παίρνουμε τά τυχαῖα δείγματα εἶναι κανονικοί. Ποῖες εἶναι οἱ δειγματικές κατανομές τῶν διαφόρων στατιστικῶν συναρτήσεων; Θά δώσουμε τήν ἀπάντηση γιά τίς στατιστικές συναρτήσεις $\overline{X}$, s^2, $(\overline{X}-\mu)\sqrt{n}/s$, s_1^2/s_2^2, ὅπου μ εἶναι ἡ μέση τιμή τοῦ πληθυσμοῦ καί s_1^2, s_2^2 οἱ διακυμάνσεις δύο τυχαίων δειγμάτων ἀπό δύο ἀνεξάρτητους κανονικούς πληθυσμούς.

Θ ε ώ ρ η μ α 8.2:

Ἔστω $X_1,\ldots,X_n$ τυχαῖο δεῖγμα ἀπό ἕνα κανονικό πληθυσμό μέ μέση τιμή μ καί διακύμανση σ^2. Τότε

(i) ἡ δειγματική κατανομή τοῦ $\overline{X}$ εἶναι κανονική μέ μέση τιμή μ καί διακύμανση σ^2/n, συμβολικά

$$\overline{X}\sim N(\mu,\sigma^2/n)\ . \qquad (8.20)$$

(*ii*) ἡ δειγματική κατανομή τοῦ $(n-1)s^2/\sigma^2$ εἶναι χί - τετράγωνο μέ $n-1$ βαθμούς ἐλευθερίας, συμβολικά

$$\frac{(n-1)s^2}{\sigma^2}\sim\chi^2_{n-1}\ . \qquad (8.21)$$

(*iii*) Τά στατιστικά $\overline{X}$ καί S^2 εἶναι ἀνεξάρτητα.

Ἀπόδειξη: Θά ἀποδείξουμε μόνο τήν πρόταση (*i*). Οἱ προτάσεις (*ii*) καί (*iii*) ἀπαιτοῦν πολυπλοκότερη ἀνάλυση καί ξεφεύγουν ἀπό τό σκοπό τοῦ βιβλίου αὐτοῦ. Ἡ Πρόταση 6.5 τοῦ Κεφαλαίου 6 μᾶς δίνει ὅτι ἡ τ.μ. $\sum_{i=1}^{n}X_i$ εἶναι κανονική μέ μέση τιμή $\sum_{i=1}^{n}\mu_i = n\mu$ καί διακύμανση $\sum_{i=1}^{n}\sigma_i^2 = n\sigma^2$ διότι οἱ τ.μ. X_i εἶναι ἰσόνομες $N(\mu,\sigma^2)$. Ἔτσι ἡ ροπογεννήτρια συνάρτηση τῆς $\overline{X}=\sum_{i=1}^{n}X_i/n$ εἶναι

$$m_{\overline{X}}(t) = m_{\Sigma X_i}(t/n) = e^{n\mu t/n+\frac{1}{2}n\sigma^2t^2/n^2}$$

$$= e^{\mu t+\frac{\sigma^2}{2n}t^2}$$

$$= e^{\mu t+\frac{1}{2}(\frac{\sigma^2}{n})t^2}\ .$$

Ἡ πρόταση βέβαια μπορεῖ ν'ἀποδειχθεῖ μέ τήν βοήθεια τοῦ θεωρήματος 8.1 (*i*), (*ii*) καί τοῦ γεγονότος ὅτι γραμμικοί συνδιασμοί κανονικῶν τυχαίων μεταβλητῶν ὁδηγοῦν σέ κανονικές κατανομές. ▼

Παράδειγμα 8.1:

Ένα τυχαίο δείγμα μεγέθους 25 λαμβάνεται από ένα κανονικό πληθυσμό μέ διακύμανση 625. Νά βρεθοῦν οἱ πιθανότητες (i) ἡ μέση τιμή τοῦ δείγματος νά διαφέρει κατά 4 ἤ καί περισσότεο ἀπό τή μέση τιμή τοῦ πληθυσμοῦ καί (ii) ἡ διακύμανση τοῦ δείγματος νά εἶναι μεγαλύτερη τοῦ 300.

Τό θεώρημα 8.2 μᾶς δίνει

$$\bar{X} \sim N(\mu_{\bar{X}} = \mu, \sigma^2_{\bar{X}} = \frac{\sigma^2}{n} = \frac{625}{25} = 25)$$

καί

$$\frac{(n-1)s^2}{\sigma^2} = \frac{24s^2}{625} \sim \chi^2_{24} \ .$$

Έτσι γιά τό πρῶτο ἐρώτημα ἔχουμε

$$\begin{aligned} P\{|\bar{X}-\mu| \geq 4\} &= 1-P\{\mu-4<\bar{X}<\mu+4\} \\ &= 1-P\{-\frac{4}{5} < \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} < \frac{4}{5}\} \\ &= 1-P\{-0,8<z<0,8\} \\ &= 1-(0,2881+0,2881) = 0,4238. \end{aligned}$$

Γιά τό δεύτερο

$$\begin{aligned} P\{s^2>300\} &= P\{\frac{24s^2}{625} \cdot \frac{24\times 300}{625} = 11,52\} \\ &= P\{\chi^2_{24} > 11,52\} \ . \end{aligned}$$

'Από τούς συνήθεις πίνακες τῆς κατανομῆς X^2 ἔχουμε

$$97,5\% < P\{X^2_{24} > 11,52\} \leq 99\%$$

διότι $$10,856 < 11,52 < 12,401 .$$

Θ ε ώ ρ η μ α 8.3:

<u>Ἔστω</u> $X_1,\ldots,X_n$ <u>τυχαῖο δεῖγμα μεγέθους</u> n <u>ἀπό ἕνα κανονικό πληθυσμό μέ μέση τιμή</u> μ <u>καί διακύμανση</u> σ^2. <u>Τότε ἡ στατιστική συνάρτηση</u> $(\bar{X}-\mu)\sqrt{n}/s$ <u>ἔχει</u> t-<u>κατανομή μέ</u> $n-1$ <u>βαθμούς ἐλευθερίας</u>. <u>Συμβολικά</u>

$$\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1} . \qquad (8.22)$$

'Απόδειξη: Βάσει τοῦ θεωρήματος 8.2 ἔχουμε:

$$\bar{X}\sim N(\mu,\sigma^2/n) \quad , \quad \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) ,$$

$$\frac{(n-1)s^2}{\sigma^2} \sim X^2_{n-1}$$

καί οἱ μεταβλητές $\bar{X}$ καί s^2 εἶναι στατιστικῶς ἀνεξάρτητες. Ἄρα ἐφαρμόζοντες τόν 'Ορισμό 8.3 παίρνουμε:

$$\left(\frac{\bar{X}-\mu}{\sigma\sqrt{n}}\right)/\sqrt{\frac{(n-1)s^2}{\sigma^2}/(n-1)} = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1} . \ \blacktriangledown$$

Πόρισμα 8.2: <u>Ἔστω</u> $(\bar{X}_1,s_1^2)$ <u>καί</u> $(\bar{X}_2,s_2^2)$ <u>οἱ μέ-</u>

σες τιμές καί διακυμάνσεις δύο τυχαίων δειγμάτων μεγέθους n_1 καί n_2 άντίστοιχα άπό δύο άνεξάρτητους κανονικούς πληθυσμούς μέ μέσες τιμές μ_1 καί μ_2 άντίστοιχα καί κοινή διακύμανση σ^2. Τότε ή στατιστική συνάρτηση

$$\frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}, \tag{8.23}$$

ὅπου

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}, \tag{8.24}$$

δηλαδή ἔχει τήν t κατανομή μέ n_1+n_2-2 βαθμούς έλευθερίας.

Θεώρημα 8.4:

Ἔστω S_1^2 καί S_2^2 οἱ διακυμάνσεις δύο τυχαίων δειγμάτων μεγέθους n_1 καί n_2 άπό δύο άνεξάρτητους κανονικούς πληθυσμούς μέ μέσες τιμές καί διακυμάνσεις (μ_1,σ_1^2), (μ_2,σ_2^2) άντίστοιχα. Τότε ή στατιστική συνάρτηση $\sigma_2^2S_1^2/\sigma_1^2S_2^2$ ἔχει F-κατανομή μέ n_1-1 καί n_2-1 βαθμούς έλευθερίας, συμβολικά

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,\, n_2-1}.$$

'Απόδειξη: 'Από τό θεώρημα 8.2 ἔχουμε:

$$\frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1} \quad \text{καί} \quad \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}\ .$$

'Εφαρμόζοντες τόν 'Ορισμό 8.4 καί τήν ἀνεξαρτησία τῶν S_1^2 καί S_2^2 ἔχουμε

$$\left(\frac{(n_1-1)S_1^2}{\sigma_1^2}/(n_1-1)\right)/\left(\frac{(n_2-1)S_2^2}{\sigma_2^2}/(n_2-1)\right) = \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2} \sim F_{n_1-1,n_2-1}\ . \blacktriangledown$$

8. 5 Κεντρικό 'Οριακό Θεώρημα

"Οταν ὁ πληθυσμός ἀπό τόν ὁποῖο παίρνουμε τό τυχαῖο δεῖγμα δέν εἶναι κανονικός ποῖες εἶναι οἱ δειγματικές κατανομές τῶν διαφόρων στατιστικῶν; "Οπως ἔχουμε ἀναφέρει ἡ εὕρεση τῶν δειγματικῶν κατανομῶν εἶναι ἀπό τά σπουδαιότερα θέματα τῆς Μαθηματικῆς Στατιστικῆς. 'Η Στατιστική Συμπερασματολογία στηρίζεται στίς δειγματικές κατανομές. 'Η ἀπάντηση στό παραπάνω ἐρώτημα ἀπαιτεῖ μεγαλύτερο μαθηματικό ὑπόβαθρο καί ξεφεύγει ἀπό τό σκοπό τοῦ βιβλίου αὐτοῦ. "Εχουμε ὅμως ἕνα σημαντικό θεώρημα τό ὁποῖο μᾶς δίδει τήν κατά προσέγγιση κατανομή μεγάλων (μέ πολλούς ὅρους) ἀθροισμάτων ἀνεξαρτήτων καί ἰσόνομων τ.μ. καί τό ὁποῖο καταδεικνύει τή σπουδαιότητα τῆς κανονικῆς κατανομῆς. 'Από τό ἄθροισμα εὔκολα προσδιορίζεται ἡ κατανομή τῆς μέσης τιμῆς. Τό θεώρημα αὐτό εἶναι τό περίφημο Κεντρικό Ὁριακό Θεώρημα (ΚΟΘ) καί ἐμφανίζεται σέ δύο μορφές.

Θεώρημα 8.5 Κεντρικό Ὁριακό Θεώρημα (Μορφή Ι).

Ἔστω $X_1,\ldots,X_n$ n ἰσόνομες καί ἀνεξάρτητες τυχαῖες μεταβλητές (τυχαῖο δεῖγμα) ἡ κάθε μία μέ πεπερασμένη μέση τιμή μ καί διακύμανση σ^2. Τότε ἄν τό n εἶναι μεγάλο, ἡ τυχαία μεταβλητή $\dfrac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$ εἶναι κατά προσέγγιση $N(0,1)$ ἤ ἰσοδύναμα, τό ἄθροισμα ΣX_i ἔχει κατά προσέγγιση κανονική κατανομή μέ μέση τιμή $n\mu$ καί διακύμανση $n\sigma^2$.

Θεώρημα 8.6 Κεντρικό Ὁριακό Θεώρημα (Μορφή ΙΙ).

Ἔστω $X_1,\ldots,X_n$ ἕνα τυχαῖο δεῖγμα ἀπό κάποιο πληθυσμό μέ πεπερασμένη μέση τιμή μ καί διακύμανση σ^2. Ἄν τό n εἶναι μεγάλο, τότε ἡ δειγματική μέση τιμή $\overline{X}$ ἔχει κατά προσέγγιση κανονική κατανομή μέ μέση τιμή μ καί διακύμανση σ^2/n ἤ ἰσοδύναμα, ἡ τυχαία μεταβλητή

$(\overline{X}-\mu)/(\sigma/\sqrt{n})$ εἶναι κατά προσέγγιση $N(0,1)$.

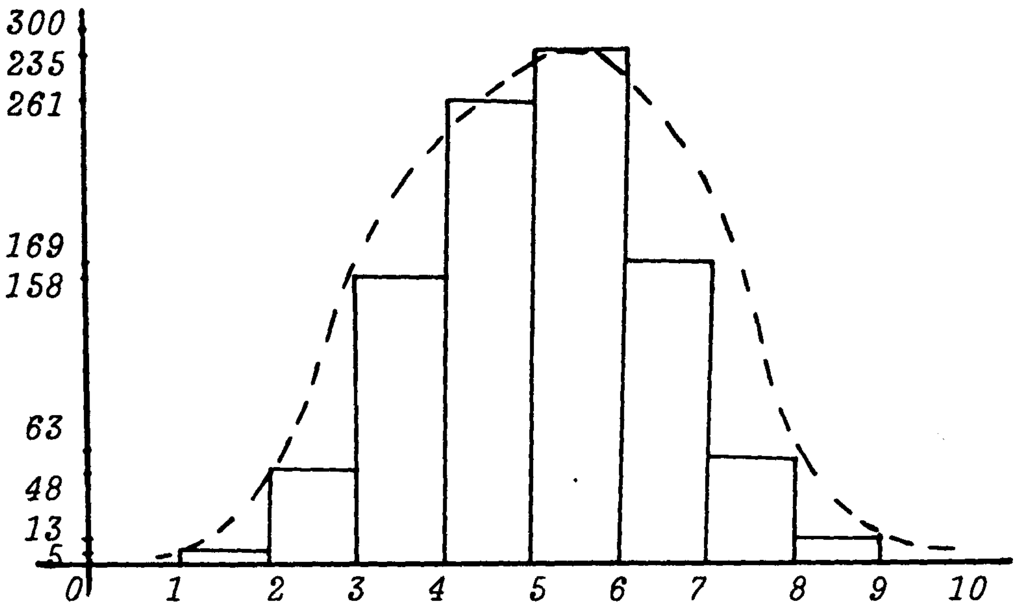
Τό ἀξιοσημείωτο τοῦ Κεντρικοῦ Ὁριακοῦ θεωρήματος εἶναι ὅτι καί ἄν ἀκόμη ἡ κατανομή τοῦ πληθυσμοῦ δέν εἶναι κανονική ἡ μέση τιμή τοῦ δείγματος ἔχει κατά π ρ ο σ έ γ γ ι σ η κανονική κατανομή ἄν τό n εἶναι μεγάλο. Αὐτό μᾶς ἐπιτρέπει νά χρησιμοποιοῦμε τή στατιστική θεωρία πού στηρίζεται στήν κανονική κατανομή ἀρκεῖ τό δεῖγμα νά εἶναι μεγάλο.

Τό Κεντρικό Ὁριακό Θεώρημα συχνά ἑρμηνεύεται ὅτι συνεπάγεται σύγκλιση τῆς κατανομῆς τοῦ $\overline{X}$ πρός μία κανονική κατανομή καθώς τό $n\to\infty$. Αὐτό δέν εἶναι σωστό διότι $Var(\overline{X}) \to 0$ καθώς τό $n\to\infty$. Τό ΚΟΘ δικαιολογεῖ τήν π ρ ο σ έ γ γ ι σ η τῆς κατανομῆς τοῦ $\overline{X}$ μέ τήν κατανομή $N(\mu,\sigma^2/n)$ ὅταν τό n εἶναι μεγάλο.

Ἡ ἀπόδειξη τοῦ Κεντρικοῦ Ὁριακοῦ Θεωρήματος ἀπαιτεῖ βαθύτερες ἔννοιες καί προτάσεις τῆς Μαθηματικῆς Στατιστικῆς καί ξεφεύγει ἀπό τό σκοπό τοῦ βιβλίου αὐτοῦ. Ἐμπειρικά τό θεώρημα μπορεῖ νά ἐπιβεβαιωθεῖ (*i*) παίρνοντας ἕνα μεγάλο ἀριθμό τυχαίων δειγμάτων μεγέθους n ἀπό μία κατανομή μέ μέση τιμή μ καί διακύμανση σ^2, (*ii*) ὑπολογίζοντας γιά κάθε δεῖγμα τή μέση του τιμή $\bar{x}$ καί (*iii*) χαράσοντας τό ἱστόγραμμα τῶν τιμῶν $\bar{x}$. Τό ἱστόγραμμα αὐτό θά πρέπει νά προσεγγίζει μία καμπύλη ἀνάλογη τῆς συναρτήσεως πυκνότητας πιθανότητας τῆς κανονικῆς κατανομῆς μέ μέση τιμή μ καί διακύμανση σ^2/n. Πράγματι *1002* τυχαῖα δείγματα μεγέθους *5* παίρνονται ἀπό τήν ὁμοιόμορφο κατανομή $U(0,10)$ μέ βάση τό διάστημα $(0,10)$. Πράγματι μέ τή βοήθεια τοῦ *minicomputer Hewlett-Packard 45* *1002* τυχαῖα δείγματα μεγέθους *5* παίρνονται ἀπό τήν ὁμοιόμορφο κατανομή $U(0,10)$ μέ προσέγγιση δευτέρου δεκαδικοῦ ψηφίου. Γιά κάθε δεῖγμα ὑπολογίζουμε τό $\overline{X}$. Κατόπιν οἱ *1002* δειγματικές μόνο μέσες τιμές ὁμαδοποιοῦνται σέ *10* διαστήματα $[0,1)$, $[1,2)$,...,$[9,10)$ μέ τά ἀκόλουθα ἀποτελέσματα:

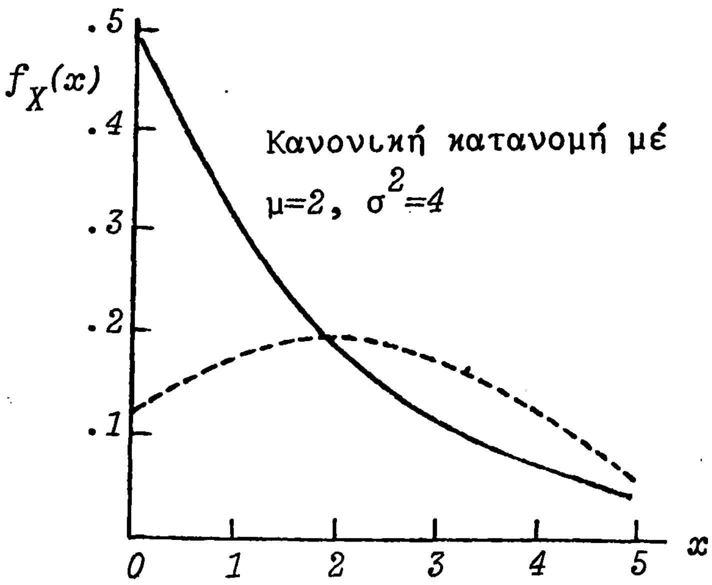
Διάστημα	Συχνότης
[0,1)	0
[1,2)	5
[2,3)	48
[3,4)	158
[4,5)	261
[5,6)	285
[6,7)	169
[7,8)	63
[8,9)	13
[9,10)	0

Τό ἱστόγραμμα τῶν τιμῶν αὐτῶν δίδεται στό Σχῆμα 8.4 ἀπό τό ὁποῖο φαίνεται ἡ προσέγγιση τῆς κατανομῆς τοῦ $\overline{X}$ μέ μία κατανομή σχήματος καμπάνας.

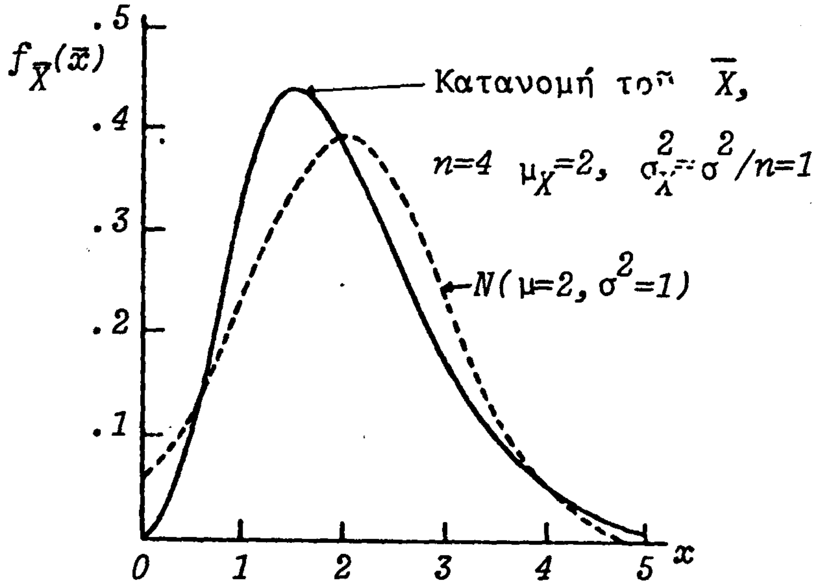


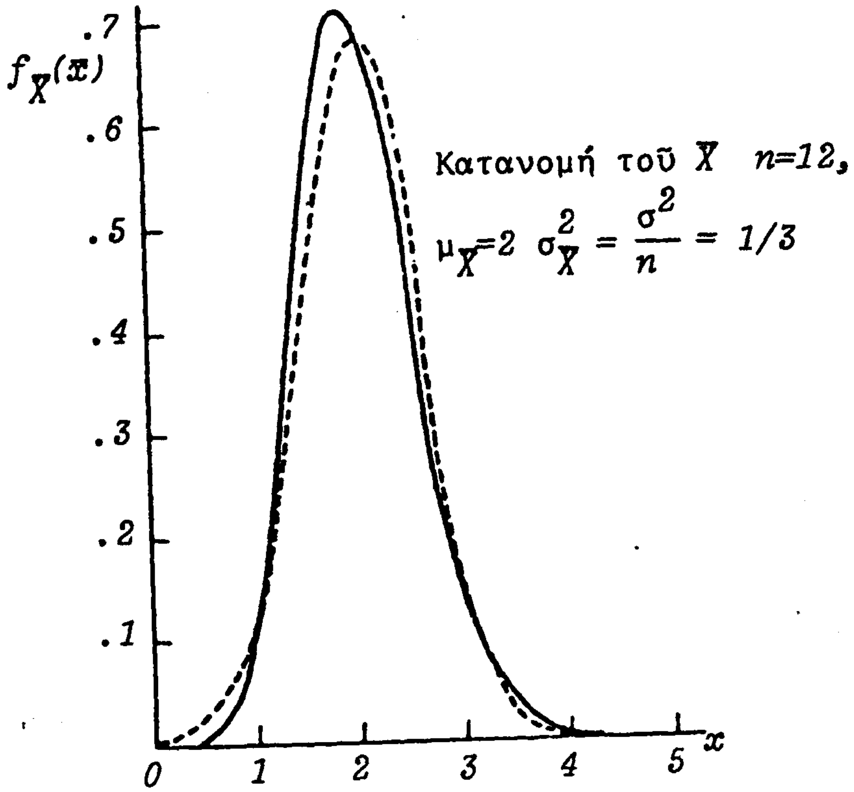
Σχῆμα 8.4 Κατανομή τοῦ $\overline{X}$ μέ δειγματοληψία ἀπό ὁμοιόμορφη κατανομή.

Εἶναι φανερό ὅτι ὅσο μεγαλύτερο τό n τόσο καλύτερη ἡ προσέγγιση τῆς κανονικῆς κατανομῆς ἀπό τήν κατανομή τοῦ $\overline{X}$. Τό Σχῆμα 8.5 ἐπεξηγεῖ ὄχι μόνο τό Κεντρικό Ὁριακό Θεώρημα ἀλλά καί τή βελτίωση τῆς προσεγγίσεως καθώς τό n μεγαλώνει. Ἔτσι τό Σχῆμα 8.5 (α) δίνει τήν πυκνότητα (συνεχής γραμμμή) τῆς ἐκθετικῆς κατανομῆς μέ $\lambda = 1/2$ ($\mu=2$, $\sigma^2=4$) καί τήν πυκνότητα (διακεκομμένη γραμμή) τῆς κανονικῆς κατανομῆς μέ τήν ἴδια μέση τιμή καί διακύμανση.



Σχῆμα 8.5 (α): Κεντρικό Ὁριακό θεώρημα μέ δειγματοληψία ἀπό ἐκθετική κατανομή





Σχῆμα 8.5(β,γ): Κεντρικό Ὁριακό θεώρημα μέ δειγματοληψία ἀπό ἐκθετική κατανομή.

Ἡ μέση τιμή $\overline{X}$ γιά δείγματα μεγεθους $n=4$ ἀπό τόν ἐκθετικό πληθυσμό τοῦ Σχήματος 8.1 (a) ἔχει μέση τιμή 2 καί διακύμανση $\sigma^2/n=1$. Τό Σχῆμα 6.3 (b) μᾶς δείχνει τήν ἀ κ ρ ι β ή κατανομή (συνεχής γραμμή) τοῦ $\overline{X}$ καί τήν ἀντίστοιχο κανονική κατανομή μέ $\mu=2$ καί $\sigma^2=1$. Οἱ δύο καμπύλες εἶναι ἀξιοσημείωτα ὅμοιες. Τά ἴδια στοιχεῖα δίδονται στό Σχῆμα 8.1 (c) ἀλλά γιά $n=12$. Ἡ προσέγγιση εἶναι ἀξιοθαύμαστη ἀκόμη καί παρά τό γεγονός ὅτι τό n δέν εἶναι ὑπερβολικά μεγάλο.

Ἀπομένει νά καθοριστεῖ τό μέγεθος τοῦ n ὥστε ἡ προσέγγιση τοῦ Κεντρικοῦ Ὁριακοῦ θεωρήματος νά εἶναι ἱκανοποιητική. Δυστυχῶς δέν ὑπάρχει ξεκάθαρη καί γενική ἀπάντηση στό ἐρώτημα αὐτό. Ἡ κατάλληλη τιμή τοῦ n ἐξαρτᾶται ἀπό τήν κατανομή τοῦ πληθυσμοῦ καί τή χρήση πού θά κάνουμε στά ἀποτελέσματα τά ὁποῖα θά πάρουμε. Κατά κανόνα τό Κεντρικό Ὁριακό Θεώρημα δίδει ἱκανοποιητικά ἀποτελέσματα ἀκόμη καί γιά μικρά δείγματα ($n \geq 25$ ἤ 30), ἀρκεῖ ἡ κατανομή τοῦ πληθυσμοῦ νά εἶναι μονοκόρυφη καί νά μήν ἔχει οὐρές μέ μεγάλη μάζα πιθανότητας.

Οἱ ἐφαρμογές τοῦ Κεντρικοῦ Ὁριακοῦ θεωρήματος εἶναι πολύπλευρες: σέ πολύπλοκες φυσικές μετρήσεις, τό σφάλμα μετρήσεως εἶναι τό ἄθροισμα ΣX_i πολλῶν ἀνεξαρτήτων τυχαίων σφαλμάτων X_i. Τό ὕψος ἑνός φυτοῦ εἶναι τό ἄθροισμα ΣX_i πολλῶν μικρῶν ἀνεξαρτήτων αὐξήσεων X_i. Τό ἴδιο ἰσχύει γιά τό ὕψος, τό βάρος κ.λ.π. ἑνός ἀτόμου, τήν ἐπίδοση στό σχολεῖο κ.λ.π. Στίς περιπτώσεις αὐτές ἡ κανονική κατανομή προσφέρει ἱκανοποιητική προσέγγιση τῆς κατανομῆς τοῦ ἀθροίσματος.

<u>Παράδειγμα 8.2</u>:

Έστω ότι σέ ένα οποιοδήποτε μ ι κ ρ ο δ ε υ τ ε ρ ό λ ε π τ ο (10^{-6} sec) ή πιθανότητα k ήλεκτρόνια νά φτάσουν στήν άνοδο μιας λυχνίας κενού είναι

$$(0,16)^k e^{-0,16}/k!\ .$$

Νά βρεθεί ή πιθανότητα ό άριθμός τών ήλεκτρονίων πού θά φθάσουν σέ ένα δ ε υ τ ε ρ ό λ ε π τ ο νά είναι μεταξύ 159000 καί 161000.

Έστω $X_1, X_2, X_3, \ldots$ ό άριθμός τών ήλεκτρονίων πού φτάνουν στήν άνοδο τό 1^o, 2^o, 3^o, κ.λ.π. μικροδευτερόλεπτο. Προφανώς $X_i \sim P(\lambda=0,16)$, $i=1,2,\ldots$ καί $\mu=EX_i=0,16$ καί $\sigma^2=VarX_i=0,16$. Τό δευτερόλεπτο έχει 10^6 μικροδευτερόλεπτα. 'Ο άριθμός τών ήλεκτρονίων πού φτάνουν στήν άνοδο σέ 1 δευτερόλεπτο είναι $X_1+X_2+\ldots+X_{10^6}$, $n=10^6$ καί ή ζητουμένη πιθανότητα είναι

$$P\{159000 \leq X_1+\ldots+X_{10^6} \leq 161000\}.$$

Θά ύπολογίσουμε τήν πιθανότητα αύτή κατά προσέγγιση χρησιμοποιώντας τό Κεντρικό 'Οριακό Θεώρημα καί τή δ ι ό ρ θ ω σ η σ υ ν ε χ ε ί α ς.

Θέτουμε $X=X_1+\ldots+X_{10^6}$ καί έχουμε

$$\mu = E(X_i)=\lambda=0,16\ ,\ \sigma^2 = Var(X_i)=\lambda=0,16\ ,$$

$$\mu_X=n\mu = 10^6\times 0,16=160000, n\sigma_X^2 = 10^6\times 0,16=160000.$$

"Ετσι

$$P\{159000 \leq X_1+\ldots+X_{10^6} \leq 161000\}$$

$$\approx P\{159999,5<X<161000,5 \mid X\sim N(\mu=160000\ \sigma^2=160000\}$$

$$\approx P\{\frac{159999,5-160000}{\sqrt{160000}}<\frac{X-\mu_X}{\sigma_X}<\frac{161000,5-160000}{\sqrt{160000}} \mid X\sim N(\mu,\sigma^2)\}$$

$$\approx P\{\frac{-1000,5}{400}<z<\frac{1000,5}{400}\}$$

$$\approx P\{-2,5<z<2,5\}= 0,9876\ .$$

8. 6 Κανονική προσέγγιση τῆς διωνυμικῆς κατανομῆς

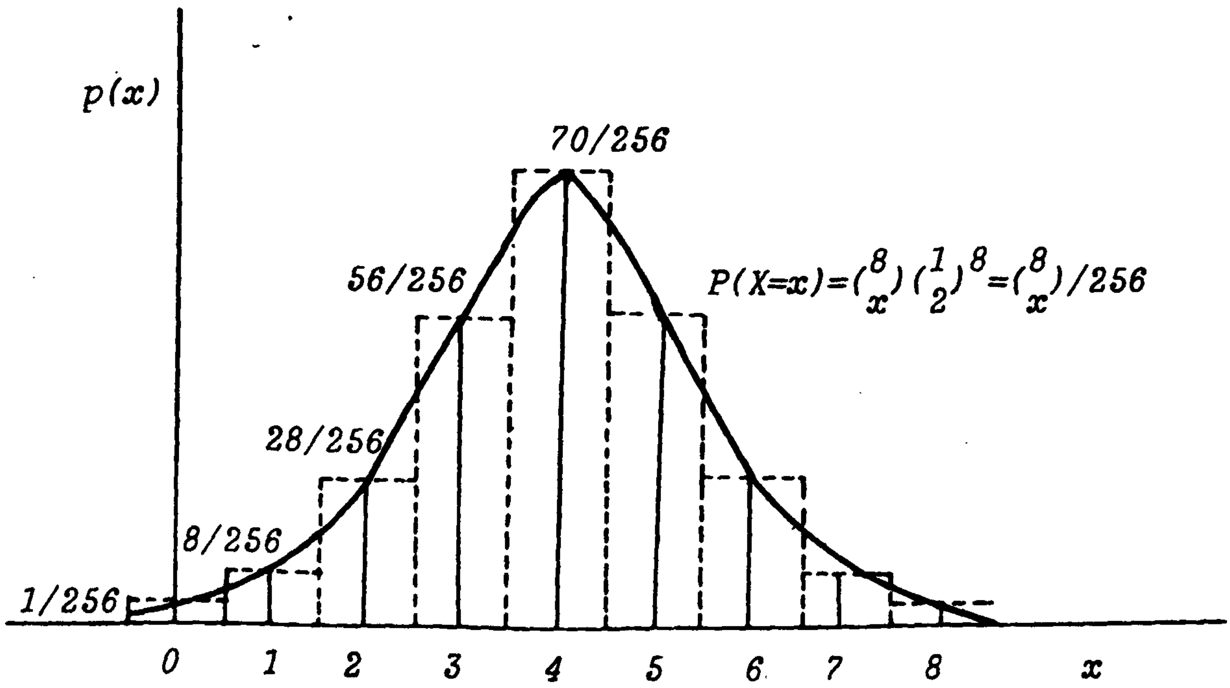
Τό Κεντρικό Ὁριακό θεώρημα μπορεῖ νά ἐφαρμοστεῖ στήν περίπτωση μίας διωνυμικῆς τυχαίας μεταβλητῆς X μέ μεγάλο ἀριθμό n δοκιμῶν καί πιθανότητα ἐπιτυχίας p. Ἡ ἐφαρμογή αὐτή ὀνομάζεται κανονική προσέγγιση τῆς διωνυμικῆς κατανομῆς.

"Εστω X_i τό ἀποτέλεσμα τῆς i δοκιμῆς τοῦ διωνυμικοῦ πειράματος, ὅπου $X_i=1$ ἄν τό ἀποτέλεσμα εἶναι ἐπιτυχία E καί $X_i=0$ ἄν τό ἀποτέλεσμα εἶναι ἀποτυχία A τότε ἔχουμε

$$X = X_1+X_2+\ldots+X_n\ .$$

Οἱ τυχαῖες μεταβλητές $X_1,\ldots,X_n$ εἶναι ἀνεξάρτητες καί ἰσόνομες μέ μέση τιμή p καί διακύμανση pq. "Ετσι γιά μεγάλο n ἡ τυχαία μεταβλητή $X=\Sigma X_i$ ἔχει κατά προσέγγιση κανονική κατανομή μέ μέση τιμή np καί διακύμανση npq.

Γιά παράδειγμα ἄς θεωρήσουμε τήν περίπτωση $X \sim B(n=8,\ p=0,5)$. Ἡ συνάρτηση πιθανότητας τοῦ X δίδεται στό Σχῆμα 8.2.



Σχῆμα 8.6 Κανονική προσέγγιση διωνυμικῆς κατανομῆς

Ἄν πάρουμε τά ἐνδιάμεσα σημεῖα $-0,5,\ 0,5,\ 1,5, \ldots, 7,5$ καί $8,5$ καί κατασκευάσουμε ὀρθογώνια μέ ὕψη τίς πιθανότητες $1/256, \ldots, 70/256, \ldots, 1/256$, τότε τά ἐμβαδά τῶν ὀρθογωνίων ἰσοῦνται μέ τίς ἀντίστοιχες πιθανότητες $p(0),\ p(1), \ldots, p(8)$. Ἔτσι

$$P(3 \leq X \leq 6) = \begin{cases} \text{ἄθροισμα ἐμβαδῶν ὀρθογωνίων μέ βάσεις} \\ (2.5, 3.5), (3.5, 4.5), (4.5, 5.5), (5.5, 6.5) \end{cases}$$

$$= \frac{56}{256} + \frac{70}{256} + \frac{56}{256} + \frac{28}{256} = \frac{210}{256} = 0.8200$$

Ἔστω τώρα μία κανονική κατανομή μέ μέση τιμή καί διακύμανση τίς ἴδιες ὅπως ἡ διωνυμική, δηλαδή μέση τιμή

$$\mu = np = 8\times(1/2) = 4 \quad \text{καί} \quad \sigma^2 = npq = 8\times(1/2)\times(1/2) = 2$$

Χαράσουμε τή συνάρτηση πυκνότητας πιθανότητας τῆς κανονικῆς πάνω στό ἴδιο σχῆμα καί βλέπουμε ὅτι τά ἐμβαδά τῶν ὀρθογωνίων προσεγγίζονται ἀπό τά ἀντίστοιχα ἐμβαδά κάτω ἀπό τήν κανονική συνάρτηση πυκνότητας πιθανότητας. Ἔτσι

$$P(2.5<X<6.5 \mid X\sim N(\mu=4, \sigma^2=2)) = P\left(\frac{2.5-4}{1.414} < \frac{X-\mu}{\sigma} < \frac{6.5-4}{1.414}\right)$$

$$= P(-1.06<z<1.77) = 0.3554+0.4616$$

$$= 0.8170$$

καί

$$P(3 \leq X \leq 6) \approx P(2.5<X<6.5 \mid X\sim N(\mu=np, \sigma^2=npq)).$$

Ἡ προσέγγιση μας ἔχει ἀκρίβεια περίπου $0,4\%$ τῆς ἀληθινῆς τιμῆς.

Ἡ αὔξηση τοῦ διαστήματος $[3,6]$ κατά $1/2$ δεξιά καί ἀριστερά λέγεται διόρθωση συνεχείας δεδομένου ὅτι διακριτές πιθανότητες ὑπολογίζονται σάν ἐμβαδά μέ τήν βοήθεια συναρτήσεων πυκνοτήτων πιθανοτήτων. Ἡ διόρθωση συνεχείας ἔχει σάν ἀποτέλεσμα τήν βελτίωση τῆς προσεγγίσεως.

Γενικά, ἄν τό διωνυμικό ἐνδεχόμενο εἶναι

$$a < X \leq b$$

τότε χρησιμοποιοῦμε τό ἐνδεχόμενο

$$a+1/2<X<b+(1/2)$$

γιά τήν προσέγγιση μέ τήν κανονική κατανομή. Ἡ προσέγγιση εἶναι καλύτερη ὅσο τό n εἶναι μεγαλύτερο καί ὅσο τό p εἶναι πλησιέστερο στό $1/2$. Ἀπό ἐμπειρικούς ὑπολογισμούς προκύπτει ὅτι ἡ προσέγγιση τῆς διωνυμικῆς κατανομῆς ἀπό τή κανονική εἶναι ἱκανοποιητική ὅταν τά n καί p ἱκανοποιοῦν τόν ἀκόλουθο π ρ α κ τ ι κ ό κ α ν ό ν α

$$0 \leq np-2\sqrt{npq}<np+2\sqrt{npq} \leq n \ .$$

<u>Παράδειγμα</u> <u>8.3</u>:

Ἕνας μεγάλος ἀριθμός σπόρων μίας ποικιλίας λουλουδιῶν ἀναμειγνύεται μέ τίς ἀκόλουθες ἀναλογίες ὡς πρός τό χρῶμα τοῦ λουλουδιοῦ πού θά παραχθεῖ: 2 κόκκινα, 2 ἄσπρα 1 μπλέ. Οἱ σπόροι ἀναμειγνύονται καί συσκευάζονται τυχαία σέ σακκοῦλες τῶν 100 σπόρων. Νά βρεθεῖ ἡ πιθανότητα μία σακκούλα νά περιέχει τό πολύ 50 "ἄσπρους" σπόρους.

Ἔστω X ὁ ἀριθμός τῶν "ἄσπρων" σπόρων μίας σακκούλας. Προφανῶς τό διωνυμικό πείραμα μπορεῖ νά υἱοθετηθεῖ καί ἄν X εἶναι ὁ ἀριθμός τῶν ἄσπρων σπόρων μίας σακκούλας τότε $X \sim B(n=100, p=40\%)$, ὅπου p=πιθανότης "ἄσπρου" σπόρου στό μεῖγμα = $2/(2+2+1)=2/5$. Τότε ἡ πιθανότητα πού ζητᾶμε εἶναι

$$\sum_{x=0}^{50} \binom{100}{x} (0,4)^x (0,6)^{100-x} \quad .$$

Ἡ πιθανότητα αὐτή μπορεῖ νά ὑπολογίστεῖ κατά προσέγγιση ὡς ἀκολούθως:

$$P(X\leq 50)$$

$$\approx P\{X<50,5 \,|\, X\sim N(\mu=100\times 0,4,\ \sigma^2=100\times 0,4\times 0,6)\}$$

$$\approx P\{\frac{X-\mu}{\sigma}< \frac{50,5-40}{\sqrt{24}} \,|\, X<\sim N(\mu=40,\ \sigma^2=24)\}$$

$$\approx P(z<2,14) = 0,9842.$$

ΑΣΚΗΣΕΙΣ

8.1. Δίδεται ὁ πληθυσμός τῶν μετρήσεων: 3, 4, 5, 6
α) Νά ὑπολογιστοῦν τά μ καί σ^2 γιά τόν πληθυσμό αὐτό. β) Νά καταγραφοῦν ὅλα τά δυνατά τυχαῖα δείγματα (μέ ἐπανάθεση) μεγέθους $n=2$.
γ) Γιά κάθε δεῖγμα νά βρεθοῦν τά $\bar{X}$ καί S^2.
δ) Νά βρεθοῦν τά $E(\bar{X})$ καί $Var(\bar{X})$ καί νά συγκριθοῦν μέ τά ἀποτελέσματα πού δίνουν γνωστές σχέσεις. ε) Νά βρεθεῖ ἡ $E(S^2)$ καί νά συγκριθεῖ μέ τό σ^2.

8.2. Ἄν $X_1,\ldots,X_n$ εἶναι τυχαῖο δεῖγμα ἀπό μιά κανονική κατανομή $N(\mu,\sigma^2)$ νά βρεθοῦν *(i)* ἡ κατανομή τοῦ $X_i-\bar{X}$ γιά i σταθερό, *(ii)* ἡ κατανομή τοῦ $n(\bar{X}-\mu)^2/\sigma^2$, *(iii)* ἡ μέση τιμή τοῦ S'^2, ὅπου $S'^2=[\Sigma(X_i-\bar{X})^2]/n$ καί *(iv)* ἡ κατανομή τοῦ $\Sigma(X_i-\mu)^2/\sigma^2$.

8.3. Ἔστω S^2 ἡ διακύμανση ἑνός τυχαίου δείγματος μεγέθους 10 ἀπό μιά κανονική κατανομή $N(\mu=0,\ \sigma^2=4)$. Νά βρεθεῖ ἡ $Var(S^2)$.

8.4. Ἄν $\bar{X}$ εἶναι ἡ μέση τιμή ἑνός τυχαίου δείγματος μεγέθους 100 ἀπό ἕνα πληθυσμό μέ κατανομή X^2_{50} νά ὑπολογιστεῖ κατά προσέγγιση ἡ πιθανότητα $P(49<\bar{X}<51)$.

8.5. Νά βρεθεῖ ἡ κατανομή τῆς τ.μ. $(\Sigma X_i^2)/\sigma^2$, ὅπου

$X_1,\ldots,X_n$ τυχαΐο δεΐγμα άπό ένα πληθυσμό μέ κανονική κατανομή $N(0,\sigma^2)$.

8.6. Έστω $X_1,\ldots,X_{25}$ καί $Y_1,\ldots,Y_{25}$ δυό τυχαΐα δείγματα άπό δυό άνεξάρτητους κανονικούς πληθυσμούς $N(0,16)$ καί $N(1,9)$ άντίστοιχα. Νά ύπολογιστεΐ ή πιθανότητα $P(\bar{X}-\bar{Y}>0)$.

8.7. Τά άποτελέσματα ένός μεγάλου άριθμοΰ άνεξάρτητων μετρήσεων καταγράφονται. Κάθε ένα άπό τά άποτελέσματα στρογγυλεύεται στόν πλησιέστερο άκέραιο. Άν τά σφάλματά πού γίνονται λόγω στρογγυλεύσεως άκολουθοΰν τήν κατανομή

$$f(x) = \begin{cases} 2(1-2|x|) , & -1/2 < x < 1/2 \\ 0 , & \text{άλλοΰ} \end{cases} ,$$

νά βρεθεΐ ή πιθανότητα τό μέσο σφάλμα στρογγυλεύσεως νά εΐναι άπολύτως μικρότερο τοΰ $1/\sqrt{3n}$.

8.8. Έστω

$$f(x) = \begin{cases} 1/x^2 , & x > 1 \\ 0 , & \text{άλλοΰ} , \end{cases}$$

ή σ.π.π. μιάς τ.μ. X καί ένα τυχαΐο δεΐγμα μεγέθους 72 άπό τήν κατανομή αύτή. Νά ύπολογιστεΐ κατά προσέγγιση ή πιθανότητα περισσότερες άπό 50 άπό τίς τιμές τοΰ τυχαίου δείγματος νά εΐναι μικρότερες τοΰ 3.

8.9. Ἔστω τυχαῖο δεῖγμα μεγέθους $n = 75$ ἀπό τήν ὁμοιόμορφη κατανομή $U(0,10)$. Νά ὑπολογιστεῖ κατά προσέγγιση ἡ πιθανότητα $P(0,45 < \bar{X} < 0,55)$.

8.10. Ἡ πιθανότητα νά γιατρευτεῖ ἕνα ἄτομο τό ὁποῖο πάσχει ἀπό μιά ἀσθένεια καί ὑποβάλλεται σέ μιά πειραματική θεραπεία εἶναι 5%. *(i)* 10 τέτοια ἄτομα ἐκλέγονται τυχαίως καί ὑποβάλλονται στήν θεραπεία. Ποιά ἡ πιθανότητα νά γιατρευτοῦν 3 ἀπό αὐτά; *(ii)* Ποιά ἡ πιθανότητα νά χρειαστοῦν 15 τέτοια ἄτομα μέχρις ὅτου 2 ἀπό αὐτά γιατρευτοῦν. *(iii)* 600 ἄτομα μέ αὐτή τήν ἀσθένεια ἐκλέγονται στήν τύχη. Ποιά ἡ πιθανότητα μετά τήν θεραπεία νά γιατρευτοῦν 25 μέχρι 35 ἄτομα;

8.11. Ἔστω τυχαῖο δεῖγμα μεγέθους 48 ἀπό τήν κατανομή

$$f(x) = \begin{cases} 1/2 & ,0 < x \leq 2 \\ 0 & ,\text{ἀλλοῦ}. \end{cases}$$

Νά ὑπολογιστεῖ κατά προσέγγιση ἡ $P(0.95<\bar{X}<1.05)$.

8.12. Ἡ Ὀλυμπιακή Ἀεροπορία (ΟΑ) γνωρίζει ὅτι 5% τῶν ἀτόμων πού κλείνουν θέση σέ μιά πτήση δέν ἐμφανίζονται τελικά γιά νά ταξιδέψουν. Ἄν ἡ ΟΑ κάνει κρατήσεις γιά 160 ἄτομα σέ μιά πτήση μέ 150 μόνο θέσεις νά βρεθεῖ ἡ πιθανότητα νά ὑπάρχει θέση διαθέσιμη γιά ὅλα τά ἄτομα πού ἔχουν κλείσει θέσεις καί ἐμφανίζονται γιά τό ταξείδι.

8.13. Ἡ μέση τιμή καί ἡ τυπική ἀπόκλιση τοῦ χρόνου ζωῆς μιᾶς λυχνίας εἶναι 1285 καί 150 ὧρες ἀντίστοιχα. Νά βρεθεῖ ἡ πιθανότητα ἡ μέση τιμή ἑνός τυχαίου δείγματος 100 λυχνιῶν νά εἶναι μεγαλύτερη ἀπό 1300 ὧρες.

8.14. Ἕνα ρολόϊ κάνει σφάλμα στήν ἔνδειξη τῆς ὥρας τό πολύ μισό $(\pm 0,5)$ λεπτό τήν ἡμέρα, ἀλλά γενικά (κατά μέσον ὅρο) πηγαίνει καλά. Ἔτσι γιά τό μοντέλο "ἀκριβείας" τοῦ ρολογιοῦ ὁρίζουμε μέ X_i τό πραγματικό σφάλμα τοῦ ρολογιοῦ (σέ λεπτά τῆς ὥρας) στίς 12 τό μεσημέρι τῆς i μέρας, $i=1,2,\ldots$, καί ὑποθέτουμε ὅτι οἱ τ.μ. X_i, $i=1,2,\ldots$ εἶναι ἀνεξάρτητες ὁμοιόμορφες τ.μ. $U(-1/2\,,\,1/2)$. Νά βρεθεῖ κατά προσέγγιση ἡ πιθανότητα μετά ἀπό 100 μέρες τό ρολόϊ νά δείχνει λάθος τό πολύ ἴσο μέ $3,46=\sqrt{12}$ λεπτά στίς 12 τό μεσημέρι.

8.15. Ἔστω ὅτι ὁ ἀριθμός τῶν ἠλεκτρονίων πού φθάνουν στήν ἄνοδο μιᾶς λυχνίας κενοῦ σέ ἕνα ὁποιοδήποτε μικροδευτερόλεπτο $(10^{-6}$ sec) ἀκολουθεῖ κατανομή τοῦ Poisson μέ παράμετρο $\lambda=0,16$. Νά βρεθεῖ ἡ πιθανότητα ὁ ἀριθμός τῶν ἠλεκτρονίων πού θά φθάσουν σέ ἕνα δευτερόλεπτο νά εἶναι μεταξύ 159.000 καί 161.000.

8.16. Ἐργοστάσιο συσκευάζει τά τρανζίστορς πού κατασκευάζει σέ κιβώτια τῶν 150. Τά βάρη τῶν τρανζίστορς εἶναι τυχαῖες μεταβλητές μέ μέση τιμή 0,8 gr. καί διακύμανση 0,49 gr. Νά βρεθεῖ ἡ πιθανότητα τό καθαρό βάρος 30 κιβωτίων νά εἶναι μεγαλύτερο ἀπό 3,7 kgr.

8.17. Τά ἄτομα πού χρησιμοποιοῦν τό ἀσανσέρ μιᾶς πολυκατοικίας θεωροῦνται ὅτι προέρχονται ἀπό κάποιο κανονικό πληθυσμό μέ μέση τιμή 60 kgr. καί τυπική ἀπόκλιση 10 kgr. Νά βρεθεῖ ἡ πιθανότητα 4 ἄτομα νά ἔχουν συνολικό βάρος μικρότερο τῶν 280 kgr. πού εἶναι τό ὅριο φορτίου τοῦ ἀσανσέρ.

8.18. Ἔστω X_i, $i=1,\ldots,5$ ἀνεξάρτητες τυπικές κανονικές τυχαῖες μεταβλητές (a) Νά βρεθεῖ ἡ μέση τιμή καί ἡ τυπική ἀπόκλιση τῆς $X_1^2+X_2^2+\ldots+X_5^2$. (b) Νά βρεθεῖ ἡ σταθερή C ἔτσι ὥστε

$$P(-C \leq 2X_5/\sqrt{X_1^2+\ldots+X_4^3} \leq C) = 0{,}90.$$

8.19. Ἄν X_1, X_2, X_3 εἶναι ἀνεξάρτητες τυχαῖες μεταβλητές ἡ κάθε μιά μέ τυπική κανονική κατανομή (a) Νά ὑπολογιστεῖ ἡ $P(X_1^2+X_2^2+X_3^2 \leq 1)$ καί (b) Νά δοθεῖ ἡ κατανομή τῆς $2X_1^2/(X_2^2+X_3^2)$.

8.20. 'Η πιθανότητα θεραπείας ένός άτόμου μέ ένα σπάνιο σύνδρομο εἶναι 1% . Σέ κάθε μιά άπό τίς άκόλουθες περιπτώσεις έκλέγεται ένα τυχαῖο δεῖγμα άπό τόν πληθυσμό τῶν άτόμων πού πάσχουν άπό τό σύνδρομο αύτό. "Αν τό μέγεθος τοῦ δείγματος εἶναι α) $n=10$, νά βρεθεῖ ή πιθανότητα θεραπείας ένός άκριβῶς άτόμου, β) $n=300$, νά βρεθεῖ ή πιθανότητα θεραπείας τριῶν τό πολύ άτόμων χρησιμοποιῶντας τήν προσέγγιση τοῦ Poisson καί γ) $n=600$, νά βρεθεῖ ή πιθανότητα θεραπείας 4 μέχρι καί 12 άτόμων χρησιμοποιῶντας τήν κανονική προσέγγιση.

8.21. "Εστω ότι ό άριθμός τῶν άεροπλάνων πού φτάνουν σέ κάποιο άεροδρόμιο σέ όποιαδήποτε εἰκοσάλεπτο περίοδο άκολουθεῖ τήν κατανομή τοῦ Poisson μέ μέση τιμή 100. *(i)* Νά βρεθεῖ ένα κάτω φράγμα τῆς πιθανότητας ό άριθμός τῶν άεροπλάνων πού φτάνουν σέ μιά δεδομένη εἰκοσάλεπτο περίοδο νά εἶναι μεταξύ 80 καί 120 χρησιμοποιῶντας τήν άνισότητα τοῦ Chebyshev. *(ii)* Νά ύπολογιστεῖ ή πιθανότητα στό *(i)* χρησιμοποιῶντας τό Κεντρικό 'Οριακό Θεώρημα.

8.22. "Αν X_1, X_2 εἶναι τυχαῖο δεῖγμα άπό μιά τυπική κανονική κατανομή $N(0,1)$ νά βρεθοῦν οἱ κατανομές τῶν στατιστικῶν συναρτήσεων

(i) $(X_2-X_1)/\sqrt{2}$, *(ii)* $(X_1+X_2)^2/(X_2-X_1)^2$

$(iii)\ (X_1+X_2)/\sqrt{(X_1-X_2)^2}$ καί $(iv)\ X_2^2\,/\,X_1^2$.

8.23. Ἔστω $X_1,\ldots,X_n$ τυχαῖο δεῖγμα ἀπό μιά τυπική κανονική κατανομή $N(0,1)$. Ἔστω ἐπίσης

$$\bar{X}_k=\frac{1}{k}\sum_{i=1}^{k}X_i \quad \text{καί} \quad \bar{X}_{n-k}=\frac{1}{n-k}\sum_{i=k+1}^{n}X_i$$

Νά βρεθοῦν οἱ κατανομές τῶν ἀκόλουθων στατιστικῶν συναρτήσεων $(i)\ (\bar{X}_k+\bar{X}_{n-k})/2$
$(ii)\ k\bar{X}_k^2+(n-k)\bar{X}_{n-k}^2$, $(iii)\ X_1^2/X_2^2$

8.24. Ἔστω X_i ὁ ἀριθμός τῶν μετεωροειδῶν πού συγκρούονται μέ ἕνα δοκιμαστικό δορυφόρο κατά τή διάρκεια τῆς i τροχιᾶς καί S_n ὁ συνολικός ἀριθμός τῶν μετεωροειδῶν πού συγκρούονται μέ τόν δορυφόρο κατά τή διάρκεια n τροχιῶν. Ἄν οἱ τυχαῖες μεταβλητές $X_1, X_2,\ldots$ εἶναι ἀνεξάρτητες, ἡ κάθε μιά Poisson μέ παράμετρο λ (i) Νά βρεθοῦν ἡ $E(S_n)$ καί ἡ $Var(S_n)$ καί (ii) Ἄν $n=100$ καί $\lambda=4$ νά βρεθεῖ κατά προσέγγιση ἡ $P(S_{100}>440)$.

ΚΕΦΑΛΑΙΟ 9

στατιστικη συμπερασματο-λογια

Συμπερασματολογία εἶναι ὁ κλάδος τῆς Λογικῆς, ὁ ὁποῖος ἀσχολεῖται μέ τήν ἐξαγωγή συμπερασμάτων. Τά συμπεράσματα μπορεῖ νά ἔχουν τή μορφή ἀποφάσεως, προβλέψεως, ἤ ἀποκτήσεως καινούργιας γνώσης, διαμορφώσεως γνώμης κ.λ.π. Καθημερινά ἀντιμετωπίζουμε καταστάσεις πού ἀπαιτοῦν λήψη προσωπικῆς ἀποφάσεως ἤ προβλέψεως γιά τό μέλλον. Ἡ κυβέρνηση ἐνδιαφέρεται νά προβλέψει τόν τιμάριθμο τό ἑπόμενο ἔτος. Ὁ ἐρευνητής θέλει νά ἀποδείξει πειραματικά τή θεωρία του. Ὁ ἰατρός ἐνδιαφέρεται νά συγκρίνει δύο διαφορετικές θεραπεῖες. Ἡ νοικοκυρά θέλει νά ξέρει ἄν τό ἀπορρυπαντικό *A* εἶναι πιό ἀποτελεσματικό ἀπό τό ἀπορρυπαντικό *B*. Γιά ὅλα αὐτά ἐξάγονται συμπεράσματα μέ τή βοήθεια ὅλων τῶν σχετικῶν πληροφοριῶν πού λέγονται παρατηρήσεις ἤ δεδόμενα.

Ὁ Στατιστικός συλλέγει τά δεδομένα μέ πειράματα ἤ δειγματοληψίες καί ἐπιδιώκει νά βγάλει σ υ μ π ε ρ ά σ μ α τ α γιά τά φαινόμενα πού ἐξετάζει. Χρησιμοποιεῖ τίς μετρήσεις του γιά ἐ κ τ ί μ η σ η τῶν τιμῶν τῶν ἀγνώστων παραμέτρων ἤ γιά τ έ σ τ (ἔλεγχο) ὑ π ο θ έ σ ε ω ν περί αὐτῶν. Ἔτσι ἡ Στατιστική Συμπερασματολογία χωρίζεται σέ δύο κλάδους: τήν Ἐ κ τ ι μ η τ ι κ ή καί τόν Ἔ λ ε γ χ ο τ ῶ ν Σ τ α τ ι σ τ ι κ ῶ ν Ὑ-π ο θ έ σ ε ω ν.

Στήν § 8.1 παρουσιάσαμε μία σφυγμομέτρηση κοινῆς γνώμης κατά τήν ὁποία πήραμε ἕνα δεῖγμα ἀπό τούς ψηφοφόρους μιᾶς πόλης πού θεωροῦνται διχασμένοι σέ ἐκείνους πού ὑποστηρίζουν τόν ὑποψήφιο A καί ἐκείνους πού ὑποστηρίζουν τόν ὑποψήφιο B. Ὑποθέσαμε ὅτι p εἶναι τό ἄγνωστο ποσοστό ψηφοφόρων πού ὑποστηρίζουν τόν ὑποψήφιο A. Σκοποί τῆς σφυγμομέτρησης μπορεῖ νά εἶναι ἡ ἐκτίμηση ἤ πρόβλεψη τοῦ p ἤ ὁ ἔλεγχος κάποιας ὑποθέσεως γιά τό p, π.χ. νά ἐλέγξουμε ὅτι $p>1/2$, δηλαδή νά προβλέψουμε βάσει τοῦ δείγματος ἄν ὁ A θά κερδίσει τίς ἐκλογές. Καί στίς δύο περιπτώσεις βγάζουμε στατιστικά συμπεράσματα γιά τό p.

Ἡ στατιστική συμπερασματολογία ἔχει ἐπαγωγικό χαρακτήρα, δηλαδή γενικεύει ἀπό ἕνα συγκεκριμένο πείραμα σέ ὅλα τά παρόμοια πειράματα, ἀπό τό δεῖγμα στόν πληθυσμό. Ἐξ αἰτίας τοῦ ἐπαγωγικοῦ της χαρακτήρα, κάθε στατιστική συμπερασματολογία περιέχει ἀβεβαιότητα. Ἡ ἀβεβαιότητα αὐτή μπορεῖ νά μετρηθεῖ μέ τή βοήθεια τῆς ἔννοιας τῆς πιθανότητας. Γι'αὐτό μελετήσαμε στά προηγούμενα σέ τόση ἔκταση τή Θεωρία Πιθανότητων. Σκοπός τῆς Στατιστικῆς εἶναι ἡ μελέτη μεθόδων συμπερασματολογίας καί τρόπων μετρήσεως τῆς ἀβεβαιότητας αὐτῆς. Ἔτσι σέ κάθε πρακτική ἐφαρμογή ἡ στατιστική συμπερασματολογία περιλαμβάνει δύο στοιχεῖα: (α) τό συμπέρασμα καί (β) τό μέτρο τῆς ὀρθότητας ἤ καταλληλότητας του.

Στό Κεφάλαιο αὐτό ἀναπτύσουμε πρῶτα τίς ἀρχές καί μεθόδους ἐκτιμήσεως στατιστικῶν παραμέτρων καί στή συνέχεια δίνουμε τή φιλοσοφία καί μεθοδολογία τοῦ ἐλέγχου τῶν στατιστικῶν ὑποθέσεων. Στόχος μας εἶναι νά δώσουμε μία ἁπλή καί σύντομη εἰσαγωγή στίς ἰδέες τῆς Στατιστικῆς Συμπερασματολογίας. Γιά πλήρη ἀνάπτυξη τῶν θεμάτων αὐτῶν ὁ ἀναγνώστης παραπέμπεται σέ προχωρημένα συγγράμματα τῆς Στατιστικῆς.

Α΄ ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΩΝ

9Α. 1 Γενικά

Ἔστω ὅτι ἔχουμε στά χέρια μας ἕνα τ.δεῖγμα $X_1, X_2, \ldots, X_n$ ἀπό κάποιο πληθυσμό μέ μία ἄγνωστη παράμετρο. Τό πρόβλημα μας εἶναι νά βροῦμε μία ἤ περισσότερες ποσότητες πού θά χρησιμοποιηθοῦν γιά τήν ἐκτίμηση τῆς ἄγνωστης παραμέτρου. Εἶναι προφανές ὅτι οἱ ποσότητες αὐτές θά πρέπει νά προέρχονται ἀπό τό δεῖγμα. Γιά παράδειγμα, ὅταν θέλουμε νά ἐκτιμήσουμε τή μέση τιμή μ ἑνός πληθυσμοῦ X, μποροῦμε νά χρησιμοποιήσουμε τήν ποσότητα $\bar{X}$. Ὅταν χρησιμοποιοῦμε μιά τέτοια ποσότητα, ἡ ποσότητα αὐτή λέγεται ἐ κ τ ι μ ή τ ρ ι α σ υ ν ά ρ τ η σ η καί ἡ ἀριθμητική τιμή της πού προκύπτει ἀπό τό δεῖγμα ἐ κ τ ι μ η τ ή ς τῆς παραμέτρου. Οἱ ἐκτιμήτριες συναρτήσεις εἶναι στατιστικές συναρτήσεις καί ἐξαρτῶνται μόνο ἀπό τό δεῖγμα καί ὄχι τήν παράμετρο πού ζητᾶμε νά ἐκτιμήσουμε. Στήν περίπτωση αὐτή ἐκτιμοῦμε ἕνα ἄγνωστο σημεῖο μέ ἕνα γνωστό σημεῖο. Ἡ σχετική μεθοδολογία λέγεται σ η μ ε ι-ο ε κ τ ι μ η τ ι κ ή.

Τήν ἄγνωστη παράμετρο μποροῦμε νά ἐκτιμήσουμε μέ ἕνα διάστημα ἀντί μέ ἕνα σημεῖο. Χρειαζόμαστε δηλαδή κατά κανόνα δύο ποσότητες πού προέρχονται ἀπό τό δεῖγμα, τά ὅρια τοῦ διαστήματος. Ἡ σχετική μεθοδολογία λέγεται ἐ κ τ ι μ η τ ι κ ή μέ δ ι α σ τ ή μ α τ α ἐ μ π ι σ τ ο σ ύ ν η ς.

Σέ πολλά προβλήματα τῆς Στατιστικῆς ἡ κατανομή τοῦ πληθυσμοῦ εἶναι γνωστῆς μορφῆς ἐκτός ἀπό τό γεγονός ὅτι περιέχει μία ἤ περισσότερες ἄγνωστες παραμέτρους. Ἡ περίπτωση αὐτή ὀνομάζεται (παραμετρικό) σ τ α τ ι σ τ ι κ ό μ ο ν τ έ λ ο. Γιά παράδειγμα ἡ κατανομή τοῦ πληθυσμοῦ μπορεῖ νά εἶναι διωνυμική $B(n,p)$ μέ n γνωστό καί $p\epsilon(0,1)$ ἄγνωστο ἤ κανονική $N(\mu,\sigma^2)$ μέ $-\infty<\mu<\infty$, $\sigma^2>0$ ἄγνωστα κ.λ.π. Γενικά οἱ ἄγνωστοι παράμετροι θά συμβολίζονται μέ θ ἤ θ_1, θ_2 κ.λ.π. καί οἱ κατανομές τοῦ πληθυσμοῦ X μέ $p(x,\theta)$ ἤ $f(x_i,\theta)$. Τό πρόβλημα μας εἶναι νά ἐκτιμήσουμε τό θ ἤ συναρτήσεις $g(\theta)$ τοῦ θ. Οἱ ἐκτιμήτριες συναρτήσεις τοῦ θ θά συμβολίζονται μέ $\hat{\theta}$ ἤ γενικά $T(X_1,\ldots,X_n)$.

9Α 2 Σημειοεκτιμητική

Τό πρόβλημα τῆς σημειοεκτιμητικῆς εἶναι διπλό: ποιά εἶναι τά κριτήρια ἤ ἀρχές μέ τίς ὁποῖες ἀξιολογοῦμε τούς ἐκτιμητές καί ὁρίζουμε τόν "καλύτερο" ἐκτιμητή μεταξύ τῶν διαφόρων ὑποψηφίων καί ποιές εἶναι οἱ μέθοδοι εὑρέσεως ἐκτιμητῶν;

Τά κριτήρια ἀξιολογήσεως ἐκτιμητῶν εἶναι πολλά. Ἀπό αὐτά ἀναφέρουμε δύο βασικά:

'Αμεροληψία: Ἕνας ἐκτιμητής $\hat{\theta}$ μιᾶς παραμέτρου θ λέγεται ἀμερόληπτος ἄν ἔχει ἀναμενόμενη τιμή θ γιά κάθε τιμή τῆς παραμέτρου θ, δηλαδή

$$E(\hat{\theta}) = \theta \qquad (9.1)$$

γιά κάθε θ.

'Η ἀμεροληψία ἑνός ἐκτιμητῆ μετρᾶ τήν ὀ ρ θ ό τ η τ α (*accuracy*) τῆς ἐκτιμήσεως καί ἐκφράζει τήν ἰδέα ὅτι σέ πολλές νοερές ἤ πραγματικές ἐπαναλήψεις τῆς δειγματοληψίας μας ἡ μέθοδος ἐκτιμήσεως πού χρησιμοποιοῦμε θά μᾶς δώσει κατά μέσο ὅρο τήν ἄγνωστη τιμή τοῦ θ.

Εἶναι εὔκολο νά δεῖ κανείς ὅτι ὑπάρχουν πολλοί ἀμερόληπτοι ἐκτιμητές γιά κάθε παράμετρο θ, π.χ. ἄν $n=50$ οἱ ἐκτιμητές

$$\bar{X}_{50} = \frac{1}{50}\sum_{i=1}^{50} X_i$$

καί

$$\bar{X}_{25} = \frac{1}{25}\sum_{i=1}^{25} X_i$$

εἶναι καί οἱ δύο ἀμερόληπτοι ἐκτιμητές τοῦ μ. Διαισθανόμαστε ὅτι ὁ $\bar{X}_{50}$ εἶναι καλύτερος ἀπό τόν $\bar{X}_{25}$ διότι προκύπτει ἀπό περισσότερες μετρήσεις, δηλ. χρησιμοποιεῖ περισσότερη πληροφορία ἀπό τόν $\bar{X}_{25}$. 'Εκεῖνο πού παίζει ρόλο εἶναι ἡ μ ε τ α β λ η τ ό τ η τ α τοῦ ἐκτιμητῆ. 'Η μεταβλητότητα τοῦ ἐκτιμητῆ μετρεῖται μέ τή διακύμανση του. 'Η διακύμανση τοῦ $\bar{X}_{50}$ εἶναι $\sigma^2/50$ [βλ.(8.2)] καί ἡ διακύμανση τοῦ $\bar{X}_{25}$ εἶναι $\sigma^2/25$, μεγαλύτερη ἀπό τήν πρώτη. Όσο μικρότερη ή διακύμανση τόσο καλύτερη. Αὐτές οἱ ἰδέες μᾶς ὁδηγοῦν στό δεύτερο κριτήριο.

Ἐ λ ά χ ι σ τ η Δ ι α κ ύ μ α ν σ η: Μεταξύ τῶν ἀμερολήπτων ἐκτιμητῶν προτιμητέος εἶναι ἐκεῖνος πού ἔχει τή μικρότερη διακύμανση.

Ἡ ἀ κ ρ ί β ε ι α *(precision)* ἑνός ἀμερόληπτου ἐκτιμητῆ $\hat{\theta}$ μετρεῖται συνήθως μέ τήν διακύμανση του.

$$\sigma_{\hat{\theta}}^{2} = Var(\hat{\theta}) = E(\hat{\theta}-\theta)^{2} \qquad (9.2)$$

ἤ τήν τυπική ἀπόκλιση του

$$\sigma_{\hat{\theta}} = \sqrt{E(\hat{\theta}-\theta)^{2}}\ .$$

Ἡ τυπική ἀπόκλιση $\sigma_{\hat{\theta}}$ λέγεται μερικές φορές καί τ υ π ι κ ό σ φ ά λ μ α τ ο ῦ ἐ κ τ ι μ η τ ῆ. Γιά τό δειγματικό μέσο $\overline{X}$ τό τυπικό σφάλμα εἶναι $\sigma/\sqrt{n}$, δηλαδή

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\ . \qquad (9.3)$$

Ἕνας ἐκτιμητής λέγεται ἀ μ ε ρ ό λ η π τ ο ς ὁ μ ο ι ό μ ο ρ φ α ἐ λ ά χ ι σ τ η ς δ ι α κ ύ μ α ν σ η ς (Α.Ο.Ε.Δ.) ὅταν εἶναι ἀμερόληπτος καί ἔχει μικρότερη διακύμανση ἀπό κάθε ἄλλο ἀμερόληπτο ἐκτιμητή.

Ἔχουμε δεῖ ὅτι ἄν ἡ ἄγνωστη παράμετρος θ εἶναι ἡ μέση τιμή τοῦ πληθυσμοῦ μ, τό $\overline{X}$ ἔχει τήν ἰδιότητα τῆς ἀμεροληψίας δηλαδή

$$E(\overline{X}) = \mu\ . \qquad (9.4)$$

Ἄν ἐπί πλέον ὁ πληθυσμός εἶναι κανονικός τό $\overline{X}$ εἶναι Α.Ο.Ε.Δ. τοῦ μ.

Ὁμοίως ἄν $X_1,\ldots,X_n$ εἶναι ἕνα τυχαῖο δεῖγμα ἀπό κάποιο πληθυσμό μέ ἄγνωστη παράμετρο σ^2 τό στατιστικό S^2 εἶναι ἀμερόληπτος ἐκτιμητής τοῦ σ^2 δηλαδή

$$E(S^2) = \sigma^2. \qquad (9.5)$$

Τό σημαντικό σημεῖο τῶν σχέσεων (9.4) καί (9.5) εἶναι ὅτι ἰσχύουν πάντα ἀσχέτως τῶν ἰδιαιτέρων τιμῶν πού τά μ καί σ^2 ἐνδέχεται νά ἔχουν. Ἡ σχέση (9.5) δικαιολογεῖ τή χρήση τοῦ S^2 καί ὄχι τοῦ S'^2 γιά τή διακύμανση τοῦ δείγματος. Γιά νά γίνει ἡ διακύμανση τοῦ δείγματος ἀμερόληπτη ἐκτιμήτρια τοῦ σ^2 διαιρέσαμε τό ἄθροισμα τῶν τετραγώνων τῶν ἀποκλίσεων (7.15) μέ τό $n-1$ καί ὄχι μέ τό n. Ἡ ποσότης S'^2 τείνει νά ὑ-π ο ε κ τ ι μ ᾶ τό σ^2. Ἡ τυπική ἀπόκλιση τοῦ δείγματος

$$S = \sqrt{\frac{1}{n-1}\,\Sigma(X_i-\overline{X})^2}$$

συνήθως χρησιμοποιεῖται ὡς ἐκτιμητής τοῦ σ παρά τό γεγονός ὅτι δέν εἶναι ἀκριβῶς ἀμερόληπτος.

Τό τυπικό σφάλμα τοῦ $\overline{X}$ εἶναι

$$\sigma_{\overline{X}} = \sigma/\sqrt{n}$$

καί ἐκτιμᾶται μέ

$$\hat{\sigma}_{\overline{X}} = S/\sqrt{n} \quad .$$

Ἐκτός ἀπό τά παραπάνω κριτήρια ἐξιολογήσεως ἐκτιμητῶν ὑπάρχουν καί ἄλλα ὅπως ἡ ἐ π ά ρ κ ε ι α, ἡ σ υ ν έ π ε ι α κλπ. Ὑπάρχουν ἐπίσης διάφοροι μέθοδοι εὑρέσεως ἐκτιμητῶν. Ἡ πλέον σημαντική ἀπό αὐτές εἶναι ἡ μέθοδος τῆς μ ε γ ί σ τ η ς π ι θ α ν ο φ ά ν ε ι α ς.

Τά θέματα αὐτά ξεφεύγουν ἀπό τό σκοπό τοῦ βιβλίου αὐτοῦ καί δέν θά ἀναπτυχθοῦν. Ἀποτελοῦν ἀντικείμενο τῆς Ἐκτιμητικῆς ἡ ὁποία γιά κάθε στατιστικό μοντέλο καί περίπτωση μᾶς δίνει τόν καλύτερο ἐκτιμητή ἤ ὑποδεικνύει τρόπους εὑρέσεως αὐτοῦ.

9Α. 3 Διαστήματα ἐπιστοσύνης

Ἔστω $X_1,\ldots,X_n$ ἕνα τυχαῖο δεῖγμα ἀπό κάποιο πληθυσμό μέ μία ἄγνωστη παράμετρο θ. Ἔστω $L=L(X_1,\ldots,X_n)$ καί $U=U(X_1,\ldots,X_n)$ τά κάτω καί ἄνω ὅρια ἑνός διαστήματος ἀντιστοίχα. Τά ὅρια αὐτά εἶναι συναρτήσεις τῶν τυχαίων μεταβλητῶν τοῦ δείγματος. Οἱ τιμές τους εἶναι $l(x_1,\ldots,x_n)$ καί $u(x_1,\ldots,x_n)$. Μποροῦμε νά χρησιμοποιήσουμε τό διάστημα (L,U) μέ τιμή (l,u) γιά τήν ἐκτίμηση τοῦ θ. Δύο πράγματα πρέπει νά ζητᾶμε: (*i*) τό διάστημα (L,U) νά περιέχει τήν ἀληθινή τιμή τοῦ θ ἕνα μεγάλο "π ο σ ο σ τ ό φ ο ρ ῶ ν" καί (*ii*) τό διάστημα νά ἔχει ὅσο τό δυνατόν μικρότερο μῆκος. Τό "ποσοστό φορῶν" πού ἕνα διάστημα (L,U) περιέχει τό θ λέγεται ἐ π ί π ε δ ο ἤ β α θ μ ό ς ἐ μ π ι σ τ ο σ ύ ν η ς, συμβολίζεται μέ $100(1-\alpha)\%$ καί φανερώνει τήν πιθανότητα τό (L,U) νά περιέχει τό θ, δηλαδή

$$\text{βαθμός ἐμπιστοσύνης} = P(L<\theta<U) = 1-\alpha \quad . \qquad (9.6)$$

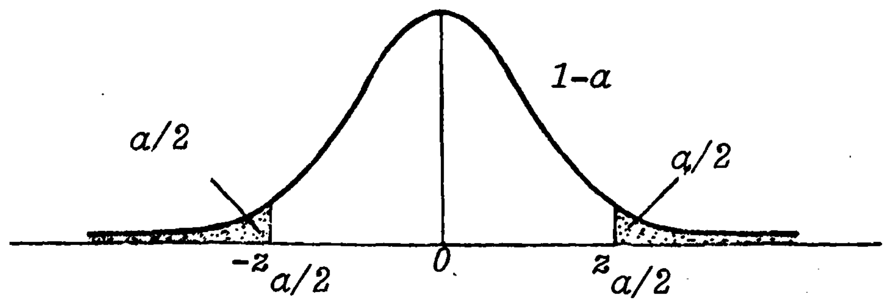
Τό διάστημα (L,U) λέγεται **διάστημα ἐμπιστοσύνης** (Δ.Ε).

Γιά νά διευκρινήσουμε τίς παραπάνω ἔννοιες θά κατασκευάσουμε ἕνα $100(1-\alpha)\%$ διάστημα ἐμπιστοσύνης γιά τή μέση τιμή μ ἑνός κανονικοῦ πληθυσμοῦ μέ βάση ἕνα τυχαῖο δεῖγμα μεγέθους n καί μέ τήν ὑπόθεση ὅτι ἡ διακύμανση σ^2 τοῦ πληθυσμοῦ εἶναι γνωστή. Ἡ τελευταία αὐτή ὑπόθεση εἶναι ἐξωπραγματική. Διευκολύνει ὅμως στήν ἀρχή τήν κατανόηση τῆς βασικῆς διαδικασίας. Διάστημα ἐμπιστοσύνης στήν περίπτωση πού τό σ^2 εἶναι ἄγνωστο δίδεται ἀργότερα.

Ἔστω $\overline{X}$ ἡ μέση τιμή τοῦ δείγματος. Γνωρίζουμε ὅτι τό $\overline{X}$ ἔχει κανονική κατανομή $N(\mu,\sigma^2/n)$. Ἔτσι γιά κάθε μ

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1). \tag{9.7}$$

Ἔστω $z_{\alpha/2}$ τό ἀντίστροφο ἑκατοστιαῖο σημεῖο τῆς τυπικῆς κανονικῆς κατανομῆς πού ὁρίστηκε στή §8.3, σχέση (8.17). Τότε ἔχουμε (βλ. Σχῆμα 9.1)



Σχῆμα 9.1 Τυπική κανονική κατανομή

$$P\{-z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\} = 1-\alpha \qquad (9.8)$$

ή

$$P\{-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\} = 1-\alpha$$

ή

$$P\{\overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\} = 1-\alpha \qquad (9.9)$$

διότι τά ἐνδεχόμενα στίς δύο τελευταῖες σχέσεις εἶναι ἰσοδύναμα μέ τό ἐνδεχόμενο $|\overline{X}-\mu|<z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. Ἔτσι τό διάστημα (L,U) μέ ὅρια

$$L = \overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{καί} \quad U = \overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad (9.10)$$

ἔχει τήν ἰδιότητα

$$P\{L<\mu<U\} = 1-\alpha \qquad (9.11)$$

γιά κάθε μ. Τό διάστημα αὐτό εἶναι διάστημα ἐμπιστοσύνης (μέ ὅρια ἐμπιστοσύνης τά L καί U) γιά τήν παράμετρο μ μέ βαθμό ἐμπιστοσύνης $100(1-\alpha)\%$.

Παρατηροῦμε ὅτι τά ὅρια (9.10) εἶναι τυχαῖες μεταβλητές καί τό διάστημα (L,U), μέ τήν ἰδιότητα (9.11) γιά κάθε μ, εἶναι τυχαῖο διάστημα. Παρατηροῦμαι ἐπίσης ὅτι τά σημεῖα $-z_{\alpha/2}$ καί $z_{\alpha/2}$ τῆς (9.8) δέν εἶναι τά μόνα πού ἱκανοποιοῦν τή σχέση (9.8). Ὑπάρχουν ἄπειρα σημεῖα z_1, z_2 τέτοια ὥστε

$$P\{z_1<z<z_2\} = 1-\alpha .$$

Ἄρα τά Δ.Ε. σταθεροῦ βαθμοῦ ἐμπιστοσύνης δέν εἶναι μονοσήμαντα. Μέ τήν ἀπαίτηση τοῦ ἐλάχιστου μήκους γίνονται μονοσήμαντα ὅπως στήν παραπάνω περίπτωση μέ $z_1 = -z_{\alpha/2}$ καί $z_2 = z_{\alpha/2}$.

Παράδειγμα 9.1:

Ἕνα διεγερτικό φάρμακο ἐλέγχεται γιά τήν ἐπίδραση του στήν πίεση τοῦ αἵματος. Οἱ πιέσεις αἵματος εἴκοσι ἀτόμων μετροῦνται πρίν ἀπό τή λήψη καί μισή ὥρα μετά τή λήψη τοῦ φαρμάκου καί λαμβάνονται οἱ ἀκόλουθες διαφορές:

$$7,\ 6,\ 0,\ 8,\ -9,\ -4,\ 0,\ 1,\ -9,\ 1$$

$$2,\ 7,\ 0,\ 6,\ -6,\ -5,\ -1,\ 6,\ -2,\ 4\ .$$

Ἀπό προηγούμενες μελέτες εἶναι γνωστό ὅτι ἡ πρό καί μετά τή λήψη τοῦ φαρμάκου διαφορά πιέσεων ἀκολουθεῖ κανονική κατανομή μέ $\sigma^2=25$. Νά κατασκευασθεῖ ἕνα 95% διάστημα ἐμπιστοσύνης τῆς μέσης διαφορᾶς μ πιέσεων αἵματος.

Πράγματι ἔχουμε: $\overline{X}=0,6$, $Z_{0,025}=1,96$, $\sigma=5$, $n=20$. Ἔτσι τά ὅρια ἐμπιστοσύνης εἶναι

$$L = \overline{X}-1,96\times \frac{5}{\sqrt{20}} = 0,6-2,19 = -1,59$$

$$U = \overline{X}+1,96\times \frac{5}{\sqrt{20}} = 0,6+2,19 = 2,79.$$

Ἄρα ἕνα διάστημα ἐμπιστοσύνης τῆς μέσης διαφορᾶς πιέσεων πρίν καί μετά τό φάρμακο μέ βαθμό ἐμπιστοσύνης 95% εἶναι τό διάστημα $(-1,59,\ 2,79)$.

Ποτέ δέν γράφουμε

$$P\{-1,59<\mu<2,79\} = 0,95$$

διότι μετά τό πείραμα ἡ ἀληθινή τιμή τοῦ μ εἴτε εὑρίσκεται ἐντός τοῦ διαστήματος $(-1,59,\ 2,79)$ εἴτε ἐκτός καί ἡ παραπάνω πιθανότητα εἴτε εἶναι 1 εἴτε 0. Λέμε ὅμως ὅτι ἔχουμε 95% ἐμπιστοσύνη ὅτι $-1,59<\mu<2,79$. Αὐτό εἶναι ἕνας τρόπος ἐκφράσεως τῆς σχέσεως (9.11). Ἡ ἐμπιστοσύνη πού ἔχουμε στά ὅρια $-1,59$ καί $2,79$ πηγάζει στήν πραγματικότητα ἀπό τήν ἐμπιστοσύνη μας στή σ τ α τ ι σ τ ι κ ή δ ι α δ ι κ α σ ί α πού ἔδωσε τά ὅρια αὐτά. Αὐτή ἡ διαδικασία ἔδωσε τυχαῖες μεταβλητές L καί U πού ἔχουν 95% πιθανότητα σέ ἐπαναληπτική δειγματοληψία νά περικλείουν τό ἀληθινό μ· ἄν οἱ είδικές τιμές τους $-1,59$ καί $2,79$ στή πραγματικότητα περικλείουν τό ἀληθινό μ δέν ἔχουμε τρόπο νά τό γνωρίζουμε.

Μιλήσαμε γιά τήν ἔννοια τοῦ διαστήματος ἐμπιστοσύνης καί δώσαμε ἕνα τέτοιο διάστημα γιά τό ἁπλούστερο ἀπό τά στατιστικά μοντέλα, τήν κανονική κατανομή μέ μ ἄγνωστο καί σ^2 γνωστό. Ἀπό τόν τρόπο εὑρέσεως τοῦ διαστήματος αὐτοῦ διαφαίνεται ἡ ἀκόλουθη γενική ἰδέα εὑρέσεως διαστημάτων ἐμπιστοσύνης: ἀρκεῖ νά εὕρουμε μία τυχαία ποσότητα ὅπως ἡ (9.7) μέ γνωστή κατανομή· τό διάστημα ἐμπιστοσύνης εὑρίσκεται τότε ἀνάλογα μέ τίς σχέσεις (9.8) καί (9.9) δηλαδή ἀντιστρέφοντας τίς ἀνισότητες καί ἀπομονώνοντας τήν ἄγνωστη παράμετρο.

Πράγματι ἡ ἰδέα αὐτή ἀποδίδει γιά πολλά στατιστικά μοντέλα. Οἱ λεπτομέρειες τοῦ θέματος αὐτοῦ μαζί μέ τή γενική θεωρία τῶν διαστήματων ἐμπιστοσύνης καί τόν τρόπο εὑρέσεως διαστημάτων ἐλαχίστου μήκους ἀποτελοῦν ἀντικείμενο παραπέρα μελέτης σέ προχωρημένα βιβλία τῆς Στατιστικῆς.

Β΄ΕΛΕΓΧΟΣ ΣΤΑΤΙΣΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ
(Στατιστικά Τέστ)

9Β. 1 Τὸ στατιστικὸ τέστ-Οἱ πρῶτες ἔννοιες

Θά ξεκινήσουμε μέ ἕνα παράδειγμα πού περιέχει ὅλα τά στοιχεῖα τοῦ θέματος.

Τό λεγόμενο "τριπλό τέστ" μᾶς δίνει ἕνα τρόπο ἐλέγχου τῶν ἱκανοτήτων ἑνός ὑποψηφίου "δοκιμαστοῦ κρασιῶν" νά διακρίνει λεπτές διαφορές στή γεύση, ἄρωμα καί ὑφή διαφόρων κρασιῶν. Τό τέστ ἐφαρμόζεται παρουσιάζοντας στόν ὑποψήφιο τρία δείγματα κρασιῶν, δύο ἀπό τά ὁποῖα προέρχονται ἀπό τό ἴδιο κρασί, καί ζητῶντας του ἀφοῦ τά δοκιμάσει νά βρεῖ τό διαφορετικό κρασί. Τά δείγματα εἶναι ὅσο τό δυνατό ὅμοια καί ἡ σειρά παρουσιάσεως τυχαία. Ἄν ὁ ὑποψήφιος δέν ἔχει πράγματι τήν ἱκανότητα νά ξεχωρίζει τά κρασιά, τότε ἔχει πιθανότητα $1/3$ νά βρεῖ τό διαφορετικό κρασί καί τό ἐρώτημα εἶναι ἄν τό ἄτομο μπορεῖ νά φέρει καλύτερα ἀποτελέσματα.

Γιά νά ἐλέγξουμε τήν ἱκανότητα τοῦ ὑποψηφίου δέν ἀρκεῖ ἕνα μόνο "τριπλό τέστ". Τοῦ δίνομε μία σειρά τέτοιων τέστ μέ δείγματα παρουσιαζόμενα κάθε φορά στήν τυχη γιά νά ἐξαλείψουμε τήν περίπτωση συστηματικῆς μεροληψίας καί ἐξασφαλίσουμε ἀνεξαρτησία ἀπό τέστ σέ τέστ. Ἔστω p ἡ πιθανότητα ὁ ὑποψήφιος νά βρεῖ σωστά τό διαφορετικό κρασί σέ μία ἁπλή ἐφαρμογή τοῦ τριπλοῦ τέστ (δοκιμή). Ἄν ὁ ὑποψήφιος ἔχει τήν ἱκανότητα διακρίσεως κρασιῶν τότε $p>1/3$, ἄν δέν ἔχει καμμία ἱκανότητα τότε $p=1/3$. Δεχόμαστε ὅτι τό ἄτομο δέν μπορεῖ νά ἔχει ἐπίδοση χειρότερη ἀπό τήν τύχη,δηλαδή ἀποκλείουμε ἐκ τῶν προτέρων τήν περίπτωση $p<1/3$. Δεχόμαστε ἐπίσης ὅτι ἡ πιθανότητα p εἶναι σταθερή γιά τό ἄτομο καί δέν μεταβάλλεται μέ τήν πάροδο τῶν τέστ. Τό p, $0\leq p\leq 1$, εἶναι ἡ ἄγνωστος στατιστική παράμετρος καί κατά κάποιο τρόπο εἶναι τό κριτήριο τῆς διακριτικῆς ἱκανότητας τοῦ ὑποψηφίου.

Ὁρίζουμε τίς δύο προηγούμενες ὑποθέσεις ὡς ἀκολούθως

Μ η δ ε ν ι κ ή ὑπόθεση $H_0: p=1/3$

Ἐ ν α λ λ α κ τ ι κ ή ὑπόθεση $H_a: p>1/3$.

Ἔτσι ἀντιμετωπίζουμε τό πρόβλημα ἐλέγχου τῆς μηδενικῆς ὑποθέσεως H_0 ἔναντι τῆς ἐναλλακτικῆς H_a. Ὑποθέτουμε ἐπί πλέον ἐκ τῶν προτέρων ὅτι δύο εἶναι οἱ δυνατές περιπτώσεις: εἴτε ἀληθεύει ἡ H_0 εἴτε ἀληθεύει ἡ H_a καί ὅτι ἡ ἐκλογή μας εἶναι μεταξύ αὐτῶν.

Ὁ ἔ λ ε γ χ ο ς σ τ α τ ι σ τ ι κ ῶ ν ὑ π ο -

θ έ σ ε ω ν μᾶς δίδει τόν τρόπο νά κρίνομε ἤ ἀποφασίσομε γιά τήν ὀρθότητα ἤ ἀποδοχή τῆς H_0. Ἡ ὅλη διαδικασία ὀνομάζεται σ τ α τ ι σ τ ι κ ό τ έ σ τ .

Γιά νά γίνει ὁ στατιστικός ἔλεγχος τῶν H_0 καί H_a πρέπει νά χρησιμοποιήσομε τίς πληροφορίες γιά τό p πού μᾶς δίνουν τά στατιστικά δεδομένα, δηλαδή τό τυχαῖο δεῖγμα. Ἄν κάνομε n δοκιμές μέ τόν ὑποψήφιο δοκιμαστή καί x εἶναι ὁ ἀριθμός τῶν σωστῶν ἀναγνωρίσεων πού κάνει, τότε ἡ τ.μ. X ἔχει διωνυμική κατανομή μέ παραμέτρους τό n καί κάποιο ἄγνωστο p:

$$P(X=x) = \binom{n}{x} p^x q^{n-x}, \quad x=0,1,\ldots,n .$$

Θέλουμε νά χρησιμοποιήσομε τό στατιστικό X γιά νά ἀποφασίσομε ἄν ἡ ἐναλλακτική ὑπόθεση H_a εἶναι πιό λογική ἤ παραδεκτή ἀπό τή μηδενική ὑπόθεση H_0, δηλαδή ἄν ὁ ὑποψήφιος ἔχει τήν ἱκανότητα διακρίσεως κρασιῶν ἤ ὄχι. Μέ ἄλλα λόγια ἄν ἡ H_0 ἀληθεύει ἤ ὄχι. Θέλομε νά κατασκευάσουμε, πρίν γίνει ὁ πειραματισμός, ἕνα κανόνα γιά νά πάρουμε ἀπόφαση. Ὁ κανόνας αὐτός εἶναι τό στατιστικό τέστ. Ἀπό ἄλλη σκοπιά θέλομε νά δοῦμε μέ τή βοήθεια ἑνός στατιστικοῦ τέστ ἄν τά ἀποτελέσματα τοῦ πειράματος στηρίζουν τήν H_0 ἤ τήν H_a. Αὐτά γίνονται κατασκευάζοντας μιά π ε ρ ι ο χ ή ἀ π ο ρ ρ ί ψ ε ω ς ἤ κ ρ ί σ ι μ η π ε ρ ι ο χ ή (κ.π.) C, δηλαδή ἕνα σύνολο τιμῶν τοῦ X πού θά μᾶς ὁδηγήσει στό νά ἀπορρίψομε τῶν H_0 καί νά προτιμήσουμε τήν H_a ἀπό τήν H_0. Ἄν τό πείραμα δώσει στό X μία ἀριθμητική τιμή πού βρίσκεται ἐντός τοῦ C, τότε ἀπορρίπτομε τήν H_0 διαφορετικά δέν ἀπορρίπτομε τήν H_0.

Γιά έπεξήγηση έστω ότι $n=10$. Θεωροῦμε δύο κρίσιμες περιοχές:

$$C_5 = \{5, 6, 7, 8, 9, 10\}$$

καί

$$C_8 = \{8, 9, 10\} \quad .$$

Ἄν χρησιμοποιήσομε τῆς C_5 τό στατιστικό τέστ εἶναι:

ἀπορρίπτομε τήν H_0 καί δεχόμαστε τήν H_a (δηλαδή ὑπάρχουν ἐνδείξεις ὅτι ὁ ὑποψήφιος ἔχει τήν ἱκανότητα διακρίσεως ἄν $X \geq 5$.

δεχόμαστε τήν H_0 (ὁ ὑποψήφιος δέν ἔχει τήν ἱκανότητα) ἄν $X<5$.

Ὁμοίως τό τέστ μέ κρίσιμο περιοχή τήν C_8 εἶναι:

ἀπορρίπτομε τήν H_0 ἄν $X \geq 8$ καί δεχόμαστε τήν H_0 ἄν $X<8$.

Οἱ παραπάνω κρίσιμες περιοχές εἶναι εὔλογες διότι προφανῶς ὅσο πιό μεγάλο εἶναι τό X τόσο μεγαλύτερη ἔνδειξη ἔχομε ὅτι $p>1/3$ καί ἐπομένως πρέπει νά ἀπορρίψωμε τήν H_0. Ἀντίθετα ἡ κρίσιμη περιοχή $C=\{1,3,5,7,9\}$ δέν εἶναι εὔλογη. Ἔτσι ἔχομε δύο προβλήματα: πῶς κατασκευάζομε κρίσιμες περιοχές καί πῶς τίς ἀξιολογοῦμε. Θά ἀναπτύξομε πρῶτα τό δεύτερο πρόβλημα.

9Β 2 Ἀξιολόγηση τῶν τέστ-Σφάλματα τύπου Ι καί τύπου ΙΙ

Ποιά ἀπό τίς κρίσιμες περιοχές C_5 καί C_8 εἶναι καλύτερη; Ὅταν κάνομε ἕνα στατιστικό τέστ μπορεῖ νά καταλήξουμε σέ λανθασμένη ἀπόφαση. Ἔτσι ἔχομε δύο εἴδη δυνατῶν σφαλμάτων, τά σφάλματα Τύπου I καί Τύπου II:

<u>Σφάλμα Τύπου I</u>: 'Απορρίπτουμε τήν H_0 ὅταν ἡ H_0 ἀληθεύει
<u>Σφάλμα Τύπου II</u>: Δεχόμαστε τήν H_0 ὅταν ἡ H_a ἀληθεύει.

Διαγραμματικά τά σφάλματα μποροῦν νά παρασταθοῦν μέ τόν ἀκόλουθο πίνακα:

	Στατιστικός (συμπέρασμα)	
	ἀποδέχεται τήν H_0	ἀπορρίπτει τήν H_0
Φύση (πραγματικότης): ἀληθεύει ἡ H_0	0 πιθανότητα = $1-\alpha$	Σφάλμα Τύπου I πιθανότητα = α
Φύση (πραγματικότης): ἀληθεύει ἡ H_a	Σφάλμα Τύπου II πιθανότητα = β	0 πιθανότητα = $1-\beta$

Σφάλματα Στατιστικῶν Τέστ

'Αρχίζοντας μέ τό ἄνω ἀριστερά τετραγωνίδιο καί πηγαίνοντας πρός τή φορά τῶν δεικτῶν τοῦ ὡρολογίου ἔχομε: ὅταν ἡ H_0 ἀληθεύει καί ὁ Στατιστικός ἀποδεχθεῖ τήν H_0 δέν γίνεται σφάλμα, ἐνῶ ἄν ὁ Στατιστικός ἀπορρίψει τήν H_0 τότε γίνεται σφάλμα Τύπου I. 'Ομοίως ὅταν ἀληθεύει ἡ H_a καί ὁ Στατιστικός ἀποδειχθεῖ τήν H_0 γίνεται σφάλμα Τύπου II, ἐνῶ στήν ἀντίθετη περίπτωση δέν γίνεται σφάλμα. Τά δύο σφάλματα εἶναι προφανῶς τυχαῖα ἐνδεχόμενα.

'Η ἐπίδοση ἑνός τέστ ἤ κρίσιμης περιοχῆς ἀξιολογεῖται μέ τίς πιθανότητες τῶν σφαλμάτων τῶν παραπάνω τύπων:

$$\alpha = P(\text{Σφάλμα Τύπου I}) = P(\text{'Απορρίπτεται ἡ } H_0 | H_0 \text{ ἀληθινή})$$

$$\beta = P(\text{Σφάλμα Τύπου II}) = P(\text{Δεκτή ἡ } H_0 | H_0 \text{ λανθασμένη}). \tag{9.12}$$

'Η πιθανότητα α τοῦ Σφάλματος Τύπου I λέγεται καί μέγεθος τῆς κρίσιμης περιοχῆς ἤ ἐπίπεδο σημαντικότητας τοῦ στατιστικοῦ τέστ. Τό $\gamma=1-\beta$ λέγεται ἰσχύς ἤ δύναμη τοῦ τέστ. Εἶναι φανερό ὅτι θέλουμε τά α καί β νά εἶναι ὅσο τό δυνατό μικρότερα.

Γιά τήν κρίσιμη περιοχή C_5, σφάλμα τύπου I γίνεται ὅταν ὁ δοκιμαστής δέν ἔχει τήν διακριτική ἱκανότητα ἀλλά ἐπιτυγχάνει 5 ἤ περισσότερες ἐπιτυχεῖς ἀναγνωρίσεις ἁπλῶς στήν τύχη. Ἔτσι ἀπορρίπτομε τήν H_0 λανθασμένα. 'Η πιθανότητα νά γίνει αὐτό εἶναι:

$$\begin{aligned}
\alpha(C_5) &= P\,(\text{ἀπορρίπτεται ἡ } H_0 | H_0 \text{ ἀληθινή}) \\
&= P(X \geq 5)\,|p=1/3) = \sum_{x=5}^{10} \binom{10}{x} \left(\tfrac{1}{3}\right)^x \left(\tfrac{2}{3}\right)^{10-x} \\
&= \frac{1}{3^{10}} \left[\binom{10}{5} 2^5 + \binom{10}{6} 2^4 + \ldots + \binom{10}{9} 2 + 1 \right] \\
&= 0{,}2131\ .
\end{aligned}$$

Γιά τήν κρίσιμη περιοχή C_8 ἡ πιθανότητα τοῦ σφάλματος τύπου I εἶναι:

$$\alpha(C_8) = P(X \geq 8\,|p=\tfrac{1}{3}) = \sum_{x=8}^{10} \binom{10}{x} \left(\tfrac{1}{3}\right)^x \left(\tfrac{2}{3}\right)^{10-x} = 0{,}0034\ .$$

Ἔτσι τό $\alpha(C_8)$ εἶναι μικρότερο τοῦ $\alpha(C_5)$. Ὅσον ἀφορᾶ τό σφάλμα τύπου I τό τέστ C_8 εἶναι καλύτερο ἀπό τό τέστ C_5.

Ἄς ὑπολογίσουμε τώρα τίς πιθανότητες τοῦ σφάλματος τύπου II. Ὅταν ἡ H_0 εἶναι λανθασμένη, δηλ. $p>1/3$, τό β ἐξαρτᾶται ἀπό τή συγκεκριμένη τιμή τοῦ p πού θά θεωρήσουμε. Ἔτσι ἔστω $p=0,70$. Τότε γιά τίς κρίσιμες περιοχές C_5 καί C_8 ἔχομε

$$\beta(C_5,p=0,70) = P\text{ (Δεκτή ἡ } H_0|p=0,7)$$

$$= P(X<5|p=0,7)= \sum_{x=0}^{4} \binom{10}{x}(0,7)^x(0,3)^{10-x}$$

$$= 0,0473$$

$$\beta(C_8,p=0,70) = P(X<8|p=0,7)= \sum_{x=0}^{7} \binom{10}{x}(0,7)^x(0,3)^{10-x}$$

$$= 0,6171 ,$$

δηλαδή $\beta(C_8,p=0,70)>\beta(C_5,p=0,70)$. Ἔτσι ὡς πρός τό σφάλμα τύπου II τό τέστ C_8 εἶναι χειρότερο τοῦ C_5.

Βλέπουμε δηλαδή ὅτι μέ γνώμονα τό σφάλμα τύπου I ἡ κρίσιμη περιοχή C_8 εἶναι προτιμώτερη ἀπό τήν C_5 καί ἀντίστροφα ἡ C_5 εἶναι προτιμώτερη ἀπό τήν C_8 ὅταν ὁ γνώμονας εἶναι τό σφάλμα τύπου II. Αὐτό εἶναι ἀναπόφευκτο: καθώς τό α μικραίνει τό β μεγαλώνει καί ἀντίστροφα ὅταν τό α μεγαλώνει τό β μικραίνει. Στήν ἀκραία περίπτωση πού ποτέ δέν κάνομε σφάλμα Τύπου I, δηλαδή ποτέ δέν ἀπορρίπτομε τήν H_0, ἡ κρίσιμη περιοχή C εἶναι τό κενό σύνολο καί τό $\alpha=0$. Ἀλλά τότε πάντοτε δεχόμαστε τήν H_0 καί $\beta=1$. Στήν ἄλλη ἀκραία περίπτωση πού ποτέ δέν κάνομε σφάλμα Τύπου II, δηλαδή ποτέ δέν δεχόμαστε τήν $H_0, \beta=0$. Ἀλλά τότε πάντοτε ἀπορρίπτομε τήν H_0 καί $\alpha=1$. Εἶναι ἀδύνατον τά α καί β νά γίνουν καί τά δύο 0 ἤ καί τά δύο 1 γιά μιά συγκεκριμένη κρίσιμη περιοχή.

Μέ ἄλλα λόγια εἶναι ἀδύνατη ἡ ταυτόχρονη, δηλαδή γιά τό ἴδιο C, ἐλαχιστοποίηση τῶν $\alpha(C)$ καί $\beta(C)$. Τό πρόβλημα αὐτό παρακάμπτεται κρατῶντας τό α σταθερό,- καθορισμένο ἐκ τῶν προτέρων καί ἐπιδιώκοντας νά βροῦμε ἐκείνη τήν κρίσιμη περιοχή C μέ μέγεθος μικρότερο ἤ ἴσο τοῦ α καί ἐλάχιστο β [ἤ ἰσοδύναμα, τή μέγιστη ἰσχύ $\gamma=1-\beta$]. Τό σταθερό προκαθορισμένο α λέγεται ἐπίπεδο σημαντικότητος τοῦ τέστ καί συνήθως παίρνει τιμές 1%, 5%, ἤ 10%. Τό τέστ λέγεται καί τέστ ἐπιπέδου ἤ μεγέθους α.

Ἔτσι στό παράδειγμα τοῦ "τριπλοῦ τέστ" ἄν πάρουμε τό $\alpha = 5\%$ εἶναι εὔκολο νά δοῦμε ὅτι ἡ κρίσιμη περιοχή

$$C_7 = \{7, 8, 9, 10\}$$

εἶναι ἡ μεγαλύτερη κρίσιμη περιοχή μέ

$$\alpha(C_7) = 0{,}0196<0{,}05 .$$

Μπορεῖ ἐπί πλέον νά ἀποδειχθεῖ ὅτι ἡ C_7 εἶναι ἡ ἰσχυρότερη (ἄριστη) κρίσιμη περιοχή ἐπιπέδου σημαντικότητας 5% γιά τόν ἔλεγχο τῆς H_0 ἔναντι τῆς H_α. Δηλαδή

$\beta(C_7,p) \leqq \beta(C,p)$ γιά κάθε C μέ $\alpha(C) \leqq 0,05$
καί κάθε $p > 1/3$.

9Β. 3 Τὰ στοιχεῖα τοῦ τέστ

Ἀπό τά παραπάνω εἶναι φανερό ὅτι κάθε στατιστικό τέστ ἀποτελεῖται ἀπό τά ἀκόλουθα στοιχεῖα:

1. Τήν μηδενική H_0 καί τήν ἐναλλακτική H_α ὑπόθεση.
2. Τό προκαθορισμένο ἐπίπεδο σημαντικότητας α.
3. Τήν στατιστική συνάρτηση τοῦ τέστ καί τήν κρίσιμο περιοχή (κ.π.)C.
4. Τήν τιμή τῆς στατιστικῆς συνάρτησης τοῦ τέστ πού προκύπτει ἀπό τά δεδομένα (μετρήσεις).
5. Τό συμπέρασμα: ἄν ἡ τιμή τῆς στατιστικῆς συνάρτησης τοῦ τέστ πέσει μέσα στήν κρίσιμη περιοχή C ἡ H_0 ἀπορρίπτεται ὑπέρ τῆς H_α· ἄν πέσει ἐκτός ἡ H_0 δέν ἀπορρίπτεται.

Γιά κάθε ἕνα ἀπό τά παραπάνω στοιχεῖα ἔχουμε τίς ἀκόλουθες παρατηρήσεις:

Στοιχεῖο 1. Οἱ ὑποθέσεις H_0 καί H_α διατυπώνονται μέ τή βοήθεια τῶν στατιστικῶν παραμέτρων τοῦ μοντέλου πού υἱοθετεῖται γιά τήν ἀνάλυση τῶν μετρήσεων.

Διακρίνονται σέ ά π λ έ ς (π.χ. $p = 1/3$) καί σ ύ ν θ ε τ ε ς (π.χ. $p>1/3$). Διακρίνονται έπίσης καί σέ μ ο ν ό π λ ε υ ρ ε ς (π.χ. $p>1/3$) καί δ ί π λ ε υ ρ ε ς (π.χ. $p\neq 1/3$). Ὑποθέσεις τύπου $\theta>\theta_0$, ὅπου θ_0 γνωστό λέγονται δ ε ξ ι ό π λ ε υ ρ ε ς καί ὑποθέσεις τύπου $\theta<\theta_0$ λέγονται ά ρ ι σ τ ε ρ ό π λ ε υ ρ ε ς.

Παρατήρηση: Ἡ μηδενική ὑπόθεση εἶναι συνήθως τό "ἀντίθετο" τοῦ ἰσχυρισμοῦ πού θέλομε νά ἐλέγξουμε. Μηδενοποιεῖ τόν ἰσχυρισμό αὐτό. Ἔτσι στό παράδειγμα τοῦ "τριπλοῦ τέστ" ἄν ὁ ἰσχυρισμός τοῦ ὑποψηφίου εἶναι ὅτι ἔχει τήν ἱκανότητα διακρίσεως (δηλαδή $p>1/3$) τότε ἡ H_0 εἶναι $p=1/3$ καί H_a:$p>1/3$. Ἄν κάνομε σφυγμομέτρηση τῆς κοινῆς γνώμης μέ σκοπό νά ἀποδείξομε ὅτι λιγότεροι τοῦ 30% τῶν πολιτῶν ὑποστηρίζουν κάποιο θέμα τότε ἡ H_0 εἶναι $p\geq 0,3$ καί ἡ H_a εἶναι $p<0,3$, ὅπου p εἶναι τό ποσοστό τῶν πολιτῶν πού ὑποστηρίζουν τό θέμα. Ὅταν θέλομε νά ἐλέγξομε τήν ὑπόθεση ὅτι ἡ μακρόχρονη ἀποθήκευση τοῦ αἵματος ἤ πλάσματος σέ καταψύκτη δέν ἀλλάζει τήν περιεκτικότητα του σέ χοληστερίνη, τριγλυκερίδια καί ἄλλες ουσίες, ἡ H_0 εἶναι $\mu=0$ καί ἡ H_a εἶναι $\mu\neq 0$, ὅπου μ εἶναι ἡ μέση διαφορά μετρήσεων περιεκτικότητος φρέσκου αἵματος καί τοῦ ἴδιου αἵματος μετά ἀπό ὁρισμένους μῆνες ἀποθηκεύσεως.

Σ τ ο ι χ ε ῖ ο 2. Τά α ἐκφράζει τό ποσοστό (πιθανότητα) σφάλματος Τύπου Ι πού εἴμαστε διατεθειμένοι νά ἀνεχτοῦμε.

Σ τ ο ι χ ε ῖ ο 3. Ἡ στατιστική συνάρτηση τοῦ τέστ καί ἡ κρίσιμη περιοχή εἶναι τά κύρια στοιχεῖα τοῦ τέστ. Στήν πιό γενική περίπτωση ἡ κ.π. τοῦ τέστ εἶναι ἕνα ὑποσύνολο τοῦ δειγματικοῦ χώρου S, ὁ ὁποῖος συνήθως εἶναι πολυδιάστατος. Ἐπειδή ὅμως συχνά χρησιμοποιοῦνται μετασχηματισμοί ἀπό τό S στήν εὐθεῖα γραμμή ἡ κ.π. δίδεται μέ τή βοήθεια ἰσοδύναμων μονοδιάστατων κρίσιμων περιοχῶν. Οἱ μετασχηματισμοί αὐτοί εἶναι οἱ στατιστικές συναρτήσεις τῶν τέστ. Ἔτσι ἡ σ.σ. τοῦ τέστ εἶναι μία τ υ χ α ί α π ο σ ό τ η ς μ έ γ ν ω σ τ ή κ α τ α ν ο μ ή ὅ τ α ν ἡ H_0 ἀ λ η θ ε ύ ε ι. Ἡ μορφή τῆς H_a καθορίζει τήν θέση τῆς κρίσιμης περιοχῆς στόν ἄξονα τῶν x. Ἄν ἡ H_a εἶναι δεξιόπλευρη ἡ κρίσιμη περιοχή τοῦ τέστ εἶναι συνήθως τῆς μορφῆς $[c_0,+\infty)$, ἄν H_a εἶναι ἀριστερόπλευρη ἡ κ.π. τοῦ τέστ εἶναι τῆς μορφῆς $(-\infty,c_0]$ καί τέλος ἄν ἡ H_a δίπλευρη ἡ κ.π. εἶναι τῆς μορφῆς $(-\infty,c_1]U[c_2,+\infty)$, ὅπου c_0,c_1,c_2 γνωστοί ἀριθμοί. Τά c_0, c_1,c_2 ὑπολογίζονται ἀπό τήν κατανομή τῆς σ.σ. τοῦ τέστ ὅταν ἀληθεύει ἡ H_0. Ἀνάλογα τά τέστ λέγονται δεξιόπλευρα, ἀριστερόπλευρα καί δίπλευρα ἤ τέστ δεξιᾶς οὐρᾶς, ἀριστερᾶς οὐρᾶς, δύο οὐρῶν τῆς κατανομῆς.

Σ τ ο ι χ ε ῖ ο 4: Οἱ μετρήσεις (τό τυχαῖο δεῖγμα) πρέπει νά "ταιριάζουν" στό στατιστικό μοντέλο πού υἱοθετήθηκε. Στό παράδειγμά μας, ἔχομε πράγματι μετρήσεις ἀπό διωνυμικό πείραμα.

Σ τ ο ι χ ε ῖ ο 5: Ὅταν ἀπορρίπτομε τήν H_0, τότε εἴτε τήν ἀπορρίπτομε σωστά εἴτε λανθασμένα. Στή δεύτερη περίπτωση τό σφάλμα μας ἐλέγχεται ἀπό τό α, ἔχει δηλαδή πιθανότητα μικρότερη ἤ ἴση τοῦ α, τό ὁποῖο εἶναι γνωστό προκαθορισμένο μέ μικρή τιμή. Όταν δέν άπορρίπτομε τήν H_0 πάλι εἴτε ένεργοῦμε σωστά εἴτε λανθασμένα. Στή δεύτερη περίπτωση τό σφάλμα μας ἐλέγχεται ἀπό τό β, τό ὁποῖο ὅμως εἶναι ἄγνωστο διότι ἐξαρτᾶται ἀπό τίς τιμές τοῦ θ πού άνήκουν στό H_a. Ἔτσι δέν ξέρομε ἄν τό σφάλμα μας ἔχει μικρή ἤ μεγάλη πιθανότητα. Στό παράδειγμα μας μέ τό "τριπλό τέστ" καί μέ κρίσιμη περιοχή τό C_7 καί $\alpha=0,05$ ἄν τό p εἶναι λίγο μεγαλύτερο ἀπό τό $1/3$, τό β εἶναι σχεδόν ἴσο μέ τό $1-\alpha(C_7)=1-0,0196=0,9804\approx 95\%$· ἐνῶ ἄν τό p εἶναι $0,95$ τότε $\beta=0,0011$. Γενικά οἱ τιμές τοῦ β εἶναι μικρές γιά μεγάλο p ἀλλά σχεδόν ἴσες μέ $1-\alpha$ γιά p κοντά στό $1/3$. Ἄν ἡ τιμή τῆς σ.σ. τοῦ τέστ πέσει μέσα στήν κ.π. δέν μποροῦμε μέ ἐμπιστοσύνη νά ἀποδεχθοῦμε τήν H_0 διότι δέν μποροῦμε μέ ἐμπιστοσύνη ν' ἀπορρίψουμε "ὁλόκληρη" τήν H_a. Ἔτσι ὑπάρχει μία ἀσυμμετρία στόν τρόπο μεταχειρήσεως τῶν H_0 καί H_a.

Ἀπό τά παραπάνω βγαίνει τό ἀκόλουθο συμπέρασμα: ε ἶ ν α ι π ρ ο τ ι μ ώ τ ε ρ ο ν ά ἀ π ο ρ ρ ί π τ ο μ ε τ ή ν H_0 π α ρ ά ν ά τ ή ν ἀ π ο δ ε χ ό μ ε θ α. Ὅταν ἀπορρίπτομε τήν H_0 τό ποσοστό σφάλματος εἶναι ἐκ τῶν προτέρων γνωστό ἤ εἶναι μικρότερο ἤ ἴσο ἑνός γνωστοῦ ὁρίου, τοῦ α. Γιά τούς ἴδιους λόγου ἀποφεύγουμε νά λέμε ὅτι "ἀποδεχόμαστε" τήν H_0 καί χρησιμοποιοῦμε τόν ὅρο "δέν ἀπορρίπτεται ἡ H_0", ἤ λέμε ὅτι τά στατιστικά δεδομένα δέν μᾶς δίνουν ἀρκετές ἐνδείξεις γιά νά ἀπορρίψομε τήν μηδενική ὑπόθεση.

Μερικές φορές τά στατιστικά τέστ λέγονται καί τ έ σ τ σ η μ α ν τ ι κ ό τ η τ α ς. Ἡ διαφορά μέ τά τέστ στατιστικῶν ὑποθέσεων εἶναι περισσότερο φιλοσοφικῆς μορφῆς· παρά οὐσιαστικῆς. Τά τέστ στατιστικῶν ὑποθέσεων εἶναι μία μεθοδολογία λήψεως ἀποφάσεως τύπου "ἀπορρίπτω -δέχομαι". Τά τέστ σημαντικότητας ἀντίθετα κρίνουν ἄν ἕνα πειραματικό ἀποτέλεσμα εἶναι στατιστικῶς σημαντικό. Ἕνα ἀποτέλεσμα εἶναι σ τ α τ ι σ τ ι κ ῶ ς σ η μ α ν τ ι κ ό ἄν ἡ πιθανότητα νά συμβεῖ τό ἀποτέλεσμα αὐτό κάτω ἀπό ὁρισμένες προϋποθέσεις H_0 εἶναι μικρή. Τό γεγονός ὅτι τό ἀποτέλεσμα πραγματοποιεῖται δίνει σημαντικές ἐνδείξεις ἐναντίον τῆς H_0 καί φανερώνει ὅτι τά πειραματικά δεδομένα δέν τήν ὑποστηρίζουν.

Στό παράδειγμα τοῦ δοκιμαστοῦ κρασιῶν ἕνα X μέ τιμές $7, 8, 9$ ἤ 10 εἶναι σ τ α τ ι σ τ ι κ ῶ ς σ η μ α ν τ ι κ ό σ τ ό ἐ π ί π ε δ ο $1,96\%$ διότι ὅταν ἡ H_0-δηλ. $p=1/3$- ἀληθεύει ἡ $P(X \geq 7)$ ἰσοῦται μέ $0,0196$. Ἀλλά καί ἡ $P(X=0)$ εἶναι μικρότερη τοῦ $1,96\%$. Εἶναι λοιπόν ἕνα X ἴσο μέ τό 0 στατιστικῶς σημαντικό στό ἐπίπεδο $1,96\%$; Ὄχι διότι ὑποθέσαμε ὅτι ὁ δοκιμαστής δέν μπορεῖ νά ἔχει ἐπίδοση χειρότερη ἀπό τήν τύχη. Ἄν λοιπόν τό $X=0$ εἶναι ἀσύνηθες ὅταν $p>1/3$, εἶναι περισσότερο ἀσύνηθες ὅταν $p>1/3$. Ἄν ὅμως τό p μπορεῖ νά εἶναι καί μικρότερο τοῦ $1/3$ τότε καί τό $X=0$ εἶναι στατιστικῶς σημαντικό δηλαδή μᾶς δίνει σημαντική ἔνδειξη ἐναντίον τῆς $H_0{:}p=1/3$.

9B. 4 Γενικὲς παρατηρήσεις

1. Εἶναι φανερό ὅτι ποτέ δέν μποροῦμε νά ἀποδείξουμε τήν ὀρθότητα μίας ὑποθέσεως. Πρέπει βέβαια νά ἐξαιρέ-

σουμε τετριμμένες ἤ ἐξεζητημένες περιπτώσεις πού μέ τήν ἐκτέλεση ἑνός ἤ περισσοτέρων πειραμάτων ἀποκαλύπτεται ὁ ἄγνωστος χαρακτῆρας τοῦ πληθυσμοῦ. Ὁμοίως καί περιπτώσεις πού δέν κυβερνοῦνται ἀπό τούς νόμους τῆς τύχης. Ἡ θέση τῆς φύσης εἶναι ἄγνωστη καί οἱοσδήποτε ἤ ὁσοσδήποτε πειραματισμός εἶναι ἀδύνατο νά τήν ἀποκαλύψει.

2. Ἡ ἀποδοχή ἤ ἀπόρριψη τῆς H_0 ἐξαρτᾶται ἀπό τούς κινδύνους (*risks*) πού εἴμαστε διατεθειμένοι νά δεχθοῦμε. Ἐξαρτᾶται ἀπό τά α καί β. Ἀλλάζοντας τό α μπορεῖ νά ἀλλάξει ἡ ἀπόφαση ἤ τό συμπέρασμα. Ἄν ἡ ἰσχύς τοῦ τέστ εἶναι γνωστή, τότε ἀποδεχόμενοι τήν H_0 ἔχομε πλήρη γνώση τῶν κινδύνων τῆς ἀποφάσεως μας. Συχνά ἀντί νά δεχτοῦμε μία μηδενική ὑπόθεση προτιμᾶμε νά ἐκτιμήσουμε τήν ἄγνωστη παράμετρο μέ ἕνα σημεῖο ἤ ἕνα διάστημα ἐμπιστοσύνης.

3. Ἡ διατύπωση τῶν ὑποθέσεων πρέπει νά γίνεται πρίν γίνει τό πείραμα καί ληφθοῦν οἱ μετρήσεις. Διαφορετικά τά συμπεράσματα δέν εἶναι ἀμερόληπτα.

Μετά τήν παρουσίαση τῆς φιλοσοφίας τῶν λογικῶν ἀρχῶν τῶν στατιστικῶν τέστ ἀπομένει νά παρουσιάσομε τή θεωρία εὑρέσεως καί τίς ἰδιότητες τῶν στατιστικῶν συναρτήσεων τῶν τέστ. Αὐτό ἀποτελεῖ ἀντικείμενο τῶν κλάδων "Ἔλεγχος Στατιστικῶν Ὑποθέσεων καί Θεωρία Στατιστικῶν Ἀποφάσεων καί εἶναι ἐκτός τοῦ σκοποῦ τοῦ παρόντος βιβλίου. Ἐδῶ θά περιοριστοῦμε στήν ἀκόλουθη ἀρχή κατασκευῆς σ.σ. τοῦ τέστ πού ἀπορρέει ἀπό τόν ὁρισμό τους: σ.σ. <u>τοῦ τέστ εἶναι μία τυχαία ποσότης μέ γνωστή κατανομή ὅταν ἀληθεύει ἡ</u> H_0.

Στή συνέχεια θά παρουσιάσομε μερικές έφαρμογές τῆς στατιστικῆς συμπερασματολογίας πού θά έμπεδώσουν τόσο τά διαστήματα έμπιστοσύνης ὅσο καί τά στατιστικά τέστ.

Γ' ΕΦΑΡΜΟΓΕΣ: ΑΠΛΗ ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ

9Γ. 1 Συμπερασματολογία γιὰ τὴ μέση τιμὴ

Ἔστω $X_1,\ldots,X_n$ ἕνα τυχαῖο δεῖγμα μεγέθους n ἀπό ἕνα πληθυσμό μέ μέση τιμή μ καί διακύμανση σ^2. Ἔστω $\overline{X}$ ἡ μέση τιμή τοῦ δείγματος καί S^2 ἡ διακύμανση του. Θέλομε νά έκτιμήσομε τό μ καί νά έλέγξομε ὑποθέσεις γιά τό μ. Ὑποθέτομε στήν ἀρχή ὅτι ὁ πληθυσμός εἶναι κ α ν ο ν ι κ ό ς. Ἀργότερα θά χαλαρώσομε τήν ὑπόθεση αὐτή.

Ὁ καλύτερος έκτιμητής τοῦ μ εἶναι ὁ

$$\hat{\mu} = \overline{X}$$

μέ έκτιμώμενο τυπικό σφάλμα

$$\hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}}$$

(Βλέπε §9Α.2). Τό καλύτερο διάστημα έμπιστοσύνης γιά τό μ εἶναι

$$L = \overline{X}-t_{\alpha/2,n-1}\,\frac{S}{\sqrt{n}}, \qquad U = \overline{X}+t_{\alpha/2,n-1}\,\frac{S}{\sqrt{n}}. \quad (9.13)$$

Τό διάστημα αὐτό ἀπορρέει ἀπό τό γεγονός ὅτι ἡ ποσότης

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

(Βλέπε Θεώρημα 8.3).

Γιά νά ἐλέγξουμε ὑποθέσεις τῆς μορφῆς

(a)	$H_0: \mu = \mu_0$	ἔναντι	$H_a: \mu > \mu_0$
(b)	$H_0: \mu \leq \mu_0$	"	$H_a: \mu > \mu_0$
(c)	$H_0: \mu = \mu_0$	"	$H_a: \mu < \mu_0$
(d)	$H_0: \mu \geq \mu_0$	"	$H_a: \mu < \mu_0$
(e)	$H_0: \mu = \mu_0$	"	$H_a: \mu \neq \mu_0$

μέ ἐπίπεδο σημαντικότητας α χρησιμοποιοῦμε τό στατιστικό

$$\boxed{t = \frac{\overline{X}-\mu_0}{S/\sqrt{n}} \quad \text{τό όποῖο ἔχει κατανομή } t_{n-1} \text{ ὅταν } \mu = \mu_0} \qquad (9.14)$$

(Βλέπε Θεώρημα 8.3) καί τίς ἀκόλουθες ἀντίστοιχες κρίσιμες περιοχές:

γιά τήν (a) $C = (t_{\alpha,n-1}, +\infty)$, δηλ. $t > t_{\alpha,n-1}$

" " (b) $C = (t_{\alpha,n-1}, +\infty)$, δηλ. $t > t_{\alpha,n-1}$

" " (c) $C = (-\infty, -t_{\alpha,n-1})$, δηλ. $t < -t_{\alpha,n-1}$

" " (d) $C = (-\infty, -t_{\alpha,n-1})$, δηλ. $t < -t_{\alpha,n-1}$

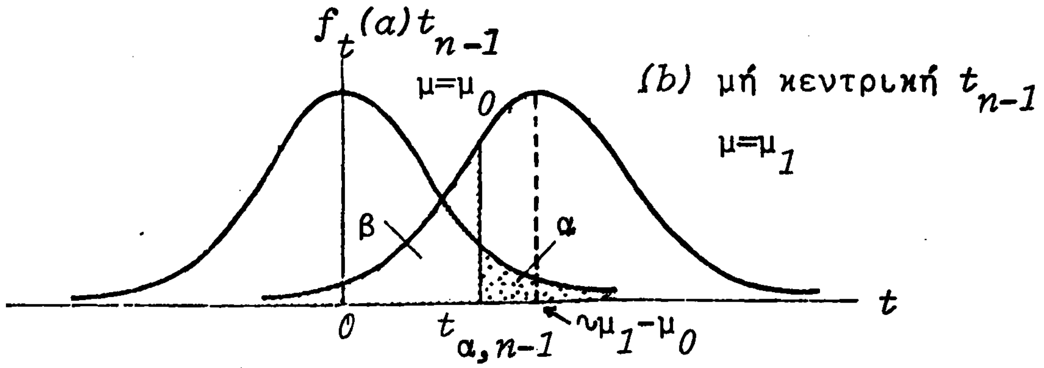
γιά τήν (ε) $C = (-\infty, -t_{\alpha/2,n-1}) U (t_{\alpha/2,n-1}, +\infty)$,

δηλ. $|t| > t_{\alpha/2,n-1}$.

τό τ έ σ τ αὐτό λέγεται t τ έ σ τ .

Οἱ παραπάνω κρίσιμες περιοχές εἶναι οἱ καλύτερες δυνατές. Αὐτό ἀποδεικνύεται στή Θεωρία Ἐλέγχου Στατιστικῶν Ὑποθέσεων. Μπορεῖ νά φανεῖ ὅμως καί διαισθητικά. Πράγματι, γιά τίς ὑποθέσεις (a) μεγάλες θετικές τιμές τοῦ $\overline{X}$ ὑποδηλώνουν ἀπόρριψη τῆς H_0 καί ἀποδοχή τῆς H_a. Ἀλλά μεγάλες θετικές τιμές τοῦ $\overline{X}$ σημαίνουν μεγάλες θετικές τιμές τοῦ t.

Τό Σχῆμα 9.2 μᾶς δίνει μιά γραφική παράσταση τῆς κρίσιμης περιοχῆς τοῦ t-τέστ, τοῦ α καί τοῦ β (γιά $\mu=\mu_1>\mu_0$) γιά τήν περίπτωση τῶν ὑποθέσεων (a)



Σχῆμα 9.2 Κρίσιμη περιοχή, α καί β γιά τό t-τέστ

Όταν ισχύει ή H_0 ή σ.σ. (9.14) τοῦ t-τέστ ἔχει κατανομή t_{n-1}. Ἄν $\mu=\mu_1>\mu_0$ ή σ.σ (9.14) ἔχει μή κεντρική t_{n-1} κατανομή ή όποία ἔχει ἀνάλογο σχῆμα μέ τήν t_{n-1} ἀλλά εἶναι μετατοπισμένη πρός τά δεξιά μέ μέση τιμή ἀνάλογη τοῦ $\mu_1-\mu_0$. Ἡ κ.π. καθορίζεται μέ τό σημεῖο t_α. Δεξιά του καί κάτω ἀπό τήν καμπύλη (a) ἔχουμε τό α, ἐνῶ ἀριστερά του καί κάτω ἀπό τήν καμπύλη (b) ἔχομε τό β.

Παράδειγμα 9.2:

Ἕνας ὑγειονομικός σταθμός θέλει νά ἐλέγχει ἄν ὁ μέσος ἀριθμός βακτηριδίων ἀνά μονάδα ὄγκου θαλασσινοῦ νεροῦ σέ μία παραλία ὑπερβαίνει τό ἐπίπεδο ἀσφαλείας τῶν 200. Δώδεκα δείγματα νεροῦ συλλέγονται καί εὑρίσκονται οἱ ἀκόλουθοι ἀριθμοί βακτηριδίων ἀνά μονάδα ὄγκου:

$$170,\quad 175,\quad 190,\quad 198,\quad 215,\quad 185$$
$$184,\quad 207,\quad 210,\quad 193,\quad 196,\quad 180.$$

Ὑπάρχει λόγος ἀνησυχίας;

Ἔστω μ ὁ μέσος ἀριθμός βακτηριδίων ἀνά μονάδα ὄγκου καί σ^2 ή διακύμανση. Δεδομένου ὅτι ή ἀποτυχία διαπιστώσεως παραβιάσεως τοῦ ὁρίου ἀσφαλείας μπορεῖ νά ἔχει σοβαρές συνέπειες ή μηδενική καί ή ἐναλλακτική ὑπόθεση εἶναι

$$H_0:\mu \geq 200 \quad \text{καί} \quad H_a:\mu< 200.$$

Ὑποθέτουμε ὅτι οἱ μετρήσεις ἀποτελοῦν ἕνα τυχαῖο δεῖγμα ἀπό μία κανονική κατανομή καί ὅτι τό ἐπίπεδο σημαντικότητας εἶναι $\alpha = 1\%$. Τό t τέστ

$$t = \frac{\overline{X}-200}{S/\sqrt{12}}$$

ἔχει *11* βαθμούς ἐλευθερίας καί κρίσιμη περιοχή

$$C = (-\infty, -t_{0,01,11} = -2,718).$$

Οἱ ὑπολογισμοί μέ τά παραπάνω δεδομένα μᾶς δίνουν

$$\overline{X} = 191,9$$

$$S = 14,03$$

$$t = \frac{191,1-200}{14,03/\sqrt{12}} = \frac{-8,9}{4,05} = -2,197.$$

Ἐπειδή ἡ παρατηρηθεῖσα τιμή $t = -2,197$ εἶναι ἐκτός τῆς κ.π. C, ἡ H_0 δέν ἀπορρίπτεται στό ἐπίπεδο τοῦ *1%*. Δέν ὑπάρχει ἔνδειξη ὅτι τό μέσο ἐπίπεδο τῶν βακτηριδίων εἶναι ἐντός τῶν ορίων ἀσφαλείας.

Στό σημεῖο αὐτό ἴσως θελήσουμε νά ἐκτιμήσουμε τό μ. Πράγματι ὁ ἐκτιμητής τοῦ μ εἶναι ὁ

$\hat{\mu}=\overline{X}=191,9$ βακτηρίδια ἀπό μονάδα ὄγκου

καί ἔχει τυπικό σφάλμα

$$\hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}} = 4,05.$$

Ἕνα *95%* διάστημα ἐμπιστοσύνης γιά τό μ εἶναι τό

$$\overline{X} \pm t_{0,025,11} \frac{S}{\sqrt{n}} = 191,9 \pm 2,201 \times 4,05 = 191,9 \pm 8,91$$

Παράδειγμα 9.3:

Σέ μιά προσπάθεια νά προσδιοριστεῖ ἄν ἡ εἰδική ἐκπαίδευση αὐξάνει ἤ ὄχι τόν δείκτη εὐφυΐας, εἴκοσι πέντε παιδιά ἐξετάζονται μέ ἕνα βασικό - τυποποιημένο τέστ εὐφυΐας. Κατόπιν τά παιδιά αὐτά παρακαολουθοῦν ἕνα εἰδικό μάθημα, ὁ σκοπός τοῦ ὁποίου εἶναι ἡ αὔξηση τοῦ δείκτη εὐφυΐας. Στό τέλος του μαθήματος ἐξετάζονται γιά δεύτερη φορά. Ἡ διαφορά μεταξύ τῶν βαθμῶν τῆς δευτέρας καί πρώτης ἐξετάσεως καταγράφεται γιά κάθε παιδί. Ἔστω ὅτι ἡ μέση τιμή τῶν διαφορῶν αὐτῶν εἶναι $\overline{X}=6$ μονάδες (σέ κλίμακα 0-100) καί ἡ διακύμανση $S^2=64$. Ἔχει ἡ εἰδική ἐκπαίδευση αὐξήσει τόν δείκτη εὐφυΐας;

Ἔστω μ ἡ μέση τιμή τῶν διαφορῶν. Ἡ μηδενική ὑπόθεση εἶναι ὅτι ἡ εἰδική ἐκπαίδευση δέν εἶναι ἀποτελεσματική, δηλ. H_0:μ=0. Ἡ ἐναλλακτική εἶναι H_a:μ>0, δηλαδή ὑπάρχει θετική αὔξηση. Θά χρησιμοποιήσομε δεξιόπλευρο t -τέστ καί θέτομε α=5%. Ἔτσι $t_{0,05,24}=1,711$.

Ἡ τιμή τῆς σ.σ. τοῦ t-τέστ εἶναι

$$t = \frac{\overline{X}-0}{S/\sqrt{25}} = \frac{6-0}{8/5} = 3,75 > 1,711 = t_{\alpha,n-1} \, .$$

Ἔτσι ἀπορρίπτεται ἡ H_0 καί βγάζουμε τό συμπέρασμα ὅτι πράγματι ἡ εἰδική ἐκπαίδευση αὐξάνει τό δείκτη εὐφυΐας. Τό ποσοστό σφάλματος μας εἶναι 5%.

Όσες φορές χρησιμοποιοῦμε τήν κατανομή t_ν γιά ὑπολογισμό διαστημάτων ἐμπιστοσύνης ἤ τέστ στατιστικῶν ὑποθέσεων ὑποθέτομε ὅτι τά δεδομένα μας εἶναι ἕνα τυχαῖο δεῖγμα ἀπό ἕνα κανονικό πληθυσμό. Ἡ πρώτη ὑποθέση τοῦ τυχαίου δείγματος εἶναι οὐσιαστική ἄν θέλομε νά διατυπώσομε πιθανοθεωρητικά συμπεράσματα γιά τά ἀποτελέσματα μας. Δέν μπορεῖ νά χαλαρωθεῖ.

Ἡ δεύτερη ὑπόθεση τοῦ κανονικοῦ πληθυσμοῦ δέν εἶναι τόσο ἰσχυρή καί μπορεῖ νά χαλαρωθεῖ ἀρκεῖ νά ἔχομε ἕνα ἀρκετά μεγάλο δεῖγμα. Αὐτό εἶναι ἄμεση συνέπεια τοῦ Κεντρικοῦ Ὁριακοῦ θεωρήματος.

Μ έ γ ε θ ο ς τ ο ῦ δ ε ί γ μ α τ ο ς: Ἡ στατιστική θεωρία προσφέρει πολύτιμη βοήθεια κατά τή σχεδίαση μελετῶν, ἐρευνῶν ἤ πειραμάτων. Μία τέτοια περίπτωση εἶναι ὁ προσδιορισμός τοῦ μεγέθους τοῦ δείγματος.

Ἡ ἀκρίβεια (ἀξιοπιστία) ἑνός ἐκτιμητῆ $\hat{\theta}$ μετρεῖται μέ τό τυπικό του σφάλμα $\sigma_{\hat{\theta}}$ τό ὁποῖο συνήθως εἶναι

$$\hat{\sigma}_{\hat{\theta}} = \sigma/\sqrt{n} \,,$$

ὅπου σ ἡ τυπική ἀπόκλιση τοῦ πληθυσμοῦ καί n τό μέγεθος τοῦ δείγματος. Ἔτσι ἄν ἡ μεταβλητότητα (δηλ. τό σ) τοῦ πληθυσμοῦ εἶναι γνωστή (π.χ. ἀπό προηγούμενες μελέτες) μποροῦμε νά προσδιορίσομε τό n ὥστε νά πετύχομε μία δεδομένη ἀκρίβεια γιά τό $\hat{\theta}$.

Ἔστω τώρα ὅτι θέλομε νά ἐκτιμήσομε τή μέση τιμή μ ἑνός πληθυσμοῦ ἀπό ἕνα μεγάλο δεῖγμα κατά τέτοιο τρόπο ὥστε ἡ μέση τιμή $\overline{X}$ τοῦ δείγματος νά εὑρίσκεται (μέ ἐμπιστοσύνη $100(1-\alpha)\%$) ἐντός w μονάδων ἀπό τό μ. Μέ ἄλλα λόγια θέλομε τό μῆκος τοῦ Δ.Ε. (*9.13*) νά εἶναι $2w$

Ἔτσι

$$U\text{-}L = 2t_{\alpha/2,n-1} \frac{S}{n} = 2w$$

ἤ

$$n = \frac{t_{\alpha/2,n-1}\, S^2}{w^2} \quad . \qquad (9.15)$$

Ἡ λύση ὡς πρός n τῆς ἐξισώσεως (9.15) δέν εἶναι εὔκολη. Στήν πράξη ὅμως μποροῦμε νά προσεγγίσομε τό $t_{\alpha/2,n-1}$ μέ τό $z_{\alpha/2}$ καί τό S μέ κάποιο σ πού προκύπτει ἀπό προσωπική γνώση, παλαιά δεδομένα, ἤ πληροφορία ἀπό τή βιβλιογραφία. Ἔτσι ἔχομε τή σχέση

$$n = \frac{z^2_{\alpha/2}\sigma^2}{w^2} \quad . \qquad (9.16)$$

Βλέπομε δηλαδή ὅτι καθώς ἡ διακύμανση αὐξάνει, τό n ἐπίσης μεγαλώνει. Πληθυσμοί μέ μεγάλη μεταβλητότητα ἀπαιτοῦν περισσότερες μετρήσεις (σέ σύγκριση μέ πληθυσμούς μέ μικρή μεταβλητότητα) γιά νά ἐπιτευχθεῖ ἡ ἐπιθημητή ἀκρίβεια. Ἐπίσης ὅσο ἡ ἐμπιστοσύνη τῆς ἐκτιμήσεως αὐξάνει (δηλ. τό α μικραίνει) τό n μεγαλώνει. Καί καθώς τό w μικραίνει τό n μεγαλώνει.

9Γ. 2 Συμπερασματολογία γιὰ τὸ διωνυμικὸ p

Ἔστω ὅτι δίνονται n ἀνεξάρτητες δοκιμές τύπου *Bernoulli* μέ πιθανότητα ἐπιτυχίας p καί ἔστω X ὁ ἀριθμός τῶν ἐπιτυχιῶν. Μέ ἄλλα λόγια δίνεται ἕνα διωνυμικό πείραμα μέ τιμή X ὅπου ἡ τυχαία μεταβλητή X ἔχει διωνυμική κατανομή $B(n,p)$.

Ὁ φυσιολογικός ἐκτιμητής $\hat{p}$ τοῦ p εἶναι ἡ παρατηθεῖσα συχνότης ἐπιτυχίας, δηλαδή

$$\hat{p} = \frac{X}{n} \ . \qquad (9.17)$$

Ὁ ἐκτιμητής αὐτός εἶναι ἀ μ ε ρ ό λ η π τ ο ς διότι

$$E(\hat{p}) = E(X)/n = np/n = p$$

καί ἔχει τ υ π ι κ ό σ φ ά λ μ α

$$\sigma_{\hat{p}} = \sqrt{p(1-p)/n} \qquad (9.18)$$

πού ἐκτιμᾶται μέ

$$\hat{\sigma}_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/n} \ . \qquad (9.19)$$

Τό Κεντρικό Ὁριακό Θεώρημα μᾶς δίνει ὅτι ἡ κατανομή τῆς τ.μ. X εἶναι κατά προσέγγιση κανονική μέ μέση τιμή np καί διακύμανση npq (βλ. ἐδάφιο 8.6). Ἡ προσέγγιση βελτιώνεται ὅσο τό n αὐξάνει. Ἕνα $100(1-\alpha)\%$ διάστημα ἐμπιστοσύνης γιά τό p κατασκευάζεται μέ τή βοήθεια τῆς προσεγγίσεως αὐτῆς:

$$L = \hat{p}-z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} \ , \ U = \hat{p}+z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} \ . \qquad (9.20)$$

Ὁ ἔλεγχος στατιστικῶν ὑποθέσεων γιά τό p γίνεται μέ τή βοήθεια τῆς στατιστικῆς συναρτήσεως X ὅπως αναπτύχθηκε στίς §9Β.1, 9Β.2 καί 9Β.3. Τό τέστ αὐτό εἶναι ἀκριβές καί γίνεται μέ τή χρήση λεπτομερῶν διωνυμικῶν πινάκων ἤ σημαντικῆς ὑπολογιστικῆς προσπάθειας. Ἕνα κατά προσέγγιση (γιά μεγάλα n) τέστ μπορεῖ νά γίνει χρησιμοποιῶντας τό στατιστικό

$$z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{X-np_0}{\sqrt{np_0(1-p_0)}}$$ τό ὁποῖο ἔχει τυπική κανονική κατανομή ὅταν $p=p_0$. (9.21)

Ἔτσι γιά νά ἐλέγξομε ὑποθέσεις τῆς μορφῆς

(a) $H_0: p=p_0$ ἔναντι $H_\alpha: p>p_0$

(b) $H_0: p\leq p_0$ ἔναντι $H_\alpha: p>p_0$

(c) $H_0: p=p_0$ ἔναντι $H_\alpha: p<p_0$

(d) $H_0: p\geq p_0$ ἔναντι $H_\alpha: p<p_0$

(e) $H_0: p=p_0$ ἔναντι $H_\alpha: p\neq p_0$

χρησιμοποιοῦμε τό z τῆς (9.19) καί τίς ἀκόλουθες κρίσιμες περιοχές

γιά τήν (a) $C = (z_\alpha, +\infty)$, δηλ. $z>z_\alpha$

" " (b) $C = (z_\alpha, +\infty)$, δηλ. $z>z_\alpha$

" " (c) $C = (-\infty, -z_\alpha)$, δηλ. $z<-z_\alpha$

" " (d) $C = (-\infty, -z_\alpha)$, δηλ. $z<-z_\alpha$

" " (e) $C = (-\infty, -z_{\alpha/2}) \cup (z_{\alpha/2}, +\infty)$

δηλ. $|Z|>z_{\alpha/2}$

Παράδειγμα 9.4:

Ἡ Ἰατρική Ἐρευνητική Ὁμάδα (*Coronary Drug Project, J. Chron, Dis., vol 27,267-285)* ἀνέφερε τό 1974 ὅτι ἀπό 2789 ἄνδρες ἡλικίας *30-64*, οἱ ὁποῖοι εἶχαν μία ἤ περισσότερες καρδιακές προσβολές, ὑποβλήθηκαν σέ θεραπεία καί ἔγιναν καλά, *358*(δηλαδή *12,8%*) πέθαναν κατά τή διάρκεια τῶν τριῶν ἐτῶν πού ἀκολούθησαν τήν θεραπεία. Νά ἐκτιμηθεῖ ἡ πιθανότητα θανάτου ἐντός τῶν τριῶν ἐτῶν μετά τή θεραπεία.

Ἔστω ὅτι γιά κάθε δοκιμή, δηλαδή γιά κάθε καρδιοπαθή, ἐπιτυχία εἶναι τό ἄτομο νά μήν πεθάνει κατά τή διάρκεια τῶν τριῶν ἐτῶν μετά τή θεραπεία. Ἔστω p ἡ πιθανότητα αὐτή. Τότε ἔχομε $n = 2789$ δοκιμές καί

$$\hat{p} = \frac{2431}{2789} = 0,872, \qquad 1-p = \frac{358}{2789} = 0,128 \ .$$

Τό $\hat{p}$ ἐκτιμᾶ τήν πιθανότητα ζωῆς κατά τήν διάρκεια τῶν τριῶν ἐτῶν καί τό $1-\hat{p}$ ἐκτιμᾶ τήν πιθανότητα θανάτου ἐντός τῶν τριῶν μετά τή θεραπεία ἐτῶν. Τό τυπικό σφάλμα τοῦ $\hat{p}$ εἶναι

$$\hat{\sigma}_{\hat{p}} = \sqrt{(0,128)(0,872)/\ 2789} = 0,006 \ .$$

Ὁμοίως

$$\hat{\sigma}_{(1-\hat{p})} = 0,006 \ .$$

Ἕνα κατά προσέγγιση 95% Δ.Ε. γιά τό p δίδεται ἀπό τήν (9.20):

$$0,872 \pm 1,96(0,006) = 0,872 \pm 0,012,$$

δηλαδή $L = 86\%$ καί $U = 88,4\%$.

Μέ όμοιο τρόπο ένα 95% Δ.Ε. γιά τό $1-p$ εἶναι

$$0,128 \pm 1,96(0,006) = 0,128 \pm 0,012,$$

δηλαδή $L = 11,6\%$ καί $U = 14\%$. Μέ ἄλλα λόγια, ὁ τριετής ρυθμός θανάτου εἶναι μεταξύ 11,6% καί 14%.

Μέ βάση τά προηγούμενα νομίζετε ὅτι εἶναι λογικό νά συμπεράνομε ὅτι ὁ ἐ τ ή σ ι ο ς ρυθμός θανάτου γιά τήν κατηγορία τῶν παραπάνω ἀτόμων εἶναι μεταξύ 3,9% καί 4,7%;

Ἡ παραπάνω ἀνάλυση ἰσχύει ὑπό τήν προϋπόθεση ὅτι τά 2789 ἄτομα ἀποτελοῦν ἕνα τυχαῖο δεῖγμα ἀπό τόν (μεγάλο) πληθυσμό καρδιοπαθῶν μίας χώρας. Στήν πραγματικότητα τά 2789 ἄτομα δέν ἐλήφθησαν μέ δειγματοληψία. Ἐλήφθησαν ἀπό μιά παρεμφερή κλινική δοκιμή τῆς ὁποίας ὁ σκοπός ἦταν ἡ σύγκριση διαφόρων φαρμάκων ἐλαττώσεως τῆς χοληστερίνης μέ τό ἀδρανές φάρμακο *placebo* ὡς προληπτικά φάρμακα καρδιακοῦ νοσήματος. Τά 2789 ἄτομα πῆραν τό *placebo* καί ἡ μελέτη ἔγινε σέ 53 νοσοκομεῖα τῶν Ἡνωμένων Πολιτειῶν τῆς Ἀμερικῆς πού ἔλαβαν μέρος στήν ἔρευνα. Βέβαια πρέπει νά γνωρίζουμε ὅτι τά ἄτομα αὐτά εἶχαν συστηματική ἰατρική παρακολούθηση μέ ἀποτέλεσμα νά ὑπάρχει ἕνα μικρό ποσοστό μεροληψίας ἔναντι τῶν τυχαίως ἐκλεγομένων καρδιοπαθῶν.

Ἔστω τώρα ὅτι θέλομε νά ἐλέγξομε τήν ὑπόθεση $H_0{:}p=0,90$ ἔναντι $H_a{:}p \neq 0,90$. Ἡ τιμή τῆς σ.σ. τοῦ τέστ (9.21) εἶναι

$$z = \frac{0,872-0,90}{\sqrt{(0,90)(0,10)/2789}} = \frac{0,028}{0,0057} = -4,91 .$$

Ἡ τιμή αὐτή εἶναι μέσα στήν κρίσιμη περιοχή μεγέθους 5%

$$c = (-\infty, -1,96) \cup (1,96, +\infty)$$

καί ἔτσι μέ ἐπίπεδο σημαντικότητας $\alpha = 5\%$ ἡ H_0 ἀπορρίπτεται. Μάλιστα δέ ἐπειδή ἡ πιθανότητα $P(Z<-4,91)$ εἶναι σχεδόν μηδέν ἡ τιμή $z = 4,91$ εἶναι στατιστικῶς πολύ σημαντική καί ἔτσι ὑπάρχουν ἰσχυρές ἐνδείξεις γιά τήν ἀπόρριψη τῆς H_0. Τό ἀποτέλεσμα αὐτό δέν μᾶς ξενίζει διότι τό $n = 2789$ εἶναι ἀρκετά μεγάλο καί τό p ἐκτιμᾶται μέ τό p μέ μεγάλη ἀκρίβεια.

Μέγεθος τοῦ δείγματος: Ἄν θέλουμε νά ἐκτιμήσουμε τό p μέ βαθμό ἐμπιστοσύνης $100(1-\alpha)\%$ καί σφάλμα ἐκτιμήσεως ἐντός $\pm w$ ἀπό Δ.Ε. (9.20) ἔχουμε

$$2w = U-L = 2z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad ἤ \quad n = \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{w}\right)^2$$

ἐπειδή $\hat{p}(1-\hat{p}) \leq 1/4$ γιά κάθε $\hat{p}$ $(0,1)$

$$n \leq \frac{1}{4}\left(\frac{z_{\alpha/2}}{w}\right)^2 .$$

9Γ.3 Συμπερασματολογία γιά τή διαφορά δύο μέσων τιμῶν

'Εδῶ ἀναλύομε δεδομένα πού ἀποτελοῦνται ἀπό δύο ἀ ν ε ξ ά ρ τ η τ α τυχαῖα δείγματα, δεῖγμα *1* ἀπό πληθυσμό *1* καί δεῖγμα *2* ἀπό πληθυσμό *2*. Συχνά τό δεῖγμα *1* εἶναι δ ε ῖ γ μ α ἐ λ έ γ χ ο υ καί τό δεῖγμα *2* εἶναι δεῖγμα "θ ε ρ α π ε ί α ς" ἤ "δ ο κ ι μ α σ ί α ς. Ἄλλες φορές τά δείγματα προκύπτουν ἀπό πειράματα συγκρίσεως τῶν ἀποτελεσμάτων δ ύ ο δ ι α φ ο ρ ε τ ι κ ῶ ν θ ε ρ α π ε ι ῶ ν.

Στή Στατιστική θ ε ρ α π ε ῖ ε ς εἶναι ὁποιεσδήποτε διαδικασίες ἤ μέθοδοι ἤ οὐσίες τά ἀποτελέσματα τῶν ὁποίων θέλομε νά ἐκτιμήσομε ἤ συγκρίνομε. Θεραπεῖες μπορεῖ νά εἶναι διαφορετικά φάρμακα ἤ διαφορετικές δόσεις ἑνός φαρμάκου διαφορετικές θεραπευτικές ἀγωγές στήν ἰατρική, διαφορετικές χημικές οὐσίες ἤ διαφορετικά ποσοστά ἐφαρμογῆς μίας χημικῆς οὐσίας, διαφορετικές μηχανές ἤ διαφορετικοί χειριστές κ.λ.π.

'Η σύγκριση δύο θεραπειῶν γίνεται ἐφαρμόζοντας τες σέ ὁμοειδές πειραματικό ὑλικό. Διαιροῦμε τό πειραματικό ὑλικό σέ π ε ι ρ α μ α τ ι κ έ ς μ ο ν ά δ ε ς καί κατόπιν ἐφαρμόζομε στήν τύχη μία ἀπό τίς δύο θεραπεῖες σέ κάθε πειραματική μονάδα. Αὐτό γίνεται εἴτε μέ ἕνα νόμισμα (κορώνα-γράμματα) εἴτε μέ τή βοήθεια πινάκων τυχαίων ἀριθμῶν. Μερικές φορές ἐπιβάλλομε καί τόν δευτερεύοντα περιορισμό κάθε θεραπεία νά ἐφαρμόζεται στόν ἴδιο ἀριθμό πειραματικῶν μονάδων. Σχέδια πειραμάτων τῆς παραπάνω μορφῆς λέγονται π λ ή ρ ω ς τ υ χ α ι ο π ο ι η μ έ-ν α σ χ έ δ ι α.

Ἀφοῦ ἐφαρμοστοῦν οἱ θεραπεῖες παίρνομε ἕνα ἀντικείμενικό μέτρο τοῦ ἀποτελέσματος κάθε θεραπείας ἀπό κάθε μονάδα καί ἔτσι καταλήγομε σέ δυό ἀνεξάρτητα δείγματα ἀπό δύο πληθυσμούς, τίς δύο θεραπεῖες.

Στόχος εἶναι νά ἐκτιμήσομε τή διαφορά $\mu_1-\mu_2$ ἤ νά ἐλέγξομε τήν ὑπόθεση $H_0: \mu_1=\mu_2$, ὅπου μ_1 καί μ_2 εἶναι οἱ μέσες τιμές τῶν πληθυσμῶν (θεραπειῶν) 1 καί 2 ἀντίστοιχα.

Ἔστω $X_{11},\ldots,X_{1n_1}$ τά δεδόμενα (τυχαῖο δεῖγμα μεγέθους n_1) ἀπό τόν πληθυσμό 1 καί $X_{21},\ldots,X_{2n_2}$ τά δεδομένα (τυχαῖο δεῖγμα μεγέθους n_2) ἀπό τόν πληθυσμό 2. Οἱ πληροφορίες πού περιέχονται στά δεδομένα αὐτά συνοψίζονται περιληπτικά στόν ἀκόλουθο πίνακα:

Δεῖγμα	Μέγεθος	Βαθμ.ἐλευθ.(β.ε.)	Μέσ.τιμή	Ἄθρ.τετρ.(SS)
1	n_1	n_1-1	$\bar{X}_1$	$\Sigma(X_{1i}-\bar{X}_1)^2$
2	n_2	n_2-1	$\bar{X}_2$	$\Sigma(X_{2i}-\bar{X}_2)^2$
Σύνολο		n_1+n_2-2		Συνολικό SS

Πίνακας 9.1 Ὑπολογισμοί δύο δειγμάτων.

Ὑπενθυμίζομε τούς τύπους μηχανῆς γιά τά ἀθροίσματα τετραγώνων καί τίς διακυμάνσεις:

$$S_1^2 = \frac{1}{n_1-1}\Sigma(X_{1i}-\bar{X}_1)^2 = \frac{1}{n_1-1}\{\Sigma X_{1i}^2 - \frac{(\Sigma X_{1i})^2}{n_1}\}$$

$$S_2^2 = \frac{1}{n_2-1}\Sigma(X_{2i}-\bar{X}_2) = \frac{1}{n_2-1}\{\Sigma X_{2i}^2 - \frac{(\Sigma X_{2i})^2}{n_2}\} . \qquad (9.22)$$

Ὁ καλύτερος ἐκτιμητής τῆς διαφορᾶς $\mu_1-\mu_2$ εἶναι ὁ

$$\widehat{\mu_1-\mu_2} = \overline{X}_1-\overline{X}_2 \,. \qquad (9.23)$$

Ἐπειδή τά τυχαῖα δείγματα εἶναι ἀνεξάρτητα, οἱ στατιστικές συναρτήσεις $\overline{X}_1$ καί $\overline{X}_2$ εἶναι ἀνεξάρτητες. Ἔτσι ἡ διακύμανση τοῦ $\overline{X}_1-\overline{X}_2$ εἶναι

$$Var(\overline{X}_1-\overline{X}_2) = Var\overline{X}_1+Var\overline{X}_2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \,. \qquad (9.24)$$

Ἀπό τή σχέση αὐτή βρίσκομε ὅτι τό τυπικό σφάλμα τοῦ ἐκτιμητῆ $\overline{X}_1-\overline{X}_2$ τῆς διαφορᾶς $\mu_1-\mu_2$ εἶναι

$$\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \,. \qquad (9.25)$$

Ἔτσι τό ἐκτιμώμενο τυπικό σφάλμα τοῦ $\overline{X}_1-\overline{X}_2$ εἶναι

$$\hat{\sigma}_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \,. \qquad (9.26)$$

Γιά νά προχωρήσομε τώρα στά διαστήματα ἐμπιστοσύνης καί τέστ γιά τή διαφορά $\mu_1-\mu_2$ χρειάζεται νά κάνουμε μερικές ὑποθέσεις γιά τίς κατανομές τῶν $\overline{X}_1$ καί $\overline{X}_2$. Ἄν λοιπόν τά δείγματα 1 καί 2 εἶναι μεγάλα τότε ἰσχύουν κατά προσέγγιση οἱ σχέσεις

$$\overline{X}_1 \sim N\left(\mu_1,\frac{\sigma_1^2}{n_1}\right) \quad \text{καί} \quad \overline{X}_2 \sim N\left(\mu_2,\frac{\sigma_2^2}{n_2}\right) . \qquad (9.27)$$

Ἔτσι στήν περίπτωση αὐτή ἕνα $100(1-\alpha)\%$ κατά προσέγγιση διάστημα ἐμπιστοσύνης γιά τή διαφορά $\mu_1-\mu_2$ δίδεται ἀπό τίς σχέσεις

$$\boxed{\begin{aligned} L &= \overline{X}_1-\overline{X}_2-Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \\ U &= \overline{X}_1-\overline{X}_2+Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\,. \end{aligned}} \qquad (9.28)$$

Ὁμοίως, στήν περίπτωση αὐτή, ὑποθέσεις τῆς μορφῆς

$$\begin{array}{llcl} (a) & H_0:\mu_1-\mu_2=\delta_0 & \text{ἔναντι} & H_\alpha:\mu_1-\mu_2>\delta_0 \\ (b) & H_0:\mu_1-\mu_2\le\delta_0 & \text{"} & H_\alpha:\mu_1-\mu_2>\delta_0 \\ (c) & H_0:\mu_1-\mu_2=\delta_0 & \text{"} & H_\alpha:\mu_1-\mu_2<\delta_0 \\ (d) & H_0:\mu_1-\mu_2\ge\delta_0 & \text{"} & H_\alpha:\mu_1-\mu_2<\delta_0 \\ (e) & H_0:\mu_1-\mu_2=\delta_0 & \text{"} & H_\alpha:\mu_1-\mu_2\ne\delta_0 \end{array} \qquad (9.29)$$

ἐλέγχονται κατά προσέγγιση μέ τό στατιστικό

$$z=\frac{\overline{X}_1-\overline{X}_2-\delta_0}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \qquad (9.30)$$

τό ὁποῖο ἔχει κατά προσέγγιση τυπική κανονική κατανομή ὅταν ἰσχύει ἡ H_0.

Οἱ ἀντίστοιχες κρίσιμες περιοχές εἶναι οἱ ἀκόλουθες

γιά τήν (a) $C=(z_\alpha,+\infty)$, δηλ. $Z>z_\alpha$

" " (b) $C=(z_\alpha,+\infty)$, δηλ. $Z>z_\alpha$

γιά τήν (c) $C = (-\infty, -z_\alpha)$, δηλ. $Z>-z_\alpha$

" " (d) $C = (-\infty, -z_\alpha)$, δηλ. $Z<-z_\alpha$

" " (e) $C = (-\infty, -z_{\alpha/2}) U (z_{\alpha/2}, +\infty)$,

δηλ. $|Z|>z_{\alpha/2}$.

Όταν τά μεγέθη n_1 καί n_2 τῶν δειγμάτων 1 καί 2 εἶναι μικρά ἡ προηγούμενη συμπερασματολογία ὅσον ἀφορᾶ τά διαστήματα ἐμπιστοσύνης καί τέστ δέν ἰσχύει. Τά διαστήματα (9.28) καί τέστ (9.30) δέν εἶναι ἀξιόπιστα. Γιά νά προχωρήσομε χρειάζονται περισσότερες ὑποθέσεις γιά τίς κατανομές τῶν πληθυσμῶν 1 καί 2 καί τίς διακυμάνσεις τους. Χρειάζεται νά ὑποθέσομε ὅτι οἱ πληθυσμοί εἶναι κανονικοί καί ἔχουν ἴσες διακυμάνσεις.

Κ α ν ο ν ι κ ο ί π λ η θ υ σ μ ο ί - ἴ σ ε ς δ ι α κ υ μ ά ν σ ε ι ς. Ἔστω λοιπόν ὅτι οἱ δύο πληθυσμοί εἶναι κ α ν ο ν ι κ ο ί καί ὅτι οἱ διακυμάνσεις τους σ_1^2 καί σ_2^2 εἶναι ἴ σ ε ς καί ἄγνωστες, δηλαδή $\sigma_1^2=\sigma_2^2=\sigma^2$. Ὅταν τό πειραματικό ὑλικό (πειραματικές μονάδες) εἶναι ὁμοειδές καί ἐλέγχονται ὅλοι οἱ παράγοντες πού εἶναι δυνατόν νά ἐπηρεάζουν τίς μετρήσεις (ἐκτός ἀπό τίς θεραπεῖες) τότε ἡ ὑπόθεση τῆς ἰσότητας τῶν διακυμάνσεων εἶναι ρεαλιστική. Βέβαια ἡ ὑπόθεση αὐτή μπορεῖ νά ἐλεγχθεῖ καί στατιστικῶς (βλέπε τέστ ἰσότητας τῶν διακυμάνσεων).

Ὁ καλύτερος ἐκτιμητής τῆς κοινῆς διακυμάνσεως σ^2 εἶναι ὁ

$$S^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} = \frac{\text{συνολικό } SS}{\text{συνολικοί β.ε.}} \qquad (9.31)$$

ἔχει n_1+n_2-2 βαθμούς ἐλευθερίας καί λέγεται ἀ θ ρ ο ι σ τ ι κ ό ς ἐ κ τ ι μ η τ ή ς τῆς διακυμάνσεως σ^2. Ἔτσι τό ἐκτιμώμενο τυπικό σφάλμα τοῦ ἐκτιμητῆ $\overline{X}_1-\overline{X}_2$ τῆς διαφορᾶς εἶναι [βλ. (9.26)]

$$\hat{\sigma}_{\overline{X}_1-\overline{X}_2}=S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \quad . \tag{9.32}$$

Τά ἐκτιμώμενα τυπικά σφάλματα τῶν ἐκτιμητῶν $\hat{\mu}_1=\overline{X}_1$ καί $\hat{\mu}_2=\overline{X}_2$ εἶναι

$$\hat{\sigma}_{\overline{X}_1}=S/\sqrt{n_1}\ , \quad \hat{\sigma}_{\overline{X}_2}=S/\sqrt{n_2} \quad . \tag{9.33}$$

Μέ τίς παραπάνω ὑποθέσεις τό στατιστικό

$$\frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \quad \text{ἔχει κατανομή } t_{n_1+n_2-2} \tag{9.34}$$

γιά κάθε μ_1,μ_2 καί σ^2.

Ἔτσι τό καλύτερο $100(1-\alpha)\%$ διάστημα ἐμπιστοσύνης γιά τή διαφορά $\mu_1-\mu_2$ δίνεται ἀπό τίς σχέσεις

$$\boxed{\begin{aligned} L &= \overline{X}_1-\overline{X}_2-t_{\alpha/2,n_1+n_2-2}\, S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \\ U &= \overline{X}_1-\overline{X}_2+t_{\alpha/2,n_1+n_2-2}\, S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \quad . \end{aligned}} \tag{9.35}$$

Οἱ ὑποθέσεις (a),(b),...,(e) τῆς (9.29) γιά τή διαφορά $\mu_1-\mu_2$ ἐλέγχονται μέ τή στατιστική συνάρτηση

$$t = \frac{X_1 - X_2 - \delta_0}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ πού ἔχει κατανομή $t_{n_1+n_2-2}$ ὅταν ἰσχύει ἡ H_0. (9.36)

Οἱ ἀντίστοιχες κρίσιμες περιοχές εἶναι οἱ ἀκόλουθες:

γιά τήν (a) $C = (t_\alpha, +\infty)$, δηλ. $t > t_\alpha$

" " (b) $C = (t_\alpha, +\infty)$, δηλ. $t > t_\alpha$

" " (c) $C = (-\infty, -t_\alpha)$, δηλ. $t < -t_\alpha$

" " (d) $C = (-\infty, -t_\alpha)$, δηλ. $t < -t_\alpha$

" " (e) $C = (-\infty, -t_{\alpha/2}) U (t_{\alpha/2}, \infty)$

δηλ. $|t| > t_{\alpha/2}$

Τά διαστήματα ἐμπιστοσύνης καί τά τέστ γιά μία ἀπό τίς μέσες τιμές μ_i χρησιμοποιοῦν τόν ἀθροιστικό ἐκτιμητή S τοῦ σ καί οἱ ποσότητες $t_{\alpha/2}$ καί t ἔχουν n_1+n_2-2 βαθμούς ἐλευθερίας.

Παράδειγμα 9.5:

Δύο ἀνεξάρτητες ὁμάδες παιδιῶν 1 καί 2 ἀποτελοῦνται ἀπό 10 παιδιά ἡ κάθε μιά. Τά παιδιά τῆς ὁμάδας 1 προέρχονται ἀπό ὑπερτασικούς γονεῖς ἐνῶ τά παιδιά τῆς ὁμάδας 2 προέρχονται ἀπό γονεῖς μέ κανονική πίεση. Ἡ συστολική πίεση τοῦ αἵματος μετριέται στά παιδιά τῶν δύο ὁμάδων μέ τά ἀκόλουθα ἀποτελέσματα:

Ὁμάδα 1: 100, 102, 96, 106, 110, 110, 120, 112, 112, 90

Ὁμάδα 2: 104, 88, 100, 98, 102, 92, 96, 100, 96, 96.

Μέ ἐπίπεδο σημαντικότητας =10% νά ἐλεγχθεῖ ἄν ὑπάρχει διαφορά μεταξύ τῆς μέσης πίεσης μ_1 παιδιῶν ἀπό ὑπερτασικούς γονεῖς καί τῆς μέσης φυσιολογικῆς πίεσης μ_2. Ὑποθέτομε ὅτι ἡ συστολική πίεση τοῦ αἵματος ἀκολουθεῖ κανονική κατανομή καί ὅτι ἡ μεταβλητότης τῶν πιέσεων εἶναι ἡ ἴδια καί γιά τούς δύο πληθυσμούς, παιδιά ὑπερτασικῶν γονέων καί παιδιά φυσιολογικῶν γονέων.

Γιά τά παραπάνω δεδομένα ἔχομε

Δεῖγμα 1	Δεῖγμα 2
$\Sigma X_{1i} = 1058$	$\Sigma X_{2i} = 972$
$\bar{X}_1 = 105{,}8$	$\bar{X}_2 = 97{,}2$
$\Sigma X_{1i}^2 = 112644$	$\Sigma X_{2i}^2 = 94680$

Ἔτσι

$$\Sigma(X_{1i}-\bar{X}_1)^2 = \Sigma X_{1i}^2 - \frac{(\Sigma X_{1i})^2}{n_1} = 112644 - \frac{(1058)^2}{10} = 707{,}6$$

$$\Sigma(X_{2i}-\bar{X}_2)^2 = 94680 - \frac{(972)^2}{10} = 201{,}6$$

καί ὁ Πίνακας 9.1 εἶναι

Δεῖγμα	n	β.ε.	μέση τιμή	SS
1	10	9	105,8	707,6
2	10	9	97,2	201,6
Σύνολο		18		909,2

'Η μέση διαφορά πιέσεων $\mu_1-\mu_2$ έκτιμᾶται μέ τήν ποσότητα

$$\widehat{\mu_1-\mu_2}=\overline{X}_1-\overline{X}_2=8{,}6\ .$$

'Επειδή οἱ πληθυσμοί εἶναι κανονικοί καί ἔχουν τήν ἴδια μεταβλητότητα $\sigma_1^2=\sigma_2^2=\sigma^2$ ὁ ἀθροιστικός ἐκτιμητής τοῦ σ^2 εἶναι [βλ.*(9.31)*]

$$S^2=\frac{909{,}2}{18}=50{,}51 \qquad (S=7{,}11).$$

'Η ἀξιοπιστία τοῦ ἐκτιμητῆ $\overline{X}_1-\overline{X}_2$ ζυγίζεται μέ τό τυπικό του σφάλμα [βλ.*(9.32)*].

$$\hat{\sigma}_{\overline{X}_1-\overline{X}_2}=\sqrt{50{,}51}\sqrt{\frac{1}{10}+\frac{1}{10}}=3{,}18\ ,$$

τό ὁποῖο εἶναι σχετικά μικρό.

'Η ὑπόθεση πού θέλομε νά ἐλέγξομε μέ ἐπίπεδο σημαντικότητας *10%* εἶναι [βλ.*(9.29) (e)* μέ $\delta_0=0$]

$$H_0:\mu_1=\mu_2 \quad \text{ἔναντι} \quad H_a:\mu_1\neq\mu_2\ .$$

Ἔτσι τό στατιστικό t μέ $\delta_0=0$ [βλ.*(9.36)*] ἔχει *18* β.ε. καί εἶναι

$$t=\frac{105{,}8-97{,}2}{7{,}11\sqrt{\frac{1}{10}+\frac{1}{10}}}=\frac{8{,}6}{3{,}18}=2{,}70\ .$$

'Η κρίσιμη περιοχή γιά $\alpha=10\%$ εἶναι

$$|t|>t_{0{,}05{,}18}=1{,}734\ .$$

'Επειδή $2{,}70>1{,}734$ ἀπορρίπτεται ἡ H_0 καί ἐξάγεται τό συμπέρασμα ὅτι ὑπάρχει στατιστικῶς σημαντική διαφορά μεταξύ τῆς μέσης πίεσης αἵματος παιδιῶν ἀπό ὑπερτασικούς γονεῖς καί τῆς μέσης φυσιολογικῆς πίεσης. Τό ποσοστό σφάλματος εἶναι *10%*.

Τ ό μ έ γ ε θ ο ς τ ῶ ν δ ε ι γ μ ά τ ω ν. Ἡ συζήτηση ἀκολουθεῖ τό σκεπτικό τοῦ προσδιορισμοῦ τοῦ μεγέθους τοῦ δείγματος πού ἀναπτύχθηκε στήν περίπτωση τοῦ ἑνός πληθυσμοῦ καί ἑνός δείγματος (βλ. §9Γ.1). Τό ἐνδιαφέρον μας εἶναι γιά τή παράμετρο $\delta=\mu_1-\mu_2$. Ὑποθέτομε ὅτι οἱ διακυμάνσεις τῶν δύο πληθυσμῶν εἶναι ἴσες, δηλαδή $\sigma_1^2=\sigma_2^2=\sigma^2$, καί ὅτι ὁ ἐρευνητής γνωρίζει ἤ μπορεῖ νά καθορίσει μία λογική τιμή τοῦ σ. Ὑπάρχουν πολλές πειραματικές περιπτώσεις πού οἱ παραπάνω προϋποθέσεις ἱκανοποιοῦνται.

Ἔστω ὅτι θέλομε νά ἐκτιμήσομε τή διαφορά $\delta=\mu_1-\mu_2$ κατά τέτοιο τρόπο ὥστε ἡ δειγματική διαφορά $\overline{X}_1-\overline{X}_2$ νά εὑρίσκεται ἐντός w μονάδων ἀπό τό δ μέ γνωστή ἐμπιστοσύνη $100(1-\alpha)\%$. Μέ ἄλλα λόγια θέλομε τό μῆκος τοῦ Δ.Ε. *(9.35)* νά εἶναι $2w$. Ἔτσι

$$U - L = 2t_{\alpha/2,n_1+n_2-2}\, S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 2w\ .$$

Ὑπάρχουν πολλοί συνδιασμοί τῶν n_1 καί n_2 πού ἱκανοποιοῦν τήν παραπάνω σχέση. Θέτοντες $S=\sigma$ καί ὑποθέτοντες ὅτι τά *2* δείγματα εἶναι μεγάλα καί ἴσα, δηλαδή $n_1=n_2=n$, καί $t_{\alpha/2,n_1+n_2-2} = z_{\alpha/2}$ ἡ προηγούμενη σχέση γίνεται

$$z_{\alpha/2}\sigma\sqrt{\frac{1}{n}+\frac{1}{n}} = w$$

καί τελικά λύνοντας την ώς πρός n έχομε:

$$n = \frac{2Z^2_{\alpha/2}\sigma^2}{2} .$$

n μετρήσεις ἐκλέγονται ἀπό τόν πληθυσμό 1 καί n ἀπό τόν πληθυσμό 2.

9Γ. 4 Συγκρίσεις κατὰ ζεύγη

Οἱ συγκρίσεις κατά ζεύγη ἀναφέρονται σέ δεδομένα πού λαμβάνονται κατά ζεύγη. Τό Παράδειγμα 9.3 εἶναι μία τέτοια περίπτωση. Τά δεδομένα τῶν συγκρίσεων κατά ζεύγη ἐμφανίζονται σ υ ν ή θ ω ς σέ πειράματα τῶν ὁποίων ὁ σκοπός εἶναι ἡ διερεύνηση τῆς ἀποτελεσματικότητας μίας θεραπείας. Ζεύγη ἀτόμων ἤ πειραματοζώων ἤ πιό γενικά πειραματικῶν μονάδων σχηματίζονται μέ βάση διάφορους παράγοντες ὅπως π.χ. ἡ ἡλικία, τό βάρος, τό φύλο, ἡ ὀξύτητα τῆς ἀσθένειας, ἡ χημική σύνθεση, ἡ κληρονομικότητα (δίδυμα), τό περιβάλλον κ.λ.π., δηλ. παράγοντες πού ἐπηρεάζουν τίς μετρήσεις. Σέ ἕνα μέλος κάθε ζεύγους ἐφαρμόζεται ἡ θεραπεία ἐνῶ τό ἄλλο μέλος χρησιμεύει γιά ἔλεγχο *(control)*. Μετά λαμβάνονται οἱ μετρήσεις. Πιό φυσιολογικά, συγκρίσεις κατά ζεύγη ἐφαρμόζονται ὅταν ἡ θεραπεία ἐφαρμόζεται σέ κάθε ἕνα ἀπό n ἄτομα ἤ πειραματικές μονάδες καί λαμβάνονται δύο μετρήσεις, μία πρίν ἀπό τή θεραπεία καί μία μετά τή θεραπεία. Τό μέλος τοῦ ζεύγους στό ὁποῖο ἐφαρμόζεται ἡ θεραπεία ἐκλέγεται στήν τύχη. Σχέδια πειραμάτων μέ τά παραπάνω χαρακτηριστικά λέγονται καί σ χ έ δ ι α τ υ χ α ι ο π ο ι η μ έ ν ω ν *block* (μέ δύο πειραματικές μονάδες σέ κάθε *block*).

Ἡ παράμετρος πού μᾶς ἐνδιαφέρει ἐδῶ εἶναι ἡ $\delta=\mu_1-\mu_2$ ὅπου μ_1 εἶναι ἡ μέση τιμή τῆς θεραπείας καί μ_2 ἡ μέση φυσιολογική τιμή (τό *control*). Ἡ συμπερασματολογία γίνεται μέ τόν ἴδιο τρόπο ὅπως στήν §9Γ.1, περίπτωση μίας μέσης τιμῆς ἑνός πληθυσμοῦ καί ἑνός δείγματος. Τό δεῖγμα εἶναι οἱ διαφορές $D_i=X_{1i}-X_{2i}$, $i=1,\ldots,n$ ὅπου $X_{11},\ldots,X_{1n}$ οἱ μετρήσεις θεραπείας καί $X_{21},\ldots,X_{2n}$ οἱ μετρήσεις *control*. Βλέπε Παράδειγμα 9.3. Ἡ διαφορά $\delta=\mu_1-\mu_2$ ἐκτιμᾶται ἀμερόληπτα μέ τό

$$\bar{D} = \bar{X}_1-\bar{X}_2\ . \tag{9.38}$$

Τό ἐκτιμώμενο τυπικό σφάλμα τοῦ $\bar{D}$ εἶναι

$$\hat{\sigma}_{\bar{D}} = S_D/\sqrt{n}\ , \tag{9.39}$$

ὅπου

$$S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i-\bar{D})^2\ .$$

Ὑποθέτοντας ὅτι οἱ διαφορές $D_1,\ldots,D_n$, ἀποτελοῦν ἕνα τυχαῖο δεῖγμα ἀπό ἕνα κανονικό πληθυσμό, ἕνα $100(1-\alpha)\%$ διάστημα ἐμπιστοσύνης γιά τή διαφορά δ εἶναι τό

$$\boxed{\begin{aligned} L &= \bar{D}-t_{\alpha/2,n-1}S_D/\sqrt{n} \\ U &= \bar{D}+t_{\alpha/2,n-1}S_D/\sqrt{n}\ . \end{aligned}} \tag{9.40}$$

Ὑποθέσεις γιά τό δ π.χ., $H_0: \delta=\delta_0$ ἐλέγχονται μέ τό στατιστικό

$$t = \frac{\overline{D}-\delta_0}{S_D/\sqrt{n}} \quad \text{τό οποίο έχει κατανομή } t_{n-1} \text{ όταν αληθεύει ή } H_0. \qquad (9.41)$$

Μέ συγκρίσεις κατά ζεύγη μποροῦμε νά συγκρίνομε καί δύο διαφορετικές θεραπεῖες. Στά πλήρως τυχαιοποιημένα πειράματα συγκρίσεως δύο θεραπειῶν (βλ. §9Γ.3), οἱ θεραπεῖες ἐφαρμόζονται στήν τύχη πάνω σέ ὅλες τίς πειραματικές μονάδες. Συνήθως ἡ τυχαιοποίηση περιορίζεται ἔτσι ὥστε κάθε θεραπεία νά ἐφαρμόζεται στόν ἴδιο ἀριθμό πειραματικῶν μονάδων (ἴσες ἐπαναλήψεις κάθε θεραπείας). Τά δεδομένα πειραμάτων τῆς μορφῆς αὐτῆς ἀναλύονται μέ τή μεθοδολογία τῶν δύο δειγμάτων -δύο πληθυσμῶν (βλ. §9Γ.3).

Γιά τά παρπάνω πειράματα ἡ ἀποτελεσματικότητα καί ἀκρίβεια τῆς ἐκτίμησης εἶναι ἀντιστρόφως ἀνάλογες πρός τόν ἀθροιστικό ἐκτιμητή S^2 τῆς κοινῆς διακυμάνσεως σ^2, ὅσο μικρότερο τό S^2 τόσο μεγαλύτερη ἡ ἀκρίβεια. Γιά νά αὐξήσομε τήν ἀκρίβεια ἑνός πειράματος, μέ ἕνα συγκεκριμένο ποσό πειραματικοῦ ὑλικοῦ, πρέπει νά ἀλλάξομε τό σχέδιο τοῦ πειράματος γιά νά ἐλαττώσομε τήν ἀνεξήγητη μεταβλητότητα τῶν δεδομένων καθώς αὐτή μετριέται μέ τό S^2. Ἐπειδή τό S^2 προκύπτει ἀπό τή μεταβλητότητα μεταξύ πειραματικῶν μενοάδων πού δέχονται τήν ἴδια θεραπεία, εἴτε πρέπει νά ἐλαττώσομε τή μεταβλητότητα αὐτή χρησιμοποιῶντας περισσότερο ὁμογενές ὑλικό, εἴτε πρέπει νά ἀπαλείψομε μερική ἀπό τή μεταβλητότητα αὐτή χρησιμοποιῶντας ἐκ τῶν προτέρων πληροφορίες γιά τίς μετρήσεις (ἀποκρίσεις) πού θά προκύψουν ἀπό τίς πειραματικές μονάδες. Οἱ ἰδέες αὐτές ὑλοποιοῦνται μέ τά σχέδια τυχαιοποιημένων *block*.

Ἄν οἱ πειραματικές μονάδες ἐμφανίζονται κατά ζεύγη ἤ μποροῦν νά ὁμαδοποιηθοῦν σέ ζεύγη ἔτσι ὥστε ἡ μεταβλητότητα τῶν ἀποκρίσεων μεταξύ τῶν μελῶν κάθε ζεύγους νά εἶναι μικρότερη ἀπό τή μεταβλητότητα μεταξύ μελῶν διαφόρων ζευγῶν, μποροῦμε νά βελτιώσομε τήν ἀποτελεσματικότητα τοῦ πειράματος μας ἐφαρμόζοντας τίς δύο θεραπεῖες τυχαῖα στά δυό μέλη κάθε ζεύγους. Κάθε μία ἀπό τίς δύο θεραπεῖες ἐφαρμόζεται σέ ἕνα μέλος κάθε ζεύγους. Ἔτσι ἐκτιμᾶται ἡ διαφορά στήν ἐπίδοση τῶν δύο θεραπειῶν γιά κάθε ζεῦγος, ἡ μεταβλητότητα μεταξύ τῶν ζευγῶν δέν ὑπεισέρχεται. Ἄν ἡ τελευταία μεταβλητότητα εἶναι μεγαλύτερη ἀπό τή μεταβλητότητα ἐκτός τῶν ζευγῶν, τό πείραμα τυχαιοποιημένων *block* θά εἶναι καλύτερο (ἀποτελεσματικότερο γιά τή σύγκριση τῶν δύο θεραπειῶν) ἀπό τό πλήρως τυχαιοποιημένο πείραμα.

<u>Παράδειγμα 9.6</u>:

Γιά νά προσδιοριστεῖ ἄν μία ἐπεξεργασία θερμότητας εἶναι ἀποτελεσματική στό νά ἐλαττώνει ἤ ὄχι τόν ἀριθμό τῶν βακτηριδίων τοῦ γάλακτος παίρνονται *10* δείγματα γάλακτος καί γιά κάθε δεῖγμα γίνεται μικροσκοπική καταμέτρηση τῶν βακτηριδίων πρίν καί μετά τήν ἐπεξεργασία. Τά δεδομένα εἶναι οἱ λογάριθμοι τῶν ἀριθμῶν τῶν βακτηριδίων.

Λογάριθμοι ἀριθμῶν βακτηριδίων

Δεῖγμα	Πρίν ἀπό τήν ἐπεξεργασία(X_1)	Μετά τήν ἐπεξεργασία (X_2)	Διαφορές $D_i = X_{1i}-X_{2i}$
1	6,88	6,85	0,03
2	7,05	6,84	0,21
3	8,24	7,07	1,17
4	5,20	5,05	0,15
5	6,16	6,18	-0,02
6	6,67	6,71	-0,04
7	7,01	6,49	0,52
8	5,46	5,24	0,22
9	5,87	5,88	-0,01
10	6,54	6,41	0,13

Γιά τά παραπάνω δεδομένα ἔχομε

$$\Sigma D_i = 2,36 \qquad \overline{D} = 0,236$$

$$\Sigma D_i^2 = 1,7742 \qquad (\Sigma D_i)^2/n = 0,5570$$

$$S_D^2 = (1,7742-0,5570)/9 = 0,1352$$

$$S_{\overline{D}}^2 = 0,1352/10 = 0,0135.$$

Θέλομε νά ἐλέγξομε τήν ὑπόθεση μή ἀποτελεσματικότητας τῆς ἐπεξεργασίας $H_0: \delta=\mu_1-\mu_2=0$ ἔναντι τῆς ὑποθέσεως ἐλαττώσεως τοῦ ἀριθμοῦ τῶν βακτηριδίων $H_a: \delta>0$.

Ἡ σ.σ. τοῦ τέστ

$$t = \frac{\overline{D}-0}{S_D/\sqrt{n}}$$

ἔχει τιμή

$$t = \frac{0,236}{\sqrt{0,0135}} = 2,0312$$

καί ἐπειδή $t = 2,0312 > t_{0,05,\,9} = 1,833$ ἀπορρίπτομε τήν H_0 μέ ἐπίπεδο σημαντικότητας 5% καί βγάζομε τό συμπέρασμα ὅτι ἡ ἐπεξεργασία θερμάνσεως ἐλαττώνει τόν ἀριθμό τῶν βακτηριδίων.

9Γ. 5 Συμπερασματολογία γιὰ τὶς διακυμάνσεις

α) Μ ί α δ ι α κ ύ μ α ν σ η

Ἔστω $X_1,\ldots,X_n$ ἕνα τυχαῖο δεῖγμα ἀπό κάποιο πληθυσμό μέ μέση τιμή μ καί διακύμανση σ^2. Ἔχομε δεῖ στά προηγούμενα ὅτι ὁ καλύτερος (ἀπό πλευρᾶς ἀμεροληψίας) ἐκτιμητής τοῦ σ^2 εἶναι ἡ δειγματική διακύμανση

$$S^2 = \frac{1}{n-1}\,\Sigma(X_i-\overline{X})^2 = \frac{1}{n-1}\left[\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}\right].$$

Ὁμοίως ὁ ἐκτιμητής τοῦ σ εἶναι τό S.

Ἔστω τώρα ὅτι ὁ πληθυσμός εἶναι κ α ν ο ν ι κ ό ς. Τό καλύτερο $100(1-\alpha)\%$ διάστημα ἐμπιστοσύνης γιά τό σ^2 εἶναι τό

$$L = \frac{(n-1)S^2}{\chi^2_{\alpha,n-1}} \qquad U = \frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}} \,. \qquad (9.42)$$

Τό ἀντίστοιχο διάστημα γιά τήν τυπική ἀπόκλιση σ εἶναι τό

$$\left(\sqrt{(n-1)S^2/X^2_{\alpha,n-1}}\ ,\ \sqrt{(n-1)S^2/X^2_{1-\alpha,n-1}}\right).$$

Τά διαστήματα αυτά απορρέουν από τό γεγονός ὅτι ἡ ποσότης

$$\frac{(n-1)S^2}{\sigma^2} \sim X^2_{n-1}$$

γιά κάθε μ καί σ (βλ. θεώρημα 8.2).

Γιά νά ἐλέγξομε ὑποθέσεις τῆς μορφῆς

(a) $H_0:\sigma^2=\sigma_0^2$ (ἤ $\sigma=\sigma_0$) ἔναντι $H_a:\sigma^2>\sigma_0^2$ (ἤ $\sigma>\sigma_0$)

(b) $H_0:\sigma^2\leqq\sigma_0^2$ (ἤ $\sigma\leqq\sigma_0$) " $H_a:\sigma^2>\sigma_0^2$ (ἤ $\sigma>\sigma_0$)

(c) $H_0:\sigma^2=\sigma_0^2$ (ἤ $\sigma=\sigma_0$) " $H_a:\sigma^2<\sigma_0^2$ (ἤ $\sigma<\sigma_0$)

(d) $H_0:\sigma^2\geqq\sigma_0^2$ (ἤ $\sigma\geqq\sigma_0$) " $H_a:\sigma^2<\sigma_0^2$ (ἤ $\sigma<\sigma_0$)

(e) $H_0:\sigma^2=\sigma_0^2$ (ἤ $\sigma=\sigma_0$) " $H_a:\sigma^2\neq\sigma_0^2$ (ἤ $\sigma\neq\sigma_0$)

μέ ἐπίπεδο σημαντικότητας α χρησιμοποιοῦμε τό στατιστικό

$$\boxed{X^2 = \frac{(n-1)S^2}{\sigma_0^2} \quad \text{τό ὁποῖο ἔχει } X^2_{n-1} \text{ κατανομή ὅταν } \sigma^2=\sigma_0^2\ .} \qquad (9.43)$$

καί τίς ἀκόλουθες κρίσιμες περιοχές:

γιά τήν (a) $C = (X^2_{\alpha,n-1},+\infty)$, δηλ. $X^2>X^2_{\alpha,n-1}$

" " (b) $C = (X^2_{\alpha,n-1},+\infty)$, δηλ. $X^2>X^2_{\alpha,n-1}$

" " (c) $C = (0,X^2_{1-\alpha,n-1})$, δηλ. $X^2<X^2_{1-\alpha,n-1}$

γιά τήν (d) $C = (0, X^2_{1-\alpha,n-1})$, δηλ. $X^2 < X^2_{1-\alpha,n-1}$

" " (e) $C = (0, X^2_{1-\alpha/2,n-1}) \cup (X^2_{\alpha/2,n-1}, +\infty)$.

Τό τέστ αὐτό λέγεται X^2 (χί - τ ε τ ρ ά γ ω ν ο).

Παράδειγμα 9.7:

Ὁ κατασκευαστής ἑνός ὀργάνου ἀκριβείας ἰσχυρίζεται ὅτι ἡ τυπική ἀπόκλιση τῶν μετρήσεων πού γίνονται μέ τό ὄργανο εἶναι $\sigma=2$. Ἄν σέ ἕνα πείραμα πάρομε μετρήσεις 4,2, 5,3, 10,3 ἀληθεύει ὁ ἰσχυρισμός τοῦ κατασκευαστοῦ; (α=5%).

Γιά τά δεδομένα τοῦ παραδείγματος ἔχομε

$$\Sigma X_i = 19{,}8 \;, \qquad \Sigma X_i^2 = 151{,}82, \quad n=3$$

$$\Sigma(X_i-\bar{X})^2 = 151{,}82-(19{,}8)^2/3 = 21{,}14$$

$$S^2 = \Sigma(X_i-\bar{X})^2/(n-1) = 21{,}14/2 = 10{,}57 \;.$$

Θέλομε νά ἐλέγξομε τήν

$$H_0 : \sigma = 2 \quad \text{ἔναντι} \quad H_a : \sigma \neq 2$$

Ἡ σ.σ. τοῦ τέστ εἶναι

$$X^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{2\times 10{,}57}{4} = 5{,}285$$

καί $X^2_{0,025,\,2} = 7{,}378$, $X^2_{0,975,\,2} = 0{,}0506$.

Έτσι έπειδή

$$X^2_{0,975,2} = 0,0506 < X^2 = 5,285 < X^2_{0,025,2} = 7,378$$

τά δεδομένα δέν μᾶς δίνουν άρκετές ένδείξεις γιά νά άπορρίψομε τόν ίσχυρισμό τοῦ κατασκευαστοῦ μέ $\alpha=5\%$. Ένα 95% διάστημα έμπιστοσύνης γιά τό σ εἶναι

$$L = \sqrt{\frac{(n-1)S^2}{X^2_{0,025,2}}} = \sqrt{\frac{21,14}{7,378}} = \sqrt{2,87} = 1,69$$

$$U = \sqrt{\frac{(n-1)S^2}{X^2_{0,975,2}}} = \sqrt{\frac{21,14}{0,0506}} = \sqrt{417,79} = 20,44$$

β) Σύγκριση δύο διακυμάνσεων

Παράδειγμα 9.8:

Δύο διαφορετικά ήλεκτρονικά ὄργανα A καί B, χρησιμοποιοῦνται γιά τή μέτρηση τῆς πιέσεως τοῦ ματιοῦ. Ένδιαφερόμαστε νά συγκρίνομε τά ὄργανα αύτά ὡς πρός τήν ποιότητα τους. Ή ποιότητα ένός όργάνου εἶναι μεγάλη ὅταν έ π α ν ε ι λ η μ έ ν ε ς μετρήσεις στό ἴδιο άντικείμενο ἔχουν μικρή μ ε τ α β λ η τ ό τ η τ α. Έτσι έπανειλημένες μετρήσεις στό ἴδιο μάτι καί τήν ἴδια περίπου χρονική στιγμή δίνουν τά άκόλουθα άποτελέσματα.

	A	B
Άριθμός μετρήσεων	$n_1=10$	$n_2=8$
Διακύμανση	$S_1^2=1,25$	$S_2^2=0,28$

Γενικά: Ἔστω δύο ἀνεξάρτητα τυχαῖα δείγματα μεγέθους n_1 καί n_2 ἀντίστοιχα ἀπό δύο κανονικούς πληθυσμούς μέ μέσες τιμές μ_1, μ_2 καί διακυμάνσεις σ_1^2, σ_2^2. Ἔστω $\overline{X}_1, S_1^2$ καί $\overline{X}_2, S_2^2$ οἱ μέσες τιμές καί διακυμάνσεις τῶν δύο δειγμάτων. Ἐνδιαφερόμαστε νά συγκρίνομε τίς δύο διακυμάνσεις. Ἡ μηδενική ὑπόθεση εἶναι $H_0{:}\sigma_1^2 = \sigma_2^2$. Σέ ἄλλες ἐφαρμογές ἡ ὑπόθεση αὐτή δίδει ἴση μεταβλητότητα σέ δύο μεθόδους μετρήσεων κ.λ.π.

Ἡ στατιστική συνάρτηση τοῦ τέστ εἶναι

$$\boxed{F = S_1^2/S_2^2 \sim F_{n_1-1,n_2-1} \quad \text{ὅταν ἰσχύει ἡ } H_0.} \tag{9.44}$$

(Βλέπε Θεώρημα 8.4). Τό τέστ αὐτό λέγεται F τέστ καί ἐφαρμόζεται γιά ὑποθέσεις τῆς ἀκόλουθης μορφῆς

(a)	$H_0{:}\sigma_1^2=\sigma_2^2$	ἔναντι	$H_\alpha{:}\sigma_1^2\neq\sigma_2^2$
(b)	$H_0{:}\sigma_1^2=\sigma_2^2$	"	$H_\alpha{:}\sigma_1^2>\sigma_2^2$
(c)	$H_0{:}\sigma_1^2=\sigma_2^2$	"	$H_\alpha{:}\sigma_1^2<\sigma_2^2$
(d)	$H_0{:}\sigma_1^2\leq\sigma_2^2$	"	$H_\alpha{:}\sigma_1^2>\sigma_2^2$
(e)	$H_0{:}\sigma_1^2\geq\sigma_2^2$	"	$H_\alpha{:}\sigma_1^2<\sigma_2^2$

μέ τίς ἀκόλουθες κρίσιμες περιοχές:

γιά τήν (a) $C = (0, F_{1-\alpha/2,n_1-1,n_2-1}) U (F_{\alpha/2,n_1-1,n_2-1}, +\infty)$

" " (b) $C = (F_{\alpha,n_1-1,n_2-1}, +\infty)$

" " (c) $C = (0, F_{1-\alpha,n_1-1,n_2-1})$

γιά τήν (d) $C = (F_{\alpha,n_1-1,n_2-1},+\infty)$

(e) $C = (0,F_{1-\alpha,n_1-1,n_2-1})$

Λύση τοῦ Παραδείγματος:

'Εδῶ ἔχομε νά ἐλέγξομε τήν $H_0:\sigma_1^2=\sigma_2^2$ ἔναντι $H_a:\sigma_1^2 \neq\sigma_2^2$.

"Εστω $\alpha=10\%$. 'Η σ.σ. τοῦ τέστ

$$F = S_1^2/S_2^2$$

ἔχει τιμή $F = 1,25/0,28 = 4,46$. 'Η κρίσιμη περιοχή μεγέθους 10% εἶναι $C = (0,F_{0,95,9,7})U(F_{0,05,9,7},+\infty)$. 'Επειδή

$$F_{1-a,n_1,n_2} = 1/F_a \;,$$

$$F_{0,95,9,7}=1/F_{0,05,7,9}=1/3,29=0,30 \text{ καί } F_{0,05,9,7}=3,68.$$

'Επειδή τώρα $F=4,46>3,68$ ἀπορρίπτομε τήν H_0 μέ ἐπίπεδο σημαντικότητας 10%. Δέν ὑπάρχει ἀμφιβολία ὅτι τό ὄργανο B εἶναι καλύτερο ἀπό τό A. Αὐτό εἶναι φανερό καί ἀπό τιμές τῶν S_1^2 καί S_2^2 πού εἶναι διαφορετικές. Δέν ἀρκεῖ ὅμως αὐτό, διότι ἄν π.χ. $S_2^2=0,35$ τό ἀντικείμενο δέν εἶναι στατιστικῶς σημαντικό καί ἡ διαφορά μεταξύ S_2^2 ὀφείλεται στήν τύχη 10%. Τό ποσοστό σφάλματος εἶναι 10%.

9Γ. 6 Συμπερασματολογία γιά τή διαφορά δύο ποσοστῶν (πιθανοτήτων ἐπιτυχίας)

α) Σύγκριση δυό ποσοστῶν κατά ζεύγη.

<u>Παράδειγμα</u> <u>9.9:</u>

Τό κάθε ἕνα ἀπό *60* δείγματα ρινικῶν ἐκκριμάτων καλλιεργεῖται σέ δύο διαφορετικά μέσα *A* καί *B* μέ στόχο τή σύγκριση τῆς ἱκανότητας τῶν δύο μέσων νά ἀνιχνεύσουν ὁρισμένους ἰούς γρίππης. Τά ἀποτελέσματα δίνονται στόν ἀκόλουθο πίνακα ὅπου + σημαίνει ὅτι γίνεται ἀνίχνευση τοῦ ἰοῦ τῆς γρίππης καί - ὅτι δέν γίνεται.

Κατανομή *60* ρινικῶν ἐκκριμάτων βάσει ἀποτελεσμάτων καλλιέργεια σέ δύο μέσα

Μέσον *A*	+	+	-	-	
Μέσον *B*	+	+	-	-	
Αριθ.πτυέλων	*25*	*14*	*4*	*17*	Σύνολο *60* .

Τά ἀποτελέσματα αὐτά μποροῦν νά περιγραφοῦν ἰσοδύναμα μέ τόν ἀκόλουθο τρόπο:

		Μέσον *B* +	Μέσον *B* -	Σύνολο
Μέσον *A*	+	*25*	*14*	*39*
	-	*4*	*17*	*21*
Σύνολο		*29*	*31*	*60*

Ἡ μηδενική ὑπόθεση εἶναι ὅτι τά δύο μέσα εἶναι ἐξ ἴσου ἀποτελεσματικά στήν ἀνίχνευση ἰῶν γρίπης.

Γενικά: Συχνά ἔχομε ἕνα πληθυσμό τά μέλη τοῦ ὁποίου ὑποβάλλονται σέ δύο "δοκιμασίες ἤ θεραπεῖες" *1* καί *2* μέ σκοπό τή σύγκριση τῆς ἐπιδόσεως ἤ ἀποτελεσματικότητας τῶν δύο δοκιμασιῶν. Ἔστω p_1 καί p_2 οἱ ἄγνωστες πιθανότητες - ποσοστά "ἐπιτυχίας" τῶν δοκιμασιῶν *1* καί *2* ἀντιστοίχως. Στόχος μας εἶναι ὁ ἔλεγχος τῆς μηδενικῆς ὑποθέσεως $H_0 : p_1 = p_2$. N μέλη τοῦ πληθυσμοῦ ἐκλέγονται τυχαῖα καί τό κάθε ἕνα ὑποβάλλεται στίς δύο δοκιμασίες μέ τά ἀκόλουθα N ζεύγη ἀποτελεσμάτων:

Δοκιμασία 1	Δοκιμασία 2	Ἀριθμός ζευγῶν
E	E	x
E	A	y
A	E	z
A	E	w
	Σύνολο	N

E = ἐπιτυχία
A = ἀποτυχία

Τά ἴδια ἀποτελέσματα μποροῦν νά παρασταθοῦν ὡς ἀκολούθως

		Δοκιμασία 2		
		E	A	
Δοκιμασία 1	E	x	y	$x+y$
	A	z	w	$z+w$
		$x+z$	$y+w$	N

Τό ποσοστό ἐπιτυχίας τῆς δοκιμασίας *1* στό δεῖγμα εἶναι

$\hat{p}_1=\frac{x+y}{N}$ καί τῆς δοκιμασίας 2, $\hat{p}_2=\frac{x+z}{N}$. Εἶναι φανερό ὅτι ἡ τυχαία μεταβλητή $x+y$ εἶναι διωνυμική $B(N,p_1)$ καί ἡ τυχαία μεταβλητή $x+z$ εἶναι $B(N,p_2)$. Ἐπί πλέον οἱ δύο αὐτές μεταβλητές εἶναι ἐξηρτημένες καί ὅταν ἰσχύει ἡ H_0

$$E[x+y-(x+z)]=Np_1-Np_2=0 \quad \text{ἤ} \quad E(y)=E(z).$$

Μέ ἄλλα λόγια ἡ μηδενική ὑπόθεση λέει ὅτι ὁ ἀναμενόμενος ἀριθμός τῶν μή "ὁμοειδῶν" ζευγῶν εἶναι ὁ ἴδιος. Ἔστω τώρα $n=y+z$. Τότε δοθέντος ὅτι ὑπάρχουν n μή "ὁμοειδή" ζεύγη ἡ τυχαία μεταβλητή y ἔχει κατανομή $B(n,1/2)$ ὅταν ἰσχύει ἡ H_0. Ἄρα ἕνα τέστ μεγάλου δείγματος (κατά προσέγγιση) λαμβάνεται χρησιμοποιῶντας τή στατιστική συνάρτηση

$$\frac{y-n/2}{\sqrt{n/2}} \sim \text{ κατά προσέγγιση } N(0,1) \text{ ὅταν ἰσχύει } p_1=p_2 \tag{9.45}$$

Ἡ προσέγγιση βελτιώνεται ὅταν χρησιμοποιήσομε τή λεγόμενη "διόρθωση συνεχείας" μειώνοντας τήν ἀπόλυτη τιμή τοῦ $y-n/2$ κατά $1/2$. Τό παραπάνω τέστ λέγεται καί τέστ τοῦ *McNemar*.

Ἕνα κατά προσέγγιση $100(1-\alpha)\%$ διάστημα ἐμπιστοσύνης τῆς διαφορᾶς p_1-p_2 τῶν δύο ποσοστῶν δίδεται ἀπό τή σχέση

$$y-z\pm z_{\alpha/2}\frac{\sqrt{y+z}}{N} \tag{9.46}$$

Λ ύ σ η τ ο ῦ Π α ρ α δ ε ί γ μ α τ ο ς: Ἡ τιμή τῆς στατιστικῆς συναρτήσεως τοῦ τέστ *(9.45)* εἶναι

$$\frac{14-18/2}{\sqrt{18/2}} = \frac{5}{2,12} = 2,36$$

καί ἡ τιμή P τοῦ τέστ εἶναι

$$P=P(|Z|>2,36) = 0,0182<0,05.$$

Ἄρα τά πειραματικά δεδομένα μαρτυροῦν ὅτι οἱ ἱκανότητες ἀνιχνεύσεως τῶν δύο μέσων εἶναι διάφορες μέ ἐπίπεδο σημαντικότητας μικρότερο τοῦ *5%*. Συγκεκριμένα τό μέσον *A* εἶναι περισσότερο ἀποτελεσματικό τοῦ *B*. Ἡ διόρθωση συνεχείας δίδει σάν τιμή τῆς στατιστικῆς συναρτήσεως *(9.45)* τήν $4,5/2,12 = 2,12$, ἀποτέλεσμα ἐπίσης στατιστικῶς σημαντικό.

Ἕνα *95%* διάστημα ἐμπιστοσύνης τῆς διαφορᾶς μεταξύ τῶν δύο ποσοστῶν τῶν θετικῶν ἀνιχνεύσεων εἶναι

$$\frac{14-4}{60} \pm 1,96\,\frac{\sqrt{18}}{60}$$

$$=0,17\pm 0,14$$

$$=0,03 \text{ ἕως } 0,31$$

Παρατηροῦμε ὅτι ἄν καί τό παραπάνω τέστ σημαντικότητας στηρίζεται μόνο στίς δύο συχνότητες y καί z, ἡ ἐκτιμώμενη διαφορά μεταξύ τῶν ποσοστῶν θετικῶν ἀνιχνεύσεων καί τό τυπικό της σφάλμα ἐξαρτῶνται καί ἀπό τό N. Μέ ἄλλα λόγια ἡ ἔνδειξη γιά τήν ὕπαρξη διαφορᾶς στηρίζεται μόνο στά μή ὁμοειδῆ ζεύγη· τό μέγεθος τῆς διαφορᾶς ἐκτιμᾶται ἀπό ὅλα τά ἀριθμητικά δεδομένα.

β) Σ ύ γ κ ρ ι σ η δ ύ ο π ο σ ο σ τ ῶ ν χ ω ρ ί ς ζ ε ύ γ η.

Παράδειγμα 9.10

Σέ μία διπλότυφλη κλινική δοκιμή πού ἀφορᾶ ἕνα ἐνδεχόμενα ὑποτασικό φάρμακο χρησιμοποιοῦνται *100* ἀσθενεῖς.*50* ἀσθενεῖς ἐκλέγονται τυχαῖα καί τούς δίδεται τό ἐνεργό φάρμακο καί στούς ὑπόλοιπους δίδεται τό *placebo*. Ἰσοδύναμα τό φάρμακο καί τό *placebo* μποροῦν νά δοθοῦν στούς *100* ἀσθενεῖς στήν τύχη ἔτσι ὥστε *50* ἀπό αὐτούς νά πάρουν τό φάρμακο καί *50* τό *placebo*. Τό σχέδιο τοῦ πειράματος αὐτοῦ λέγεται π λ ή ρ ω ς τ υ χ α ι ο π ο ι η μ έ ν ο σ χ έ δ ι ο. Ἡ ἀντίδραση τοῦ ἀσθενοῦς στήν θεραπεία χαρακτηρίζεται σάν εὐνοϊκή καί μή εὐνοϊκή βάσει τοῦ βαθμοῦ καί τῆς διάρκειας τῆς πιέσεως τοῦ αἵματος. Μᾶς ἐνδιαφέρει νά διαπιστώσομε ἄν τό φάρμακο ἔχει τήν ἱκανότητα νά ἐλαττώσει τήν πίεση τοῦ αἵματος. Ἡ ἀντίδραση σέ *34* ἀπό τούς *50* ἀσθενεῖς πού πῆραν τό φάρμακο ἦταν εὐνοϊκή καί σέ *9* ἀπό τούς *50* πού πῆραν τό *placebo* ἦταν ἐπίσης εὐνοϊκή. Ἡ μηδενική ὑπόθεση εἶναι ὅτι τό φάρμακο δέν ἔχει κανένα ἀποτέλεσμα.

Γ ε ν ι κ ά: Ἔστω δύο πληθυσμοί τά μέλη τῶν ὁποίων μποροῦν νά καταταχθοῦν σέ δύο κατηγορίες A καί B, π.χ. A εἶναι ἡ κατηγορία τῶν μελῶν πού ἔχουν ἕνα χαρακτηριστικό γνώρισμα a καί B ἡ κατηγορία ἐκείνων πού δέν ἔχουν τό a. Ἔστω p_1 καί p_2 οἱ πιθανότητες τῶν μελῶν τοῦ πληθυσμοῦ νά ἀνήκουν στίς κατηγορίες A καί B ἀντιστοίχως.

Ένα τυχαίο δείγμα μεγέθους n_1 ἐκλέγεται ἀπό τόν πρῶτο πληθυσμό μέ ἐπανατοποθέτηση, ἤ χωρίς ἐπανατοποθέτηση ἄν ὁ πληθυσμός εἶναι ἄπειρος καί περιέχει x_1 μέλη τῆς κατηγορίας A. Ὁμοίως ἕνα ἀνεξάρτητο τυχαῖο δεῖγμα μεγέθους n_2 ἐκλέγεται ἀπό τόν δεύτερο πληθυσμό καί περιέχει x_2 μέλη τῆς κατηγορίας B. Τό ποσοστό p_1 ἐκτιμᾶται μέ

$$\hat{p}_1 = x_1/n_1 \quad \text{καί τό} \quad p_2 \quad \text{μέ} \quad \hat{p}_2 = x_2/n_2 \,.$$

Προφανῶς

$$E(\hat{p}_1-\hat{p}_2) = p_1-p_2$$

καί

$$Var(\hat{p}_1-\hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} \,. \qquad (9.47)$$

Ἡ διακύμανση τῆς διαφορᾶς $\hat{p}_1-\hat{p}_2$ ἐκτιμᾶται ἀπό τήν παραπάνω σχέση ἀντικαθιστῶντας τά p_1 καί p_2 μέ $\hat{p}_1$ καί $\hat{p}_2$ ἀντιστοίχως, δηλαδή

$$\hat{Var}(\hat{p}_1-\hat{p}_2) = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} \,.$$

Ένα κατά προσέγγιση $100(1-\alpha)\%$ διάστημα ἐμπιστοσύνης τῆς διαφορᾶς p_1-p_2 τῶν δύο ποσοστῶν δίδεται ἀπό τή γνωστή σχέση

$$\hat{p}_1-\hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \,. \qquad (9.49)$$

῎Εστω τώρα ἡ μηδενική ὑπόθεση $H_0: p_1 = p_2$. ῎Εστω ἐπίσης p ἡ κοινή τιμή τῶν ποσοστῶν. Τά $\hat{p}_1$ καί $\hat{p}_2$ εἶναι καί τά δύο ἐκτιμητές τοῦ p καί δέν ἔχει νόημα νά ἐκτιμοῦμε τό p μέ δύο διαφορετικές ποσότητες. ῞Οταν ἀληθεύει ἡ H_0 τά δύο δείγματα προέρχονται στήν πραγματικότητα ἀπό τόν ἴδιο πληθυσμό καί ὁ ἄριστος ἐκτιμητής τοῦ p εὑρίσκεται ἑνώνοντας τά δείγματα, δηλαδή

$$\hat{p} = \frac{x_1+x_2}{n_1+n_2} \quad . \tag{9.50}$$

῎Ετσι ὁ ἐκτιμητής τῆς διακυμάνσεως τῆς $\hat{p}_1 - \hat{p}_2$ εἶναι

$$Var(\hat{p}_1-\hat{p}_2) = \hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right). \tag{9.51}$$

῾Η μηδενική ὑπόθεση ἐλέγχεται κατά προσέγγιση μέ τή στατιστική συνάρτηση

$$z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1) \text{ ὅταν ἰσχύει ἡ } H_0. \tag{9.52}$$

<u>Λύση</u> <u>τοῦ</u> <u>παραδείγματος</u>:

Τό ποσοστό τῶν ἀσθενῶν πού ἀντιδροῦν εὐνοϊκά στό φάρμακο εἶναι $\hat{p}_1 = 34/50 = 0,68$ δεδομένου ὅτι $n_1 = 50$ καί $x_1 = 34$. ῾Ομοίως τό ποσοστό τῶν ἀσθενῶν πού ἀντιδροῦν εὐνοϊκά μέ τό *placebo* εἶναι

$\hat{p}_2 = 9/50 = 0,18$ δεδομένου ὅτι $n_2 = 50$ καί $x_2 = 9$.

Τό κοινό ποσοστό εὐνοϊκῆς ἀντιδράσεως εἶναι

$$\hat{p} = \frac{39+4}{50+50} = \frac{43}{100} = 0,43.$$

Ἡ μηδενική ὑπόθεση εἶναι ὅτι τό φάρμακο δέν ἔχει κανένα ἀποτέλεσμα. Ἀκόμη πιό ἀρνητικά θά λέγαμε ὅτι τό φάρμακο ἔχει κανένα ἤ ἔχει χειρότερο ἀποτέλεσμα ἀπό τό *placebo*. Ἔτσι ἡ H_0 εἶναι p_1 p_2. Ἡ στατιστική συνάρτηση τοῦ τέστ (9.52) ἔχει τιμή

$$z = \frac{0,68-0,18}{\sqrt{0,43(1-0,43)(\frac{1}{50}+\frac{1}{50})}} = \frac{0,50}{\sqrt{0,43\times 0,57\times\frac{2}{50}}} = \frac{0,50}{0,099} = 5,05.$$

Μεγάλες τιμές τοῦ z ὑποδεικνύουν ἀπόρριψη τῆς H_0. Πράγματι ἡ τιμή P τοῦ τέστ εἶναι

$$P = P(z>5,05)\approx 0,0<0,05.$$

Ἄρα μέ ἐπίπεδο σημαντικότητας $\alpha=5\%$ τά πειραματικά δεδομένα ὁδηγοῦν στήν ἀπόρριψη τῆς ὑπόθεσης ὅτι τό φάρμακο δέν εἶναι ἀποτελεσματικό. Ὑπάρχουν ἀρκετές ἐνδείξεις ὅτι τό φάρμακο ἔχει τήν ἱκανότητα νά ἐλαττώσει τήν πίεση.

Ἡ διαφορά μεταξύ τῶν δύο ποσοστῶν ἀποτελεσματικότητας τοῦ φαρμάκου καί τοῦ *placebo* εἶναι $0,68-0,18=0,5$. Ἕνα 95% διάστημα ἐμπιστοσύνης τῆς διαφορᾶς p_1-p_2 ἔχει ὅρια

$$0,5 \pm 1,96\sqrt{\frac{0,68(1-0,68)}{50}+\frac{0,18(1-0,18)}{50}} = 0,5\pm 0,17.$$

9Γ. 7 Ἄλλα θέματα

Τά θέματα τῆς Στατιστικῆς Συμπερασματολογίας εἶναι πολλά καί πολύπλευρα. Τά θέματα αὐτά ἀποτελοῦν ἀντικείμενο παραπέρα μελέτης τῆς Στατιστικῆς καί ξεφεύγει ἀπό τό σκοπό τοῦ βιβλίου. Ἀναφέρουμε χαρακτηριστικά μερικούς τέτλους: τέστ καλῆς προσαρμογῆς, πίνακες συναφείας, ἀνάλυση k δειγμάτων, μή παραμετρική συμπερασματολογία, ἀνάλυση τῶν τάξεων τῶν μετρήσεων, γραμμική παλινδρόμηση καί συσχέτηση, ἀνάλυση τῆς διακυμάνσεως, πολυδιάστατη ἀνάλυση, κλπ. Σέ ὅλα αὐτά ἔχουμε ἐκτιμητική καί ἔλεγχο ὑποθέσεων. Ὁ ἀναγνώστης παραπέμπεται σέ προχωρημένα βιβλία τῆς Στατιστικῆς.

ΑΣΚΗΣΕΙΣ

9.1. Ένας πληθυσμός άποτελεῖται άπό τίς άκόλουθες τρεῖς τιμές $4, 5, 6$. Νά βρεθεῖ ὁ βαθμός έμπιστοσύνης τοῦ διαστήματος έμπιστοσύνης (L, U) γιά τή μέση τιμή τοῦ πληθυσμοῦ μ, ὅπου $L = \bar{X} - (\sigma^2_{\bar{X}})/2$, $U = \bar{X} + (\sigma^2_{\bar{X}})/2$ καί $\bar{X}$ ἡ μέση τιμή δείγματος μέ έπανάθεση μεγέθους $n = 2$.

9.2. Δώδεκα φυτά έκλέγονται στήν τύχη καί άπό κάθε τέτοιο φυτό έκλέγεται ἕνα φύλλο στήν τύχη. Γιά κάθε φύλλο προσδιορίζεται τό περιεχόμενο του σέ άσκορβικό όξύ μέ τά άκόλουθα άποτελέσματα:

9,35	8,65	11,68	12,77	8,81	9,52
8,68	9,82	10,29	10,99	10,76	10,55.

(i) Νά έκτιμηθεῖ τό μέσο περιεχόμενο άσκορβικοῦ όξέως καί νά βρεθεῖ τό τυπικό σφάλμα τοῦ έκτιμητῆ *(ii)* Νά βρεθεῖ ἕνα 95% διάστημα έμπιστοσύνης τοῦ μέσου περιεχομένου *(iii)* νά βρεθεῖ ἕνα διάστημα έμπιστοσύνης τοῦ σ.

9.3. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_{n_1})$ καί $\underset{\sim}{Y} = (Y_1, \ldots, Y_{n_2})$ δύο άνεξάρτητα τυχαῖα δείγματα άπό δύο κανονικές κατανομές μέ μέσες τιμές καί διακυμάνσεις (μ_x, σ^2) καί (μ_y, σ^2) άντίστοιχα. Νά δειχθεῖ ὅτι γιά κάθε $W \epsilon [0,1]$ ἡ στατιστική συνάρτηση $U = U(X, Y) = WS^2_x + (1-W)S^2_y$ εἶναι ἕνας άμερόληπτος έκτιμητής τοῦ σ^2.

9.4. Δίνεται τυχαῖο δεῖγμα $X_1, X_2, \ldots, X_9$ ἀπό ἕνα πληθυσμό μέ κανονική κατανομή $N(\mu, 1)$ καί θέλουμε νά ἐλέγξουμε τήν ὑπόθεση $H_0: \mu = 20$ ὡς πρός τήν ἐναλλακτική $H_a: \mu = 21{,}5$, χρησιμοποιώντας τό ἀκόλουθο κριτήριο: Ἀπορρίπτουμε τήν H_0 ἄν $\bar{X} \geq 21$ καί τήν δεχόμαστε ἄν $\bar{X} < 21$. Ὑπολογίστε τίς πιθανότητες σφαλμάτων τύπου I καί II καθώς καί τήν δύναμη (ἰσχύ) τοῦ κριτηρίου (τέστ).

9.5. 10 ἄτομα ἐκλέγονται τυχαίως ἀπό ἕνα πληθυσμό καί ἡ συστολική πίεση τοῦ αἵματός τους βρίσκεται $100, 102, 96, 106, 110, 110, 120, 112, 112, 90$. Εἶναι γνωστό ὅτι ἡ συστολική πίεση ἀκολουθεῖ τήν κανονική κατανομή μέ μέση τιμή μ καί τυπική ἀπόκλιση σ. *(i)* Κατασκευᾶστε ἕνα 95% διάστημα ἐμπιστοσύνης γιά τό μ. *(ii)* Νά ἐλεγχθεῖ ἡ μηδενική ὑπόθεση $H_0: \mu = 100$ μέ ἐπίπεδο σημαντικότητας 5%.

9.6. Δίνεται μιά παρατήρηση X ἀπό τήν κανονική κατανομή $N(\mu, 1)$ καί θέλουμε νά ἐλέγξουμε τήν ὑπόθεση $H_0: \mu = 10$ ὡς πρός τήν ἐναλλακτική $H_a: \mu = 14$ βασιζόμενοι στό ἀκόλουθο κριτήριο: Ἀπορρίπτουμε τήν H_0, ἐάν $X > 13$ καί δεχόμαστε τήν H_0 ὁπουδήποτε ἀλλοῦ. Ὑπολογίστε τίς πιθανότητες τῶν σφαλμάτων τύπου I καί II.

9.7. Δύο έργαστηριακές μέθοδοι A καί B διατίθενται γιά τή μέτρηση ένός μεγέθους καί εἶναι άμερόληπτοι. Ἡ διακύμανση τῶν μετρήσεων A εἶναι $\sigma_A^2 = 4$ καί τῶν μετρήσεων B $\sigma_B^2 = 3$. Οἱ μετρήσεις A στοιχίζουν 20 δραχμές ἡ μία καί οἱ μετρήσεις B 50 δραχμές ἡ μία. Διατίθενται 500 δραχμές. Ποιά μέθοδος εἶναι προτιμητέα;

9.8. Σέ 9 γυναῖκες έφαρμόζεται ὁρισμένη δίαιτα γιά τήν έλάττωση τοῦ βάρους τους. Ὁ παρακάτω πίνακας δίνει τά βάρη σέ Kgr πρό καί μετά τή δίαιτα.

Πρό	67	70	56	62	63	68	60	58	59
Μετά	62	60	58	56	54	56	59	52	52

Μέ έπίπεδο σημαντικότητας $0,01$ νά έλεγχτεῖ έάν ἡ δίαιτα έπέφερε έλάττωση τοῦ βάρους.

9.9. Ἕνας χημικός παρασκεύασε μία διάλυση πού σκοτώνει τό 70% τῶν έντόμων κάποιου εἶδους. Ποιό θά πρέπει νά εἶναι τό μέγεθος τοῦ δείγματος ὥστε, μέ βαθμό έμπιστοσύνης 95% , τά άποτελέσματα τῆς δειγματοληψίας νά εἶναι έντός $0,02$ τοῦ άληθινοῦ κλάσματος τῶν έντόμων πού σκοτώνονται;

9.10. Ἄν τό ὕψος τοῦ άνθρώπου άκολουθεῖ κανονική κατανομή μέ $\sigma = 6,35$ cm, ποιό θά πρέπει νά εἶναι τό μέγεθος τοῦ δείγματος ὥστε, μέ βαθμό ἐμπιστοσύνης 95% , ἡ μέση τιμή τοῦ δείγματος νά μή διαφέρει ἀπό τήν ἀληθινή μέση τιμή τοῦ πληθυσμοῦ περισσότερο ἀπό *1,27 cm* σέ ἀπόλυτο τιμή;

9.11. Ἀπό τά στοιχεῖα ἑνός ἐργαστηρίου φαίνονται τά ἀκόλουθα: 60 *transistors* τύπου *A* σέ ἕνα δεῖγμα χιλίων *transistors* βρέθηκαν ἐλαττωματικά καί 40 *transistors* τύπου *B* σέ ἕνα ἄλλο δεῖγμα ἐπίσης χιλίων *transistors* βρέθηκαν ἐλαττωματικά. Τά παραπάνω δεδομένα δίνουν ἀρκετές ἐνδείξεις πού νά φαίνεται ὅτι ὑπάρχει διαφορά στήν ποιότητα μεταξύ *transistors* τύπου *A* καί τύπου *B*;

9.12. Ἕνας νέος τύπος λαπτήρων, οἱ ὁποῖοι χρησιμοποιοῦνται σέ φλάς φωτογραφικῶν μηχανῶν, πρόκειται νά δοκιμαστεῖ προκειμένου νά ἐκτιμηθεῖ ἡ πιθανότητα p ὁ νέος τύπος λαμπτήρων νά δίδει τό ἀπαιτούμενο φῶς στήν κατάλληλη στιγμή. Γιά τό σκοπό αὐτό ἐλήφθη ἕνα δεῖγμα ἀπό 1000 λαμπτῆρες τοῦ τύπου αὐτοῦ καί εὑρέθη ὅτι 920 ἔδιναν τό ἀπαιτούμενο φῶς στήν κατάλληλη στιγμή. Σέ ποιά ὅρια θά κεῖται τό p μέ πιθανότητα 99%;

9.13. Δύο ἀνεξάρτητες ὁμάδες παιδιῶν *A* καί *B* ἀποτελοῦνται ἀπό 10 παιδιά ἡ κάθε μία. Τά παιδιά τῆς *A* ὁμάδας προέρχονται ἀπό ὑπερτασικούς γονεῖς ἐνῶ τά παιδιά τῆς *B* ὁμάδας ἀπό γονεῖς μέ κανονική πίεση. Ἡ συστολική πίεση τοῦ αἵματος μετριέται στά παιδιά τῶν δυό ὁμάδων μέ τά ἀκόλουθα ἀποτελέσματα:

Ὁμάδα A: *100, 102, 96, 106, 110, 110, 120, 112, 112, 90*

Ὁμάδα B: *104, 88, 100, 98, 102, 92, 100, 96 96, 96.*

Ὑποθέτουμε ὅτι ἡ συστολική πίεση ἀκολουθεῖ τήν κανονική κατανομή μέ μέση τιμή μ_A γιά τήν ὁμάδα A καί μ_B γιά τήν ὁμάδα B καί μέ κοινή διασπορά σ^2 καί γιά τίς δυό ὁμάδες. Ἐλέγξτε ἄν ὑπάρχει διαφορά ἀνάμεσα στά μ_A καί μ_B μέ ἐπίπεδο σημαντικότητας $\alpha = 10\%$.

9.14. Γιά τή σύγκριση δύο μεθόδων φυσιοθεραπείας ἕνα δεῖγμα ἀπό *18* ἀσθενεῖς ἐκλέγεται τυχαίως. Στούς *9* ἀσθενεῖς ἐφαρμόζεται ἡ παλαιά μέθοδος καί μετρᾶται ὁ χρόνος θεραπείας μέχρι νά πετύχουμε ἕνα ἐπιθυμητό ἀποτέλεσμα. Οἱ μετρήσεις ἔδωσαν (σέ μέρες θεραπείας) *32, 37, 35, 28, 41, 44, 35, 31, 34*. Στούς ὑπόλοιπους ἀσθενεῖς ἐφαρμόζεται ἡ καινούργια μέθοδος καί οἱ χρόνοι (μέρες) θεραπείας μέχρι νά πετύχουμε τό ἴδιο ἀποτέλεσμα εἶναι *35, 31, 29, 25, 34, 40, 27, 32, 31*. Ἐάν ὑποτεθεῖ ὅτι οἱ μετρήσεις ἀκολουθοῦν τήν κανονική κατανομή μέ μέσους μ_1 καί μ_2 ἀντίστοιχα καί μέ τήν ἴδια διασπορά σ^2 νά ἐλεγχθεῖ ἡ ὑπόθεση ὅτι οἱ δυό μέθοδοι δίνουν τά ἴδια ἀποτελέσματα μέ ἐπίπεδο σημαντικότητας *5%*;

9.15. Γιά νά ἐλέγξουμε τήν ἀποτελεσματικότητα ἑνός νέου λιπάσματος γιά τήν παραγωγή σίτου, χρησιμοποιοῦμε ἕνα ἀγρό ὁ ὁποῖος διαιρεῖται σέ *60* τετράγωνα μέ τά ἴδια ἀκριβῶς πλεονεκτήματα τό καθ'ἕνα. Τό νέο λίπασμα ἐφαρμόζεται τυχαίως σέ *30* τετράγωνα τό δέ παλιό ἐπίσης τυχαίως στά ὑπόλοιπα. Ἡ μέση ἀπόδοση τῶν *30* τετραγώνων στά ὁποῖα εἶχε χρησιμοποιηθεῖ τό νέο λίπασμα εἶναι *18* τόννοι καί ἡ τυπική ἀπόκλιση *0,6*. Τά ἀντίστοιχα μεγέθη γιά τά ὑπόλοιπα τετράγωνα στά ὁποῖα χρησιμοποιήθηκε τό παλιό λίπασμα εἶναι *17* τόννοι καί *0,5* ἀντίστοιχα. Μέ ἐπίπεδο σημαντικότητας *0,05* νά ἐλεγχθεῖ ἡ ὑπόθεση ὅτι δέν ὑπάρχει διαφορά μεταξύ τῶν δυό λιπασμάτων.

9.16. Τό *1974* μία ἰατρική ἐρευνητική ὁμάδα ἀνέφερε ὅτι ἀπό *3.000* ἄνδρες ἡλικίας *30 - 64*, οἱ ὁποῖοι εἶχαν μία ἤ περισσότερες καρδιακές προσβολές, ὑποβλήθηκαν σέ θεραπεία καί ἔγιναν καλά, *357* πέθαναν κατά τή διάρκεια τῶν *3* ἐτῶν πού ἀκολούθησαν τή θεραπεία. *(i)* Νά ἐκτιμηθεῖ ἡ πιθανότητα θανάτου ἐντός τριῶν ἐτῶν μετά τή θεραπεία καί νά ὑπολογιστεῖ τό σφάλμα ἐκτίμησης καί *(ii)* Νά ἐλεγχθεῖ ἡ ὑπόθεση ὅτι ἡ πιθανότητα θανάτου εἶναι

95% έναντι H_a:$p \neq 95\%$ μέ επίπεδο σημαντικότητας 5%.

9.17. Πιστεύεται ότι τά νεογέννητα μωρά τά όποία άσκοϋνται καί ύποβοηθοϋνται στό περπάτημα περπατοϋν τελικά κατά μέσο όρο στούς 9,75 μήνες. Έξι μωρά άσκοϋνται έντατικά καί παρατηροϋνται οί άκόλουθες ήλικίες περπατήματος (σέ μήνες): 9, 9,5 , 9,75, 10, 13, 19,5 . Νά έλεγχθεϊ στατιστικώς ό ίσχυρισμός τών 9,75 μηνών μέ $\alpha = 5\%$ ύποθέτοντες ότι ή ήλικίες περπατήματος άκολουθοϋν τήν κανονική κατανομή.

9.18. Ένα μικρό καί ένα μεγάλο φροντιστήριο έχουν ποσοστά έπιτυχίας στίς πανελλήνιες έξετάσεις τοϋ 1980 70% καί 80% άντίστοιχα. Τό μικρό φροντιστήριο έχει 30 ύποψήφιους καί τό μεγάλο 100. Είναι πράγματι τό μεγάλο φροντιστήριο καλύτερο άπό τό μικρό; ($\alpha = 5\%$).

9.19. Δυό ήλεκτρονικά όργανα A καί B χρησιμοποιοϋνται γιά τή μέτρηση τής πίεσης ένός συστήματος. Ή ποιότητα τών όργάνων έξαρτάται άπό τή μεταβλητότητα έπανειλημμένων μετρήσεων στό ίδιο σύστημα. Μία σειρά άπό άνεξάρτητες τέτοιες μετρήσεις τήν ίδια περίπου χρονική στιγμή έδωσε τά άκόλουθα άποτελέσματα

ὄργανο	A	B
ἀριθμ. μετρήσεων	$n_A = 10$	$n_B = 8$
διακύμανση	$S_A^2 = 1,25$	$S_B^2 = 0,28$

Μέ $\alpha = 10\%$ ὑπάρχει ἔνδειξη ποιοτικῆς διαφορᾶς τῶν ὀργάνων;

9.20. Ἔστω $X_1, \ldots, X_n$ τυχαῖο δεῖγμα ἀπό μιά Poisson κατανομή μέ παράμετρο λ. Νά βρεθοῦν δύο ἀμερόληπτοι ἐκτιμητές τοῦ λ.

9.21. Τά παρακάτω δεδομένα δείχνουν τήν πίεση τοῦ αἵματος ἐννέα ἀσθενῶν πού μετρήθηκε μέ δυό διαφορετικά ὄργανα A καί B.

		'Ασθενεῖς								
		1	2	3	4	5	6	7	8	9
ὄργανα	A	144	165	125	149	141	118	131	126	147
	B	147	167	124	152	146	120	135	126	149

Μέ ἐπίπεδο σημαντικότητας $0,01$ ἐξετᾶστε τόν ἰσχυρισμό ὅτι κατά μέσο ὅρο τό ὄργανο B δίνει μεγαλύτερες ἐνδείξεις ἀπό τό A.

9.22. 'Ο Pauling τό 1971 ἰσχυρίστηκε ὅτι μεγάλες ἡμερήσιες δόσεις ἀσκορβικοῦ ὀξέως μποροῦν νά ἐμποδίσουν ἤ νά θεραπεύσουν τό κοινό κρυολόγημα. Ἔτσι παρατηρήθηκε αὔξηση τῆς χρήσης βιταμίνης *C*. Ὁ Keith τό *1974* μελέτησε τήν ἐπίδραση μαζικῶν δόσεων ἀσκορβικοῦ ὀξέως στήν κατανάλωση τροφῆς, ἀνάπτυξη, βάρος νεφρῶν, μορφολογία καί σωματική σύνθεση νεαρῶν ἀρσενικῶν χοιριδίων. Τούς ἔδωσε ἡμερήσιες συμπληρωματικές δόσεις *250*, *500* καί *1000 mg*. ἀσκορβικοῦ ὀξέως γιά *9* ἑβδομάδες. Παρόμοια χοιρίδια χρησιμοποιήθηκαν γιά control. Ὁ παρακάτω πίνακας δίνει τήν περιεκτικότητα σέ λιπίδια (σέ ποσοστό βάρους σώματος) *8* χοιριδίων (*4* control καί *4* τῆς ὁμάδας τῶν *1000 mg.*).

Περιεκτικότητα λιπιδίων (%)

Control	(X)	23,8	15,4	21,7	18,0
1000 mg	(Y)	13,8	9,3	17,2	15,1

Ὑποθέτοντας ὅτι ἡ περιεκτικότητα ἀκολουθεῖ κανονική κατανομή νά ἐλεγχθεῖ ἡ ὑπόθεση ὅτι τό ἀσκορβικό ὀξύ δέν μεταβάλλει τό ποσοστό λιπιδίων, μέ ἐπίπεδο σημαντικότητας $\alpha = 5\%$.

9.23. Δυό διαφορετικά κράματα μετάλλων *I*, *II* χρησιμοποιοῦνται γιά τήν κατασκευή ἐξαρτημάτων μίας οἰκιακῆς συσκευῆς. Ἑκατό δείγματα ἀπό τό κάθε ἕνα ὑπόκεινται σέ ἕνα τέστ ἀντοχῆς σέ κάμψεις. Ἐλαττώματα παρατηρήθηκαν σέ *18* ἀπό ἐκεῖνα πού γίνονται μέ τό κράμα *I* καί σέ *26* ἀπό ἐκεῖνα πού γίνονται μέ τό κράμα *II*. Μποροῦμε νά συμπεράνουμε ὅτι τά κράματα ἔχουν τήν ἴδια ἀνθεκτικότητα ἤ ὅτι τό κράμα τύπου I εἶναι ἀνθεκτικώτερο τοῦ κράματος τύπου II, $\alpha = 5\%$.

9.24. Δυό μηχανές M_1 καί M_2 αὐτομάτου συσκευασίας ἔχουν ρυθμιστεῖ ἔτσι ὥστε νά γεμίζουν πακέτα μέ βάρος $1kgr$. 'Υπάρχουν ὑπόνοιες ὅτι οἱ δυό μηχανές δέν λειτουργοῦν ὁμοιόμορφα ὡς πρός τό βάρος τοῦ περιεχομένου. Γιά τό λόγο αὐτό λαμβάνονται 100 πακέτα ἀπό τήν παραγωγή κάθε μηχανῆς M_1, M_2 καί εὑρίσκεται ὅτι οἱ δειγματικοί μέσοι εἶναι ἀντίστοιχα 1,07 καί 1,18 kgr. 'Από προηγούμενες παρατηρήσεις γνωρίζουμε ὅτι $\sigma_1 = 0{,}10$ καί $\sigma_2 = 0{,}12$ kgr. α) 'Υπάρχει πράγματι διαφορά μεταξύ τῶν δυό μηχανῶν μέ ἐπίπεδο σημαντικότητας $\alpha = 0{,}05$; β) Νά βρεθεῖ ἡ ἰσχύς τοῦ τέστ πού χρησιμοποιεῖται στό α) ὅταν τό μέσο βάρος τοῦ πακέτου ἀπό τή μηχανή M_1 εἶναι κατά δ kgr μεγαλύτερο ἀπό τό μέσο βάρος τοῦ πακέτου γιά τή μηχανή M_2.

9.25. Νά κατασκευαστεῖ ἕνα διάστημα ἐμπιστοσύνης μέ ἐπίπεδο σημαντικότητας 95% γιά τήν παράμετρο $\theta = 3\mu + 7$ βάσει παρατηρηθέντος δείγματος μέ τιμές 8, 14, 11 ὅπου μ εἶναι ἡ μέση τιμή μιᾶς κανονικῆς κατανομῆς $N(\mu, \sigma^2)$ μέ σ^2 ἄγνωστο.

9.26. Οἱ τιμές κλεισίματος τῶν μετοχῶν δυό ἀνωνύμων ἑταιρειῶν παρατηρήθηκαν γιά μιά περίοδο 16 ἡμερῶν. Οἱ μέσες τιμές καί διακυμάνσεις ἦταν

$$\bar{y}_1 = 40{,}33\ ,\ \bar{y}_2 = 42{,}54\ ,\ s_1^2 = 1{,}54\ ,\ s_2^2 = 2{,}96.$$

Ἐξετᾶστε ἐάν τά παραπάνω ἀριθμητικά δεδομένα δίνουν ἀρκετές ἐνδείξεις πού νά μαρτυροῦν διαφορά στήν μ ε τ α β λ η τ ό τ η τ α τῶν τιμῶν τῶν μετοχῶν τῶν δυό πληθυσμῶν πού συσχετίζονται μέ τά παραπάνω δείγματα. Ὑποθέστε ὅτι οἱ κατανομές τῶν πληθυσμῶν εἶναι οἱ κανονικές $N(\mu_1, \sigma_1^2)$ καί $N(\mu_2, \sigma_2^2)$ ἀντίστοιχα $(\alpha = 10\%)$.

9.27. Ἀπό τά στοιχεῖα ἑνός νοσοκομείου φαίνεται ὅτι *60* ἄνδρες σέ ἕνα δεῖγμα ἀπό *1000* ἄνδρες ἔναντι *40* γυναικῶν σέ ἕνα ἄλλο δεῖγμα ἀπό *1000* γυναῖκες νοσηλεύονται στό νοσοκομεῖο αὐτό ἐξ αἰτίας καρδιακοῦ νοσήματος. Τά παραπάνω δεδομένα δίνουν ἀρκετές ἐνδείξεις πού νά φαίνεται ὅτι ὑπάρχει διαφορά στό ποσοστό καρδιακῶν παθήσεων μεταξύ τῶν ἀνδρῶν καί γυναικῶν πού μπῆκαν στό νοσοκομεῖο; $(\alpha = 5\%)$.

9.28. Ἕνα πείραμα σύγκρισης τῶν χρόνων ἀντίδρασης δυό φαρμάκων *A*, *B* ἔδωσε τά κάτωθι ἀποτελέσματα (σέ δευτερόλεπτα) γιά ἕνα τυχαῖο δεῖγμα *16* ἀτόμων.

Φάρμακο *A*	*1 3 2 1 2 1 3 2*
Φάρμακο *B*	*4 2 3 3 1 2 3 3*

Ἐάν οἱ κατανομές τῶν χρόνων ἀντίδρασης τῶν φαρμάκων *A* καί *B* εἶναι $N(\mu_A, \sigma^2)$ καί $N(\mu_B, \sigma^2)$ ἀντίστοιχα νά βρεθεῖ ἕνα *90%* διάστημα ἐμπιστοσύνης γιά τή διαφορά $\mu_A - \mu_B$.

ΠΑΡΑΡΤΗΜΑ

Σ τ α τ ι σ τ ι κ ο ί Π ί ν α κ ε ς

I. Διωνυμική Κατανομή

II. Κατανομή τοῦ Poisson

III. Τυπική Κανονική Κατανομή

IV. Κατανομή X^2

V. Κατανομή t

VI. Κατανομή F

ΠΙΝΑΚΑΣ Ι

Διωνυμική Κατανομή

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$$

n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
						p					
1	0	.9500	.9000	.8500	.8000	.7500	.7000	.6500	.6000	.5500	.5000
	1	.0500	.1000	.1500	.2000	.2500	.3000	.3500	.4000	.4500	.5000
2	0	.9025	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025	.2500
	1	.0950	.1800	.2550	.3200	.3750	.4200	.4550	.4800	.4950	.5000
	2	.0025	.0100	.0225	.0400	.0625	.0900	.1225	.1600	.2025	.2500
3	0	.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250
	1	.1354	.2430	.3251	.3840	.4219	.4410	.4436	.4320	.4084	.3750
	2	.0071	.0270	.0574	.0960	.1406	.1890	.2389	.2880	.3341	.3750
	3	.0001	.0010	.0034	.0080	.0156	.0270	.0429	.0640	.0911	.1250
4	0	.8145	.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915	.0625
	1	.1715	.2916	.3685	.4096	.4219	.4116	.3845	.3456	.2995	.2500
	2	.0135	.0486	.0975	.1536	.2109	.2646	.3105	.3456	.3675	.3750
	3	.0005	.0036	.0115	.0256	.0469	.0756	.1115	.1536	.2005	.2500
	4	.0000	.0001	.0005	.0016	.0039	.0081	.0150	.0256	.0410	.0625
5	0	.7738	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503	.0312
	1	.2036	.3280	.3915	.4096	.3955	.3602	.3124	.2592	.2059	.1562
	2	0214	.0729	.1382	.2048	.2637	.3087	.3364	.3456	.3369	.3125
	3	.0011	.0081	.0244	.0512	.0879	.1323	.1811	.2304	.2757	.3125
	4	.0000	.0004	.0022	.0064	.0146	.0284	.0488	.0768	.1128	.1562
	5	.0000	.0000	.0001	.0003	.0010	.0024	.0053	.0102	.0185	.0312
6	0	.7351	.5314	.3771	.2621	.1780	.1176	.0754	.0467	.0277	.0156
	1	.2321	.3543	.3993	.3932	.3560	.3025	.2437	.1866	.1359	.0938
	2	.0305	.0984	.1762	.2458	.2966	.3241	.3280	.3110	.2780	.2344
	3	.0021	.0146	.0415	.0819	.1318	.1852	.2355	.2765	.3032	.3125
	4	.0001	.0012	.0055	.0154	.0330	.0595	.0951	.1382	.1861	.2344
	5	.0000	.0001	.0004	.0015	.0044	.0102	.0205	.0369	.0609	.0938
	6	.0000	.0000	.0000	.0001	.0002	.0007	.0018	.0041	.0083	.0156
7	0	6983	.4783	.3206	.2097	.1335	.0824	.0490	.0280	.0152	.0078
	1	.2573	.3720	.3960	.3670	.3115	.2471	.1848	.1306	.0872	.0547
	2	.0406	.1240	.2097	.2753	.3115	.3177	.2985	.2613	.2140	.1641
	3	.0036	.0230	.0617	.1147	.1730	.2269	.2679	.2903	.2918	.2734
	4	.0002	.0026	.0109	.0287	.0577	.0972	.1442	.1935	.2388	.2734
	5	.0000	.0002	.0012	.0043	.0115	.0250	.0466	.0774	.1172	.1641
	6	.0000	.0000	.0001	.0004	.0013	.0036	.0084	.0172	.0320	.0547
	7	.0000	.0000	.0000	.0000	.0001	.0002	.0006	.0016	.0037	.0078
8	0	.6634	.4305	.2725	.1678	.1001	.0576	.0319	.0168	.0084	.0039
	1	.2793	.3826	.3847	.3355	.2670	.1977	.1373	.0896	.0548	.0312
	2	.0515	.1488	.2376	.2936	.3115	.2965	.2587	.2090	.1569	.1094
	3	.0054	.0331	.0839	.1468	.2076	.2541	.2786	.2787	.2568	.2188
	4	.0004	.0046	.0185	.0459	.0865	.1361	.1875	.2322	.2627	.2734

ΠΙΝΑΚΑΣ Ι: Διωνυμική κατανομή (συνέχεια)

		p									
n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
8	5	.0000	.0004	.0026	.0092	.0231	.0467	.0808	.1239	.1719	.2188
	6	.0000	.0000	.0002	.0011	.0038	.0100	.0217	.0413	.0703	.1094
	7	.0000	.0000	.0000	.0001	.0004	.0012	.0033	.0079	.0164	.0312
	8	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0007	.0017	.0039
9	0	.6302	.3874	.2316	.1342	.0751	.0404	.0207	.0101	.0046	.0020
	1	.2985	.3874	.3679	.3020	.2253	.1556	.1004	.0605	.0339	.0176
	2	.0629	.1722	.2597	.3020	.3003	.2668	.2162	.1612	.1110	.0703
	3	.0077	.0446	.1069	.1762	.2336	.2668	.2716	.2508	.2119	.1641
	4	.0006	.0074	.0283	.0661	.1168	.1715	.2194	.2508	.2600	.2461
	5	.0000	.0008	.0050	.0165	.0389	.0735	.1181	.1672	.2128	.2461
	6	.0000	.0001	.0006	.0028	.0087	.0210	.0424	.0743	.1160	.1641
	7	.0000	.0000	.0000	.0003	.0012	.0039	.0098	.0212	.0407	.0703
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0013	.0035	.0083	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0020
10	0	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025	.0010
	1	.3151	.3874	.3474	.2684	.1877	.1211	.0725	.0403	.0207	.0098
	2	.0746	.1937	.2759	.3020	.2816	.2335	.1757	.1209	.0763	.0439
	3	.0105	.0574	.1298	.2013	.2503	.2668	.2522	.2150	.1665	.1172
	4	.0010	.0112	.0401	.0881	.1460	.2001	.2377	.2508	.2384	.2051
	5	.0001	.0015	.0085	.0264	.0584	.1029	.1536	.2007	.2340	.2461
	6	.0000	.0001	.0012	.0055	.0162	.0368	.0689	.1115	.1596	.2051
	7	.0000	.0000	.0001	.0008	.0031	.0090	.0212	.0425	.0746	.1172
	8	.0000	.0000	.0000	.0001	.0004	.0014	.0043	.0106	.0229	.0439
	9	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016	.0042	.0098
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010
11	0	.5688	.3138	.1673	.0859	.0422	.0198	.0088	.0036	.0014	.0005
	1	.3293	.3835	.3248	.2362	.1549	.0932	.0518	.0266	.0125	.0054
	2	.0867	.2131	.2866	.2953	.2581	.1998	.1395	.0887	.0513	.0269
	3	.0137	.0710	.1517	.2215	.2581	.2568	.2254	.1774	.1259	.0806
	4	.0014	.0158	.0536	.1107	.1721	.2201	.2428	.2365	.2060	.1611
	5	.0001	.0025	.0132	.0388	.0803	.1321	.1830	.2207	.2360	.2256
	6	.0000	.0003	.0023	.0097	.0268	.0566	.0985	.1471	.1931	.2256
	7	.0000	.0000	.0003	.0017	.0064	.0173	.0379	.0701	.1128	.1611
	8	.0000	.0000	.0000	.0002	.0011	.0037	.0102	.0234	.0462	.0806
	9	.0000	.0000	.0000	.0000	.0001	.0005	.0018	.0052	.0126	.0269
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0007	.0021	.0054
	11	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0005
12	0	.5404	.2824	.1422	.0687	.0317	.0138	.0057	.0022	.0008	.0002
	1	.3413	.3766	.3012	.2062	.1267	.0712	.0368	.0174	.0075	.0029
	2	.0988	.2301	.2924	.2835	.2323	.1678	.1088	.0639	.0339	.0161
	3	.0173	.0852	.1720	.2362	.2581	.2397	.1954	.1419	.0923	.0537
	4	.0021	.0213	.0683	.1329	.1936	.2311	.2367	.2128	.1700	.1208
	5	.0002	.0038	.0193	.0532	.1032	.1585	.2039	.2270	.2225	.1934
	6	.0000	.0005	.0040	.0155	.0401	.0792	.1281	.1766	.2124	.2256
	7	.0000	.0000	.0006	.0033	.0115	.0291	.0591	.1009	.1489	.1934
	8	.0000	.0000	.0001	.0005	.0024	.0078	.0199	.0420	.0762	.1208
	9	.0000	.0000	.0000	.0001	.0004	.0015	.0048	.0125	.0277	.0537

ΠΙΝΑΚΑΣ Ι: Διωνυμική κατανομή (συνέχεια)

		p									
n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
12	10	.0000	.0000	.0000	.0000	.0000	.0002	.0008	.0025	.0068	.0161
	11	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010	.0029
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002
13	0	.5133	.2542	.1209	.0550	.0238	.0097	.0037	.0013	.0004	.0001
	1	.3512	.3672	.2774	.1787	.1029	.0540	.0259	.0113	.0045	.0016
	2	.1109	.2448	.2937	.2680	.2059	.1388	.0836	.0453	.0220	.0095
	3	.0214	.0997	.1900	.2457	.2517	.2181	.1651	.1107	.0660	.0349
	4	.0028	.0277	.0838	.1535	.2097	.2337	.2222	.1845	.1350	.0873
	5	.0003	.0055	.0266	.0691	.1258	.1803	.2154	.2214	.1989	.1571
	6	.0000	.0008	.0063	.0230	.0559	.1030	.1546	.1968	.2169	.2095
	7	.0000	.0001	.0011	.0058	.0186	.0442	.0833	.1312	.1775	.2095
	8	.0000	.0000	.0001	.0011	.0047	.0142	.0336	.0656	.1089	.1571
	9	.0000	.0000	.0000	.0001	.0009	.0034	.0101	.0243	.0495	.0873
	10	.0000	.0000	.0000	.0000	.0001	.0006	.0022	.0065	.0162	.0349
	11	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0012	.0036	.0095
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
14	0	.4877	.2288	.1028	.0440	.0178	.0068	.0024	.0008	.0002	.0001
	1	.3593	.3559	.2539	.1539	.0832	.0407	.0181	.0073	.0027	.0009
	2	.1229	.2570	.2912	.2501	.1802	.1134	.0634	.0317	.0141	.0056
	3	.0259	.1142	.2056	.2501	.2402	.1943	.1366	.0845	.0462	.0222
	4	.0037	.0349	.0998	.1720	.2202	.2290	.2022	.1549	.1040	.0611
	5	.0004	.0078	.0352	.0860	.1468	.1963	.2178	.2066	.1701	.1222
	6	.0000	.0013	.0093	.0322	.0734	.1262	.1759	.2066	.2088	.1833
	7	.0000	.0002	.0019	.0092	.0280	.0618	.1082	.1574	.1952	.2095
	8	.0000	.0000	.0003	.0020	.0082	.0232	.0510	.0918	.1398	.1833
	9	.0000	.0000	.0000	.0003	.0018	.0066	.0183	.0408	.0762	.1222
	10	.0000	.0000	.0000	.0000	.0003	.0014	.0049	.0136	.0312	.0611
	11	.0000	.0000	.0000	.0000	.0000	.0002	.0010	.0033	.0093	.0222
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0019	.0056
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0009
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
15	0	.4633	.2059	.0874	.0352	.0134	.0047	.0016	.0005	.0001	.0000
	1	.3658	.3432	.2312	.1319	.0668	.0305	.0126	.0047	.0016	.0005
	2	.1348	.2669	.2856	.2309	.1559	.0916	.0476	.0219	.0090	.0032
	3	.0307	.1285	.2184	.2501	.2252	.1700	.1110	.0634	.0318	.0139
	4	.0049	.0428	.1156	.1876	.2252	.2186	.1792	.1268	.0780	.0417
	5	.0006	.0105	.0449	.1032	.1651	.2061	.2123	.1859	.1404	.0916
	6	.0000	.0019	.0132	.0430	.0917	.1472	.1906	.2066	.1914	.1527
	7	.0000	.0003	.0030	.0138	.0393	.0811	.1319	.1771	.2013	.1964
	8	.0000	.0000	.0005	.0035	.0131	.0348	.0710	.1181	.1647	.1964
	9	.0000	.0000	.0001	.0007	.0034	.0116	.0298	.0612	.1048	.1527
	10	.0000	.0000	.0000	.0001	.0007	.0030	.0096	.0245	.0515	.0916
	11	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0074	.0191	.0417
	12	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0016	.0052	.0139
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010	.0032
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

ΠΙΝΑΚΑΣ Ι: Διωνυμική κατανομή (συνέχεια)

n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
16	0	.4401	.1853	.0743	.0281	.0100	.0033	.0010	.0003	.0001	.0000
	1	.3706	.3294	.2097	.1126	.0535	.0228	.0087	.0030	.0009	.0002
	2	.1463	.2745	.2775	.2111	.1336	.0732	.0353	.0150	.0056	.0018
	3	.0359	.1423	.2285	.2463	.2079	.1465	.0888	.0468	.0215	.0085
	4	.0061	.0514	.1311	.2001	.2252	.2040	.1553	.1014	.0572	.0278
	5	.0008	.0137	.0555	.1201	.1802	.2099	.2008	.1623	.1123	.0667
	6	.0001	.0028	.0180	.0550	.1101	.1649	.1982	.1983	.1684	.1222
	7	.0000	.0004	.0045	.0197	.0524	.1010	.1524	.1889	.1969	.1746
	8	.0000	.0001	.0009	.0055	.0197	.0487	.0923	.1417	.1812	.1964
	9	.0000	.0000	.0001	.0012	.0058	.0185	.0442	.0840	.1318	.1746
	10	.0000	.0000	.0000	.0002	.0014	.0056	.0167	.0392	.0755	.1222
	11	.0000	.0000	.0000	.0000	.0002	.0013	.0049	.0142	.0337	.0667
	12	.0000	.0000	.0000	.0000	.0000	.0002	.0011	.0040	.0115	.0278
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0008	.0029	.0085
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0018
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
17	0	.4181	.1668	.0631	.0225	.0075	.0023	.0007	.0002	.0000	.0000
	1	.3741	.3150	.1893	.0957	.0426	.0169	.0060	.0019	.0005	.0001
	2	.1575	.2800	.2673	.1914	.1136	.0581	.0260	.0102	.0035	.0010
	3	.0415	.1556	.2359	.2393	.1893	.1245	.0701	.0341	.0144	.0052
	4	.0076	.0605	.1457	.2093	.2209	.1868	.1320	.0796	.0411	.0182
	5	.0010	.0175	.0668	.1361	.1914	.2081	.1849	.1379	.0875	.0472
	6	.0001	.0039	.0236	.0680	.1276	.1784	.1991	.1839	.1432	.0944
	7	.0000	.0007	.0065	.0267	.0668	.1201	.1685	.1927	.1841	.1484
	8	.0000	.0001	.0014	.0084	.0279	.0644	.1134	.1606	.1883	.1855
	9	.0000	.0000	.0003	.0021	.0093	.0276	.0611	.1070	.1540	.1855
	10	.0000	.0000	.0000	.0004	.0025	.0095	.0263	.0571	.1008	.1484
	11	.0000	.0000	.0000	.0001	.0005	.0026	.0090	.0242	.0525	.0944
	12	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0081	.0215	.0472
	13	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0021	.0068	.0182
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0016	.0052
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
18	0	.3972	.1501	.0536	.0180	.0056	.0016	.0004	.0001	.0000	.0000
	1	.3763	.3002	.1704	.0811	.0338	.0126	.0042	.0012	.0003	.0001
	2	.1683	.2835	.2556	.1723	.0958	.0458	.0190	.0069	.0022	.0006
	3	.0473	.1680	.2406	.2297	.1704	.1046	.0547	.0246	.0095	.0031
	4	.0093	.0700	.1592	.2153	.2130	.1681	.1104	.0614	.0291	.0117
	5	.0014	.0218	.0787	.1507	.1988	.2017	.1664	.1146	.0666	.0327
	6	.0002	.0052	.0301	.0816	.1436	.1873	.1941	.1655	.1181	.0708
	7	.0000	.0010	.0091	.0350	.0820	.1376	.1792	.1892	.1657	.1214
	8	.0000	.0002	.0022	.0120	.0376	.0811	.1327	.1734	.1864	.1669
	9	.0000	.0000	.0004	.0033	.0139	.0386	.0794	.1284	.1694	.1855

ΠΙΝΑΚΑΣ I: Διωνυμική κατανομή (συνέχεια)

n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
							p				
18	10	.0000	.0000	.0001	.0008	.0042	.0149	.0385	.0771	.1248	.1669
	11	.0000	.0000	.0000	.0001	.0010	.0046	.0151	.0374	.0742	.1214
	12	.0000	.0000	.0000	.0000	.0002	.0012	.0047	.0145	.0354	.0708
	13	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0045	.0134	.0327
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011	.0039	.0117
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0009	.0031
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0006
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
19	0	.3774	.1351	.0456	.0144	.0042	.0011	.0003	.0001	.0000	.0000
	1	.3774	.2852	.1529	.0685	.0268	.0093	.0029	.0008	.0002	.0000
	2	.1787	.2852	.2428	.1540	.0803	.0358	.0138	.0046	.0013	.0003
	3	.0533	.1796	.2428	.2182	.1517	.0869	0422	.0175	.0062	.0018
	4	.0112	.0798	.1714	.2182	.2023	.1491	.0909	.0467	.0203	.0074
	5	.0018	.0266	.0907	.1636	.2023	.1916	.1468	.0933	.0497	.0222
	6	.0002	.0069	.0374	.0955	.1574	.1916	.1844	.1451	.0949	.0518
	7	.0000	.0014	.0122	.0443	.0974	.1525	.1844	.1797	.1443	.0961
	8	.0000	.0002	.0032	.0166	.0487	.0981	.1489	.1797	.1771	.1442
	9	.0000	.0000	.0007	.0051	.0198	.0514	.0980	.1464	.1771	.1762
	10	.0000	.0000	.0001	.0013	.0066	.0220	.0528	.0976	.1449	.1762
	11	.0000	.0000	.0000	.0003	.0018	.0077	.0233	.0532	.0970	.1442
	12	.0000	.0000	.0000	.0000	.0004	.0022	.0083	.0237	.0529	.0961
	13	.0000	.0000	.0000	.0000	.0001	.0005	.0024	.0085	.0233	.0518
	14	.0000	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0082	.0222
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0022	.0074
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0018
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
20	0	.3585	.1216	.0388	.0115	.0032	.0008	.0002	.0000	.0000	.0000
	1	.3774	.2702	.1368	.0576	.0211	.0068	.0020	.0005	.0001	.0000
	2	.1887	.2852	.2293	.1369	.0669	.0278	.0100	.0031	.0008	.0002
	3	.0596	.1901	.2428	.2054	.1339	.0716	.0323	.0123	.0040	.0011
	4	.0133	.0898	.1821	.2182	.1897	.1304	.0738	.0350	.0139	.0046
	5	.0022	.0319	.1028	.1746	.2023	.1789	.1272	.0746	.0365	.0148
	6	.0003	.0089	.0454	.1091	.1686	.1916	.1712	.1244	.0746	.0370
	7	.0000	.0020	.0160	.0545	.1124	.1643	.1844	.1659	.1221	.0739
	8	.0000	.0004	.0046	.0222	.0609	.1144	.1614	.1797	.1623	.1201
	9	.0000	.0001	.0011	.0074	.0271	.0654	.1158	.1597	.1771	.1602
	10	.0000	.0000	.0002	.0020	.0099	.0308	.0686	.1171	.1593	.1762
	11	.0000	.0000	.0000	.0005	.0030	.0120	.0336	.0710	.1185	.1602
	12	.0000	.0000	.0000	.0001	.0008	.0039	.0136	.0355	.0727	.1201
	13	.0000	.0000	.0000	.0000	.0002	.0010	.0045	.0146	.0366	.0739
	14	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0049	.0150	.0370
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0049	.0148
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0046
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	20	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

ΠΙΝΑΚΑΣ II

Κατανομή τοῦ Poisson

$$P(X=x) = e^{-\lambda}\lambda^{x}/x!$$

	λ									
x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	.9048	.8187	.7408	.6703	.6065	.5488	.4966	.4493	.4066	.3679
1	.0905	.1637	.2222	.2681	.3033	.3293	.3476	.3595	.3659	.3679
2	.0045	.0164	.0333	.0536	.0758	.0988	.1217	.1438	.1647	.1839
3	.0002	.0011	.0033	.0072	.0126	.0198	.0284	.0383	.0494	.0613
4	.0000	.0001	.0002	.0007	.0016	.0030	.0050	.0077	.0111	.0153
5	.0000	.0000	.0000	.0001	.0002	.0004	.0007	.0012	.0020	.0031
6	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0003	.0005
7	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

	λ									
x	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	.3329	.3012	.2725	.2466	.2231	.2019	.1827	.1653	.1496	.1353
1	.3662	.3614	.3543	.3452	.3347	.3230	.3106	.2975	.2842	.2707
2	.2014	.2169	.2303	.2417	.2510	.2584	.2640	.2678	.2700	.2707
3	.0738	.0867	.0998	.1128	.1255	.1378	.1496	.1607	.1710	.1804
4	.0203	.0260	.0324	.0395	.0471	.0551	.0636	.0723	.0812	.0902
5	.0045	.0062	.0084	.0111	.0141	.0176	.0216	.0260	.0309	.0361
6	.0008	.0012	.0018	.0026	.0035	.0047	.0061	.0078	.0098	.0120
7	.0001	.0002	.0003	.0005	.0008	.0011	.0015	.0020	.0027	.0034
8	.0000	.0000	.0001	.0001	.0001	.0002	.0003	.0005	.0006	.0009
9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0002

	λ									
x	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	.1225	.1108	.1003	.0907	.0821	.0743	.0672	.0608	.0550	.0498
1	.2572	.2438	.2306	.2177	.2052	.1931	.1815	.1703	.1596	.1494
2	.2700	.2681	.2652	.2613	.2565	.2510	.2450	.2384	.2314	.2240
3	.1890	.1966	.2033	.2090	.2138	.2176	.2205	.2225	.2237	.2240
4	.0992	.1082	.1169	.1254	.1336	.1414	.1488	.1557	.1622	.1680
5	.0417	.0476	.0538	.0602	.0668	.0735	.0804	.0872	.0940	.1008
6	.0146	.0174	.0206	.0241	.0278	.0319	.0362	.0407	.0455	.0504
7	.0044	.0055	.0068	.0083	.0099	.0118	.0139	.0163	.0188	.0216
8	.0011	.0015	.0019	.0025	.0031	.0038	.0047	.0057	.0068	.0081
9	.0003	.0004	.0005	.0007	.0009	.0011	.0014	.0018	.0022	.0027
10	.0001	.0001	.0001	.0002	.0002	.0003	.0004	.0005	.0006	.0008
11	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0002	.0002
12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

ΠΙΝΑΚΑΣ ΙΙ: Κατανομή τοῦ Poisson (συνέχεια)

λ

x	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	.0450	.0408	.0369	.0334	.0302	.0273	.0247	.0224	.0202	.0183
1	.1397	.1304	.1217	.1135	.1057	.0984	.0915	.0850	.0789	.0733
2	.2165	.2087	.2008	.1929	.1850	.1771	.1692	.1615	.1539	.1465
3	.2237	.2226	.2209	.2186	.2158	.2125	.2087	.2046	.2001	.1954
4	.1734	.1781	.1823	.1858	.1888	.1912	.1931	.1944	.1951	.1954
5	.1075	.1140	.1203	.1264	.1322	.1377	.1429	.1477	.1522	.1563
6	.0555	.0608	.0662	.0716	.0771	.0826	.0881	.0936	.0989	.1042
7	.0246	.0278	.0312	.0348	.0385	.0425	.0466	.0508	.0551	.0595
8	.0095	.0111	.0129	.0148	.0169	.0191	.0215	.0241	.0269	.0298
9	.0033	.0040	.0047	.0056	.0066	.0076	.0089	.0102	.0116	.0132
10	.0010	.0013	.0016	.0019	.0023	.0028	.0033	.0039	.0045	.0053
11	.0003	.0004	.0005	.0006	.0007	.0009	.0011	.0013	.0016	.0019
12	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005	.0006
13	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0002	.0002
14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

λ

x	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
0	.0166	.0150	.0136	.0123	.0111	.0101	.0091	.0082	.0074	.0067
1	.0679	.0630	.0583	.0540	.0500	.0462	.0427	.0395	.0365	.0337
2	.1393	.1323	.1254	.1188	.1125	.1063	.1005	.0948	.0894	.0842
3	.1904	.1852	.1798	.1743	.1687	.1631	.1574	.1517	.1460	.1404
4	.1951	.1944	.1933	.1917	.1898	.1875	.1849	.1820	.1789	.1755
5	.1600	.1633	.1662	.1687	.1708	.1725	.1738	.1747	.1753	.1755
6	.1093	.1143	.1191	.1237	.1281	.1323	.1362	.1398	.1432	.1462
7	.0640	.0686	.0732	.0778	.0824	.0869	.0914	.0959	.1002	.1044
8	.0328	.0360	.0393	.0428	.0463	.0500	.0537	.0575	.0614	.0653
9	.0150	.0168	.0188	.0209	.0232	.0255	.0280	.0307	.0334	.0363
10	.0061	.0071	.0081	.0092	.0104	.0118	.0132	.0147	.0164	.0181
11	.0023	.0027	.0032	.0037	.0043	.0049	.0056	.0064	.0073	.0082
12	.0008	.0009	.0011	.0014	.0016	.0019	.0022	.0026	.0030	.0034
13	.0002	.0003	.0004	.0005	.0006	.0007	.0008	.0009	.0011	.0013
14	.0001	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005
15	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0002

λ

x	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0
0	.0061	.0055	.0050	.0045	.0041	.0037	.0033	.0030	.0027	.0025
1	.0311	.0287	.0265	.0244	.0225	.0207	.0191	.0176	.0162	.0149
2	.0793	.0746	.0701	.0659	.0618	.0580	.0544	.0509	.0477	.0446
3	.1348	.1293	.1239	.1185	.1133	.1082	.1033	.0985	.0938	.0892
4	.1719	.1681	.1641	.1600	.1558	.1515	.1472	.1428	.1383	.1339

ΠΙΝΑΚΑΣ ΙΙ: Κατανομή τοῦ Poisson (συνέχεια)

	λ									
x	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0
5	.1753	.1748	.1740	.1728	.1714	.1697	.1678	.1656	.1632	.1606
6	.1490	.1515	.1537	.1555	.1571	.1584	.1594	.1601	.1605	.1606
7	.1086	.1125	.1163	.1200	.1234	.1267	.1298	.1326	.1353	.1377
8	.0692	.0731	.0771	.0810	.0849	.0887	.0925	.0962	.0998	.1033
9	.0392	.0423	.0454	.0486	.0519	.0552	.0586	.0620	.0654	.0688
10	.0200	.0220	.0241	.0262	.0285	.0309	.0334	.0359	.0386	.0413
11	.0093	.0104	.0116	.0129	.0143	.0157	.0173	.0190	.0207	.0225
12	.0039	.0045	.0051	.0058	.0065	.0073	.0082	.0092	.0102	.0113
13	.0015	.0018	.0021	.0024	.0028	.0032	.0036	.0041	.0046	.0052
14	.0006	.0007	.0008	.0009	.0011	.0013	.0015	.0017	.0019	.0022
15	.0002	.0002	.0003	.0003	.0004	.0005	.0006	.0007	.0008	.0009
16	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002	.0003	.0003
17	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001

	λ									
x	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7.0
0	.0022	.0020	.0018	.0017	.0015	.0014	.0012	.0011	.0010	.0009
1	.0137	.0126	.0116	.0106	.0098	.0090	.0082	.0076	.0070	.0064
2	.0417	.0390	.0364	.0340	.0318	.0296	.0276	.0258	.0240	.0223
3	.0848	.0806	.0765	.0726	.0688	.0652	.0617	.0584	.0552	.0521
4	.1294	.1249	.1205	.1162	.1118	.1076	.1034	.0992	.0952	.0912
5	.1579	.1549	.1519	.1487	.1454	.1420	.1385	.1349	.1314	.1277
6	.1605	.1601	.1595	.1586	.1575	.1562	.1546	.1529	.1511	.1490
7	.1399	.1418	.1435	.1450	.1462	.1472	.1480	.1486	.1489	.1490
8	.1066	.1099	.1130	.1160	.1188	.1215	.1240	.1263	.1284	.1304
9	.0723	.0757	.0791	.0825	.0858	.0891	.0923	.0954	.0985	.1014
10	.0441	.0469	.0498	.0528	.0558	.0588	.0618	.0649	.0679	.0710
11	.0245	.0265	.0285	.0307	.0330	.0353	.0377	.0401	.0426	.0452
12	.0124	.0137	.0150	.0164	.0179	.0194	.0210	.0227	.0245	.0264
13	.0058	.0065	.0073	.0081	.0089	.0098	.0108	.0119	.0130	.0142
14	.0025	.0029	.0033	.0037	.0041	.0046	.0052	.0058	.0064	.0071
15	.0010	.0012	.0014	.0016	.0018	.0020	.0023	.0026	.0029	.0033
16	.0004	.0005	.0005	.0006	.0007	.0008	.0010	.0011	.0013	.0014
17	.0001	.0002	.0002	.0002	.0003	.0003	.0004	.0004	.0005	.0006
18	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002
19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

	λ									
x	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
0	.0008	.0007	.0007	.0006	.0006	.0005	.0005	.0004	.0004	.0003
1	.0059	.0054	.0049	.0045	.0041	.0038	.0035	.0032	.0029	.0027
2	.0208	.0194	.0180	.0167	.0156	.0145	.0134	.0125	.0116	.0107
3	.0492	.0464	.0438	.0413	.0389	.0366	.0345	.0324	.0305	.0286
4	.0874	.0836	.0799	.0764	.0729	.0696	.0663	.0632	.0602	.0573
5	.1241	.1204	.1167	.1130	.1094	.1057	.1021	.0986	.0951	.0916
6	.1468	.1445	.1420	.1394	.1367	.1339	.1311	.1282	.1252	.1221
7	.1489	.1486	.1481	.1474	.1465	.1454	.1442	.1428	.1413	.1396
8	.1321	.1337	.1351	.1363	.1373	.1382	.1388	.1392	.1395	.1396
9	.1042	.1070	.1096	.1121	.1144	.1167	.1187	.1207	.1224	.1241

ΠΙΝΑΚΑΣ II: Κατανομή τοῦ Poisson (συνέχεια)

x	λ 7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
10	.0740	.0770	.0800	.0829	.0858	.0887	.0914	.0941	.0967	.0993
11	.0478	.0504	.0531	.0558	.0585	.0613	.0640	.0667	.0695	.0722
12	.0283	.0303	.0323	.0344	.0366	.0388	.0411	.0434	.0457	.0481
13	.0154	.0168	.0181	.0196	.0211	.0227	.0243	.0260	.0278	.0296
14	.0078	.0086	.0095	.0104	.0113	.0123	.0134	.0145	.0157	.0169
15	.0037	.0041	.0046	.0051	.0057	.0062	.0069	.0075	.0083	.0090
16	.0016	.0019	.0021	.0024	.0026	.0030	.0033	.0037	.0041	.0045
17	.0007	.0008	.0009	.0010	.0012	.0013	.0015	.0017	.0019	.0021
18	.0003	.0003	.0004	.0004	.0005	.0006	.0006	.0007	.0008	.0009
19	.0001	.0001	.0001	.0002	.0002	.0002	.0003	.0003	.0003	.0004
20	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002
21	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001

x	λ 8.1	8.2	8.3	8.4	8.5	8.6	8.7	8.8	8.9	9.0
0	.0003	.0003	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001
1	.0025	.0023	.0021	.0019	.0017	.0016	.0014	.0013	.0012	.0011
2	.0100	.0092	.0086	.0079	.0074	.0068	.0063	.0058	.0054	.0050
3	.0269	.0252	.0237	.0222	.0208	.0195	.0183	.0171	.0160	.0150
4	.0544	.0517	.0491	.0466	.0443	.0420	.0398	.0377	.0357	.0337
5	.0882	.0849	.0816	.0784	.0752	.0722	.0692	.0663	.0635	.0607
6	.1191	.1160	.1128	.1097	.1066	.1034	.1003	.0972	.0941	.0911
7	.1378	.1358	.1338	.1317	.1294	.1271	.1247	.1222	.1197	.1171
8	.1395	.1392	.1388	.1382	.1375	.1366	.1356	.1344	.1332	.1318
9	.1256	.1269	.1280	.1290	.1299	.1306	.1311	.1315	.1317	.1318
10	.1017	.1040	.1063	.1084	.1104	.1123	.1140	.1157	.1172	.1186
11	.0749	.0776	.0802	.0828	.0853	.0878	.0902	.0925	.0948	.0970
12	.0505	.0530	.0555	.0579	.0604	.0629	.0654	.0679	.0703	.0728
13	.0315	.0334	.0354	.0374	.0395	.0416	.0438	.0459	.0481	.0504
14	.0182	.0196	.0210	.0225	.0240	.0256	.0272	.0289	.0306	.0324
15	.0098	.0107	.0116	.0126	.0136	.0147	.0158	.0169	.0182	.0194
16	.0050	.0055	.0060	.0066	.0072	.0079	.0086	.0093	.0101	.0109
17	.0024	.0026	.0029	.0033	.0036	.0040	.0044	.0048	.0053	.0058
18	.0011	.0012	.0014	.0015	.0017	.0019	.0021	.0024	.0026	.0029
19	.0005	.0005	.0006	.0007	.0008	.0009	.0010	.0011	.0012	.0014
20	.0002	.0002	.0002	.0003	.0003	.0004	.0004	.0005	.0005	.0006
21	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002	.0002	.0003
22	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0001

x	λ 9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10
0	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0000
1	.0010	.0009	.0009	.0008	.0007	.0007	.0006	.0005	.0005	.0005
2	.0046	.0043	.0040	.0037	.0034	.0031	.0029	.0027	.0025	.0023
3	.0140	.0131	.0123	.0115	.0107	.0100	.0093	.0087	.0081	.0076
4	.0319	.0302	.0285	.0269	.0254	.0240	.0226	.0213	.0201	.0189

ΠΙΝΑΚΑΣ II: Κατανομή τοῦ Poisson (συνέχεια)

x	λ 9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10
5	.0581	.0555	.0530	.0506	.0483	.0460	.0439	.0418	.0398	.0378
6	.0881	.0851	.0822	.0793	.0764	.0736	.0709	.0682	.0656	.0631
7	.1145	.1118	.1091	.1064	.1037	.1010	.0982	.0955	.0928	.0901
8	.1302	.1286	.1269	.1251	.1232	.1212	.1191	.1170	.1148	.1126
9	.1317	.1315	.1311	.1306	.1300	.1293	.1284	.1274	.1263	.1251
10	.1198	.1210	.1219	.1228	.1235	.1241	.1245	.1249	.1250	.1251
11	.0991	.1012	.1031	.1049	.1067	.1083	.1098	.1112	.1125	.1137
12	.0752	.0776	.0799	.0822	.0844	.0866	.0888	.0908	.0928	.0948
13	.0526	.0549	.0572	.0594	.0617	.0640	.0662	.0685	.0707	.0729
14	.0342	.0361	.0380	.0399	.0419	.0439	.0459	.0479	.0500	.0521
15	.0208	.0221	.0235	.0250	.0265	.0281	.0297	.0313	.0330	.0347
16	.0118	.0127	.0137	.0147	.0157	.0168	.0180	.0192	.0204	.0217
17	.0063	.0069	.0075	.0081	.0088	.0095	.0103	.0111	.0119	.0128
18	.0032	.0035	.0039	.0042	.0046	.0051	.0055	.0060	.0065	.0071
19	.0015	.0017	.0019	.0021	.0023	.0026	.0028	.0031	.0034	.0037
20	.0007	.0008	.0009	.0010	.0011	.0012	.0014	.0015	.0017	.0019
21	.0003	.0003	.0004	.0004	.0005	.0006	.0006	.0007	.0008	.0009
22	.0001	.0001	.0002	.0002	.0002	.0002	.0003	.0003	.0004	.0004
23	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002
24	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

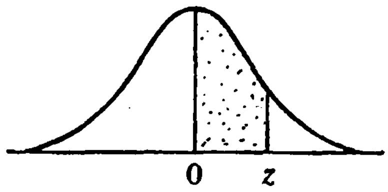
x	λ 11	12	13	14	15	16	17	18	19	20
0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2	.0010	.0004	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0000
3	.0037	.0018	.0008	.0004	.0002	.0001	.0000	.0000	.0000	.0000
4	.0102	.0053	.0027	.0013	.0006	.0003	.0001	.0001	.0000	.0000
5	.0224	.0127	.0070	.0037	.0019	.0010	.0005	.0002	.0001	.0001
6	.0411	.0255	.0152	.0087	.0048	.0026	.0014	.0007	.0004	.0002
7	.0646	.0437	.0281	.0174	.0104	.0060	.0034	.0018	.0010	.0005
8	.0888	.0655	.0457	.0304	.0194	.0120	.0072	.0042	.0024	.0013
9	.1085	.0874	.0661	.0473	.0324	.0213	.0135	.0083	.0050	.0029
10	.1194	.1048	.0859	.0663	.0486	.0341	.0230	.0150	.0095	.0058
11	.1194	.1144	.1015	.0844	.0663	.0496	.0355	.0245	.0164	.0106
12	.1094	.1144	.1099	.0984	.0829	.0661	.0504	.0368	.0259	.0176
13	.0926	.1056	.1099	.1060	.0956	.0814	.0658	.0509	.0378	.0271
14	.0728	.0905	.1021	.1060	.1024	.0930	.0800	.0655	.0514	.0387
15	.0534	.0724	.0885	.0989	.1024	.0992	.0906	.0786	.0650	.0516
16	.0367	.0543	.0719	.0866	.0960	.0992	.0963	.0884	.0772	.0646
17	.0237	.0383	.0550	.0713	.0847	.0934	.0963	.0936	.0863	.0760
18	.0145	.0256	.0397	.0554	.0706	.0830	.0909	.0936	.0911	.0844
19	.0084	.0161	.0272	.0409	.0557	.0699	.0814	.0887	.0911	.0888
20	.0046	.0097	.0177	.0286	.0418	.0559	.0692	.0798	.0866	.0888
21	.0024	.0055	.0109	.0191	.0299	.0426	.0560	.0684	.0783	.0846
22	.0012	.0030	.0065	.0121	.0204	.0310	.0433	.0560	.0676	.0769
23	.0006	.0016	.0037	.0074	.0133	.0216	.0320	.0438	.0559	.0669
24	.0003	.0008	.0020	.0043	.0083	.0144	.0226	.0328	.0442	.0557

ΠΙΝΑΚΑΣ ΙΙ: Κατανομή τοῦ Poisson (συνέχεια)

	λ									
z	11	12	13	14	15	16	17	18	19	20
25	.0001	.0004	.0010	.0024	.0050	.0092	.0154	.0237	.0336	.0446
26	.0000	.0002	.0005	.0013	.0029	.0057	.0101	.0164	.0246	.0343
27	.0000	.0001	.0002	.0007	.0016	.0034	.0063	.0109	.0173	.0254
28	.0000	.0000	.0001	.0003	.0009	.0019	.0038	.0070	.0117	.0181
29	.0000	.0000	.0001	.0002	.0004	.0011	.0023	.0044	.0077	.0125
30	.0000	.0000	.0000	.0001	.0002	.0006	.0013	.0026	.0049	.0083
31	.0000	.0000	.0000	.0000	.0001	.0003	.0007	.0015	.0030	.0054
32	.0000	.0000	.0000	.0000	.0001	.0001	.0004	.0009	.0018	.0034
33	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0005	.0010	.0020
34	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0006	.0012
35	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0007
36	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0004
37	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002
38	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
39	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

ΠΙΝΑΚΑΣ III

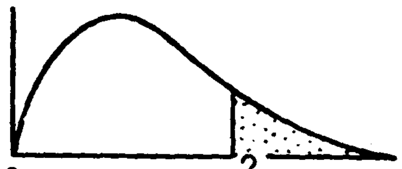
Τυπική Κανονική Κατανομή



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

ΠΙΝΑΚΑΣ IV Κατανομή χ^2

Τιμές $\chi^2_{\alpha,\nu}$:

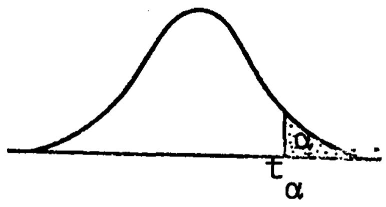


0 χ^2_α

ν	α = .995	α = .99	α = .975	α = .95	α = .05	α = .025	α = .01	α = .005	ν
1	.0000393	.000157	.000982	.00393	3.841	5.024	6.635	7.879	1
2	.0100	.0201	.0506	.103	5.991	7.378	9.210	10.597	2
3	.0717	.115	.216	.352	7.815	9.348	11.345	12.838	3
4	.207	.297	.484	.711	9.488	11.143	13.277	14.860	4
5	.412	.554	.831	1.145	11.070	12.832	15.086	16.750	5
6	.676	.872	1.237	1.635	12.592	14.449	16.812	18.548	6
7	.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278	7
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955	8
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589	9
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188	10
11	2.603	3.053	3.816	4.575	10.675	21.920	24.725	26.757	11
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300	12
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819	13
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319	14
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801	15
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267	16
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718	17
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156	18
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582	19
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997	20
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401	21
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796	22
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181	23
24	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.558	24
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928	25
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290	26
27	11.808	12.879	14.573	16.151	40.113	43.194	46.963	49.645	27
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993	28
29	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.336	29
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672	30

ΠΙΝΑΚΑΣ V
Κατανομή t

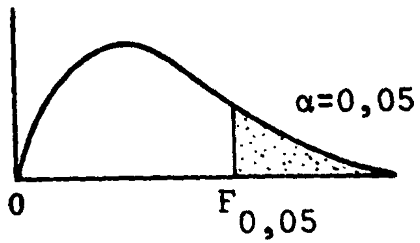
Τιμές $t_{\alpha,\nu}$:



ν	α = .10	α = .05	α = .025	α = .01	α = .005	ν
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
inf.	1.282	1.645	1.960	2.326	2.576	inf.

ΠΙΝΑΚΑΣ VI Κατανομή F

Τιμές $F_{0,05},\ \nu_1,\ \nu_2$:



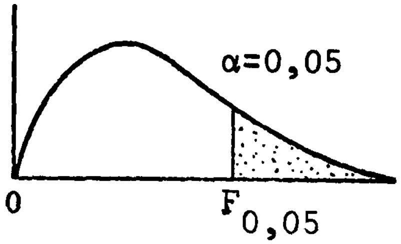
ν_1 = Βαθμοί ἐλευθερίας ἀριθμητῆ

ν_2 = Βαθμοί ἐλευθερίας παρονομαστῆ

	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

ΠΙΝΑΚΑΣ VI Κατανομή F (συνέχεια)

Τιμές $F_{0,05,\ \nu_1,\ \nu_2}$:



ν_1 = βαθμοί ἐλευθερίας ἀριθμητῆ

ν_1 = βαθμοί ἐλευθερίας παρονομαστῆ

	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4,052	5,000	5,403	5,625	5,764	5,859	5,928	5,982	6,023	6,056	6,106	6,157	6,209	6,235	6,261	6,287	6,313	6,339	6,366
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.4	26.3	26.2	26.1
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.7	13.6	13.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.53	2.45	2.36	2.27	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

ΒΙΒΛΙΟΓΡΑΦΙΑ

Α. ΕΛΛΗΝΙΚΑ

1. Ἀθανασοπούλου, Δ. (1975). Εἰσαγωγή εἰς τήν Στατιστικήν καί τήν θεωρίαν τῶν Πιθανοτήτων Ι, Ἀθήνα.

2. Δρακάτου, Κ. (1979). Στατιστική Ι, ΙΙ, ΙΙΙ, IV, Ἀθήνα.

3. Κάκουλλου, Θ. (1969). Μαθήματα Θεωρίας Πιθανοτήτων, Ἀθήνα.

4. Κάκουλλου, Θ. (1972). Στατιστική Θεωρία καί Ἐφαρμογαί, Ἀθήνα.

5. Κουνιᾶ, Σ. (1977). Μαθήματα Θεωρίας Πιθανοτήτων καί Στατιστικῆς, Μέρος Ι Πιθανότητες, Θεσσαλονίκη.

6. Κεβόρκ, Κ. (1972). Στατιστική Ι, ΙΙ, ΙΙΙ, Ἀθήνα.

7. Παναγιωτόπουλος, Δ.Χ. (1979). Ἐφαρμογές Πιθανοτήτων καί Στατιστικῆς στή μελέτη καί προγραμματισμό Δημοσίων Ἔργων, Ξάνθη. (Μετάφραση τοῦ Β.1).

8. Περσίδης, Σ.Κ. (1977). Πιθανότητες καί Στατιστική. ΕΣΠΙ, Ἀθήνα.(Μετάφραση τοῦ Β. 18).

9. Ρούσσα, Γ. (1973). Στοιχεῖα Πιθανοθεωρίας μετ' Ἐφαρμογῶν, Πάτρα.

10. Ρούσσα, Γ. (1976). Στατιστική Συμπερασματολογία 1, 2, Πάτρα.

Β. ΞΕΝΑ

1. Ang, A.H.S. and Tang, W.H. (1975). Probability Concepts in Engineering Planning and Design, Vol. I, John Wiley and Sons, Inc.
2. Chung, C.L. (1974). A Course in Probability Theory, Second Ed., Academic Press, New York.
3. David, F.N. (1962). Games, Gods and Gambling, Charles Griffin and Co. Ltd, London.
4. Feller, W. (1968). "An Introduction to Probability Theory and its Applications", Vol. I, Third Ed., John Wiley and Sons, Inc., New York.
5. Fisz, M. (1963). Probability Theory and Mathematical Statistics, Third Ed., John Wiley and Sons, Inc., New York.
6. Freund, J. (1971). Mathematical Statistics , Second Ed., Prentice - Hall, Inc., New Jersey.
7. Hogg, R. and Craig, A. (1970). Introduction to Mathematical Statistics, Third Ed., The Macmillan Co., New York.
8. Huntsberger, D.V., Billingsley, P. (1973). Elements of Statistical Inference, Third Edition, Allyn and Bacon, Inc., Boston.
9. Kempthorne, O. and Folks, L. (1971). Probability, Statistics and Data Analysis, Iowa State University Press, Ames, Iowa.

10. Lindgren, B.W. (1976). Statistical Theory, Third Edition, Macmillan Publishing Co., Inc., N.York.

11. Maistrov, L.E. (1974). Probability Theory -A Historical Sketch, Academic Press, New York.

12. Mendenhall, W. (1975). Introduction to Probability and Statistics, Fourth Edition, Duxbury Press, Massachussetts.

13. Mendenhall, W. and Scheaffer, R. (1973). Mathematical Statistics with Applications, Duxbury Press, Massachussetts.

14. Mood, A., Graybill, F., Boes, D. (1974). Introduction to the Theory of Statistics, Third Ed., McGraw-Hill, New York.

15. Moran, P.A.P. (1968). An Introduction to Probability Theory. Clarendon Press, Oxford.

16. Renyi, A. (1970). Probability Theory, North Holland Publishing Co., Amsterdam.

17. Roussas, G. (1973). A First Course in Mathematical Statistics, Addison-Wesley Publishing Co.,Reading Massachussetts.

18. Spiegel, M.R. (1972). Statistics, Schaum's Outline Series, McGraw-Hill Book Co., New York.

ΕΥΡΕΤΗΡΙΟ

16
×12
32
16
192

19

500

509 505
192
313

Γ. ΤΣΟΛΗΣ
ΤΥΠΟΓΡΑΦΕΙΑ–OFFSET–ΦΩΤΟΣΥΝΘΕΣΗ
Γ. ΜΙΧΑΗΛΙΔΗ 4 – ΙΩΑΝΝΙΝΑ – ΤΗΛ. 22877–26536