ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ

Η ΚΑΝΟΝΙΚΗ ΚΑΤΑΝΟΜΗ

Μαρίτσα Ευαγ. Γκόνου

ΕΡΓΑΣΙΑ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος
Συμπληρωματικής Ειδίκευσης στη Στατιστική
Μερικής Παρακολούθησης (Part-time)

Αθήνα
Ιούνιος 2003

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

Η ΚΑΝΟΝΙΚΗ ΚΑΤΑΝΟΜΗ

Μαρίτσα Ευαγ. Γκόνου

ΕΡΓΑΣΙΑ
Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος
Συμπληρωματικής Ειδίκευσης στη Στατιστική
Μερικής Παρακολούθησης (Part-time)

Αθήνα
Ιούνιος 2003

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

Εργασία που υποβλήθηκε ως μέρος των απαιτήσεων για την απόκτηση Μεταπτυχιακού Διπλώματος Συμπληρωματικής Ειδίκευσης στη Στατιστική Μερικής Παρακολούθησης (Part-time)

Η ΚΑΝΟΝΙΚΗ ΚΑΤΑΝΟΜΗ

Μαρίτσα Ευαγ. Γκόνου

Υπεύθυνο μέλος ΔΕΠ:
Α. Δημάκη
Επίκουρη Καθηγήτρια

Ο Διευθυντής Μεταπτυχιακών Σπουδών

Μιχαήλ Ζαζάνης
Αναπληρωτής Καθηγητής

Αφιερώνεται
Στην Οικογένειά μου
και στην Θεά Τύχη
για να την εξευμενίσω

ΕΥΧΑΡΙΣΤΙΕΣ

Αισθάνομαι την ανάγκη να εκφράσω τις ειλικρινείς ευχαριστίες μου, σε όλους εκείνους που με τον ένα ή τον άλλο τρόπο βοήθησαν στην αποπεράτωση αυτής της εργασίας.

Πρωτίστως εκφράζω τις ειλικρινέστατες ευχαριστίες μου στην επιβλέπουσα καθηγήτρια μου κυρία Αικατερίνη Δημάκη επίκουρο καθηγήτρια του Τμήματος Στατιστικής του Οικονομικού Πανεπιστημίου για την αδιάκοπη ηθική συμπαράσταση και τις πολύτιμες επιστημονικές συμβουλές και υποδείξεις της κατά τη διάρκεια της εργασίας μου, αλλά και για τη συνεχή και ακούραστη επιστημονική της καθοδήγηση.

Θα ήταν παράλειψη, στο σημείο αυτό, να μην εκφράσω τις ευχαριστίες μου προς όλους τους διδάσκοντες καθηγητές του Μεταπτυχιακού Τμήματος Στατιστικής του Οικονομικού Πανεπιστημίου Αθηνών στους οποίους οφείλω την επιστημονική μου κατάρτιση κατά τις μεταπτυχιακές μου σπουδές.

Θερμά ευχαριστώ όλα τα μέλη του Εργαστηρίου Στατιστικής καθώς και την Γραμματεία του Μεταπτυχιακού για την υποστήριξη και την πολύτιμη προσφορά τους, καθ΄ όλη τη διάρκεια των μεταπτυχιακών μου σπουδών.

Θέλω, ακόμη να πω ένα μεγάλο ευχαριστώ στους γονείς μου και στον αδελφό μου για την κατανόηση και την υπομονή τους καθ' όλη τη διάρκεια της εργασίας μου, των οποίων η αμέριστη συμπαράσταση και η συνεχής ενθάρρυνση ήταν απαραίτητη προϋπόθεση για να φτάσει σε πέρας αυτή η εργασία.

Ευχαριστώ θερμά όλους όσους από το ευρύτερο οικογενειακό και φιλικό περιβάλλον, με τον τρόπο τους, με βοήθησαν και με στήριξαν κατά το χρονικό διάστημα εκπόνησης της παρούσης εργασίας.

Τέλος ευχαριστώ όλους εκείνους που κατά κάποιο τρόπο συνέβαλαν στην πραγματοποίηση αυτής της εργασίας και τους οποίους ίσως παρέλειψα να αναφέρω.

ΒΙΟΓΡΑΦΙΚΟ ΣΗΜΕΙΩΜΑ

Η Μαρία Γκόνου γεννήθηκε και μεγάλωσε στο Αρτεμίσιο –Μαντινείας - Αρκαδίας, όπου και τελείωσε το δημοτικό. Τις Γυμνασιακές της σπουδές τις έκανε στην Τρίπολη, όπου ζει μέχρι σήμερα. Αποφοίτησε από το 3ο Λύκειο Τρίπολης.

Τις Πανεπιστημιακές της σπουδές τις παρακολούθησε στην Πάτρα, όπου και είναι πτυχιούχος του Μαθηματικού τμήματος του Πανεπιστημίου Πατρών.

Μετά την ολοκλήρωση των σπουδών της εργάστηκε σαν αναπληρώτρια καθηγήτρια σε πολλές πόλεις της Πελοποννήσου. Έχει διδάξει επί σειρά ετών Μαθηματικά σε μαθητές Γυμνασίου και Άλγεβρα, Γεωμετρία, Ανάλυση σε μαθητές Λυκείου.

Ταυτόχρονα παρακολούθησε μια σειρά από επιμορφωτικά σεμινάρια στα ''Μαθηματικά'', "στους Ηλεκτρονικούς Υπολογιστές και την Πληροφορική", στη "Διδακτική των Μαθηματικών" και στα "Παιδαγωγικά".

Είναι μέλος της Μαθηματικής Εταιρείας και ιδρυτικό μέλος του παραρτήματος Τρίπολης, όπου και διετέλεσε μέλος της Διοικούσης επιτροπής. Έχει παρακολουθήσει τα περισσότερα από τα συνέδρια που διοργανώνει η Μαθηματική Εταιρεία σε διάφορες πόλεις της Ελλάδος, και τις διημερίδες.

Διορίστηκε σαν μόνιμη καθηγήτρια στην Τήνο, όπου έκανε και την πρώτη της έρευνα "σχετικά με τις διατροφικές συνήθειες των κατοίκων του νησιού". Αυτή ήταν και η πρώτη επαφή της με την Στατιστική έρευνα. Την επόμενη χρονιά παρακολούθησε εξαμηνιαίο Σεμινάριο στη «ΣΤΑΤΙΣΤΙΚΗ» στο Οικονομικό Πανεπιστήμιο Αθηνών, όπου γοητεύθηκε από το νέο αντικείμενο και θέλησε να μάθει πιο πολλά για αυτό. Έτσι παρακολούθησε το Μεταπτυχιακό Πρόγραμμα «Συμπληρωματικής Ειδίκευσης στη Στατιστική» που διοργανώνει το Οικονομικό Πανεπιστήμιο Αθηνών για εργαζόμενους και εκπαιδευτικούς.

Σήμερα εργάζεται σαν μόνιμη καθηγήτρια Μαθηματικών στο Γυμνάσιο – Λύκειο Δημητσάνας.

Τα ερευνητικά της ενδιαφέροντα επικεντρώνονται στις Κατανομές, τα Γραμμικά μοντέλα, την Βιοστατιστική και την Ανάλυση δεδομένων.

ABSTRACT

Maritsa Ag. Gonou

The Normal Distribution

June 2003

It is known that many variables expressing natural characteristics or phenomena are, at least approximately, normally distributed. Due to the importance of the Normal Distribution special emphasis is given to its course of development over time. Chapter 3 presents and proves when necessary, its basic properties. The next chapters present algorithms approximating the normal distribution by others while special emphasis is given to its characteristic properties. Finally this dissertation focuses on the possibility of approximating certain distributions, like Poisson Binomial or Hypergeometric by the Normal Distribution.

ΠΕΡΙΛΗΨΗ

Μαρίτσα Ευαγ. Γκόνου

Κανονική Κατανομή

Ιούνιος 2003

Είναι γνωστό ότι πολλά φυσικά χαρακτηριστικά κατανέμονται, έστω και προσεγγιστικά, σύμφωνα με την κανονική κατανομή. Λόγω της ξεχωριστής θέσης που κατέχει, δίνεται ιδιαίτερη έμφαση στην ιστορική της διαδρομή. Στο τρίτο κεφάλαιο περιγράφονται και αποδεικνύονται, όπου κρίνεται σκόπιμο, ιδιότητες της κατανομής. Στη συνέχεια παρατίθενται αλγόριθμοι προσέγγισης της κανονικής κατανομής από άλλες κατανομές ενώ ιδιαίτερη έμφαση δίνεται στις χαρακτηριστικές της ιδιότητες. Τέλος, η μελέτη εστιάζεται στη δυνατότητα προσέγγισης κατανομών, όπως η Poisson, η διωνυμική και η υπεργεωμετρική από την κανονική κατανομή.

ΚΑΤΑΛΟΓΟΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΚΕΦΑΛΑΙΟ 4°

ΠΙΝΑΚΕΣ, ΕΠΕΚΤΑΣΕΙΣ, ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ

ΚΕΦΑΛΑΙΟ 5°

ΧΑΡΑΚΤΗΡΙΣΜΟΙ

ΚΕΦΑΛΑΙΟ 6°

ΚΕΝΤΡΙΚΟ ΟΡΙΑΚΟ ΘΕΩΡΗΜΑ

ΚΕΦΑΛΑΙΟ 7°

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

ΚΑΤΑΛΟΓΟΣ ΓΡΑΦΗΜΑΤΩΝ

ΚΕΦΑΛΑΙΟ 1$^{\text{ο}}$

Εισαγωγή

Η κανονική κατανομή είναι η πιο βασική και περισσότερο γνωστή κατανομή της Θεωρίας Πιθανοτήτων και της Στατιστικής. Αποτελεί τη θεμελιώδη κατανομή πάνω στην οποία στηρίζεται ολόκληρο το επιστημονικό οικοδόμημα αυτών των κλάδων. Έχει μοναδικές μαθηματικές, πιθανοθεωρητικές και στατιστικές ιδιότητες.

Ανακαλύφθηκε το 1733 από τον Abraham De Moivre ως το όριο της διωνυμικής κατανομής. Το 1774 ο Pierre Laplace εμελέτησε τις μαθηματικές της ιδιότητες. Παρ' όλα αυτά αποδίδεται στον Gauss ο οποίος την αναφέρει για πρώτη φορά σε μία εργασία του το 1809.

Ο Gauss αντιλήφθηκε την κανονική κατανομή ως καμπύλη των λαθών. Για πολλά έτη ήταν διευθυντής του αστρονομικού παρατηρητήριου του Goettingen. Προσέλκυσε την προσοχή του το γεγονός ότι οι διαφορετικές αναγνώσεις των μεταβάσεων των αστέρων προκλήθηκαν από τις μεμονωμένες διαφορές στο χρόνο αντίδρασης των παρατηρητών. Παρατήρησε ότι τα μοντέλα (κατανομές)των σφαλμάτων των μετρήσεων μπορούσαν να προσεγγιστούν ικανοποιητικά από μία συνεχή καμπύλη. Η καμπύλη αυτή ονομαζόταν "κανονική καμπύλη των σφαλμάτων" και αποδιδόταν στους νόμους της τύχης.

Μια άλλη σημαντική περιοχή των εφαρμογών της κανονικής κατανομής προτάθηκε από τον Quetelet που χρησιμοποίησε την κανονική κατανομή ως μια βάση της έννοιάς του "μέσου ατόμου", "I homme money". Από αυτή την άποψη, οι αποκλίσεις των αποτελεσμάτων από το μέσο όρο της κανονικής κατανομής χρησιμοποιούνται για την ταξινόμηση των ατόμων βασισμένη στις ερμηνείες των αποτελεσμάτων δοκιμής.

Η "καμπύλη των λαθών" μετονομάστηκε "κανονική καμπύλη" από τον Karl Pearson. Γράφοντας για την κανονική κατανομή ο Galton βεβαίωσε ότι "εάν οι Έλληνες την ήξεραν, θα είχαν αυτή. Βασιλεύει με την ηρεμία και την πληρότητα της στη πιο άγρια σύγχυση. Όσο πιο τεράστιος ο όχλος και όσο μεγαλύτερη η προφανής αναρχία, τόσο τελειότερη είναι η ταλάντευσή του. Είναι ο ανώτατος νόμος του μη αιτιατού. Όταν ένα μεγάλο δείγμα των χαοτικών στοιχείων αντιλαμβάνεται της τάξεως του μεγέθους της, μια ανυποψίαστη και ομορφότερη μορφή τακτικότητας αποδεικνύεται λανθάνουσα σε όλα εμπρός.

Όπως καταλαβαίνουμε η κανονική κατανομή είναι μία πολύ σημαντική κατανομή.

Υπάρχουν πολλές εξηγήσεις για την πανταχού παρουσία της κανονικής κατανομής. Η πλειοψηφία των φυσικών και νοητικών μεγεθών (χαρακτηριστικών) κατανέμεται ώστε να την προσεγγίζει. Εκτεινόμενη από το μείον έως το συν άπειρο και καλύπτοντας την μοναδιαία περιοχή, η κανονική κατανομή είναι ένα ιδανικό θεωρητικό πρότυπο του κόσμου όπου τα περισσότερα πράγματα είναι δυνατά, αλλά όχι πιθανά.

"Εντούτοις, μέσα στο μόνο πάρα πολύ πραγματικό κόσμο μας, η κανονική κατανομή συμβολίζει την ιστορτκή στρατηγική που είναι γνωστή στα ζωντανά όντα για να συνεχίζεται το γένος τους".

Όπως αναφέρθηκε η κανονική κατανομή διαδραματίζει έναν πολύ κυρίαρχο ρόλο. Μερικές από τις εφαρμογές της προκύπτουν μόνες τους με φυσικό τρόπο. Για παράδειγμα, περιμένουμε μερικές μεταβλητές να έχουν κανονική κατανομή, όπως το ύψος και το βάρος των ανθρώπων, ο δείκτης νοημοσύνης, το μήκος, ο όγκος, η επίδοση, οι βαθμοί και άλλα. Αρχικά πολλοί πληθυσμοί που αντιμετωπίζονται κατά την διάρκεια της έρευνας σε διάφορους τομείς φαίνονται να έχουν κανονική κατανομή σε ένα καλό βαθμό προσέγγισης. Επίσης δεδομένου του γεγονότος ότι το δείγμα είναι μεγάλο (πάνω από 30 παρατηρήσεις)η κατανομή του μέσου όρου θα είναι κανονική ανεξάρτητα από τον πληθυσμό προέλευσης. Αυτή είναι η βασική ιδέα του κεντρικού οριακού θεωρήματος. Η συνέπεια αυτού του αποτελέσματος είναι ότι σε ένα αριθμό στατιστικών μεθόδων δεν είναι απαραίτητο να προβληματιζόμαστε με την κατανομή του πληθυσμού προέλευσης για την οποία παίρνουμε ένα δείγμα. Δεν υπονοούμε εδώ ότι οι περισσότερες κατανομές που αντιμετωπίζονται στην πράξη είναι κανονικές, αλλά αυτές που συναντάμε στην πράξη οι περισσότερες είναι κανονικές.

Μια άλλη εκτίμηση που ευνοεί την κανονική κατανομή είναι το γεγονός ότι οι κατανομές δειγματοληψίας βασισμένες σε μια κανονική κατανομή του γεννήτορα πληθυσμού είναι αρκετά εύχρηστες αναλυτικά. Στη εξαγωγή των συμπερασμάτων για τους πληθυσμούς από τα δείγματα, είναι απαραίτητο να υπάρξουν οι κατανομές για τις διάφορες συναρτήσεις των δειγματικών παρατηρήσεων. Το μαθηματικό πρόβλημα παίρνοντας αυτές τις κατανομές είναι συχνά ευκολότερο για τα δείγματα από έναν κανονικό πληθυσμό απ' ό,τι από οποιουσδήποτε άλλον.

Στην εφαρμογή των στατιστικών μεθόδων βασισμένων στην κανονική κατανομή, ο ερευνητής πρέπει να ξέρει, τουλάχιστον προσεγγιστικά, τη γενική μορφή της συνάρτησης κατανομής που τα στοιχεία του ακολουθούν. Εάν είναι κανονική, μπορεί να χρησιμοποιήσει τις μεθόδους άμεσα.Εάν δεν είναι, μπορεί μερικές φορές να μετασχηματίσει τα στοιχεία του έτσι ώστε οι μετασχηματισμένες παρατηρήσεις να ακολουθούν μια κανονική κατανομή. Όταν ο ερευνητής δεν ξέρει τη μορφή κατανομής του πληθυσμού του, τότε μπορεί να χρησιμοποιήσει άλλες γενικότερες αλλά συνήθως λιγότερο ισχυρές, μεθόδους μεθόδων ανάλυσης αποκαλούμενες μη παραμετρικές .

Τέλος, η κανονική κατανομή εφαρμόζεται σε πολλές στατιστικές τεχνικές που χρησιμοποιούνται στην στατιστική συμπερασματολογία όπως είναι, η ανάλυση παλινδρόμησης, η ανάλυση διασποράς και άλλες.

Στους ελέγχους υποθέσεων που αναφέρονται στην μέση τιμή ενός πληθυσμού γίνεται η υπόθεση ότι ο δειγματικός μέσος κατανέμεται κανονικά. Στους ελέγχους υποθέσεων που αναφέρονται στις παραμέτρους του γραμμικού μοντέλου, ο προσθετέος που αντιστοιχεί στο τυχαίο σφάλμα θεωρείται ότι κατανέμεται σύμφωνα με την κανονική κατανομή. Τέλος, στους ελέγχους υποθέσεων στο πλαίσιο προβλημάτων ανάλυσης διασποράς, υποθέτουμε ότι η εντός των επιδράσεων μεταβλητότητα ακολουθεί την κανονική κατανομή.

Η εργασία αποτελείται συνολικά από εννέα κεφάλαια μαζί με την εισαγωγή.

◙Στο δεύτερο κεφάλαιο γίνεται μία ιστορική ανασκόπηση της μέχρι τώρα επιστημονικής έρευνας για την προέλευση, την ονομασία και την επέκταση της κανονικής κατανομής. Λόγω της μεγάλης σημασίας της κανονικής κατανομής, έχει δοθεί ιδιαίτερη προσοχή στην ιστορική ανάπτυξή της.

◙Στο τρίτο κεφάλαιο δίνεται η συνάρτηση πυκνότητας πιθανότητας της γενικής κανονικής και της τυποποιημένης κατανομής. Περιγράφονται λεπτομερώς όλες οι ιδιότητες της γενικής, αλλά και της τυποποιημένης κατανομής που προκύπτουν από την γραφική παράσταση και όχι μόνο. Οι ροπές, η ροπογεννήτρια συνάρτηση, η χαρακτηριστική συνάρτηση και πολλές άλλες ιδιότητες που είναι χαρακτηριστικά γνωρίσματα της κανονικής κατανομής. Σε αυτό το κεφάλαιο δίνονται επίσης οι μαθηματικές αποδείξεις για κάποιες ιδιότητες που κρίθηκαν απαραίτητες.

◙Το τέταρτο κεφάλαιο είναι το κεφάλαιο με τις προσεγγίσεις. Περιλαμβάνει α)την ιστορική προέλευση των Πινάκων β) τους αλγόριθμους προσέγγισης της αθροιστικής συνάρτησης κατανομής Φ(x)και του ποσοστιαίου σημείου x_p γ)την αναφορά στην αντικατάσταση της όψης μιας κανονικής κατανομής από άλλες κατανομές.

◙Στο πέμπτο κεφάλαιο παρουσιάζονται οι χαρακτηρισμοί της κανονικής κατανομής. Οι χαρακτηρισμοί έχουν παίξει ένα προεξέχοντα ρόλο στην θεωρητική ανάπτυξη της κατανομής. Μερικοί από τους πιο μαθηματικούς και θεωρητικούς χαρακτηρισμούς έχουν παραλειφθεί. Οι αποδείξεις παραλείπονται επειδή είναι έξω από το πνεύμα της εργασίας. Εν τούτοις έχουμε αναφέρει αρκετούς από τους πιο παλαιούς έως τους πιο πρόσφατους που θεωρούνται περισσότερο γνωστοί. Οι χαρακτηρισμοί αυτοί αφορούν:
α)Γραμμικές μορφές β)Γραμμικές αλλά και τετραγωνικές μορφές γ)Αυτοί που είναι βασισμένοι στα σύνολα ανεξαρτήτων τυχαίων μεταβλητών δ) Οι χαρακτηρισμοί που είναι βασισμένοι στις υπό συνθήκη κατανομές και στις ιδιότητες παλινδρόμησης ε)τις χαρακτηριστικές συναρτήσεις στ)Τέλος καλύπτονται οι ειδικοί χαρακτηρισμοί.

■Το έκτο κεφάλαιο αναφέρεται στο Κεντρικό Οριακό Θεώρημα. Το κεφάλαιο αυτό κρίθηκε απαραίτητο, διότι τα περισσότερα θεωρητικά πορίσματα για την χρήση της κανονικής κατανομής είναι βασισμένα σε μορφές κεντρικών οριακών θεωρημάτων.
Σε αυτό το κεφάλαιο εξηγείται η σημασία και ο ρόλος της κανονικής κατανομής σε πολλά φυσικά φαινόμενα. Για την καλλίτερη κατανόηση έγινε προσομοίωση της κανονικής κατανομής από την συσκευή του Galton Board.Επίσης έγινε με την χρήση του υπολογιστή παράδειγμα προσομοίωσης, όσον αναφορά την κατανομή του μέσου ενός τυχαίου δείγματος.

■Στο έβδομο κεφάλαιο γίνεται προσέγγιση μερικών κατανομών από την κανονική κατανομή. Συγκεκριμένα γίνεται προσέγγιση α) της διωνυμικής από την κανονική. Αναφέρεται πρακτικός αλλά και θεωρητικοί τρόποι. β) της Poisson από την κανονική, όπου δίνεται η απόδειξη γ) της Υπεργεωμετρικής από την κανονική κατανομή. Υπάρχουν ακόμη πολλές κατανομές όπως η Αρνητική διωνυμική, η Beta, η Von Mises, η X^2,η t,η F και άλλες που προσεγγίζονται από την κανονική κατανομή, αλλά δεν γίνεται αναφορά σε αυτές.

■Στο όγδοο κεφάλαιο γίνεται η μίξη δύο κανονικών κατανομών. Οι σύνθετες κανονικές κατανομές διαμορφώνονται με την απόδοση μιας κατανομής στην μία ή και στις δύο από τις παραμέτρους μ, σ της συνάρτησης πυκνότητας πιθανότητας της f(x).

■Στο ένατο κεφάλαιο συνοψίζονται τα συμπεράσματα που προέκυψαν από την παρούσα εργασία και δίνονται οι επεκτάσεις του θέματος αυτού για περαιτέρω έρευνα.

■Στο τέλος της παρούσης εργασίας παρατίθεται η χρησιμοποιηθείσα βιβλιογραφία και το απαραίτητο παράρτημα.

ΚΕΦΑΛΑΙΟ 2^0

Γένεση: Ιστορικό υπόβαθρο

Ξέρω ότι , τίποτα δεν είναι τόσο ικανό να κεντρίσει τη φαντασία όσοι η θαυμάσια μορφή κοσμικής διαταγής που εκφράζεται από το " νόμο της συχνότητας του λάθους ". Ο νόμος θα είχε προσωποποιηθεί από τους Έλληνες και θα τον είχαν αποθεώσει, εάν ήξεραν για αυτόν. Βασιλεύει με ηρεμία και με πλήρη μετριοφροσύνη στη μέση της πιο άγριας σύγχυσης.

Sir Francis Galton

Έτσι έγραψε ο Sir Francis Galton (1889) για την κανονική κατανομή, σε μια εποχή όπου η αναζήτηση της επιστήμης χρωματίστηκε από το ρομαντισμό του δέκατου ενάτου αιώνα.

Το δέκατο έβδομο αιώνα ο Γαλιλαίος (μετάφραση 1953,1962) διατύπωσε τα συμπεράσματά του σχετικά με τη μέτρηση των αποστάσεων στα αστέρια από τους αστρονόμους (Maistrov, 1974). Παρατήρησε ότι τα τυχαία λάθη είναι αναπόφευκτα στις παρατηρήσεις, ότι τα μικρά λάθη είναι πιθανότερο να εμφανιστούν από τα μεγάλα, ότι οι μετρήσεις είναι εξίσου επιρρεπείς σε σφάλματα προς τη μια κατεύθυνση (προς τα πάνω) ή προς την άλλη (προς τα κάτω), και ότι η πλειοψηφία των παρατηρήσεων τείνει να συγκεντρωθεί γύρω από την πραγματική τιμή. Ο Γαλιλαίος αποκάλυψε εδώ πολλά από τα χαρακτηριστικά του κανονικού νόμου της κατανομής πιθανότητας, και επίσης επιβεβαίωσε ότι (τα τυχαία) λάθη που γίνονται στην παρατήρηση διακρίνονται από (τα συστηματικά) τελικά λάθη που προκύπτουν από τον υπολογισμό.

Αν και η μελέτη της πιθανότητας άρχισε πολύ νωρίς, το πρώτο μεγάλο άλμα της σύγχρονης στατιστικής έγινε με τη δημοσίευση το 1713 του Ars Conjectandi από τον Jacob Bernoulli, στο οποίο ο Bernoulli απέδειξε τον ασθενή νόμο των μεγάλων αριθμών. Η κανονική κατανομή πρωτοεμφανίστηκε το 1733 σαν μία προσέγγιση στην διωνυμική κατανομή, όταν ο Abraham de Moivre γνωστοποίησε την εργασία του σε μερικούς από τους συγχρόνους του. Μια αναζήτηση από τους Daw και Pearson (1972) επιβεβαίωσε ότι διάφορα αντίγραφα αυτής της δημοσίευσης είχαν πολύ στενή σχέση με τα αντίγραφα βιβλιοθηκών του Miscellanea Analytica de Moivre's που τυπώθηκαν το 1733 ή αργότερα.

Το θεώρημα εμφανίστηκε πάλι στο βιβλίο του De Moivre, The Doctrine of Chances (1738, 1756, 1967 δείτε επίσης David, το 1962, όπου εμφανίζεται ως παράρτημα). Αν και το κύριο αποτέλεσμα καλείται συνήθως "Οριακό θεώρημα του De Moivre-Laplace", η ίδια προσέγγιση στις δυωνυμικές πιθανότητες πάρθηκε και από τον Daniel Bernoulli το 1770-1771, αλλά επειδή δημοσίευσε την εργασία του μέσω της Αυτοκρατορικής Ακαδημίας των Επιστημών στο St.Petersburg, αυτή παρέμεινε εκεί σχεδόν απαρατήρητη μέχρι πρόσφατα (Sheynin, 1970). Ο Bernoulli σύνταξε επίσης τον πιο παλαιό γνωστό πίνακα της καμπύλης $y = \exp\left(-\frac{\mu^2}{100}\right)$ (δείτε πίνακα 2.1)

Η φυσική ανάπτυξη της θεωρίας Πιθανοτήτων στις μαθηματικές στατιστικές πραγματοποιήθηκε από τον Pierre Simon de Laplace, ο οποίος "ήταν πιο κατάλληλος για την αρχική ανάπτυξη των μαθηματικών στατιστικών από οποιοδήποτε άλλο άτομο" (Stigler, 1975). Ο Laplace (1810,1812 1878-1912) ανέπτυξε τη χαρακτηριστική συνάρτηση ως εργαλείο για τη θεωρία μεγάλων δειγμάτων και απέδειξε το πρώτο γενικό κεντρικό οριακό θεώρημα. Τα κεντρικά οριακά θεωρήματα δείχνουν ότι τα αθροίσματα των τυχαίων μεταβλητών όταν τυποποιούνται ώστε να έχουν μέσο μηδέν και διασπορά μονάδα τείνουν να συμπεριφερθούν, όπως οι τυποποιημένες κανονικές μεταβλητές καθώς το μέγεθος του δείγματος γίνεται μεγάλο. Αυτό συμβαίνει, παραδείγματος χάριν, όταν προέρχονται από τυχαία δείγματα που έχουν ληφθεί από "καλά-συμπεριφερόμενες " κατανομές.

Ο Laplace έδειξε ότι μια κατηγορία γραμμικών αμερόληπτων εκτιμητών των συντελεστών γραμμικής παλινδρόμησης κατανέμεται κατά προσέγγιση κανονικά εάν το μέγεθος του δείγματος είναι μεγάλο. Το 1812 (Laplace, 1812), απέδειξε ότι η κατανομή πιθανότητας της πιθανής διάρκειας ζωής σε οποιαδήποτε καθορισμένη ηλικία τείνει στην κανονική κατανομή. (Seal, 1967).

Πίνακας 2.1

Μερικοί αρχικοί πίνακες των κανονικών συναρτήσεων, με βήμα και ακρίβεια

Προέλευση	**Συνάρτηση**	**Κάλυψη του πεδίου τιμών με βήμα**	**Ακρίβεια**
D.Bernoulli (1770-1771)	$\exp\left(-\frac{x^2}{100}\right)$	x = 1(1)5(5)30	4 ακρ.
Kramp (1799)	$\log_{10}\left\{\sqrt{\pi}\left[1-\Phi\sqrt{2}x\right]\right\}$	x= 0(0.01)3	7 δεκ.
Legendre (1826)	$2\sqrt{\pi}\left[1-\Phi\left(\sqrt{2}x\right)\right]$	x = 0(0.01)5 $\exp(-x^2)$ = 0(0.01)0.8	10 δεκ.
de Morgan (1837) **de Morgan (1838)**	$2\Phi\left(\sqrt{2}x\right)-1$	x = 0(0.01)2 x = 0(0.01)3	7 δεκ. 7 δεκ.
Glaisher (1871) **Markov (1888)**	$\sqrt{\pi}\left[1-\Phi\left(\sqrt{2}x\right)\right]$	x=3(0.01)4.5 x = 0(0.001)3(0.01)4.8	11 δεκ. 11 δεκ.
Burgess (1898)	$2\Phi\left(\sqrt{2}x-1\right)$	x = 0(0.0001)1(0.002)3 (0.1)5(0.5)6	15 δεκ.
Sheppard (1898)	$\sqrt{2\pi}\phi(x)$	x= 0(0.05)4(0.1)5	5 δεκ.
Sheppard (1898)	Z_a	$2\alpha-1=\begin{cases}0.1(0.1)0.9\\0(0.01)0.99\end{cases}$	10 δεκ. 5 δεκ.
Sheppard (1903) [a]	$\Phi(x)$, $\phi(x)$	x= 0(0.01)6	7 δεκ.
Sheppard (1907) [a]	Z_a	α = 0.5(0.001)0.999	4 δεκ.

[a] Ενσωματωμένος σε PEARSON και Hartley (1958)

Πηγή: Greenwood και Hartley, 1962.

Το 1818 έδειξε ότι όταν η αρχική κατανομή είναι κανονική, ο εκτιμητής ελαχίστων τετραγώνων (LSE) έχει τη μικρότερη διασπορά από οποιονδήποτε γραμμικό συνδυασμό παρατηρήσεων (Stigler,1973). Κατά τη διάρκεια της παραγωγής αυτού του αποτελέσματος, ο Laplace έδειξε ότι η ασυμπτωτική κοινή κατανομή του εκτιμητή ελαχίστων τετραγώνων και του εκτιμητή στατιστικής σειράς είναι διμετάβλητη κανονική, και παίρνει την ελάχιστη διασπορά. Η ιδιότητα αυτή του εκτιμητή ελαχίστων τετραγώνων κάτω από την κανονικότητα, προσπαθεί να συνδυάσει και τους δύο εκτιμητές για να μειώσει τη διασπορά.

Τα παραπάνω αποτελέσματα αναφέρονται στην εργασία του Laplace όσον αφορά την κανονική κατανομή. Το πεδίο των εργασιών του πάντως είναι πολύ ευρύτερο, για περαιτέρω πληροφορίες, δείτε Stigler (1973, 1975a).

Τα προβλήματα που προκύπτουν από τη συλλογή των παρατηρήσεων στην αστρονομία οδήγησαν τον Legendre το 1805 να διατυπώσει την αρχή των ελαχίστων τετραγώνων, δηλαδή την αρχή της ελαχιστοποίησης του αθροίσματος των τετραγώνων "των λαθών" των παρατηρήσεων. Από τον Legendre διατυπώθηκαν επίσης οι κανονικές εξισώσεις.

Το 1809, ο Carl Friedrich Gauss δημοσίευσε το Theoria Motus Corporum Coelestium, δηλώνοντας ότι είχε χρησιμοποιήσει την αρχή των ελαχίστων τετραγώνων από 1795. Αυτό οδήγησε σε κάποια διαμάχη ως προς την προτεραιότητα βλέπε Gauss, Laplace, Legendre, και διάφορους συναδέλφους του Gauss (Plackett, 1972).Όλοι αυτοί στηρίχθηκαν στον ισχυρισμό, εάν δηλαδή η δημοσίευση πρέπει να είναι το κριτήριο για να διευθετήσει το ζήτημα της προτεραιότητας ή όχι. Στο δέκατο ένατο αιώνα, η έρευνα γινόταν συχνά ανεξάρτητα, χωρίς γνώση των ευρημάτων άλλων ερευνητών, όπως θα δούμε αργότερα. Δεν αποτελεί λοιπόν έκπληξη, ότι, ο Gauss δεν ήξερε τίποτα για την προηγούμενη εργασία του Legendre όταν δημοσίευσε το Theoria Motus.

Σε αυτήν την εργασία, ο Gauss έδειξε ότι η κατανομή των λαθών, υποτίθεται συνεχής, πρέπει να είναι κανονική εάν η παράμετρος θέσης έχει (πάλι στη σύγχρονη ορολογία) μία ομοιόμορφη προγενέστερη κατανομή, έτσι ώστε ο αριθμητικός μέσος όρος να είναι η επικρατούσα τιμή της μεταγενέστερης κατανομής (Seal, 1967). Το γραμμικό μοντέλο των ελαχίστων τετραγώνων του Gauss είναι συνεπώς κατάλληλο, όταν τα "λάθη" προέρχονται από μια κανονική κατανομή. Ένας αμερικανός μαθηματικός, ο Robert Adrain (1808), που δεν ήξερε τίποτα για την εργασία του Gauss αλλά που μπορεί να είχε δει το βιβλίο του Legendre, παρήγαγε τις μονομετάβλητες και διμετάβλητες κανονικές κατανομές ως κατανομές των λαθών, και ως εκ τούτου τη μέθοδο ελαχίστων τετραγώνων (Stigler, 1977), αλλά η εργασία του δεν επηρέασε την ανάπτυξη του θέματος.

Η μελέτη των ελαχίστων τετραγώνων, ή της θεωρίας των λαθών, συνεχίστηκε για αρκετές δεκαετίες. Η κανονική κατανομή δεν είχε βρει ακόμα τη θέση της ούτε στους θεωρητικούς ούτε στους εφαρμοσμένους κλάδους του θέματος, και ο Gauss έδωσε λίγη παραπάνω προσοχή σε αυτήν (Seal,1967). Εντούτοις, επισημαίνει (Maistrov, 1974) ότι βάσει του κανονικού νόμου, λάθη οποιουδήποτε μεγέθους είναι δυνατά. Μόλις έγινε αποδεκτή η καθολικότητα του κανονικού νόμου και, οι επιστήμονες υπέθεσαν ότι όλες οι παρατηρήσεις πρέπει επομένως να διατηρηθούν, με συνέπεια μια καθυστέρηση στην ανάπτυξη των μεθόδων για αναγνώριση και αποβολή ακραίων σημείων. Για μια καλή περίληψη των συνεισφορών του Gauss στην Στατιστική και τη θεωρία των ελαχίστων τετραγώνων, δείτε Sprott (1978) ή Whittaker και Robinson (1926).

Ο αστρονόμος Friedrich Wilhelm Bessel (1818) δημοσίευσε μια σύγκριση σε περισσότερες από 300 μετρήσεις γωνιακών συντεταγμένων των αστέρων, των παρατηρηθέντων υπολοίπων και εκείνων που αναμένονται από τον κανονικό νόμο των λαθών του Gauss και διατύπωσε μια εντυπωσιακή συμφωνία σε 300. Ο Bessel το 1838 οδηγήθηκε στην ανάπτυξη της υπόθεσης των στοιχειωδών λαθών από μια εφαρμογή του κανονικού νόμου ως προσέγγιση της κατανομής του συνολικού λάθους, όταν υποτίθεται ότι αυτό το λάθος είναι το αποτέλεσμα ενός απείρως μεγάλου αριθμού ίσων, αλλά εξίσου πιθανών θετικών ή αρνητικών στοιχειωδών λαθών.

Ο Bessel κατέληξε έτσι στον κανονικό νόμο ως προσέγγιση για το συνολικό λάθος, που υποτίθεται ότι τώρα ήταν το άθροισμα ενός μεγάλου αριθμού αμοιβαία ανεξάρτητων, αλλά όχι όμοια κατανεμημένων στοιχειωδών λαθών με τις καλά-συμπεριφερόμενες ιδιότητες, συμπεριλαμβανομένης της συμμετρικής κατανομής γύρω από το μηδέν.

Η υπόθεση των στοιχειωδών λαθών αναπτύχθηκε σταθερά και επαληθεύθηκε ιδιαίτερα από τους αστρονόμους όπως ο G.B.Airy (1861), ο οποίος ερμήνευσε το κεντρικό οριακό θεώρημα του Laplace ακριβώς από μια τέτοια οπτική γωνία.Στην πραγματικότητα, η ανάπτυξη του θεωρήματος του Laplace δεν εξαρτάται από την υπόθεση "στοιχειωδών λαθών", αν και θα μπορούσε να ήταν έτσι. Για περαιτέρω πληροφορίες, δείτε τον Adams (1974).

Το 1860 ο Σκωτσέζος μαθηματικός και φυσικός James Clerk Maxwell δημοσίευσε το πρώτο από τα δύο μεγάλα έργα του για την κινητική θεωρία των αερίων. Χρησιμοποιώντας τις γεωμετρικές εκτιμήσεις, παρήγαγε την κανονική κατανομή ως κατανομή των ορθογώνιων τμημάτων της ταχύτητας των μορίων που κινούνται ελεύθερα στο κενό (Maxwell,1860).Τα αποτελέσματά του οδηγούν, μέσω της εργασίας Boltzmann, στη σύγχρονη θεωρία των στατιστικών μηχανικών και είναι αξιοσημείωτο ότι ως πρώτη προσπάθεια η κίνηση των αερίων περιγράφηκε από μια στατιστική συνάρτηση παρά από μια αιτιοκρατική. "Οι ταχύτητες κατανέμονται μεταξύ των μορίων" έγραψε, "σύμφωνα με τον ίδιο νόμο όπως τα λάθη κατανέμονται μεταξύ των παρατηρήσεων στην θεωρία της μεθόδου των ελαχίστων τετραγώνων ".

Ένας από τους πρώτους που προσάρμοσε μια κανονική καμπύλη σε δεδομένα έξω από τον τομέα της αστρονομίας ήταν ο Βέλγος επιστήμονας Adolphe Quetelet (1846), ο οποίος προσάρμοσε την κανονική κατανομή, στο ύψος 100.000 νεοσύλλεκτων Γάλλων (Stigler, 1975b). Ο Quetelet εξοικειώθηκε με το κεντρικό οριακό θεώρημα του Laplace, αλλά αυτή η έμμεση προσέγγιση απέφυγε τη χρήση μαθηματικής ανάλυσης ή πινάκων της κανονικής πιθανότητας. Μερικοί από τους πρώτους πίνακες που συνδέονται με την κανονική κατανομή αναφέρονται στον πίνακα 2.1.

Ο πρώτος επιστήμονας στην Αγγλία που χρησιμοποίησε την πρώτη εργασία για την ήπειρο ήταν ο Francis Galton, ο οποίος είχε μια αξιοπρόσεκτη σταδιοδρομία στην εξερεύνηση, στη γεωγραφία, στη μελέτη της μετεωρολογίας και, προ πάντων, στην ανθρωπομετρία, δηλαδή στη μελέτη της ανθρωπολογίας μέσω της ανάλυσης των φυσικών μετρήσεων. Στη φυσική κληρονομιά του (1889) (Natural Inheritance), ο Galton επικέντρωσε την προσοχή του στην εργασία του Quetelet, που αφορά την προσαρμογή κανονικής καμπύλης στις ανθρώπινες μετρήσεις και εφάρμοσε ένα μοντέλο για να περιγράψει την εξάρτηση τέτοιων μετρήσεων με εκείνες των γονέων ενός απογόνου.

Ο Galton σχεδίασε τα δεδομένα του σε δύο διαστάσεις. Ανέπτυξε το γραμμικό μοντέλο παλινδρόμησης, την έννοια της συσχέτισης (1888), και την εξίσωση της διμεταβλητής κανονικής κατανομής (1886), με τη βοήθεια ενός μαθηματικού, του Hamilton Dickson.Οι διμετάβλητες και τριμετάβλητες κανονικές κατανομές ήταν ήδη αναπτυγμένες ανεξάρτητα από τον Bravais

(1846), με την εφαρμογή σε στοιχεία από τον Schols (1875), αλλά αν και ενήμεροι για την έννοια της συσχέτισης, οι μετέπειτα ερευνητές δεν έδωσαν το βαθμό σπουδαιότητας που τους έδωσε ο Galton.

Ένας καλός απολογισμός για το πώς ο Galton έφθασε στις ελλειπτικές καμπύλες και τη διμεταβλητή κανονική κατανομή δίνεται από τον Pearson (1920, 1970), συμπεριλαμβανομένου του διαγράμματος από το οποίο ανακάλυψε παρατηρώντας τη μορφή της διμετάβλητης κανονικής επιφάνειας.

Η δημοσίευση της Natural Inheritance ήταν ο καταλύτης για την αγγλική σχολή της βιομετρίας ώστε να αρχίσει να παρουσιάζει σημαντικά αλματώδη πρόοδο. Ο ζωολόγος W. F.R. Weldon επιδίωξε τη βοήθεια ενός εφαρμοσμένου μαθηματικού, του Karl Pearson, που συνειδητοποίησε ότι οι στατιστικές μέθοδοι θα μπορούσαν να δώσουν στοιχεία για να υποστηριχθεί η θεωρία της φυσικής επιλογής του Δαρβίνου. Ο "νόμος του λάθους", όπως ο Galton υποστήριξε για κάποιο διάστημα, προέτρεψε τον Weldon να υποθέσει αρχικά ότι όλοι οι φυσικοί χαρακτήρες στους ομοιογενείς ζωικούς πληθυσμούς κατανέμονται κανονικά (Pearson, 1965,1970).

Ο Galton αντελήφθηκε, εντούτοις, ότι τα σύνολα δεδομένων μπορούν να ακολουθήσουν και κάποιον άλλο νόμο συχνότητας. Ο γεωμετρικός μέσος όρος ενός συνόλου παρατηρήσεων, έγραψε, μπορεί να αντιπροσωπεύσει καλύτερα την πιθανότερη τιμή μιας κατανομής, και σε αυτή την περίπτωση, οι λογάριθμοι των παρατηρήσεων να υποτεθεί, ότι ακολουθούν τον κανονικό νόμο. Αυτό οδήγησε στη λογαριθμική κανονική κατανομή, αλλά υποκίνησε τον Pearson να αναπτύξει ένα σύστημα καμπυλών συχνότητας, ανάλογο με ένα σύνολο παραμέτρων, που θα κάλυπτε όλες τις κατανομές που εμφανίζονται στη φύση, τουλάχιστον εκείνες που είναι συνεχείς (Pearson, 1967, 1970).

Ο Karl Pearson γενίκευσε και τη συζήτηση του Galton για τη συσχέτιση με την ανάπτυξη της θεωρίας της πολλαπλής συσχέτισης και της πολλαπλής παλινδρόμησης. Μια τέτοια θεωρία, παρατήρησε, θα ήταν απαραίτητη για να απαντήσει στο είδος ερωτήσεων που έθεσε ο Weldon. Εργαζόμενος με δύο, τρεις, και τέσσερις μεταβλητές, ο Francis Edgeworth (1892) έκανε μια πρώτη αναφορά στην πολυμετάβλητη κανονική κατανομή και ο Pearson (1896) έδωσε μια σαφή προέκταση σε αυτό (Seal, 1967).

Είναι αξιοσημείωτο ότι η εργασία του Legendre, του Gauss, και της Σχολής ελαχίστων τετραγώνων πέρασε απαρατήρητη κατά τη διάρκεια των πρώτων ετών της βιομετρίας στην Αγγλία (1885-1908), ενδεχομένως λόγω της ενασχόλησης του Pearson με την πολυμεταβλητή κανονική κατανομή (Seal, 1967). Η προφητική εξίσωση παλινδρόμησης του Pearson (1896) δεν φάνηκε να είναι ίδια στη μορφή και τη λύση με το μοντέλο του Gauss του 1809. Ο έντονο ερευνητικό ενδιαφέρον στη θεωρία των ελαχίστων τετραγώνων συνεχίστηκε στην εργασία του R.. A. Fisher και άλλων στο Rothamsted για αρκετά χρόνια μετά από το 1930 (Seal, 1967).

Στην προσπάθεια να προσαρμοστούν τα δεδομένα στις καμπύλες συχνότητάς του, ο Pearson βρέθηκε αντιμέτωπος με την ανάγκη να εξεταστεί η καλή προσαρμογή. Ο Ernst Abbe (1863, 1906) είχε παράγει την κατανομή του $\sum X_i^2$ όπου $X_1,\ldots,X_n$ είναι ένα τυχαίο δείγμα από μια κανονική κατανομή με μέσο όρο μηδέν (Sheynin, 1966, Kendall, 1971), και ο Helmert (1876) παρήγαγε την κατανομή $\sum\left(X_i-\overline{X}\right)^2$ όπου $\overline{X}$ είναι ο αριθμητικός μέσος όρος τους. Οι ίδιοι παράγουν τη X^2 κατανομή με n και n - 1 βαθμούς ελευθερίας, αντίστοιχα. Η εργασία του Abbe πέρασε απαρατήρητη έως το 1966, πίστωση πριν από τον χρόνο που είχε δοθεί στο Helmert. Ο μετασχηματισμός πινάκων του Helmert, εντούτοις, χρησιμοποιείται ακόμα ως εκπαιδευτικό εργαλείο για τη κατανομή $\sum\left(X_i-\overline{X}\right)^2$ και για επαλήθευση η ανεξαρτησία του $\overline{X}$ και η δειγματική διασπορά $\dfrac{\sum\left(X_i-\overline{X}\right)^2}{(n-1)}$ σε ένα κανονικό δείγμα.

Το πρόβλημα είναι αρκετά διαφορετικό εάν ο γεννήτορας πληθυσμός δεν είναι κανονικός. Ο Lancaster (1966) αποδεικνύει πώς ο Laplace παρείχε τις απαραίτητες τεχνικές στον Bienaymé το 1838 και το 1852 για να θεωρηθεί η X^2 κατανομή ως ασυμπτωτική κατανομή μεγάλων-δειγμάτων χωρίς οποιαδήποτε υπόθεση της κανονικότητας. Ο Bienaymé έλαβε κάτι παρεμφερές με το "$\dfrac{\sum(\text{παρατηρηθείσα - αναμενόμενη})^2}{\text{αναμενόμενη}}$" που είναι το X^2 στατιστικό του Pearson, άθροισμα που δεν λαμβάνει υπ' όψιν του τις κατηγορίες.

Ο Weldon είχε εντοπίσει μια αξιοπρόσεχτη εξαίρεση στην προσαρμογή της κανονικής καμπύλης σε ένα σύνολο δεδομένων για το σχετικό μετωπικό εύρος (Pearson,1965,1970).Χρησιμοποιώντας αυτό θα μπορούσε να συνδυάσει δύο υποκατηγορίες και να προσαρμόσει μια σύνθετη κανονική κατανομή ή " διπλή- καμπύλη," όπως αυτός την ονόμασε, αλλά η προσαρμογή έγινε καθαρά από τη δοκιμή και το λάθος του 1892. Το πρώτο στατιστικό άρθρο του Karl Pearson το 1894 αντιμετώπισε αυτό το πρόβλημα, εισάγοντας τη μέθοδο των ροπών σαν μία τεχνική για την προσαρμογή μιας κατανομής συχνοτήτων. Αλλά το θέμα εάν η προσαρμοσμένη κατανομή ήταν λογική οδήγησε τον Pearson στην εργασία που κατέληξε στο άρθρο του 1900 και καθιέρωσε την X^2 ως πολύ καλή προσαρμογή, σταθερό κριτήριο και ακρογωνιαίο λίθο των σύγχρονων στατιστικών. Ειρωνικά, η πρώτη ιερή αγελάδα που έπεσε με αυτό το νέο εργαλείο ήταν ο νόμος των λαθών. Ο Pearson επέπληξε τον αστρονόμο G.B.Airy (1861), ο οποίος είχε προσπαθήσει να επεξηγήσει την καθολικότητα του κανονικού νόμου με ένα ατυχές σύνολο δεδομένων. Χρησιμοποιώντας τα ίδια δεδομένα και τη νέα X^2 κατανομή,ο Pearson έδειξε ότι ο κανονικός νόμος έδωσε τελικά μια απαράδεκτη προσαρμογή . "Πόσο υγιές είναι το πνεύμα του σκεπτικισμού", έγραψε, "σε όλες τις έρευνες σχετικά με τη συμφωνία της θεωρίας και της φύσης". (Pearson, 1900).

Ο Pearson και η Αγγλική Σχολή της Βιομετρίας προτιμούν να εργάζονται με μεγάλα σύνολα δεδομένων για την στατιστική ανάλυση των ποικίλων προβλημάτων που αντιμετωπίζουν.

Αλλά ο W. S. Gosset ("Student") αναγκάστηκε από τις περιστάσεις που επικρατούσαν στην επιχείρηση ζυθοποιείων του Guiness στο Δουβλίνο να χρησιμοποιήσει μικρά δείγματα για να λύσει τα προβλήματα που προκύπτουν μόνο από τέτοια δείγματα . Ένα έτος εργασίας στο βιομετρικό εργαστήριο κάτω από τον Pearson οδήγησε στο διάσημο άρθρο του "Το πιθανό λάθος του μέσου" (Student, 1908) στον οποίο συμπέρανε η κατανομή t, η αναλογία του δειγματικού μέσου σημαίνει τυπική απόκλιση σε ένα κανονικό δείγμα. Ο Pearson έδειξε μικρό ενδιαφέρον για τα αποτελέσματα του Gosset, ίσως επειδή αισθάνθηκε διστακτικός να αφήσει οποιοδήποτε βιολόγο ή ιατρικό ερευνητή να θεωρήσει ότι υπήρξε μια απλή μέθοδος τα συμπεράσματα από λίγα δεδομένα.

Ο Gossett ήταν απληροφόρητος για την παραγωγή από τον Helmert της κατανομής δειγματοληψίας του $\sum(X_i - \overline{X})^2$, αλλά κατέληξε στην κατανομή του και την παρουσίασε, ώστε να μην συνδέεται με το $\overline{X}$. Μια άλλη εργασία από τον Gosset, δείτε Pearson και Wishart (1958).

Η Αγγλική Σχολή της Βιομετρίας δεν ανέλαβε καμμία σοβαρή μελέτη της θεωρίας πιθανοτήτων, αντίθετα με τη Ρώσικη σχολή του Pafnuti Lvovich Tchebyshev και τους μαθητές του Markov και Lyapunov.

Από τα μέσα του δέκατου ενάτου αιώνα σε αυτή τη σχολή εφάρμοσαν τη μαθηματική ακαμψία στους νόμους των μεγάλων αριθμών, των εξαρτώμενων γεγονότων, και των κεντρικών οριακών ιδιοτήτων.Με την εισαγωγή της έννοιας μιας τυχαίας μεταβλητής, ήταν σε θέση να καθιερώσουν ικανοποιητικές συνθήκες για τα τυποποιημένα αθροίσματα που εξαρτώνται, καθώς επίσης και για τις ανεξάρτητες τυχαίες μεταβλητές ώστε να συγκλίνουν στον κανονικό νόμο. Η πρώτη σαφής διατύπωση του προβλήματος, μαζί με μια απόδειξη η οποία αργότερα απαιτούσε αναθεωρήσεις και την πρόσθετη απόδειξη (Maistrov, 1974) δόθηκε από τον Tchebyshev το 1887 (ανατυπωμένος 1890) χρησιμοποιώντας τη μέθοδο των ροπών. Η σημασία της προσέγγισης Tchebyshev στα κεντρικά οριακά θεωρήματα εντοπίζεται στο σαφώς καθορισμένο μαθηματικό χαρακτήρα που απέδωσε στις τυχαίες μεταβλητές. Αυτό το έκανε εισαγωγή περιορισμών στη καταλληλότητα εφαρμογής των αποτελεσμάτων στη θεωρία πιθανοτήτων, έτσι ώστε σε κάθε περίσταση κάποιος να καθορίζει εάν τα οριακά θεωρήματα εξακολουθούν να ισχύουν ή όχι. Αφέθηκε στον Andrei Andreevich Markov (1898) να διορθώσει το θεώρημα Tchebyshev, και στον Αλέξανδρο Mikhailovich Lyapunov (1901) να παραγάγουν ένα κεντρικό οριακό θεώρημα μεγάλης γενικότητας, που να αποδεικνύεται αυστηρά με τα εργαλεία της κλασσικής ανάλυσης, συμπεριλαμβανομένης αυτής των χαρακτηριστικών συναρτήσεων. Οι τυπικές αυστηρές μαθηματικές συζητήσεις των κεντρικών οριακών θεωρημάτων (για τις ανεξάρτητες τυχαίες μεταβλητές) μπορούμε να πούμε ότι άρχισαν με την εργασία του Lyapunov (1900).

Ένα χρήσιμο θεώρημα που συνδέεται με το όνομά του δηλώνει ότι όταν οι X_1, X_2..., Xn είναι ανεξάρτητες, όμοια κατανεμημένες τυχαίες μεταβλητές με πεπερασμένη μέση τιμή και τυπική απόκλιση η κατανομή του τυποποιημένου αθροίσματος

$$\frac{\sum_{j=1}^{n} X_j - n\mathrm{E}[X]}{\sqrt{nVar(X)}}$$

τείνει στην κανονική κατανομή όταν το n τείνει στο άπειρο.

Ο Lyapunov έθεσε επίσης έναν ανώτερο όριο για το μέγεθος της διαφοράς μεταξύ των αθροιστικών συναρτήσεων κατανομής του τυποποιημένου αθροίσματος και της μοναδιαίας κανονικής. Αυτό το ανώτερο όριο ήταν της μορφής $Cn^{-\frac{1}{2}}\log n$, όπου C είναι μία σταθερά βασισμένη στις διασπορές και στις τρίτες ροπές του X. Στη συνέχεια το όριο αυτό βελτιώθηκε αρκετά από τους, Berry (1941), Esseen (1942), Zahl (1966), και Zolotarev (1967).

Για την περίπτωση που οι μεταβλητές Xj κατανέμονται ομοιόμορφα το ανώτερο όριο που λαμβάνεται από τον Zolotarev (1967) είναι $0{,}82\left(\frac{v_3}{\sigma^3}\right)n^{-\frac{1}{2}}$

όπου $\sigma^2 = Var(X_i)$ και $v_3 = \mathrm{E}\left[|X_i - \mathrm{E}[X_i]|^3\right]$.

Αυτό το αποτέλεσμα ήταν μια βελτίωση στο πιο παλιό αποτέλεσμα του Wallace (1959) [που διορθώνει ένα αποτελέσμα του Berry (1941)].

Ο Zahl (1966) ορίζει το ανώτερο όριο ως $0{,}65\left(\frac{v_3}{\sigma^3}\right)n^{-\frac{1}{2}}$ το οποίο μπορεί να ληφθεί, εφ' όσον $\frac{v_3}{\sigma^3} \geq \frac{3}{\sqrt{2}} = 2{,}22$

Μπορεί να δειχθεί από τη μελέτη των ιδιαίτερων περιπτώσεων ότι το ανώτερο όριο πρέπει να είναι τουλάχιστον $C\left(\frac{v_3}{\sigma^3}\right)n^{-\frac{1}{2}}$ με

$$C = \frac{\sqrt{13}}{6\sqrt{2\pi}} = 0{,}40974$$

Ο Zolotarev (1967) έχει δείξει ότι εάν η διασπορά και η απόλυτη τρίτη κεντρική ροπή του X_j είναι σ_j^2, και v_{3j}, αντίστοιχα (j = 1,2..., n),τότε ένα ανώτερο όριο για το μέγεθος της διαφοράς μεταξύ των αθροιστικών συναρτήσεων κατανομής του τυποποιημένου αθροίσματος και της μοναδιαίας κανονικής είναι: $0.9051\left(\sum_{j=1}^{n} v_{3j}\right)\left(\sum_{j=1}^{n}\sigma_j^2\right)^{-\frac{3}{2}}$

Για τη γενική περίπτωση ανεξάρτητων (αλλά όχι απαραιτήτως όμοια κατανεμημένων) μεταβλητών, ο Lindeberg (1922) έδειξε ότι βάζοντας $Var(X_i)=\sigma_i^2$, $\sigma_{(n)}^2 = \sum_{i=1}^{n}\sigma_i^2$ και εάν επιπλέον

$\lim_{n\to\infty}\sigma_{(n)}^{-2}\sum_{i=1}^{n}\left(P_r\left\{\left|X_i - \mathrm{E}[X_i]\right| \geq t\sigma_{(n)}\right\}\times \mathrm{E}\left|X_i - \mathrm{E}[X_i]\right| \geq t\sigma_{(n)}\right) = 0$ για τα όλα $t > 0$,

τότε η κατανομή του τυποποιημένου αθροίσματος : $\sigma_{(n)}^{-1}\sum_{i=1}^{n}(X_i - \mathrm{E}[X_i])$ τείνει στην κανονική κατανομή μονάδων όταν το n τείνει στο άπειρο.

Η ανάγκη του Lindeberg για κάποια συνθήκη ικανοποιήθηκε από τον Feller (1935).

Δεδομένου ότι αυτό το θεώρημα δεν απαιτεί την ύπαρξη των ροπών οποιασδήποτε τάξης, μπορεί να φανεί ότι η μέθοδος των ροπών του Tchebyshev είχε επιζήσει περισσότερο. Προφανώς επειδή αυτό του κέντρισε το ενδιαφέρον, ο Markov (1913,1951) αποδεικνύει το θεώρημα του Lyapunov με τη μέθοδο των ροπών χρησιμοποιώντας το γνωστό επινόημα με τις περικομμένες τυχαίες μεταβλητές. Για μια παρουσίαση αυτών και άλλων οριακών θεωρημάτων καθώς αναπτύχθηκαν ιστορικά, δείτε Uspensky (1937). Η εργασία του Markov συνεχίστηκε και τέθηκε σε σταθερή αξιωματική βάση από τους Bernstein, Khinchine, και Kolmogorov. Για τους περαιτέρω απολογισμούς της πρώτης ιστορίας της Ρωσικής σχολής, δείτε Maistrov (1974) και Adams (1974).

Στο ίδιο άρθρο, ο Tchebyshev (1890) ανέπτυξε μια σειρά επέκτασης και ένα άνω όριο για τη διαφορά μεταξύ της αθροιστικής συνάρτησης κατανομής $F_n(x)$ και του τυποποιημένου αθροίσματος $\frac{\left(\sum X_i\right)}{\left(\sqrt{n}\sigma\right)}$ μιας τυποποιημένης κανονικής μεταβλητής, όπου σ είναι η τυπική απόκλιση σε ένα πληθυσμό που μας ενδιαφέρει. Το 1905, ο Charlier εισήγαγε μία σειρά για να βελτιώσει την προσέγγιση του κεντρικού οριακού θεωρήματος.

Στη συνάρτηση πυκνότητας του $\frac{\left(\sum X_i\right)}{\left(\sqrt{n}\sigma\right)}$ από την άποψη της τυποποιημένης κανονικής πυκνότητας φ, ο Edgeworth (1905) ανέπτυξε την επέκταση της σειράς του για το $F_n(x)$ με τον ίδιο σκοπό στο μυαλό, αν και είχε παραγάγει τις εκδόσεις για τις συμμετρικές πυκνότητες από 1883.(Δείτε Gnedenko και Kolmogorov (1968), και Stigler (1978)).

Η πρώτη ιστορία της κανονικής κατανομής είναι κατά ένα μεγάλο μέρος η ιστορία της Στατιστικής ως επιστήμη. Αφήνουμε την ιστορία σε αυτό το σημείο με τον R.A.Fisher καθώς οι σύγχρονες στατιστικές αρχίζουν να διακλαδίζονται προς τα έξω και να επιταχύνονται.

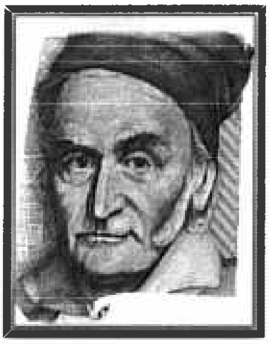
Από το 1915 και μετά ο Fisher μελέτησε τη κατανομή του συντελεστή συσχέτισης, την απόλυτη απόκλισης του $\frac{\sum\left|X_i - \bar{X}\right|}{n}$ στα κανονικά δείγματα, τους συντελεστές παλινδρόμησης, τις αναλογίες συσχέτισης, πολλαπλής παλινδρόμησης και μερικούς συντελεστές συσχέτισης, και την αναλογία F των διασπορών των δειγμάτων σε δύο κανονικούς πληθυσμούς. Συγχρόνως ανέπτυξε τις ιδέες του για εκτίμηση, επάρκεια, πιθανότητα, συμπερασματολογία, ανάλυση διακύμανσης, και πειραματικό σχέδιο (Savage, 1976).

Οι πολυδιάστατες επεκτάσεις των κεντρικών οριακών θεωρημάτων έχουν ερευνηθεί από τους Bergström (1945), Essen (1958), Sadikova (1966), και Sazanov (1967), ενώ συζητώντας τις ιστορικές λεπτομέρειες των εννοιών της διασποράς και του σφάλματος στα στατιστικά του Quetelet, έχουν φέρει προς τα έξω το ζωτικής σημασίας ρόλο που η κανονική κατανομή παίζει στα μαθηματικά της κοινωνίας. Ο Wilf (1988) έχει σχολιάσει εν συντομία την γενική αναζήτηση για την κανονικότητα. Ο Read το(1985) παρείχε μια λεπτή αναθεώρηση των διάφορων σημαντικών αναπτύξεων για την κανονική κατανομή. Ο Stigler (1982) έχει προτείνει νέα πρότυπα για την κανονική κατανομή.

Όπως θα ανέμενε κάποιος, έχουν υπάρξει πολλές πτυχές στην εξέληξη της κανονικής κατανομής. Συνεπώς διάφορα βιβλία και μονογραφίες έχουν κατά καιρούς εμφανιστεί ασχολούμενα συγκεκριμένα με την εξαγωγή συμπερασμάτων, τα χαρακτηριστικά, τα όρια ανοχής, την πρόβλεψη, την

καλή ποιότητα των προσαρμογών. Κανένα από αυτά δεν είναι εφικτό ούτε απαραίτητο να περιέχει όλες αυτές τις λεπτομέρειες των εξελίξεων.
Οι υποθέσεις κανονικότητας έχουν διαδραματίσει έναν βασικό ρόλο στη στατιστική ανάλυση μέσω των ετών, αλλά από τη δεκαετία του '60 αρκετά μεγαλύτερη προσοχή έχει δοθεί στην αμφισβήτηση αυτών των υποθέσεων, που απαιτούν εκτιμητές που είναι αυτοδύναμοι όταν παραβιάζονται τέτοιες υποθέσεις, και στην επινόηση περαιτέρω ενημερωτικών δοκιμών για την αξιοπιστία τους.

2.1 Μια τελική λέξη για την ονοματολογία



Ο Gauss όπως είναι σε νόμισμα 10 Μάρκων.

Η κανονική κατανομή έχει λάβει το όνομα διαφόρων επιστημόνων, συμπεριλαμβανομένου του Laplace και του Gauss.
Η πιο παλαιά δημοσιευμένη παραγωγή της κανονικής κατανομής (ως προσέγγιση σε μια δυωνυμική κατανομή) εμφανίζεται σε ένα τεύχος του de Moivre στις 12 Νοεμβρίου 1733. Αυτό το τεύχος ήταν γραμμένο στα λατινικά αλλά το 1738 ο de Moivre δημοσίευσε μια αγγλική μετάφραση, με μερικές προσθήκες.
Η εφαρμογή όμως της κανονικής κατανομής στην Στατιστική άρχισε μετά από τις εργασίες του αστρονόμου και μαθηματικού Pierres Laplace και του μαθηματικού Karl Gauss, οι οποίοι τη μελέτησαν ανεξάρτητα ο ένας από τον άλλο και χωρίς να έχουν υπόψη τους την εργασία του De Moivre και αφορούσε τη θεωρία σφαλμάτων στις διάφορες παρατηρήσεις, γι' αυτό και ονομάζεται και κατανομή σφαλμάτων ή κατανομή των Gauss- Laplace.

Άλλα ονόματα για την κανονική κατανομή είναι Δεύτερος Νόμος του Laplace, Laplace,Γκαουσσιανή, De Moivre.

O Kac (1975) υπενθυμίζει ότι έχει ονομαστεί επίσης μετά από τους Quetelet και Maxwell. Αν και μερικοί σύγχρονοι συγγραφείς αναφέρονται σε αυτό ως κατανομή Laplace ή Laplace - Gauss, και οι μηχανικοί της δίνουν το όνομα του Gauss, όμως κανένα σύγχρονο όνομα συγγραφέως μετά από το δημιουργό του Abraham De Moivre, μέχρι τότε ήταν γνωστό. O Francis Galton την ονόμασε με διάφορα ονόματα, δείχνοντας έτσι ότι πριν από το 1900 κανένας όρος δεν είχε γίνει κοινά αποδεκτός. Μεταξύ αυτών των ονομάτων ήταν ο νόμος της συχνότητας του λάθους και ο εκθετικός νόμος (1875),ο νόμος της απόκλισης από το μέσο όρο, και ο νόμος των λαθών της παρατήρησης (1869).

Το 1877,ο Galton χρησιμοποίησε αρχικά το όνομα κανονικός νόμος υπό την έννοια ότι αντιμετωπίζεται συνήθως στις στατιστικές. Πιο παλιά γινόταν χρήση της γνωστής ονοματολογίας, και υπό την ίδια έννοια χρησιμοποιήθηκε το 1873, από τον αμερικανό Charles Sanders Peirce (Kruskal, 1978).

Η ονομασία κανονικός νόμος ή κανονική κατανομή κέρδισε την αποδοχή της Αγγλικής σχολής της βιομετρίας. O Karl Pearson (1921-1933) απαίτησε να έχει όφελος από αυτό, αλλά ήταν προφανώς απληροφόρητος για τις προηγούμενες χρήσεις.

Παρατηρούμε λοιπόν, ότι λόγω της σημασίας της κανονικής κατανομής και της κομψής της φύσης έχει δοθεί ιδιαίτερη προσοχή στην ιστορική ανάπτυξή της.

ΚΕΦΑΛΑΙΟ 3^{o}

Κανονική κατανομή:Ορισμός και Ιδιότητες

Ακούμε συχνά να γίνεται συζήτηση για την "κανονική κατανομή" ή την "κωδωνοειδή καμπύλη". Σκεφτόμαστε ποια είναι η μορφή της, αλλά και τι είναι μια κανονική κατανομή; Η κανονική κατανομή, είναι γνωστή ως κατανομή του Gauss (γκαουσσιανή κατανομή).

Η κανονική ή γκαουσσιανή κατανομή είναι μια πανταχού παρούσα και εξαιρετικά σημαντική συνεχής κατανομή πιθανότητας που εξετάζεται στις στατιστικές. Είναι πραγματικά μία οικογένεια κατανομών της ίδιας γενικής μορφής, που διαφέρει μόνο στις παραμέτρους θέσης και διασποράς, αποκαλούμενες συνήθως μέση και τυπική απόκλιση.

Η κανονική καμπύλη καλείται οικογένεια κατανομών. Κάθε μέλος της οικογένειας καθορίζεται με τον καθορισμό των παραμέτρων (μ και σ) του προτύπου. Επειδή η παράμετρος μ μπορεί να πάρει οποιαδήποτε τιμή, θετική ή αρνητική, και η παράμετρος σ μπορεί να πάρει οποιαδήποτε θετική τιμή, η οικογένεια των κανονικών καμπυλών είναι αρκετά μεγάλη, αποτελούμενη από έναν άπειρο αριθμό μελών. Αυτό καθιστά την κανονική καμπύλη γενικής χρήσης έναν πρότυπο, ικανό να περιγράψει έναν μεγάλο αριθμό φυσικά εμφανιζόμενων φαινομένων.

Διακρίνουμε δύο περιπτώσεις :

(α)Την γενική κανονική κατανομή και

(β) Την τυπική (ή τυποποιημένη) κανονική κατανομή.

Η τυποποιημένη κανονική κατανομή είναι η κανονική κατανομή με μέσο όρο μηδέν και τυπική απόκλιση ένα.

3.1 Γενική κανονική κατανομή

3.1.1Συνάρτηση πυκνότητας πιθανότητας [Probability density function (pdf)]

Η μαθηματική μορφή της κανονικής κατανομής δίνεται από την παρακάτω συνάρτηση πυκνότητας πιθανότητας:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\chi-\mu}{\sigma}\right)^2}$$

όπου: **μ** είναι ο αριθμητικός μέσος της τυχαίας μεταβλητής Χ, που προσδιορίζει τη θέση της καμπύλης στον άξονα των x και μπορεί να πάρει τιμές από -∞ έως +∞ δηλαδή: $-\infty \leq \mu \leq +\infty$

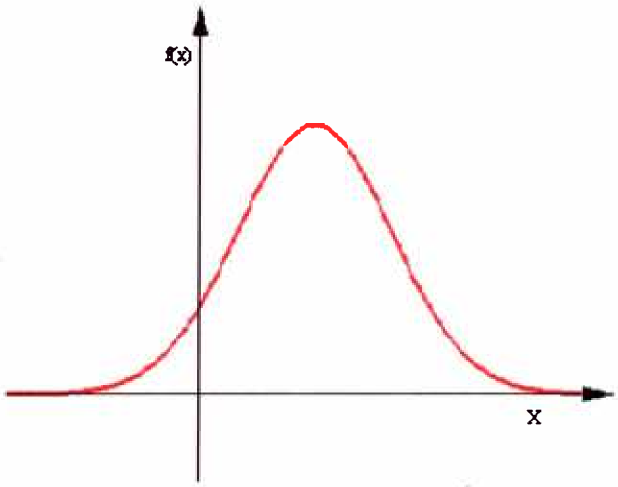
σ είναι η τυπική απόκλιση της τυχαίας μεταβλητής Χ, που παίρνει μόνο θετικές τιμές (**σ>0**) και προσδιορίζει το σχήμα της κανονικής κατανομής

e=2,71828, η βάση των νεπέρειων λογαρίθμων.

π=3,14159, ο λόγος της περιφέρειας ενός κύκλου ως προς την διάμετρό του

3.1.2Γραφική παράσταση της συνάρτηση πυκνότητας πιθανότητας

Η καμπύλη που παριστάνει την συνάρτηση πυκνότητας πιθανότητας της σχέσης $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\chi-\mu}{\sigma}\right)^2}$ έχει την μορφή καμπάνας, είναι συμμετρική και έχει ασύμπτωτο τον άξονα των τετμημένων (άξονα x),όπως φαίνεται στο παρακάτω διάγραμμα 3.1.



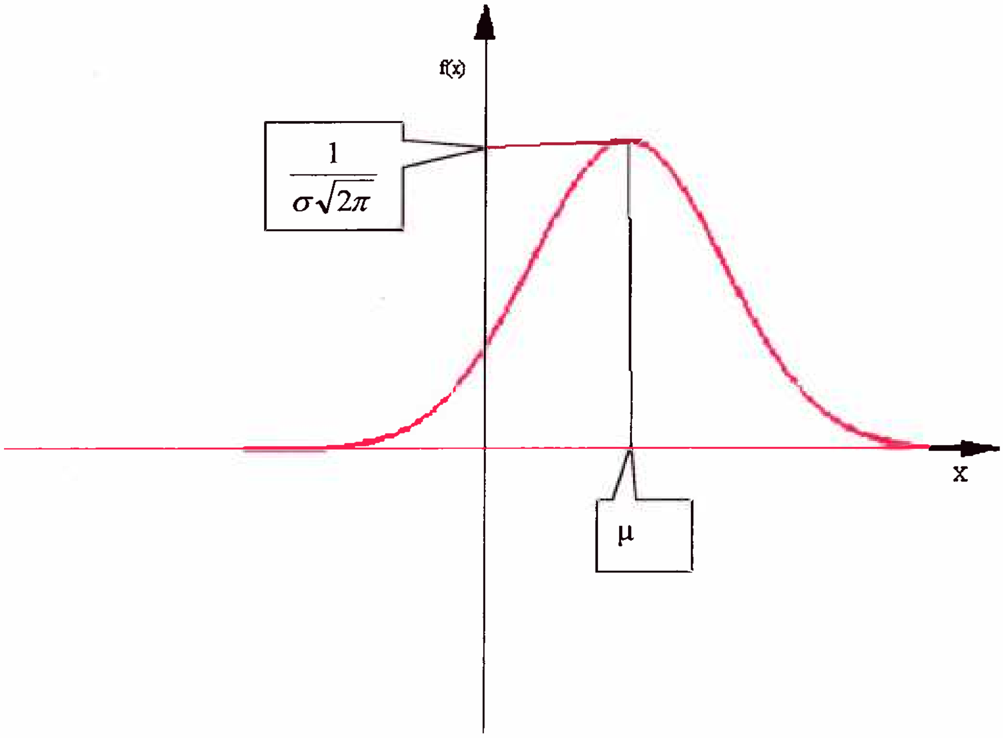
Διάγραμμα 3.1: *Γραφική παράσταση της συνάρτησης πυκνότητας πιθανότητας της γενικής κανονικής κατανομής*

3.1.3 Ιδιότητες της κανονικής κατανομής

• Μπορούμε εύκολα να δούμε ότι η συνάρτηση πυκνότητας πιθανότητας μιας κανονικής κατανομής τυχαίας μεταβλητής X είναι συμμετρική γύρω από το μ, δηλαδή $f(x-\mu)=f(x+\mu)$ για κάθε $x \in R$.

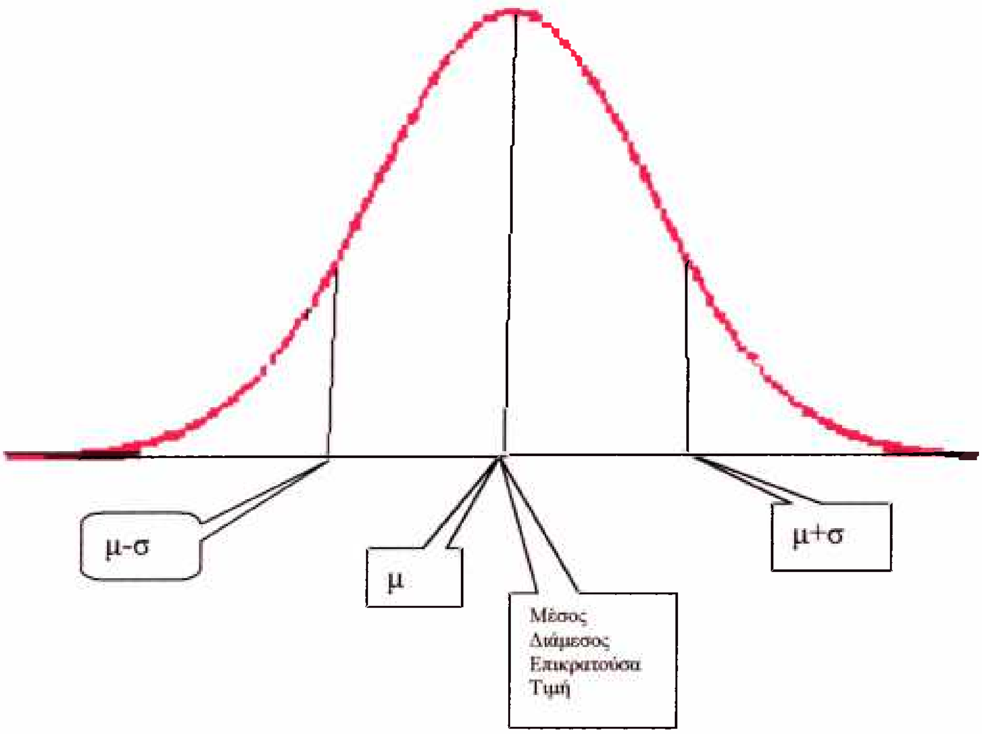
• Παρουσιάζει μέγιστο για x=μ, με τιμή : $\dfrac{1}{\sigma\sqrt{2\pi}}$

•Η $f(x) \to 0$ όταν $x \to \infty$ και όταν $x \to -\infty$ δηλαδή ο άξονας των x είναι ασύμπτωτος της γραφικής παράστασης.



Διάγραμμα 3.2: *Γραφική παράσταση της συνάρτηση πυκνότητας πιθανότητας της γενικής κανονικής κατανομής όπου είναι σημειωμένο το μέγιστο*

- Η κατανομή είναι μονοκόρυφη με επικρατούσα τιμή στο σημείο x=μ, που είναι επίσης διάμεσος και μέσος της κατανομής. Δηλαδή, ο αριθμητικός μέσος, η διάμεσος και η επικρατούσα τιμή συμπίπτουν όπως βλέπουμε και στο επόμενο σχήμα.
- Η f είναι αύξουσα για $x < \mu$ και φθίνουσα $x > \mu$.
- Η συνάρτηση $f(x)$ παρουσιάζει σημεία καμπής στις τιμές x=μ-σ και x=μ+σ
- Η f στρέφει τα κοίλα άνω για $x < \mu - \sigma$ και $x > \mu + \sigma$ και στρέφει τα κοίλα κάτω για $\mu - \sigma < x < \mu + \sigma$
- Η κανονική κατανομή έχει συντελεστή ασυμμετρίας του Pearson a_3=0 και συντελεστή κυρτότητας a_4=3. Αν διαπιστώσουμε ότι μια κατανομή συχνοτήτων έχει a_3=0 και a_4=3, τότε συμπεραίνουμε ότι ακολουθεί την κανονική κατανομή.



Διάγραμμα 3.3: *Γραφική παράσταση της συνάρτηση πυκνότητας πιθανότητας της γενικής κανονικής κατανομής όπου είναι σημειωμένα τα σημεία καμπής, ο μέσος, η διάμεσος και η επικρατούσα τιμή.*

• Η αναμενόμενη τιμή και η διακύμανση της κανονικής κατανομής αποδεικνύεται ότι είναι:

$$\mathrm{E}(X)=\int_{-\infty}^{+\infty} x\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx=\mu$$

$$\sigma^2=Var(X)=\mathrm{E}\left[(X-\mu)^2\right]=\mathrm{E}\left[X^2-2X\mu+\mu^2\right]=$$
$$\mathrm{E}\left(X^2-2\mu\mathrm{E}(X)+\mu^2\right)=\mathrm{E}\left(X^2\right)-2\mu^2+\mu^2=\mathrm{E}\left(X^2\right)-\mu^2$$

Άρα : $\mathrm{E}\left(X^2\right)=\mu^2+\sigma^2$

$$Var(X)=\mathrm{E}\left(X^2\right)-[\mathrm{E}(X)]^2=\mathrm{E}\left(X^2\right)-\mu^2=$$

$$\int_{-\infty}^{+\infty} x^2\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx-\mu^2=\mu^2+\sigma^2-\mu^2=\sigma^2$$

Ορισμός:

Αν έχουμε μία συνεχή τυχαία μεταβλητή X που μπορεί να πάρει τιμές από $-\infty$ έως $+\infty$ και εάν η συνάρτηση πυκνότητας πιθανότητας της X δίνεται από την σχέση :

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\chi-\mu}{\sigma}\right)^2} \qquad (3.1)$$

τότε λέμε ότι η τυχαία μεταβλητή X ακολουθεί την κανονική κατανομή με μέσο μ και διακύμανση σ^2, όπου $\sigma>0$.

●Η κανονική κατανομή με παραμέτρους μ,σ^2 συμβολίζεται ως εξής:

$X\sim N(\mu,\sigma^2)$

Σημείωση:

Για να είναι η κανονική κατανομή μια *καλά ορισμένη* κατανομή πρέπει να ικανοποιεί τις συνθήκες:

1. $f(x)\geq 0$ (σχεδόν παντού)

2. $\int_{-\infty}^{+\infty} f(x)dx=\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\chi-\mu}{\sigma}\right)^2}dx=1$

Απόδειξη:

1. $f(x)\geq 0 \Leftrightarrow \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\chi-\mu}{\sigma}\right)^2}\geq 0$ που ισχύει πάντα.

2. $\int_{-\infty}^{+\infty} f(x)dx=\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\chi-\mu}{\sigma}\right)^2}dx=2.\int_{\mu}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\chi-\mu}{\sigma}\right)^2}dx$

Θέτουμε: $\boxed{\frac{\chi-\mu}{\sigma}=y}$

$\frac{\chi}{\sigma}-\frac{\mu}{\sigma}=y\Leftrightarrow\frac{1}{\sigma}dx=dy\Leftrightarrow dy=\frac{1}{\sigma}dx$

Αλλαγή ορίων:

Για χ=μ, στο $\frac{\chi-\mu}{\sigma}=y$ *έχουμε* $\frac{\mu-\mu}{\sigma}=y\Leftrightarrow y=0$

Όταν $x \to +\infty, y \to +\infty$

οπότε έχουμε:

$$2.\int_{\mu}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 2.\int_{\mu}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

Θέτουμε: $\boxed{\frac{y^2}{2} = u}$

Οπότε $y^2 = 2u \Leftrightarrow \boxed{y = \sqrt{2u}}$

$$\frac{y^2}{2} = u \text{ οπότε } \frac{2ydy}{2} = du \Leftrightarrow ydy = du \Leftrightarrow \sqrt{2u}.dy = du \Leftrightarrow \boxed{dy = \frac{1}{\sqrt{2u}}du}$$

Αλλαγή ορίων:

Για y=0, u=0

Για $y \to +\infty, u \to +\infty$

$$\text{Τότε:}2.\int_{\mu}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = 2.\int_{0}^{+\infty} \frac{1}{\sqrt{2\pi}}.\frac{1}{\sqrt{2u}} e^{-u} du = 2.\frac{1}{\sqrt{2\pi}} \int_{0}^{+\infty} .\frac{1}{\sqrt{2u}} e^{-u} du =$$

$$= 2.\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2}} \int_{0}^{+\infty} \frac{1}{\sqrt{u}}.e^{-u} du = \frac{1}{\sqrt{\pi}} \int_{0}^{+\infty} e^{-u} u^{-\frac{1}{2}} du$$

Η συνάρτηση γάμμα ορίζεται ως εξής:

$$\Gamma(t) = \int_{0}^{+\infty} e^{-x} x^{t-1} dx \text{ για } t>0$$

$$\text{για } t = \frac{1}{2} \text{ έχουμε } \Gamma\left(\frac{1}{2}\right) = \int_{0}^{+\infty} e^{-x} x^{-\frac{1}{2}} dx \text{ για } t>0$$

Αποδεικνύεται ότι $: \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

$$\text{Οπότε το : } \frac{1}{\sqrt{\pi}} \int_{0}^{+\infty} e^{-u} u^{-\frac{1}{2}} du = \frac{1}{\sqrt{\pi}}.\Gamma\left(\frac{1}{2}\right) = \frac{1}{\sqrt{\pi}}.\sqrt{\pi} = 1$$

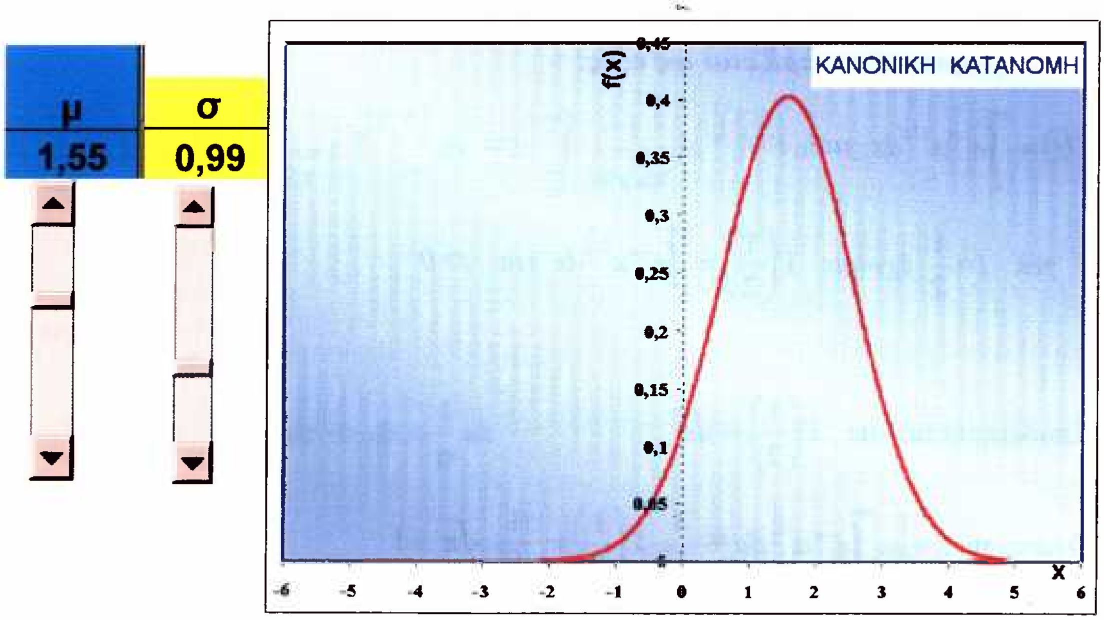
Άρα $\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\chi-\mu}{\sigma}\right)^2} dx = 1$

Επομένως η κανονική κατανομή είναι μία καλά ορισμένη κατανομή. Ακόμη συμπεραίνουμε ότι το εμβαδόν κάτω από την καμπύλη ισούται με την μονάδα.

- **Οικογένεια κατανομών κανονικής κατανομής**

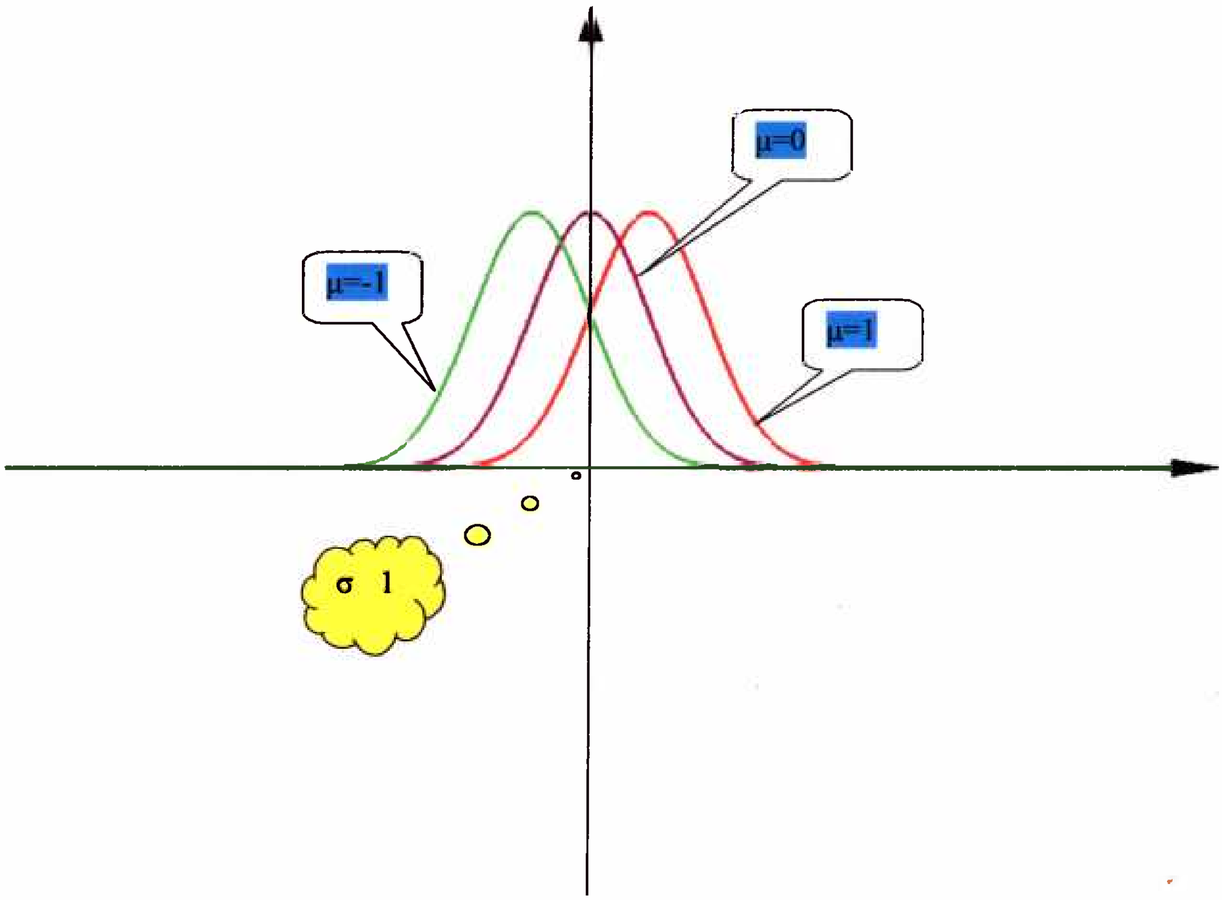
Σε κάθε μεταβολή των μ και σ η σχέση $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\chi-\mu}{\sigma}\right)^2}$ μας δίνει μια οικογένεια από καμπύλες κανονικής κατανομής. Είναι στην ουσία μία οικογένεια κατανομών της ίδιας γενικής μορφής, που διαφέρει μόνο στις παραμέτρους μέση τιμή και τυπική απόκλιση.

Με βάση τα παραπάνω έγινε ένα *πρόγραμμα στο Ecxel δυναμικής γραφικής παράστασης της κανονικής κατανομής*, όπου μεταβάλλοντας τη μέση τιμή και την τυπική απόκλιση παρατηρούμε τις διαφορετικές καμπύλες της κανονικής κατανομής και τη δημιουργία των οικογενειών κατανομών. Η δυναμική αυτή γραφική παράσταση, παρακολουθείται, όταν είμαστε συνδεμένοι με ηλεκτρονικό υπολογιστή.



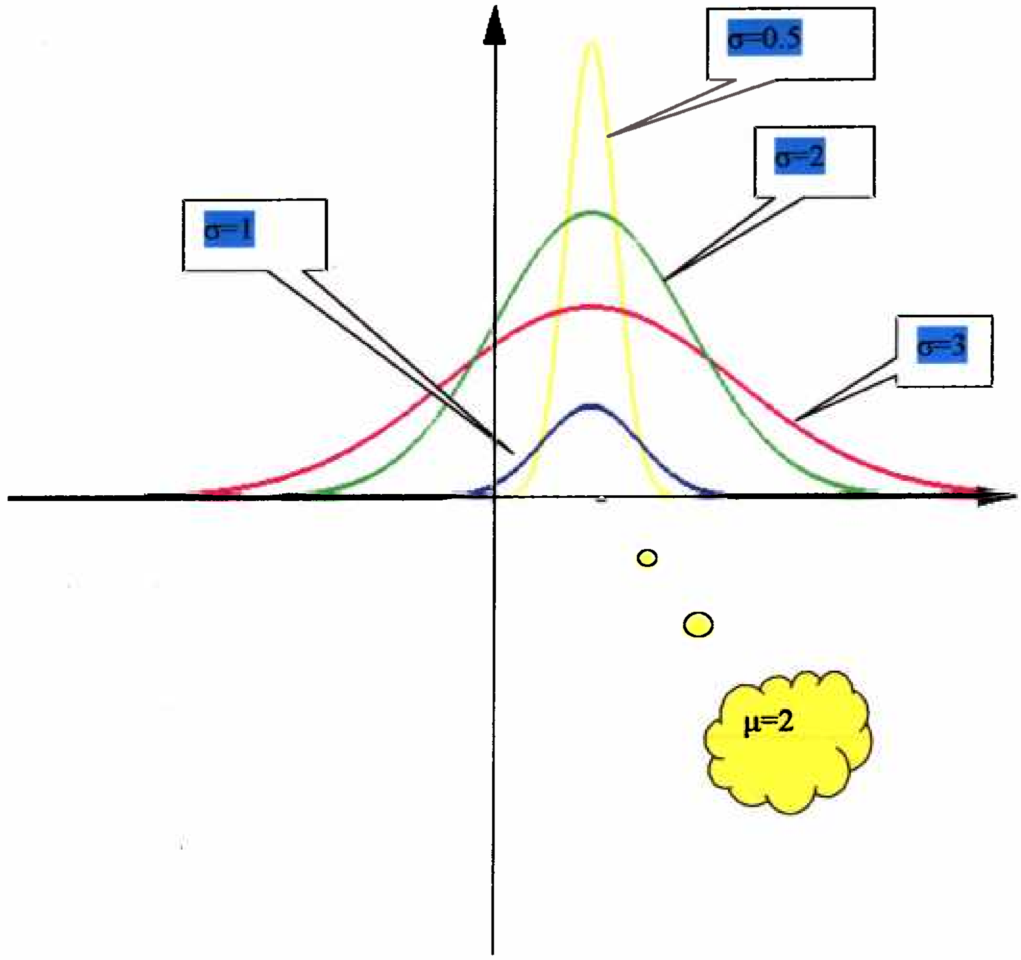
Διάγραμμα 3.4: *Δυναμική γραφική παράστασης της κανονικής κατανομής*

Για καλλίτερη κατανόηση παραθέτουμε διαγράμματα με τις διαφορετικές θέσεις της καμπύλης, όταν μεταβάλουμε τις παραμέτρους.
Συγκεκριμένα σε μια μεταβολή της μέσης τιμής μ(που είναι η παράμετρος θέσης) με σταθερή την τυπική απόκλιση σ παρατηρούμε ότι έχουμε μία μετακίνηση της καμπύλης κατά μήκος του άξονα των x (Διάγραμμα 3.5)



Διάγραμμα 3.5: *Οικογένεια καμπύλων κανονικής κατανομής, όταν μεταβάλλεται η μέση τιμή μ και η τυπική απόκλιση σ παραμένει σταθερή.*

Επίσης αν κρατήσουμε σταθερή τη μέση τιμή και μεταβάλουμε την τυπική απόκλιση, παρατηρούμε ότι καθώς αυξάνεται η τυπική απόκλιση η καμπύλη γίνεται πλατύκυρτη, ενώ όταν μικραίνει η τυπική απόκλιση η καμπύλη γίνεται λεπτόκυρτη,(Διάγραμμα 3.6)



Διάγραμμα 3.6: *Οικογένεια καμπύλων κανονικής κατανομής, όταν μεταβάλλεται η τυπική απόκλιση σ και η μέση τιμή μ παραμένει σταθερή.*

3.1.4 Αθροιστική συνάρτηση κατανομής

Η αθροιστική συνάρτηση κατανομής μιας μεταβλητής Χ που ακολουθεί την κανονική κατανομή που παριστάνει την πιθανότητα η τυχαία μεταβλητή Χ να πάρει τιμή μικρότερη από την x και δίνεται από την σχέση :

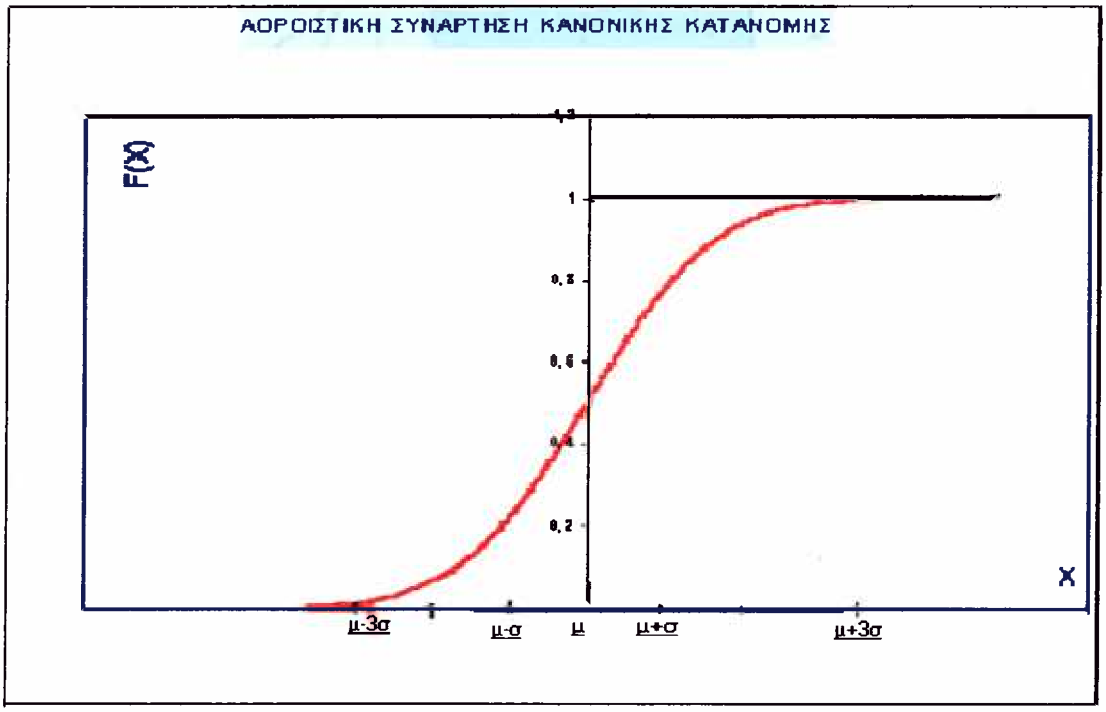
$$F(x) = P\{X \leq x\} = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Αν μεταβληθεί το ανώτερο όριο x του ολοκληρώματος, η γραφική παράσταση της $F(x)$είναι μια καμπύλη αύξουσα που περιέχεται μεταξύ 0 και 1.

● Η πιθανότητα να πάρει μια τυχαία μεταβλητή Χ, που ακολουθεί την κανονική κατανομή, τιμή μικρότερη ή ίση με x δίνεται από το ολοκλήρωμα:

$$F(x) = P\{X \le x\} = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\chi-\mu}{\sigma}\right)^2} dx$$

που εκφράζει την αθροιστική συνάρτηση κατανομής. Στο παρακάτω σχήμα βλέπουμε την γραφική παράσταση της αθροιστικής συνάρτησης της γενικής κανονικής κατανομής.

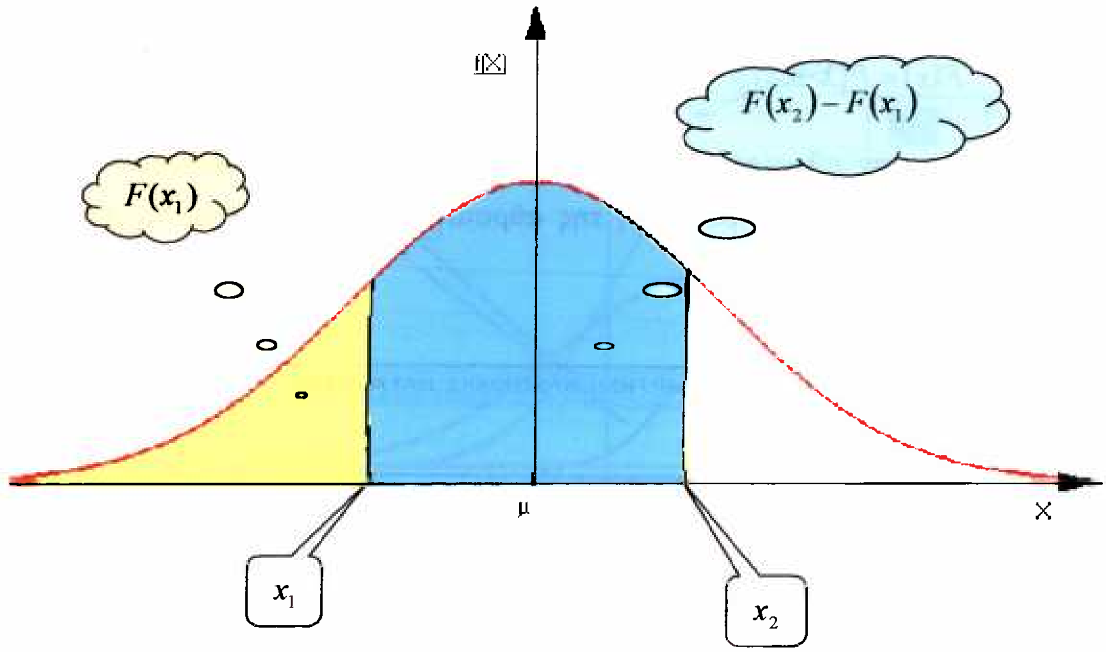


Διάγραμμα 3.7: *Γραφική παράσταση αθροιστικής συνάρτησης γενικής κανονικής κατανομής*

● Η πιθανότητα η τυχαία μεταβλητή Χ να πάρει τιμή μεταξύ των x_1 και x_2 δίνεται από το ολοκλήρωμα :

$$P\{x_1 \le X \le x_2\} = \int_{x_1}^{x_2} f(x)dx = F(x_2) - F(x_1)$$

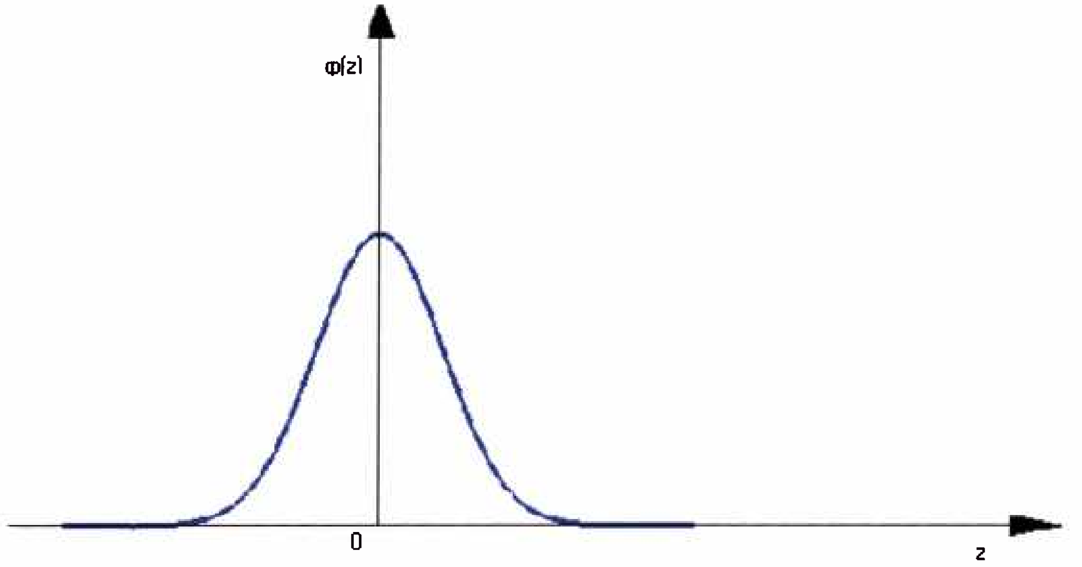
Στο παρακάτω σχήμα βλέπουμε την γραφική αυτού του ολοκληρώματος.



Διάγραμμα 3.8: *Γραφική παράσταση της πιθανότητας η τυχαία μεταβλητή X να πάρει τιμή μεταξύ των τιμών x_1 και x_2*

● Έστω X μία τυχαία μεταβλητή που ακολουθεί κανονική κατανομή με N(μ,σ^2), και Y=αX+β, όπου α και β είναι πραγματικοί αριθμοί. Τότε η τυχαία μεταβλητή Y ακολουθεί την κανονική κατανομή N(αμ+β,$\alpha^2\sigma^2$).

3.2 Τυποποιημένη κανονική κατανομή



Διάγραμμα 3.9: *Γραφική παράσταση της συνάρτησης πυκνότητας πιθανότητας της τυπικής κανονικής κατανομής*

Είδαμε προηγουμένως ότι οι κανονικές κατανομές είναι άπειρες, όσα είναι και τα δυνατά ζεύγη των παραμέτρων μ, σ. Συνεπώς, ο υπολογισμός της πιθανότητας μια τυχαία μεταβλητή Χ που ακολουθεί την κανονική κατανομή να πάρει τιμή $x_1 \leq X \leq x_2$ δίνεται από το ολοκλήρωμα

$$P\{x_1 \leq X \leq x_2\} = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Για κάθε ζεύγος των παραμέτρων μ και σ θα χρειαστεί να υπολογίσουμε ένα νέο ολοκλήρωμα της παραπάνω μορφής. Ο υπολογισμός όμως του παραπάνω ολοκληρώματος για κάθε μια κανονική κατανομή παρουσιάζει πολύ μεγάλες δυσκολίες στις εφαρμογές.

Για να αντιμετωπίσουμε τις δυσκολίες αυτές, κάνουμε τον παρακάτω γραμμικό μετασχηματισμό, αντικαθιστώντας τη μεταβλητή X με μια νέα μεταβλητή $Z = \frac{X-\mu}{\sigma}$,η οποία λέγεται κανονικοποιημένη τιμή της X.

Άρα η συνάρτηση πυκνότητας πιθανότητας της τυποποιημένης κανονικής κατανομής δίνεται από τη σχέση:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad \text{όπου } -\infty < z < \infty \tag{3.2}$$

Αποδεικνύεται ότι, αν μια τυχαία μεταβλητή X ακολουθεί την κανονική κατανομή που έχει μέσο μ και διακύμανση σ^2, τότε και η τυχαία μεταβλητή $Z = \frac{X-\mu}{\sigma}$ ακολουθεί την κανονική κατανομή που έχει μέσο μηδέν και διακύμανση 1.Δηλαδή:

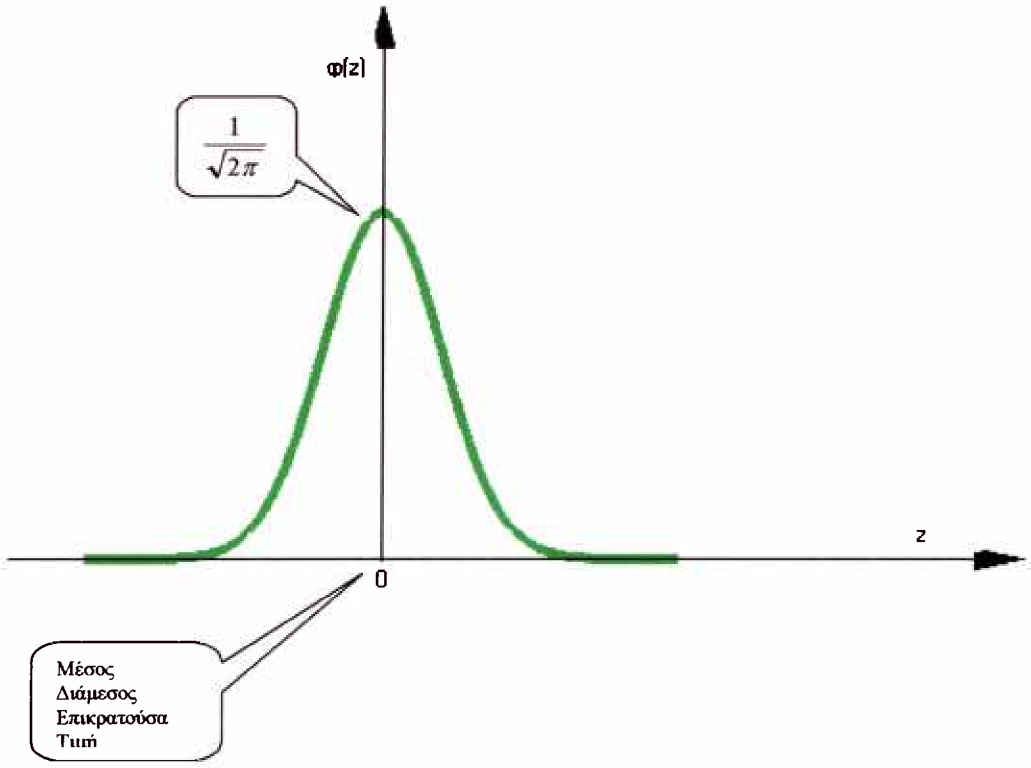
$$\mathrm{E}(Z) = \mathrm{E}\left(\frac{X-\mu}{\sigma}\right) = \frac{\mathrm{E}(X-\mu)}{\sigma} = \frac{1}{\sigma}[\mathrm{E}(X) - \mathrm{E}(\mu)] = \frac{1}{\sigma}(\mu - \mu) = 0$$

$$Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}[Var(X) - Var(\mu)] = \frac{1}{\sigma^2}(\sigma^2 - 0) = \frac{\sigma^2}{\sigma^2} = 1$$

Άρα $\mathrm{E}(Z) = \mu = 0$ και $Var(Z) = \sigma^2 = 1$

Η κανονική κατανομή που έχει μέσο 0 και διακύμανση 1 λέγεται τυποποιημένη κανονική κατανομή και συμβολίζεται με N(0,1).Αν λύσουμε ως προς X τον τύπο της κανονικοποίησης βρίσκουμε ότι X=μ+σZ. Επομένως η κανονικοποιημένη τιμή μιας τυχαίας μεταβλητής καθορίζει πόσες τυπικές αποκλίσεις απέχει η τιμή αυτή της X από την μέση τιμή της.

Η γραφική της παράσταση της τυποποιημένης κανονικής κατανομής φαίνεται στο παρακάτω σχήμα, από το οποίο προκύπτουν αντίστοιχες ιδιότητες με αυτές της γενικής κανονικής κατανομής.



Διάγραμμα 3.10: *Γραφική παράσταση τυπικής κανονικής κατανομής στην οποία έχει σημειωθεί το μέγιστο, ο μέσος, η διάμεσος και η επικρατούσα τιμή*

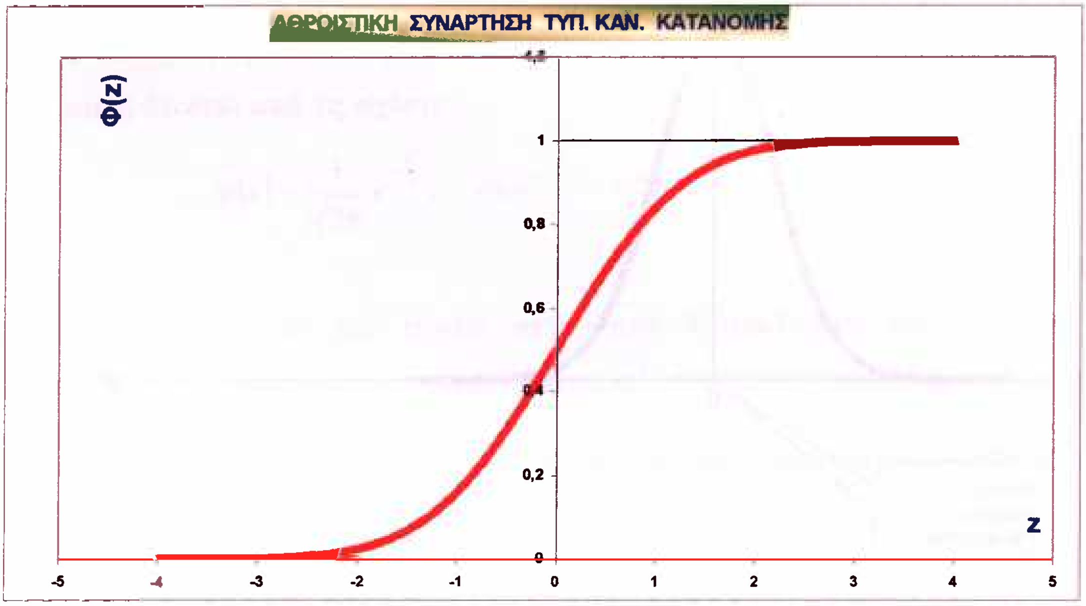
- Η συνάρτηση $\varphi(x)$ παρουσιάζει μέγιστο για $z = 0$ με τιμή : $\frac{1}{\sqrt{2\pi}}$
- Η συνάρτηση $\varphi(x)$ παρουσιάζει σημεία καμπής στις τιμές z=-1 και z=1
- Η φ είναι συμμετρική γύρω από το $z = 0$.
- Η φ είναι αύξουσα για $z < 0$ και φθίνουσα για z>0

- Ο μέσος, η διάμεσος και η επικρατούσα τιμή συμπίπτουν στο z=0

- Η φ στρέφει τα κοίλα άνω για $z < -1$ και z >1 και στρέφει τα κοίλα κάτω για -1 <z <1
- Η φ(z) → 0 όταν z → ∞ και όταν z → −∞ δηλαδή ο άξονας των x είναι ασύμπτωτος της γραφικής παράστασης.
- Η αθροιστική συνάρτηση κατανομής της τυποποιημένης κανονικής είναι:

$$\Phi(z)=P_r(-\infty \le U \le z)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \quad \text{όπου } -\infty<z<\infty$$

Η γραφική παράσταση της δίνεται από το παρακάτω σχήμα.



Διάγραμμα 3.11: *Γραφική παράσταση αθροιστικής συνάρτησης τυπικής κανονικής κατανομής*

• Επειδή η κατανομή της μεταβλητής Z δεν περιέχει παραμέτρους, δηλαδή είναι ανεξάρτητη από τις παραμέτρους μ και σ, μπορούμε να κατασκευάσουμε πίνακες οι οποίοι δίνουν τις διάφορες τιμές της αθροιστικής συνάρτησης κατανομής $\Phi(z)$ για κάθε τιμή της z. Οι πίνακες της $\Phi(z)$χρησιμοποιούνται για να μας δίνουν την τιμή του ολοκληρώματος:

$$\Phi(z)=P_r(-\infty \le Z \le z)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

Υπάρχουν έτοιμοι πίνακες που μας δίνουν για z από –4 έως 4 τις τιμές της αθροιστικής συνάρτησης κατανομής $\Phi(z)$ της N(0,1).Οι πίνακες μας δίνουν συνήθως τις τιμές της $\Phi(z)$ για θετικές μόνο τιμές της μεταβλητή z, γιατί για αρνητικές τιμές της z χρησιμοποιούμε τις σχέσεις: $\varphi(z)=\varphi(-z)$ και $\Phi(-z)=1-\Phi(z)$, οι οποίες ισχύουν λόγω της συμμετρικότητας της $\varphi(z)$ γύρω από το z=0.Ακόμη λόγω της συμμετρικότητας ισχύουν οι σχέσεις : $z_\beta=-z_{1-\beta}$ $\quad \Phi(z_\beta)-\Phi(z_{1-\beta})=1-2\beta$, $\quad 0<\beta<0.5$

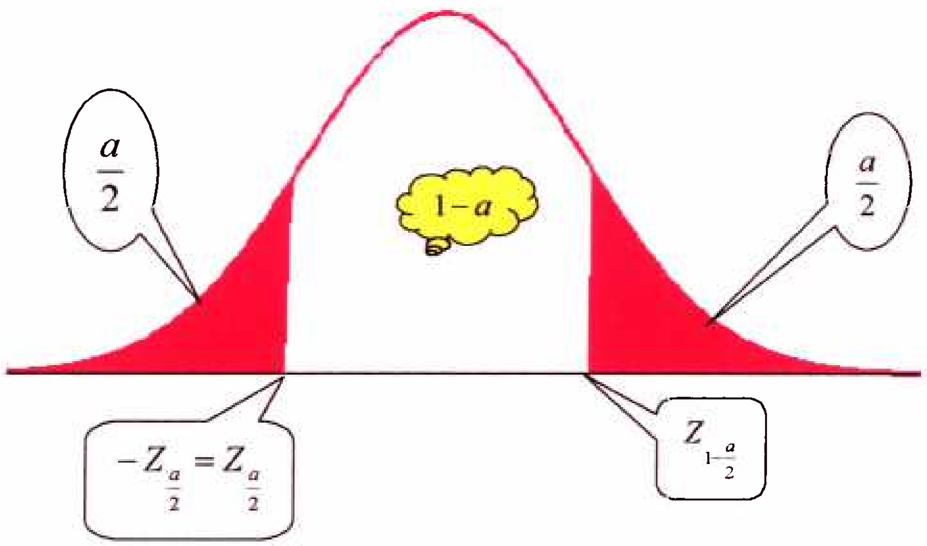
• Για τις τυχαίες μεταβλητές X και $Z = \dfrac{X-\mu}{\sigma}$ ισχύουν οι σχέσεις :

$$F(x) = \Phi\left[\left(\frac{x-\mu}{\sigma}\right)\right] \qquad \text{και} \qquad \Phi(z) = F(\mu + \sigma z)$$

3.2.1Υπολογισμός πιθανοτήτων τυποποιημένης κανονικής κατανομής

Γενική περίπτωση :

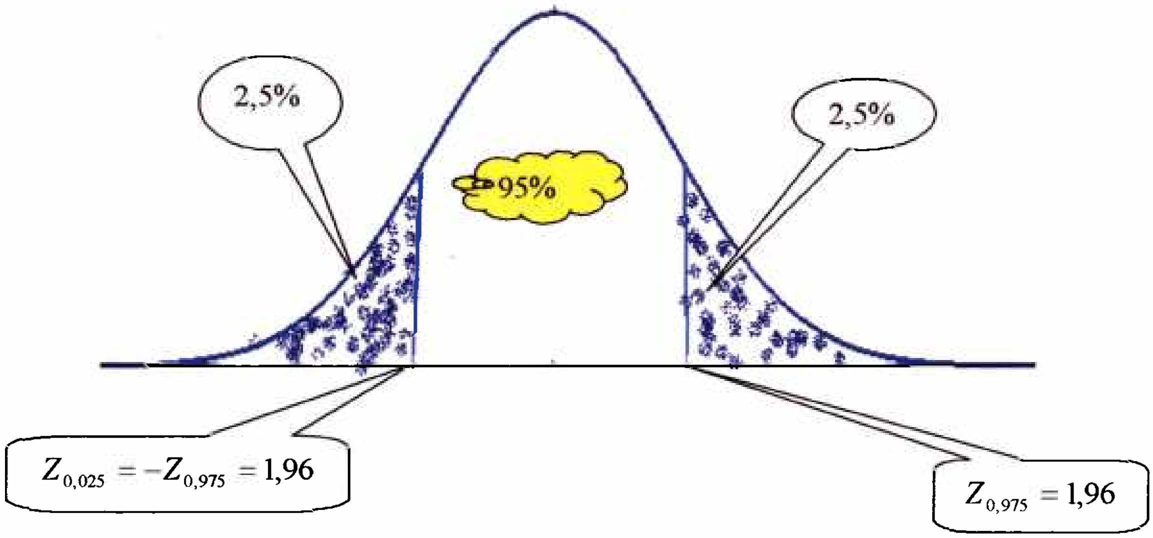
(Όπου α είναι το επίπεδο σημαντικότητας)



$$P\left(Z < Z_{\frac{a}{2}}\right) = \frac{a}{2} \qquad\qquad P\left(Z < Z_{1-\frac{a}{2}}\right) = 1 - \frac{a}{2}$$

Διάγραμμα 3.12: *Υπολογισμός πιθανοτήτων τυποποιημένης κανονικής κατανομής για επίπεδο σημαντικότητας α%*

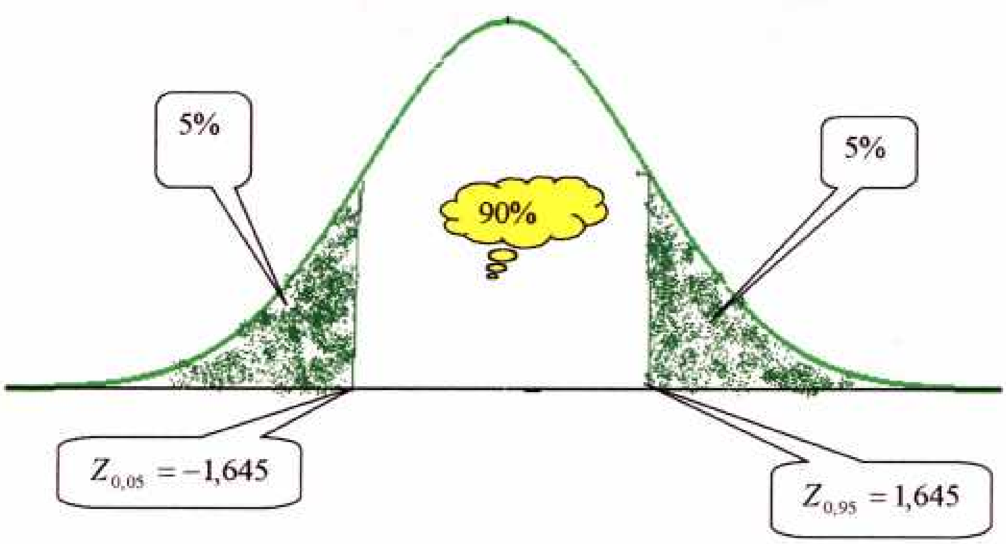
α)



$P(Z < -1{,}96) = 0{,}025$ $\qquad$ $P(Z < 1{,}96) = 0{,}975$

Διάγραμμα 3.13: *Υπολογισμός πιθανοτήτων τυποποιημένης κανονικής κατανομής για επίπεδο σημαντικότητας 5%*

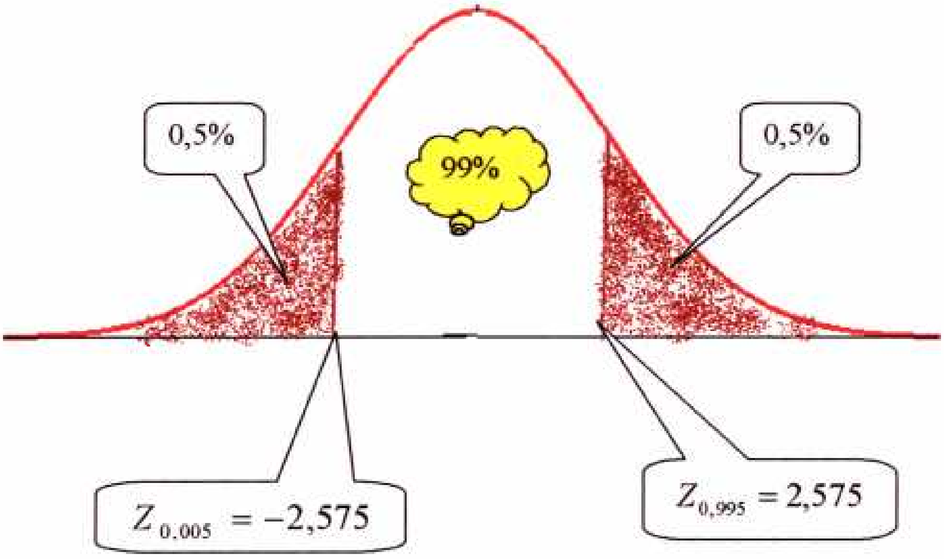
β)



$P(Z < -1{,}645) = 0{,}05$ $\qquad$ $P(Z < 1{,}645) = 0{,}95$

Διάγραμμα 3.14: *Υπολογισμός πιθανοτήτων τυποποιημένης κανονικής κατανομής για επίπεδο σημαντικότητας 10%*

γ)



$P(Z < -2{,}575) = 0{,}005$ $\qquad\qquad$ $P(Z < 2{,}575) = 0{,}995$

Διάγραμμα 3.15: *Υπολογισμός πιθανοτήτων τυποποιημένης κανονικής κατανομής για επίπεδο σημαντικότητας 1%*

Πρόταση: Αν X~ N(μ,σ^2) τότε $F(x) = P_r(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

Χρησιμοποιώντας την ιδιότητα αυτή είμαστε σε θέση να καθορίσουμε μια σταθερά έτσι ώστε αν X~ N(μ,σ^2) η $P_r(X \leq c) = a$

Αυτό ισχύει διότι $P_r(X \leq c) = a \Leftrightarrow \Phi\left(\frac{c-\mu}{\sigma}\right) = a$

π.χ αν μ=10, σ^2=4 και α=0,95

$P_r(X \leq c) = 0{,}95 \Leftrightarrow \Phi\left(\frac{c-10}{2}\right) = 0{,}95 \Leftrightarrow c = 13{,}29$

Παρατήρηση: Σχετικά με την ανισότητα Chebyshev ισχύει ότι για οποιαδήποτε τυχαία μεταβλητή τουλάχιστον τα $\frac{3}{4}$ των τιμών της βρίσκονται στο διάστημα $[\mu - 2\sigma, \mu + 2\sigma]$ ενώ στο διάστημα $[\mu - 3\sigma, \mu + 3\sigma]$ βρίσκονται τουλάχιστον τα $\frac{8}{9}$ των παρατηρήσεων.

Ειδικότερα όμως για την κανονική κατανομή όμως έχουμε

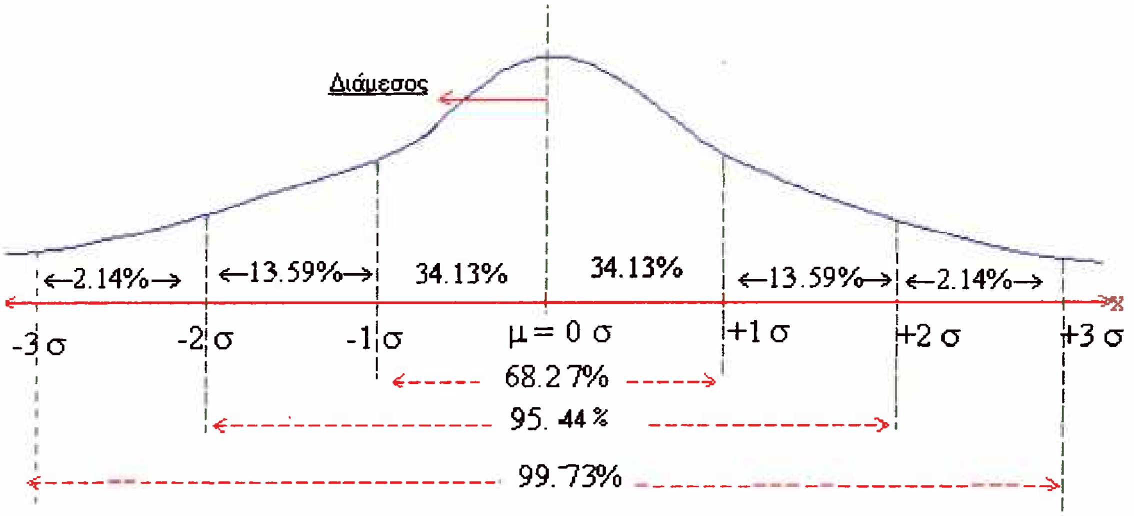
$$P_r(\mu-\sigma<X<\mu+\sigma)=\Phi\left(\frac{\mu+\sigma-\mu}{\sigma}\right)-\Phi\left(\frac{\mu-\sigma-\mu}{\sigma}\right)=\Phi(1)-\Phi(-1)=0.6827$$

Επίσης

$$P_r(\mu-2\sigma<X<\mu+2\sigma)=\Phi\left(\frac{\mu+2\sigma-\mu}{\sigma}\right)-\Phi\left(\frac{\mu-2\sigma-\mu}{\sigma}\right)=\Phi(2)-\Phi(-2)=0.9544$$

ενώ

$$P_r(\mu-3\sigma<X<\mu+3\sigma)=\Phi\left(\frac{\mu+3\sigma-\mu}{\sigma}\right)-\Phi\left(\frac{\mu-3\sigma-\mu}{\sigma}\right)=\Phi(3)-\Phi(-3)=0.9973$$



Διάγραμμα 3.16: *Ποσοστό του εμβαδού της κανονικής κατανομής που περιλαμβάνεται ανάμεσα σε μία, σε δύο και σε τρεις τυπικές αποκλίσεις γύρω από το μέσο.*

Συμπέρασμα: Παρατηρούμε από το παραπάνω διάγραμμα, ότι σε απόσταση μιας τυπικής απόκλισης γύρω από το μέσο, περιλαμβάνεται το 68,27%του συνολικού εμβαδού της κανονικής κατανομής. Σε απόσταση δύο τυπικών αποκλίσεων από το μέσο περιλαμβάνεται το 95,44% του συνολικού εμβαδού της κανονικής κατανομής. Τέλος σε απόσταση τριών τυπικών αποκλίσεων από το μέσο περιλαμβάνεται το 99,73%του συνολικού εμβαδού.

3.3 Άλλες συναρτήσεις

3.3.1 Συνάρτηση σφάλματος

- Η συνάρτηση σφάλματος erf(x) ορίζεται ως εξής:

$$erf(x) = 2\pi^{-\frac{1}{2}} \int_0^x e^{-t^2} dt = 2\Phi\left(x\sqrt{2}\right) - 1 \quad \text{για } x \geq 0$$

- Η συμπληρωματική συνάρτηση σφάλματος είναι : $erfc(x) = 1 - erf(x)$.

- Η ελλιπής συνάρτηση λόγου των $P(a,x)$ και $\varphi(x)$ συνδέεται με την σχέση

$$P\left(\frac{1}{2}, x\right) = 2\Phi\left(\sqrt{2x}\right) - 1 \;, x \geq 0 \;, \text{όπου } P(a,x) = \frac{\int_0^x t^{a-1} e^{-t} dt}{\Gamma(a)}$$

3.3.2 Επαναληπτικά παράγωγα του $\varphi(x)$.Πολυώνυμο του Hermite

$$\left(-\frac{d}{dx}\right)^r \varphi(x) = \mathrm{H}_r(x)\varphi(x)$$

όπου $\mathrm{H}_r(x)$ είναι το πολυώνυμο του Hermite r βαθμού.

$$\mathrm{H}_r(x) = x^r - \frac{r^{(2)}}{2.1!} x^{r-2} + \frac{r^{(4)}}{2^2.2!} x^{r-4} - \frac{r^{(6)}}{2^3.3!} x^{r-6} + \ldots$$

όπου : $r^{(a)} = r(r-1)..(r-a+1)$

Τα πρώτα εννέα πολυώνυμα δίνονται από τις παρακάτω σχέσεις.

$\mathrm{H}_0(x) = 1$

$\mathrm{H}_1(x) = x$

$\mathrm{H}_2(x) = x^2 - 1$

$\mathrm{H}_3(x) = x^3 - 3x$

$\mathrm{H}_4(x) = x^4 - 6x^2 + 3$

$\mathrm{H}_5(x) = x^5 - 10x^3 + 15x$

$\mathrm{H}_6(x) = x^6 - 15x^4 + 45x^2 - 15$

$\mathrm{H}_7(x) = x^7 - 21x^5 + 105x^3 - 105x$

$\mathrm{H}_8(x) = x^8 - 28x^6 + 210x^4 - 420x^2 + 105$

Εκφράσεις του $\mathrm{H}_r(x)$ μπορούμε να παρατηρήσουμε χρησιμοποιώντας τον αναγωγικό τύπο: $\mathrm{H}_r(x) = x\mathrm{H}_{r-1}(x) - (r-1)\mathrm{H}_{r-2}(x)$

3.4 Ροπές και άλλες ιδιότητες

3.4. 1 Ροπές

►Ορισμοί:

- **Ροπή r τάξης περί** το μηδέν

Έστω X μία συνεχής τυχαία μεταβλητή και r ένας θετικός αριθμός. Η ποσότητα $\mu_r' = \mathrm{E}(X^r) = \int_{-\infty}^{\infty} x^r f(x)dx$, λέγεται ροπή r τάξης περί το μηδέν της **τυχαίας** μεταβλητής X ή **απλή** ροπή r τάξης της X.

Η απλή ροπή r τάξης υπάρχει αν $\int_{-\infty}^{\infty} |x^r| f(x)dx < \infty$.

Επίσης $\boxed{\mu_1' = \mathrm{E}(X) = \mu}$ και $\mu_2' = \mathrm{E}(X^2)$.

- Ροπή r **τάξης περί την μέση τιμή ή κεντρική ροπή r τάξης**

Έστω X μία τυχαία μεταβλητή και r ένας θετικός αριθμός. Η ποσότητα $\mu_r = \mathrm{E}[(X-\mu)^r] = \int_{-\infty}^{\infty} (x-\mu)^r f(x)dx$, αν X συνεχής, λέγεται ροπή r τάξης περί την μέση τιμή ή κεντρική ροπή r τάξης της τυχαίας μεταβλητής X.

Η κεντρική ροπή r τάξης υπάρχει αν $\int_{-\infty}^{\infty} |(x-\mu)^r| f(x)dx < \infty$.

- Από τον τύπο της κεντρικής ροπής r τάξης απορρέουν τα εξής :

Για r=1, $\mu_1 = \mathrm{E}[(X-\mu)] = \int_{-\infty}^{\infty} (x-\mu) f(x)dx = 0$. Άρα $\mu_1 = 0$

Για r=2, $\mu_2 = \mathrm{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \mathrm{Var}(x)$. Άρα $\boxed{\mu_2 = Var(x) = \sigma^2}$

Για r=3, $\mu_3 = \mathrm{E}[(X-\mu)^3] = \frac{\mu_3}{\sigma^3}$, όπου μ_3 ο συντελεστής ασυμμετρίας

Για r=4, $\mu_4 = \mathrm{E}[(X-\mu)^4] = \frac{\mu_4}{\sigma^4}$, όπου μ_4 ο συντελεστής κύρτωσης

- Η σχέση που συνδέει τις ροπές γύρω από το μηδέν με τις ροπές γύρω από την μέση τιμή είναι η εξής: $\mu_2 = \mu_2' - \mu_1'^2$, η οποία προκύπτει από την σχέση $Var(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$

► **Υπολογισμός ροπών κανονικής κατανομής**

Προς απλούστευση των υπολογισμών χρησιμοποιώ την τυποποιημένη κανονική κατανομή N(0,1).

Η συνάρτηση πυκνότητας πιθανότητας της τυποποιημένης κανονικής κατανομής, όπως γνωρίζουμε δίνεται από την σχέση : $\boxed{\varphi(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}}$

Έστω ότι η Z ακολουθεί την κανονική κατανομή. Δεδομένου ότι η κατανομή είναι συμμετρική γύρω από το Z=0, έχουμε ότι E[Z]=0.

Επομένως η ροπή περί την μέση τιμή μ_r θα είναι ίση με την ροπή μ_r' περί το μηδέν.

Δηλαδή $\mu_r=\mu_r'=\mathrm{E}(Z^r)=\int_{-\infty}^{+\infty}x^r.\frac{1}{\sqrt{2\pi}}.e^{-\frac{x^2}{2}}dx=(2\pi)^{-1}\int_{-\infty}^{+\infty}x^r.e^{-\frac{x^2}{2}}.dx$

Άρα η ροπή περί την μέση τιμή μ_r είναι ίση με την ροπή μ_r' περί το μηδέν

• **Εάν r είναι περιττός :**

$$\mu_r=\mu_r'=\mathrm{E}(Z^r)=\left(\sqrt{2\pi}\right)^{-1}\int_{-\infty}^{+\infty}x^r.e^{-\frac{x^2}{2}}dx=0$$

Άρα όταν r είναι περιττός $\boxed{\mu_r=0}$.

• **Εάν r είναι άρτιος :**

$$\mu_r=\mathrm{E}(Z^r)=\int_{-\infty}^{+\infty}x^r.\frac{1}{\sqrt{2\pi}}.e^{-\frac{x^2}{2}}dx=2.\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}x^r..e^{-\frac{x^2}{2}}dx=\sqrt{\frac{4}{2\pi}}\int_{0}^{+\infty}x^re^{-\frac{x^2}{2}}dx$$

οπότε $\mu_r==\sqrt{\frac{2}{\pi}}\int_0^{+\infty}x^re^{-\frac{x^2}{2}}dx$

Θέτουμε : $\boxed{\mathrm{I}=\int_0^{+\infty}x^re^{-\frac{x^2}{2}}dx}$

Οπότε: $\mu_r=\sqrt{\frac{2}{\pi}}.\mathrm{I}$

Υπολογισμός του $\mathrm{I}=\int_0^{+\infty}x^re^{-\frac{x^2}{2}}dx$

Θέτουμε: $\boxed{\frac{x^2}{2}=t}$ οπότε $\frac{x^2}{2}=t \Leftrightarrow x^2=2t \Leftrightarrow x=\sqrt{2t} \Leftrightarrow x=(2t)^{\frac{1}{2}}$. Άρα $\boxed{x=(2t)^{\frac{1}{2}}}$

$$\frac{x^2}{2}=t \Leftrightarrow \frac{2x}{2}dx=dt \Leftrightarrow xdx=dt \Leftrightarrow (2t)^{\frac{1}{2}}dx=dt \Leftrightarrow dx=\frac{dt}{(2t)^{\frac{1}{2}}} \Leftrightarrow dx=(2t)^{-\frac{1}{2}}.dt.$$

Άρα $\boxed{dx=(2t)^{-\frac{1}{2}}.dt}$

Συνεπώς το

$$\mathrm{I}=\int_0^{+\infty} x^r e^{-\frac{x^2}{2}}dx=\int_0^{+\infty}(2t)^{\frac{r}{2}}.e^{-t}.(2t)^{-\frac{1}{2}}.dt=\int_0^{+\infty}(2t)^{\frac{r-1}{2}}.e^{-t}.dt=2^{\frac{r-1}{2}}\int_0^{+\infty}(2t)^{\frac{r-1}{2}}.e^{-t}.dt$$

Η συνάρτηση γάμμα ορίζεται ως εξής:

$$\Gamma(t)=\int_0^{+\infty} e^{-x}x^{t-1}dx \quad \text{για} \quad t>0$$

$$\Gamma(r)=\int_0^{+\infty} e^{-t}t^{r-1}dt \quad \text{για} \quad \tau>0$$

επίσης υπολογίζουμε το : $\Gamma\left(\frac{r+1}{2}\right)=\int_0^{+\infty} e^{-t}t^{\frac{r+1}{2}-\frac{2}{2}}dt$ για $\tau>0$

Άρα : $\boxed{\Gamma\left(\frac{r+1}{2}\right)=\int_0^{+\infty} e^{-t}t^{\frac{r-1}{2}}dt}$

Ακόμη αν n ακέραιος ισχύει: $\boxed{\Gamma\left(n+\frac{1}{2}\right)=\frac{1.3.5....(2n-1)}{2^n}.\sqrt{\pi}}$

Οπότε: $\mathrm{I}=2^{\frac{r-1}{2}}\int_0^{+\infty}(2t)^{\frac{r-1}{2}}.e^{-t}.dt \Leftrightarrow \mathrm{I}=2^{\frac{r-1}{2}}.\Gamma\left(\frac{r+1}{2}\right)$

Επομένως : $\mu_r=\sqrt{\frac{2}{\pi}}.\mathrm{I}$

$$\mu_r=\sqrt{\frac{2}{\pi}}.2^{\frac{r-1}{2}}.\Gamma\left(\frac{r+1}{2}\right)=\sqrt{2}.2^{\frac{r-1}{2}}.\frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi}}=2^{\frac{1}{2}}.2^{\frac{r}{2}-\frac{1}{2}}.\frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi}}=2^{\frac{r}{2}}.\frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}$$

$$\mu_r = 2^{\frac{r}{2}} . \frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}$$

Όμως επειδή r άρτιος τότε $\frac{r}{2}$ ακέραιος

Οπότε έχουμε: $\Gamma\left(\frac{r+1}{2}\right) = \Gamma\left(\frac{r}{2}+1\right) = \frac{1.3.5....(r-3).(r-1)}{2^{\frac{r}{2}}} . \sqrt{\pi}$

Οπότε : $\mu_r = \dfrac{2^{\frac{r}{2}} . \frac{1.3.5.....(r-3)(r-1)}{2^{\frac{r}{2}}} . \sqrt{\pi}}{\sqrt{\pi}}$

Άρα όταν r είναι άρτιος $\boxed{\mu_r = 1.3.5.....(r-3).(r-1)}$

3.4.2 Ροπογεννήτρια Συνάρτηση

Ορισμός:Έστω X τυχαία μεταβλητή. Εάν υπάρχει h>0 τέτοιο ώστε $\mathrm{E}\left(e^{tx}\right) < \infty$ για κάθε t στην περιοχή $|t| < h$, τότε η ποσότητα $\mathrm{M}(t) = \mathrm{E}\left(e^{tx}\right) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$,αν X συνεχής ονομάζεται ροπογεννήτρια της τυχαίας μεταβλητής X.

► **Υπολογισμός της ροπογεννήτριας συνάρτησης κανονικής κατανομής**

Έστω $\mathrm{M}(t) = \mathrm{E}\left(e^{tx}\right) = \int_{-\infty}^{\infty} e^{tx} . f(x) dx$,οπότε έχουμε:

$$\mathrm{M}(t) = \mathrm{E}\left(e^{tx}\right) = \mathrm{E}\left(e^{tx+t\mu-t\mu}\right) = \mathrm{E}\left(e^{tx+t\mu-t\mu}\right) = \mathrm{E}\left(e^{t\mu} . e^{tx-t\mu}\right) = e^{t\mu}\mathrm{E}\left(e^{tx-t\mu}\right) = e^{t\mu}\mathrm{E}\left(e^{t(x-\mu)}\right)$$

$$\mathrm{M}(t) = e^{t\mu} \int_{-\infty}^{\infty} e^{t(x-\mu)} . \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx =$$

$$= e^{t\mu} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{t(x-\mu)} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = e^{t\mu} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{t(x-\mu)-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx =$$

$$= e^{t\mu}\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{2\sigma^2 t(x-\mu)-(x-\mu)^2}{2\sigma^2}}dx = e^{t\mu}\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}\left[(x-\mu)^2-2\sigma^2 t(x-\mu)\right]}dx$$

$$= e^{t\mu}\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}\left[(x-\mu)^2-2\sigma^2 t(x-\mu)+\sigma^4t^2-\sigma^4t^2\right]}dx = e^{t\mu}\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}\left[(x-\mu-\sigma^2t)^2-\sigma^4t^2\right]}dx=$$

$$= e^{t\mu}\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu-\sigma^2t)^2}{2\sigma^2}}e^{\frac{\sigma^4t^2}{2\sigma^2}}dx = e^{t\mu}e^{\frac{\sigma^4t^2}{2\sigma^2}}\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu-\sigma^2t)^2}{2\sigma^2}}dx=$$

$$= e^{t\mu}e^{\frac{\sigma^2t^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{[x-(\mu+\sigma^2t)]^2}{2\sigma^2}}dx$$

Το ολοκλήρωμα $\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{[x-(\mu+\sigma^2t)]^2}{2\sigma^2}}dx = 1$ διότι πρόκειται για κανονική κατανομή με μέση τιμή $\mu+\sigma^2 t$ και διακύμανση σ^2.

Οπότε έχουμε: $M(t) = e^{t\mu}e^{\frac{\sigma^2t^2}{2}}.1 = e^{t\mu}e^{\frac{\sigma^2t^2}{2}}$

Άρα η ροπογεννήτρια συνάρτηση είναι : $\boxed{M(t) = e^{t\mu+\frac{\sigma^2t^2}{2}}}$

3.4.3 Χαρακτηριστική συνάρτηση

Η Χαρακτηριστική συνάρτηση της N(μ,σ²) είναι η $\varphi(t) = E\left(e^{itx}\right) = \int_{-\infty}^{\infty} e^{itx}.f(x)dx$

όπου $i=\sqrt{-1}, t \in R$. Δηλαδή η χαρακτηριστική συνάρτηση **είναι ένας μετασχηματισμός** Fourier στο σημείο –t.

Όμως $M(t) = E\left(e^{tx}\right) = \int_{-\infty}^{\infty} e^{tx}.f(x)dx$

Οπότε : $\varphi(t) = M(it)$

Επειδή $M(t) = e^{t\mu+\frac{\sigma^2t^2}{2}}$

Έχουμε $\varphi(t) = M(it) = e^{it\mu+\frac{\sigma^2(it)^2}{2}} = e^{it\mu+\frac{\sigma^2i^2t^2}{2}} = e^{it\mu+\frac{\sigma^2(-1)t^2}{2}} = e^{it\mu-\frac{\sigma^2t^2}{2}}$

Άρα η χαρακτηριστική συνάρτηση είναι: $\boxed{\varphi(t) = e^{it\mu-\frac{\sigma^2t^2}{2}}}$

•**Η αθροιστική ροπογεννήτρια συνάρτηση** δίδεται από την σχέση: $\log \mathrm{E}\left(e^{itX}\right) = it\mu - \frac{t^2\sigma^2}{2}$

3.4.4 Λοιπές ιδιότητες

•*Η* ***τροποποιητική αξία*** της συνάρτησης πυκνότητας πιθανότητας είναι $\left(\sqrt{2\pi}\right)^{-1} = 0{,}3979$.

•Για όλα τα r > 2, ***η αθροιστική*** κ_r είναι μηδέν. Αυτή η ιδιότητα χαρακτηρίζει τις κανονικές κατανομές.

• Η ***μέση απόκλιση της Χ*** είναι $\sigma\sqrt{\frac{2}{\pi}} = 0{,}798\sigma$ Για όλες τις κανονικές κατανομές

Ισχύει : $\frac{\text{Mean deviation}}{\text{Standard deviation}} = \sqrt{\frac{2}{\pi}} = 0.798$

• Η γενέτειρα πληροφοριών συνάρτηση της Χ είναι:

$$\left(\sqrt{2\pi\sigma}\right)^{-u}\left(\frac{\sqrt{2\pi\sigma}}{\sqrt{u}}\right) = \left(\sqrt{2\pi\sigma}\right)^{-(u-1)} u^{-\frac{1}{2}}$$

•Η ***εντροπία*** είναι: $\log\left(\sqrt{2\pi\sigma}\right) + \frac{1}{2}$

•Είναι επίσης ενδιαφέρον να σημειωθεί ότι η συνάρτηση πυκνότητας πιθανότητας $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\chi-\mu}{\sigma}\right)^2}$ μπορεί να εκφραστεί στην ***αριθμητική μορφή*** $0.3979(0.6065)^{\left[\frac{(x-\mu)}{\sigma}\right]^2}$.

• Εάν $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες, κανονικά κατανεμημένες τυχαίες μεταβλητές, τότε οποιαδήποτε γραμμική συνάρτηση αυτών των μεταβλητών κατανέμεται επίσης κανονικά.

• Είναι ενδιαφέρον να σημειωθεί ότι εάν X_1 και X_2 είναι ανεξάρτητα, και κάθε ένα κατανέμεται κανονικά με μηδενική αναμενόμενη τιμή, τότε το $X_1X_2\left(X_1^2+X_2^2\right)^{-\frac{1}{2}}$ είναι επίσης κανονικά κατανεμημένο.

Εάν επιπλέον var(X_1)=var(X_2),τότε $\dfrac{\left(X_1^2-X_2^2\right)}{\left(X_1^2+X_2^2\right)}$ είναι επίσης κανονικά κατανεμημένο.

• Έστω $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες τυχαίες μεταβλητές που κάθε μια κατανέμεται κανονικά, με την εφαρμογή του μετασχηματισμού.:

$$X_1 = \overline{X} + (1.2)^{-\frac{1}{2}}.Z_2\sigma + (2.3)^{-\frac{1}{2}}Z_3\sigma + \ldots + [(n-1)n]^{-\frac{1}{2}}Z_n\sigma$$

$$X_2 = \overline{X} - (1.2)^{-\frac{1}{2}}.Z_2\sigma + (2.3)^{-\frac{1}{2}}Z_3\sigma + \ldots + [(n-1)n]^{-\frac{1}{2}}Z_n\sigma$$

$$X_3 = \overline{X}\ldots\ldots\ldots\ldots\ldots\ldots - 2(2.3)^{-\frac{1}{2}}Z_3\sigma + \ldots + [(n-1)n]^{-\frac{1}{2}}Z_n\sigma$$

$$X_n = \overline{X}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots - (n-1)[(n-1)n]^{-\frac{1}{2}}Z_n\sigma$$

μπορεί να δειχθεί ότι:

1. $\overline{X} = \dfrac{\sum_{j=1}^{n} X_j}{n}$ έχει μια κανονική κατανομή με αναμενόμενη τιμή μ και τυπική απόκλιση $\dfrac{\sigma}{\sqrt{n}}$.

2. Κάθε Z_j (j = 2.. , n) είναι μια κανονική μεταβλητή μονάδων

3. $\overline{X}, Z_2, \ldots, Z_n$ είναι ένα αμοιβαία ανεξάρτητο σύνολο μεταβλητών, και από αυτό προκύπτει το παρακάτω.

4. $\sum_{j}^{n}(X_j-\overline{X})^2=\sigma^2\sum_{j=2}^{n}Z_j^2$ κατανέμεται ως σ^2 (χ^2 με (ν - 1) βαθμούς ελευθερίας).Αυτό το τελευταίο αποτέλεσμα αποδείχθηκε από τον Helmert το 1875-76.

Ο παραπάνω μετασχηματισμός καλείται μετασχηματισμό Helmert.

5. $\overline{X}$ **και οποιαδήποτε συνάρτηση g(X_i -** $\overline{X}$, ..., Xn - $\overline{X}$) είναι αμοιβαία **ανεξάρτητα. Σε αυτό το αποτέλεσμα είναι χρήσιμες** οι ροπές υπολογισμού **και οι κατανομές των στατιστικών** $\overline{X}\left[Range(X_1,...,X_n)\right]^{-1}$ και

$$\overline{X}\left[n^{-1}\sum_{j=1}^{n}\left|X_j-\overline{X}\right|\right]^{-1}$$

Από αυτό μπορεί επίσης **να δειχθεί ότι:**

6. $\sum_{j}^{n}(X_j-\overline{X})^2$ **και οποιαδήποτε** συνάρτηση των **αναλογιών**

$$\left\{(X_i-\overline{X})\left[\sum_{j=1}^{n}(X_j-\overline{X})^2\right]^{-\frac{1}{2}}\right\}$$

είναι αμοιβαία ανεξάρτητα.

ΚΕΦΑΛΑΙΟ $4^{Ο}$

Πίνακες, Επεκτάσεις και Αλγόριθμοι

Η τυποποιημένη κανονική αθροιστική συνάρτηση κατανομής Φ(x) και οι σχετικές συναρτήσεις, όπως τα ποσοστιαία σημεία και η αναλογία του Mills, μπορούν να επεκταθούν σε σειρές και σε άλλες μορφές όπως οι συνεχείς συναρτήσεις. Τέτοιες εκφράσεις δίνουν συχνά προσεγγίσεις στο Φ(x) και σε άλλες συναρτήσεις, με την περικοπή μιας επέκτασης σε ένα σημείο που παρέχει έναν κατάλληλο βαθμό ακρίβειας

Η κύρια χρήση τέτοιων επεκτάσεων είναι στην παροχή των κατάλληλων αλγορίθμων για τις αριθμητικές τιμές των συναρτήσεων του ενδιαφέροντος μας. Επομένως στο κεφάλαιο έχουμε περιλάβει τους διαθέσιμους πίνακες ως λίστα επεκτάσεων και προσεγγίσεων. Μερικά από τα παραπάνω χρησιμοποιήθηκαν με τη μία ή την άλλη μορφή για να υπολογίσουν τις τιμές που εμφανίζονται στους δημοσιευμένους πίνακες της κανονικής κατανομής.

Όπου οι προσεγγίσεις διερευνούνται για την ακρίβεια, ή συγκρίνονται με κάτι άλλο, συγγραφείς έχουν εξετάσει γενικά το απόλυτο σφάλμα ή το απόλυτο σχετικό σφάλμα. Στην προσέγγιση της αθροιστικής συνάρτησης κατανομής Φ(x), παραδείγματος χάριν, αυτό θα ήταν $|G(x)-\Phi(x)|$ ή $|G(x)-\Phi(x)|/\Phi(x)$ αντίστοιχα, όπου το G(x) είναι η προσεγγίζουσα συνάρτηση.

Σε αυτό το κεφάλαιο, το απόλυτο σφάλμα θα αποδειχτεί εν συντομία με το |σφάλμα |,και το απόλυτο σχετικό σφάλμα με το |σχετικό σφάλμα |.

Ένας μεγάλος αριθμός πηγών συμπεριλαμβάνεται σε αυτό το κεφάλαιο, και μερικές από αυτές περιέχουν μερικά σφάλματα που εμφανίζονται στις πιο πρόσφατες διορθώσεις. Δεν έχουμε υποστηρίξει ότι έχουμε ανιχνεύσει όλα αυτά τα σφάλματα, αλλά εκείνα που έγιναν μεταξύ 1943 και 1969 μπορούν να αναφερθούν στο συσσωρευτικό δείκτη στο " Mathematics of Computation" τεύχη 1 έως 23. Επόμενα σφάλματα, ανεξάρτητα από την πηγή, αναφέρονται συχνά στο Mathematics of Computation.

4.1 Πίνακες

Οι πίνακες σχετικά με την κανονική κατανομή μονάδων είναι ένα απαραίτητο συστατικό για οποιοδήποτε εγχειρίδιο στη στατιστική θεωρία ή τις εφαρμογές του. Αυτό γίνεται επειδή για πολλές δεκαετίες η κανονική κατανομή κράτησε μία κεντρική θέση στις στατιστικές. Όπως επισημαίνεται, οι πίνακες της κανονικής κατανομής μονάδων αρκούν για τους υπολογισμούς συσχέτισης σε όλες τις κανονικές κατανομές.
Μια μεγάλη επιλογή των πινάκων που δίνουν τις πιθανότητες κάτω από την τυποποιημένη κανονική καμπύλη είναι διαθέσιμη. Η αθροιστική συνάρτηση κατανομής $\Phi(x)$ είναι συνήθως ταξινομημένη σε πίνακες στα εγχειρίδια εξετάζοντας τις εισαγωγικές στατιστικές και την ανάλυση δεδομένων, για τη γενική μελέτη και για τις εφαρμογές στους ειδικούς τομείς όπως τη γεωργία, τις επιχειρήσεις, την ιατρική και την ψυχολογία. Αυτοί οι πίνακες συνήθως έχουν ακρίβεια τεσσάρων δεκαδικών θέσεων.

(α) Όταν περιγράψουμε τη μεγάλη επιλογή των διαθέσιμων πινάκων των $\Phi(x)$, $\varphi(x)$, z_α, των παραγώγων των $\varphi(x)$, της αναλογία του Mills' $R(x)$, και τα λοιπά, θα απαριθμήσουμε μερικές από τις πιο προσιτές πηγές, που συγκεντρώνονται στις πρόσφατες πηγές, συμπεριλαμβανομένου ενός διασταυρωμένου τμήματος των αμερικανικών, βρετανικών, ρωσικών, ινδικών, και ιαπωνικών πινάκων.
Ο Greenwood και Hartley (1962) παραθέτουν έναν περιεκτικό κατάλογο πινάκων που δημοσιεύθηκαν μέχρι το 1958. Τελευταία ο Fletcher και λοιποί (1962b) περιλαμβάνουν μια ειδική συζήτηση των σφαλμάτων που εμφανίζονται στους δημοσιευμένους πίνακες που αναφέρονται, μαζί με τις διορθώσεις. Οι εξαιρετικά λεπτομερείς πίνακες της συνάρτησης του σφάλματος και τα πρώτα 20 παράγωγά τους είναι δημοσιευμένα στην

U.S.S.R (ΕΣΣΔ) δείτε Smimov (1960, 1965) Επίσης υπάρχει μια άριστη περιεκτική συζήτηση στους Johnson και Kotz (1970a).

(β) Εδώ δίνουμε έναν κατάλογο μόνο των ευκολότερα διαθέσιμων πινάκων. Ο πληρέστερος κατάλογος δίνεται στο Εθνικό Γραφείο των Προτύπων (1952) Οι συναρτήσεις που ταξινομούνται συχνότερα σε πίνακες είναι Φ(x), φ(x), και z_α, αλλά υπάρχουν παραλλαγές για τις ειδικές χρήσεις.

(γ) Ο Πίνακας 4. 1 συνοψίζει τις πληροφορίες που βρίσκονται σε επτά πηγές μερικές από αυτές μπορούν να φανούν για να είναι πολύ λεπτομερείς, ιδιαίτερα για την αθροιστική συνάρτηση κατανομής Φ(x). Ο ακριβέστερος πίνακας των ποσοστιαίων σημείων z_a, [όπου δίνεται το 1-α,ή το $\Phi(z_a)$]που μπορέσαμε να βρούμε ήταν αυτός του White (1970), δίνοντας το z_α σε 20 δεκαδικές θέσεις για 1-α=0.50(0.005)0.995 και για $a = 5\times10^{-k}$,2.5×10^{-k}, και $k = 1(1)20$. Δείτε επίσης Kelley (1948), με z_α σε 8 θέσεις.

(δ) Είναι ιστορικού ενδιαφέροντος ο κατάλογος μερικών από τους πιο παλαιούς γνωστούς πίνακες σχετικά με την κανονική κατανομή. Αν και αυτοί ήταν επιρρεπείς σε σφάλματα, είναι αξιοπρόσεκτο ότι τέτοιοι βαθμοί ακρίβειας λήφθηκαν χωρίς τη βοήθεια των υπολογιστών αλλά, ως επί το πλείστον, με μόνη τη χρήση των πιο πρωτόγονων υπολογιστών χειρός .

Πίνακας 4.1: Πίνακες της μοναδιαίας κανονικής όπου cdf Φ, pdf φ, Ποσοστιαία σημεία z_a [$\Phi(z_a) = 1 - a$], Παράγωγα των φ, και την αναλογία του Mill R(x), [$R(x) = \{1 - \Phi(x)\}/\varphi(x)$]

Προέλευση	Συνάρτηση	Κάλυψη	Δεκαδικές Θέσεις	Significant figures
Abramowitz and Stegun (1964)[α]	$\Phi(x), \varphi(x), \varphi'(x)$	x=0.0(0.02)3.00	15	
	$\Phi(x)$	x=3.0(0.05)5.0	10	
	$\varphi(x)$	x=3.0(0.05)5.0	9	
	$\varphi'(x)$	x=3.0(0.05)5.0	7	
	$\varphi^{(r)}(x)$, r=2,…,6	x=0.0(0.02)3.00(0.05)5.0	7-10	
	$-\log_{10}[1-\Phi(x)]$	x=5(1)50(10)100(50)500	5	
	$\varphi^{(r)}(x)$, r=7,…,12	x=0.0(0.1)5.0	7	
	$\varphi(z_\alpha)$, z_α	1-α=0.500(0.001)0.999	5	
	z_α	1- α=0.975(0.0001)0.9999	5	
	z_α	$\alpha=10^{-r}$, r=4,….,23	5	
Owen (1962)	$\Phi(x), \varphi(x)$	x=0.0(0.01)3.99	6	
	$1-\Phi(x)$	x=3.0(0.1)6.0(0.2)10(1)20, 25,30(10)100(25)200(50)500		
	z_α, $\varphi(z_\alpha)$	1-α=0.50(0.01)0.90(0.005).99 (0.001)0.999(0.00001)0.9999 κ.λ.π σε $1-10^9$	5	5
	$\Phi(x)/\varphi(x)$	x=0(0.1)3.99	4	
	$R(x)$, $\varphi'(x)$, $\varphi''(x)$, $\varphi'''(x)$	x=0.0(0.01)3.99	5	
	$R(x)$	x=3.0(0.1)6.0(0.2)10(1)20, 25,30(10)100(25)200(50)500		5

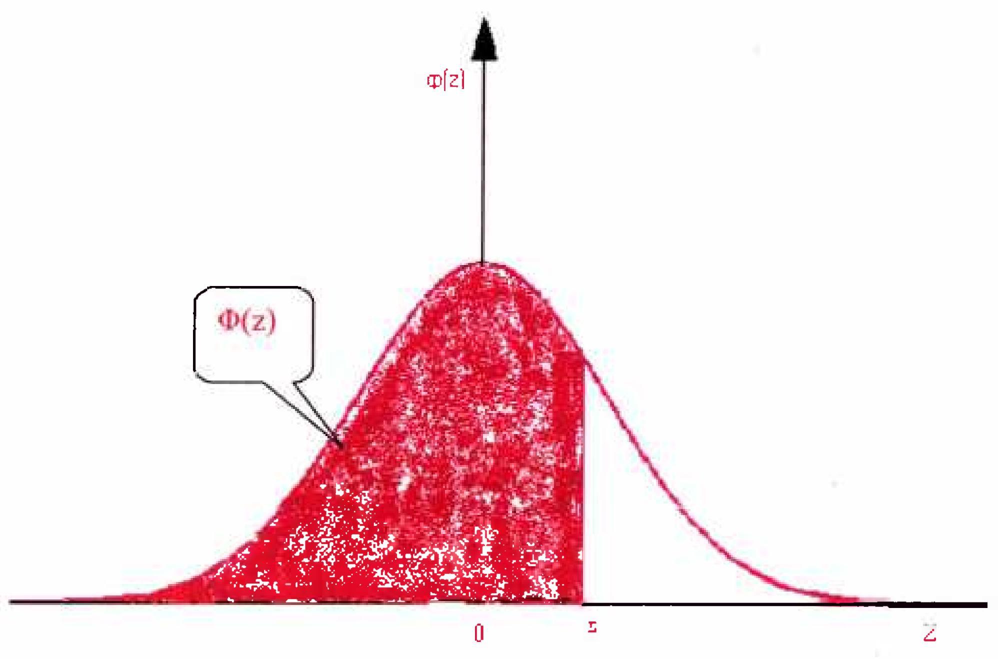
α: Οι πίνακες έχουν συμπληρωθεί από τους Zelen N.C.Severo

Καμία πραγματική δυσκολία δεν παρουσιάζεται από την εκτεταμένη χρήση των πινάκων της κανονικής κατανομής μονάδων. Οι πίνακες αυτοί είτε δίνουν το Φ(z) είτε μόνο το ολοκλήρωμα από μηδέν έως z.

Όταν δίδεται το Φ(z) οι πίνακες μας δίνουν τις διάφορες τιμές της αθροιστικής συνάρτησης κατανομής Φ(z) για κάθε τιμή του z.Δηλαδή μας δίνουν την τιμή του ολοκληρώματος: $\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$

Υπάρχουν έτοιμοι πίνακες που μας δίνουν, για z από –4 έως 4, τις τιμές της αθροιστικής συνάρτησης κατανομής Φ(z) της N(0,1). Η αθροιστική συνάρτηση κατανομής Φ(z) παριστάνεται γραφικά με το γραμμοσκιασμένο τμήμα του παρακάτω διαγράμματος 4.1 πού μας δίνει το εμβαδόν από -∞ έως z,που το συμβολίζουμε με Φ(z).

Οι περισσότεροι από τους πίνακες μόνο θετικές τιμές της μεταβλητής δίνουν, γιατί για τις αρνητικές χρησιμοποιούμε τις σχέσεις: Φ (z) =1 -Φ (-z) και φ(z) = φ(-z).



Διάγραμμα 4.1:*Γραφική παράσταση εμβαδού τυπικής κανονικής κατανομής από $-\infty$ έως z*

Προέλευση	Συνάρτηση	Κάλυψη	Δεκαδικές Θέσεις	Significant figures
Pearson and Hartley (1966)	$\Phi(x),\varphi(x)$	x=0.0(0.01)4.50 x=4.50(0.01)6.00	7 10	
	$-\log_{10}[1-\Phi(x)]$	x=5(1)50(10)100(50)500	5	
	z_α	1-α=0.50(0.001)0.999 0.98(0.0001) 0.9999 $1-10^{-r}$, r=4,….,9	4 4 4	
	$\varphi(z_\alpha)$	1-α=0.50(0.001)0.999	5	
Pearson and Hartley (1972)	z_α, $\varphi(z_\alpha)$	1-α=0.50(0.001)0.999	10	
	z_α	1-α=0.999(0.001)0.9999	8	
	$\varphi(z_\alpha)$	1-α=0.999(0.001)0.9999	9	
	$\Phi(x)-0.5$, $\varphi(x)$ $\varphi'(x)$, $\varphi''(x)$	x=0.0(0.02)6.20	6	
	$\varphi^{(r)}(x)$, r=3,…,9	x=0.0(0.02)6.20	5 σε1	
Rao (1966)	$\Phi(x)-0.5$	x=0.00(0.001)3.0(0.01) 4(0.01)4.9	6	
	$\varphi(x)$	x=0.00(0.001)3.0(0.01) 4(0.01)4.9	6	
	z_α	2(1-α)=0.01(0.01)0.99 10^{-r}, r=3,….,9	5	
	$2[1-\Phi(x)]$	x=0.25, 0.5(0.5)5.0	6	5
Smirnov (1965)	$\Phi(x)-0.5$, $\varphi(x)$	x=0.00(0.001)2.50(0.002) 3.40(0.005)4.00(0.01) 4.50 x=4.50(0.01)6.00	7 10	
	$-\log_{10}[1-\Phi(x)]$	x=5(1)50(10)100(50)500	5	
Yamuti (1972)	$\varphi(x)$	x=0.0(0.01)4.99		5
	$1-\Phi(x)]$	x=0.0(0.01)4.99		5
	$1-\Phi(x)]$	x=0.1(0.1)10.0		35
	z_α	α=0.0(0.001)0.499	5	

Όταν δίνεται το ολοκλήρωμα από μηδέν έως z, λόγω της συμμετρίας της κανονικής κατανομής, από τις τιμές αυτές μπορούμε να υπολογίσουμε την Φ(z) για κάθε z.Ισχύουν οι ακόλουθες σχέσεις : Φ (z) =1 -Φ (-z) για κάθε z και

$$\Phi(z)=\begin{cases}\frac{1}{2}+\int_0^z \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx & \text{για } z\geq 0\\ \frac{1}{2}-\int_0^{-z} \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx & \text{για } z\leq 0\end{cases}$$

4.1.1 Πρόγραμμα υπολογισμού της αθροιστικής συνάρτηση κατανομής Φ(z)

Εμείς υπολογίσαμε την τιμή του ολοκληρώματος $\Phi(z)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$

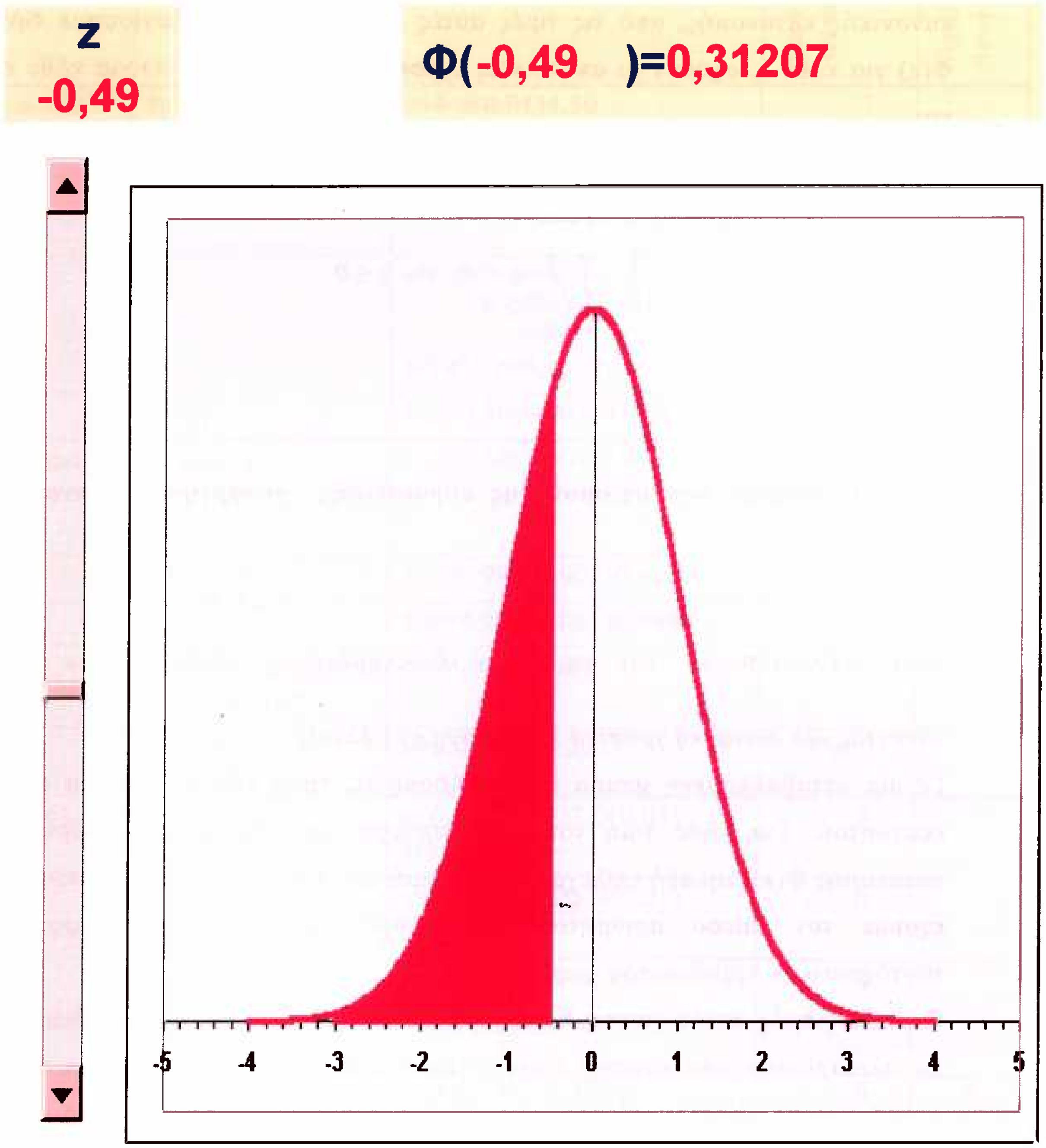
κάνοντας μία δυναμική γραφική παράσταση στο Excel.

Σε μία μεταβαλλόμενη μπάρα τοποθετήσαμε τις τιμές του z, με βήμα ενός εκατοστού. Για κάθε τιμή του z, βρίσκουμε την αθροιστική συνάρτηση κατανομής Φ(z).Δηλαδή επιλέγοντας την τιμή του z με προσέγγιση εκατοστού έχουμε τον άμεσο αριθμητικό υπολογισμό του Φ(z) καταγράφοντας ταυτόχρονα το εμβαδόν του χωρίου.

Ο υπολογισμός αυτός γίνεται καλλίτερα κατανοητός εάν είμαστε συνδεμένοι με ηλεκτρονικό υπολογιστή, όπου μετακινώντας το βελάκι επάνω στην μπάρα παρατηρούμε ταυτόχρονα το εμβαδόν που διαγράφεται και πόσο είναι αριθμητικά αυτό το εμβαδόν.

Συνεπώς με αυτό τον τρόπο παράγουμε μόνοι μας τις τιμές του ολοκληρώματος Φ(z) που μπορούμε να παίρνουμε έτοιμες από πίνακες. Έτσι έχουμε όλες τις τιμές του ολοκληρώματος για κάθε τιμή του z από –4 έως 4 και μάλιστα με ταυτόχρονη καταγραφεί του εμβαδού του χωρίου της κατανομής.

Στο παρακάτω διάγραμμα 4.2 παίρνουμε μία εικόνα του πως εργαζόμαστε.



Διάγραμμα 4.2:*Δυναμική γραφική παράσταση υπολογισμού του εμβαδού από* $-\infty$ *έως z της τυπικής κανονικής κατανομής*

4.1.2 Εκατοστιαία σημεία της κανονικής κατανομής

Πίνακας 4. 2:Εκατοστιαία σημεία της κανονικής κατανομής, ως τυποποιημένες αποκλίσεις (τιμές της Z_a)

a	Z_a
0.5	0.000000
0.6	0.253347
0.7	0.524401
0.75	0.674490[α]
0.8	0.841621
0.9	1.281552
0.95	1.644854
0.975	1.959964
0.99	2.326348
0.995	2.575829
0.9975	2.807034
0.999	3.090232

[α] Η τιμή $Z_{0,75}$ (= 0.6745), στο ανώτερο τεταρτημόριο της μοναδιαίας κανονικής κατανομής, καλείται περιστασιακά πιθανό σφάλμα της κατανομής, αν και αυτή η ονοματολογία χρησιμοποιείται σπάνια αυτή τη στιγμή. Το πιθανό σφάλμα της κατανομής (1) είναι, φυσικά, $Z_{0,75}\ \sigma$ ''

Υπάρχουν πολλές άλλες δημοσιεύσεις που περιέχουν τις διάφορες μορφές των πινάκων της κανονικής κατανομής. Υπάρχει ανάγκη για τους εκτενείς πίνακες της κανονικής κατανομής που δίνονται. Εδώ περιοριζόμαστε, στον πίνακα 4.2, σε μερικές συνήθως χρησιμοποιημένες τιμές Z_a. Οι πίνακες με τις αποκλίσεις της τυχαίας μοναδιαίας κανονικής κατανομής (απεικονίζουν τις τιμές μιας τυχαίας μεταβλητής που έχει μια κανονική κατανομή μονάδων) έχουν κατασκευαστεί από τους πίνακες των τυχαίων αριθμών (απεικονίζουν τις τιμές μιας τυχαίας μεταβλητής, που έχει μια διακριτή ορθογώνια κατανομή με βάση τους ακέραιους αριθμούς 0-9). Το 1948 ο Wold (1948) δημοσίευσε ένα σύνολο από 25000 τυχαίες μοναδιαίες κανονικές αποκλίσεις(με 3 δεκαδικές θέσεις), βασισμένο στον Kendall και σε πίνακες τυχαίων αριθμών. Ένα σύνολο από 10400 τυχαίες μοναδιαίες κανονικές αποκλίσεις (επίσης με 3 δεκαδικές θέσεις), που βασίζονται στον πίνακα Tippett (1927) των τυχαίων αριθμών, δημοσιεύθηκε από Sengupta και Bhattacharya (1958). Αυτοί αντικατέστησαν ένα πιο παλαιό σύνολο πινάκων, που πρωτοεμφανίστηκε το 1936 [Mahalanobis et Al (1934)] που βρέθηκαν για να ελέγξουν ένα αριθμό σφαλμάτων.

Ένα σύνολο από 100000 τυχαίες μοναδιαίες κανονικές αποκλίσεις, με 3 δεκαδικές θέσεις, που βασίζονται στο πρώτο μισό-εκατομμύριο τυχαίων παραχθέντων αριθμών το 1947, δημοσιεύθηκε από τον RAND(1955).Στον Buslenko και λοιποί (1966) υπάρχει ένας πίνακας με 1000 τυχαίες μοναδιαίες κανονικές αποκλίσεις, με 4 δεκαδικές θέσεις. Αυτοί υπολογίστηκαν από τις τιμές πέντε ανεξάρτητων τυχαίων μεταβλητών R_1..., R_5,όπου κάθε μια κατανέμεται τυχαία, μέσα στο εύρος 0 έως 1, χρησιμοποιώντας τους τύπους $Z = X - 0{,}01(3X - X^3)$ όπου

$$X = \frac{1}{\sqrt{5}}\sum_{j=1}^{5}\sqrt{3}(2R_j - 1)$$

Σημειώνουμε ότι $\sqrt{3}(2R_j - 1)$ έχει μια τυποποιημένη ορθογώνια κατανομή.

4.2 Αλγόριθμοι προσεγγίσεων

Η πιο κοινή χρήση της κανονικής κατανομής είναι ως μία προσέγγιση, όπου είτε η κανονικότητα αποδίδεται στο πώς να είναι η κατανομή στην κατασκευή ενός μοντέλου ή μια γνωστή κατανομή αντικαθίσταται από μια κανονική κατανομή με την ίδια αναμενόμενη τιμή και τυπική απόκλιση.

Ένα διαφορετικό είδος στην προσέγγιση είναι αναπτυγμένο σε σχέση σύνδεση με τον υπολογισμό των συναρτήσεων Φ(.), Z(.) στους υπολογιστές. Αυτές οι προσεγγίσεις υιοθετούν συνήθως τις πολυωνυμικές εκφράσεις. Δίνουν αρκετά υψηλή ακρίβεια, μερικές φορές μόνο μέσα στα καθορισμένα όρια των τιμών της μεταβλητής. Έξω από αυτά τα όρια μπορούν να δώσουν αρκετά πτωχή προσέγγιση.

4.2.1 Ο Zelen και ο Severo (1964) αναφέρουν, μεταξύ άλλων, τους ακόλουθους τύπους, οι οποίοι είναι βασισμένοι στους τύπους που δίνονται από τον **Hastings (1955):**

$$\Phi(x) = 1 - \left(a_1 t + a_2 t^2 + a_3 t^3\right) Z(x) \qquad (4.1)$$

με $t = (0.33267x)^{-1}$,$a_1 = 0.4361836$,$a_2 = -0.1201676$ και $a_3 = 0.9372980$

Το λάθος στο $\Phi(x)$ για $x \geq 0$ είναι μικρότερο από 1×10^{-5}

$$\Phi(x) = 1 - \frac{1}{2}\left(1 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4\right)^{-4} \qquad (4.2)$$

με a_1=0.196854, a_2=0.115194, a_3=0.000344, a_4=0.0195227

Το λάθος στο $\Phi(x)$ για $x \geq 0$ είναι μικρότερο από 2.5×10^{-4}

$$Z(x) = \left(a_0 + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_6 x^6\right)^{-1} \qquad (4.3)$$

με a_0=2.490895, a_2=1.466003 a_4=-0.024393, a_4=0.178257

Το λάθος στο $Z(x)$ για $x \geq 0$ είναι μικρότερο από 2.7×10^{-3}

4.2.2 Τα πολύ ακριβή αποτελέσματα μπορούν να επιτευχθούν με τον τύπο**[Hart (1966)]**

$$1 - \Phi(x) = \left(x\sqrt{2\pi}\right)^{-1}\left[\exp\left(-\frac{x^2}{2}\right)\right] \times$$

$$\left[1 - \frac{\left(1 + bx^2\right)^{\frac{1}{2}}}{\left(1 + ax^2\right)}\left\{x\sqrt{\frac{\pi}{2}} + \left[\frac{1}{2}\pi x^2 + \frac{\left(1 + bx^2\right)^{\frac{1}{2}}}{\left(1 + ax^2\right)} e^{-\frac{x^2}{2}}\right]^{\frac{1}{2}}\right\}^{-1}\right] \qquad (4.4)$$

με

$$a = \frac{1}{2\pi}\left[1 + \left(1 + 6\pi - 2\pi^2\right)^{\frac{1}{2}}\right] = 0.212024$$

$$b = \frac{1}{2\pi}\left[1 + \left(1 + 6\pi - 2\pi^2\right)^{\frac{1}{2}}\right]^2 = 0.282455$$

4.2.3 Για x> 2, **ο Schucany και ο Gray (1968)** έχουν κατασκευάσει έναν απλούστερο τύπο

$$1-\Phi(x)=\left[\left(x^2+2\right)\sqrt{2\pi}\right]^{-1}\times x\left[\exp\left(-\frac{x^2}{2}\right)\right]\frac{x^6+6x^4+14x^2-28}{x^6+5x^4-20x^2-4} \tag{4.5}$$

Ο οποίος είναι ακόμα καλύτερος από τον (4.4) για x> 3. [δίνει ανάλογο λάθος με τον (4.4) για $5\le x\le 10$ είναι περίπου 0.5×10^{-5} όπου στο (4.5) μειώνεται από 0.39×10^{-5} για x = 5 έως 0.42×10^{-7} για x = 10.]
Με την χρήση των μάλλον επιμελημένων τύπων, αρκετά αξιοπρόσεκτη ακρίβεια μπορεί να επιτευχθεί. Ο Strecock (1968) δίνει τους τύπους για τις τιμές του erf (X) [$erf(x)=2\pi^{-\frac{1}{2}}\int_0^x e^{-t^2}dt=2\Phi\left(x\sqrt{2}\right)-1$] και της αντίστροφης συνάρτησης [inver f(y) όπου erf (inverf(y)) =y],ο οποίος δίνει ακρίβεια 22 δεκαδικών ψηφίων για $|x|$ (ή $|inverf(y)|$) λιγότερο από 7.85.

4.2.4 Ο McGillivray και ο Kaller (1966) έχουν εξετάσει την απόκλιση μεταξύ $\Phi(x)$ και $\Phi(x)+a_{2r}Z(x)\mathrm{H}_{2r-1}(x)$ όπου $\mathrm{H}_{2r-1}(x)$ είναι το πολυώνυμο του Hermite της τάξης $2r-1$και α_{2r} είναι μια σταθερά που επιλέγεται έτσι ώστε το $1+a_{2r}\mathrm{H}_{2r}(x)$ να μην μπορεί να είναι αρνητικό. Οι μέσοι των a_{2r} πρέπει να κυμαίνονται μεταξύ μηδενός και του $A_{2r}=\left|\inf_x H_{2r}(x)\right|^{-1}$

Η δεύτερη συνάρτηση $\Phi(x)+a_{2r}Z(x)H_{2r-1}(x)$είναι η αθροιστική συνάρτηση κατανομής μιας συμμετρικής κατανομής που έχει τις πρώτες κεντρικές ροπές r άρτιες, (και φυσικά όλες τις περιττές ροπές),ίδιες με αυτές της κανονικής κατανομής μονάδων.
Η απόκλιση δεν μπορεί να υπερβεί $A_{2r}\sup_x\{Z(x)|H_{2r-1}(x)|\}$

Οι τιμές αυτής της ποσότητας, για r = 2,3, και 4 είναι 0.10, 0.03, και 0.005, αντίστοιχα.[Φυσικά οι άλλες κατανομές με τις ίδιες (μηδενικές) περιττές κεντρικές ροπές και τις πρώτες κεντρικές ροπές r άρτιες, θα μπορούσαν να έχουν μεγαλύτερες αποκλίσεις, αλλά αυτά τα αποτελέσματα δίνουν μια χρήσιμη ιδέα της ακρίβειας που λαμβάνεται από την εξίσωση των ροπών.]

4.2.5 Η προσέγγιση

$$\Phi(x)=\frac{1}{2}\left[1+\left\{1-\exp\left(\frac{-2x^2}{\pi}\right)\right\}^{\frac{1}{2}}\right] \tag{4.6}$$

επιτεύχθηκε από τον **Pólya (1945).** Αυτή έχει ένα μέγιστο λάθος 0.003 όταν x = 1.6. Ο **Cadwell (1951)** τροποποιείσε το (4.6) σε

$$\Phi(x)=\frac{1}{2}\left[1+\left\{1-\exp\left(-2\pi^{-1}x^2-\frac{2}{3}\pi^{-2}(\pi-3)x^4\right)\right\}^{\frac{1}{2}}\right] \tag{4.7}$$

Πέρα από το εύρος 0 <x < 3.5, το μέγιστο λάθος του (4.7) είναι 0.0007, όταν x=2.5.Η μορφή (4.7) θα μπορούσε να μην χρησιμοποιηθεί για τις μεγάλες τιμές του x. Ο Canwell προτείνει για εμπειρικούς λόγους, την προσθήκη των όρων

$$-0.0005x^6+0.00002x^8$$

στον εκθέτη στον (4.7). Αυτό μειώνει το μέγιστο λάθος σε 0.00005.

4.2.6 Ο Carta(1975), προεκτείνει προσεγγίσεις παρόμοιες με αυτές του Hastings στο (4.2), της μορφής

$$\Phi(x)=1-\frac{1}{2}\left(a_1+a_2x+\ldots+a_nx^{n-1}\right)^{-2q} \text{ για } \chi\geq 0 \tag{4.8}$$

Παρατηρούμε την προστιθέμενη προσαρμοστικότητα της μεταβλητής του πρώτου συντελεστή, σε αντιδιαστολή με τη σταθερή τιμή 1 στο (4.2). Για τις διαφορετικές επιλογές του n και του q, ο Carta (1975) έχει παρουσιάσει τους συντελεστές a_i οι οποίοι παρέχουν το ελάχιστο απόλυτο σφάλμα. Για παράδειγμα, από πίνακες του Carta, έχουμε τους συντελεστές που αντιστοιχούν σε n = 6 και q = 4 όπως

0.9999998582	0.0487385796	0.02109811045
0.003372948927	-0.00005172897742	0.00008569579420

Το απόλυτο σφάλμα για αυτήν την προσέγγιση (για όλα τα x≥ 0) είναι μικρότερο από 1.2×10^{-6}

Σημειώστε ότι σε αυτή την περίπτωση (και επίσης για μερικές άλλες επιλογές των n,q), ο κύριος συντελεστής είναι πολύ κοντά 1 [όπως στο (4.2)]
Ο Carta επίσης έχει παρουσιάσει παρόμοιες προσεγγίσεις για x περιορισμένο στα διαστήματα [0, 3.09], [0,4.00], και [0,5.20].

4.2.7 Ο **Badhe (1976)** παρουσίασε την ακόλουθη προσέγγιση που είναι εύκολα εκτελέσιμη σε έναν υπολογιστή τσέπης :

$$\Phi(x)=1-\frac{Z(x)}{x}\left[1-\frac{1}{Y}\left(1+\frac{1}{Y}\left\{7+\frac{1}{Y}\left[55+\frac{1}{Y}\left(445+\frac{3745Q_1(x)}{Y}\right)\right]\right\}\right)\right] \tag{4.9}$$

όπου $Y=x^2+10$ και

$$Q_1(x)=8.5\left(x^2-0.4284639753x^{-2}+1.240964109\right)^{-1}+1$$

Ο Badhe (1976) έχει επισημάνει ότι η προσέγγιση στο (4.9) είναι καλή όταν x > 4, αλλά βεβαίως ακατάλληλη για x ≤ 2. Για την περίπτωση όπου x ≥ 2, ο Badhe έχει παρουσιάσει μια έβδομου βαθμού πολυωνυμική προσέγγιση (που επιτεύχθηκε από τον Chebyshev) και δίνεται από

$$\Phi(x)=\frac{1}{2}+x\{a+Y[b+Y(c+Y\{d+Y[e+Y(f+Y(g+hY))]\})]\}$$

όπου :

$$Y=\frac{x^2}{32}$$

a=0.3989422784

b=-2.127690079

c=10.2125662121

d=-38.8830314909

e=120.2836370787

f=-303.2973153419

g=575.073131917

h=-603.9068092058

Το μέγιστο απόλυτο **σφάλμα για αυτήν** την προσέγγιση, για $x\in[0,2]$ είναι 0.2×10^{-8}. Με **την χρησιμοποίηση της** επέκτασης του Hermite:

$$\sqrt{2\pi}\int_0^x Z(t)dt = 2e^{-\frac{x^2}{8}}\sum_{n=0}^{\infty}\frac{1}{(2n+1)!}\left(\frac{x}{2}\right)^{2n+1}\mathrm{H}_{2n}\left(\frac{x}{2}\right) \tag{4.10}$$

όπου $\mathrm{H}_n(x)$ είναι το νιοστό πολυώνυμο Hermite, και η γνωστή επαναληπτική σχέση :

$H_{n+1}(x) = xH_n(x) - nH_{n-1}(x)$ για n = 1,2,... όπου $H_0(x) \equiv 1$ και $H_1(x) = x$

4.2.8 O Kerridge και ο Cook (1976) προτείνουν χρησιμοποίηση της σειράς

$$\sqrt{2\pi}\int_0^x Z(t)dt = xe^{-\frac{x^2}{8}}\sum_{n=0}^{\infty}\frac{1}{2n+1}\theta_{2n}\left(\frac{x}{2}\right) \tag{4.11}$$

για τον υπολογισμό **του** $\Phi(x)$σε έναν υπολογιστή. Στο (11), $\theta_n(x) = \dfrac{x^n H_n(x)}{n!}$ το οποίο υπολογίζεται **εύκολα χρησιμοποιώντας** την **επαναληπτική σχέση**

$$\theta_{n+1} = \frac{x^2(\theta_n - \theta_{n-1})}{n+1} \quad , \mathrm{n} = 1,2,...$$

βλέπε Odeh και Evans (1974).

4.2.9 O Page (1977) εξετάζει απλές προσεγγίσεις της μορφής

$$\Phi(x) = \frac{e^{2y}}{1+e^{2y}} \tag{4.12}$$

όπου $y = a_1 x(1 + a_2 x^2)$και καθορισμένα(σταθερά) a_1 = 0.7988 και a_2 = 0.04417 για να παρέχει σε μια προσέγγιση το μέγιστο σφάλμα 0.14×10^{-3}

O Page έχει παρουσιάσει επίσης μια παρόμοια απλή προσέγγιση για το ποσοστιαίο σημείο που δίνει δύο δεκαδικά ακρίβεια.

4.2.10 O Derenzo (1977) παρείχε μια προσέγγιση στην αθροιστική συνάρτηση κατανομής της μοναδιαίας κανονικής ως

$$\Phi(x)=1-\frac{1}{2}\exp\left\{-\frac{(83x+351)x+562}{\left(\frac{703}{x}\right)+165}\right\} \quad , \quad x>0 \qquad (4.13)$$

με ένα μέγιστο απόλυτο σφάλμα του 0.042% για $x \in (0,5.5]$. Μια άλλη προσέγγιση που παρουσιάζει ο Derenzo είναι :

$$\Phi(x)=1-\frac{1}{\sqrt{2\pi}x}\exp\left(-\frac{x^2}{2}-\frac{0.94}{x^2}\right) \qquad (4.14)$$

με ένα μέγιστο απόλυτο σφάλμα του 0.040% για $x \geq 5.5$. Ο Derenzo (1977) επίσης παρείχε μια προσέγγιση για το ποσοστιαίο σημείο x_p (για όλα τα p) όπως

$$x_p=\left\{\frac{[(4y+100)y+205]y^2}{[(2y+56)y+192]y+131}\right\}^{\frac{1}{2}} \qquad (4.15)$$

όπου, $y=-\log\left(1-\frac{p}{2}\right)$ με ένα μέγιστο απόλυτο σφάλμα για 1.3×10^{-4} για $x \in (0,5.2)$. Για $x \in [5.2,2.6)$ ο Derenzo έχει δώσει την προσέγγιση

$$x_p=\left\{\frac{[(2y+280)y+572]y}{(y+144)y+603}\right\}^{\frac{1}{2}}$$

με ένα μέγιστο απόλυτο σφάλμα του 4×10^{-4}.

Απλούστερες (αλλά όχι απαραιτήτως ακριβέστερες για όλα τα x) προσεγγίσεις για την αθροιστική συνάρτηση κατανομής $\Phi(x)$ έχουν δοθεί από πολλούς άλλους. Έχουν κάνει μελέτες πολλών από τους αλγορίθμους που είναι διαθέσιμοι για τον υπολογισμό της αθροιστικής συνάρτησης κατανομής $\Phi(x)$.

4.2.11 Ο Moran (1980), με ελαφρώς τροποποιημένο έναν τύπο του Strecock (1968) που δόθηκε για τη συνάρτηση λάθους, βρήκε την προσέγγιση.

$$\Phi(x) = \frac{1}{2} + \frac{1}{\pi}\sum_{n=0}^{12}\left(n+\frac{1}{2}\right)^{-1} e^{-\frac{\left(n+\frac{1}{2}\right)^2}{9}} \sin\left\{\frac{\sqrt{2}}{3}\left(n+\frac{1}{2}\right)x\right\} \tag{4.16}$$

Η συγκρίνοντας την προσέγγιση που δίνεται από τον Strecock, αυτή η προσέγγιση είναι απλούστερη στη χρησιμοποίηση και έχει ακρίβεια εννέα δεκαδικών ψηφίων για $|x| \le 7$.

4.2.12 O Shore (1982), με τη χρησιμοποίηση μιας "betalike" αθροιστικής κατανομής G(X) [με G(-∞) =0, G(∞) = 1, και $G(0) = \frac{1}{2}$ και

$$\frac{dG}{dx} = \begin{cases} cG^{k_1}(1-G)^{k_2} & , 0 \le G \le 1 \\ c(1-G)^{k_1} G^{k_2} & , \frac{1}{2} \le G \le 1 \end{cases}$$

όπου $k_1 > 0$] για να προσεγγίσει το $\Phi(x)$, παρήγαγε τις ακόλουθες τρεις προσεγγίσεις για το ποσοστιαίο σημείο x_p:

$$x_p = -5.5310\left\{\left(\frac{1-p}{p}\right)^{0.1193} - 1\right\}, \qquad p > \frac{1}{2} \tag{4.17}$$

$$x_p = -0.4115\left\{\frac{1-p}{p} + \ln\left(\frac{1-p}{p}\right) - 1\right\}, \qquad p \ge \frac{1}{2} \tag{4.18}$$

$$x_p = -a\ln\left(\frac{1-p}{p}\right) + b, \qquad p \ge \frac{1}{2} \tag{4.19}$$

όπου a και b πρέπει να ικανοποιούν τη σχέση:

$$b = \left[1 - 1.3682a^2\right]^{\frac{1}{2}} - 1.3862a$$

Διαφορετικά η απλούστερη κατά προσέγγιση μορφή του $b = 1.3086 - 2.3735a$ (που είναι ευκολότερο να χρησιμοποιηθεί, όταν επιδιώκεται τιμή του α για την οποία δίνει καλύτερη προσαρμογή για οποιαδήποτε επιθυμητό εύρος του x).

Η προσέγγιση στο (4.17) είναι η ακριβέστερη των τριών με μια μέγιστη απόλυτη διαφορά 0.0073 (0.5%) για $0 \le x \le 2.3$. Είναι ενδιαφέρον να σημειωθεί ότι για $b = 0$ έχουμε $a = 0.5513 = \frac{\sqrt{3}}{\pi}$, οπότε σ' αυτή την περίπτωση η απλή προσέγγιση στο (4.19) γίνεται η λογιστική προσέγγιση (με την παράμετρο θέσης 0 και την παράμετρος μορφής $\frac{\sqrt{3}}{\pi}$).

Για $-0.5 \le x \le 2.2$ η καλύτερη προσέγγιση του (4.19) είναι με $a = 0.495$ και $b = 0.1337$, για x> 2. 2, η καλύτερη προσέγγιση λαμβάνεται με $a = 0.4506$ και $b = 0.\ 2252$.

Ο Shore το (1982) έχει λάβει επίσης μια καλή προσέγγιση για την τυποποιημένη κανονική συνάρτηση πυκνότητας πιθανότητας όπως:

$$Z(x) = 1.4184(1-p)^{0.8632} p \ , \quad \mathrm{p} = \Phi(\mathrm{x}) \ge \frac{1}{2} \tag{4.20}$$

Αυτή η προσέγγιση παρέχει πλησιέστερες αποδόσεις στην προσέγγιση της αναλογίας του Mills όπως:

$$\frac{Z(x)}{1-\Phi(x)} = \begin{Bmatrix} \left[1.4184 p^{0.8632}\right]^{-1} , & \mathrm{p} \le \frac{1}{2} \\ \left[1.4184(1-p)^{-0.1368} p\right]^{-1}, & p \ge \frac{1}{2} \end{Bmatrix} \tag{4.21}$$

Ο Shore συζητά επίσης την ακρίβεια αυτής της προσέγγισης.

4.2.13 O Shah (1985) πρότεινε την ακόλουθη προσέγγιση για

$$\Phi(x) - \frac{1}{2} \ , \ \mathrm{x} \ge 0$$

$$\begin{cases} \dfrac{x(4.4-x)}{10} & , \ 0 \le \mathrm{x} \le 2.2 \\ 0.49 & , \ 2.2 < \mathrm{x} < 2.6 \\ 0.50 & , \ \mathrm{x} \ge 2.6 \end{cases} \tag{4.22}$$

Ακόμα κι αν αυτή η προσέγγιση είναι απλή στην χρήση, δεν ήταν σκόπιμο να προσεγγίσει δεξιές ουρές για $x \geq 2.6$.

4.2.14 Για αυτόν τον λόγο ο **Norton (1989)** πρότεινε ως προσεγγίσεις για το $1-\Phi(x)$ τις

$$\frac{1}{2}e^{-\frac{(x^2+x)}{2}} \quad , \quad 0 \leq x \leq 2.6 \tag{4.23}$$

$$\frac{1}{2}e^{-\frac{(x^2+1.2x^{0.8})}{2}} \quad ,0 \leq x \leq 2.7 \tag{4.24}$$

$$\frac{Z(x)}{x} \quad , \quad x > 2.7 \tag{4.25}$$

Σε αυτές τις προσεγγίσεις που δίνονται από τον Norton, έχουν γίνει πολλά σχόλια και κριτικές από πολλούς και επίσης έχουν δοθεί απαντήσεις σε αυτά τα σχόλια από τον Norton.

4.2.15 O Riffenburgh (1967) έχει προτείνει να προσεγγιστεί μια συμμετρική περικομμένη κανονική κατανομή μονάδων από τη συνάρτηση πυκνότητας

$$\frac{\frac{1}{2}(Z(x)-Z(c))}{\Phi(c)-\frac{1}{2}-cZ(c)} \qquad -c \leq x \leq c$$ όπου - c, c είναι τα σημεία της αποκοπής.

Η χρήση αυτής της προσέγγισης συστήνεται μόνο όταν το c υπερβαίνει το 1 (κατά προτίμηση $c \geq 1.5$). Οι πίνακες διασποράς (με 3 δεκαδικά ψηφία) της προσεγγίζουσας κατανομής είναι δοσμένοι από τον Riffenburgh (1967) για c= 0.8(0.1)1.2(0.05)4.00 και επίσης της $P_r[X \leq x]-\frac{1}{2}$ (με 4 δεκαδικά ψηφία) για c=1.2(0.1)3.0 και x σε διαστήματα του 0.05. (O Riffenburgh έχει επίσης διαδικασίες δοκιμής βασισμένες σε αυτήν την κατανομή).

4.2.16 O Beasley και ο Springer (1977) έχουν παράσχει έναν αλγόριθμο για τον υπολογισμό του ποσοστιαίου σημείου x_p για μια καθορισμένη τιμή, έτσι ώστε

$$p = \int_{-\infty}^{x_p} Z(t)dt$$

Η υπορουτίνα FORTRAN αντικαθιστά το p με το $q = p - \frac{1}{2}$ και κατόπιν συγκρίνει το $|q|$ με το 0.42.Εάν $|q| \le 0.42$,το x_p καθορίζεται από μια λογική προσέγγιση

$$x_p = \frac{q\mathrm{A}(q^2)}{\mathrm{B}(q^2)} \tag{4.26}$$

όπου Α και Β είναι πολυώνυμα 3^{ου} και 4^{ου} βαθμού αντίστοιχα, ενώ εάν $|q|$>0.42 μια βοηθητική μεταβλητή, $r = \left\{\ln\left(\frac{1}{2} - |q|\right)\right\}^{\frac{1}{2}}$ διαμορφώνεται αρχικά και στη συνέχεια το x_p ως

$$x_p = \pm\frac{C(r)}{D(r)} \tag{4.27}$$

όπου το C και το D είναι πολυώνυμα 3^{ου} και 2^{ου} βαθμού αντίστοιχα, και ως πρόσημο λαμβάνεται αυτό του q.

4.2.17 Συζητάμε τώρα μερικά όρια στην τιμή του Φ(x). Οι διάφορες ανισότητες για την αναλογία του **Mills's** μπορούν επίσης να ερμηνευθούν ως όρια για το Φ(x) ή το Z(x). Χρησιμοποιώντας ένα απλό γεωμετρικό επιχείρημα (βασισμένο στην κοινή κατανομή δύο ανεξάρτητων κανονικών μεταβλητών μονάδων) μπορεί να δειχθεί ότι

$$\frac{1}{2}\left[1 + \left(1 - e^{-\frac{x^2}{2}}\right)^{\frac{1}{2}}\right] \le \Phi(x) \le \frac{1}{2}\left[1 + \left(1 - e^{-x^2}\right)^{\frac{1}{2}}\right] \tag{4.28}$$

[δείτε **D'Ortenzio(1965)**]. Με μία συνέχεια του επιχειρήματος, στην αριστερή πλευρά του (4.28) το $\left(1 - e^{-\frac{x^2}{2}}\right)$ μπορεί να αντικατασταθεί από το

$1-e^{-\frac{x^2}{2}}+\left(2\pi^{-1}-\frac{1}{2}\right)^2 e^{-x^2}$ και στην δεξιά πλευρά το $\left(1-e^{-x^2}\right)$μπορεί να αντικατασταθεί από το $1-e^{-x^2}-\left(1-2\pi^{-1}\right)^2 e^{-x^2}$

O Schonfelder (1978) συζήτησε περαιτέρω στις επεκτάσεις Chebyshcv για το λάθος και τις σχετικές συναρτήσεις, συμπεριλαμβανομένου του Φ(x).O Hamaker (1978) παρουσίασε απλές προσεγγίσεις για την αθροιστική συνάρτηση κατανομής Φ(x), καθώς επίσης και για το ποσοστιαίο σημείο x_p.O Lin (1988) συζήτησε τις εναλλακτικές λύσεις στις προσεγγίσεις Hamakers, Schmeiser (1979) έδωσε τις εύκολες προσεγγίσεις για το ποσοστιαίο σημείο x_p όπου μπορούσαν να χρησιμοποιηθούν και υπολογιστές τσέπης, ο Lin (1989) παρουσίασε τις προσεγγίσεις στην κανονική πιθανότητα ουρών καθώς επίσης και τις αντίστροφές τους που είναι εύχρηστες ακόμη και με υπολογιστή τσέπης.

O Lin (1990) έχει προτείνει επίσης μία απλούστερη λογιστική προσέγγιση του Φ(x) και του x_p [επίσης δείτε το σχόλιο μετά από (4.19)].

Ακόμα κι αν όλες αυτές οι προσεγγίσεις είναι γρήγορες και εύχρηστες, οι σχετικές ακρίβειές τους μπορούν να ποικίλουν εντυπωσιακά, και επομένως η χρήση τους πρέπει να κριθεί με βάση την ιδιαίτερη κατάσταση που κάποιος σκοπεύει να χρησιμοποιήσει. Ο σύγχρονος προσωπικός υπολογιστής με τη μεγάλη δύναμη και τη μνήμη του κάνει μέρος της εργασίας (εάν όχι όλη) άσκοπη, όπως οι ακόλουθες παρατηρήσεις δείχνουν.

O Fleming (1989) έχει υποστηρίξει τη χρήση της αριθμητικής ολοκλήρωσης για να προσεγγίσει το Φ(x),και να μπορεί να αποδειχθεί εύκολα χρησιμοποιώντας ένα πρόγραμμα υπολογισμών με λογιστικό φύλλο (spreadsheet) και έναν προσωπικό υπολογιστή. Παραδείγματος χάριν, το Φ(x) μπορεί να υπολογιστεί χρησιμοποιώντας Lotus 1-2-3 για να προσεγγίσει την περιοχή κάτω από την τυποποιημένη κανονική συνάρτηση πυκνότητας πιθανότητας Z(x). O Fleming (1989) έχει σημειώσει ότι μια ορθογώνια ολοκλήρωση δίνει αποτελέσματα με ακρίβεια τέσσερις δεκαδικές θέσεις για $0 \leq x \leq 3$, όπου το πλάτος του διαστήματος λαμβάνεται ως 0.01.

Όπως έχει επισημάνει, άλλες μέθοδοι ολοκλήρωσης (όπως ο τραπεζοειδής κανόνας, ο κανόνας Simpson, ο κανόνας three-eighths του Newton) είναι επίσης εύκολα προσαρμόσιμες και θα είναι ακόμα ακριβέστερες στην αξιολόγηση του Φ(x). Όταν ένας ισχυρός προσωπικός υπολογιστής με ένα πρόγραμμα υπολογισμών σε λογιστικό φύλλο (spreadsheet) είναι διαθέσιμος, πρέπει να μπορεί να χρησιμοποιηθεί κατά προτίμηση, όταν στηρίζεται σε απλές στην χρήση προσεγγίσεις.

4.3 Χαρτί πιθανοτήτων

Μία μηχανική μέθοδος για τον σχεδιασμό μιας κανονικής καμπύλης πυκνότητας πιθανότητας έχει περιγραφεί από τον Edwards (1963). Το κανονικό έγγραφο πιθανότητας είναι έγγραφο γραφικών παραστάσεων με μια φυσική κλίμακα στην οριζόντια κατεύθυνση (άξονα τετμημένων), ενώ οι αποστάσεις στην κάθετη κλίμακα (άξονα τεταγμένων) είναι ανάλογες προς

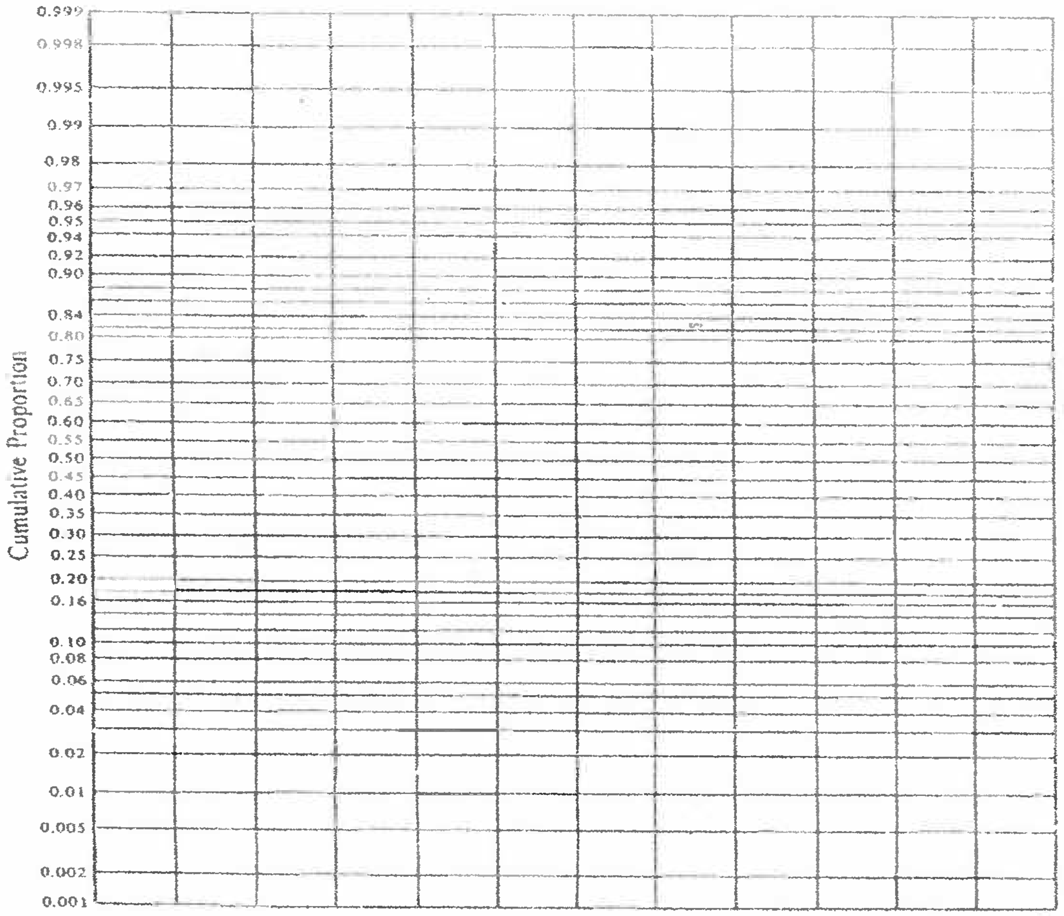


Figure 13.2 Normal Probability Paper

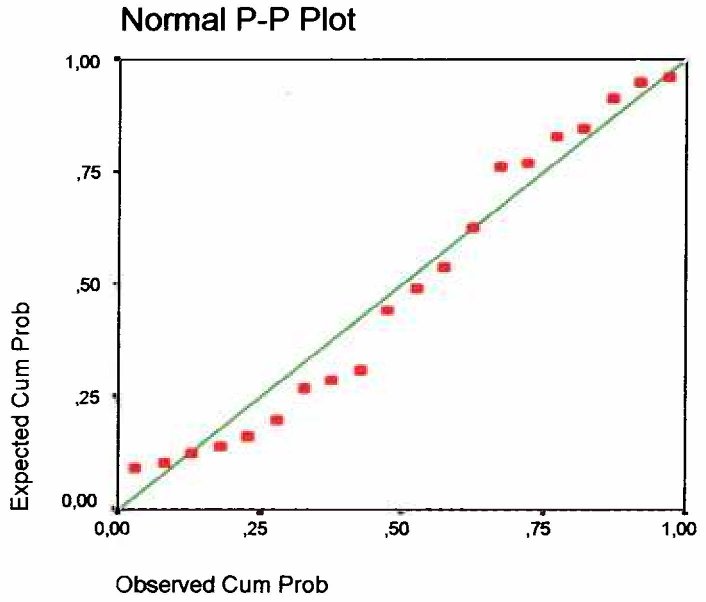
την αντίστοιχη κανονική απόκλιση. Η κάθετη κλίμακα είναι συνήθως χαρακτηρισμένη σε ποσοστά. Κατά συνέπεια 50% αντιστοιχούν στον οριζόντιο άξονα, 25% και 75% είναι στις αποστάσεις 0.6745 κάτω από και επάνω από αυτήν την γραμμή, 5% και 95% είναι στις αποστάσεις 1.9600) κάτω από και επάνω από αυτήν την γραμμή, και τόσο επάνω (όπως φαίνεται στο παραπάνω σχήμα). Ο Barnett (1976) έχει συζητήσει τις κατάλληλες θέσεις χάραξης πιθανότητας, ενώ ο Nelson (1976) έχει διαμορφώσει την κατασκευή του κανονικού εγγράφου πιθανότητας. Πρόσφατα ο Nelson διαβιβάζει μια σταθεροποιημένη κανονική τεχνική χάραξης πιθανότητας: Ο Rouncefield (1990), μεταξύ πολλών άλλων εξήγησε πώς κάποιος θα μπορούσε να χρησιμοποιήσει το κανονικό έγγραφο πιθανότητας για να αξιολογήσει την ισχύ της υπόθεσης της κανονικής κατανομής για ένα δείγμα.

Εάν X ακολουθεί την κανονική κατανομή και $P_r[X \leq x]$ σχεδιάζεται (ως τεταγμένη) ενάντια στο x (ως τετμημένη), τότε παίρνουμε μια ευθεία γραμμή. Η κλίση αυτής της γραμμής είναι σ^{-1} και η παρεμπόδισή της στον οριζόντιο άξονα είναι x=μ. Εάν οι παρατηρηθείσες συχνότητες των γεγονότων ($X \leq x$) χρησιμοποιούνται αντί των πραγματικών πιθανοτήτων, μια περίπου ευθεία- γραμμή διαγράμματος μπορεί να αναμένεται. Η ευθεία γραμμή που προσαρμόζεται σε αυτά τα παρατηρηθέντα σημεία δίνει τις εκτιμήσεις των σ και μ. Τέτοιες γραφικές μέθοδοι εκτίμησης μπορούν να δώσουν μία καλή πρακτική ακρίβεια.

Εάν η οριζόντια κλίμακα είναι λογαριθμική, έχουμε λογαριθμικό έγγραφο πιθανότητας. Το έγγραφο πιθανότητας της μισοκανονικής είναι απλά κανονικό έγγραφο πιθανότητας .Αυτό χρησιμοποιείται στις τεχνικές ανάλυσης διασποράς που αναπτύσσονται από τον Daniel (1959).

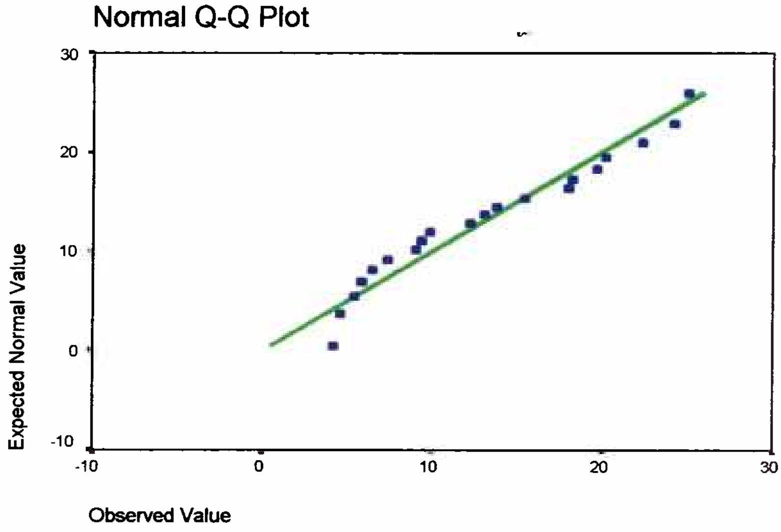
►Σήμερα αντί να χρησιμοποιούμε χαρτί πιθανοτήτων για να εξετάσουμε αν τα δεδομένα ενός δείγματος ακολουθούν κανονική κατανομή αυτό γίνεται στον υπολογιστή με την χρήση των στατιστικών πακέτων (SPSS, MINITAB). Ο βαθμός στον οποίο τα σημεία που προκύπτουν βρίσκονται σε μία ευθεία γραμμή καθορίζει το πόσο καλά μπορεί να προσεγγιστεί η κατανομή μας από μία κανονική κατανομή. Τα παρακάτω σχήματα έχουν γίνει με το SPSS για έλεγχο κανονικότητας σε ένα σετ δεδομένων.

α)Στον άξονα των τεταγμένων σχεδιάζεται η αναμενόμενη σχετική αθροιστική συχνότητα (επί %).



Διάγραμμα 4.3:*Έλεγχος κανονικότητας με διάγραμμα P-P*

β)Στον άξονα των τεταγμένων σχεδιάζονται οι αναμενόμενες συχνότητες των γεγονότων



Διάγραμμα 4.4:*Έλεγχος κανονικότητας με διάγραμμα Q-Q*

Με βάση τα παραπάνω σχήματα μπορούμε να συμπεράνουμε και με τους δύο τρόπους, ότι τα σημεία βρίσκονται κοντά στην ευθεία, και επομένως τα δεδομένα μας έχουν πράγματι κανονική κατανομή.

4.4 Προσεγγίζοντας την κανονική κατανομή από άλλες κατανομές

Ένας αριθμός συγγραφέων έχει επιδιώξει να προσαρμόσει άλλες κατανομές στην κανονική, και έπειτα να προσεγγίσει την κανονική αθροιστική συνάρτηση κατανομής και τα ποσοστιαία σημεία από αυτές. Οι επιτυχέστερες αυτών των προσπαθειών είναι του Burr και του Weibull που θα δούμε παρακάτω.

Σχετικά με την αντικατάστασης της όψης μιας κανονικής κατανομής από μια άλλη κατανομή σημειώνουμε ότι:

4.4.1 Η λογαριθμοκανονική κατανομή με συνάρτηση πυκνότητας πιθανότητας

$$f(x)=\begin{cases}\dfrac{1}{xb\sqrt{2\pi}}e^{-\frac{(\ln x-\ln a)^2}{2\beta^2}} & ,x>0 \quad a,b>0\\ 0 & ,x\leq 0\end{cases}$$

μπορεί να δώσει μια καλή αντιπροσώπευση της κανονικής κατανομής, όταν ο συντελεστής μεταβολής έχει μικρή απόλυτη τιμή (έστω, λιγότερο από 0.25).Επίσης εάν η τυχαία μεταβλητή X ακολουθεί την λογαριθμοκανονική κατανομή, τότε η τυχαία μεταβλητή $\ln X$ ακολουθεί την κανονική κατανομή.

4.4.2 Η ιδιαίτερη μορφή **λογιστικής κατανομής** είναι πολύ κοντά σε μια κανονική κατανομή

4.4.3 Ο Raab και ο Green (1961) έχουν προτείνει ότι η κατανομή με συνάρτηση πυκνότητας πιθανότητας

$$f(x)=(2\pi)^{-1}(1+\cos x),\quad -\pi<x<\pi \tag{4.29}$$

μπορεί να χρησιμοποιηθεί στο να αντικαταστήσει μια κανονική κατανομή. Η αντιστοιχία δεν είναι πολύ ακριβής (παρατηρούμε στον πίνακα 4.3 ότι η σύγκριση του τυποποιημένου εκατοστημορίου αποκλίνει στις δύο κατανομές) αλλά θα δώσει μερικές φορές χρήσιμα αναλυτικά αποτελέσματα. Η αντικατάσταση θα χρησιμοποιούταν, μόνο εάν η ουσιαστική απλοποίηση στην ανάλυση πραγματοποιούταν με αυτόν τον τρόπο.

Η αναμενόμενη τιμή και η τυπική απόκλιση μιας τυχαίας μεταβλητής με κατανομή την (4.29) είναι μηδέν και $\left(\frac{1}{3}\pi^2 - 2\right)^{\frac{1}{2}} = 1.14$ αντίστοιχα. Το τυποποιημένο εύρος της κατανομής $(-\pi,\pi)$ είναι έτσι από - 2.77 έως 2.77 τυπικές αποκλίσεις, και προφανώς η αντικατάσταση δίνει μία πτωχή προσαρμογή στις ουρές.

Πίνακας 4.3

Τυποποιημένη κατανομή σημείων εκατοστημορίου και η κανονική κατανομή

Αθροιστική Πιθανότητα	**Τυποποιημένες τιμές**	
	Κανονική κατανομή	**Κατανομή με συνάρτηση Πυκνότητας Πιθανότητας** $f(x) = (2\pi)^{-1}(1+\cos x), -\pi < x < \pi$
0.5	0.000	0.000
0.6	0.253	0.279
0.75	0.674	0.732
0.9	1.282	1.334
0.95	1.645	1.649
0.975	1.960	1.888
0.99	2.326	2.124
β_2	3.000	2.406

4.4.4 Ο Bell (1962) έχει περιγράψει ακόμα πιο απλές προσεγγίσεις, χρησιμοποιώντας τις τριγωνικές κατανομές. Επισήμανε ότι τέτοιες προσεγγίσεις μπορούν να θεωρηθούν ως δεύτερο στάδιο σε μία σειρά των προσεγγίσεων από τις κατανομές των μέσων των αυξανομένων αριθμών των ανεξάρτητων ορθογώνια κατανεμημένων μεταβλητών (η μέθοδος αυτή χρησιμοποιείται από τον Buslenko και λοιπούς (1966) για την κατασκευή "τυχαίων κανονικών αποκλίσεων").

4.4.5 Ο Steffensen (1937) έχει προτείνει τη χρήση της κατανομής ενός πολλαπλασίου μιας τυχαίας μεταβλητής X (π.χ. cXv), όπου v αρκετά μεγάλο Κάλεσε αυτό "ημικανονική" κατανομή.

4.4.6 Ο Chew (1968) περιλαμβάνει 2,4, και 5 σε έναν κατάλογο πέντε πιθανών αντικαταστάσεων για τις κανονικές κατανομές. Ο Chew έχει προσαρμόσει την ομοιόμορφη, την τριγωνική, την συνημιτονοειδή, την λογιστική, και την Laplace κατανομές, όλες συμμετρικές γύρω από το μηδέν, στην τυποποιημένη κανονική, επιλέγοντας τις τιμές παραμέτρου για να ταιριάξει με τη δεύτερη ροπή, όπου είναι δυνατόν. Κοντά σε μηδέν, καμία από αυτές τις προσαρμογές δεν είναι πολύ καλή. Ο Chew διεκδικεί ότι πέρα από περίπου μια τυπική απόκλιση οι συναρτήσεις πυκνότητας πιθανότητας αυτών των κατανομών είναι "αρκετά κοντά" μαζί με την Φ(x), αλλά τα σχετικά σφάλματα μπορούν να είναι αρκετά μεγάλα. Αυτές είναι πολύ ακατέργαστες προσεγγίσεις.

4.4.7 Ο Hoyt (1968) προσεγγίζει την συνάρτηση πυκνότητας πιθανότητας της N(0,1) με το άθροισμα S τριών αμοιβαία ανεξάρτητων τυχαίων μεταβλητών, όπου κάθε μία έχει μια ομοιόμορφη κατανομή στο (-1.1). Το S έχει συνάρτηση πυκνότητας πιθανότητας το g(s), που δίνεται από το

$$g(s)=\begin{Bmatrix} (3-s^2)/8\,, & |s|\le 1 \\ (3-|s|)^2/16\,, & 1\le |s|\le 3 \\ 0\,, & |s|\ge 3 \end{Bmatrix}$$

Το μέγιστο σφάλμα στη συνάρτηση πυκνότητας πιθανότητας είναι 0.024 όταν s = 0, και στην αθροιστική συνάρτηση κατανομής είναι 0.010 όταν s = 0.60.

4.4.8 Προσέγγιση από τις κατανομές Burr.

Εξετάζουμε την οικογένεια των κατανομών με αθροιστική συνάρτηση κατανομής G που δίνεται από την σχέση:

$$G(x) = 1 - \left(1 + x^{c}\right)^{-k} \qquad x \geq 0, c > 0, k > 0$$

O Burr (1967) προσεγγίζει την Φ(x) με τα $G_1(x)$ και $G_2(x)$,όπου

$$G_1(x) = \begin{cases} 1 - \left[1 + \{(0.6446.93) + (0.1619.84)x\}^{4.874}\right]^{-6.158} & ,x > -3.9799,80 \\ 0 & ,x < -3.9799,80 \end{cases}$$

$$G_2(x) = \frac{\{G_1(x) + 1 - G_1(-x)\}}{2}$$

Το $G_2(x)$ παρέχει την καλλίτερη προσέγγιση.

Η μεγαλύτερη διαφορά μεταξύ $G_2(x)$ και Φ(x) είναι 4.6×10^{-4} όπου x = ±0.6.Εάν εμείς λύσουμε την εξίσωση $G_2(y_p) = 1 - p$ για το y_p όταν p δίνεται, λαμβάνουμε μια προσέγγιση για το ποσοστιαίο σημείο z_p όπου $\Phi(z_p) = 1 - p$

O Burr το (1967) δίνει την προσέγγιση

$$z_p \cong y_p \cong \left[\left(p^{-\frac{1}{6.158}} - 1\right)^{\frac{1}{4.874}} - \left\{(1-p)^{-\frac{1}{6.158}} - 1\right\}^{\frac{1}{4.874}}\right] \Big/ 0.323968$$

Οπότε $0.008 < p < \frac{1}{2}$, $|error| < 3.5 \times 10^{-3}$,και όταν $0.045 < p < 0.50$, η προσέγγιση συγκρίνεται ευνοϊκά με αυτή η οποία είχε δοθεί από τον Hastings (1955).

4.4.9 Προσέγγιση από Weibull κατανομές.

Η μορφή της κατανομής Weibull με διαμόρφωση της παραμέτρου περίπου στο 3.25 όπως θα δούμε, είναι σχεδόν ίδια με την κανονική κατανομή μονάδων. Η οικογένεια Weibull έχει αθροιστική συνάρτηση κατανομής το W(X) που δίνεται από

$$W(x) = P_r(X \leq x) = 1 - \exp\left[-\{(x-a)/\theta\}^{m}\right], x > a, \theta > 0, m > 0$$

όπου a,θ, και m είναι παράμετροι θέσης, κλίμακας, και μορφής, αντίστοιχα. Εάν c≈3.60, η τρίτη κεντρική ροπή είναι μηδέν, και εάν c≈ 2.20, η κύρτωση είναι 3.0, όπως για μία κανονική συνάρτηση πυκνότητας πιθανότητας.

(α) O Plait (1962) διαπίστωσε ότι η κατανομή Weibull για την οποία m =3.25, a = 0, και θ=1 προσαρμόζεται σε $N(0.8963, 0.303^2)$ κατανομή ακριβώς.

(β)Με την εξίσωση διάφορων μέτρων ασυμμετρίας στο μηδέν, ο Dubey (1967) φθάνει σε τέσσερις προσεγγίσεις στο Φ(x), από το οποίο δίνουμε δύο που εμφανίζονται να προσαρμόζονται στο μέγιστο εύρος τιμών του x, στο διάστημα [- 3.3].

Οι προσεγγίσεις $W_1(X)$ και $W_2(X)$ που ακολουθούν παρήχθησαν με την τοποθέτηση των μέτρων (επικρατούσα τιμή -μέση τιμή) και της τρίτης κεντρικής ροπής μ_3, αντίστοιχα, ίσες με το μηδέν.

	m	a	θ	x:προνομιούχο εύρος
$W_1(x)$	3.25889	-2.96357	3.30589	[-1.5,-1.1],[0.2,0.9],[2.2,3.0]
$W_2(x)$	3.60232	-3.24311	3.59893	[-3.0,-1.6],[-1.0,0.1],[1.0,2.1]

Το επιθυμητό εύρος ελαχιστοποιεί το |σφάλμα|, αν και ο Dubey (1967) βρίσκει μερικές τιμές του x για τις οποίες οι άλλες προσεγγίσεις του είναι ελαφρώς καλύτερες. Ο συνδυασμός των ανωτέρω, εντούτοις, έχει έναν μέγιστο |σφάλμα| του 0.0078 όταν x = -0,9.

(γ) O Macino (1984) κατατάσσει αυτά και σε άλλες αναφορές. Αυτός παρατηρεί όταν m=3,43938,η Weibull συνεισφέρει μία ενδιαφέρουσα ιδιότητα μαζί με την κανονική. Αυτή λέει ότι ο μέσος και η διάμεσος είναι ίσες. Επιπλέον όταν η τιμή του m είναι σχεδόν ίδια με τιμή του μέσου της διακινδυνευμένης αναλογίας (mean hazard rates –m.h.r's),των δύο κατανομών, τότε αυτές έχουν ίσες διακυμάνσεις. Για συνάρτηση πυκνότητας πιθανότητας g(.) και αθροιστική συνάρτηση κατανομής G(.) το m.h.r ορίζεται ως εξής :

$$\int_{-\infty}^{\infty} \frac{g(x)}{1-G(x)} \cdot g(x) dx$$

ΚΕΦΑΛΑΙΟ 5^{Ο}

Χαρακτηρισμοί

Τα τελευταία χρόνια έχει γίνει μεγάλη συζήτηση για τα χαρακτηριστικά των κατανομών, γενικά και ειδικότερα για τα χαρακτηριστικά των κανονικών κατανομών.Στην πραγματικότητα ως ένα ορισμένο βαθμό, τα χαρακτηριστικά των κανονικών κατανομών έχουν γίνει ένας κλάδος των μαθηματικών, με έμφαση στις συναρτησιακές εξισώσεις και στα χαρακτηριστικά των συναρτήσεων, αλλά με περιορισμένο αντίκτυπο στην Εφαρμοσμένη Στατιστική.

Εάν μια κανονικά κατανεμημένη μεταβλητή ή εάν ένα τυχαίο δείγμα από μια κανονική κατανομή έχει κάποια ιδιότητα P, είναι ενδιαφέρον να γνωρίζουμε εάν μια τέτοια ιδιότητα ορίζει μονοσήμαντα τον κανονικό νόμο. Κατά συνέπεια ως χαρακτηριστικές ιδιότητες της κανονικής κατανομής θεωρούμε τις ικανές και αναγκαίες συνθήκες οι οποίες όταν ισχύουν προκύπτει , ότι η κατανομή της τυχαίας μεταβλητής X είναι η κανονική κατανομή.

Σοβαρή έρευνα σε αυτήν την περιοχή έχει γίνει από το 1935, αλλά τα περισσότερα αποτελέσματα χρονολογούνται από το 1950 ή αργότερα.

Είμαστε ανίκανοι να παρέχουμε ένα περιεκτικό, ή ακόμα και πλήρως αντιπροσωπευτικό, απολογισμό των χαρακτηρισμών της κανονικής κατανομής σε αυτή την εργασία λόγω του περιορισμένου χώρου. Πολλά από τα πιο πρόσφατα αποτελέσματα δεν έχουν μεγάλη εφαρμογή στην πράξη. Η εκμετάλλευση των πιο παλαιών αποτελεσμάτων δεν έχει γίνει ακόμη.

Αν κάποιος επιθυμεί μπορεί να βρει μερικές από τις ειδικές περιπτώσεις πρακτικότερου ενδιαφέροντος από τα πιο γενικά θεωρήματα. Μερικοί από τους πιο μαθηματικούς και θεωρητικούς χαρακτηρισμούς έχουν παραλειφθεί, αλλά μπορούν να βρεθούν στα βιβλία και τα επεξηγηματικά έγγραφα για το θέμα. Αυτοί περιέχονται σους Stuart και Ord (το 1987), Kotz (1974) Lukacs (1956) Lukacs και Laha (1964) Mathai και Pederzoli (1977) και Kagan και λοιποί. (1973). Η αναθεώρηση του αναφερομένου στο τέλος από τον Diaconis και τους λοιπούς (1977) δίνει μια άριστη περίληψη και μια έκθεση του τομέα.

Ενώ οι χαρακτηρισμοί της εκθετικής κατανομής περιλαμβάνουν κατά ένα μεγάλο μέρος σειρά στατιστικών, εκείνοι του κανονικού νόμου συχνά χαρακτηρίζονται κατανομές ή ιδιότητες ανεξαρτησίας γραμμικών και τετραγωνικών μορφών. Οι γραμμικές μορφές εξετάζονται μόνες τους στο [5.1], και επίσης μαζί με τις τετραγωνικές μορφές στο [5.2], αυτές είναι βασισμένες στα σύνολα ανεξάρτητων και μερικές φορές ανεξάρτητων και ισόνομα κατανεμημένων τυχαίων μεταβλητών.

Οι χαρακτηρισμοί που είναι βασισμένοι στις υπό συνθήκη κατανομές και στις ιδιότητες παλινδρόμησης είναι επίσης κυρίως βασισμένοι στις γραμμικές και τετραγωνικές μορφές αυτοί δίνονται στο [5.3].Μπορούμε να σημειώσουμε ότι τα στατιστικά σε όλους αυτούς τους χαρακτηρισμούς είναι αμετάβλητα γραμμικά. Τα ειδικά θέματα καλύπτονται στο [5.4] (χαρακτηριστικές συναρτήσεις),[5.6](πολικοί ισότιμοι μετασχηματισμοί) Παραλείπονται οι χαρακτηρισμοί που είναι βασισμένοι στην επάρκεια, τις πληροφορίες, και τις έννοιες από τη θεωρία απόφασης και τη στατιστική συμπερασματολογία), και την Διάταξη στατιστικών. Ακόμη παραλείπεται το θέμα για το πώς κοντά σε μια χαρακτηριστική ιδιότητα μια κατανομή μπορεί να είναι σχεδόν κανονική (αλλά δείτε επίσης Diaconis και λοιποί. 1977).

5.1 Χαρακτηρισμοί από γραμμικά στατιστικά

5.1.1 (α) Έστω X_1 και X_2 ανεξάρτητες,όχι εκφυλισμένες τυχαίες μεταβλητές, όχι απαραιτήτως όμοια κατανεμημένες. Έστω επίσης, $Y= X_1 + X_2$.
Τότε Y κατανέμεται κανονικά εάν και μόνο εάν X_1 και X_2 κατανέμονται κανονικά (Mathai και Pederzoli 1977, Cramer, 1936).

(β) Έστω X και Y είναι τυχαίες μεταβλητές. Εάν οι συντελεστές a και b υπάρχουν έτσι ώστε $a^2+b^2>0$ και εάν X και $aX+bY$ είναι ανεξάρτητες σε κάθε ισότιμο σύστημα. Τότε X και Y είναι ανεξάρτητες κανονικές τυχαίες μεταβλητές με την ίδια διασπορά (Berk, 1986).

Εδώ, "σε κάθε ισότιμο σύστημα" σημαίνει "κάτω από κάθε ορθογώνια περιστροφή των αξόνων".

Ο Berk επεκτείνει αυτό το αποτέλεσμα:

Έστω $X_1, X_2,\ldots,Xn$ είναι τυχαίες μεταβλητές και $(a_1,a_2,\ldots,a_n)$,$(b_1,b_2,\ldots,b_n)$ διαφορετικά από το μηδέν διανύσματα των συντελεστών. Οπότε εάν $\sum_i a_i X_i$ και $\sum_i b_i X_i$ είναι ανεξάρτητα σε κάθε ισότιμο σύστημα στο Ευκλείδιο ν-χώρο, τότε είτε τα X_i είναι εκφυλισμένα είτε $\sum_i a_i X_i=0$ και $X_1, X_2,\ldots,Xn$ είναι αμοιβαία ανεξάρτητες N(.,σ^2) τυχαίες μεταβλητές με κοινή διασπορά. Δείτε επίσης Hartman και Wintner (1940). Αυτό το αποτέλεσμα για n = 3 λήφθηκε από τον James Clerk Maxwell.

(γ) Έστω το άθροισμα $X_1 + X_2 + \ldots + Xn$ των n ανεξάρτητων τυχαίων μεταβλητών (n = 2,3,...). Αυτό κατανέμεται κανονικά εάν και μόνο εάν κάθε συνιστώσα μεταβλητή $X_1, X_2,\ldots,Xn$ η ίδια κατανέμεται κανονικά (Cramer, 1946).

(δ) Έστω $X_1, X_2,\ldots,Xn$ είναι ένα τυχαίο δείγμα από έναν ορισμένο πληθυσμό, και έστω $L=a_1X_1+a_2X_2+\ldots+a_nX_n$ ένα γραμμικό στατιστικό, όπου $a_1,a_2,\ldots,a_n$ είναι σταθερές διάφορες του μηδενός. Τότε ο πληθυσμός είναι κανονικός εάν και μόνο εάν το L είναι κανονικά κατανεμημένο (Lukacs, 1956).

(ε) Έστω $X_1, X_2,\ldots,Xn$ είναι ένα τυχαίο δείγμα από έναν ορισμένο πληθυσμό και έστω οι σταθερές $a_1,a_2,\ldots,a_n$ που δίνονται, έτσι ώστε $A_s=\sum_{j=1}^{n} a_j^s \neq 0$ για όλες τις θετικές ακέραιες τιμές του s. Τότε το στατιστικό $a_1X_1+a_2X_2+\ldots+a_nX_n$ κατανέμεται κανονικά με μέσο μ και διασπορά σ^2 αν και μόνο αν κάθε ένα από τα $X_1, X_2,\ldots,Xn$ κατανέμεται κανονικά με μέσο $\frac{\mu}{A_1}$ και διασπορά $\frac{\sigma^2}{A_2}$ (Lukacs, 1956).

5.1.2 Έστω $X_1, X_2,\ldots,X_n$ είναι ένα τυχαίο δείγμα από έναν ορισμένο πληθυσμό και υποθέτουμε ότι όλες οι απόλυτες ροπές $E(|X_i|^k)$ υπάρχουν ($i=1,\ldots,n$ και $k=1,2,\ldots$).Θεωρούμε τα γραμμικά στατιστικά

$$L_1 = a_1X_1 + a_2X_2 + \ldots + a_nX_n \text{ και } L_2 = b_1X_1 + b_2X_2 + \ldots + b_nX_n$$

Υποθέτουμε ότι $\{|a_1|,\ldots,|a_n|\}$ δεν είναι ένας συνδυασμός των $\{|b_1|,\ldots,|b_n|\}$ και ότι :

$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i$, $\sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} b_i^2$ Τότε ο αρχικός πληθυσμός κατανέμεται κανονικά αν και μόνο αν L_1 και L_2 κατανέμονται πανομοιότυπα (Lukacs 1956). Ο Marcinkiewicz (1939) δίνει ένα αποτέλεσμα που γενικεύει αυτό στα άπειρα αθροίσματα.

5.1.3 Έστω $X_1, X_2,\ldots,X_n$ είναι ένα τυχαίο δείγμα από έναν ορισμένο πληθυσμό. Τότε κάθε γραμμικό στατιστικό $a_1X_1 + a_2X_2 + \ldots + a_nX_n$ κατανέμεται όπως η τυχαία μεταβλητή $(a_1^2 + \ldots + a_n^2)^{\frac{1}{2}} X_1$ εάν και μόνο εάν ο αρχικός πληθυσμός είναι κανονικός. Αυτό είναι μια ειδική περίπτωση ενός θεωρήματος του Lukacs (1956). Ο Shimizu (1962) έχει ένα παρόμοιο αποτέλεσμα, που απαιτεί την υπόθεση των πεπερασμένων διασπορών.

5.1.4 (α) Έστω X_1 και X_2 ανεξάρτητες τυχαίες μεταβλητές με πεπερασμένες διασπορές. Τότε το άθροισμα $X_1 + X_2$ και η διαφορά $X_1 - X_2$ είναι ανεξάρτητες εάν και μόνο εάν X_1 και X_2 κατανέμονται κανονικά (Bernstein,1941).

(β) Έστω $X_1, X_2,\ldots,X_n$ είναι ένα δείγμα από έναν ορισμένο πληθυσμό και επί πλέον τα γραμμικά στατιστικά

$$L_1 = a_1X_1 + a_2X_2 + \ldots + a_nX_n \text{ και } L_2 = b_1X_1 + b_2X_2 + \ldots + b_nX_n$$

έτσι ώστε : $\sum_{j=1}^{n} a_j b_j = 0$, ενώ $\sum_{j=1}^{n} (a_j b_j)^2 \neq 0$ Τότε ο αρχικός πληθυσμός είναι κανονικός εάν και μόνο εάν L_1 και L_2 κατανέμονται ανεξάρτητα (Lukacs, 1956). Αυτό το αποτέλεσμα δεν απαιτεί οποιαδήποτε υπόθεση της ύπαρξης ροπών.

(γ) Το θεώρημα του Darmois- Skitovich:

Έστω $X_1,\ldots,X_n$ είναι αμοιβαία ανεξάρτητες τυχαίες μεταβλητές (όχι απαραίτητα όμοια κατανεμημένες). Έστω ότι L_1 και L_2 είναι γραμμικά στατιστικά $\sum_{i=1}^{n} a_i X_i$ και $\sum_{i=1}^{v} b_i X_i$ όπου $a_1, a_2,\ldots,a_n, b_1, b_2,\ldots,b_n$ είναι σταθεροί συντελεστές.

Εάν L_1 και L_2 είναι ανεξάρτητα, τότε οι τυχαίες μεταβλητές X_i για τις οποίες το γινόμενο $a_i b_i \neq 0$ είναι όλες κανονικά κατανεμημένες (Kagan και λοιποί. 1973, Mathai και Pederzoli, 1977, Darmois, 1951, Skitovitch, 1953.

Η μορφή του αντιστρόφου δίνεται ως εξής: Εάν $\sum_{j=1}^{n} \sigma_j^2 a_j b_j = 0$ [όπου $\mathrm{var}(X_j) = \sigma_j^2$] και αυτά τα X_i για τα οποία $a_i b_i \neq 0$ κατανέμονται κανονικά, τότε L_1 και L_2 είναι ανεξάρτητα.

5.1.5 Ένας ορθογώνιος μετασχηματισμός των n ανεξάρτητων ισόνομα και κανονικά κατανεμημένων τυχαίων μεταβλητών διατηρεί αμοιβαία ανεξαρτησία και κανονικότητα. Ο Lancaster (1954) έδειξε ότι αυτή η ιδιότητα χαρακτηρίζει την κανονική κατανομή.

Έστω $X_1, X_2,\ldots,X_n$ αμοιβαία ανεξάρτητες τυχαίες μεταβλητές που κάθε μία έχει μέσο μηδέν και διασπορά μονάδα και έστω ο σημαντικός γραμμικός μετασχηματισμός :

$$Y_j = \sum_{i=1}^{n} a_{ij} X_j$$

μπορεί να κάνει, εκείνα τα οποία αποκλείουν του τύπου $X_i + a = cY_j$, και έτσι ώστε τα $Y_1, Y_2,\ldots,Y_n$ να είναι αμοιβαία ανεξάρτητα.

Τότε $X_1, X_2,\ldots,X_n$ κατανέμονται κανονικά και ο μετασχηματισμός είναι ορθογώνιος. Οι Stuart και Ord (1987) δηλώνουν ότι αυτός ο χαρακτηρισμός της κανονικής κατανομής "διασαφηνίζει τη θέση της κύριας σημασίας που κρατάει στη στατιστική θεωρία".

5.1.6 Στο ακόλουθο αποτέλεσμα εξετάζουμε ένα άπειρο σύνολο μεταβλητών (Kagan και λοιποί 1973, Mathai και Pederzoli 1977).

Έστω $X_1, X_2,\ldots$ μια ακολουθία ανεξάρτητων τυχαίων μεταβλητών και $\{a_j\},\{b_j\}$ είναι δύο ακολουθίες πραγματικών σταθερών έτσι ώστε :

(ι) Οι ακολουθίες $\left\{\frac{a_j}{b_j} : a_j b_j \neq 0\right\}$ και $\left\{\frac{b_j}{a_j} : a_j b_j \neq 0\right\}$ είναι και οι δύο οριακές.

(ii) $\sum_{j=1}^{n} a_j X_j$ και $\sum_{j=1}^{v} b_j X_j$ συγκλίνουν με κάποια πιθανότητα στις τυχαίες μεταβλητές U και V, αντίστοιχα.

(iii) Uκαι Vείναι ανεξάρτητες.

Τότε για κάθε j έτσι ώστε $a_j b_j \neq 0$, η X_j είναι κανονικά κατανεμημένη.

5.2 Γραμμικοί και τετραγωνικοί χαρακτηρισμοί

5.2.1 (α) Έστω $X_1, X_2, \ldots, X_n$ ένα τυχαίο δείγμα από έναν ορισμένο πληθυσμό, $\overline{X} = \frac{\sum X_i}{n}$ ο μέσος του δείγματος, και $S^2 = \frac{\sum (X_i - \overline{X})^2}{(n-1)}$ η διασπορά του δείγματος. Για $n \geq 2$, μία αναγκαία και ικανή συνθήκη για την ανεξαρτησία των $\overline{X}$ και S^2 είναι ότι ο αρχικός πληθυσμός είναι κανονικός (Kagan και λοιποί. 1973, Stuart και Ord, 1987, Lukacs 1956). Αυτό το αποτέλεσμα επιτεύχθηκε αρχικά από τον Geary (1936) και από τον Lukacs (1942) υπό πιο περιοριστικούς όρους που περιλαμβάνουν την ύπαρξη των ροπών.

(β) Έστω $X_1, X_2, \ldots, X_n$ ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές με κοινό πεπερασμένο μέσο όρο και διασπορά. Έστω η πρώτη σειρά διαδοχικών διαφορών μέσων στο τετράγωνο που καθορίζονται από

$$D_k^2 = \frac{1}{2}(n-k)^{-1} \sum_{i=1}^{n-k} (X_{i+k} - X_i)^2 \quad , \quad k = 1,2,\ldots,n-1$$

και έστω $\overline{X} = \frac{\sum X_i}{n}$. Τότε για κάθε k μία αναγκαία και ικανή συνθήκη για το $\overline{X}$ και D_k^2 για να είναι ανεξάρτητα είναι ότι ο αρχικός πληθυσμός είναι κανονικός (Geisser, 1956).

(γ) Το ακόλουθο θεώρημα από τον Rao (1958) γενικεύει τα αποτελέσματα στα (α) και (β) ανωτέρω. Έστω $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές τυχαίες μεταβλητές με κοινούς πεπερασμένο μέσο όρο και διασπορά, και έστω:

$$\delta^2 = \left(\sum_{t=1}^{m} \sum_{j=1}^{n} h_{tj}^2 \right)^{-1} \sum_{t=1}^{m} (h_{t1} X_1 + \ldots + h_{tn} X_n)^2 \quad , \ \mathrm{m} \geq 1$$

όπου $\sum_{j=1}^{n} h_{tj} = 0$ όταν t=1,2,…,m. Τότε μία αναγκαία και επαρκής συνθήκη για τον αρχικό πληθυσμό για να είναι κανονικός είναι ότι τα $\overline{X}$ και δ^2 είναι ανεξάρτητα. Δείτε επίσης Laha (1953).

Στο (α), $m = n$ και $\mathrm{h}_{\mathrm{tj}} = 1 - n^{-1}$, $\mathrm{h}_{\mathrm{tj}} = -n^{-1}$ $\mathrm{t} \neq \mathrm{j}$

Στο (β), $m = n - k$ και $\mathrm{h}_{\mathrm{jj}} = -1$, $\mathrm{h}_{\mathrm{tj}} = 1$ εάν $\mathrm{j} = \mathrm{t} + \mathrm{k}$ και $\mathrm{h}_{\mathrm{tj}} = 0$

Σε διαφορετική περίπτωση:

(δ) Έστω $X_1, X_2, \ldots, Xn$ ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές ($n \geq 4$), και έστω $Y = \dfrac{(X_1 - X_2)}{S}$. Εάν Y είναι στοχαστικά ανεξάρτητη από το ζευγάρι $(\overline{X}, S^2)$, όπου $\overline{X} = \dfrac{\sum X_i}{n}$ και $S^2 = \dfrac{\sum (X_i - \overline{X})^2}{(n-1)}$, τότε ο αρχικός πληθυσμός είναι κανονικός (Kelkar και Matthes, 1970).

(ε) Ο Kaplansky (1943) αποδυνάμωσε την συνθήκη για το χαρακτηρισμό στο (α). Έστω $X_1, X_2, \ldots, Xn$ ένα τυχαίο δείγμα από συνεχή πληθυσμό με δειγματικό μέσο $\overline{X}$ και δειγματική διασπορά S^2. Τότε η κοινή συνάρτηση πυκνότητας πιθανότητας των $X_1, X_2, \ldots, Xn$ είναι μια συνάρτηση $h(\mathrm{\bar{x}}, \mathrm{s})$ των τιμών του $\overline{X}$ και του S και επίσης η κοινή συνάρτηση πυκνότητας πιθανότητας του $\overline{X}$ και του S δίνεται από $h(\mathrm{\bar{x}}, \mathrm{s}) s^{n-2}$ εάν και μόνο εάν ο αρχικός πληθυσμός είναι κανονικός ($n \geq 3$).

5.2. 2 (α) Έστω $X_1, X_2, \ldots, Xn$ τυχαίο δείγμα από έναν ορισμένο πληθυσμό. Έστω $L = \sum_{i=1}^{n} a_i X_i$ και $Q = \sum_{i=1}^{n} X_i^2 - L^2$, όπου $\sum_{i=1}^{n} a_i^2 = 1$. Τότε μία αναγκαία και επαρκής συνθήκη για την ανεξαρτησία των L και Q είναι ότι ο αρχικός πληθυσμός είναι κανονικός (Kagan και λοιποί 1973, Mathai και Pederzoli 1977).

(β) Έστω $X_1, X_2,\ldots,X_n$ είναι τυχαίο δείγμα από έναν πληθυσμό όπου η υπάρχει η δεύτερη ροπή και έστω $\overline{X} = \frac{\sum X_i}{n}$, $Q = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}X_iX_j$ όπου

$$\sum_{i=1}^{n} a_{ii} \neq 0 \text{ και } \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} = 0$$

Τότε $\overline{X}$ και Q είναι ανεξάρτητα εάν και μόνο εάν ο αρχικός πληθυσμός είναι κανονικός (Mathai και Pederzoli, 1977, Kagan και λοιποί 1973). Δείτε επίσης Lukacs και Laha (1964).

(γ) Έστω $X_1, X_2,\ldots,X_n$ ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές, και έστω $Q = \sum_{j=1}^{n}\sum_{k=1}^{n} a_{ij}X_jX_k + \sum_{j=1}^{n} b_jX_j$ όπου $\sum_{j=1}^{n} a_{jj} \neq 0$ και $\sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk} = 0$. Τότε τα $X_1, X_2,\ldots,X_n$ και το Q είναι ανεξάρτητα εάν και μόνο εάν

(i) ο αρχικός πληθυσμός είναι κανονικός,

(ii) $\sum_{k=1}^{n} a_{jk} = 0$ όταν k=1,...,n

(iii) $\sum_{j=1}^{n} b_j = 0$ (Lukacs και Laha, 1964).

5.2.3 Βασισμένη σε ένα τυποποιημένο κανονικό δείγμα, η κατανομή του $\sum_{i=1}^{n}(X_i + a_i)^2$ εξαρτάται από τα $a_1, a_2,\ldots,a_n$, ωστόσο η μη κεντρική X^2 δείχνει μη κεντρικότητα των παραμέτρων $a_1^2 + \ldots + a_n^2$. Αυτή η ιδιότητα είναι ένας χαρακτηρισμός του κανονικού νόμου.

(α) Έστω $X_1, X_2,\ldots,X_n$ είναι ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές ($n \geq 2$), και η κατανομή του στατιστικού $\sum_{i=1}^{n}(X_i + a_i)^2$ εξαρτάται από τα $a_1, a_2,\ldots,a_n$ μόνο μέσω του $\sum_{i=1}^{n} a_i^2$ όπου $a_1, a_2,\ldots,a_n$ είναι πραγματικοί, τότε ο αρχικός πληθυσμός είναι κανονικός (Kagan και λοιποί 1973).

(β)Εάν το τυχαίο διάνυσμα $(X_1, X_2,\ldots,X_m)$ είναι ανεξάρτητο από το τυχαίο διάνυσμα $(X_{m+1},\ldots,X_n)$, $1 \leq m < n$, και εάν για τις αυθαίρετες σταθερές $\sum_{i=1}^{n}(X_i + a_i)^2$ η κατανομή του $\sum_{j=1}^{n}(X_j + a_j)^2$ εξαρτάται από τα $a_1, a_2,\ldots,a_n$ μόνο μέσω του $\sum_{j=1}^{n} a_j^2$, τότε τα $X_1, X_2, \ldots, X_n$ είναι αμοιβαία ανεξάρτητα και κανονικά κατανεμημένα με μέσο μηδέν και κοινή διακύμανση. (Kotz, 1974).

5.2.4 Τα ακόλουθα σχετικά αποτελέσματα οφείλονται στο Geisser (1973).

(α) Έστω X και Y ανεξάρτητες τυχαίες μεταβλητές, και έστω η X ακολουθεί μία N(0,1) κατανομή. Τότε η τυχαία μεταβλητή $\frac{(aX+bY)^2}{(a^2+b^2)}$ ακολουθεί μια X_1^2 κατανομή με έναν βαθμό ελευθερίας για κάποια διαφορετικά από το μηδέν α και β εάν και μόνο αν Y^2 ακολουθεί την κατανομή X_1^2.

(β) Έστω X και Y ανεξάρτητα όπου κάθε ένα τα X^2 και Y^2 ακολουθεί μια X_1^2 κατανομή. Τότε η τυχαία μεταβλητή $\frac{(aX+bY)^2}{(a^2+b^2)}$ ακολουθεί μια X_1^2 κατανομή, για κάποια διαφορετικά από το μηδέν α και β, εάν και μόνο εάν τουλάχιστον ένα από το X και Y ακολουθεί μια N(0,1) κατανομή.

(γ) Έστω X και Y ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές. Τότε X και Y είναι N(0,1) τυχαίες μεταβλητές εάν και μόνο εάν, για κάποια διαφορετικά από το μηδέν α και β, οι τυχαίες μεταβλητές $\frac{(aX+bY)^2}{(a^2+b^2)}$ και $\frac{(aX-bY)^2}{(a^2+b^2)}$ ακολουθούν X_1^2 κατανομές.

5.2.5 Η κατανομή της δειγματικής διασποράς εξασφαλίζει ένα χαρακτηρισμό της κανονικότητας.

(α) Έστω $X_1, X_2, \ldots, X_n$ ($n \geq 2$)ένα τυχαίο δείγμα από μία μη εκφυλισμένη συμμετρική κατανομή με πεπερασμένη διασπορά σ^2 και δειγματικό μέσο $\overline{X}$.

Τότε το $\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}$ κατανέμεται ως X^2 με (n-1) βαθμούς ελευθερίας X_{n-1}^2 εάν και μόνο εάν η κατανομή του αρχικού πληθυσμού είναι κανονική (Ruben, 1974).

(β) Έστω $X_1, X_2, \ldots$ μια ακολουθία από ανεξάρτητες και ισόνομα κατανεμημένες μη εκφυλισμένες τυχαίες μεταβλητές με πεπερασμένη διασπορά σ^2 και έστω m και n διακριτοί ακέραιοι αριθμοί όχι μικρότεροι από το 2. Σημειώνουμε $\sum_{i=1}^{m}\frac{X_i}{m}$ με $\overline{X}_m$ και $\sum_{i=1}^{n}\frac{X_i}{n}$ με $\overline{X}_n$.

Τότε $\frac{\sum_{i=1}^{m}(X_i-\overline{X}_m)^2}{\sigma^2}$ και $\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}$ κατανέμονται όπως η X_{m-1}^2 και X_{n-1}^2 αντίστοιχα, εάν και μόνο εάν η αρχική κατανομή είναι κανονική (Ruben, 1975, Bondesson, 1977).

(γ) Σε έναν άλλο χαρακτηρισμό όταν $n \geq 2$, τα X_i ορίζονται όπως στο (α) είναι κατανεμημένα ως N(0,1) εάν και μόνο εάν $\sum_{i=1}^{n}X_i^2$ και $n\overline{X}_n^2$ κατανέμονται όπως X_n^2 και X_1^2 αντίστοιχα (Ahsanullah, 1987).

(δ) Εξετάζουμε ένα διάνυσμα παρατηρήσεων μετά από τις συνηθισμένες υποθέσεις του γενικού γραμμικού μοντέλου (Rao, 1973,), έχοντας ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές και συμμετρικά κατανεμημένες συνιστώσες σφάλματος. Έστω σ^2 είναι η διασπορά του σφάλματος και ν οι βαθμοί ελευθερίας του αθροίσματος τού υπόλοιπου στο τετράγωνο (RSS). Τότε η κατανομή των $\frac{RSS}{\sigma^2}$ ακολουθεί κατανομή X_v^2 εάν και μόνο εάν οι συνιστώσες του σφάλματος κατανέμονται κανονικά. Ο Ruben (1976,) δίνει μια επίσημη παρουσίαση και μια απόδειξη.

5.3 Χαρακτηρισμοί από τις υπό συνθήκη κατανομές και τις ιδιότητες Παλινδρόμησης

5.3.1 (α) Έστω X και Y τυχαίες μεταβλητές, κάθε μία με μία διαφορετική από το μηδέν συνάρτηση πυκνότητας πιθανότητας στο μηδέν. Εάν η υπό συνθήκη κατανομή του X, που δίνεται από X +Y = x+ y, είναι κανονική κατανομή με μέσο $\frac{x+y}{2}$ για όλα τα x και y, τότε X και Y είναι ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές (Patil και Seshadri, 1964,).

(β) Έστω X και Y είναι ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές με πυκνότητα f. Αν οι η υπό συνθήκη συναρτήσεις πυκνότητας πιθανότητας των X+Y, δοθέντος X-Y=t, υπάρχουν και είναι ίσες για όλα τα t σε ένα σύνολο Borel E που έχει θετικό μέτρο Lebesgue. Τότε η f είναι μια κανονική πυκνότητα (Ramasubramaniam, 1985).

5.3.2 Έστω X_1 και X_2 ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές με μέσο μηδέν. Εάν $E(X_1 - aX_2 \mid X_1 - \beta X_2)=0$ και $E(X_1 + \beta X_2 \mid X_1 - \alpha X_2)=0$, όπου $a \neq 0, \beta \neq 0$ τότε X_1 και X_2 είναι κανονικές (ενδεχομένως εκφυλισμένες) όταν βα=1, και εκφυλίζονται σε οποιαδήποτε άλλη περίπτωση (Kagan και λοιποί, 1973).

5.3.3 Έστω $X_1, X_2, \ldots, X_k$ ανεξάρτητες τυχαίες μεταβλητές και $a_j, b_j (j = 1,\ldots,k)$ σταθερές διάφορες του μηδενός έτσι ώστε, όταν $i \neq j, a_i b_i^{-1} + a_j b_j^{-1} \neq 0$. Εάν η υπό συνθήκη κατανομή των $\sum_{j=1}^{n} a_j X_j$, δοθέντος $\sum_{j=1}^{k} b_j X_j$, είναι συμμετρική, τότε $X_1, X_2, \ldots, X_k$ είναι κανονικές τυχαίες μεταβλητές (ενδεχομένως εκφυλισμένες). Εάν η χαρακτηριστική συνάρτηση **των** X_i είναι $\exp(itA_j - B_j t^2)$, με A_j πραγματικό και B_j μη αρνητικό, τότε $\sum_{j=1}^{k} A_j a_j = 0$ και $\sum_{j=1}^{k} B_j a_j b_j = 0$ (Kagan και λοιποί 1973)

5.3.4 (α) Το θεώρημα Kagan –Linnik-Rao (Kagan και λοιποί. 1973).

Έστω $X_1, X_2, \ldots, X_n$ ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές $(n \geq 3)$ με μέσο μηδέν, και έστω $\overline{X} = \frac{\sum X_i}{n}$.

Τότε εάν $E(\overline{X} \mid X_1 - \overline{X}, \ldots, X_n - \overline{X}) = 0$, ο αρχικός πληθυσμός είναι κανονικός. Μία εναλλακτική συνθήκη, που δίνεται από τον Kagan και λοιπούς (1965), είναι ότι $E(\overline{X} \mid X_2 - X_1, X_3 - X_1, \ldots, X_n - X_1) = 0$. Εάν n= 2, αυτή η ιδιότητα ισχύει για οποιαδήποτε αρχική συμμετρική κατανομή.

(β) Έστω $X_1, X_2, \ldots, X_n$ $(n \geq 3)$ αμοιβαία ανεξάρτητες τυχαίες μεταβλητές, κάθε μία με μέσο μηδέν, και έστω $L_1, \ldots, L_n$ γραμμικά ανεξάρτητες γραμμικές συναρτήσεις των $X_1, X_2, \ldots, X_n$, με όλους τους συντελεστές στο L_1 διαφορετικούς από το μηδέν. Τότε, εάν $E(L_1 \mid L_2, \ldots, L_n) = 0$, οι τυχαίες μεταβλητές $X_1, \ldots, X_n$ είναι όλες κανονικά κατανεμημένες. (Kagan και λοιποί 1973). Αυτή είναι μια γενίκευση του αποτελέσματος στο (α).

(γ) Έστω $X_1, X_2, \ldots, X_n$ ανεξάρτητες και ισόνομα κατανεμημένες μη εκφυλισμένες τυχαίες μεταβλητές, και $L_1 = \sum_{j=1}^{n} a_{ij} X_j$ (i=1,...,n-1) είναι γραμμικά ανεξάρτητες γραμμικές συναρτήσεις των $X_1, X_2, \ldots, X_n$ ενώ $L_n \sum_j a_{nj} X_j s$ είναι μια γραμμική μορφή έτσι ώστε το διάνυσμα $(a_{ni}, \ldots, a_{nn})$ δεν είναι πολλαπλάσιο οποιουδήποτε διανύσματος με συνιστώσες 0,1,-1. Τότε από τις συνθήκες $\mathrm{E}(L_i \mid L_n) = 0 \quad (i = 1, \ldots, n-1)$, συμπεραίνουμε είναι ότι ο αρχικός πληθυσμός είναι κανονικός εάν και μόνο εάν $\sum_{j=1}^{n} a_{ij} a_{nj} = 0 \quad (i = 1, \ldots, n-1)$ (Kagan και λοιποί 1973).

(δ) Cacoullos (1967b,) δίνει ένα αποτέλεσμα ανάλογο με αυτό στο (γ)
Έστω $X_1, X_2, \ldots, X_n$ ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές με μέσο μηδέν και θετική πεπερασμένη διασπορά. Οι μορφές $L_1, \ldots, L_{n-1}$ είναι γραμμικά ανεξάρτητα στατιστικά που ορίζονται στο (γ). Ορίζουμε $Ln = b_1X_1 + \ldots + bnXn$ (χωρίς τους περιορισμούς στους συντελεστές $b_1 + \ldots + bn$).Τότε εάν $\mathrm{E}(L_i \mid L_n) = 0$ (i=1,...,n-1), η αρχική κατανομή είναι κανονική.

(ε) $X_1, X_2, \ldots, X_n$ ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές με μέσο μηδέν και διασπορά μονάδα. Εάν $\mathrm{E}(\overline{X}^2 \mid X_2 - X_1, X_3 - X_1, \ldots, X_n - X_1) = 0$ τότε τα X_i κατανέμονται κανονικά (Wesolowski, 1987). Αυτό το αποτέλεσμα συσχετίζεται με αυτό στο (α).

5.3.5 (α) Το ακόλουθο αποτέλεσμα συσχετίζεται με το θεώρημα Darmois – Skitovich του [(c) 5.1.4], αποδυναμώνει την συνθήκη της ανεξαρτησίας των γραμμικών στατιστικών και αφ' ετέρου, απαιτεί την υπόθεση των πεπερασμένων διασπορών.

Έστω $X_1, X_2, \ldots, X_n$ ένα τυχαίο δείγμα από έναν πληθυσμό με μέσο μηδέν και διασπορά σ^2. Καθορίζουμε τα γραμμικά στατιστικά

$$Y_1 = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n \text{ και } Y_2 = b_1 X_1 + b_2 X_2 + \ldots + b_n X_n$$

έτσι ώστε $a_n \neq 0$, $|b_n| > \max(|b_1|, \ldots, |b_{n-1}|)$, και $\mathrm{E}(Y_1 \mid Y_2) = 0$.

Εάν $\sigma^2 < \infty$, $\sum_{i=1}^{n} a_i b_i = 0$ $\left(\acute{\eta}\,\sigma^2 > 0\right)$, και $\frac{a_i b_i}{\alpha_n b_n} < 0$ (i=1,...,n-1), τότε ο αρχικός πληθυσμός είναι κανονικός (Rao, 1967).

(β) Μια ειδική περίπτωση του(α) είναι η ακόλουθη (Rao, 1967). Έστω $X_1, X_2,\dots,X_n$ ένα τυχαίο δείγμα από έναν πληθυσμό με μέσο μηδέν και πεπερασμένη διασπορά ($n \geq 3$), και έστω $\overline{X}$ είναι ο μέσος του δείγματος. Τότε εάν, για οποιαδήποτε σταθερή τιμή του i $(1 \leq i \leq n)$ $\mathrm{E}\left(\overline{X} \mid X_i - \overline{X}\right) = 0$ ο αρχικός πληθυσμός είναι κανονικός.

(γ) Pathak και Phillai (1968,) δίνει έναν χαρακτηρισμό που αφαιρεί την υπόθεση μιας πεπερασμένης διαφοράς στο (α). Έστω $X_1, X_2,\dots,X_n$ είναι n+1 ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές με κοινό μέσο μηδέν. Έστω

$$Y_1 = X_0 - a_1 X_1 - a_2 X_2 - \dots - a_n X_n \text{ και } Y_2 = X_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n$$

όπου $a_i b_i > 0$ και $|b_i| < 1 (i = 1,\dots,n)$, και έτσι ώστε $\mathrm{E}(Y_1 \mid Y_2) = 0$. Τότε η αρχική κατανομή είναι κανονική εάν και μόνο εάν $\sum_{i=1}^{n} a_i b_i = 1$.

(δ) Rao (1967) επέκτεινε το χαρακτηρισμό στο (α) ως εξής. Έστω

$X_1, X_2,\dots,X_n$ ($n \geq 3$) αμοιβαία ανεξάρτητες (αλλά όχι απαραιτήτως όμοια κατανεμημένες) τυχαίες μεταβλητές, και υποθέτουμε ότι υπάρχουν n γραμμικά στατιστικά.

$$Y_i = a_{i1} X_1 + \dots + a_{in} X_n \quad , \; \mathrm{i} = 1,\dots,\mathrm{n}$$

έτσι ώστε η ορίζουσα $\left|\left(a_{ij}\right)\right| \neq 0$ και $a_{11},\dots,a_{1n}$ (όλα διαφορετικά από το μηδέν). Εάν $\mathrm{E}(X_i) = 0 (i = 1,\dots,n)$ και $\mathrm{E}(Y_1 \mid Y_2,\dots,Y_n) = 0$,τότε $X_1, X_2,\dots,X_n$ είναι όλες κανονικά κατανεμημένες.

5.3.6 Το ακόλουθο αποτέλεσμα μπορεί να συγκριθεί με αυτό στο [5. 1. 6]. Έστω $X_1, X_2,\dots,X_n$ μια ακολουθία ανεξάρτητων και ισόνομα κατανεμημένων μη εκφυλισμένων τυχαίων μεταβλητών με μέσο μηδέν και ροπές όλων των βαθμών.

Υποθέτουμε ότι $\{a_i\}$ και $\{b_i\}$είναι ακολουθίες πραγματικών σταθερών έτσι ώστε $\sum_{i=1}^{\infty}|a_i|$ συγκλίνει, $\sum_{i=1}^{\infty} b_i X_i$ συγκλίνει με πιθανότητα ένα, και

$$E\left(\sum_{i=1}^{\infty} a_i X_i \mid \sum_{i=1}^{\infty} b_i X_i\right) = 0$$

Τότε ο αρχικός πληθυσμός είναι κανονικός (Kagan και λοιποί. 1973).

5.3.7 Γραμμική Παλινδρόμηση και Ομοσκεδαστικότητα

Έστω $X_1, X_2, \dots, X_n$ αμοιβαία ανεξάρτητες τυχαίες μεταβλητές με πεπερασμένη $\sigma_i^2(i=1,\dots,n)$.

Έστω $L_1 = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$ και $L_2 = b_1 X_1 + b_2 X_2 + \dots + b_n X_n$ όπου $a_i b_i \neq 0(i=1,\dots,n)$.

Τότε $E(L_1 \mid L_2) = a + \beta L_2$ και $Var(L_1 \mid L_2) = \sigma_0^2$ για κάποιες σταθερές α και β, εάν και μόνο εάν

(i) X_i είναι κανονική όποτε $b_i \neq \beta a_i$, και

(ii) $\beta = \dfrac{\Sigma' a_i b_i \sigma_i^2}{\Sigma' a_i^2 \sigma_i^2}$, $\quad \sigma_0^2 = \Sigma'\{(b_i - \beta a_i)^2 \sigma_i^2\}$

όπου Σ' δείχνει το άθροισμα όλων των ανωτέρω i έτσι ώστε $b_i \neq \beta a_i$, (Kagan και λοιποί. 1973, Lukacs και Laha, 1964).

5.3.8 Έστω $X_1, X_2, \dots, X_n$ τυχαίες μεταβλητές με πεπερασμένη διασπορά σ^2. Εάν η υπό συνθήκη ελπίδα οποιουδήποτε αμερόληπτου τετραγωνικού εκτιμητή $c\sigma^2$ (c≠0), λαμβάνοντας υπόψη το σταθερό άθροισμα $X_1 + \cdots + X_n$, δεν περιλαμβάνει τα τελευταία, η κατανομή του αρχικού πληθυσμού είναι κανονική (Laha, 1953). Δείτε επίσης [5.1. 2] (α).

5.3.9 Στο υπόλοιπο αυτού του τμήματος υποθέτουμε την έννοια της σταθερότητας της παλινδρόμησης. Μία τυχαία μεταβλητή Y με πεπερασμένο μέσο καθορίζεται για να έχει σταθερή παλινδρόμηση σε μία τυχαία μεταβλητή X εάν E(Y|X) =E(Y) με μια πιθανότητα όσον αφορά τη κατανομή πιθανότητας του X (Lukacs, 1956). Ένας άλλος τρόπος καθορίζει, ότι η υπό συνθήκη ελπίδα του Υ, δεδομένου ότι X =x, είναι ανεξάρτητη του x, σχεδόν παντού.

5.3.10 Έστω $X_1, X_2, \ldots, X_n$ τυχαίο δείγμα από έναν ορισμένο πληθυσμό με πεπερασμένη διασπορά. Έστω L και Q γραμμικά και τετραγωνικά στατιστικά, που καθορίζονται από

$$L = X_1 + \ldots + X_n \quad \text{και} \quad Q = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}X_iX_j + \sum_{i=1}^{n} b_iX_i$$

όπου $\sum_{i=1}^{n} a_{ii} \neq 0$, $\sum_{i=1}^{n}\sum_{i=1}^{n} a_{ij} = 0$ και $\sum_{i=1}^{n} b_i = 0$. Τότε ο αρχικός πληθυσμός είναι κανονικός εάν και μόνο εάν το Q έχει σταθερή παλινδρόμηση στο L. (Kagan και λοιποί. 1973, Mathai και Pederzoli 1977, Lukacs και Laha, 1964). Δείτε επίσης Lukacs (1956).

5.3.11 (α) Έστω $X_1, X_2, \ldots, X_n$ ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές με ροπές μέχρι το βαθμό m. Ορίζουμε τα γραμμικά στατιστικά $L_1 = \sum_{i=1}^{n} a_iX_i$, $L_2 = \sum_{i=1}^{n} b_iX_i$ **όπου** $\sum_{i=1}^{n} a_ib_i = 0$ και $a_ia_j > 0(i, j = 1, \ldots, n)$.

Τότε το L_2^2 έχει σταθερή παλινδρόμηση στο L_1 εάν και μόνο εάν η αρχική κατανομή είναι κανονική (Cacoullos, 1967a, δείτε επίσης Cacoullos, 1967b).

(β) Έστω $X_1, X_2, \ldots, X_n$ τυχαίες μεταβλητές όπως στο (α), με μια συμμετρική κατανομή. Ορίζουμε L_1 και L_2 όπως στο (α),όπου $a_i \neq 0(i = 1, \ldots, n)$ και $\sum_{i=1}^{n} a_ib_i = 0$. Τότε το L_2^2 έχει σταθερή παλινδρόμηση στο L_1 εάν και μόνο εάν η αρχική κατανομή είναι κανονική (Cacoullos, 1967a).

5.3.12 Έστω $X_1, X_2, \ldots, X_n$ ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές έτσι ώστε $E\left(|X_1|^p\right) < \infty$ $(p \geq 2)$. Έστω $P = P(X_1, \ldots, X_n)$ ένα ομοιογενές πολυωνυμικό στατιστικό βαθμού p, το όποιο είναι ένας αμερόληπτος εκτιμητής του p-βαθμού αθροιστικού κ_p. Τότε P έχει σταθερή παλινδρόμηση στο άθροισμα $X_1 + \ldots + Xn$ εάν και μόνο εάν ο θεμελιώδης πληθυσμός είναι κανονικός (Kagan και λοιποί 1973).

5.4 Ανεξαρτησία μερικών Στατιστικών

5.4.1 Μία φυσική ερώτηση που υποβάλλεται είναι εάν η ανεξαρτησία του δειγματικού μέσου $\overline{X}$ και μερικών συναρτήσεων των $X_1-\overline{X},...,X_n-\overline{X}$ σε ένα τυχαίο δείγμα από έναν πληθυσμό χαρακτηρίζει τον κανονικό νόμο. Για μια σχετική συζήτηση, δείτε Johnson και Kotz (1970)

5.4.2 Έστω $X_1, X_2,...,X_n$ ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές και έστω $M_p=\dfrac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^p}{n}$, η p-βαθμού δειγματική ροπή (p= 2,3,...).Αν το (p-1)! δεν είναι διαιρετό με (n-1), τότε M_p και $\overline{X}$ είναι ανεξάρτητα εάν και μόνο εάν ο αρχικός πληθυσμός είναι κανονικός (Lukacs και Laha, 1964).

5.4. 3 Ένα πολυώνυμο $P(x_1,...,x_n)$ βαθμού p λέγεται ότι δεν είναι μοναδικό εάν περιέχει τη p-βαθμού δύναμη τουλάχιστον της μιας μεταβλητής και εάν για όλους τους θετικούς ακέραιους αριθμούς k, όπου Π (k) είναι το διαμορφωμένο πολυώνυμο, εάν αντικαθιστούμε στο $P(x_1,...,x_n)$ κάθε θετική δύναμη του x_s^j με το k(k-1) ... (k-j-1) (Lukacs και Laha, 1964).

Έστω $X_1, X_2,...,X_n$ είναι ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές, και το $P=P(X_1,...,X_n)$ δεν είναι μοναδικό πολυωνυμικό στατιστικό. Εάν P και X_1 +... + Xn είναι ανεξάρτητα, τότε η αρχική κατανομή είναι κανονική (Lukacs και Laha, 1964).

5.4. 4 Έστω X και Y είναι ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές. Έστω ότι το πηλίκο X/Y ακολουθεί το νόμο Cauchy, κατανέμεται συμμετρικά γύρω από το μηδέν και είναι ανεξάρτητο από το $U=X^2+Y^2$. Τότε οι τυχαίες μεταβλητές X και Y ακολουθούν τον κανονικό νόμο (Seshadri, 1969).

5.4.5 Έστω $X_0,X_1,...,X_n$ ($n\geq 2$) n+1 ανεξάρτητες τυχαίες μεταβλητές τέτοιες ώστε $\Pr(X_i=0)=0$ (i =0,1,...,n) και με συμμετρικές κατανομές γύρω από το μηδέν. Έστω ακόμη ότι

$$Y_1=\frac{X_1}{|X_0|},\quad Y_2=\frac{X_2\sqrt{2}}{\left(X_0^2+X_1^2\right)^{\frac{1}{2}}},\quad Y_n=\frac{X_n\sqrt{n}}{\left(X_0^2+X_1^2+...+X_{n-1}^2\right)^{\frac{1}{2}}}$$

Μία αναγκαία και ικανή συνθήκη για τα $X_0, X_1, \ldots, X_n$ να είναι ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές $N(0,\sigma^2)$ είναι ότι τα $Y_0, Y_1, \ldots, Y_n$ είναι αμοιβαία ανεξάρτητες τυχαίες μεταβλητές που έχουν την κατανομή Student t με $1,2,\ldots,n$ βαθμούς ελευθερίας, αντίστοιχα (Kotlarski, 1966, Mathai και Pederzoli, 1977).

5.4. 6 Για έναν ορισμό του k-στατιστικού στον ακόλουθο χαρακτηρισμό, δείτε Stuart και Ord (1987). Έστω $X_1, \ldots, X_n$ ένα τυχαίο δείγμα από έναν πληθυσμό με αθροιστική συνάρτηση κατανομής G(x), και έστω p ένας ακέραιος αριθμός μεγαλύτερος από το ένα, έτσι ώστε η p-βαθμού ροπή της G υπάρχει. Η κατανομή G είναι κανονική εάν και μόνο εάν το k-στατιστικό της τάξης p είναι ανεξάρτητο από το δειγματικό μέσο $\overline{X}$ (Lukacs, 1956).

5.5 Χαρακτηριστικές συναρτήσεις και ροπές

5.5.1.1 Μία κατανομή είναι χαρακτηρισμός εάν $\mathrm{E}\left[\{\min(X_1, X_2, \ldots, X_n)\}^k\right]$ ορίζεται για όλα τα $k \geq 1$, όπου $X_1, X_2, \ldots, X_n$ είναι ένα τυχαίο δείγμα από τη κατανομή που μας ενδιαφέρει (Chan, 1967). Δείτε επίσης Arnold και Meeden (1975).

5.5.1.2 Μία τυχαία μεταβλητή X έχει μία N(θ, 1) κατανομή εάν και μόνο εάν, για όλες τις πραγματικές τιμές διαφορίσιμων συναρτήσεων $f(\cdot)$ έτσι ώστε $\mathrm{E}|f'(X)| < \infty$, η ταυτότητα $\mathrm{E}\{(X-\theta)f(X)\} = \mathrm{E}\{f'(X)\}$ ισχύει (Prakasa Rao, 1979).

5.5.1.3 Έστω X μία συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας $f(\cdot)$, μέσο μ και διασπορά σ^2. Έστω g πραγματική τιμή, απολύτως συνεχής και διαφορίσιμη.

Έστω ακόμη ότι

$$U(X) = \sup_g \left\{ \frac{Var[g(X)]}{\sigma^2 \mathrm{E}[g'(X)]^2} \right\} \quad \text{και} \quad V(X) = \inf_g \left\{ \frac{Var[g(X)]}{\sigma^2 \mathrm{E}^2[g'(X)]} \right\}$$

Τότε εάν το $U(X) \equiv 1$, η X ακολουθεί κανονική κατανομή (Borovkov και Utev, 1983). Εάν $V(X) \equiv 1$, τότε η X ακολουθεί κανονική κατανομή (Cacoullos και Papathanasiou, 1989). Δείτε επίσης Konwar (1991)

5.5.2 Έστω m_1, m_2 και n είναι ακέραιοι αριθμοί τέτοιοι ώστε: $0 \le m_i \le n$,n⟩1 $(i = 1,2)$. Υποθέτουμε ότι $\psi_1(t)$ και $\psi_2(t)$ είναι δύο χαρακτηριστικές συναρτήσεις τέτοιοι ώστε:

$$\psi_1(t)\psi_2(t) = \exp(ict)\left\{\psi_1\left(\frac{t}{\sqrt{n}}\right)\right\}^{m_1}\left\{\psi_1\left(-\frac{t}{\sqrt{n}}\right)\right\}^{n-m_1} \cdot \left\{\psi_2\left(\frac{t}{\sqrt{n}}\right)\right\}^{m_2}\left\{\psi_2\left(-\frac{t}{\sqrt{n}}\right)\right\}^{n-m_2}$$

για όλες τις τιμές του t. Τότε $\psi_1(t)$ και $\psi_2(t)$ αντιστοιχούν στις κανονικές κατανομές (Blum, 1956).

5.5.3 Το ακόλουθο αποτέλεσμα είναι χρήσιμο στη μελέτη κατανομών μεγέθους σημείου. Έστω Y μία τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας $p(x)$ και ροπογεννήτρια συνάρτηση $M(t)$, και έστω X τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας $g(x)$, όπου

$$g(x) = \frac{\exp(tx)p(x)}{M(t)} \quad , -\infty < x < \infty$$

Τότε $Var(X) = Var(Y)$ για όλες τις τιμές του t εάν και μόνο εάν η Y έχει μία κανονική συνάρτηση πυκνότητας πιθανότητας (Ziegler, 1965). Υπό αυτούς τους όρους, εάν Y ακολουθεί κατανομή $N(\mu,\sigma^2)$, τότε η X ακολουθεί κατανομή $N(\mu+\sigma^2 t,\sigma^2)$.

5.5.4 Έστω ψ είναι μια μη τετριμμένη χαρακτηριστική συνάρτηση έτσι ώστε, σε κάποια γειτονιά προέλευσης όπου δεν μηδενίζεται.

$$\psi(t) = \prod_{j=1}^{\infty}\left[\{\psi(\pm\beta_j t)\}^{v_j}\right], 0 < \beta_j < 1 , v_j > 0 \ (j = 1,2,...)$$

Τότε (α) $\sum_{j=1}^{\infty} v_j \beta_j^2 \le 1$ και (β) ψ είναι η χαρακτηριστική συνάρτηση ενός κανονικού νόμου, εάν και μόνο εάν $\sum_{j=1}^{\infty} v_j \beta_j^2 \le 1$ (Kagan και λοιποί. 1973).

5.5.5 Η τυχαία μεταβλητή W^2 ακολουθεί μία X_1^2 κατανομή εάν και μόνο εάν η χαρακτηριστική συνάρτηση του W, $\psi(t)$ ικανοποιεί την εξίσωση

$$\psi(t)+\psi(-t)=2\exp\left(-\frac{t^2}{2}\right)$$

(Geisser, 1973). Η κατανομή του W είναι τότε N(0, 1).

5.5.6 Έστω Z_1, Z_2, και Z_3 τρεις αμοιβαία ανεξάρτητες τυχαίες μεταβλητές, που κατανέμονται συμμετρικά γύρω από το μηδέν με αθροιστική συνάρτηση κατανομής συνεχή στο μηδέν. Τότε η κοινή χαρακτηριστική συνάρτηση των αναλογιών $\frac{Z_1}{Z_3}$ και $\frac{Z_2}{Z_3}$ είναι $\exp\left[-\left(t_1^2+t_2^2\right)^{\frac{1}{2}}\right]$ εάν και μόνο εάν Z_1, Z_2, και Z_3 είναι ανεξάρτητες και ισόνομα κατανεμημένες κανονικές τυχαίες μεταβλητές με μέσο μηδέν (Kotz, 1974, Pakshirajan και Mohan, 1969).

5.5.7 Έστω X, Y και Z ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές, με πεπερασμένη απόλυτη ροπή της καθορισμένη σειράς p, p>0, p περιττός. Αν για κάποια θετική σταθερά A η εξίσωση $\mathrm{E}\left(|aX+bY+cZ|^p\right)=\mathrm{A}\left(a^2+b^2+c^2\right)^{\frac{p}{2}}$ ικανοποιείται για οποιουσδήποτε πραγματικούς αριθμούς a, b και c, τότε κάθε μια από τις X, Y και Z είναι κανονικά κατανεμημένη με μέσο μηδέν (Braverman, 1985).

5.6 Χαρακτηρισμοί από τις ιδιότητες των μετασχηματισμών

5.6.1 Έστω $x_1, x_2, \ldots, x_n$ οι καρτεσιανές συντεταγμένες, όχι όλες μηδέν, από ενός σημείου στον Ευκλείδειο χώρο των n διαστάσεων. Μπορούμε να μετασχηματίσουμε στις πολικές συντεταγμένες $(r, \theta_1, \theta_2, \ldots, \theta_{n-1})$, εάν $n \geq 2$, ως εξής: Ορίζουμε το μετασχηματισμό T με :

$$x_1 = r\cos\theta_1 \;,\; x_2 = r\sin\theta_1\cos\theta_2$$

$$x_3 = r\sin\theta_1\sin\theta_2\cos\theta_3, \ldots,$$

$$x_{n-1} = r\sin\theta_1\sin\theta_2\ldots\sin\theta_{n-2}\cos\theta_{n-1}$$

$$x_n = r\sin\theta_1\ldots\sin\theta_{n-2}\sin\theta_{n-1} \;,\; -\infty < x_i < \infty \;\; (i=1,\ldots,n)$$

$$r \geq 0 \,,\; 0 \leq \theta_1 \leq 2\pi \;\; (i=1,\ldots,n-1)$$

Εάν X_1, X_2,…,X_n είναι τυχαίες μεταβλητές με κοινή συνάρτηση πυκνότητας πιθανότητας, $p_1(x_1,\ldots,x_n)$ ο μετασχηματισμός T οδηγεί σε πολικές τυχαίες μεταβλητές $R, \Theta_1, \ldots, \Theta_{n-1}$ με κοινή συνάρτηση πυκνότητας πιθανότητας $p_2(r, \theta_1, \theta_2, \ldots, \theta_{n-1})$ που δίνεται από το

$$p_2(r, \theta_1, \theta_2, \ldots, \theta_{n-1}) = p_1(x_1, x_2, \ldots, x_{n-1}) r^{n-1} \left|\sin^{n-2}\theta_1 \sin^{n-3}\theta_2 \ldots \sin\theta_{n-2}\right|$$

Μπορούμε να μιλήσουμε για το R ως ακτινική μεταβλητή, και για τα $\Theta_1, \ldots, \Theta_{n-1}$ σαν γωνιακές μεταβλητές σημειώνουμε ότι $R^2 = X_1^2 + \ldots + X_n^2$

(α) Έστω $X_1, X_2, \ldots, X_n$ ένα τυχαίο δείγμα από έναν ορισμένο πληθυσμό, ο οποίος είναι απολύτως συνεχής. Τότε η κοινή συνάρτηση πυκνότητας πιθανότητας του δείγματος, $p_1(x_1, x_2, \ldots, x_n)$, είναι σταθερή στη "σφαίρα" $x_1^2 + x_2^2 + \ldots + x_n^2 = r^2$ εάν και μόνο εάν ο αρχικός πληθυσμός είναι τυποποιημένος κανονικός (Bartlett, 1934, Stuart και Ord, 1987).

(β) Έστω $X_1, X_2, \ldots, X_n$ $(n \geq 2)$ τυχαίες μεταβλητές. Υποθέτουμε ότι δεδομένου του μετασχηματισμού T η αναλογία

$$p_2(r, \theta_1, \ldots, \theta_{n-1}) = p_1(x_1, \ldots, x_n) r^{n-1} \left|\sin^{n-2}\theta_1 \sin^{n-3}\theta_2 \ldots \sin\theta_{n-2}\right|$$

είναι καλά ορισμένη και διαφορετική από το μηδέν παντού, συνεχής στο r και ικανή στο $p_1(x_1, x_2, \ldots, x_n)$, που είναι στη συνέχεια συνεχής σε κάθε x_i $(i = 1, \ldots, n)$ Τότε οι $X_1, X_2, \ldots, X_n$ είναι αμοιβαία ανεξάρτητες και το P είναι ανεξάρτητο από τα $\Theta_1, \ldots, \Theta_{n-1}$ εάν και μόνο εάν $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές με κατανομή $N(0,\sigma^2)$ (Tamhanker, 1967, Kotz, 1974, Johnson και Kotz, 1970). Αυτό το αποτέλεσμα θέτει ουσιαστικά τους όρους κάτω από τους οποίους η ανεξαρτησία των ακτινικών και (κοινών) γωνιακών μεταβλητών χαρακτηρίζει την κανονικότητα.Ο Tamhanker δίνει μετασχηματισμένη συνάρτηση πυκνότητας πιθανότητας $p_2(r, \theta_1, \theta_2, \ldots, \theta_{n-1})$ χωρίς συμβολισμό απόλυτης τιμής. Για έναν πολύ σχετικό χαρακτηρισμό δείτε Alam (1971).

5.6. 2 Έστω X_1, X_2 ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές με (συνεχή) συνάρτηση πυκνότητας πιθανότητας g(x).

Έστω Y_1, Y_2 τυχαίες μεταβλητές που ορίζονται από το μετασχηματισμό

$$Y_1 + iY_2 = (X_1 + iX_2)^k \left(X_1^2 + X_2^2\right)^{-\frac{(k-1)}{2}}, i = \sqrt{-1}$$

όπου k είναι ένας ακέραιος αριθμός όχι μικρότερος από το 2. Τότε X_1 και X_2 είναι ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές $N(0,\sigma^2)$ εάν και μόνο εάν Y_1 και Y_2 είναι ανεξάρτητες με την ίδια κατανομή με τις X_1 και X_2 (Beer και Lukacs, 1973, Kotz, 1974).

5.6. 3 Έστω X_1 και X_2 ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές με (συνεχή) συνάρτηση πυκνότητας πιθανότητας g(x). Έστω Y_1 και Y_2 τυχαίες μεταβλητές που ορίζονται από το μετασχηματισμό

$$Y_1 = X_1 \cos\{a(X_1^2 + X_2^2)\} + X_2 \sin\{a(X_1^2 + X_2^2)\}$$

$$Y_2 = -X_1 \cos\{a(X_1^2 + X_2^2)\} + X_2 \cos\{a(X_1^2 + X_2^2)\}$$

όπου a είναι μια σταθερά διάφορη του μηδενός. Τότε X_1 και X_2 είναι ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές $N(0,\sigma^2)$ εάν και μόνο εάν Y_1 και Y_2 είναι ανεξάρτητες με την ίδια κατανομή με τις X_1 και X_2 (Beer και Lukacs, 1973).

5.6.4 (α) Το ακόλουθο αποτέλεσμα σχεδιάστηκε για τη δοκιμή της κανονικότητας. Έστω n= 2k + 3, όπου $k \geq 2$. Έστω $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές με (άγνωστο) μέσο μ και (άγνωστη) διασπορά σ^2, $|\mu|\langle\infty$, $\sigma^2\rangle 0$. Έστω ακόμη ότι

$$Z_1 = \frac{(X_1 - X_2)}{\sqrt{2}} \quad , \quad Z_2 = (X_1 + X_2 - 2X_3)\sqrt{3.2}, \ldots,$$

$$Z_{n-1} = \frac{\{X_1 + \ldots + X_{n-1} - (n-1)X_n\}}{\sqrt{n(n-1)}}$$

$$Z_n = X_1 + \ldots + X_n$$

$$Y_1 = Z_1^2 + Z_2^2 \quad , \quad Y_2 = Z_3^2 + Z_4^2 , \ldots,$$

$$Y_{k+1} = Z_{n-2}^2 + Z_{n-1}^2 \quad , \quad S_{k+1} = Y_1 + Y_2 + \ldots + Y_k$$

όπου $k+1 = \frac{1}{2}(n-1)$. Τέλος, έστω

$$U_{(r)} = \frac{(Y_1 + \ldots + Y_r)}{S_{k+1}} \quad , \quad r = 1, \ldots, k$$

Τότε $U_{(1)}, U_{(2)}, \ldots, U_{(k)}$ συμπεριφέρονται όπως το στατιστικό k τάξης των k ανεξάρτητων και ισόνομα κατανεμημένων τυχαίων μεταβλητών που έχουν μια ομοιόμορφη κατανομή στο (0, 1), εάν και μόνο εάν η αρχική κατανομή των $X_1, X_2,\ldots,X_n$ είναι $N(\mu,\sigma^2)$ (Csörgö και Seshadri, 1971, Kotz, 1974).

(β) Οι πηγές στο (α) δίνουν επίσης έναν ανάλογο χαρακτηρισμό που είναι κατάλληλος όταν μ είναι γνωστό, εδώ $n = 2k$, $Z_i = X_i - \mu$ $(i = 1,\ldots,n)$, το Y_j καθορίζεται ομοίως, με k+1 που αντικαθίσταται από το k. $S_k = Y_1 + \ldots + Y_k$ και $U_{(r)} = \dfrac{(Y_1 + \ldots + Y_r)}{S_k}$, όπου τώρα $r = 1,\ldots,k-1$. Το συμπέρασμα είναι άμεσα ανάλογο, αντικαθιστώντας το k από το k-1.

5.6.5 Υποθέτουμε ότι υπάρχει ένας μη σημαντικός γραμμικός μετασχηματισμός

$$\begin{pmatrix} W \\ Z \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

του ενός ζευγαριού των ανεξάρτητων τυχαίων μεταβλητών (X,Y) σε ένα άλλο ζευγάρι ανεξάρτητων τυχαίων μεταβλητών (W, Z). Τότε :

i) κάθε μεταβλητή έχει μια πεπερασμένη διασπορά,

ii) και οι τέσσερις μεταβλητές κατανέμονται κανονικά ή όλες είναι εκφυλισμένες (Lancaster, 1987).

Αυτός ο χαρακτηρισμός δεν απαιτεί την υπόθεση των πεπερασμένων διασπορών.

5.6.6 Στους ακόλουθους δύο χαρακτηρισμούς X και Y είναι ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες μεταβλητές, και $U = \min(X,Y)$.

Και για τα δύο αποτελέσματα δείτε Ahsanullah και Hamedani (1988).

(α) X και Y είναι συμμετρικές γύρω από το μηδέν, με συνάρτηση πυκνότητας πιθανότητας $g(\cdot)$. Εάν U^2 κατανέμεται ως X^2 με έναν βαθμό ελευθερίας δηλαδή ως X_1^2, τότε οι X και Y τυποποιημένες κανονικές.

(β) X και Y έχουν μια απολύτως συνεχή κατανομή. Εάν το U^2 κατανέμεται ως X_1^2 και X/Y ως Cauchy με συνάρτηση πυκνότητας πιθανότητας $\left[\pi\left(1+x^2\right)\right]^{-1}$, τότε οι X και Y είναι τυποποιημένες κανονικές.

ΚΕΦΑΛΑΙΟ 6^{Ο}

Κεντρικό Οριακό Θεώρημα

Εκφράσεις όπως "μακροπρόθεσμα" και "κατά μέσο όρο" είναι: συνηθισμένες στην καθημερινή μας ζωή και εκφράζουν την πεποίθηση μας ότι οι μέσοι των αποτελεσμάτων επαναλαμβανόμενων πειραμάτων επιδεικνύουν λιγότερη τυχαία κύμανση(μεταβλητότητα) καθώς σταθεροποιούνται σε κάποιο όριο.

Όταν μελετάμε απλά φαινόμενα, τα παρατηρούμε μαζί με όλες τις ιδιαιτερότητες που τα συνοδεύουν. Οι ιδιαιτερότητες αυτές εμποδίζουν να αντιληφθούμε τους νόμους που διέπουν τη μελέτη μεγάλου αριθμού παρόμοιων φαινομένων. Από πολύ παλιά είχε παρατηρηθεί ότι παράγοντες οι οποίοι δεν συνδέονται με την ουσία της διεργασίας που προκάλεσε το φαινόμενο ως σύνολο και οι οποίο εμφανίζονται μόνο σε ορισμένες μεμονωμένες περιπτώσεις εξουδετερώνονται αμοιβαία, *όταν θεωρηθεί ο μέσος ενός μεγάλου αριθμού παρατηρήσεων.* Αργότερα, αυτό το εμπειρικό αποτέλεσμα παρατηρήθηκε κατ' επανάληψη χωρίς όμως να γίνει ποτέ κάποια προσπάθεια για να εξηγηθεί θεωρητικά. Επιπλέον, σύμφωνα με πολλούς συγγραφείς δεν χρειαζόταν καμιά ερμηνεία, γιατί η παρουσία "τάξης" (κανόνων) τόσο στα φυσικά όσο και στα κοινωνικά φαινόμενα δεν ήταν για αυτούς παρά το αποτελέσμα θεϊκής επέμβασης.

Αργότερα από την επιστημονική έρευνα του Chebyshev, του Markov καθώς και άλλων ερευνητών βλέπουμε ότι οι ερευνητές αυτοί, εκτός του ότι ανίχνευσαν την εμπειρική σταθερότητα των μέσων, επιπλέον βρήκαν τις γενικές προϋποθέσεις των οποίων η ικανοποίηση αναδεικνύει τη στατιστική σταθερότητα των μέσων.

Η έννοια της σύγκλισης στην θεωρία πιθανοτήτων παίζει πολύ σημαντικό ρόλο και υπάρχουν αρκετοί τρόποι να επεκταθεί η έννοια της σύγκλισης μιας ακολουθίας πραγματικών αριθμών σε ακολουθίες τυχαίων μεταβλητών.

Εκείνο που μας ενδιαφέρει είναι η σύγκλιση μιάς οικογένειας τυχαίων μεταβλητών και μάλιστα η μελέτη της ακολουθίας $S_n = X_1 + X_2 + ... + X_n$ μερικών αθροισμάτων μιας ακολουθίας ανεξάρτητων και ισόνομων τυχαίων μεταβλητών. Συχνά ενδιαφερόμαστε για μελλοντικές τιμές διαδικασιών οι οποίες συμβαίνουν στον κόσμο, και επομένως, για την μακροπρόθεσμη συμπεριφορά του μαθηματικού μοντέλου. Γι αυτό χρειαζόμαστε να αποδείξουμε οριακά θεωρήματα για ακολουθίες τυχαίων μεταβλητών.

Ο νόμος των τυχαίων αριθμών προσδιορίζει τη μορφή σύγκλισης ακολουθιών τέτοιων τυχαίων μεταβλητών, οι οποίες μπορούν να εκφρασθούν ως μερικά αθροίσματα άλλων ανεξαρτήτων τυχαίων μεταβλητών. Συγκεκριμένα, αναφέρει ότι τυχαίες μεταβλητές της μορφής $S_n = X_1 + X_2 + ... + X_n$ συγκλίνουν, με κάποια έννοια σύγκλισης, στην μέση τους τιμή, όπως αυτή ορίζεται από τη σχέση. Έστω $\mathrm{E}(S_n) = \sum_{j=1}^{n} \mathrm{E}(X_j)$

Το επόμενο ερώτημα, που όπως είναι φυσικό ανακύπτει, έχει να κάνει με την κατανομή της τυχαίας μεταβλητής S_n. Το Κεντρικό Οριακό θεώρημα απαντά στο ερώτημα αυτό. Συγκεκριμένα, αποδεικνύει ότι η κατανομή της τυχαίας μεταβλητής S_n είναι κατά προσέγγιση η κανονική κατανομή. Το γνωστό αυτό θεώρημα έχει γίνει αντικείμενο έντονης ερευνητικής δραστηριότητας με στόχο την ανακάλυψη των πιο γενικών συνθηκών κάτω από τις οποίες αυτό ισχύει. Από την άλλη πλευρά, το θεώρημα αυτό έχει αποτελέσει τη βάση της εφαρμοσμένης έρευνας. Στη θεωρία της μέτρησης των σφαλμάτων, το παρατηρούμενο σφάλμα μπορεί να εκφρασθεί ως άθροισμα ενός μεγάλου αριθμού ανεξάρτητων τυχαίων ποσοτήτων οι οποίες συνεισφέρουν στο αποτέλεσμα. Επίσης, στην Στατιστική, ο μέσος n ανεξάρτητων τυχαίων μεταβλητών (που είναι ίσος με μία σταθερά (n) επί το άθροισμα των μεταβλητών αυτών) είναι κεντρικής σημασίας.

Σε τέτοιες περιπτώσεις, η υπόθεση μιας κανονικής κατανομής μπορεί να είναι κατάλληλη. Στη συνέχεια θα θεωρήσουμε μια μορφή του Κεντρικού Οριακού Θεωρήματος κάτω από υποθέσεις οι οποίες είναι εύλογες στις περισσότερες πρακτικές εφαρμογές.

6.1 Το Κεντρικό Οριακό θεώρημα

Θεώρημα *(Κεντρικό Οριακό θεώρημα):* Έστω $\{X_n\}$ μία ακολουθία ανεξάρτητων τυχαίων μεταβλητών με πεπερασμένες μέσες τιμές και πεπερασμένες διασπορές. Ορίζουμε την ακολουθία $\{S_n\}$ των μερικών αθροισμάτων $S_n = \sum_{i=1}^{n} X_i$, για κάθε $n \geq 1$, για την οποία ισχύει ότι

$$\mathrm{E}(S_n) = \sum_{j=1}^{n} \mathrm{E}(X_j)$$

και $V(S_n) = \sum_{i=1}^{n} V(X_i)$ για κάθε $n \geq 1$

Έστω $Z_n = \dfrac{S_n - \mathrm{E}(S_n)}{\sqrt{V(S_n)}}$

και F_n η συνάρτηση κατανομής της τυχαίας μεταβλητής Z_n.

Τότε, $\lim_{n \to \infty} F_n(z) = \Phi(z)$, $z \in (-\infty, +\infty)$

όπου $\Phi(z)$ είναι η συνάρτηση κατανομής της τυποποιημένης κανονικής κατανομής.

► Θα δώσουμε μια απλούστερη μορφή του Κεντρικού Οριακού Θεωρήματος με την οποία το θεώρημα αυτό είναι περισσότερο γνωστό. Η μορφή αυτή αντιστοιχεί στην περίπτωση όπου: $\mathrm{E}(X_n) = \mu$ και $V(X_n) = \sigma^2$ για κάθε n.

Η ειδική αυτή περίπτωση οδηγεί στην εξής μορφή του Κεντρικού Οριακού Θεωρήματος.

Θεώρημα: Έστω $\{X_n\}$ μία ακολουθία ανεξάρτητων τυχαίων μεταβλητών με ίσες πεπερασμένες μέσες τιμές και ίσες πεπερασμένες διασπορές, δηλαδή

$\mathrm{E}(X_n) = \mu$ και $V(X_n) = \sigma^2$ για κάθε n.

Έστω επί πλέον η τυχαία μεταβλητή $Z_n = \dfrac{S_n - \mathrm{E}(S_n)}{\sqrt{V(S_n)}}$

με συνάρτηση κατανομής F_n.

Τότε $\lim\limits_{n\to\infty} F_n(z) = \Phi(z)$, $\quad z \in (-\infty, +\infty)$

όπου Φ(z) είναι η συνάρτηση κατανομής της τυποποιημένης κανονικής κατανομής.

<u>Σκιαγράφηση της απόδειξης :</u>

Η απόδειξη στηρίζεται στη μέθοδο των ροπογεννητριών, την οποία θα παραλείψουμε. Για περισσότερη πληρότητα θα κάνουμε μία σκιαγράφηση της απόδειξης.

$$S_n = X_1 + X_2 + \ldots + X_n$$

$$\mathrm{E}(S_n) = \mathrm{E}(X_1) + \mathrm{E}(X_2) + \ldots + \mathrm{E}(X_n)$$

$$\mathrm{E}(S_n) = \mu + \mu + \ldots + \mu$$

$$\mathrm{E}(S_n) = n\mu$$

$$V(S_n) = V(X_1) + V(X_2) + \ldots + V(X_n)$$

$$V(S_n) = \sigma^2 + \sigma^2 + \ldots + \sigma^2$$

$$V(S_n) = n\sigma^2$$

$$S_n \sim \mathrm{N}(n\mu, n\sigma^2)$$

Έστω η τυχαία μεταβλητή $\overline{X} = \frac{S_n}{n} = \frac{X_1 + X_2 + ... + X_n}{n} \Leftrightarrow S_n = n\overline{X}$

Τότε έχουμε : $\frac{S_n - \mathrm{E}(Sn)}{\sqrt{V(S_n)}} = \frac{n\overline{X} - n\mu}{\sqrt{n\sigma^2}} = \frac{n\overline{X} - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}n(\overline{X} - \mu)}{\sigma n} = \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma}$

Οπότε : $P\left[\frac{S_n - \mathrm{E}(S_n)}{\sqrt{V(S_n)}} \leq z\right] \underset{n\to+\infty}{\to} \Phi(z)$ όπου: $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$, $z \in (-\infty, +\infty)$

Ή $P\left[\frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \leq z\right] \underset{n\to+\infty}{\to} \Phi(z)$

Χρησιμοποιώντας ανάλογους συμβολισμούς για αρκούντως μεγάλες τιμές του n έχουμε:

$P\left[\frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \leq z\right] \approx \Phi(z)$ για $z \in (-\infty, +\infty)$

ή $\frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \sim \mathrm{N}(0,1)$

ή $\frac{(\overline{X} - \mu)}{\frac{\sigma}{\sqrt{n}}} \sim \mathrm{N}(0,1)$

ή $\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

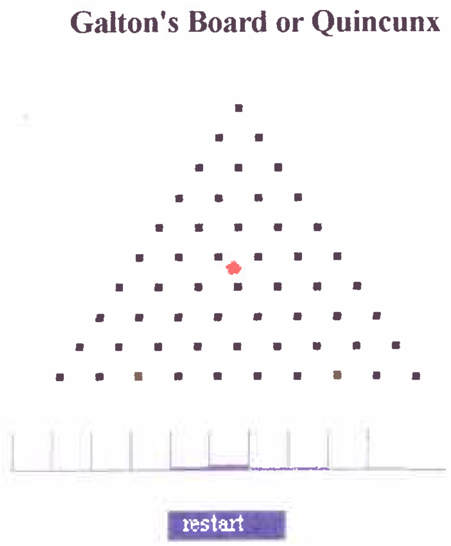
Επομένως, όταν το μέγεθος του δείγματος είναι μεγάλο (n>30), η τυχαία μεταβλητή $\overline{X}$ θα ακολουθεί την κανονική κατανομή $\mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$,ανεξάρτητα από την μορφή της κατανομής του γεννήτορα πληθυσμού από την οποία πήραμε το δείγμα.

Διαφορετικά θα μπορούσαμε να πούμε ότι για αρκετά μεγάλες τιμές του n το άθροισμα S_n, n ανεξάρτητων τυχαίων μεταβλητών $X_1, X_2, \ldots, X_n$ με μέση τιμή μ και διασπορά σ^2,έχει περίπου την κανονική κατανομή με μέση τιμή nμ και διασπορά $n\sigma^2$. Δηλαδή $P(S_n \leq s) \cong \Phi\left(\frac{s - n\mu}{\sigma\sqrt{n}}\right)$

Ένα άμεσο συμπέρασμα που μπορεί να εξαχθεί από το Κεντρικό Οριακό Θεώρημα είναι ότι για οποιοδήποτε δοθέν διάστημα [α,β] ισχύει ότι:

$$P\left(\alpha \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \beta\right) \cong \Phi(\beta) - \Phi(\alpha)$$

6.2 Προσομοίωση του Πίνακα Galton



Σχήμα 6.1: *Σχήμα προσομοίωσης του Πίνακα Galton ή χιαστό*

Αυτό το σχήμα που μοιάζει με μήλο μιμείται τον πίνακα Galton, στον οποίο οι σφαίρες πέφτουν μέσω μιας τριγωνικής σειράς καρφιών. Αυτή η συσκευή καλείται επίσης χιαστό. Κάθε φορά που χτυπά μια σφαίρα ένα καρφί έχει πιθανότητα 50 τοις εκατό να πέσει αριστερά του καρφιού και πιθανότητα 50 τοις εκατό να πέσει δεξιά του.

Οι σωροί των σφαιρών που συσσωρεύονται στις αυλακώσεις κάτω από το τρίγωνο μοιάζουν με δυωνυμική κατανομή. Για να φθάσει στο δοχείο η σφαίρα αρκετά αριστερά, πρέπει να πέσει αριστερά κάθε φορά που χτυπά ένα καρφί.

Επειδή ο πίνακας Galton αποτελείται από μια σειρά πειραμάτων οι σωροί στις αυλακώσεις είναι το ποσό 10 τυχαίων μεταβλητών.

Επομένως, αυτή η προσομοίωση παρέχει επίσης μια απεικόνιση του Κεντρικού Οριακού Θεωρήματος, το οποίο δηλώνει ότι η κατανομή του αθροίσματος των n τυχαίων μεταβλητών πλησιάζει την κανονική κατανομή όταν το n είναι μεγάλο. Όταν προσθέτουμε περισσότερες σειρές καρφιών στον πίνακα η προσέγγιση θα είναι καλύτερη.

6.3 Αλγόριθμοι Προσομοίωσης

Πολλοί αλγόριθμοι για τους ψευδοτυχαίους αριθμούς από έναν κανονικό πληθυσμό έχουν αναπτυχθεί κατά τη διάρκεια των ετών. Φυσικά οποιοσδήποτε ομοιόμορφα τυχαίος αριθμός που παράγει τον αλγόριθμο μπορεί να χρησιμοποιηθεί επιτυχώς με την αντίστροφη αθροιστική συνάρτηση κατανομής (ή μια αποδοτική προσέγγιση από αυτή) για να παραγάγει τις ψευδοτυχαίες κανονικές παρατηρήσεις. Παράλληλα έχουν αναπτυχθεί και άλλες απλούστερες, αποδοτικότερες, και γρηγορότερες μέθοδοι προσομοίωσης, μία από τις οποίες περιγράφουμε παρακάτω.

6.3.1 Μέθοδος Box-Muller

Αρχίζοντας από δύο ανεξάρτητες τυποποιημένες κανονικές μεταβλητές X_1 και X_2, οι Box-Muller (1958) εξέτασαν το μετασχηματισμό

$$Y_1 = \exp\left\{-\frac{1}{2}\left(X_1^2 + X_2^2\right)\right\} \text{ και } Y_2 = \frac{1}{2\pi}\tan^{-1}\left(\frac{X_1}{X_2}\right) \tag{6.1}$$

δείχνοντας ότι οι μεταβλητές Y_1 και Y_2 είναι ανεξάρτητες, ομοιόμορφα κατανεμημένες στο (0,1). Δεδομένου ότι Y_2 κατανέμεται ομοιόμορφα στο (0,1), εύκολα συνάγεται ότι $\frac{X_1}{X_2}$ έχει μια τυποποιημένη κατανομή Cauchy. Ως εκ τούτου, μετά παραγάγει δύο ψευδοτυχαίες ομοιόμορφες στο (0,1) παρατηρήσεις Y_1, και Y_2, μπορεί να χρησιμοποιηθεί για να λάβει τις απαραίτητες ψευδοτυχαίες κανονικές παρατηρήσεις X_1 και X_2.

Εναλλακτικά, εάν εξετάσουμε τον πολικό μετασχηματισμό

$$X_1 = r\sin\theta \text{ και } X_2 = r\cos\theta \tag{6.2}$$

Μπορεί εύκολα να ελεγχθεί ότι η κοινή πυκνότητα των r και θ είναι

$$\rho(r,\theta) = \frac{1}{2\pi}e^{-\frac{r^2}{2}}r, \text{ όπου } 0 \le r < \infty, 0 \le \theta \le 2\pi \tag{6.3}$$

Από το θεώρημα παραγοντοποίησης, έχουμε αμέσως ότι οι μεταβλητές r και θ είναι στατιστικά ανεξάρτητες. Επιπλέον οι μεταβλητές

$$U_1 = e^{-\frac{r^2}{2}} \quad \text{και} \quad U_2 = \frac{1}{2\pi}\theta \tag{6.5}$$

είναι ανεξάρτητες ομοιόμορφα κατανεμημένες στο (0,1). [Αυτές είναι ακριβώς ίδιες με τις Y_1 και Y_2 στο (6.1)]

Αναστρέφοντας το μετασχηματισμό, παίρνουμε:

$$\begin{aligned} X_1 &= \sqrt{-2\ln U_1}\cos(2\pi U_2) \\ X_2 &= \sqrt{-2\ln U_1}\sin(2\pi U_2) \end{aligned} \tag{6.6}$$

ένα ζευγάρι των ψευδοτυχαίων τυποποιημένων κανονικών παρατηρήσεων.

6.4 Η κανονική κατανομή ως δειγματική κατανομή

Όπως ήδη αναφέρθηκε, η κανονική κατανομή χρησιμοποιείται για την περιγραφή πολλών ποσοτικών φαινομένων. Αποτελεί όμως χρήσιμο εργαλείο και στην πειραματική έρευνα.

Όπως ήδη αναφέρθηκε, μέσω της χρήσης του Κεντρικού Οριακού Θεωρήματος, η κανονική κατανομή μπορεί να χρησιμοποιηθεί για την συναγωγή συμπερασμάτων όσο αφορά την ακρίβεια με την οποία ο μέσος $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ ενός δείγματος n ανεξαρτήτων παρατηρήσεων X_1, X_2, …,Xn από μια οποιαδήποτε κατανομή εκτιμά την μέση τιμή μ της κατανομής αυτής.

Συγκεκριμένα, η χρήση της κανονικής κατανομής κάνει εφικτό τον προσδιορισμό της πιθανότητας $P(-\varepsilon \le \overline{X} - \mu \le \varepsilon)$, δηλαδή της πιθανότητας με την οποία η εκτίμηση που παρέχει ο $\overline{X}$ δεν θα απέχει από την πραγματική αλλά άγνωστη τιμή της μέσης τιμής μ περισσότερο από ε.

Πιο συγκεκριμένα μια από τις σημαντικότερες εφαρμογές του Κεντρικού Οριακού Θεωρήματος συναντάται στην περιοχή της Στατιστικής Συμπερασματολογίας, όπου συχνά ενδιαφερόμαστε να εκτιμήσουμε την άγνωστη μέση τιμή μ μίας τυχαίας μεταβλητής X. Έστω ότι για το σκοπό αυτό πραγματοποιήσαμε n ανεξάρτητες παρατηρήσεις X_1, X_2, …,X_n πάνω στην τυχαία μεταβλητή X. Υποθέτουμε δηλαδή ότι ελήφθησαν παρατηρήσεις από n ανεξάρτητες επαναλήψεις του τυχαίου πειράματος που παρήγαγε την τυχαία μεταβλητή X. Ένας φυσικός και προφανής τρόπος εκτίμησης της μέσης τιμής μ είναι μέσω του μέσου των παρατηρήσεων, που συνήθως ονομάζεται δειγματικός μέσος:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Επειδή από τον ορισμό του ο δειγματικός μέσος είναι μία τυχαία μεταβλητή, δεν θα μπορούσε να περιμένει κανείς ότι η τιμή του θα είναι πάντα ίση με την άγνωστη τιμή μ, αφού κάθε σύνολο n ανεξάρτητων παρατηρήσεων πάνω στην X θα οδηγεί σε διαφορετική τιμή του $\overline{X}$. Μπορούμε όμως να "μετρήσουμε" πόσο κοντά, γύρω από την άγνωστη τιμή μ, κυμαίνονται οι τιμές του $\overline{X}$ με τον υπολογισμό της πιθανότητας $P(-\varepsilon \leq \overline{X} - \mu \leq \varepsilon)$, δηλαδή με τον υπολογισμό της πιθανότητας ότι μία συγκεκριμένη ακρίβεια ε>0 επιτυγχάνεται. Η κατανομή του $\overline{X}$, επομένως, και η ζητούμενη πιθανότητα μπορεί να είναι δύσκολο να προσδιορισθούν κυρίως όταν η κατανομή της τυχαίας μεταβλητής X δεν είναι απλής μορφής. Από το Κεντρικό Οριακό Θεώρημα ισχύει ότι, ανεξάρτητα από τη μορφή της κατανομής της τυχαίας μεταβλητής X, καθώς ο αριθμός n των ανεξαρτήτων δοκιμών αυξάνει, η παραπάνω πιθανότητα μπορεί να προσεγγισθεί από μια πιθανότητα που υπολογίζεται από την τυποποιημένη κανονική κατανομή:

$$P(-\varepsilon \leq \overline{X} - \mu \leq \varepsilon) \cong P\left(\frac{-\varepsilon\sqrt{n}}{\sigma} \leq Z \leq \frac{-\varepsilon\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1$$

όπου Z, είναι μία τυποποιημένη κανονική μεταβλητή, και σ είναι η διασπορά της X. Γενικότερα, ισχύει ότι:

$$P(a \leq \overline{X} - \mu \leq \beta) \cong \Phi\left(\frac{\beta\sqrt{n}}{\sigma}\right) - \Phi\left(\frac{\alpha\sqrt{n}}{\sigma}\right)$$

6.5 Παράδειγμα κατανομής δειγματικού μέσου

Θεωρούμε την κατανομή των επιπέδων χοληστερόλης στο αίμα για όλους τους άνδρες της Ελλάδας, από 20 έως 74 ετών. Η κατανομή αυτή είναι περίπου κανονική. Η μέση τιμή της χοληστερόλης αυτού του πληθυσμού είναι μ=211mg/100ml και η τυπική απόκλιση είναι σ=46mg/100ml.

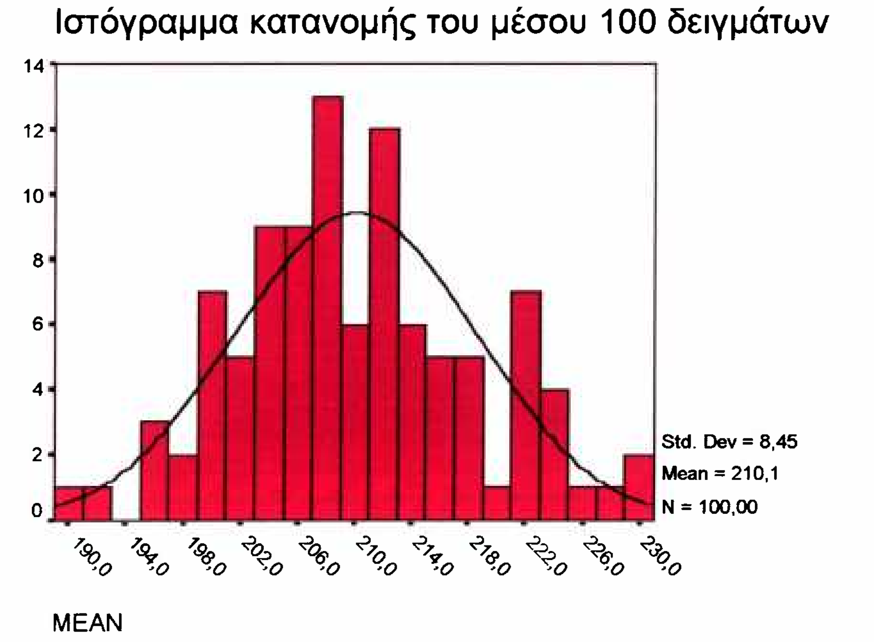
Η κατανομή της χοληστερόλης στο αίμα για όλους τους άνδρες της Ελλάδος, από 20 έως 74 ετών, μπορεί να χρησιμοποιηθεί για να επιδειχθεί μία εφαρμογή του Κεντρικού Οριακού Θεωρήματος. Κατ' αρχήν πρέπει να επιλέξουμε πολλαπλά δείγματα μεγέθους n από τον πληθυσμό και να εξετάσουμε την κατανομή των μέσων αυτών των δειγμάτων.

Αντί όμως να πάρουμε όντως δείγματα από τον πληθυσμό, μπορούμε να χρησιμοποιήσουμε τον υπολογιστή για να διεξαγάγουμε μία προσομοίωση. Σε μία προσομοίωση, ένα πρόγραμμα στον υπολογιστή χρησιμοποιείται για την αναπαραγωγή ενός πειράματος ή διαδικασίας σύμφωνα με μία συγκεκριμένη κατανομή πιθανότητας.

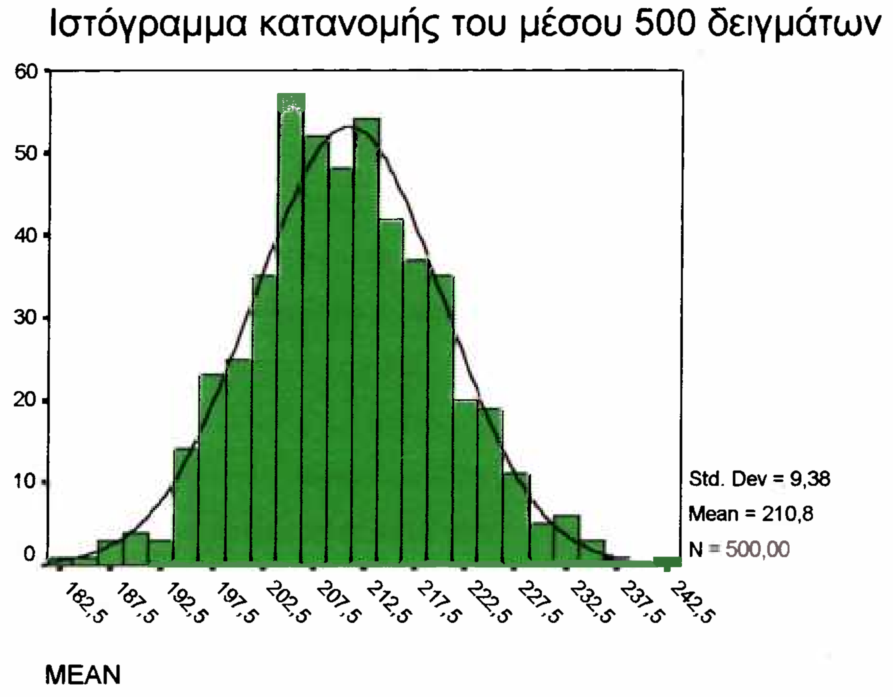
Για να επεξηγήσουμε αυτή την τεχνική παράγουμε αντίστοιχα 100, 500 και 1000 τυχαία δείγματα μεγέθους 25 από την κανονική κατανομή με μέσο μ=211mg/100ml και τυπική απόκλιση σ=46mg/100ml. Οι αριθμοί που παράγονται είναι τυχαίοι και συνεπώς διαφορετικοί κάθε φορά που εκτελείται η διαδικασία. Για να παραχθούν οι ίδιοι τυχαίοι αριθμοί κάθε φορά, θα πρέπει να δοθεί η ίδια αρχική τιμή στη γεννήτρια τυχαίων αριθμών. Κατόπιν εκτιμούμε τη μέση τιμή και την τυπική απόκλιση του πληθυσμού (οι πραγματικές τιμές του πληθυσμού είναι 211 και 46 αντίστοιχα) από κάθε ένα από αυτά τα δείγματα. Τέλος βρίσκουμε την μέση τιμή και την τυπική απόκλιση της νέας μεταβλητής, που είναι οι μέσες τιμές όλων των δειγμάτων *Στην ουσία μελετούμε την συμπεριφορά της κατανομής του μέσου τυχαίου δείγματος.*

Αυτή η διαδικασία μπορεί να γίνει με την βοήθεια των στατιστικών πακέτων SPSS και MINITAB. Εμείς κάναμε χρήση του SPSS.

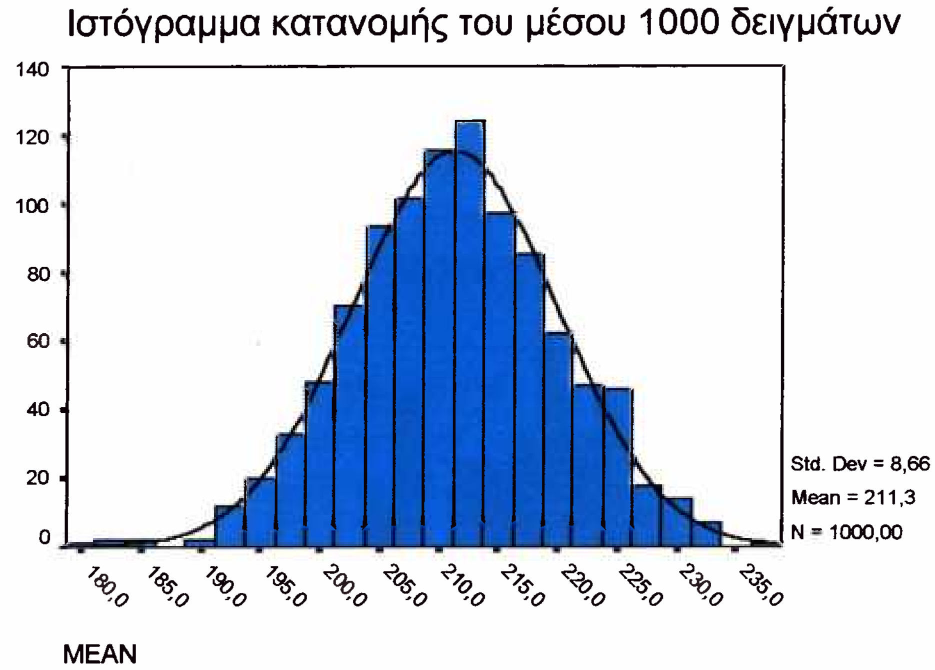
Ιστογράμματα της κατανομής του μέσου των 100, 500 και 1000 τυχαίων δειγμάτων παρουσιάζονται παρακάτω.



Διάγραμμα 6.2: *Ιστόγραμμα της κατανομής του μέσου 100 δειγμάτων*



Διάγραμμα 6.3: *Ιστόγραμμα της κατανομής του μέσου 500 δειγμάτων*



Διάγραμμα 6.4: *Ιστόγραμμα της κατανομής του μέσου 1000 δειγμάτων*

Σχόλιο:

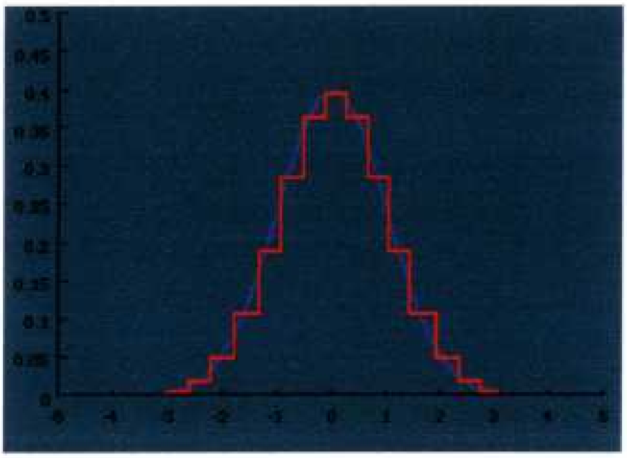
Σύμφωνα με το Κεντρικό Οριακό Θεώρημα, η κατανομή των μέσων των δειγμάτων μεγέθους 25 διαθέτει τρεις ιδιότητες. Πρώτον, ο μέσος της πρέπει να ισούται με το μέσο του πληθυσμού μ=211mg/100ml.Δεύτερο, αναμένουμε ότι η τυπική απόκλιση της κατανομής θα είναι $\frac{\sigma}{\sqrt{n}} = \frac{46}{\sqrt{25}} = 9.2\,\text{mg}/100\text{ml}$.Τέλος, η κατανομή των δειγματικών μέσων πρέπει να είναι προσεγγιστικά κανονική.

Από ότι παρατηρούμε, καθώς το πλήθος των δειγμάτων αυξάνεται, ο μέσος των δειγμάτων πλησιάζει προς τον μέσο του πληθυσμού. Αντίστοιχα καθώς το πλήθος των δειγμάτων αυξάνεται, η τυπική απόκλιση των δειγμάτων τείνει στο τυπικό σφάλμα του μέσου του πληθυσμού.

Όσο για την γραφική παράσταση παρατηρούμε,ότι καθώς το πλήθος των δειγμάτων αυξάνεται η μορφή του ιστογράμματος προσεγγίζει όλο και περισσότερο την κανονική κατανομή που υπερτίθεται του ιστογράμματος.

Όταν το πλήθος των δειγμάτων γίνει 1000,έχουμε την καλλίτερη προσέγγιση. Αν συνεχίσουμε να αυξάνουμε τις επαναλήψεις των δειγμάτων πετυχαίνουμε ακόμη καλλίτερες προσεγγίσεις. Στην προκειμένη περίπτωση δεν κρίνεται σκόπιμο να συνεχίσουμε, διότι έγιναν αντιληπτές οι ιδιότητες της κατανομής του μέσου ενός τυχαίου δείγματος.

6.6 Γιατί πολλά φυσικά ποσοτικά φαινόμενα ακολουθούν περίπου κανονική κατανομή;



Διάγραμμα 6.5: *Ιστόγραμμα της κατανομής ενός νομίσματος (εμφάνιση κεφαλής) σε 100 ρίψεις*

Η εφαρμογή του Κεντρικού Οριακού Θεωρήματος στο άθροισμα των τιμών n τυχαίων μεταβλητών βοηθά στην ερμηνεία του γεγονότος, ότι τόσα πολλά ποσοτικά φαινόμενα έχουν (τουλάχιστον κατά προσέγγιση) κανονική κατανομή.

Το Κεντρικό Οριακό Θεώρημα είναι ένα από τα πιο αξιοπρόσεκτα αποτελέσματα της θεωρίας των πιθανοτήτων. Στην απλούστερη μορφή του, το θεώρημα δηλώνει ότι το άθροισμα ενός μεγάλου αριθμού ανεξάρτητων παρατηρήσεων από την ίδια κατανομή έχει, υπό ορισμένους γενικούς όρους, μια κατά προσέγγιση κανονική κατανομή. Επιπλέον, η προσέγγιση βελτιώνεται σταθερά καθώς ο αριθμός παρατηρήσεων αυξάνεται. Το θεώρημα θεωρείται καρδιά της θεωρίας πιθανοτήτων, αν και ένα καλύτερο όνομα θα ήταν κανονικό θεώρημα σύγκλισης.

Παραδείγματος χάριν, υποθέτουμε ότι ένα συνηθισμένο νόμισμα ρίπτεται 100 φορές και ο αριθμός κεφαλών μετριέται. Αυτό είναι ισοδύναμο με το να σημειώνουμε 1 για κεφαλή και 0 για γράμματα για τον υπολογισμό του συνολικού αποτελέσματος. Κατά συνέπεια, ο συνολικός αριθμός κεφαλών είναι το άθροισμα των 100 ανεξάρτητων, όμοια κατανεμημένων τυχαίων μεταβλητών. Από το Κεντρικό Οριακό Θεώρημα, η κατανομή του συνολικού αριθμού κεφαλών θα είναι, σε έναν πολύ υψηλό βαθμό προσέγγισης, κανονική. Αυτό επεξηγείται γραφικά με να επαναλάβουμε αυτό το πείραμα πολλές φορές. Τα αποτελέσματα αυτού του πειράματος παρουσιάζονται στο παραπάνω Διάγραμμα (6.5).

Το ποσοστό που υπολογίζεται από τον αριθμό πειραμάτων τοποθετείται κατά μήκος του κάθετου άξονα, και το συνολικό αποτέλεσμα ή ο αριθμός κεφαλών τοποθετείται κατά μήκος του οριζόντιου άξονα. Μετά από έναν μεγάλο αριθμό επαναλήψεων δημιουργείται μια καμπύλη η οποία φαίνεται ως κανονική καμπύλη.

Εμπειρικά έχει παρατηρηθεί ότι τα διάφορα φυσικά μεγέθη, όπως τα ύψη των ατόμων, ακολουθούν μια περίπου κανονική κατανομή.

Ας υποθέσουμε, για παράδειγμα, ότι ενδιαφερόμαστε για το ύψος Y ενός ατόμου. Μπορούμε να θεωρήσουμε ότι το ύψος αυτό είναι το αποτέλεσμα άθροισης $Y = \sum_{i=1}^{n} X_i$ ενός μεγάλου αριθμού μικρών (απειροστών) αυξήσεων X_i, όπου κάθε μικρή αύξηση λαβαίνει χώρα μέσα σε κάποιο μικρό χρονικό διάστημα. Εύλογο είναι να υποθέσουμε ότι οι αυξήσεις X_i που αντιστοιχούν σε διαφορετικά μικρά χρονικά διαστήματα είναι αμοιβαία ανεξάρτητες και έχουν μία κοινή κατανομή (είναι ισόνομες).

Στην περίπτωση αυτή η εφαρμογή του Κεντρικού Οριακού θεωρήματος μας επιτρέπει να συμπεράνουμε ότι το ύψος $Y = S_n = \sum_{i=1}^{n} X_i$ έχει, τουλάχιστον κατά προσέγγιση μια κανονική κατανομή.

Μια προτεινόμενη εξήγηση είναι ότι αυτά τα φαινόμενα είναι άθροισμα ενός μεγάλου αριθμού ανεξάρτητων τυχαίων αποτελεσμάτων και ως εκ τούτου κατανέμονται περίπου κανονικά με βάση το Κεντρικό Οριακό Θεώρημα.

Είναι προφανές ότι η ακρίβεια της προσέγγισης της κατανομής Z_n από την κανονική κατανομή με τη βοήθεια του Κεντρικού Οριακού Θεωρήματος μεταβάλλεται με την τιμή του η. Ας υποθέσουμε, χωρίς βλάβη της γενικότητας, ότι οι μεταβλητές $X_1, X_2, \ldots, Xn$ είναι ισόνομες. Επειδή η κανονική κατανομή είναι συμμετρική γύρω από την μέση τιμή της, είναι φανερό ότι όσο περισσότερο ασύμμετρη (στρεβλή) είναι η κοινή κατανομή των τυχαίων μεταβλητών $X_1, X_2, \ldots, X_n$, τόσο μεγαλύτερο μέγεθος δείγματος απαιτείται για να επιτευχθεί μια καλή προσέγγιση της συνάρτησης κατανομής $F_{Z_n}(z)$ από την συνάρτηση κατανομής της τυποποιημένης κανονικής κατανομής.

Συνεπώς η ακρίβεια της προσέγγισης εξαρτάται από το μέγεθος του δείγματος και τη μορφή της κατανομής. Η προσέγγιση λειτουργεί καλύτερα για τις συμμετρικές κατανομές.

6.7 Γενικά Συμπεράσματα από το κεντρικό οριακό θεώρημα

Για πρακτικούς λόγους, η κύρια ιδέα του Κεντρικού Θεωρήματος Ορίου (CLT) είναι ότι ο μέσος όρος δειγμάτων παρατηρήσεων που προέρχονται από κάποιο πληθυσμό με οποιαδήποτε μορφή κατανομής είναι κατανεμημένος ως περίπου κανονική κατανομή εάν ορισμένοι όροι ικανοποιούνται. Στην θεωρητική στατιστική υπάρχουν διάφορες εκδόσεις του Κεντρικού Οριακού Θεωρήματος ανάλογα με το πώς αυτοί οι όροι διευκρινίζονται.

Οι όροι αυτοί αναφέρονται στις υποθέσεις που γίνονται για τη κατανομή του γεννήτορα πληθυσμού (πληθυσμό από τον οποίο προέρχεται το δείγμα) και την πραγματική διαδικασία δειγματοληψίας.

Μια από τις απλούστερες εκδόσεις του θεωρήματος αναφέρει ότι εάν έχουμε ένα τυχαίο δείγμα μεγέθους n (έστω, n μεγαλύτερο από 30) από έναν άπειρο πληθυσμό, με πεπερασμένη τυπική απόκλιση, τότε ο μέσος του τυποποιημένου δείγματος συγκλίνει σε μια τυποποιημένη κανονική κατανομή ή, ισοδύναμα, ο δειγματικός μέσος προσεγγίζει μια κανονική κατανομή με μέσο όρο ίσο με τον μέσο του πληθυσμού και τυπική απόκλιση ίση με τη τυπική απόκλιση του πληθυσμού που διαιρείται με την τετραγωνική ρίζα του μεγέθους του δείγματος n. Στις εφαρμογές πάντως του Κεντρικού Οριακού Θεωρήματος στα πρακτικά προβλήματα στη Στατιστική Συμπερασματολογία, οι στατιστικοί ενδιαφέρονται για το πόσο πολύ η κατά προσέγγιση κατανομή του μέσου του δείγματος, ακολουθεί μια κανονική κατανομή για τα πεπερασμένα μεγέθη δειγμάτων, από την ίδια την περιοριστική κατανομή.

Η αρκετά στενή συμφωνία με μια κανονική κατανομή επιτρέπει στους στατιστικούς να χρησιμοποιήσουν την κανονική θεωρία για τη εξαγωγή συμπερασμάτων για τις παραμέτρους του πληθυσμού (όπως ο μέσος) χρησιμοποιώντας το δειγματικό μέσο, *ανεξάρτητα από την πραγματική μορφή του γεννήτορα πληθυσμού.*

Είναι ευρέως γνωστό ότι για οποιοδήποτε γεννήτορα πληθυσμό, η τυποποιημένη μεταβλητή θα έχει κατανομή έναν μέσο όρο 0 και τυπική απόκλιση 1 κάτω από την τυχαία δειγματοληψία. Επιπλέον, εάν ο γεννήτορα πληθυσμός είναι κανονικός, τότε κατανέμεται ακριβώς ως τυποποιημένη κανονική μεταβλητή για οποιοδήποτε θετικό ακέραιο αριθμό n. Το Κεντρικό Οριακό Θεώρημα εκφράζει το αξιοπρόσεκτο αποτέλεσμα ότι, ακόμα και όταν ο γεννήτορας πληθυσμός είναι *μη-κανονικός*, η τυποποιημένη μεταβλητή είναι περίπου κανονική εάν το μέγεθος δειγμάτων είναι αρκετά μεγάλο (έστω > 30). Γενικά δεν είναι δυνατό να δηλωθούν οι συνθήκες κάτω από τις οποίες η προσέγγιση που δίνεται από τις εργασίες του Κεντρικού Οριακού Θεωρήματος τα μεγέθη δειγμάτων που απαιτούνται πριν από την προσέγγιση για να αποβεί αυτή αρκετά καλή. Γενικά, οι στατιστικοί θεωρούν ότι εάν η κατανομή του γεννήτορα πληθυσμού είναι συμμετρική και σχετικά μικρής ουράς, τότε ο δειγματικός μέσος φθάνει στην κατά προσέγγιση κανονικότητα για μικρότερα δείγματα, από ότι θα έφθανε εάν ο γεννήτορας πληθυσμός ήταν λοξός ή με μακριά ουρά.

Τώρα,θα μελετήσουμε τη συμπεριφορά του μέσου όρου δειγμάτων διαφορετικών μεγεθών που προέρχονται από ποικίλους γεννήτορες πληθυσμούς. Εξετάζοντας τις κατανομές δειγματοληψίας των μέσων των δειγμάτων που υπολογίζονται από δείγματα διαφορετικών μεγεθών που προέρχονται από ποικίλες κατανομές, αποκτούμε κάποια επίγνωση για τη συμπεριφορά του μέσου του δείγματος κάτω από εκείνους τους συγκεκριμένους όρους και επί πλέον εξετάζουμε την ισχύ των οδηγιών που προαναφέρθηκαν, για να χρησιμοποιήσουμε το Κεντρικό Οριακό Θεώρημα στην πράξη.

- Υπό ορισμένους όρους, στα μεγάλα δείγματα, η κατανομή δειγματοληψίας του μέσου του δείγματος μπορεί να προσεγγιστεί από μια κανονική κατανομή. Το μέγεθος δειγμάτων που απαιτείται για την προσέγγιση για να είναι επαρκές εξαρτάται κατά κύριο λόγο από τη μορφή της κατανομής του γεννήτορα πληθυσμού.
- Η συμμετρία (ή έλλειψη αυτής) είναι ιδιαίτερα σημαντική. Για μια συμμετρική κατανομή του γεννήτορα πληθυσμού, ακόμα κι αν είναι πολύ διαφορετική από τη μορφή μιας κανονικής κατανομής, μια επαρκής

προσέγγιση μπορεί να ληφθεί με τα μικρά δείγματα (π.χ. 10 ή 12 για την ομοιόμορφη κατανομή).

● Για τις συμμετρικές και με μικρή ουρά κατανομές του γεννήτορα πληθυσμού, ο δειγματικός μέσος φθάνει στην κατά προσέγγιση κανονικότητα για τα μικρότερα δείγματα από ότι θαέφθανε εάν ο γεννήτορας πληθυσμός ήταν λοξός και με μακριά ουρά.

●Σε μερικές ακραίες περιπτώσεις (π.χ. στη διωνυμική) απαιτούνται μεγάλα μεγέθη δειγμάτων που υπερβαίνουν τις τυπικές οδηγίες (π.χ.30) για μια επαρκή προσέγγιση.

●Για μερικές κατανομές χωρίς πρώτες και δεύτερες ροπές (π.χ Cauchy), το κεντρικό οριακό θεώρημα δεν ισχύει.

ΚΕΦΑΛΑΙΟ 7^{O}

Κανονικές προσεγγίσεις σε κατανομές

Πολλές από τις κατανομές που δεν μπορούν να αξιολογηθούν σε περιορισμένη μορφή μπορούν να προσεγγιστούν, και δεν είναι περίεργο ότι πολλές από αυτές τις προσεγγίσεις είναι βασισμένες στον κανονικό νόμο. Οι καλύτερες είναι εκείνες που είναι ασυμπτωτικές, συγκλίνουν δηλαδή υπό κάποια έννοια σε κανονικότητα. Μία κλάση κατανομών αυτού του είδους είναι εκείνες που μπορούν να αντιπροσωπευθούν (ως Binomial, Poisson, και Γάμμα κατανομές) όπως το άθροισμα των ανεξάρτητων και ισόνομα κατανεμημένων τυχαίων μεταβλητών, έτσι ώστε το Κεντρικό Οριακό Θεώρημα να μπορεί να χρησιμοποιηθεί για να παρέχει τις κατάλληλες προσεγγίσεις.

Μια άλλη πηγή κατανομών που συνεισφέρει στις κανονικές προσεγγίσεις είναι εκείνες των οποίων συναρτήσεις κατανομής μπορούν να είναι εκφρασμένες απόλυτα στις συνθήκες των ελλιπών συναρτήσεων γάμμα και βήτα αναλογιών . Η τελευταία είναι η πιθανότητα της δεξιάς ουράς της διωνυμικής (που έχει την ιδιότητα του κεντρικού οριακού θεωρήματος), και μπορεί να είναι απόλυτα συνδεμένη με τις αθροιστικές συναρτήσεις κατανομής της αρνητικής διωνυμικής, της βήτα, της t-student ,και της F κατανομής. Ισχύει λοιπόν, συχνά (αλλά όχι πάντα), η ιδιότητα, ότι οι προσεγγίσεις που είναι καλές για τη διωνυμική είναι καλές υπό κάποια έννοια για τις άλλες κατανομές.

Εμείς προσπαθήσαμε να απαριθμήσουμε όλες τις προσεγγίσεις που είναι βασισμένες στην κανονικότητα. Κάποια φτωχή προσέγγιση μπορεί να δοθεί εάν είναι ιστορικού ενδιαφέροντος, αλλά σε γενικές γραμμές, ο στόχος ήταν να παρουσιαστούν εκείνες οι προσεγγίσεις που συνδυάζουν την ακρίβεια με την απλότητα που έχει η χρήση ενός υπολογιστή ή ενός απλού αλγορίθμου . Τα κριτήρια για την ακρίβεια ή το απόλυτο και σχετικό σφάλμα αναφέρονται, όποτε μια σαφής αναφορά για την ακρίβεια που επιτυγχάνεται μπορεί να δοθεί.

Σε κάθε περίπτωση, η κατάλληλη εντολή είναι η ακόλουθη: απλές προσεγγίσεις στην αθροιστική συνάρτηση κατανομής, ακριβείς προσεγγίσεις στην αθροιστική συνάρτηση κατανομής, ακριβείς κανονικές αποκλίσεις και προσεγγίσεις στα εκατοστημόρια. Τα όρια συμπεριλαμβάνονται σε αυτό το κεφάλαιο, όταν αυτά βασίζονται στις ιδιότητες κανονικότητας.

Οι κατανομές που θα καλύψουμε σε αυτό το κεφάλαιο είναι οι εξής:Η Διωνυμική, η Poisson και η Υπεργεωμετρική.

Η βιβλιογραφία περιέχει πολύ περισσότερες πληροφορίες από αυτές που δίνονται εδώ. Εκτενείς συζητήσεις περιλαμβάνονται στα βιβλία των Johnson και Kotz (1969, 1970a, 1970b), Molenaar (1970), Peizer και Pratt (1968).

7.1 Προσεγγίση της διωνυμικής από την κανονική κατανομή

Ορισμός: Έστω g (y,n,p), η συνάρτηση πιθανότητας της διωνυμικής τυχαίας μεταβλητής Y όπου:

$$P_r(Y=y)=g(y,n,p)=\binom{n}{y}p^y(1-p)^{n-y}\quad,\quad y=0,1,\ldots,n$$

$$0<p<1\quad,\ q=1-p$$

Δηλώνουμε ότι η αθροιστική συνάρτηση κατανομής της Y με

$$G\ (y;n,p)=Pr(Y\le y)$$

και σημειώνουμε ότι $G(k,n,p)=1-G(n-k-1,n,q)$ είναι ένα αποτέλεσμα που μπορεί να δώσει μια επιλογή για κάθε προσέγγιση, στην προσπάθεια να μειωθεί το λάθος.

7.1.1 Προσέγγιση της κανονικής κατανομής μέσω της δυωνυμικής εξίσωσης $\underset{n\to\infty}{(p+q)^n}$

Πιθανότητα

Η πιθανότητα ορίζεται ως μια αναλογία του αριθμού αναμενόμενων (ευνοϊκών ή επιθυμητών) εκβάσεων προς τον αριθμό των πιθανών εκβάσεων.

Δυαδικό γεγονός

Μια ρίψη νομισμάτων είναι ένα παράδειγμα ενός δυαδικού γεγονότος. Υπάρχουν μόνο δύο πιθανές εκβάσεις (κεφάλι ή γράμματα) σε κάθε δοκιμή και η πιθανότητα κάθε έκβασης είναι 0.5. Οι δύο πιθανότητες αθροίζονται στο 1.

Ρίψη ενός "ιδανικού" νομίσματος

<u>Όλες οι πιθανές εκβάσεις</u>

Ο αριθμός πιθανών εκβάσεων, n, δίνεται από την εξίσωση $n=2^k$ όπου το k δείχνει τον αριθμό των καθοριστικών παραγόντων και η βάσης 2 αντιπροσωπεύει δύο αμοιβαία αποκλειστικές εκβάσεις που συνδέονται με κάθε καθοριστικό παράγοντα.

Η Πιθανότητα κάθε μοναδικής έκβασης δίνεται από την εξίσωση $p = \frac{1}{2^k}$

Ρίχνουμε ένα νόμισμα μία φορά : k = 1

X_1
1 : Κεφάλι
2:Γράμματα

Για έναν καθοριστικό παράγοντα, υπάρχουν δύο πιθανές εκβάσεις $2^1=2$

Πιθανότητα κάθε μοναδικής έκβασης: $\frac{1}{4} = 0.25$

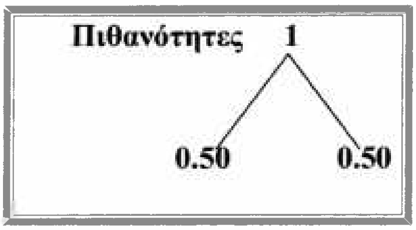
- **Αριθμός κεφαλιών:**

X_1	Κεφάλι
1	1
0	0

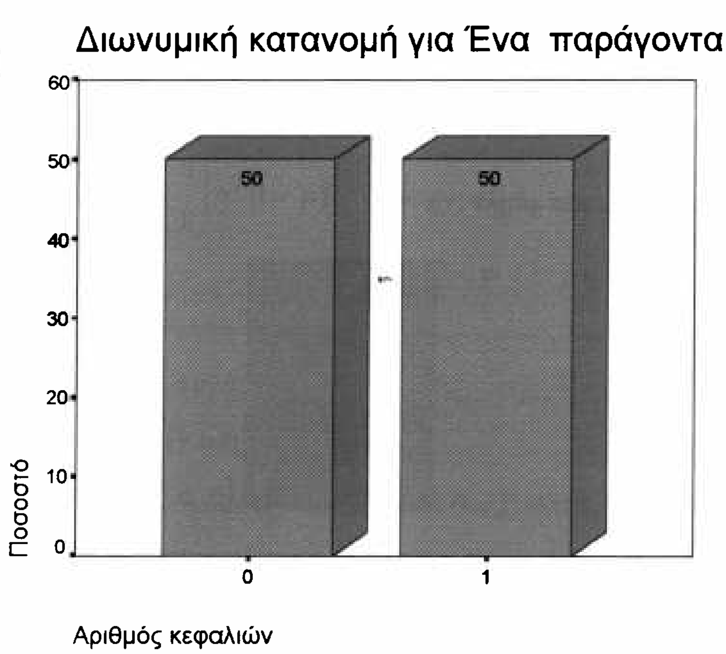
Συχνότητα και ποσοστά

Κεφάλι	Συχνότητα	Ποσοστά
1	1	½=0.50
0	1	½=0.50

❖ **Πιθανότητες που συνδέονται με το τρίγωνο του PASCAL**



Απεικόνιση Διωνυμικής κατανομής για ένα καθοριστικό παράγοντα



Διάγραμμα 7.1: *Ραμβόγραμμα Διωνυμικής κατανομής για την ρίψη, ενός ιδανικού νονίσματος*

❖ **Ρίψη δύο "ιδανικών" νομισμάτων.**

Όλες οι πιθανές εκβάσεις

X_1	X_2
1	1
1	0
0	1
0	0

Υπάρχουν 4 δυνατές εκβάσεις : $2^2=4$

Πιθανότητα κάθε μοναδικής έκβασης: $\frac{1}{4}=0.25$

Κάθε μια από τις πιθανές εκβάσεις είναι εξίσου πιθανή.

Αριθμός κεφαλιών: Υπολογισμός του αθροίσματος των γραμμών.

X_1	X_2	Κεφάλι
1+1=2		
1+0=1		
0+1=1		
0+0=0		

Σημειώνουμε ότι ένα κεφάλι μπορεί να εμφανιστεί με δύο τρόπους

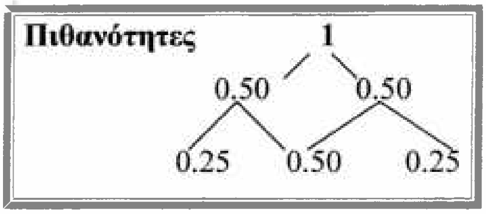
Κεφάλι	Συχνότητα	Ποσοστό
2	1	¼=0.25
1	2	2/4 =0.50
0	1	¼=0.25

Η πιθανότητα να φέρει δύο κεφάλια είναι 0.25

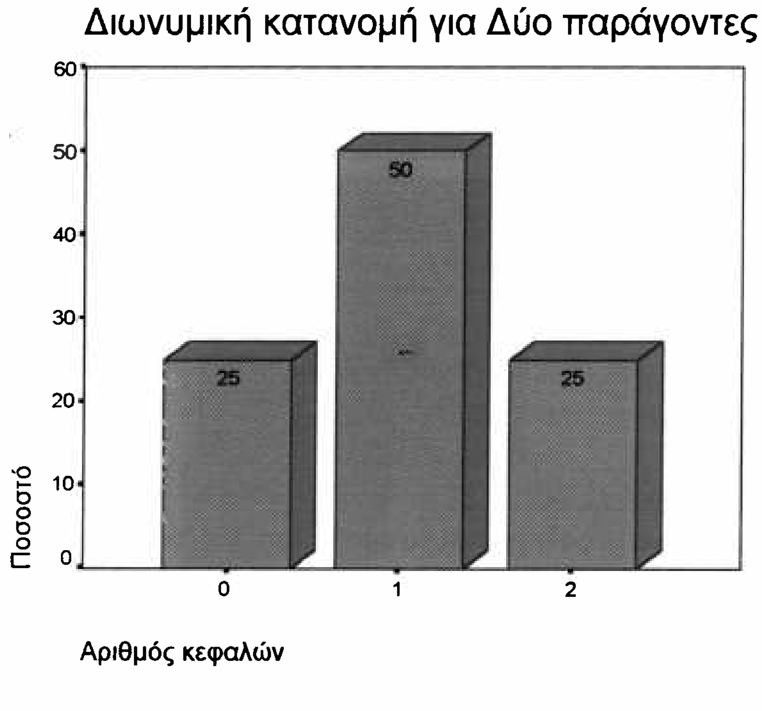
Η πιθανότητα να φέρει ένα κεφάλι είναι 0.50

Η πιθανότητα να μην φέρει κανένα κεφάλι είναι 0.25.

❖ Πιθανότητες που συνδέονται με το τρίγωνο του PASCAL



Απεικόνιση Δυωνυμικής κατανομής για δύο καθοριστικούς παράγοντες



Διάγραμμα 7.2: *Ραμβόγραμμα Διωνυμικής κατανομής για την ρίψη, δύο ιδανικών νονίσματων*

3. Ρίψη τριών "ιδανικών" νομισμάτων

0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1

Δυαδικές εκβάσεις κατασκευάσματος τριών καθοριστικών παραγόντων

Υπάρχουν 8 πιθανά αποτελέσματα: 2*2*2 = 8.

Πιθανά αποτελέσματα	X_1	X_2	X_3
1	0	0	0
2	0	0	1
3	0	1	0
4	0	1	1
5	1	0	0
6	1	0	1
7	1	1	0
8	1	1	1

Πιθανότητα κάθε μοναδικής έκβασης 1/8 =0.125

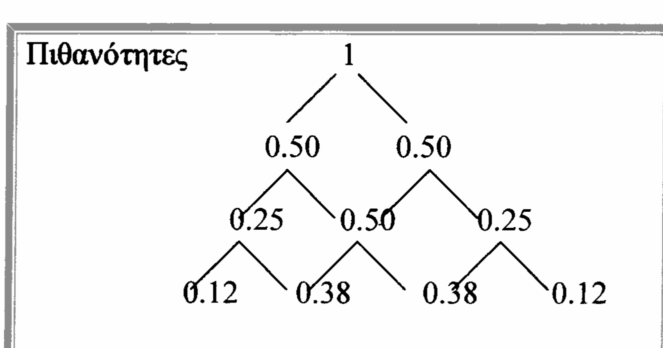
Αριθμός κεφαλιών: Υπολογίζουμε το άθροισμα των γραμμών

Πιθανά αποτελέσματα	X_1	X_2	X_3	X
1	0	0	0	0
2	0	0	1	1
3	0	1	0	1
4	0	1	1	2
5	1	0	0	1
6	1	0	1	2
7	1	1	0	2
8	1	1	1	3

Συχνότητα και ποσοστά

Αριθμός κεφαλών	Συχνότητα	Ποσοστό
0	1	0.125
1	3	0.375
2	3	0.375
3	1	0.125

Πιθανότητες που συνδέονται με το τρίγωνο του PASCAL



Απεικόνιση Διωνυμικής κατανομής για τρεις καθοριστικούς παράγοντες

Διάγραμμα 7.3: *Ραμβόγραμμα Διωνυμικής κατανομής για την ρίψη, τριών ιδανικών νονίσματων*

Οι προσεγγίσεις στην κανονική κατανομή

Η κανονική κατανομή μπορεί να περιγραφεί ως επέκταση της δυωνυμικής

εξίσωσης $\left(\underset{n \to \infty}{p + q} \right)^n$ όπου $p = q = 0.5$ και το n τείνει στο άπειρο.

Θεωρούμε σε σειρά διάταξης τα διωνυμικά αναπτύγματα με αύξηση της δύναμης από (0 έως 4), όπως παρουσιάζεται κατωτέρω.

Διωνυμικά αναπτύγματα

$$(a+b)^0 = 1$$

$$(a+b)^1 = 1a^1b^0 + 1a^0b^1$$

$$(a+b)^2 = 1a^2b^0 + 2a^1b^1 + 1a^0b^0$$

$$(a+b)^3 = 1a^3b^0 + 3a^2b^1 + 3a^1b^2 + 1a^0b^3$$

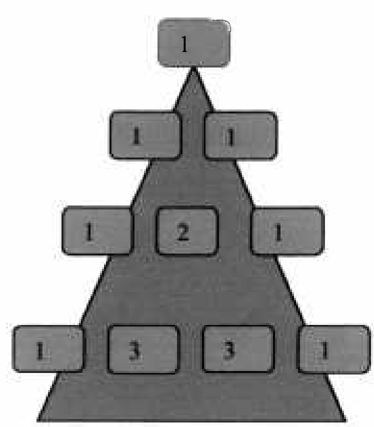
$$(a+b)^4 = 1a^4b^0 + 4a^3b^1 + 6a^2b^2 + 4a^1b^3 + 1a^0b^4$$

Για να καταλάβουμε τις αρχές πίσω από τα αναπτύγματα των διωνύμων, χωρίζουμε την δεξιά πλευρά των εξισώσεων στον παραπάνω πίνακα ανωτέρω σε δύο μέρη.
Οι εκθέτες στο πρώτο συστατικό (a) αυτής της ανάλυσης μειώνονται ενώ του (b) αυξάνονται όπως εμφανίζονται παρακάτω.

Οι συντελεστές

Το δεύτερο συστατικό περιέχει τους συντελεστές των δυωνυμικών επεκτάσεων του αποκαλούμενου τριγώνου του PASCAL, που εμφανίζονται κατωτέρω.
Παραδείγματος χάριν, σε ρίψη τριών νομισμάτων ποια είναι η πιθανότητα στο να μην έλθει κεφάλι. Ο δυωνυμικός συντελεστής είναι 1 και ο όρος του αθροίσματος είναι a^0b^3 ή $1(.5)^0(.5)^3 = .125$

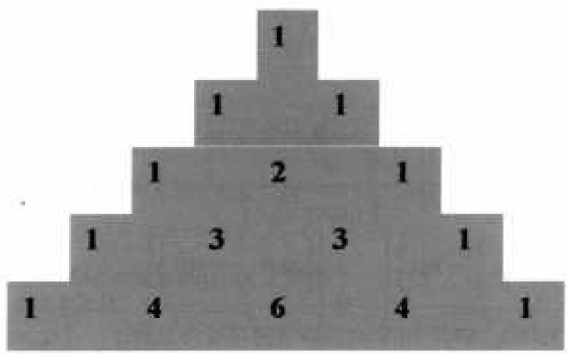
Τρίγωνο του Pascal



Κατάταξη των συντελεστών της δυωνυμικής

Το τρίγωνο του PASCAL δεν χρειάζεται να ληφθεί από την επέκταση των διωνύμων, μπορεί να κατασκευαστεί. Αρχίζουμε με 1 στην κορυφή, βάζουμε 1 σε κάθε πλευρά, και προσθέτουμε τους μέσους όρους για κάθε διαδοχική γραμμή.

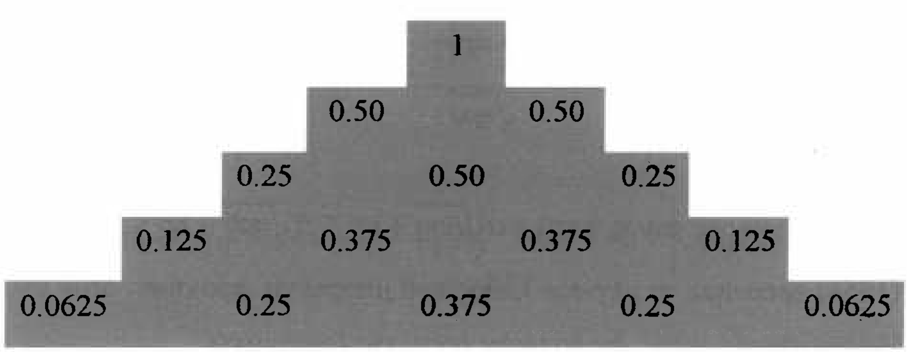
Διωνυμικοί Συντελεστές



Οι αριθμοί στη στήλη στη δεξιά πλευρά του ανωτέρω αριθμού είναι άθροισμα των συντελεστών σε κάθε γραμμή του τριγώνου του PASCAL Τα αθροίσματα είναι δυνάμεις του $2^0, 2^1, 2^2, 2^3, \ldots,$

Μετασχηματισμός των συντελεστών σε Πιθανότητες

Πιθανότητες



Αυτές οι πιθανότητες έχουν τη δυνατότητα να προσεγγίσουν τη κατανομή των πιθανοτήτων κάτω από την κανονική καμπύλη. Η ακρίβεια αυτών των προσεγγίσεων αυξάνεται όσο περισσότερες σειρές προστίθενται στο τρίγωνο του PASCAL.

7.1.2 Κλασσική κανονική προσέγγιση με διόρθωση συνέχειας

Αυτή η προσέγγιση προέρχεται άμεσα από το Οριακό θεώρημα του De Moivre – Laplace :

$$G(y,n,p) \cong \Phi\left\{\frac{\left(y+\frac{1}{2}-np\right)}{\sqrt{npq}}\right\} \quad , \mathrm{y}=0,1,2,\ldots,\mathrm{n}$$

$$\mathrm{P_r}(a \le Y \le b) \cong \Phi\left\{\frac{\left(b+\frac{1}{2}-np\right)}{\sqrt{npq}}\right\} - \Phi\left\{\frac{\left(a+\frac{1}{2}-np\right)}{\sqrt{npq}}\right\} \quad ,0 \le \mathrm{a} \le \mathrm{b} \le \mathrm{n} \quad , \qquad (7.1)$$

όπου a και b ακέραιοι

Η κανονική κατανομή είναι κατάλληλη προσέγγιση στη διωνυμική, όταν η πιθανότητα επιτυχίας είναι $0.1 \le p \le 0.9$ και $np \ge 5$ ή $n(1-p) \ge 5$.

Εάν a,b ή y δεν είναι οι ακέραιοι αριθμοί (αλλά αν ισχύει $0 \le$ a, b, y $\le$ n), πρέπει να αντικατασταθούν στην κανονική προσέγγιση από το ακέραιο μέρος των [a], [b], [y].

Μια εναλλακτική προσέγγιση δίνεται από την

$$P_r\left\{\alpha \le \frac{(Y-np)}{\sqrt{npq}} \le \beta\right\} = \Phi\left\{\frac{\left(\left[\beta\sqrt{npq}+np\right]+\frac{1}{2}-np\right)}{\sqrt{npq}}\right\} - \Phi\left\{\frac{\left(\left[\alpha\sqrt{npq}+np\right]+\frac{1}{2}-np\right)}{\sqrt{npq}}\right\}$$

όπου [x] ακέραιο μέρος του x, και $0 \le \alpha\sqrt{npq}+np \le \beta\sqrt{npq}+np \le n$

Ο Raff (1956) ερεύνησε το μέγιστο λάθος που μπορεί να προκύψει στον υπολογισμό οποιουδήποτε αθροίσματος διαδοχικών όρων της διωνυμικής.

Εάν αυτό είναι M(n,p), τότε $M(n,p) \le \frac{0.140}{\sqrt{npq}}$ πάντα, και εάν $np^{\frac{3}{2}} > 1.07$ τότε M(n, p)$\le$ 0.05.

Εάν το n είναι σταθερό, το M(n, p) μειώνεται καθώς το p αυξάνεται σε 0.5, εκτός από το ότι η ροπή αντιστρέφεται όταν το p είναι κλειστό στο μηδέν.

Εάν το p είναι σταθερό, τότε M(n, p) μειώνεται καθώς το n αυξάνεται

7.1.3 Άλλοι Επιστημονικοί τρόποι

7.1.3.1 Από το τοπικό οριακό θεώρημα του de Moivre-Laplace

$$g(y,n,p)=\binom{n}{p}p^{y}(1-p)^{n-y}\cong\Phi\left[\frac{(y-np)}{\sqrt{npq}}\right]\cong f(y;np,npq) \qquad (7.\ 2)$$

Τότε $\sum_{k=0}^{n}|g(y;n,p)-f(y;np,npq)|=\dfrac{\left\{1+4\exp\left(-\frac{3}{2}\right)\right\}}{\left(3\sqrt{2\pi npq}\right)}+0\left(n^{-1}p^{-1}q^{-1}\right)$

(Johnson και Kotz, 1969 Govindarajulu, 1965).

Εναλλακτικά, η $g(y;n,p)$ μπορεί να προσεγγιστεί όπως στην περίπτωση (7.1).

Με $b=y$, $a=y-1$

7.1.3.2 Πρόταση Molenaar : Για έναν συνδυασμό ακρίβειας και απλότητας, ο Molenaar (1970) συστήνει τους εναλλακτικούς τύπους.

$$\Phi\left[\sqrt{4k+a}\sqrt{q}-\sqrt{4n-4k+b}\sqrt{p}\right] \qquad (7.3)$$

στην προσέγγιση $G(k;n,p)$ $k=0,1,\ldots,n$, όπου $a=4$, $b=0$

Υποθέτουμε ότι p είναι κλειστό στο $\frac{1}{2}$ οπότε έχουμε:

Εαν $n=3$, $0.25\le p\le 0.75$

Εαν $n=30$, $0.40\le p\le 0.60$

Εαν $n=300$, $0.46\le p\le 0.54$

Τότε :

$$G(k;n,p)\cong\Phi\left[\sqrt{4k+3}\sqrt{q}-\sqrt{4n-4k-1}\sqrt{p}\right] \quad \text{εαν } 0.45\le\left|G(k;n,p)-\frac{1}{2}\right|\le 0.495$$

και

$$G(k;n,p)\cong\Phi\left[\sqrt{4k+2.5}\sqrt{q}-\sqrt{4n-4k-1.5}\sqrt{p}\right] \quad \text{εαν } 0.05\le|G(k;n,p)|\le 0.93$$

Εάν στην παραπάνω περίπτωση p δεν είναι κλειστό στο $\frac{1}{2}$, τότε

$$G(k;n,p) \cong \Phi\left[\sqrt{4k+4}\sqrt{q}-\sqrt{4n-4k}\sqrt{p}\right] \quad \text{εαν } 0.45 \le \left|G(k;n,p)-\frac{1}{2}\right| \le 0.495$$

$$G(k;n,1) \cong \Phi\left[\sqrt{4k+3}\sqrt{q}-\sqrt{4n-4k-1}\sqrt{p}\right] \quad \text{εαν } 0.05 \le \left|G(k;n,p)\right| \le 0.93$$

O Mo1enaar (1970) δίνει πίνακες τιμών των σχετικού σφάλματος σε αυτές τις προσεγγίσεις για ορισμένες τιμές του n μεταξύ 5 και 100, όταν p = 0.05, 0.20, 0.40 και 0.50. Το απόλυτο σφάλμα είναι $0\left(\frac{1}{\sqrt{n}}\right)$ εαν $p \neq \frac{1}{2}$, και $0\left(\frac{1}{n}\right)$ εαν $p = \frac{1}{2}$.

7.1.3.3 Οι βελτιωμένες κλασικές κανονικές προσεγγίσεις δίνονται από το:

$$G(k;n,p) \cong \Phi\left[(k+c-np)\{(n+d)pq+\delta\}^{-\frac{1}{2}}\right] \tag{7.4}$$

(στην περίπτωση (1) $c = \frac{1}{2}$, $d = \delta = 0$).

Εάν $c = \frac{(2-p)}{3}$, $d = \frac{1}{3}$, και $\delta = 0$ το μέγιστο λάθος στην προσέγγιση μειώνεται περίπου στο μισό από αυτό που υπήρχε στην περίπτωση (7.1) εκτός αν p είναι κλειστό στο 1/2, οπότε αυτή η ιδιότητα αντιστρέφεται (Gebhardt, 1969). Το πρόβλημα της επιλογής του c,d, και δ αναφέρεται στον Mo1enaar (1970) και είναι πολύπλοκο .

7.1.3.4 Η Γωνιακή προσέγγιση

(α) O Raff (1956) δίνει την προσέγγιση

$$G(k;n,p) = \Phi\left[2\sqrt{n}\left\{\arcsin\sqrt{\frac{\left(k+\frac{1}{2}\right)}{n}}\right\}-\arcsin\sqrt{p}\right] \tag{7.5}$$

Εάν M(n,p) είναι το μέγιστο λάθος όπως καθορίζεται στην (7.1), τότε

$$M(n,p) \le \frac{0.140}{\sqrt{npq}}$$

Αυτή η προσέγγιση είναι προτιμότερη από αυτή στην (7.1) εκτός αν p είναι κλειστό στο 1/2.

(β) Μία βελτιωμένη γωνιακή προσέγγιση δίνεται από

$$G(k;n,p) \cong \Phi\left[2\sqrt{n+\delta}\left\{\arcsin\sqrt{\frac{\left(k+\frac{1}{2}+\beta\right)}{n+\gamma}}-\arcsin\sqrt{p}\right\}\right] \qquad (7.6)$$

όπου $\beta = q\gamma$, $\beta = \frac{\gamma}{2}$, ή $\beta = p\gamma$ (Molenaar, 1970).

Στο (7.5) ανωτέρω, β =γ =δ =0. Κατάλληλες επιλογές των β, γ και δ οδηγούν σε μια προσέγγιση ακριβέστερη από εκείνη στην (7. 4), Molenaar (1970).

Ο Gebhardt (1969) θεωρεί $\beta = \frac{1}{6}$, $\gamma = \frac{1}{3}$, και $\delta = \frac{1}{3}$

το μέγιστο λάθος είναι το μισό στο ένα δέκατο (όταν np είναι μικρό) από αυτό στην προσέγγιση στο (7.5).

Οι Johnson και Kotz (1969) θεωρούν : $\beta = \frac{3}{8}$, $\gamma = \frac{3}{4}$, και $\delta = 0$

7.1.3.5 Η Camp- Paulson προσέγγιση είναι ακριβέστερη από εκείνες που αναφέρθηκαν παραπάνω, αλλά παράλληλα και περισσότερο πολύπλοκη (Camp, 1951):

$$G(k;n,p) \cong \Phi\left\{-\frac{x}{3\sqrt{z}}\right\}$$

$$x = \left[\frac{(n-k)p}{(k+1)q}\right]^{\frac{1}{3}}\left(9-\frac{1}{n-k}\right)+\frac{1}{k+1}-9$$

$$z = \left[\frac{(n-k)p}{(k+1)q}\right]^{\frac{2}{3}}\left(\frac{1}{n-k}\right)+\frac{1}{k+1}$$

Το λάθος είναι $0\left((npq)^{-1}\right)$για όλα τα p, και $0\left((\mathrm{npq})^{-\frac{3}{2}}\right)$ εάν $p=\frac{1}{2}$ ή $p=0.042$

(Molenaar, 1970).Το μέγιστο απόλυτο σφάλμα M(n, p), καθορισμένο με τον ίδιο τρόπο όπως στην (7.1), είναι μικρότερο από 0.0122 για οποιεσδήποτε τιμές του n και του p, ελαττώνεται καθώς το np αυξάνεται επάνω από την τιμή του 0.02 Περαιτέρω, $M(n.p)\leq\frac{0.007}{\sqrt{npq}}$(Raff, 1956). Δείτε επίσης τις παρατηρήσεις παρακάτω στην (7.1.3.6) .

7.1.3.6 Η προσέγγιση Borges (Borges, 1970)

Έστω k ακέραιος αριθμός, $0\leq k\leq n$.

Τότε

$$G(k;n,p)\cong\Phi\left(y_{k+\left(\frac{1}{2}\right)}\right)$$

$$y_k=(pq)^{-\frac{1}{6}}\sqrt{n+\frac{1}{3}}\int_p^{h(k,n)}\{s(1-s)\}^{-\frac{1}{3}}ds$$

$$h(k,n)=\frac{\left(k+\frac{1}{6}\right)}{\left(n+\frac{1}{3}\right)}$$

Το μέγιστο απόλυτο σφάλμα $M(n,p)$, που ορίζεται όπως στην (7.1), είναι

$0\left(\frac{1}{(npq)}\right)$ για όλες τις τιμές του p, και $0\left(\frac{1}{(\mathrm{npq})^{\frac{3}{2}}}\right)$ εαν $p=\frac{1}{2}$(Molenaar,1970).

Ο Gebhardt (1971) έχει παρουσιάσει το ελλιπές βήτα ολοκλήρωμα

$$\mathrm{B}\left(x;\frac{2}{3},\frac{2}{3}\right)=\int_{\frac{1}{2}}^{x}\{s(1-s)\}^{-\frac{1}{3}}ds=-\mathrm{B}\left(1-x;\frac{2}{3},\frac{2}{3}\right)$$

για όλα τα $x=0.50(0.001)1.00$ και $x=0.990(0.0001)1.000$ σε 5 δεκαδικές θέσεις

Επίσης παρουσιάζει $(pq)^{\frac{1}{6}}$ για $p=0.00(0.001)0.500$ σε 5 δεκαδικές θέσεις.

Ο Mo1enaar (1970) προτείνει την προσέγγιση :

$$\int_0^x \{s(1-s)\}^{-\frac{1}{3}} ds \cong 1.5x^{\frac{2}{3}} \frac{(60-17x)}{(60-25x)}, 0 < x \le \frac{1}{2}$$

$$= J(x) \text{ , όπως}$$

$$\text{και} \quad \cong 2.0533902 - J(1-x) \quad , \frac{1}{2} < x < 1$$

Η συμφωνία είναι καλή σε 2 δεκαδικές θέσεις εάν $x = \frac{1}{2}$, βελτίωση έχουμε σε 5 δεκαδικές θέσεις εάν x= 0.10.

Και οι δύο προσεγγίσεις του Borges και του Camp- Paulson είναι ανώτερες από εκείνες του (7.4) και του (7.6)

(arcsine) με $\beta = \frac{1}{6}$,$\gamma = \frac{1}{3}$ και $\delta = 0$ (Gebhardt 1969).

Η προσέγγιση Camp- Paulson είναι καλύτερη από την προσέγγιση Borges (αλλά όχι πάρα πολύ) για όλες τις τιμές, όταν $n \le 20$ όπως, και εάν $0.2 < p < 0.8$ όταν $n \ge 50$, (Molenaar, 1970, Gebhardt, 1969), αν και αυτά τα συμπεράσματα θεωρούν το ποσό υπολογισμού καθώς επίσης και σχετικού λάθους ως κριτήρια με βάση το μέγιστο απόλυτο σφάλμα μόνο, η προσέγγιση Borges είναι στην πραγματικότητα προτιμητέα εάν $n \le 20$ και np είναι πλησίον στο 2 ή στο 3.

7.1.3.7 Η προσέγγιση Peizer και Pratt είναι η ακριβέστερη προσέγγιση από οποιαδήποτε άλλη που περιγράφθηκε μέχρι τώρα, αλλά και η πιο δύσχρηστη. Οφείλεται στους Peizer και Pratt (1968) και είναι ακριβής στο

$$0\left(\frac{1}{(npq)^{\frac{3}{2}}}\right) \text{ εαν } p \ne \frac{1}{2} \quad \text{και στο } 0\left(\frac{1}{(npq)^{2}}\right) \text{ εαν } p = \frac{1}{2}$$

Δίνεται από το $G(k;n.p) = \Phi(z)$ όπου :

$$z = \frac{d\left[1 + qT\left\{\frac{\left(k+\frac{1}{2}\right)}{(np)}\right\} + pT\left\{\frac{\left(n-k-\frac{1}{2}\right)}{(np)}\right\}\right]^{\frac{1}{2}}}{\sqrt{pq\left(n+\frac{1}{6}\right)}}$$

$$d = k + \frac{2}{3} - \left(n + \frac{1}{3}\right)p + 0.02\left(\frac{q}{k+1} - \frac{p}{n-k} + \frac{q - \frac{1}{2}}{n+1}\right)$$

$$T(x) = \frac{\left(1 - x^2 + 2x\log x\right)}{\left(1 - x^2\right)} \quad x \neq 1$$

$$T(1) = 0$$

Δείτε και Molenaar (1970) για έναν καθαρισμό του d.Για τα ανωτέρω (Peizer και Pratt, 1968)

$$\left|G(k;n,p) - \Phi(z)\right| \leq \begin{Bmatrix} 0.001 & \text{εάν } \mathrm{k} \geq 1, \mathrm{n} - \mathrm{k} \geq 2 \\ 0.01 & \text{εάν } \mathrm{k} \geq 0, \mathrm{n} - \mathrm{k} \geq 1 \end{Bmatrix}$$

7.1.3.8 Προσέγγιση Molenaar: Η συνιστώμενη προσέγγιση, ακριβέστερη από εκείνη που δίνεται στο (7.3), έχει προταθεί από τον Molenaar (1970):

$$G(k;n,p) \cong \Phi\left\{2\sqrt{(k+1)q + \mathrm{A}} - 2\sqrt{(n-k)p + \mathrm{B}}\right\}, \ \mathrm{p} \neq \frac{1}{2}$$

$$\mathrm{A} = \left\{\frac{\left(4 - 10p + 7p^2\right)\left(k + \frac{1}{2} - np\right)^2}{(36npq)}\right\} - \frac{\left(8 - 11p + 5p^2\right)}{18}$$

$$\mathrm{B} = \left\{\frac{\left(1 - 4p + 7p^2\right)\left(k + \frac{1}{2} - np\right)^2}{(36npq)}\right\} - \frac{\left(2 + p + 5p^2\right)}{18}$$

$$G\left(k;n,\frac{1}{2}\right) \cong \Phi\left(\sqrt{2k + 2 + \beta} - \sqrt{2n - 2k + \beta}\right)$$

$$\beta = \frac{\left\{(2k + 1 - n)^2 - 10n\right\}}{(12n)}$$

Αυτή είναι ακριβής στο $0\left(\frac{1}{(npq)^{\frac{3}{2}}}\right)$, εαν $\mathrm{p} \neq \frac{1}{2}$ και σε $0\left(\frac{1}{(\mathrm{npq})^2}\right)$ εαν $\mathrm{p} = \frac{1}{2}$.

Δείτε Molenaar (1970) για τις ελαφρώς βελτιωμένες αλλά πιο δύσκολες εναλλακτικές λύσεις.

7.2 Προσέγγιση της Poisson από την κανονική κατανομή

Ορισμός : Έστω ότι g(y,λ) είναι η συνάρτηση πιθανότητας από μία Poisson τυχαία μεταβλητή Y όπου :

$$P_r(Y=y)=g(y,\lambda)=\frac{e^{-\lambda}\lambda^y}{y!}\ ,\ \ y=0,1,2,...$$

Σημειώνουμε ότι η αθροιστική συνάρτηση κατανομής της Y δίδεται από το

$$G(y;\lambda)=\Pr(Y\le y)$$

7.2.1 Προσεγγίση της Poisson από την Διωνυμική

Θα δείξουμε ότι η κατανομή Poisson προκύπτει από μία διαδικασία προσέγγισης της Διωνυμικής κατανομής.

Εστω η συνάρτηση πυκνότητας πιθανότητας της Διωνυμικής κατανομής με παραμέτρους n και p

$$P_r(Y=y)=g(y,n,p)=\binom{n}{y}p^y(1-p)^{n-y}\ ,\qquad y=0,1,...,n$$

$$0<p<1$$

Εάν η παράμετρος n τείνει στο άπειρο και το p τείνει το μηδέν, έτσι ώστε το np να είναι σταθερό και ίσο με το μηδέν, δηλαδή **np=λ** (όπου λ ο μέσος αριθμός επιτυχιών)τότε :

$$\binom{n}{y}p^y(1-p)^{n-y}\rightarrow\frac{e^{-\lambda}.\lambda^y}{y!}$$

<u>Απόδειξη:</u>

$P_r(Y=y)=\binom{n}{y}p^y(1-p)^{n-y}$ την μετασχηματίζω αν θέσω $p=\dfrac{\lambda}{n}$

$$P_r(Y=y)=\binom{n}{y}\left(\frac{\lambda}{n}\right)^{y}\left(1-\frac{\lambda}{n}\right)^{n-y}$$

$$=\frac{n!}{y!(n-y)!}\left(\frac{\lambda}{n}\right)^{\lambda}\left(1-\frac{\lambda}{n}\right)^{n-y}$$

$$=\frac{(n-y)!(n-y-1)\ldots(n-1)n}{y!(n-y)!}\left(\frac{\lambda}{n}\right)^{y}\left(1-\frac{\lambda}{n}\right)^{n-y}$$

$$=\frac{n(n-1)\ldots(n-y+1)}{y!}\frac{\lambda^{y}}{n^{y}}\left(1-\frac{\lambda}{n}\right)^{n-y}$$

$$=\frac{\lambda^{y}}{y!}\frac{n(n-1)\ldots(n-y+1)}{n^{y}}\frac{\left(1-\frac{\lambda}{n}\right)^{n}}{\left(1-\frac{\lambda}{n}\right)^{y}}$$

$$=\frac{\lambda^{y}}{y!}\cdot\frac{n}{n}\frac{n-1}{n}\cdots\cdots\frac{n-y+1}{n}\frac{\left(1-\frac{\lambda}{n}\right)^{n}}{\left(1-\frac{\lambda}{n}\right)^{y}}$$

$$=\frac{\lambda^{y}}{y!}\left(1-\frac{\lambda}{n}\right)^{n}\left\{\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-y+1}{n}\right\}\frac{1}{\left(1-\frac{\lambda}{n}\right)^{y}}$$

$$=\frac{\lambda^y}{y!}\left(1-\frac{\lambda}{n}\right)^n\left\{1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)..\left(1-\frac{y-1}{n}\right)\right\}\frac{1}{\left(1-\frac{\lambda}{n}\right)^y}$$

$$=\frac{\lambda^y}{y!}e^{-\lambda}.1=e^{-\lambda}\frac{\lambda^y}{y!}$$

διότι όταν το n τείνει στο άπειρο ισχύει:

$$\frac{\kappa}{n}\to 0 \text{ , για κ} = 1,2,\ldots,y-1$$

$$\left(1-\frac{\lambda}{n}\right)^{-y}\to 1$$

$$\left(1-\frac{\lambda}{n}\right)^{n}\to e^{-\lambda}$$

Άρα, για μεγάλες τιμές του n και μικρό p η διωνυμική κατανομή

$P_r(Y=y)=\binom{n}{y}p^y(1-p)^{n-y}$ μπορεί να προσεγγιστεί από την κατανομή Poisson

$P_r(Y=y)=e^{-np}\frac{(np)^y}{y!}$ διότι έχουμε θέσει **λ=np**

Σχόλιο: Η ωφέλεια από αυτή την προσέγγιση είναι ευδιάκριτη διότι η διωνυμική κατανομή περιέχει δύο παραμέτρους, ενώ η Poisson μόνο μία.

Εφόσον υπάρχει μία σχέση ανάμεσα στη διωνυμική και την κανονική κατανομή, συνεπάγεται ότι θα υπάρχει μία σχέση ανάμεσα στην κανονική κατανομή και την κατανομή Poisson.

Στην πραγματικότητα μπορεί να δειχθεί ότι η κατανομή Poisson πλησιάζει την κανονική κατανομή με τυποποιημένη μεταβλητή $\frac{Y-\lambda}{\sqrt{\lambda}}$ όταν το λ τείνει προς το άπειρο. Επιτυγχάνουμε καλύτερη προσέγγιση όταν $n \geq 30$ και np<5.

Θεώρημα : Έστω η τυχαία μεταβλητή Y ακολουθεί την Poisson κατανομή με παράμετρο λ, τότε για σταθερά α<β ισχύει:

$$P\left[a < \frac{Y-\lambda}{\sqrt{\lambda}} < b\right] = P\left[\lambda + a\sqrt{\lambda} < Y < \lambda + b\sqrt{\lambda}\right] \rightarrow \Phi(b) - \Phi(a) \text{ καθώς το } \lambda \rightarrow \infty$$

(Για να αποδείξουμε το συγκεκριμένο θεώρημα χρησιμοποιούμε το κεντρικό οριακό θεώρημα)

7.2.2 Κλασσική κανονική προσέγγιση με διόρθωση συνέχειας

$$G(k,\lambda) = P_r(Y \leq y) \cong \Phi\left[\frac{(k+1/2-\lambda)}{\sqrt{\lambda}}\right], \text{k} = 0,1,2,... \qquad (7.7)$$

Η ασυμπτωτική συμπεριφορά δείχνει ότι, για μεγάλο λ η διόρθωση του 1/2 πρέπει μόνο να χρησιμοποιηθεί για τις πιθανότητες μεταξύ 0.057 και 0.943 εάν λ= 10, για τις πιθανότητες μεταξύ 0.067 και 0.953 και εάν λ = 4, μεταξύ 0.073 και 0.958.

Η προσέγγιση μπορεί να υπερεκτιμήσει τις πιθανότητες λιγότερο από 0.16 και να υποτιμήσει το 1- Pr(Y≤y) όταν το τελευταίο είναι μικρότερο από 0.16. Δείτε Molenaar (1970) για τη συζήτηση αυτών και άλλων σημείων.

$$P_r(a \leq Y \leq b) \cong \Phi\left[\frac{(b+1/2-\lambda)}{\sqrt{\lambda}}\right] - \Phi\left[\frac{(a+1/2-\lambda)}{\sqrt{\lambda}}\right]$$

$$g(k,\lambda) \cong \Phi\left[\frac{(k+1/2-\lambda)}{\sqrt{\lambda}}\right] - \Phi\left[\frac{(k-1/2-\lambda)}{\sqrt{\lambda}}\right]$$

Προσέχοντας αυτά, για τους μη αρνητικούς ακέραιους αριθμούς a και b.

7.2.3 Άλλοι επιστημονικοί τρόποι

7.2.3.1 Συνιστώμενη απλή προσέγγιση, με ακρίβεια σε $0(\lambda^{-\frac{1}{2}})$(Molenaar,1970)

$$G(k,\lambda) \cong \Phi\left(2\sqrt{k+1} - 2\sqrt{\lambda}\right) \text{ για τις ουρές}$$

$$G(k,\lambda) \cong \Phi\left(2\sqrt{k+3/4} - 2\sqrt{\lambda}\right), \quad 0.09 < G(k,\lambda) < 0.94 \ , \lambda \geq 15$$

$$0.05 < G(k,\lambda) < 0.93 \ , \lambda < 15$$

Αυτή η προσέγγιση είναι ακριβέστερη από αυτή της (7.7), και είναι βασισμένη σε αυτήν των Freeman και Tukey (1950). O Molenaar (1970) συζητά $\Phi\left(2\sqrt{k+a} - 2\sqrt{\lambda+\beta}\right)$ για τις διαφορετικές επιλογές των α και β σαν μία προσέγγιση στο $G(k,\lambda)$.

7.2.3.2 Υπάρχουν επίσης σταθεροποιητές διασποράς που χρησιμοποιούν τον μετασχηματισμό $\sqrt{Y+a}$ και $\sqrt{Y} + \sqrt{Y+1}$ ο οποίος μπορεί να χρησιμοποιηθεί σε κανονικές προσεγγίσεις.

7.2.3.3 (α) Μια προσέγγιση του Makabe και του Morimura (1955) έχει όρια που δίνουν επίσης ακρίβεια $0(\frac{1}{\lambda})$, όταν $\lambda \to \infty$, Govindarajulu (1965). Έστω το k και το m είναι μη αρνητικοί ακέραιοι αριθμοί, τότε εάν $k \leq m$ και $\lambda \geq 1$.

$$G(m,\lambda) - G(k,\lambda) = \Phi(b) - \Phi(a) + \frac{\left\{\phi(a)\left(1-b^2\right) - \phi(a)\left(1-a^2\right)\right\}}{\left(6\sqrt{\lambda}\right) + R_1}$$

$$|R_1| < \frac{(0.0544)}{\lambda} + \frac{(0.0108)}{\lambda^{3/2}} + \frac{(0.2743)}{\lambda^2} + \frac{(0.0065)}{\lambda^{5/2}} + \left(1 + \left(\frac{1}{2}\right)\lambda^{-1/2}\right)\exp\left(-2\sqrt{\lambda}\right)$$

$$a = \frac{(k - 1/2 - \lambda)}{\sqrt{\lambda}} \ , \ `b = \frac{(m + 1/2 - \lambda)}{\sqrt{\lambda}}$$

Δείτε Molenaar (1970) και (γ) κατωτέρω.

(β) Cheng (1949) δίνει, ομοίως

$$G(k,\lambda) = \Phi(c) + \frac{(1-c^2)\phi(c)}{(6\sqrt{\lambda})} + R_2$$

$$c = \frac{(k + 1/2 - \lambda)}{\sqrt{\lambda}}$$

$$|R_2| < \frac{0.076}{\lambda} + \frac{0.043}{\lambda^{3/2}} + \frac{0.13}{\lambda^2}$$

(γ) Εάν $\lambda \geq 1$

$$g(k,\lambda) = \phi(y)\left\{\frac{1}{\sqrt{\lambda}} + \frac{(3y - y^3)}{6\lambda}\right\} + R_3$$

$$|R_3| = \frac{0.0748}{\lambda^{3/2}} + \frac{0.00554}{\lambda^2} + \frac{0.3724}{\lambda^{5/2}} - \frac{0.5595}{\lambda^3} + \{1 + (1/2)\lambda^{-1/2}\}\lambda^{-1/4}\exp(-2\sqrt{\lambda})$$

Δείτε τις πηγές που απαριθμούνται στο (α).

7.2.3.4 Η προσέγγιση των Wilson- Hilferty (1931) στη X^2 καταλήγει σε αυτό εδώ

$$G(k,\lambda) \cong \Phi\left[3\sqrt{k+1} - \{3\sqrt{k+1}\}^{-1} - 3\lambda^{1/3}(k+1)^{1/6}\right]$$

Εάν $G(k,\lambda) = \Phi(z)$ το σφάλμα στην προσέγγιση (Molenaar, 1970) είναι:

$$\phi(z)(3z - z^3)/(108\lambda) + 0\left(\lambda^{-\frac{3}{2}}\right)$$

7.2.3.5 Μία συνιστώμενη ακριβής προσέγγιση, με λάθος $0(\lambda^{-3/2})$ (Molenaar, 1970)

$$G(k,\lambda) \cong \Phi\left\{2\sqrt{k + \frac{(t+4)}{9}} - 2\sqrt{\lambda + \frac{(t-8)}{36}}\right\}$$

$$t = \frac{(k-\lambda+1/6)^2}{\lambda}$$

Το σφάλμα, εάν $G(k,\lambda)=\Phi(z)$ είναι:

$$\frac{\phi(z)(-6z^4+26z^2+7)}{(6480\lambda^{3/2})}+0(\lambda^{-2})$$

7.2.3.6 Η Peizer and Pratt προσέγγιση Μία άλλη ακριβής προσέγγιση (εκτός από το κοντινό k=0 όταν $\lambda \leq 5$)είναι

αυτό των Peizer and Pratt (1968) Δείτε ακόμη (Molenaar, 1970).Κατά συνέπεια

$$G(k,\lambda) \cong \Phi(u)$$

Όπου:

$$u = \left\{k-\lambda+\frac{2}{3}+\frac{\varepsilon}{(k+1)}\right\}\left[1+T\left\{\frac{\left(k+\frac{1}{2}\right)}{\lambda}\right\}\right]^{1/2}\sqrt{\lambda}$$

$$T(x)=\frac{(1-x^2+2x\log x)}{(1-x)^2},\quad x \neq 1$$

$$T(1)=0$$

Εάν οι πιθανότητες ουρών προσεγγίζονται, οι ε αυξήσεις ευνοούνται από 0.02077 έως 0.02385 καθώς το α ή το (1 - α) μειώνεται από 0.10 έως 0.005.
Εάν $G(k,\lambda) \cong \Phi(u)$ ακριβώς, το λάθος είναι (Molenaar,1970)

$$\frac{\phi(z)(-z^2+1620\varepsilon-32)}{(1620\lambda^{3/2})}+0(\lambda^{-2})$$

Όταν ε=0.02, το μέγιστο απόλυτο σφάλμα είναι 0.001 εάν $k \geq 1$ και είναι 0.01 εάν $k \geq 0$ (Peizer and Pratt, 1968)

7.2.3.7 Η προσέγγιση Riordan και Molenaar

α) Εάν $P_r(Y \le k-1) = G[(k-1),\lambda] = \Phi(x)$, τότε η ακριβής απόσταση του x δίνεται από τους (Riordan,1949 και Molenaar, 1970)

$$x = u + \frac{u^2-1}{3k^{1/2}} + \frac{7u^3-u}{36k} + \frac{219u^4-14u^2-13}{1620k^{3/2}} + \frac{3993u^5-152u^3+119u}{40320k^2} + 0\left(k^{-5/2}\right)$$

$$u = \frac{(k-\lambda)}{\sqrt{k}}$$

β) Ο Molenaar (1970) δίνει δύο παραστάσεις σε δυνάμεις του $\lambda^{-\frac{1}{2}}$, και εμείς αναπαράγουμε μία από τις δύο. Εάν $G(k,\lambda) = \Phi(z)$, τότε

$$z = v + \frac{-v^2+1}{6\lambda^{1/2}} + \frac{5v^3-2v}{72\lambda} + \frac{-249v^4+79v^2+128}{6480\lambda^{3/2}} + 0\left(\lambda^{-2}\right)$$

$$v = \frac{(k+1/2-\lambda)}{\sqrt{\lambda}}$$

7.2.3.8 Προσέγγιση με προσδιορισμένη πιθανότητα

Για μία προμελετημένη προσέγγιση που μπορεί να είναι περισσότερο ακριβής με μία προσδιορισμένη πιθανότητα α ή 1-α, όπου $0<\alpha<1/2$ και $\Phi(z_\alpha) = 1-\alpha$, έχουμε

$$G(k,\lambda) = \Phi(z) \cong \Phi(u)$$

$$u = 2\left\{k + \frac{z_a^2+11}{18}\right\}^{1/2} - 2\left\{\lambda - \frac{(z_a^2+2)}{36}\right\}^{1/2}$$

Το σφάλμα είναι : $\left(z_a^2 - z^2\right)\phi(z)\left(12\sqrt{\lambda}\right) + 0\left(\lambda^{-3/2}\right)$

Δείτε Molenaar (1970) για μία συζήτηση για αυτές και για άλλες προσεγγίσεις

7.3 Προσέγγιση της υπεργεωμετρικής από την κανονική κατανομή

Ορισμός: Έστω $g(y;n,M,N)$ η συνάρτηση πιθανότητας από μία υπεργεωμετρική τυχαία μεταβλητή Y ,όπου $g(y;n,M,N)=\dfrac{\binom{M}{y}\binom{N-M}{n-y}}{\binom{N}{n}}$

$\max(0,n-N+M)\leq y\leq \min(n,M)$, y ένας μη αρνητικός ακέραιος αριθμός.

Υποθέτουμε ότι $n\leq M\leq \frac{N}{2}$, η εξίσωση της κατανομής μπορεί να ρυθμιστεί, έτσι ώστε αυτές οι ανισότητες να ισχύουν πάντα (Molenaar, 1970).

Έστω $G(y;n,M,N)$ είναι η αθροιστική συνάρτηση κατανομής της Y.

Στην πράξη, ένα τυχαίο δείγμα λαμβάνεται χωρίς αντικατάσταση από έναν πεπερασμένο πληθυσμό μεγέθους N, του οποίου τα M αντικείμενα έχουν μια ορισμένη ιδιότητα και τα υπόλοιπα N-M αντικείμενα δεν την έχουνε. Το δείγμα είναι μεγέθους n, όπου $g(y;n,M,N)$ είναι η πιθανότητα όπου ο αριθμός των Y αντικειμένων στο δείγμα έχουν την ιδιότητα να είναι ίσα με το y.

7.3.1 Προσέγγιση της υπεργεωμετρικής από την Διωνυμική

Πρόταση : Έστω η τυχαία μεταβλητή Y ακολουθεί την υπεργεωμετρική κατανομή, τότε η αναμενόμενη τιμή της είναι $\mathrm{E}(Y)=np$, και η διασπορά της $Var(Y)=np(1-p)\dfrac{N-n}{N-1}$, όπου $p=\dfrac{M}{N}$

Παρατήρηση: Αν συγκρίνουμε την προηγούμενη πρόταση με την αντίστοιχη πρόταση για τη διωνυμική κατανομή όπου, εάν η τυχαία μεταβλητή Y ακολουθεί την διωνυμική κατανομή, τότε έχει αναμενόμενη τιμή $\mathrm{E}(Y)=np$, και διασπορά $Var(Y)=np(1-p)$.Παρατηρούμε ότι η διωνυμική και η υπεργεωμετρική κατανομή έχουν την ίδια μέση τιμή ενώ η διασπορά της υπεργεωμετρικής είναι μικρότερη από την αντίστοιχη διασπορά της διωνυμικής κατά ένα παράγοντα $\frac{N-n}{N-1}$.Ο παράγοντας αυτός λέγεται διόρθωση πεπερασμένου πληθυσμού.

Επειδή $\frac{N-n}{N-1} \underset{n\to\infty}{\to} 1$είναι προφανές ότι η διασπορά της υπεργεωμετρικής κατανομής συγκλίνει στην διασπορά της διωνυμικής κατανομής όταν το N αυξάνει.

Αποδεικνύεται ότι για αρκετά μεγάλο N η υπεργεωμετρική κατανομή μπορεί να προσεγγισθεί από την διωνυμική κατανομή.

Δηλαδή $\frac{\binom{M}{y}\binom{N-M}{n-y}}{\binom{N}{n}} \underset{N\to\infty}{\to} \binom{n}{y}\left(\frac{M}{N}\right)^{y}\left(1-\frac{M}{N}\right)^{n-y}$

για σταθερά y, n και σταθερό $\frac{M}{N}=p$

Εφόσον υπάρχει μία σχέση ανάμεσα στη διωνυμική και την κανονική κατανομή, συνεπάγεται ότι θα υπάρχει μία σχέση ανάμεσα στην υπεργεωμετρική και την κανονική κατανομή.

7.3.2 Κλασσική κανονική προσέγγιση με διόρθωση συνέχειας.

Εάν το n είναι μεγάλο, αλλά $\frac{M}{N}$δεν είναι μικρό, τότε

$$G(y;n,M,N) \cong \Phi(u) \tag{7.8}$$

όπου :

$$u=\frac{k+\frac{1}{2}-n\frac{M}{N}}{\left\{\frac{N-n}{N-1}n\frac{M}{N}\left(1-\frac{M}{N}\right)\right\}^{\frac{1}{2}}}=\frac{k+\frac{1}{2}-\mathrm{E}(Y)}{\sqrt{Var(Y)}}$$

(Hemelrijk, 1967, Molenaar, 1970).

7.3.3 Άλλοι επιστημονικοί τρόποι

7.3.3.1 Μια συνιστώμενη απλή προσέγγιση (Molenaar,1970),

$$G(y;n,M,N) \cong \Phi(u) \tag{7.9}$$

Όπου:

$$u = \begin{cases} \dfrac{2\left[\sqrt{(k+1)(N-M-n+k+1)} - \sqrt{(n-k)(M-k)}\right]}{\sqrt{N-1}} & \text{για πιθανότητες ουράς} \\[2ex] \dfrac{2\left[\sqrt{\left(k+\frac{3}{4}\right)\left(N-M-n+k+\frac{3}{4}\right)} - \sqrt{\left(n-k-\frac{1}{4}\right)\left(m-k-\frac{1}{4}\right)}\right]}{\sqrt{N}}, & \\ & 0.05 < G(k;n,M,N) < 0.93 \end{cases}$$

Αυτή η προσέγγιση είναι γενικά ανώτερη από αυτήν στο (7.8), αλλά όχι από αυτές στο (7.10) και (7.11).

7.3.3.2 Έστω $\dfrac{n}{N} \to t$, $\dfrac{M}{N} \to p$, $\dfrac{(N-M)}{N} \to q$, $h(k-np) \to \chi$ καθώς $N \to \infty$, όπου $\dfrac{1}{h} = \{Npqt(1-t)\}^{\frac{1}{2}}$. Οπότε, όταν N είναι μεγάλο (Feller 1968), $g(k;n;M,N) \cong h\phi(x)$

7.3.3.3 Nicholson (1956) Τα παραγόμενα **ανώτερα και κατώτερα** όρια σε $P_r(b \le Y \le c)$, η διαφορά μεταξύ **αυτών των ορίων είναι** $0(\sigma^{-1})$, όπου $\sigma^2 = np^*(1-p^*)\left(1-\dfrac{n}{N}\right)$, και $p^* = \dfrac{M}{N}$ δείτε επίσης Molenaar (1970).

7.3.3.4 Μία συνιστώμενη ακριβής προσέγγιση δίνεται από τον Molenaar (1985): έστω $p^* = \dfrac{M}{N}$, $s^* = \dfrac{n}{N}$, $r^2 = \dfrac{(N-n)nM(N-M)}{N^3} = \dfrac{NVar(Y)}{(N-1)}$ και

$z_k = \dfrac{\left(k+\frac{1}{2}-np^*\right)}{r}$. Οπότε

$$G(c;n,M,N) \cong \Phi(u) \tag{7.10}$$

όπου:

$$u = z_c + \left(z_c^2 - 1\right)\left[-\frac{\left(2s^* - 1\right)\left(2p^* - 1\right)}{6\tau} + z_c \frac{1 - 3s^* + 3s^{*2}}{48\tau^2}\right]$$

Αυτή η προσέγγιση συστήνεται ως περισσότερο ακριβής, εκτός αν $s^* \leq p^* \leq 0.25$ για την τελευταία περίπτωση, δείτε Molenaar (1970). Δείτε επίσης Ling και Pratt (1984)

7.3.3.5 Συνιστώμενη ακριβέστερη προσέγγιση.

Ο Ling και ο Pratt (1984) δημοσίευσαν την ακόλουθη προσέγγιση, βασισμένη στις αδημοσίευτες χειρόγραφες σημειώσεις από το David Peizer πριν από τον Αύγουστο του 1966, αυτό θα είναι κατάλληλο για να αλλαχθούν οι σημείωσεις μας για να ακολουθήσουν τις δικές τους.
Έστω η εικόνα 1.

Εικόνα 1.

y	n-y	n
M-y	(N-M)-(n-y)	N-n
M	N-M	N

→

a	b	n
c	d	m
r	s	N

Επιδιώκουμε έτσι μια προσέγγιση στην $P(Y \leq a \mid n, r, N) = G(a; n, r, N)$, η οποία είναι ισοδύναμη με την $1 - P\left(Y \leq b - \frac{1}{n}, s, N\right)$. Χωρίς απώλεια της γενικότητας ο πίνακας μπορεί να τακτοποιηθεί έτσι ώστε $a \leq d$ και α $a < b \leq c$, έτσι ώστε $2a + 1 \leq n \leq r \leq N - n$. Τρεις ρυθμίσεις γίνονται στον πίνακα, στον οποίο τα περιθώρια μπορούν να μην είναι εντελώς ίσα με τα περιθώρια σύνολα. Έστω η εικόνα 2. Ορίζουμε

Εικόνα 2.

$A = a + \frac{1}{2}$	$B = b - \frac{1}{2}$	n
$C = c - \frac{1}{2}$	$D = d + \frac{1}{2}$	m
r	s	N

$\longrightarrow$

$A' = A + \frac{1}{6}$	$B' = B + \frac{1}{6}$	$n' = n + \frac{1}{6}$
$C' = C + \frac{1}{6}$	$D' = D + \frac{1}{6}$	$m' = m + \frac{1}{6}$
$r' = r + \frac{1}{6}$	$s' = s + \frac{1}{6}$	$N' = N - \frac{1}{6}$

$\longrightarrow$

$A'' = A' + \frac{0.02}{A+0.5} + \frac{0.01}{n+1} + \frac{0.01}{r+1}$	$B'' = B' + \frac{0.02}{B+0.5} + \frac{0.01}{n+1} + \frac{0.01}{s+1}$	n'
$C'' = C' + \frac{0.02}{C+0.5} + \frac{0.01}{m+1} + \frac{0.01}{r+1}$	$D'' = D' + \frac{0.02}{D+0.5} + \frac{0.01}{m+1} + \frac{0.01}{s+1}$	m'
r'	s'	N'

$$L = A\log\left(\frac{AN}{nr}\right) + B\log\left(\frac{BN}{ns}\right) + C\log\left(\frac{CN}{mr}\right) + D\log\left(\frac{DN}{ms}\right)$$

(α)Τότε :

$$G(a;n,r,N) \cong \Phi(u) \tag{7.11}$$

όπου :

$$u = \frac{A''D'' - B''C''}{|AD - BC|}\left(2L\frac{mnrsN'}{m'n'r's'N}\right)^{\frac{1}{2}}$$

(β)

$$G(a;n,r,N) \cong \Phi(v) \qquad (7.12)$$

όπου το v καθορίζεται όπως το u στο (7.11), με $A''D'' - B''C''$ που αντικαθίσταται από το $A'D' - B'C'$.

Οι προσεγγίσεις (7.11)και (7.12) είναι ακριβέστερες από εκείνες στο (7.8) ή (7.9).

Το (7.11) είναι ακριβέστερο από το (7.12) και είναι καλύτερο από διάφορες προσεγγίσεις του τύπου Peizer που εξετάζονται στην ίδια μελέτη. Δείτε επίσης Johnson και λοιποί. (1992).

7.3.3.6 Η ακριβής κανονική αποκλίνει z που καθορίζεται από το

$$G(c;n,r,N) \cong \Phi(z)$$

ικανοποιεί :

$$z = z_c + \frac{(2s^* - 1)(2p^* - 1)(1 - z_c^2)}{6\tau}$$
$$+\left[\begin{array}{l} z_c^3\{5 - 14s^*(1-s^*) - 14p^*(1-p^*) + 38s^*(1-s^*)p^*(1-p^*)\} \\ + z_c\{-2 + 2s^*(1-s^*) + 2p^*(1-p^*) + 10s^*(1-s^*)p^*(1-p^*)\}\end{array}\right] / (72\tau^2) + 0(\tau^{-2})$$

όπου, $z_c = \dfrac{\left(c + \frac{1}{2} - np^*\right)}{\tau}$ και όπου p^*, s^* και τ ορίζονται όπως στον τρόπο

(7.3.2.4), δηλαδή : $p^* = \dfrac{M}{N}$, $s^* = \dfrac{n}{N}$, $r^2 = \dfrac{(N-n)nM(N-M)}{N^3} = \dfrac{NVar(Y)}{(N-1)}$

δείτε Molenaar (1970), που αναστρέφει επίσης αυτήν την επέκταση που εκφράζει z_c από την άποψη του z.

ΚΕΦΑΛΑΙΟ 8^{Ο}

Μίξη κατανομών

Η έκφραση μίξη κανονικών κατανομών έχει χρησιμοποιηθεί υπό δύο διαφορετικές έννοιες από διαφορετικούς συγγραφείς. Υπό την πρώτη έννοια, εάν η υπό όρους συνάρτηση πυκνότητας πιθανότητας μιας τυχαίας μεταβλητής Χ, και λαμβάνοντας υπόψη την μια ή και τις δύο παραμέτρους μ και σ^2, είναι κανονική, τότε η οριακή συνάρτηση πυκνότητας πιθανότητας της Χ, εάν αυτή υπάρχει, είναι ένα μίγμα κανονικών κατανομών. Ένα μίγμα διαμορφώνεται όταν μια ή και οι δύο παράμετροι της κανονικής κατανομής είναι συνεχείς τυχαίες μεταβλητές.

Υπό τη συχνότερα εμφανιζόμενη έννοια και αυτή που περιγράφεται εδώ, οι σύνθετες ή μικτές κατανομές προκύπτουν από τη "μίξη" δύο ή περισσότερων συστατικών κατανομών. Οι μελέτες των σύνθετων κανονικών κατανομών επιστρέφουν από αυτή την άποψη στον Karl Pearson (1894),το ζευγάρι (μ, σ) μπορεί να θεωρηθεί σαν δειγματοληψίες από μια πολυμεταβλητή (multinomial) κατανομή. Εμείς σε αυτό το κεφάλαιο θα ασχοληθούμε με την δεύτερη περίπτωση.

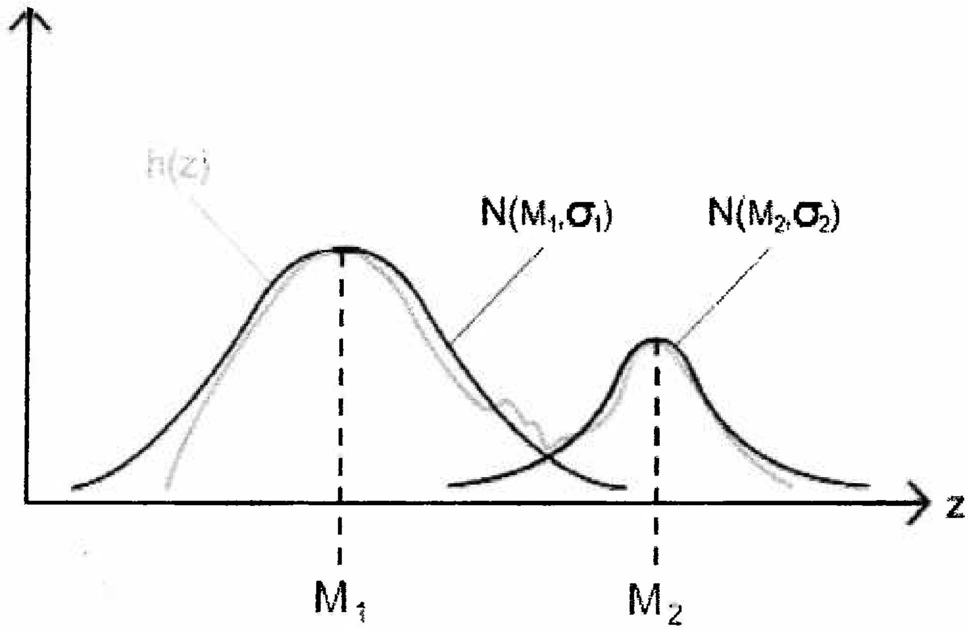
8.1 Μίξη κατανομών γενικά

Οι σύνθετες κανονικές κατανομές διαμορφώνονται με την απόδοση μιας κατανομής στην μία. ή και στις δύο, από τις παραμέτρους μ, σ της συνάρτησης πυκνότητας πιθανότητας της κανονικής κατανομής, $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

Υπάρχουν δύο ευδιάκριτα είδη κατανομών: εκείνες που λαμβάνονται με τη χρησιμοποίηση του μ ή και του σ δεδομένου ότι οι συνεχείς τυχαίες μεταβλητές είναι μεθοδολογικού και θεωρητικού ενδιαφέροντος, ενώ όταν (μ,σ), παίρνει μόνο έναν πεπερασμένο (συνήθως μικρό) αριθμό πιθανών τιμών, η προσαρμογή των αντίστοιχων κατανομών θεωρείται συνήθως ως "διαμέριση" ενός ετερογενούς πληθυσμού στα πιο ομοιογενή "μέρη.

Χρησιμοποιώντας ότι το μ είναι τυχαία μεταβλητή, παρατηρούμε ότι η κατανομή

$$\boxed{\text{Κανονική}(\mu,\sigma)\underset{\mu}{\wedge}\text{Κανονική}(\mu,\sigma')}$$



Διάγραμμα 8.1 : *Μίξη δύο κανονικών κατανομών*

είναι επίσης μια κανονική κατανομή, με αναμενόμενη τιμή μ και τυπική απόκλιση $\sqrt{\sigma^2+\sigma'^2}$, δηλαδή $\mathrm{N}\left(\mu,\sqrt{\sigma^2+\sigma'^2}\right)$. Αυτό μπορεί να αποδειχθεί από την άμεση ολοκλήρωση, ή απλά με την εκτίμηση της σύνθετης κατανομής όπως αυτή της $(\mu+Z'\sigma')+Z\sigma$ όπου το Z,Z' είναι ανεξάρτητες μοναδιαίες κανονικές μεταβλητές.

Επίσης αυτό μπορεί να παρουσιαστεί ως εξής:

$$\boxed{\text{Κανονική}(\mu,\sigma)\underset{\sigma^{-2}}{\wedge}\text{Γάμμα}\left(cX_\nu^2\right)}$$

είναι ισοδύναμος με έναν τύπο κατανομής του Pearson.

Στην πραγματικότητα :

$$(2c)^{-\frac{\nu}{2}}\left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-1}\int_0^x\left[\sqrt{2\pi}\sigma\right]^{-1}\left(\sigma^{-2}\right)^{\left(\frac{\nu}{2}\right)-1}\times\exp\left[-\left(2c\sigma^2\right)^{-1}-\left(2\sigma^2\right)^{-1}(x-\mu)^2\right]d\sigma^{-2}=$$

$$=\frac{c^{\frac{1}{2}}}{B\left(\frac{1}{2},\frac{\nu}{2}\right)}\left[1+\frac{(x-\mu)^2}{c}\right]^{-\frac{(\nu+1)}{2}} \tag{8.1}$$

(Μπορεί να διαπιστωθεί ότι η κατανομή που αποδίδεται στο σ^{-2} είναι ενός είδους μερικές φορές αποκαλούμενου βάση αναφοράς, το οποίο αποκτά μορφή αντίστροφης δήλωσης "V κατανέμεται σαν $X_\nu^2\sigma^2$να γίνει" σ^{-2}κατανέμεται ως$V^{-1}X_\nu^2$).O Teichroew (1957) έχει μελετήσει την κατανομή

$$\boxed{\text{Κανονική}\left(0,\sigma^2\right)\bigwedge_{\sigma^2}\text{Γάμμα}\left(\text{c}X_\nu^2\right)}$$

Η κατανομή έχει μια περίπλοκη μορφή, αν και η χαρακτηριστική συνάρτηση της είναι απλή $\left(1+ct^2\right)^{-\frac{\nu}{2}}$.

Η κατανομή

$$\boxed{\text{Κανονική}(\mu,\sigma)\bigwedge_{\mu}\text{Ορθογώνιος}}$$

έχει μελετηθεί από τον Bhattacharjee και λοιπούς. (1963). O Clow, ο Hansen, και McNolty (1974) έχουν θεωρήσει αυτήν την κατανομή ως κατανομή για τον παλμό-συν-στάσιμο γκαουσσιανό θόρυβο (κάτω από ένα σταθερό εύρος).

8.2 Μίξη δύο κανονικών κατανομών

Ερχόμενοι τώρα στις μίξεις ενός πεπερασμένου αριθμού (κ) κανονικών συνιστωσών, μια γενική μορφή για τη συνάρτηση πυκνότητας πιθανότητας είναι

$$g(x)=\sum_{i=1}^{k}\omega_1\left(\sqrt{2\pi}\sigma_1\right)^{-1}\exp\left[-\frac{1}{2}\left\{\frac{(x-\mu_1)^2}{\sigma_1^2}\right\}\right] \tag{8.2}$$

Οι ποσότητες $\omega_1,\omega_2,\ldots,\omega_k$ $\left(0<\omega_t:\sum_{t=1}^{k}\omega_t=1\right)$ καλούνται βάρη των συνιστωσών των κανονικών κατανομών. Θα εξετάσουμε λεπτομερώς μόνο την περίπτωση για k=2. Με τον αυξανόμενο αριθμό συνιστωσών η γενική περίπτωση γίνεται γρήγορα εξαιρετικά περίπλοκη, εν τούτοις απλοποιήσεις (όπως να υποθέσουμε όλα τα σ_t να είναι ίσα) μπορεί μερικές φορές να χρησιμοποιηθούν για να καταστήσουν την ανάλυση πιο εύχρηστη.

Θα εξετάσουμε λεπτομερώς μόνο την περίπτωση για k=2.

Με k=2, και μ_1 και μ_2 αρκετά διαφορετικά, είναι δυνατό για τον (8.2) για να αντιπροσωπεύσει τις δικόρυφες κατανομές [Helguero (1904) Prasad (1955) Teichroew (1957)].

Μια συστηματική μελέτη των όρων κάτω από την οποία αυτό είναι έτσι έχει γίνει από Eisenberger (1964) [δείτε επίσης Wessels (1964)]. Συνοψίζει τα αποτελέσματά του ως εξής:

1.Εάν

$$(\mu_1-\mu_2)^2<\frac{27\sigma_1^2\sigma_2^2}{4(\sigma_1^2+\sigma_2^2)}$$

η κατανομή δεν μπορεί να είναι δικόρυφη (ειδικότερα, εάν $\mu_1=\mu_2$).

2.Εάν

$$(\mu_1-\mu_2)^2>\frac{8\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$$

υπάρχουν τιμές ω_1 και ω_2 (ω_2=1-ω_1) για τις οποίες η κατανομή είναι δικόρυφη.

3. Για κάθε ένα σύνολο των τιμών των $\mu_1,\mu_2\sigma_1$ και σ_2 υπάρχουν τιμές των ω_1 και ω_2, για τις οποίες η κατανομή είναι μονοκόρυφη. (Αυτό είναι αρκετά εμφανές, εάν θεωρήσουμε ότι $\omega_1=0$ ή $\omega_1=1$, παίρνουμε μια κανονική κατανομή, που είναι μονοκόρυφη.

Πίνακες των ροπών του (8.2) με k = 2 [τυπική απόκλιση σ^{-1},a_3 και (α_4-3) έως 3 δεκαδικές θέσεις για $\frac{\omega_1}{\omega_2}=0.1, 0.9$ και 1.0, $\frac{\sigma_2}{\sigma_1}=1.0(0.5)3.0$ και $\frac{(\mu_2-\mu_4)}{\sigma_1}=0.0(0.5)3.0$] έχουν δοθεί από τον Linders (1930). Εάν ω_1 είναι σχεδόν 1 και έτσι ω_t, (για t> 1) είναι μικρό, η μικτή κατανομή (8.2) καλείται μερικές φορές παραλλαγμένη (contaminated) κανονική κατανομή [Tukey (1949)].

Έχει χρησιμοποιηθεί ως πρότυπο για να αξιολογήσει τις δοκιμές για την απόρριψη των απομακρυσμένων παρατηρήσεων, που λαμβάνονται στα δείγματα από μια υποθετικά κανονική κατανομή, και επίσης στις αυτοδύναμες μελέτες για τις διάφορες διαδικασίες συμπερασματολογίας.

8.2.1 Εκτίμηση των παραμέτρων ω_1, μ_1, μ_2, σ_1 και σ_2

Στη γενική περίπτωση (8.2), η r τάξεως ροπή του X γύρω από το μηδέν είναι

$$\mu_r'(X) = \sum_{t=1}^{k} \mathrm{E}\left[X^r \mid \mu_t, \sigma_t\right] = \sum_{t=1}^{k} \mathrm{E}\left[(\mu_t + Z\sigma_t)^r\right] = \sum_{j=0}^{\left[\frac{r}{2}\right]} \binom{r}{2j} \mathrm{E}\left[Z^{2j}\right] \sum_{t=1}^{k} \mu_t^{r-2j} \sigma_t^{2j}$$

όπου το Z είναι μια κανονική μεταβλητή μονάδων (γνωρίζουμε ότι $\mathrm{E}\left[Z^j\right] = 0$ εάν το j είναι περιττός).

Υποθέτουμε ότι το k είναι ίσο με το 2, και ότι θέλουμε να υπολογίσουμε τις 5 παραμέτρους ω_1 (ω_1=1-ω_2) και μ_1, μ_2, σ_1 και σ_2 με τη μέθοδο των ροπών. Πέντε ροπές θα απαιτηθούν.

Από τις εξισώσεις

$$\begin{aligned}
\mu_1' &= \omega_1\mu_1 + (1-\omega_1)\mu_2 \\
\mu_2' &= \omega_1\left(\mu_1^2 + \sigma_1^2\right) + (1-\omega_1)\left(\mu_2^2 + \sigma_2^2\right) \\
\mu_3' &= \sum_{t=1}^{2} \omega_t\left(\mu_t^3 + 3\mu_t\sigma_t^2\right) \\
\mu_4' &= \sum_{t=1}^{2} \omega_t\left(\mu_t^4 + 6\mu_t^2\sigma_t^2 + 3\sigma_t^4\right) \\
\mu_5' &= \sum_{t=1}^{2} \omega_t\left(\mu_t^5 + 10\mu_t^3\sigma_t^2 + 15\mu_t\sigma_t^4\right)
\end{aligned} \qquad (8.3)$$

προσπαθούμε να βρούμε τιμές για τα ω_1 και μ_1, μ_2, σ_1 και σ_2

Αυτό το πρόβλημα εξετάστηκε από τον Pearson το 1894. Στη συνέχεια διάφορες βελτιώσεις έχουν πραγματοποιηθεί [Charlier και Wicksell (1924), Helguero (1905)]. Τα έγγραφα από Molenaar (1965) και Cohen (1967) δίνουν μια χρήσιμη περιγραφή αυτής της εργασίας. Η ακόλουθη περίληψη είναι βασισμένη σε πληροφορίες από αυτά τα έγγραφα.

Θέτοντας $\theta_j = \mu_j - \mu_1'$ $(j = 1,2)$ οι κεντρικές ροπές $\mu_2, \mu_3 \mu_4$ και μ_5 λαμβάνονται από (8.3) με την αντικατάσταση του μ_t από το θ_t $(t = 1,2)$ Από τις προκύπτουσες εξισώσεις **μπορεί να εξαχθεί μια εξίσωση του ένατου** βαθμού για $\phi = \theta_1\theta_2$:

$$\sum_{j=0}^{9} \alpha_j \phi^j = 0 \qquad (8.4)$$

όπου:

$$\alpha_0 = -24\mu_3^6$$

$$\alpha_1 = -96\mu_3^4\kappa_4$$

$$\alpha_2 = -63\mu_3^2\kappa_4^2 - 72\mu_3^3\kappa_5$$

$$\alpha_3 = 288\mu_3^4 - 108\mu_3\kappa_4\kappa_5 + 27\kappa_4^3$$

$$\alpha_4 = 444\mu_3^2\kappa_4 - 18\kappa_5^2$$

$$\alpha_5 = 90\kappa_4^2 + 72\mu_3\kappa_5$$

$$\alpha_6 = 36\mu_3^2$$

$$\alpha_7 = 84\kappa_4$$

$$\alpha_9 = 24$$

με

$$\kappa_4 = \mu_4 - 3\mu_2^2$$

$$\kappa_5 = \mu_5 - 10\mu_2\mu_3$$

Στην εφαρμογή, οι τιμές των μ_r αντικαθίστανται από τις δειγματικές τιμές αυτών των ροπών.

Επειδή η εξίσωση (8.4) μπορεί να έχει τουλάχιστον εννέα ρίζες, μπορεί να υπάρξει δυσκολία στην επιλογή της "σωστής " ρίζας. Έκτοτε η μ_1' βρίσκεται μεταξύ μ_1 και μ_2, αυτό ακολουθεί (εκτός αν $\mu_1 = \mu_2$,) ότι θ_1 και θ_2 είναι αντιθέτων πρόσημων, και έτσι $\phi = \theta_1\theta_2 < 0$. Ως εκ τούτου μόνο αρνητικές ρίζες στο (8.4) χρειαστήκαμε να εξετάσουμε. Για την ευκολία του υπολογισμού η ακόλουθη μέθοδος εμφανίζεται να είναι κατάλληλη.

Εάν η τιμή του $\phi' = \theta_1 + \theta_2$ είναι γνωστή τότε ϕ είναι μια αρνητική ρίζα της κυβικής εξίσωσης

$$6\phi^3 - 2\phi'^2\phi^2 + (3\kappa_4 - 4\phi'\mu_3)\phi + \mu_3^2 = 0 \qquad (8.5)$$

και αυτή η εξίσωση έχει μόνο μια τέτοια ρίζα. Χρησιμοποιώντας το ϕ' και την τιμή του ϕ που βρήκαμε από το (8.5), οι τιμές θ_1 και θ_2 μπορούν να καθοριστούν, και από αυτά τα μ_1, μ_2 υπολογίζονται ως εξής:

$$\mu_j = \overline{X} + \theta_j, \quad j=1,2 \tag{8.6}$$

($\overline{X}$ είναι ο μέσος του δείγματος) και ω_1 ,ως

$$\omega_1 = \frac{\theta_2}{\theta_2 - \theta_1} \tag{8.7}$$

Τελικά έχουμε :

$$\sigma_j^2 = \frac{1}{3}\theta_j\left(2\phi' - \frac{\mu_3}{\phi}\right) + \mu_3 - \theta_j^2, \qquad j = 1,\ 2 \tag{8.8}$$

Χρησιμοποιώντας τις τιμές της παραμέτρου που αποκτήθηκαν έτσι, μια τιμή για το μ_5 μπορεί να υπολογιστεί ως $\mu_5(\phi')$. Από την αντίστροφη παρεμβολή από μια σειρά τέτοιων τιμών της $\mu_5(\phi')$, μια τιμή για το ϕ' (και ως εκ τούτου ένα σύνολο τιμών και για τις πέντε παραμέτρους) μπορεί να υπολογιστεί.

[Αυτό δεν αποκλείει τη δυνατότητα ότι περισσότερες από μια τιμές του $\mu_5(\phi')$ ισούνται με την τιμή του δείγματος, έτσι αυτό μπορεί ακόμα να είναι απαραίτητο για να κάνει διάκριση μεταξύ τέτοιων τιμών. Ο Pearson (1894) πρότεινε να επιλεχτεί η τιμή που δίνει την πιο στενή συμφωνία μεταξύ των έκτων ροπών δείγματος και πληθυσμού. Είναι εύκολο, εντούτοις, να σκεφτεί κάποιος άλλα κριτήρια, για παράδειγμα επιλέγοντας την τιμή που δίνει τη μικρότερη τιμή της X^2 ή κάποιο άλλο κριτήριο καλής προσαρμογής].

Λαμβάνοντας υπόψη την πιθανή ανακρίβεια στον υπολογισμό της έκτης κεντρικής ροπής, φαίνεται προτιμητέο να χρησιμοποιηθούν οι πρώτες και οι τρίτες απόλυτες κεντρικές ροπές, ν_1 (μέση απόκλιση) και ν_3, μαζί με τη διασπορά μ_2. Από τις εξισώσεις

$$\nu_1 = \left[\omega_1\sigma_1 + (1-\omega_1)\sigma_2\right]\sqrt{\frac{2}{\pi}}$$

$$\mu_2 = \omega_1\sigma_1^2 + (1-\omega_1)\sigma_2^2$$

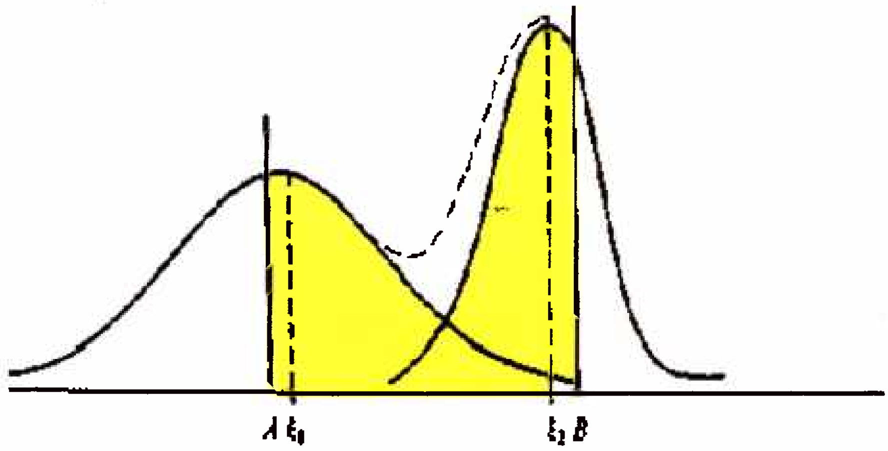
$$\nu_3 = \left[\omega_1\sigma_1^3 + (1-\omega_1)\sigma_2^3\right]\left(2\sqrt{\frac{2}{\pi}}\right)$$

παίρνουμε σ_1, σ_2 ως ρίζες της εξίσωσης

$$\left(\mu_2 - \frac{1}{2}\pi\nu_1^2\right)z^2 - \left(\frac{1}{2}\sqrt{\frac{\pi}{2}}\nu_3 - \sqrt{\frac{\pi}{2}}\nu_1\mu_2\right)z + \frac{1}{4}\pi\nu_1\nu_3 - \mu_2^2 = 0$$

Μερικές φορές οι απλούστερες διαδικασίες μπορούν να δώσουν επαρκή αποτελέσματα. Εάν η διασπορά μεταξύ των μέσων $|\mu_1 - \mu_2|$ είναι αρκετά μεγάλη, τότε οι αριστερές και δεξιές ουρές της κατανομής προέρχονται σχεδόν εξ ολοκλήρου από τα χωριστά (και διαφορετικά) συστατικά του μίγματος. Στο διάγραμμα 8.1 απεικονίζεται μια τέτοια κατάσταση, με $\mu_1 < \mu_2$ (και $\sigma_1 < \sigma_2$).Σε τέτοιες περιπτώσεις μια περικομμένη κανονική κατανομή μπορεί να προσαρμοστεί σε κάθε ουρά χωριστά. Αυτό δίνει τις εκτιμήσεις των μ_1, μ_2, σ_1 και σ_2. Τέλος, το ω_1 καθορίζεται από την εξίσωση:

$$\omega_1\mu_1 + (1-\omega_1)\mu_2 = \overline{X}$$



Διάγραμμα 8.2 : *Μίξη δύο κατανομών, όπου δεδομένα στην περιοχή κάτω από το A χρησιμοποιήθηκαν για την εκτίμηση του συνθετικού με αναμενόμενη τιμή* μ_1*(στο σχήμα* ξ_1*) και δεδομένα άνωθεν του B χρησιμοποιήθηκαν για την εκτίμηση του συνθετικού με αναμενόμενη τιμή* μ_2*(στο σχήμα* ξ_2*)*

Μια σημαντική δυσκολία σε αυτήν την μέθοδο είναι η επιλογή των σημείων της αποκοπής. Κάποιος έλεγχος είναι δυνατός με το να κινήσει τα σημεία προς το εσωτερικό, προς τον όγκο της κατανομής, εφ' όσον οι εκτιμητές των μ_j παραμένουν "εύλογα" συνεπείς. Διάφορες γραφικές μέθοδοι εκτίμησης έχουν αναπτυχθεί [Molenaar (1965), Taylor (1965)]. Ο Rivest (1981) έχει συζητήσει το άθροισμα δύο παραλλαγμένων κανονικών.

8.2.2 Συμμετρικότητα

Εάν $\mu_1 = \mu_2$, τότε η κατανομή είναι συμμετρική για οποιαδήποτε τιμή των ω_1. Μια συμμετρική κατανομή επίσης πετυχαίνουμε (ακόμη με $\mu_1 \neq \mu_2$) εάν $\sigma_1 = \sigma_2$ και $\omega_1 = \frac{1}{2}$. Αυτό μπορεί να διακριθεί από τη συμμετρική κατανομή που λαμβάνεται με $\mu_1 = \mu_2$ (και $\sigma_1 \neq \sigma_2$), μιας και στην προηγούμενη περίπτωση $\kappa_4 < 0$ ενώ στην τελευταία περίπτωση $\kappa_4 > 0$.

Συνεπώς η μικτή κανονική κατανομή είναι συμμετρική σε δύο περιπτώσεις:

(α) $\omega = \frac{1}{2}$ και $\sigma_1 = \sigma_2 = \sigma$ ή

(β) $\mu_1 = \mu_2$

Στην (β) περίπτωση η κατανομή είναι πάντοτε μονοκόρυφη. Στην περίπτωση (α) είναι μονοκόρυφη εάν $|\mu_1 - \mu_2| \leq 2\sigma$, και δικόρυφη εάν $|\mu_1 - \mu_2| > 2\sigma$ (Cohen, 1967)

Οι εκτιμητές ροπής για την περίπτωση $\mu_1 = \mu_2$ έχουν συζητηθεί λεπτομερώς στην περίπτωση που $\mu_1 = \mu_2$ από τον Agard (1961). Μια διαδικασία εκτίμησης μέγιστης πιθανότητας, για την περίπτωση που $\sigma_1 = \sigma_2$, περιγράφηκε από τον Molenaar (1965), ο οποίος κατασκεύασε επίσης ένα πρόγραμμα υπολογιστών για αυτήν την διαδικασία.

Από το 1970 η εκτενής εργασία έχει συνεχιστεί όσον αφορά τα συμπεράσματα, τη διαμόρφωση, και τις πτυχές εφαρμογής στο μίγμα-κανονικών κατανομών. Τα βιβλία από τους McLachlan και Basford (1987) παρέχουν επιμελημένες συζητήσεις σχετικά με όλες αυτές τις εξελίξεις.

ΚΕΦΑΛΑΙΟ 9^{O}

Επίλογος

Στη διεξαχθείσα εργασία επιχειρήθηκε η μελέτη της κανονικής κατανομής, μία από τις πιο βασικές και περισσότερο γνωστές κατανομές, όπως παρουσιάζεται στην ελληνική και ξένη βιβλιογραφία.

9.1 Συμπεράσματα

Συνοπτικά τα κύρια συμπεράσματα αυτής της μελέτης είναι τα ακόλουθα:

■ Λόγω της μεγάλης σοβαρότητας και χρησιμότητας της αρκετοί ερευνητές ασχολήθηκαν με την ιστορική προέλευση της. Τελικά επεκράτησαν δύο απόψεις σχετικά με την κανονική κατανομή η Αγγλική Σχολή της Βιομετρίας και η Ρώσικη σχολή.

■ Αρχικά υπήρξε αρκετός σκεπτικισμός σχετικά με την συμφωνία της θεωρίας και της φύσης.

■Η πρώτη ιστορία της κανονικής κατανομής είναι κατά ένα μεγάλο μέρος η ιστορία των αρχών της Στατιστικής ως επιστήμη.

■ Η ονομασία έγινε από πολλούς ερευνητές συμπεριλαμβανομένου του Laplace,του Gauss, του Quetelet, του Maxwell, του Galton, του Person. Τελικά το όνομα κανονικός νόμος ή κανονική κατανομή κέρδισε την αποδοχή της Αγγλικής σχολής.

■ Όταν λέμε ότι ένα ιδιαίτερο χαρακτηριστικό πληθυσμών κατανέμεται κανονικά, σημαίνει τα εξής:

1. Η συχνότερη τιμή σε μια κανονική κατανομή είναι ο μέσος όρος, με τις μισές από τις τιμές να μειώνονται κάτω του μετρίου και τις μισές επάνω από αυτό.

2.Η κανονική καμπύλη, αποκαλούμενη συχνά καμπύλη καμπάνας, είναι τέλεια συμμετρική. Επομένως ο μέσος (ο αριθμητικός μέσος όρος),η επικρατούσα τιμή (η συχνότερη τιμή), και η διάμεσος(η μεσαία τιμή) θα συμπέσουν στο κέντρο της καμπύλης (στο υψηλότερο σημείο).

3. Οι συχνότητες θα μικραίνουν από κάθε πλευρά του υψηλού κεντρικού σημείου.

4. Επειδή τα δύο μισά από κάθε πλευρά του κέντρου είναι ακριβώς συμμετρικά, η συχνότητα των τιμών επάνω από το μέσο όρο θα ταιριάξει με ακριβώς τις συχνότητες των τιμών κάτω από το μέσο όρο, υπό τον όρο ότι οι αποστάσεις μεταξύ των τιμών από το μέσο είναι ίδιες. Κατά συνέπεια, η συχνότητα μιας τιμής 3 μονάδες δεξιά του μέσου όρου θα είναι ίδια με τη συχνότητα μιας τιμής 3 μονάδες αριστερά του μέσου όρου.

5. Η συνολική συχνότητα όλων των τιμών στον πληθυσμό θα περιληφθεί από την περιοχή κάτω από την καμπύλη. Αυτό είναι αρκετά προφανές, δεδομένου ότι η συνολική περιοχή κάτω από την καμπύλη αντιπροσωπεύει όλα τα πιθανά περιστατικά κάποιου χαρακτηριστικού.

6.Οι διάφορες περιοχές κάτω από την καμπύλη επομένως θα δείξουν το ποσοστό της συνολικής συχνότητας. Παραδείγματος χάριν, 50 τοις εκατό της περιοχής κάτω από την καμπύλη βρίσκονται αριστερά του μέσου όρου (δηλαδή τα μισά από όλα τα κανονικά κατανεμημένα αποτελέσματα θα εμπέσουν σε αυτήν την περιοχή), και 50 τοις εκατό της περιοχής κάτω από την καμπύλη βρίσκονται δεξιά του μέσου όρου. Επομένως, 50 τοις εκατό όλων των αποτελεσμάτων θα βρεθούν αριστερά και 50 τοις εκατό δεξιά του μέσου όρου. Οι ίσες περιοχές κάτω από την καμπύλη αντιπροσωπεύουν την ισότητα στις συχνότητες.

7. Οι κανονικές καμπύλες μπορούν να έχουν διαφορετικές μορφές (π.χ ψηλές και λεπτές, κοντές και χαμηλές, και τα λοιπά). Εκείνο που θα καθορίζει την γενική μορφή της συμμετρικής καμπύλης θα είναι η τιμή του μέσου και της τυπικής απόκλισης του πληθυσμού (αυτά θα καθορίζουν την μορφή με ο ίδιο τρόπο που το κέντρο και η ακτίνα καθορίζουν έναν κύκλο). Αλλά τα γενικά χαρακτηριστικά που απαριθμούνται ανωτέρω θα παραμείνουν τα ίδια.

■ Είναι ιστορικού ενδιαφέροντος να σημειωθεί ότι ο κατάλογος από μερικούς από τους πιο παλαιούς γνωστούς πίνακες σχετικά με την κανονική κατανομή είναι υψηλής ακρίβειας. Αν και αυτοί ήταν επιρρεπή σε σφάλματα, είναι αξιοπρόσεκτο ότι τέτοιοι βαθμοί ακρίβειας λήφθηκαν χωρίς τη βοήθεια των υπολογιστών και, ως επί το πλείστον, με τη χρήση των πιο πρωτόγονων υπολογιστών τσέπης.

■ Ένας αριθμός συγγραφέων έχει επιδιώξει να αντικαταστάσει την όψιν μιας κανονικής κατανομής από μια άλλη κατανομή, και έπειτα να προσεγγίσει την κανονική αθροιστική συνάρτηση κατανομής και τα ποσοστιαία σημεία από αυτή. Οι επιτυχέστερες αυτών των προσπαθειών είναι του Burr και του Weibull.

■ Τα τελευταία χρόνια έχει υπάρξει ουσιαστική αύξηση του ερευνητικού ενδιαφέροντος σχετικά με τα χαρακτηριστικά των κανονικών κατανομών. Είμαστε ανίκανοι να παρέχουμε ένα περιεκτικό, ή ακόμα και πλήρως αντιπροσωπευτικό, απολογισμό στο περιορισμένο χώρο. Μια αρκετά μεγάλη αναλογία των πιο πρόσφατων αποτελεσμάτων είναι μικρής αξίας εφαρμοσμένης εργασίας. Η εκμετάλλευση των πιο πρόωρων αποτελεσμάτων δεν έχει γίνει ακόμη.

■ Το άθροισμα ενός μεγάλου αριθμού ανεξάρτητων παρατηρήσεων από την ίδια κατανομή έχει, υπό ορισμένους γενικούς όρους, μια κατά προσέγγιση κανονική κατανομή. Επιπλέον, η προσέγγιση βελτιώνεται σταθερά καθώς ο αριθμός παρατηρήσεων αυξάνεται.

■ Εμπειρικά έχει παρατηρηθεί ότι τα διάφορα φυσικά φαινόμενα, όπως το ύψος και το βάρος των ανθρώπων, ο δείκτης νοημοσύνης, το μήκος, ο όγκος, η επίδοση, οι βαθμοί και άλλα των ατόμων, ακολουθούν περίπου κανονική κατανομή.

■ Η ανάπτυξη των υπολογιστών και οι δυνατότητες που μας παρέχουν τα στατιστικά πακέτα μας επιτρέπουν να επαναλαμβάνουμε κάποιες διαδικασίες πολλές φορές 100,1000,10000 και λοιπά, πράγμα αδύνατο να γίνει με υπολογισμούς με το χέρι, και να βγάλουμε κάποια συμπεράσματα.
Ακόμη μπορούμε να χρησιμοποιήσουμε τον υπολογιστή για να διεξαγάγουμε μία προσομοίωση. Σε μία προσομοίωση, ένα πρόγραμμα στον υπολογιστή χρησιμοποιείται για την αναπαραγωγή ενός πειράματος ή διαδικασίας σύμφωνα με μία συγκεκριμένη κατανομή πιθανότητας.

■ Στο παράδειγμα που έγινε για την κατανόηση της κατανομής δειγματικού μέσου επιλέξαμε κατ' επανάληψη δείγματα μεγέθους n από τον πληθυσμό και εξετάσαμε την κατανομή των μέσων αυτών των δειγμάτων.
Παρατηρήσαμε καθώς το πλήθος των δειγμάτων αύξανε, ο μέσος των δειγμάτων πλησίαζε τον μέσο του πληθυσμού, ενώ η τυπική απόκλιση των δειγμάτων έτεινε στο τυπικό σφάλμα του μέσου του πληθυσμού.
Όσο αφορά την γραφική παράσταση παρατηρήσαμε, ότι καθώς το πλήθος των δειγμάτων αύξανε η μορφή του ιστογράμματος προσέγγιζε όλο και περισσότερο την κανονική κατανομή που υπερτίθετο του ιστογράμματος.

■Ο δειγματικός μέσος προσεγγίζει μια κανονική κατανομή με μέσο όρο ίσο με τον μέσο του πληθυσμού και τυπική απόκλιση ίση με τη τυπική απόκλιση του πληθυσμού που διαιρείται με την τετραγωνική ρίζα του μεγέθους του δείγματος n.

■Το κεντρικό οριακό θεώρημα δηλώνει το αξιοπρόσεκτο αποτέλεσμα ότι, ακόμα και όταν ο γεννήτορας πληθυσμός είναι *μη-κανονικός*, η τυποποιημένη μεταβλητή είναι περίπου κανονική εάν το μέγεθος του δείγματος είναι αρκετά μεγάλο (έστω > 30). Η συνέπεια αυτού του αποτελέσματος είναι ότι σε ένα αριθμό στατιστικών μεθόδων δεν είναι απαραίτητο να προβληματιζόμαστε για την κατανομή του πληθυσμού προέλευσης για την οποία παίρνουμε ένα δείγμα.

■ Η ακρίβεια της προσέγγισης εξαρτάται από το μέγεθος του δείγματος και τη μορφή της κατανομής. Η προσέγγιση λειτουργεί καλύτερα για τις συμμετρικές κατανομές.

■ Η κανονική κατανομή μπορεί να χρησιμοποιηθεί για να προσεγγίσει πολλές κατανομές. Εμείς δεν προσπαθήσαμε να απαριθμήσουμε όλες τις προσεγγίσεις που είναι βασισμένες στην κανονικότητα. Σε γενικές γραμμές, ο στόχος μας ήταν να παρουσιαστούν εκείνοι που συνδυάζουν την ακρίβεια με την απλότητα, που έχει η χρήση με τον υπολογιστή ή χρησιμοποιώντας ένα απλό αλγόριθμο. Προς τούτο παρουσιάσαμε την προσέγγιση της κανονικής στην διωνυμική, στην Poissson και στην Υπεργεωμετρική.

■ Η κανονική κατανομή μπορεί να περιγραφεί ως επέκταση της δυωνυμικής εξίσωσης $\underset{n\to\infty}{(p+q)^n}$ όπου p = q =0. 5 και το n τείνει στο άπειρο. Οι πιθανότητες που προκύπτουν από αυτό το διώνυμο σχηματίζουν ένα τρίγωνο του PASCAL,έτσι έχουν την δυνατότητα να προσεγγίσουν τη κατανομή των πιθανοτήτων κάτω από την κανονική καμπύλη. Η ακρίβεια αυτών των προσεγγίσεων αυξάνεται όταν προστίθεται περισσότερες σειρές στο τρίγωνο του PASCAL.

■ Όταν έχουμε μείξη δύο κανονικών κατανομών μια σημαντική δυσκολία βρίσκεται στην επιλογή των σημείων της αποκοπής.

9.2 Ο ρόλος της κανονικής κατανομής

Η κανονική κατανομή είναι η σπουδαιότερη από όλες τις κατανομές πιθανότητας. Αυτό γιατί:

- Πολλά πειράματα μπορούν να περιγραφούν μέσω τυχαίων μεταβλητών που ακολουθούν την κανονική κατανομή.
- Οι κατανομές δειγματοληψίας που είναι βασισμένες σε μια κανονική κατανομή του γεννήτορα πληθυσμού είναι αρκετά εύχρηστες αναλυτικά. Το μαθηματικό πρόβλημα

παίρνοντας αυτές τις κατανομές είναι συχνά ευκολότερο για τα δείγματα από έναν κανονικό πληθυσμό απ' ό,τι από οποιουσδήποτε άλλον.

• Η κανονική κατανομή μπορεί να χρησιμοποιηθεί σαν προσέγγιση πολλών άλλων κατανομών.

• Η κατανομή αυτή αποτελεί την βάση πολλών στατιστικών τεχνικών που χρησιμοποιούνται στην στατιστική συμπερασματολογία. όπως είναι, οι έλεγχοι υποθέσεων, η ανάλυση παλινδρόμησης, η ανάλυση διασποράς και άλλες.

■ Από την όλη παρουσίαση της εργασία έγινε αντιληπτό, ότι η κανονική κατανομή διαδραματίζει έναν πολύ σημαντικό ρόλο στην Στατιστική, αλλά και στην ζωή.

■Όπως λέει ο *William Youden*

Ο κανονικός νόμος του λάθους διακρίνεται στην εμπειρία της ανθρωπότητας ως μια από τις ευρύτερες γενικεύσεις της φυσικής φιλοσοφίας. Αυτός χρησιμεύει ως όργανο καθοδήγησης στους ερευνητές στις φυσικές και κοινωνικές επιστήμες και στην ιατρική, και τη γεωργία, και την εφαρμοσμένη μηχανική Είναι ένα απαραίτητο εργαλείο για την ανάλυση και την ερμηνεία βασικών δεδομένων που λαμβάνονται από την παρατήρηση και το πείραμα.

THE
NORMAL
LAW OF ERROR
STANDS OUT IN THE
EXPERIENCE OF MANKIND
AS ONE OF THE BROADEST
GENERALIZATIONS OF NATURAL
PHILOSOPHY □ IT SERVES AS THE
GUIDING INSTRUMENT IN RESEARCHES
IN THE PHYSICAL AND SOCIAL SCIENCES AND
IN MEDICINE, AGRICULTURE, AND ENGINEERING □
IT IS AN INDISPENSABLE TOOL FOR THE ANALYSIS AND THE
INTERPRETATION OF THE BASIC DATA OBTAINED BY OBSERVATION AND
EXPERIMENT

William Youden

9.3 Επεκτάσεις

Κλείνοντας, παρατίθενται κάποιες σκέψεις, οι οποίες θα μπορούσαν να υλοποιηθούν, στα πλαίσια μιας, ενδεχόμενης, επέκτασης της παρούσας εργασίας.

Έτσι, επειδή το θέμα είναι πολύ ευρύ θα μπορούσε κάποιος, αν ήθελε να ασχοληθεί με κάποιο από τα ακόλουθα θέματα:

1. Διμετάβλητη κανονική γενικά., αλλά και η διμετάβλητη κανονική σε δειγματικές κατανομές.

2. Πολυμετάβλητη κανονική κατανομή

3. Κατανομή Rayleigh

4. Κατανομή Maxwell

5. Λογαριθμοκανονική κατανομή

6. Εκτιμητική:Σημειακή εκτίμηση. Διαστήματα εμπιστοσύνης, Διαστήματα Tolerance,Διαστήματα πρόβλεψης.

7. Έλεγχοι υποθέσεων

8. Χαρακτηρισμοί. Αποδείξεις χαρακτηρισμών

9. Προσεγγίσεις σε άλλες κατανομές που δεν αναφέρονται στην παρούσα εργασία(όπως στην Αρνητική διωνυμική,F, t,X^2,βήτα,γάμμα και άλλες).

10. Άλλα στατιστικά από κανονικά δείγματα

11. Άλλες σχετικές κατανομές

12. Γκαουζιανές ανελίξεις.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

Ζαχαροπούλου, Χ.(1998). *Στατιστική. Μέθοδοι- Εφαρμογές.* Τόμος Α΄, Θεσσαλονίκη

Κιόχος, Π.Α.(1993). *Στατιστική* .Εκδόσεις INTERBOOKS, Αθήνα

Ξεκαλάκη, Ε.& Πανάρετος, Ι.(1993). *Πιθανότητες και Στοιχεία Στοχαστικών Ανελίξεων.(3η 'Εκδοση)*,Αθήνα

Πανάρετος, Ι., & Ξεκαλάκη Ε. (2000). *Εισαγωγή στη Στατιστική Σκέψη.* Τόμος ΙΙ, (Εισαγωγή στις Πιθανότητες και στην Στατιστική Συμπερασματολογία), Αθήνα

Παπαδήμας, Ο. Γ.(1998). *Στατιστική ΙΙ.* Μακεδονικές Εκδόσεις, Αθήνα

Παπαϊωάννου, Τ. (1981). *Εισαγωγή στις πιθανότητες και την Στατιστική.* Ιωάννινα

Ρούσσας, Γ. Γ. (1973). *Στοιχεία Πιθανοθεωρίας μετ'Εφαρμογών.* Πάτρα

Ξενόγλωσση

Abbe, E. (1906; 1863). Uber die Gesetzmassigkeit in der Vertheilung der Fehler bei Beobachtungsreihen, Gesammelte Abhandlungen, Vol. II, Jena: Gustav Fisher. English translation (1968), ref. P. B. 191, 292, T., from National Technical Information Service, Springfield, Va

Adams, W. J. (1974). *The Life and Times of the Central Limit Theorem.* Caedmon, New York

Adrain, R. (1808). Research concerning the probabilities of errors which happen in making observations, etc., *The Analyst; or Mathematical Museum,* 1(4), 93-109

Agard, J. (1961). Mélange de deux populations normales et étude de quelques fonctions f(x,*y)* de variables normales x,y, R*evue de Statistique Appliquée,* 9, No. 4, 53-70

Ahsanullah, M. and Hamedani, G. G. (1988). Some characterizations of normal distribution, *Calcutta Statistical Association Bulletin, 37,* 95-99

Ahsanullah, M. (1987). A note on the characterization of the normal distribution, *Biometrika Journal 29,* 885-888

Airy, G. B. (1861). *On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations.* Macmillan, London

Alam, K. (1971). A characterization of normality. *Annals of the Institute of Statistical Mathematics, 23,* 523-525

Arnold, B. C. and Meeden, G. (1975). Characterization of distributions by sets of moments of order statistics, *Annals of Statistics, 3,* 754-758

Arnold, B. C., and Isaacson, D. L. (1978). On normal characterizations by the distribution of linear forms, assuming finite variance, *Stochastic Processes and Their Applications,* 7, 227-230

Badhe, S. K. (1976). New approximation of the normal distribution function. *Communications in Statistics—Simulation and Computation, 5,* 173-176

Barnett, V. (1976). Convenient probability plotting positions for the normal distribution. *Applied Statistics,* 25, 47-50

Bartlett, M. S. (1934). The vector representation of a sample. *Proceedings of the Cambridge Philosophical Society, 30,* 327-340

Basu, D. (1955). On statistics independent of a complete sufficient statistic, *Sankhyā, 15,* 377-380

Basu, D. and Laha. R. G. (1954). On some characterizations of the normal distribution. *Sankhyā,* 13, 359-362.

Beasley. J. D. and Springer. S. G. (1977). The percentage points of the normal distribution, *Applied Statistics, 26,* 118-121

Beer, S. and Lukacs, E. (1973). Characterizations of the normal distribution by suitable transformations, *Journal of Applied Probability,10,* 100-108.

Bell, S. (1962). *Approximating the normal distribution with the triangular,* Santia, Corporation Report No. 494

Bergström, H. (1945). On the central limit theorem in the space R^{k}, $k>1$, *Skandinarisk Akluarictidskrift,* 28, 106-127

Berk, R.H. (1986). Sphericity and the normal law, *Annals of Probability, 14,696-701*

Bernoulli, D. (1770-1771). Mensura sortis ad fortuitam successionem rerum naturaliter contingentium applicata, *Novi Commentarii Academiae*

Scientiarum Imperialis Petropolitanae, 14, 26-45; 15, 3-28

Bernoulli, J. (1713). Ars Conjectandi; German version (1899). *Wahrscheinlichkeitsrechnung,* Engelmann, Leipzig

Bernstein, S. (1941). Sur une propriété caractéristique de la loi de Gauss, *Transactions of the Leningrad Polytechnic Institute, 3,* 21-22; reprinted (1964) in *Collected Works, Vol. 4,* 314-315

Berry, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates, *Transactions of the American Mathematical Society,* 49,122-136

Bessel, F. W. (1818). Astronomiae pro anno MDCCLV deducta ex observationibus viri incomporabilis James Bradley specula Grenovicensi per annos 1750-1762 institutis, Regiomonti

Bessel, F. W. (1838). Untersuchungen über der Wahrscheinlichkeit der Beobachtungsfehler, *Astronomische Nachrichten,* 15, 368-404

Bhattacharjee, G. P., Pandit, S. N. N. and Mohan, R. (1963). Dimensional chains involving rectangular and normal error-distributions, *Technometrics, 5,* 404-406

Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics.* Holden-Day, San Francisco

Bienaymé, J. (1838). Sur la probabilité des résultats moyens des observations; démonstration directe de la règle de Laplace, *Mémoires des Savans Etrangers, Académie (Royale) des Sciences de l'Institut de France,* Paris, 5, 513-558

Bienayme, J. (1852). Sur la probabilité des erreurs d'après la méthode des moindres carres; reprinted (1858): *Mémoires des Savans Etrangers, Académie (Royale) des Sciences de l'Institut de France,* Paris, 15, 615-663

Blum, J. R. (1956). On a characterization of the normal distribution, *Skandinavisk Aktuarietidakrift, 39,* 59-62

Bondesson, L. (1977). The sample variance, properly normalized, is χ^2-distributed for the normal law only, *Sankyā A, 39,* 303-304

Borges, R. (1970). Eine Approximation der Binomiahverteilung durch die Normalverteilung der Ordnung 1/n, *Zeitschrift für*

Wahrscheinlichkeitstheorie und verwandte Gebiete, 14, 189-199

Borovkov, A. A. and Utev, S. A. (1983). On an inequality and a related characterization of the normal distribution, *Theory of Probability and its Applications, 28*, 219-228

Box, G. E. P. and Muller, M. E. (1958). A note on the generation of random normal deviates, *Annals of Mathematical Statistics*, 29, 610-611

Bravais, A. (1846). Analyse mathématique sur les probabilités des erreurs de situation d'un point, *Mémoires des Savans Etrangers, Académie (Royale) des Sciences de l'Institut de France*, Paris, 9, 255-332

Braverman, M. Sh. (1985). The characteristic properties of normal and stable distributions, *Theory of Probability and its Applications, 30*, 465-474

Bryc, W. (1990). *Normal distributions and characterizations*. University of Cincinnati, Unpublished manuscript Cincinnati, OH

Bryc, W. (1994). *Normal distributions characterizations with applications.* University of Cincinnati, Unpublished manuscript Cincinnati, OH

Bryc, W. and Szablowski. P. J. (1990). Some characteristics of normal distribution by conditional moments, *Bulletin of the Polish Academy of Sciences: Mathematics*, 38, 209-218

Burgess, J. (1898). On the definite integral... [i.e., ...], with extended tables of values, Transactions *of the Royal Society of Edinburgh*, 39, 257-321

Burr, I. W. (1967). A useful approximation to the normal distribution function, with application to simulation, *Technometrics*, 9, 647-651

Buslenko, N. P., Golenko. D. I., Shreidcr. Yu. A., Sobol. I. M. and Sragovich. V. G. (1966). *The Monte Carlo Method.* Pergamon, Oxford (Original Russian edition published 1962)

Cacoullos, T. (1967b). Characterizations of normality by constant regression of linear statistics on another linear statistic, *Annals of Mathematical Statistics, 38*, 1894-1898

Cacoullos, T. (1967a). Some characterizations of normality, *Sankhyā A, 29*, 399-404

Cacoullos, T. and Papathanasiou, V. (1989). Characterization of distributions by variance bounds, *Statistics and Probability Letters, 7*, 351-356

Cadwell, J. H. (1951). The bivariate normal integral. *Biometrika,* 38, 475-479

Camp, B. H. (1951). Approximation to the point binomial. Annals of *Mathematical Statistics,* 22, 130-131

Carta, D. G. (1975). Low-order approximations for the normal probability integral and the error function, *Mathematics of Computation,* 29, 856-862

Chan, L. K. (1967). On a characterization of distribution by expected values of extreme order statistics, *American Mathematical Monthly, 74,* 950-951

Charlier, C. V. L., and Wicksell, S. D. (1924). On the dissection of frequency functions, *Arkir för Matematik, Astronomi och Fysik,* 18, No. 6, 1-64

Charlier, C. V. L. (1905). Über die Darstellung willkürlicher Funktionen, *Arkiv för* Mathematik, *Astronomin och Fysik,* 2(20), 1-35

Cheng, T. T. (1949) .The normal approximation to the Poisson distribution and a proof of a conjecture of Ramanujan, *Bulletin of the American Mathematical Society,* 55, 396-401

Chew, V. (1968). Some useful alternatives to the normal distribution. *The American Statistician, 22,* No. 3, 22-24

Clow, R., Hansen, E., and McNolty, F. (1974). Bayesian density functions for Gaussian pulse shapes in Gaussian noise, *Proceedings of the IEEE,* 62, 134-136

Cohen, A. C. (1967). Estimation in mixtures of two normal distributions, *Technometrics,* 9, 15-28

Cramer, H. (1936). Über eine Eigenschaft der normalen Verteilungsfunktion. *Mathematische Zeitschrift,* 41, 405-414

Cramer, H. (1946). *Mathematical Methods of Statistics.* Princeton, N.J.: Princeton University Press

Csörgö, M. and Seshadri, V. (1971). Characterizing the Gaussian and exponential laws via mappings onto the unit interval, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 18,* 333-339

Csörgö, M., Seshadri. V. and Yalovsky, M. (1975). Application of characterizations in the area of goodness-of-fit, In *Statistical Distributions in Scientific Work* (G. P.Patil, ed.), Dordrecht, Netherlands: Reidel, 2, 79-90

Daniel, C.(1959). Use of half-normal plots in interpreting factorial two-level experiments, *technometrics,*1,311-341

Darmois, G. (1951). Sur une propriété caractéristique de la loi de probabilité de Laplace, *Comptes Rendus de I'Académie des Sciences* (Paris) 232, 1999-2000

Darmois, G. (1953). Analyse gencrale des liasons stochastiques. Etude particuliaire de l'analyse factorielle lincaire, *Rente de I'lnstiiut Internationale de Statislique,* 21, 2-8

David, F. N. (1962). Games, Gods, and Gambling. Hafner, New York

Daw, R. H. and Pearson, E. S. (1972). Studies in the history of probability and statistics, XXX: Abraham de Moivre's 1733 derivation of the normal curve: A bibliographical note, *Biometrika,* 59, 677-680

De Moivre, A. (1733). Approximatio ad Summan Ferminorum Banomii $(a+b)^n$ in Seriem expansi

De Moivre, A. (1738, 1756). *The Doctrine of Chances;* reprint (1967). Chelsea, New York

De Morgan, A. (1837). Theory of probabilities, in *Encyclopedia, Metropolitana, 2,* 359-468

De Morgan, A. (1838). An essay on probabilities and on their application to life contingencies and insurance offices. Longmans, London

Derenzo, S. E. (1977). Approximations for hand calculators using small integer coefficients, *Mathematics of Computation,* 31, 214-225

Diaconis, P., Olkin, I. and Ghurye, S. G. (1977). Review of *Characterization Problems in Mathematical Statistics* (by A. M.Kagan, Yu. V. Linnik, and C. R. Rao), *Annals of Statistics, 5,* 583-592

D'Ortenzio, R. J. (1965). Approximating the normal distribution function, *Systems Design,* 9, 4-7

Dubey, S.D.(1967). Normal and Weibull distributions, *Naval research Logistic Quarterly*, 14,69-79

Dynkin, E. B. (1961). Necessary and sufficient statistics for a family of probability distributions, *Selected Translations in Mathematics, Statistics, and Probability,* 1, 17-40

Edgeworth, F. Y. (1883). The law of error, *Philosophical Magazine,* 16, 300-309

Edgeworth, F. Y. (1892). Correlated averages, *Philosophical Magazine* Ser. 5, 34, 190-204

Edgeworth, F. Y. (1905). The law of error, Proceedings *of the Cambridge Philosophical Society,* 20, 36-65

Edwards, A. W. F. (1963). A linkage for drawing the normal distribution, *Applied Statistics,* 12, 44-45

Eisenberger, I. (1964). Genesis of bimodal distributions. *Technometrics,* 6, 357-363

Esseen, C. G. (1942). On the Liapounoff limit of error in the theory of probability, *Arkiv för Mathematik, Aslronomi och Fysik,* 28A, 1-19

Esseen. C. G. (1958). On mean central limit theorems, *Kunglige Tekniska Högskolans Handlingar,* No. 121, 1-30

Feller, W. (1935). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung, *Mathematische Zeitschrift,* 40, 521-559

Feller, W. (1968). *Introduction to Probability Theory and Its Applications,* Vol. 1 (3rd ed.).Wiley, New York

Ferguson, T. S. (1967). *Mathematical Statistics-A Decision Theoretic Approach.* Academic, New York

Fleming, N. S. (1989). Approximating the normal cumulative distribution function using a spreadsheet program. *The American Statistician,* 43, 68, Reply.Ibid.,43,290-291

Fletcher, A. Miller, J. C. P., Rosenhead, L. and Comrie, L. J. (1962B). *An Index of Mathematical Tables,* Vol.II (2nd ed.),Reading, Mass .:Addison –Wesley, Scientific Computing Servise

Fox, C. (1965). A family of distributions with the same ratio property as normal distribution, *Canadian Mathematical Bulletin,* 8, 631-636

Freeman, M. F., and Tukey, J. W. (1950). Transformations related to the angular and the square root, *Annals of Mathematical Statistics,* 21,607-

611

Galileo, G. (trans. 1953; 2nd ed. 1962). *Dialogue Concerning the Two Chief* World Systems-Ptolemaic and Copernican (trans. S. Drake), University of California Press, Berkeley

Galton, F. (1869). *Hereditary Genius.* Macmillan, London

Galton, F. (1875). Statistics by intercomparison, with remarks on the law of frequency of error, *Philosophical Magazine*, 49, 33-46

Galton, F. (1886). Family likeness in stature, *Proceedings of the Royal Society,* 40, 42-73

Galton, F. (1888). Correlations and their measurement, chiefly from anthropometric data, Proceedings of the *Royal Society,* 45, 135-145

Galton, F. (1889). *Natural Inheritance* .Macmillan, London

Gauss, C. F. (1809). *Theoria Motus Corporum Coelestium*, Lib. 2, Sec. III, 205-224, Hamburg: Perthes u. Besser

Gauss, C. F. (1816). Bestimmung der Genauigkeit der Beobachtungen, *Zeitschrift Astronomi* ,1, 185-197

Geary, R. C. (1936). Distribution of "Student's" ratio for nonnormal samples, *Journal of the Royal Statistical Society B3,* 178-184

Gebhardt, F. (1969). Some numerical comparisons of several approximations to the binomial distribution, *Journal of the American Statistical Association,* 64, 1638-1646

Gebhardt, F. (1971). Incomplete Beta-integral B(x;2/3,2/3) and $[p(1 - p)]^{-1/6}$ for use with Borges' approximation of the binomial distribution, *Journal of the American Statistical Association,* 66, 189-191

Geisser, S. (1956). A note on the normal distribution, *Annals of Mathematical Statistics, 27,* 858-859

Geisser, S. (1973). Normal characterizations via the square of random variables, *Sankyā A,35, 494-494*

Geisser, S. (1973). Normal characterization via the squares of random variables, *Sankhyā. Series A,* 35,492-495

Ghurye, S. G., and Olkin, I. (1973). Identically distributed linear forms and the normal distribution, *Advances in Applied Probability, 5,* 138-152

Glaisher, J. W. L. (1871). On a class of definite integrals-Part II, *Philosophical Magazine*, 42, 421-436.

Gnedenko, B. V., and Kolmogorov, A. N. (1968). *Limit Distributions for Sums of Independent Random Variables* (trans. from Russian), Addison-Wesley ,Reading, Mass

Govindarajulu, Z. (1966). Characterization of normal and generalized truncated normal distributions using order statistics, *Annals of Mathematical Statistics, 37*, 1011-1015

Govindarajulu, Z. (1965). Normal approximations to the classical discrete distributions, Sankhyā A,27, 143-172

Greenwood, J. A., and Hartley, H. 0. (1962). *Guide* to *Tables in Mathematical Statistics*, Princeton University Press, Princeton, N.J

Hamaker, H. C. (1978). Approximating the cumulative normal distribution and its inverse, *Applied Statistics*, 27, 76-77

Hart, R. G. (1966). A close approximation related to the error function. *Mathematics of Computation*, 20, 600-602

Hartman, P. and Wintner (1940). On the spherical approach to the normal law, *American Journal of Mathematics*, 62, 759-779

Hastings, C. *(1955). Appimations for Digital Computers.* Prinecton NJ, Princeton University Press

Helguero, F. dc (1904). Sui massima delle curve dimorfiche, *Biometrika*, 3. 84-98

Helguero, F. dc (1905). Per la risoluzione delle curve dimorfiche, *Biometrika*, 4,230-231

Helmert, F. R. (1876). Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers director Beobachtungen tungen gleicher Genauigkeit, *Astronomische Nachrichten*,88,113-120

Hemelrijk, J. (1967). The hypergeometric, the normal and chi-squared, *Statistica Neerlandica*, 21, 225-229

Hombas, V. C. (1985). Characterizing the normal density as a solution of a differential equation, *Statistica Neerlandica, 39*, 387-388

Hoyt, J. P. (1968). A simple approximation to the standard normal probability density function, *The American Statistician*, 22, No. 2, 25-26

Johnson, N. L .,Kotz, S.and Kemp, A. W.(1992).*Univariate Discrete Distributions*(2nd edn.).Wiley, New York

Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions, Vol. 1.*Wiley, New York

Johnson, N. L. and Kotz, S. (1969). *Distributions in Statistics: Discrete Distributions.* Wiley, New York

Johnson, N.L. and Kotz, S.(1970a).*Distributions in Statistics: Continuous Univariate Distributions*, Vol. 1 . Wiley, New York

Johnson, N.L. and Kotz, S.(1970b).*Distributions in Statistics: Continuous Univariate Distributions*, Vol. 2 .Wiley, New York

Kac, M. (1975). Some reflections of a mathematician on the nature and the role of statistics, Proceedings *of the Conference on Directions for Mathematical Statistics,* 5-11, Applied Probability Trust

Kagan, A. M., Linnik, Yu. V. and Rao, C. R. (1965). On a characterization of the normal law based on a property of the sample average. *Sankya A,27,* 405-406

Kagan, A. M., Linnik, Yu. V. and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics* (trans. B. Ramachandran). Wiley, New York

Kagan A. M. and Shalayevskii. O.V. (1967). Characterization of the normal law by a property of the non-central χ^2-distribution. *Lietutos Matematikos Rinkinys,* 7, 57-58 (In Russian)

Kaplansky, I. (1943). A characterization of the normal distribution, *Annals of Mathematical Statistics, 14,* 197-198

Kelker, D. and Matthes, R. K. (1970). A sufficient statistics characterization of the normal distribution. *Annals of Mathematical Statistics,* 41, 1086-1090

Kelley, T. L. (1948). *The Kelley Statistical Tables.* Harvard University Press, Cambridge

Kendall, M. G. (1963). Ronald Aylmer Fisher, 1890-1962, *Biometrika* 50, 1-15; reprinted (1970) in Studies in *the History of Probability and Statistics* (E. S. Pearson and M. G. Kendall, eds.), 439-454. Hafner, New York

Kendall, M. G. (1971). Studies in the history of probability and statistics, XXVI: The work of Ernst Abbe, *Biometrika,* 58, 369-373; Corrigendum (1972): *Biometrika,* 59, 498

Kerridge, D. F., and Cook, G. W. (1976). Yet another series for the normal integral, *Biomelrika,* 63, 401-403

Khatri, C. G. and Rao, C. R. (1972). Functional equations and characterization of probability laws through linear functions of random variables, *Journal of Multivariate Analysis ,2,* 162-173

Khatri, C. G. (1979). Characterization of multivariate normality, II: Through linear regressions, *Journal of Multivariate Analysis ,9,* 589-598

Khatri, C. G. and Rao, C. R. (1976). Characterizations of multivariate normality, I: Through independence of some statistics, *Journal of Multivariate Analysis, 6,* 81-94

Konwar, R. M. (1991). On characterizations of distributions by mean absolute deviation and variance bounds, *Annals of the Institute of Statistical Mathematics, 43,* 287-295

Kotlarski, I. (1966). On characterizing the normal distribution by Student's law, *Biometrika, 53,* 603-606

Kotlarski, I. (1967). On characterizing the gamma and the normal distribution, *Pacific Journal of Mathematics, 20,* 69-76

Kotz, S. (1974). Characterization of statistical distributions: A supplement to recent surveys, *International Statistical Review, 42,* 39-65

Kramp, C. (1799). Analyse des réfractions astronomiques et terrestres, Leipsic: Schwikkert, Koenig, Paris

Kruskal, W. (1978). Formulas, numbers, words: Statistics in prose, *The American Scholar,* 47, 223-229

Laha, **R. G. (1953).** On an extension of Geary's theorem, *Biometrika, 40,* 228-229

Lancaster, H. 0. (1954). Traces and cumulants of quadratic forms in normal variables, *Journal of the Royal Statistical Society B,* 16, 247-254

Lancaster, H. 0. (1966). Forerunners of the Pearson χ^2, *Australian Journal of Statistics,* 8, 117-126

Lancaster, H. 0. (1987). Finiteness of the variances in characterizations of the normal distribution, Australian *Journal of Statistics, 29,* 101-106

Laplace, P. S. (1809-1810). Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilités, *Mémoires* de *la Classe des Sciences mathématiques et physiques de l'Institut,* Paris, 353-415, 559-565, et. passim (see Oeuvres Complètes ,12, 301-353)

Laplace, P. S. (1810). Mémoire sur les intégrales définies et leur application aux probabilités, *Mémoires de la Classe des Sciences mathématiques et physiques de l'Institut,* Paris, 279-347, et. passim (1810-1812: see Oeuvres *Complètes,* 12, 357-412)

Laplace, P. S. (1812). Tnéorie *Analytique des Probabilités,* Paris (Oeuvres Complètes, Vol. 7)

Laplace, P. S. (1878, 1912). Oeuvres Complètes **de** *Laplace,* 14 volumes. Gauthier –Villars, Paris

Legendre, A. M. (1805). *Nouvelles Méthodes pour la Détermination* des *Orbites des Comètes,* Paris

Legendre, A. M. (1826). Traite des *Fonctions elliptiques et des Intégrales Euleriennes, avec des Tables pour en faciliter le Calcul numérique,* Vol. 2, Huzard –Courcier, Paris

Lehmann, E. L. (1986). *Testing Statistical Hypotheses* (2nd edn.).Wiley, New York

Lin, J. T. (1988). Alternatives to Hamaker's approximations to the cumulative normal distribution and its inverse, *The Statistician,* 37, 413-414

Lin, J. -T. (1989). Approximating the normal tail probability and its inverse for use on a pocket calculator, *Applied Statistics,* 38, 69-70

Lindeberg,J.W.(1922). Einie neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung, *Mathematische Zeitschrift,*15,211-225

Linders, F. J. (1930). On the addition of two normal frequency curves, *Nordic Statistical Journal, 2,* 63-73

Ling, R . F.and Pratt, J. W. (1984). The accuracy of Peizer approximations to the Hypergeometric Distribution, with comparisons to some other approximations, *Journal of the American Statistical* Association ,79, 49-60

Linnik, Yu. V. (1952). Linear statistics and the normal distribution law. *Doklady Akademii Nauk SSSR,* 83, 353-355. (In Russian. English translation published be American Mathematical Society. 1961.)

Lukacs, E. and Laha, R. G. (1964). *Applications of Characteristic Functions.* Hafner, New York

Lukacs, E. (1956). Characterization of populations by properties of suitable statistics, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 2,* 195-214

Lukacs, E. (1942). A characterization of the normal distribution, *Annals of Mathematical Statistics, 13,* 91-93

Lukacs. E. (1955). Applications of Faà di Bruno's formula in mathematical statistics. *American Mathematical Monthly,62,* 340-348

Lyapunov, A. M. (1900). Sur une proposition de la théorie des probabilités. *Izvstiya Akademii Nauk SSSr, series V,*13,359-386.[Also Lyapunov, A.(1954).*Collected Works,*1,125-151,Moscow :Akademia Nauk SSSR.(In Russian)]

Lyapunov, A. M. (1901). Nouvelle forme du théorème sur la limite de probabilité, *Mémoires de* l'Academic *Impériale des Sciences de St. Pétersbourg,* 12, 1-24

Lyapunov, A. M. (1954-1965). *Izbrannye Trudi* (Selected Works), Academy of Sciences, USSR

Macino,T.(1984). Mean hazard rate and itw application to the notmal approximation of the Weibull distribution, *Naval Research Logistics Quarterly,* 31,1-8

Mahalanobis, P. C, Bose, S. S., Roy. P. R., and Banerjee, S. K. (1934). Tables of random samples from a normal distribution, *Sankhya,* 1, 289-328

Maistrov, L. E. (1967, 1974). Probability *Theory: A Historical Sketch* (trans. by S. Kotz), New York: Academic Press

Makabe, H., and Morimura, H. (1955). A normal approximation to the Poisson distribution, *Reports on Statistical Applications Research* (Union of Japanese Scientists and Engineers) ,4, 37-46

Marcinkiewicz, J. (1939). Sur une propriété de la loi de Gauss, *Mathematische Zeitschrift ,44,* 612-618

Markov, A. A. (1888). Table des valeurs de l'integrale..., St. Pétersbourg: Académie Impériale des Sciences

Markov, A. A. (1899-1900). The law of large numbers and the method of least squares, *Izvestia Physiko -mathematicheskago Obschvestva pri Imperatorskom Kazanskom Universitet,* 8, 110-128; 9, 41-43

Markov, A. A. (1913). A probabilistic limit theorem for the cases of Academician A. M. Lyapounov, *Izvlyechenye iz knigi ischislyenge veroyatnostyei* (supplement to "Theory of Probability"), 4-e

Markov, A. A. (1951). *Izbrannye Trudi* (Selected Works), Academy of Sciences, USSR

Mathai, A. M., and Pederzoli, G. (1977). *Characterizations of the Normal Probability Law.* Wiley, New York

Maxwell, J. C. (1860). Illustrations of the dynamical theory of gases, *Philosophical Magazine,* 19, 19-32; 20, 21-37; (1952): *Scientific Papers of James Clerk Maxwell,* 377-409

McGillivray, W. R., and Kaller, C. L. (1966). A characterization of deviation from normality under certain moment assumptions, *Canadian Mathematical Bulletin,* 9,509-514

McLachlan, **G. J., and Basford, K. E. (1987).** *Mixture Models: Inference and Applications to Clustering.* Marcel Dekker, New York

Meshalkin, L. D. (1968). On the robustness of some characterizations of the normal distribution, *Annals of Mathematical Statistics ,39,* 1747-1750

Molenaar, W. (1970). *Approximations to the Poisson, Binomial, and Hypergeometric Distribution Functions,* Mathematical Centre Tracts, Vol. 31.Mathematisch Centrum, Amsterdam

Molenaar, I. W. (1985). Normal approximations to the Poisson, Binomial, negative binomial and Hypergeometric Distribution, *Encyclopedia of Statistical Sciences,*Vol. 6(S.Kotz, N.L.Johnson and C.B.Read,eds.),340-347. Wiley, New York

Molenaar. **W. (1965).** Survey of separation methods for two normal distributions . *Statislica Neerlandica,* 19, 249-263

Mood, A. M.., Graybill, F. A. and Boes, D. C. (1974). *Introduction to the*

Theory of Statistics. McGraw-Hill, New York

Moran, P. A. P. (1980). Calculation of the normal distribution function. *Biometrika,* 67, 675-676

Nelson, L. S. (1976). Constructing normal probability paper, *Journal of Quality Technology,* 8, 56-57

Nhu, H. H. (1968). On the stability of certain characterizations of a normal population, *Theory of Probability and Its Applications, 13,* 299-304

Nicholson, W. L. (1956). On the normal approximation to the hyper-geometric distribution, Annals *of Mathematical Statistics , 27,* 471-483

Norton, R. M. (1989). Pocket-calculator approximation for areas under the standard normal curve, *The American Statistician,* 43, 24-26. Reply, *Ibid.,* 43, 290

Nye, H. H. (1966). On the stability of certain theorems characterizing normal distribution, *International Congress of Mathematicians: Information Bulletin,* 6, 5

Odeh, R. E. and Evans, J. O. (1974). Algorithm AS 70. The percentage points of the normal distribution, *Applied Statistics,* 23, 96-97

Pagano, M. and Gauvreau, K.(1996). *Αρχές Βιοστατιστικής.* Μετάφραση - Επιμέλεια: Ουρανία Δάφνη (2000), Εκδόσεις “ΕΛΛΗΝ”, Αθήνα

Page, E. (1977). Approximations to the cumulative normal function and its inverse for use on a pocket calculator, *Applied Statistics, 26,* 75-76

Pakshirajan, **P., and Mohan, N. C. (1971).** A characterization of the normal law, *Annals of the Institute of Statistical Mathematics,* 21, 529-532

Pakshirajan, R. P. and Mohan, N. R. (1969). A characterization of the normal law, *Annals of the Institute of Statistical Mathematics, 21,* 529-532

Patel, J. K. and Read, C. B. (1981). *Handbook of the Normal Distribution.* Marcel Dekker, New York

Patel, J. K. and Read, C. B. (1996). *Handbook of the Normal Distribution.* Marcel Dekker, New York

Pathak, P. K. and Phillai, R. N. (1968). On a characterization of the normal law, *Sankya A,30,* 141-144

Patil, G. P. and Seshadri, V. (1964). Characterization theorems for some univariate probability distributions, *Journal of the Royal Statistical Society B,26,* 286-292

Pearson, E. S. (1965). Some incidents in the early history of Biometry and Statistics, *Biometrika* 52, 3-18; reprinted (1970) in *Studies in the History of Statistics and Probability* (E. S. Pearson and M. G. Kendall, eds.), 323-338. Hafner, New York

Pearson, E. S. (1967). Some reflexions on continuity in the development of mathematical statistics, 1885-1920, *Biometrika* 54, 341-355; reprinted (1970), as in Pearson (1965), 339-353

Pearson, E. S. and Wishart, J., eds. (1958). "Student's" Collected *Papers.* Cambridge University Press, Cambridge

Pearson, K. (1894). Contributions to the mathematical study of evolution. *Philosophical Transactions of the Royal Society of London, Series A,* 185, 71-110

Pearson, K. (1896). Mathematical contributions to the theory of evolution, III: Regression, heredity, and panmixia, *Philosophical Transactions of the Royal Society of London* A,187, 253-318

Pearson, K. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philosophical Magazine* (5)50, 157-175

Pearson, K. (1921-1933). The *History of Statistics in the 17th and18th Centuries, against the Changing Background of Intellectual, Scientific, and Religious Thought* (lectures, E. S. Pearson, ed., 1978). Macmillan, New York

Pearson, K. (1920). Notes on the history of correlation, *Biometrika* 13, 25-45; reprinted (1970) in Studies in the *History of Statistics and Probability* (E. S. Pearson and M. G. Kendall, eds.), 185-205. Hafner, New York

Peizer, D. B. and Pratt, J. W. (1968). A normal approximation for binomial, F, beta, and other common, related tail probabilities, I, *Journal of the American Statistical Association,* 63, 1416-1456

Plackett, R. L. (1972). Studies in the history of probability and statistics, XXIX: The discovery of the method of least squares, Biometrika, 59, 239-251

Plait, A. (1962). The Weibull distribution-with tables, *Industrial Quality control* ,19(5),17-26

Pólya, G. (1932). Verleitung des Gauss' chen Fehlergesetzes aus einer Funktional- gleichung. *Mathematische Zeitschrift,* 18, 96-108

Pólya, G. (1945). Remarks on computing the probability integral in one and two dimensions. *Proceedings of the 1st Berkeley Symposium on Mathematical Statistics and Probability,* 63-78

Prakasa Rao, B.L.S. (1979). Characterizations of distributions through some identities, *Journal of Applied Probability, 16,* 903-909

Prasad, A. (1955). Bi-modal distributions derived from the normal distribution, *Sankltya,* 14,369-374

Quetelet, L. A. J. (1846). Lettres a *S. A. R. Le Due Régnant de Saxe-Cobourg et Gotha, sur la Théorie des Probabilités appliquée aux Sciences Morales et Politiques* (English trans., 1849). Hayez ,Brussels

Raab, D. H. and Green. E. H. (1961). A cosine approximation to the normal distribution, *Psychometrika, 26,* 447-450

Raff, M. S. (1956). On approximating the point binomial, Journal *of the American Statistical Association,* 51, 293-303

Ramasubramaniam, S. (1985). A characterisation of the normal distribution, *Sankhyā A,* 410-414

RAND Corporation (1955). *A Million Random Digits with 100,000 Normal Deviates Glencoe, IL: Free Press*

Rao, C. R. (1967). On some characterizations of the normal law, *Sankya A*,29, 1-14

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications* (2nd edn.).Wiley, New York

Rao, J. N. K. (1958). A characterization of the normal distribution. *Annals of Mathematical Statistics,* 29, 914-919

Read, C. B. (1985). Normal distribution, *Encyclopedia of Statistical Sciences,* 6. S. Kotz, N. L. Johnson, and C. B. Read (editors), 347-359. Wiley, New York:

Riffenburgh, R. H. (1967). Transformation for statistical distribution approximately normal but of finite sample range, Report NUWC-TP-19, Naval Undersea Warfare Center, San Diego, CA [Abbreviated version *Technometrics,* 11, (1969), 47-59]

Riordan, J. (1949). Inversion formulas in normal variable mapping, *Annals of Mathematical Statistics,* 20, 417-425

Rivest, L. P. (1981). On the sum of contaminated normals. *The American Statistician* ,35, 155-156

Rouncefield, M. (1990). Is it normal. How to use probability (or normal) graph paper, *Teaching Statistics,* 12, 6-8

Ruben, H. (1974). A new characterization of the normal distribution through the sample variance, *Sankya A,*36, 379-388

Ruben, H. (1975). A further characterization of normality through the sample variance, *Sankya AS,7,* 72-81

Ruben, H. (1976). A characterization of normality through the general linear model, *Sankya A,*38, 186-189

Sadikova, S. M.(1966). On two-dimensional analogs of an inequality of C.G. Essen and their application to the central limit theorem, *Teoriya Veroyatnostei i ee Primeneniya,*11,370-380.(In Russian)

Savage, L. J. (1976). On rereading R. A. Fisher (with discussion), *Annuls of Statistics,* 4, 441-500

Sazanov, V. V.(1967). On the rate of convergence in the multidimensional central limit theorem, *Teoriya Veroyatnostei i ee Primeneniya,*12,82-95.(In Russian)

Schmeiser, B. W. (1979). Approximations to the inverse cumulative normal function for use on hand calculators, *Applied Statistics,* 28, 175-176.

Schols, C. M. (1875). Over de theorie des fouten in de ruimte en in het platte vlak, *Verhandelungen der koninklijke Akademie van Wetenschappen,* Amsterdam, 15, 1-75

Schonfelder, J. L. (1978). Chebyshev expansions for the error and related functions, *Mathematics of Computation,* 32, 1232-1240

Schucany, W. R. and Gray, H. L. (1968). A new approximation related to the error function, *Mathematics of Computation,* 22, 201-202

Seal, H. L. (1967). Studies in the history of probability and statistics, XV: The historical development of the Gauss linear model, Biometrika, 54, 1-24; reprinted (1970) in *Studies in the History of Probability and Statistics* (E. S. Pearson and M. G. Kendall, eds.), 207-230. Hafner, New York

Sengupta, J. M. and Bhattacharya, N. (1958). Tables of random normal deviates, *Sankhya*, 20, 250-286

Seshadri, V. (1969). A characterization of the normal and Weibull distributions, *Canadian Mathematical Bulletin* ,12, 257-260

Shah, A. K. (1985). A simpler approximation for areas under the standard normal curve, *The American Statistician*, 39, 80. Correction, *Ibid.*, 39, 327

Sheynin, 0. B. (1966). Origin of the theory of errors, Nature 211, 1003-1004

Sheynin, 0. B. (1968). Studies in the history of probability and statistics, XXI: On the early history of the law of large numbers, Biometrika, 55, 459-467; reprinted (1970) in Studies in the *History of Statistics and Probability* (E. S. Pearson and M. G. Kendall, eds.), 231-239. Hafner, New York

Sheynin, 0. B. (1970). Studies in the history of probability and statistics, XXIII: Daniel Bernoulli on the normal law, Biometrika, 57, 199-202

Shimizu, R. (1961). A characterization of the normal distribution, *Annals of the Institute of Statistical Mathematics* , 13, 53-56

Shimizu, R. (1962). Characterization of the normal distribution, II, *Annals of the Institute of Statistical Mathematics*, 14, 173-178

Shore, H. (1982). Simple approximations for the inverse cumulative function, the density function and the loss integral of the normal distribution. *Applied Statistics*, 31, 108-114

Skitovitch, V. P. (1953). On a property of the normal distribution, *Doklady Akademii Nauk SSSR*, 89, 217-219 (in Russian)

Smirnov, **N. V. (ed.) (1960, 1965).** *Tables of the Normal Probability Integral, Normal Density, and Its Normalized Derivatives.* Macmillan, New York (Translation)

Spiegel, M.R. and Stephens L.J.(1999,1988,1961). *Στατιστική.* (3η έκδοση), McGraw-Hill: New York. Μετάφραση:Εκδόσεις ΤΖΙΟΛΑ, Θεσσαλονίκη

Spiegel, M. R. (1975). *Πιθανότητες και Στατιστική.* McGraw-Hill: New York Μετάφραση: Περσίδης Σ. Κ.(1977).Εκδόσεις ΕΣΠΙ, SCHAUM 3,Αθήνα

Sprott, D. A. (1978). Gauss's contributions to statistics, *Historia Mathematica,* 5, 183-203

Steffensen, J. F. (1937). On the semi-normal distribution, *Skandinarisk Akttiarietidskrift,* 20, 60-74

Stepniak, C. (1991). On characterization of the normal law in the Gauss Markov model, *Sankya* A, 115-117

Stigler, S. M. (1975a). Studies in the history of probability and Statistics, XXXIV: Napoleonic statistics: The work of Laplace, *Biometrika,* 62, 503-517

Stigler, S. M. (1975b). The transition from point to distribution estimation, Bulletin *of the International Statistical Institute* 46 (Proceedings of the 40th Session, Book 2), 332-340

Stigler, S. M. (1977). An attack on Gauss, published by Legendre in 1820, *Historia* Mathematics ,4, 31-35

Stigler, S. M. (1978). Francis Ysidro Edgeworth, statistician, Journal *of the Royal Statistical Society* A,141, 287-313, followed by discussion

Stigler, S. M. (1982). A modest proposal: A new standard for the normal, *The American Statistician,* 36, 137-138

Stigler, S. M. (1973). Laplace, Fisher, and the discovery of the concept of sufficiency, *Biometrika,* 60, 439-445

Strecock, A. J. (1968). On the calculation of the inverse of the error function. *Mathematics of Computation,* 22, 144-158

Stuart, A. and Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics, Vol. 1* (5th edn.). Oxford University Press, New York

Student (1908). The probable error of a mean, *Biometrika,* 6, 1-25

Tamhankar, M. V. (1967). A characterization of normality, *Annals of Mathematical Statistics,* 38, 1924-1927

Taylor, B. J. R. (1965). The analysis of polymodal frequency distributions, *Journal of Animal Ecology,* 34, 445-452

Tchebyshev, P. L. (1890). Sur deux théorèmes relatifs aux probabilités, Acta

Mathematica, 14, 305-315; reprinted (1962) in Oeuvres, Vol. 2. Chelsea, New York

Teicher, H. (1961). Maximum likelihood characterization of distributions, *Annals of Mathematical Statistics,* 32, 1214-1222

Teichroew, D. (1957). The mixture of normal distributions with different variances, *Annals of Mathematical Statistics,* 28, 510-512

Tippett, L. H. C. (1927). *Random Sampling Numbers* (Tracts for Computers XV).Cambridge University Press, Cambridge

Tranquilli, G. B. (1966). Sul teorema di Basu-Darmois, *Giornale dell'Istituto Italiano degliAttuari,29,135-152*

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics,* 5, 99-114

Uspensky, J. V. (1937). Introduction to Mathematical Probability. McGraw-Hill, New York

Wallace, D.(1959).*A corrected computation of Berry's bound for the central limit theorem error.* Statistics Research Center, University of Chicago

Wesolowski, J. (1987). A regressional characterization of the normal law, *Statistics and Probability Letters,* 6, 11-12

Wesolowski, J. (1990). Personal communication (to W. Bryc)

Wessels, J. (1964). Multimodality in a family of probability densities, with application to a linear mixture of two normal densities, *Statistica Neerlandica,* 18, 267-282

White, J. S. (1970). Tables of normal percentile points. *Journal of the American Statistical Association,* 65, 635-638

Whittaker, E. T., and Robinson, G. (1926). *The Calculus of Observations.* Blackie, London

Wilf, H. S. (1988). The quest for normality, *Educational and Psychological Measurement,* 48, 711-712

Wilson, E. B., and Hilferty. M. M. (1931). The distribution of chi-square, *Proceedings of the National Academy of Science* ,17, 684-688

Wold, H. (1948). *Random Normal Deviates* (Tracts for Computers, XXV). Cambridge University Press, Cambridge (Also *Statistica Uppsala,* 3)

Yanushkyavichyus, R. V. (1989). Doctoral thesis. *Abstracts, Vilnius,* *Lithuaniya,* Vilnius University

Zahl, S(1966). Bounds for the central limit theorem error, SIAM Journal of Applied Mathematics,14,1225-1245

Zelen, M., and Severo, N. C. (1964). Probability functions, *Handbook of Mathematical Functions,* M. Abramowitz and I. A. Stegun (editors), 925-995, U. S. Department of Commerce, *Applied Mathematics Series,* 55

Ziegler, R. K. (1965). A uniqueness theorem concerning moment distributions, *Journal of the American Statistical Association,* 60, 1203-1206

Zolotarev,V.M.(1967).A sharpening of the inequality of Berry -Esseen, *Zeitschrift für Wahrscheinlichkeistheorie und verwandte Gebiete,* 8,332-342

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

Ζαχαροπούλου, Χ.(1998). *Στατιστική. Μέθοδοι- Εφαρμογές.* Τόμος Α΄, Θεσσαλονίκη

Κιόχος, Π.Α.(1993). *Στατιστική* .Εκδόσεις INTERBOOKS, Αθήνα

Ξεκαλάκη, Ε.& Πανάρετος, Ι.(1993). *Πιθανότητες και Στοιχεία Στοχαστικών Ανελίξεων.(3η 'Εκδοση)*,Αθήνα

Πανάρετος, Ι., & Ξεκαλάκη Ε. (2000). *Εισαγωγή στη Στατιστική Σκέψη.* Τόμος ΙΙ, (Εισαγωγή στις Πιθανότητες και στην Στατιστική Συμπερασματολογία), Αθήνα

Παπαδήμας, Ο. Γ.(1998). *Στατιστική ΙΙ.* Μακεδονικές Εκδόσεις, Αθήνα

Παπαϊωάννου, Τ. (1981). *Εισαγωγή στις πιθανότητες και την Στατιστική.* Ιωάννινα

Ρούσσας, Γ. Γ. (1973). *Στοιχεία Πιθανοθεωρίας μετ'Εφαρμογών.* Πάτρα

Ξενόγλωσση

Abbe, E. (1906; 1863). Uber die Gesetzmassigkeit in der Vertheilung der Fehler bei Beobachtungsreihen, Gesammelte Abhandlungen, Vol. II, Jena: Gustav Fisher. English translation (1968), ref. P. B. 191, 292, T., from National Technical Information Service, Springfield, Va

Adams, W. J. (1974). *The Life and Times of the Central Limit Theorem.* Caedmon, New York

Adrain, R. (1808). Research concerning the probabilities of errors which happen in making observations, etc., *The Analyst; or Mathematical Museum,* 1(4), 93-109

Agard, J. (1961). Mélange de deux populations normales et étude de quelques fonctions f(x,*y)* de variables normales x,y, R*evue de Statistique Appliquée,* 9, No. 4, 53-70

Ahsanullah, M. and Hamedani, G. G. (1988). Some characterizations of normal distribution, *Calcutta Statistical Association Bulletin, 37,* 95-99

Ahsanullah, M. (1987). A note on the characterization of the normal distribution, *Biometrika Journal 29,* 885-888

Airy, G. B. (1861). *On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations.* Macmillan, London

Alam, K. (1971). A characterization of normality. *Annals of the Institute of Statistical Mathematics, 23,* 523-525

Arnold, B. C. and Meeden, G. (1975). Characterization of distributions by sets of moments of order statistics, *Annals of Statistics, 3,* 754-758

Arnold, B. C., and Isaacson, D. L. (1978). On normal characterizations by the distribution of linear forms, assuming finite variance, *Stochastic Processes and Their Applications,* 7, 227-230

Badhe, S. K. (1976). New approximation of the normal distribution function. *Communications in Statistics—Simulation and Computation, 5,* 173-176

Barnett, V. (1976). Convenient probability plotting positions for the normal distribution. *Applied Statistics,* 25, 47-50

Bartlett, M. S. (1934). The vector representation of a sample. *Proceedings of the Cambridge Philosophical Society, 30,* 327-340

Basu, D. (1955). On statistics independent of a complete sufficient statistic, *Sankhyā, 15,* 377-380

Basu, D. and Laha. R. G. (1954). On some characterizations of the normal distribution. *Sankhyā,* 13, 359-362.

Beasley. J. D. and Springer. S. G. (1977). The percentage points of the normal distribution, *Applied Statistics, 26,* 118-121

Beer, S. and Lukacs, E. (1973). Characterizations of the normal distribution by suitable transformations, *Journal of Applied Probability,10,* 100-108.

Bell, S. (1962). *Approximating the normal distribution with the triangular,* Santia, Corporation Report No. 494

Bergström, H. (1945). On the central limit theorem in the space $R^{k,}$ $k > 1$, *Skandinarisk Akluarictidskrift,* 28, 106-127

Berk, R.H. (1986). Sphericity and the normal law, *Annals of Probability, 14,696-701*

Bernoulli, D. (1770-1771). Mensura sortis ad fortuitam successionem rerum naturaliter contingentium applicata, *Novi Commentarii Academiae*

Scientiarum Imperialis Petropolitanae, 14, 26-45; 15, 3-28

Bernoulli, J. (1713). Ars Conjectandi; German version (1899). *Wahrscheinlichkeitsrechnung*, Engelmann, Leipzig

Bernstein, S. (1941). Sur une propriété caractéristique de la loi de Gauss, *Transactions of the Leningrad Polytechnic Institute, 3*, 21-22; reprinted (1964) in *Collected Works, Vol. 4*, 314-315

Berry, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates, *Transactions of the American Mathematical Society*, 49,122-136

Bessel, F. W. (1818). Astronomiae pro anno MDCCLV deducta ex observationibus viri incomporabilis James Bradley specula Grenovicensi per annos 1750-1762 institutis, Regiomonti

Bessel, F. W. (1838). Untersuchungen über der Wahrscheinlichkeit der Beobachtungsfehler, *Astronomische Nachrichten*, 15, 368-404

Bhattacharjee, G. P., Pandit, S. N. N. and Mohan, R. (1963). Dimensional chains involving rectangular and normal error-distributions, *Technometrics, 5*, 404-406

Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden-Day, San Francisco

Bienaymé, J. (1838). Sur la probabilité des résultats moyens des observations; démonstration directe de la règle de Laplace, *Mémoires des Savans Etrangers, Académie (Royale) des Sciences de l'Institut de France*, Paris, 5, 513-558

Bienayme, J. (1852). Sur la probabilité des erreurs d'après la méthode des moindres carres; reprinted (1858): *Mémoires des Savans Etrangers, Académie (Royale) des Sciences de l'Institut de France*, Paris, 15, 615-663

Blum, J. R. (1956). On a characterization of the normal distribution, *Skandinavisk Aktuarietidakrift, 39*, 59-62

Bondesson, L. (1977). The sample variance, properly normalized, is χ^2-distributed for the normal law only, *Sankyā A, 39*, 303-304

Borges, R. (1970). Eine Approximation der Binomiahverteilung durch die Normalverteilung der Ordnung 1/n, *Zeitschrift für*

Wahrscheinlichkeitstheorie und verwandte Gebiete, 14, 189-199

Borovkov, A. A. and Utev, S. A. (1983). On an inequality and a related characterization of the normal distribution, *Theory of Probability and its Applications, 28,* 219-228

Box, G. E. P. and Muller, M. E. (1958). A note on the generation of random normal deviates, *Annals of Mathematical Statistics,* 29, 610-611

Bravais, A. (1846). Analyse mathématique sur les probabilités des erreurs de situation d'un point, *Mémoires des Savans Etrangers, Académie (Royale) des Sciences de l'Institut de France,* Paris, 9, 255-332

Braverman, M. Sh. (1985). The characteristic properties of normal and stable distributions, *Theory of Probability and its Applications, 30,* 465-474

Bryc, W. (1990). *Normal distributions and characterizations.* University of Cincinnati, Unpublished manuscript Cincinnati, OH

Bryc, W. (1994). *Normal distributions characterizations with applications.* University of Cincinnati, Unpublished manuscript Cincinnati, OH

Bryc, W. and Szablowski. P. J. (1990). Some characteristics of normal distribution by conditional moments, *Bulletin of the Polish Academy of Sciences: Mathematics,* 38, 209-218

Burgess, J. (1898). On the definite integral... [i.e., ...], with extended tables of values, Transactions *of the Royal Society of Edinburgh,* 39, 257-321

Burr, I. W. (1967). A useful approximation to the normal distribution function, with application to simulation, *Technometrics,* 9, 647-651

Buslenko, N. P., Golenko. D. I., Shreidcr. Yu. A., Sobol. **I. M.** and **Sragovich. V. G. (1966).** *The Monte Carlo Method.* Pergamon, Oxford (Original Russian edition published 1962)

Cacoullos, T. (1967b). Characterizations of normality by constant regression of linear statistics on another linear statistic, *Annals of Mathematical Statistics, 38,* 1894-1898

Cacoullos, T. (1967a). Some characterizations of normality, *Sankhyā A, 29,* 399-404

Cacoullos, T. and Papathanasiou, V. (1989). Characterization of distributions by variance bounds, *Statistics and Probability Letters, 7,* 351-356

Cadwell, J. H. (1951). The bivariate normal integral. *Biometrika,* 38, 475-479

Camp, B. H. (1951). Approximation to the point binomial, Annals of *Mathematical Statistics,* 22, 130-131

Carta, IX G. (1975). Low-order approximations for the normal probability integral and the error function, *Mathematics of Computation,* 29, 856-862

Chan, L. K. (1967). On a characterization of distribution by expected values of extreme order statistics, *American Mathematical Monthly, 74,* 950-951

Charlier, C. V. L., and Wicksell, S. D. (1924). On the dissection of frequency functions, *Arkir för Matematik, Astronomi och Fysik,* 18, No. 6, 1-64

Charlier, C. V. L. (1905). Über die Darstellung willkürlicher Funktionen, *Arkiv för* Mathematik, *Astronomin och Fysik,* 2(20), 1-35

Cheng, T. T. (1949) .The normal approximation to the Poisson distribution and a proof of a conjecture of Ramanujan, *Bulletin of the American Mathematical Society,* 55, 396-401

Chew, V. (1968). Some useful alternatives to the normal distribution. *The American Statistician, 22,* No. 3, 22-24

Clow, **R., Hansen, E., and McNolty, F. (1974).** Bayesian density functions for Gaussian pulse shapes in Gaussian noise, *Proceedings of the IEEE,* 62, 134-136

Cohen, **A. C. (1967).** Estimation in mixtures of two normal distributions, *Technometrics,* 9, 15-28

Cramer, H. (1936). Über eine Eigenschaft der normalen Verteilungsfunktion. *Mathematische Zeitschrift,* 41, 405-414

Cramer, H. (1946). *Mathematical Methods of Statistics.* Princeton, N.J.: Princeton University Press

Csörgö, **M. and Seshadri, V. (1971).** Characterizing the Gaussian and exponential laws via mappings onto the unit interval, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 18,* 333-339

Csörgö, **M., Seshadri. V. and Yalovsky, M. (1975).** Application of characterizations in the area of goodness-of-fit, In *Statistical Distributions in Scientific Work* (G. P.Patil, ed.), Dordrecht, Netherlands: Reidel, 2, 79-90

Daniel, C.(1959). Use of half-normal plots in interpreting factorial two-level experiments, *technometrics,* 1,311-341

Darmois, G. (1951). Sur une propriété caractéristique de la loi de probabilité de Laplace, *Comptes Rendus de I'Académie des Sciences* (Paris) 232, 1999-2000

Darmois, G. (1953). Analyse gencrale des liasons stochastiques. Etude particuliaire de I'analyse factorielle lincaire, *Rente de I'lnstiiut Internationale de Statislique,* 21, 2-8

David, F. N. (1962). Games, Gods, and Gambling. Hafner, New York

Daw, R. H. and Pearson, E. S. (1972). Studies in the history of probability and statistics, XXX: Abraham de Moivre's 1733 derivation of the normal curve: A bibliographical note, *Biometrika,* 59, 677-680

De Moivre, A. (1733). Approximatio ad Summan Ferminorum Banomii $(a+b)^n$ in Seriem expansi

De Moivre, A. (1738, 1756). *The Doctrine of Chances;* reprint (1967). Chelsea, New York

De Morgan, A. (1837). Theory of probabilities, in *Encyclopedia, Metropolitana, 2,* 359-468

De Morgan, A. (1838). An essay on probabilities and on their application to life contingencies and insurance offices. Longmans, London

Derenzo, S. E. (1977). Approximations for hand calculators using small integer coefficients, *Mathematics of Computation,* 31, 214-225

Diaconis, P., Olkin, I. and Ghurye, S. G. (1977). Review of *Characterization Problems in Mathematical Statistics* (by A. M.Kagan, Yu. V. Linnik, and C. R. Rao), *Annals of Statistics, 5,* 583-592

D'Ortenzio, R. J. (1965). Approximating the normal distribution function, *Systems Design,* 9, 4-7

Dubey, S.D.(1967). Normal and Weibull distributions, *Naval research Logistic Quarterly*, 14,69-79

Dynkin, E. B. (1961). Necessary and sufficient statistics for a family of probability distributions, *Selected Translations in Mathematics, Statistics, and Probability,* 1, 17-40

Edgeworth, F. Y. (1883). The law of error, *Philosophical Magazine,* 16, 300-309

Edgeworth, F. Y. (1892). Correlated averages, *Philosophical Magazine* Ser. 5, 34, 190-204

Edgeworth, F. Y. (1905). The law of error, Proceedings *of the Cambridge Philosophical Society,* 20, 36-65

Edwards, **A. W. F. (1963).** A linkage for drawing the normal distribution, *Applied Statistics,* 12, 44-45

Eisenberger, I. (1964). Genesis of bimodal distributions. *Technometrics,* 6, 357-363

Esseen, C. G. (1942). On the Liapounoff limit of error in the theory of probability, *Arkiv för Mathematik, Aslronomi och Fysik,* 28A, 1-19

Esseen. C. G. (1958). On mean central limit theorems, *Kunglige Tekniska Högskolans Handlingar,* No. 121, 1-30

Feller, W. (1935). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung, *Mathematische Zeitschrift,* 40, 521-559

Feller, W. (1968). *Introduction to Probability Theory and Its Applications,* Vol. 1 (3rd ed.).Wiley, New York

Ferguson, T. S. (1967). *Mathematical Statistics-A Decision Theoretic Approach.* Academic, New York

F**leming, N. S. (1989).** Approximating the normal cumulative distribution function using a spreadsheet program. *The American Statistician,* 43, 68, Reply.Ibid.,43,290-291

Fletcher, **A. Miller, J. C. P.,** Rosenhead, **L. and Comrie, L. J. (1962B).** *An Index of Mathematical Tables,* Vol.II (2nd ed.),Reading, Mass .:Addison –Wesley, Scientific Computing Servise

Fox, C. (1965). A family of distributions with the same ratio property as normal distribution, *Canadian Mathematical Bulletin,* 8, 631-636

Freeman, M. F., and Tukey, J. W. (1950). Transformations related to the angular and the square root, *Annals of Mathematical Statistics,* 21,607-

611

Galileo, G. (trans. 1953; 2nd ed. 1962). *Dialogue Concerning the Two Chief* World Systems-Ptolemaic and Copernican (trans. S. Drake), University of California Press, Berkeley

Galton, F. (1869). *Hereditary Genius*. Macmillan, London

Galton, F. (1875). Statistics by intercomparison, with remarks on the law of frequency of error, *Philosophical Magazine*, 49, 33-46

Galton, F. (1886). Family likeness in stature, *Proceedings of the Royal Society*, 40, 42-73

Galton, F. **(1888).** Correlations and their measurement, chiefly from anthropometric data, Proceedings of the *Royal Society*, 45, 135-145

Galton, F. **(1889).** *Natural Inheritance* .Macmillan, London

Gauss, C. **F. (1809).** *Theoria Motus Corporum Coelestium*, Lib. 2, Sec. III, 205-224, Hamburg: Perthes u. Besser

Gauss, C. F. (1816). Bestimmung der Genauigkeit der Beobachtungen, *Zeitschrift Astronomi* ,1, 185-197

Geary, R. C. (1936). Distribution of "Student's" ratio for nonnormal samples, *Journal of the Royal Statistical Society B3*, 178-184

Gebhardt, F. (1969). Some numerical comparisons of several approximations to the binomial distribution, *Journal of the American Statistical Association*, 64, 1638-1646

Gebhardt, F. (1971). Incomplete Beta-integral B(x;2/3,2/3) and $[p(1 - p)]^{-1/6}$ for use with Borges' approximation of the binomial distribution, *Journal of the American Statistical Association*, 66, 189-191

Geisser, S. (1956). A note on the normal distribution, *Annals of Mathematical Statistics*, *27*, 858-859

Geisser, S. **(1973).** Normal characterizations via the square of random variables, *Sankyā A,35, 494-494*

Geisser, S. (1973). Normal characterization via the squares of random variables, *Sankhyā. Series A*, 35,492-495

Ghurye, S. G., and Olkin, I. (1973). Identically distributed linear forms and the normal distribution, *Advances in Applied Probability*, *5*, 138-152

Glaisher, J. W. L. **(1871).** On a class of definite integrals-Part II, *Philosophical Magazine*, 42, 421-436.

Gnedenko, B. V., and Kolmogorov, A. N. (1968). *Limit Distributions for Sums of Independent Random Variables* (trans. from Russian), Addison-Wesley ,Reading, Mass

Govindarajulu, Z. (1966). Characterization of normal and generalized truncated normal distributions using order statistics, *Annals of Mathematical Statistics, 37*, 1011-1015

Govindarajulu, Z. (1965). Normal approximations to the classical discrete distributions, Sankhyā A,27, 143-172

Greenwood, J. A., and Hartley, H. 0. (1962). *Guide* to *Tables in Mathematical Statistics*, Princeton University Press, Princeton, N.J

Hamaker, **H. C.** (1978). Approximating the cumulative normal distribution and its inverse, *Applied Statistics*, 27, 76-77

Hart, **R. G. (1966).** A close approximation related to the error function. *Mathematics of Computation*, 20, 600-602

Hartman, P. and Wintner (1940). On the spherical approach to the normal law, *American Journal of Mathematics*, 62, 759-779

Hastings, C. ***(1955).*** *Appimations for Digital Computers.* Prinecton NJ, Princeton University Press

Helguero, F. dc (1904). Sui massima delle curve dimorfiche, *Biometrika*, 3. 84-98

Helguero, F. dc (1905). Per la risoluzione delle curve dimorfiche, *Biometrika*, 4,230-231

Helmert, **F. R. (1876).** Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers director Beobachtungen tungen gleicher Genauigkeit, *Astronomische Nachrichten*, 88,113-120

Hemelrijk, J. (1967). The hypergeometric, the normal and chi-squared, *Statistica Neerlandica*, 21, 225-229

Hombas, V. C. (1985). Characterizing the normal density as a solution of a differential equation, *Statistica Neerlandica, 39*, 387-388

Hoyt, J. P. (1968). A simple approximation to the standard normal probability density function, *The American Statistician*, 22, No. 2, 25-26

Johnson, N. L .,Kotz, S.and Kemp, A. W.(1992).*Univariate Discrete Distributions*(2nd edn.).Wiley, New York

Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions, Vol. 1.*Wiley, New York

Johnson, N. L. and Kotz, S. (1969). *Distributions in Statistics: Discrete Distributions.* Wiley, New York

Johnson, N.L. and Kotz, S.(1970a).*Distributions in Statistics: Continuous Univariate Distributions*, Vol. 1 . Wiley, New York

Johnson, **N.L. and Kotz, S.(1970b).***Distributions in Statistics: Continuous Univariate Distributions*, Vol. 2 .Wiley, New York

Kac, **M. (1975).** Some reflections of a mathematician on the nature and the role of statistics, Proceedings *of the Conference on Directions for Mathematical Statistics,* 5-11, Applied Probability Trust

Kagan, A. M., Linnik, Yu. V. and Rao, C. R. (1965). On a characterization of the normal law based on a property of the sample average. *Sankya A,27,* 405-406

Kagan, A. M., Linnik, Yu. V. and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics* (trans. B. Ramachandran). Wiley, New York

Kagan A. M. and Shalayevskii. O.V. (1967). Characterization of the normal law by a property of the non-central χ^2-distribution. *Lietutos Matematikos Rinkinys,* 7, 57-58 (In Russian)

Kaplansky, I. (1943). A characterization of the normal distribution, *Annals of Mathematical Statistics, 14,* 197-198

Kelker, D. and Matthes, R. K. (1970). A sufficient statistics characterization of the normal distribution. *Annals of Mathematical Statistics,* 41, 1086-1090

Kelley, **T. L. (1948).** *The Kelley Statistical Tables.* Harvard University Press, Cambridge

Kendall, M. G. (1963). Ronald Aylmer Fisher, 1890-1962, *Biometrika* 50, 1-15; reprinted (1970) in Studies in *the History of Probability and Statistics* (E. S. Pearson and M. G. Kendall, eds.), 439-454. Hafner, New York

Kendall, M. G. (1971). Studies in the history of probability and statistics, XXVI: The work of Ernst Abbe, *Biometrika,* 58, 369-373; Corrigendum (1972): *Biometrika,* 59, 498

Kerridge, D. F., and Cook, G. W. (1976). Yet another series for the normal integral, *Biomelrika,* 63, 401-403

Khatri, C. G. and Rao, C. R. (1972). Functional equations and characterization of probability laws through linear functions of random variables, *Journal of Multivariate Analysis ,2,* 162-173

Khatri, C. G. (1979). Characterization of multivariate normality, II: Through linear regressions, *Journal of Multivariate Analysis ,9,* 589-598

Khatri, C. G. and Rao, C. R. (1976). Characterizations of multivariate normality, I: Through independence of some statistics, *Journal of Multivariate Analysis, 6,* 81-94

Konwar, R. M. (1991). On characterizations of distributions by mean absolute deviation and variance bounds, *Annals of the Institute of Statistical Mathematics, 43,* 287-295

Kotlarski, I. (1966). On characterizing the normal distribution by Student's law, *Biometrika, 53,* 603-606

Kotlarski, I. (1967). On characterizing the gamma and the normal distribution, *Pacific Journal of Mathematics, 20,* 69-76

Kotz, S. (1974). Characterization of statistical distributions: A supplement to recent surveys, *International Statistical Review, 42,* 39-65

Kramp, C. (1799). Analyse des réfractions astronomiques et terrestres, Leipsic: Schwikkert, Koenig, Paris

Kruskal, W. (1978). Formulas, numbers, words: Statistics in prose, *The American Scholar,* 47, 223-229

Laha, R. G. (1953). On an extension of Geary's theorem, *Biometrika, 40,* 228-229

Lancaster, H. 0. (1954). Traces and cumulants of quadratic forms in normal variables, *Journal of the Royal Statistical Society B,*16, 247-254

Lancaster, H. 0. (1966). Forerunners of the Pearson χ^2, *Australian Journal of Statistics,* 8, 117-126

Lancaster, H. 0. (1987). Finiteness of the variances in characterizations of the normal distribution, Australian *Journal of Statistics, 29,* 101-106

Laplace, P. S. (1809-1810). Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilités, *Mémoires* de *la Classe des Sciences mathématiques et physiques de l'Institut,* Paris, 353-415, 559-565, et. passim (see Oeuvres Complètes ,12, 301-353)

Laplace, P. S. (1810). Mémoire sur les intégrales définies et leur application aux probabilités, *Mémoires de la Classe des Sciences mathématiques et physiques de l'Institut,* Paris, 279-347, et. passim (1810-1812: see Oeuvres *Complètes,* 12, 357-412)

Laplace, P. S. (1812). Tnéorie *Analytique des Probabilités,* Paris (Oeuvres Complètes, Vol. 7)

Laplace, P. S. (1878, 1912). Oeuvres Complètes de *Laplace,* 14 volumes. Gauthier –Villars, Paris

Legendre, A. M. (1805). *Nouvelles Méthodes pour la Détermination* des *Orbites des Comètes,* Paris

Legendre, A. M. (1826). Traite des *Fonctions elliptiques et des Intégrales Euleriennes, avec des Tables pour en faciliter le Calcul numérique,* Vol. 2, Huzard –Courcier, Paris

Lehmann, **E. L. (1986).** *Testing Statistical Hypotheses* (2nd edn.).Wiley, New York

Lin, J. T. (1988). Alternatives to Hamaker's approximations to the cumulative normal distribution and its inverse, *The Statistician,* 37, 413-414

Lin, J. -T. (1989). Approximating the normal tail probability and its inverse for use on a pocket calculator, *Applied Statistics,* 38, 69-70

Lindeberg,J.W.(1922). Einie neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung, *Mathematische Zeitschrift,*15,211-225

Linders, F. J. (1930). On the addition of two normal frequency curves, *Nordic Statistical Journal, 2,* 63-73

Ling, **R . F.and Pratt, J. W. (1984).**The accuracy of Peizer approximations to the Hypergeometric Distribution, with comparisons to some other approximations, *Journal of the American Statistical* Association ,79, 49-60

Linnik, Yu. V. (1952). Linear statistics and the normal distribution law. *Doklady Akademii Nauk SSSR,* 83, 353-355. (In Russian. English translation published be American Mathematical Society. 1961.)

Lukacs, E. and Laha, R. G. (1964). *Applications of Characteristic Functions.* Hafner, New York

Lukacs, E. (1956). Characterization of populations by properties of suitable statistics, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 2,* 195-214

Lukacs, E. (1942). A characterization of the normal distribution, *Annals of Mathematical Statistics, 13,* 91-93

Lukacs. E. (1955). Applications of Faà di Bruno's formula in mathematical statistics. *American Mathematical Monthly,62,* 340-348

Lyapunov, A. M. (1900). Sur une proposition de la théorie des probabilités. *Izvstiya Akademii Nauk SSSr, series V,*13,359-386.[Also Lyapunov, A.(1954).*Collected Works,*1,125-151,Moscow :Akademia Nauk SSSR.(In Russian)]

Lyapunov, **A. M. (1901).** Nouvelle forme du théorème sur la limite de probabilité, *Mémoires de* l'Academic *Impériale des Sciences de St. Pétersbourg,* 12, 1-24

Lyapunov, **A. M. (1954-1965).** *Izbrannye Trudi* (Selected Works), Academy of Sciences, USSR

Macino,T.(1984). Mean hazard rate and itw application to the notmal approximation of the Weibull distribution, *Naval Research Logistics Quarterly,* 31,1-8

Mahalanobis, P.C, Bose, S. S., Roy. P. R., and Banerjee, S. K. (1934). Tables of random samples from a normal distribution, *Sankhya,* 1, 289-328

Maistrov, L. E. (1967, 1974). Probability *Theory: A Historical Sketch* (trans. by S. Kotz), New York: Academic Press

Makabe, H., and Morimura, H. (1955). A normal approximation to the Poisson distribution, *Reports on Statistical Applications Research* (Union of Japanese Scientists and Engineers) ,4, 37-46

Marcinkiewicz, J. (1939). Sur une propriété de la loi de Gauss, *Mathematische Zeitschrift* ,*44*, 612-618

Markov, A. A. (1888). Table des valeurs de l'integrale..., St. Pétersbourg: Académie Impériale des Sciences

Markov, A. A. (1899-1900). The law of large numbers and the method of least squares, *Izvestia Physiko -mathematicheskago Obschvestva pri Imperatorskom Kazanskom Universitet*, 8, 110-128; 9, 41-43

Markov, A. A. (1913). A probabilistic limit theorem for the cases of Academician A. M. Lyapounov, *Izvlyechenye iz knigi ischislyenge veroyatnostyei* (supplement to "Theory of Probability"), 4-e

Markov, A. A. (1951). *Izbrannye Trudi* (Selected Works), Academy of Sciences, USSR

Mathai, A. M., and Pederzoli, G. (1977). *Characterizations of the Normal Probability Law*. Wiley, New York

Maxwell, J. C. (1860). Illustrations of the dynamical theory of gases, *Philosophical Magazine*, 19, 19-32; 20, 21-37; (1952): *Scientific Papers of James Clerk Maxwell*, 377-409

McGillivray, W. R., and Kaller, C. L. (1966). A characterization of deviation from normality under certain moment assumptions, *Canadian Mathematical Bulletin*, 9,509-514

McLachlan, G. J., and Basford, K. E. (1987). *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, New York

Meshalkin, L. D. (1968). On the robustness of some characterizations of the normal distribution, *Annals of Mathematical Statistics* ,*39*, 1747-1750

Molenaar, W. (1970). *Approximations to the Poisson, Binomial, and Hypergeometric Distribution Functions*, Mathematical Centre Tracts, Vol. 31.Mathematisch Centrum, Amsterdam

Molenaar, I. W. (1985). Normal approximations to the Poisson, Binomial, negative binomial and Hypergeometric Distribution, *Encyclopedia of Statistical Sciences*,Vol. 6(S.Kotz, N.L.Johnson and C.B.Read,eds.),340-347. Wiley, New York

Molenaar. W. (1965). Survey of separation methods for two normal distributions . *Statislica Neerlandica*, 19, 249-263

Mood, A. M.., Graybill, F. A. and Boes, D. C. (1974). *Introduction to the*

Theory of Statistics. McGraw-Hill, New York

Moran, P. A. P. (1980). Calculation of the normal distribution function. *Biometrika,* 67, 675-676

Nelson, L. S. (1976). Constructing normal probability paper, *Journal of Quality Technology,* 8, 56-57

Nhu, H. H. (1968). On the stability of certain characterizations of a normal population, *Theory of Probability and Its Applications, 13,* 299-304

Nicholson, W. L. (1956). On the normal approximation to the hyper-geometric distribution, Annals *of Mathematical Statistics , 27,* 471-483

Norton, R. M. (1989). Pocket-calculator approximation for areas under the standard normal curve, *The American Statistician,* 43, 24-26. Reply, *Ibid.,* 43, 290

Nye, H. H. (1966). On the stability of certain theorems characterizing normal distribution, *International Congress of Mathematicians: Information Bulletin,* 6, 5

Odeh, R. E. and Evans, J. O. (1974). Algorithm AS 70. The percentage points of the normal distribution, *Applied Statistics,* 23, 96-97

Pagano, **M. and Gauvreau, K.(1996).** *Αρχές Βιοστατιστικής.* Μετάφραση - Επιμέλεια: Ουρανία Δάφνη (2000), Εκδόσεις "ΕΛΛΗΝ", Αθήνα

Page, E. (1977). Approximations to the cumulative normal function and its inverse for use on a pocket calculator, *Applied Statistics, 26,* 75-76

Pakshirajan, P., and Mohan, N. C. (1971). A characterization of the normal law, *Annals of the Institute of Statistical Mathematics,* 21, 529-532

Pakshirajan, R. P. and Mohan, N. R. (1969). A characterization of the normal law, *Annals of the Institute of Statistical Mathematics, 21,* 529-532

Patel, **J. K. and Read, C. B. (1981).** *Handbook of the Normal Distribution.* Marcel Dekker, New York

Patel, J. K. and Read, C. B. (1996). *Handbook of the Normal Distribution.* Marcel Dekker, New York

Pathak, **P. K. and Phillai, R. N. (1968).** On a characterization of the normal law, *Sankya A,30,* 141-144

Patil, G. P. and Seshadri, V. (1964). Characterization theorems for some univariate probability distributions, *Journal of the Royal Statistical Society B,26,* 286-292

Pearson, E. S. (1965). Some incidents in the early history of Biometry and Statistics, *Biometrika* 52, 3-18; reprinted (1970) in *Studies in the History of Statistics and Probability* (E. S. Pearson and M. G. Kendall, eds.), 323-338. Hafner, New York

Pearson, E. S. (1967). Some reflexions on continuity in the development of mathematical statistics, 1885-1920, *Biometrika* 54, 341-355; reprinted (1970), as in Pearson (1965), 339-353

Pearson, E. S. and Wishart, J., eds. (1958). "Student's" Collected *Papers.* Cambridge University Press, Cambridge

Pearson, K. (1894). Contributions to the mathematical study of evolution. *Philosophical Transactions of the Royal Society of London, Series A,* 185, 71-110

Pearson, K. (1896). Mathematical contributions to the theory of evolution, III: Regression, heredity, and panmixia, *Philosophical Transactions of the Royal Society of London* A,187, 253-318

Pearson, K. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philosophical Magazine* (5)50, 157-175

Pearson, K. (1921-1933). The *History of Statistics in the 17th and18th Centuries, against the Changing Background of Intellectual, Scientific, and Religious Thought* (lectures, E. S. Pearson, ed., 1978). Macmillan, New York

Pearson, K. (1920). Notes on the history of correlation, *Biometrika* 13, 25-45; reprinted (1970) in Studies in the *History of Statistics and Probability* (E. S. Pearson and M. G. Kendall, eds.), 185-205. Hafner, New York

Peizer, D. B. and Pratt, J. W. (1968). A normal approximation for binomial, F, beta, and other common, related tail probabilities, I, *Journal of the American Statistical Association,* 63, 1416-1456

Plackett, R. L. (1972). Studies in the history of probability and statistics, XXIX: The discovery of the method of least squares, Biometrika, 59, 239-251

Plait, A. (1962). The Weibull distribution-with tables, *Industrial Quality control* ,19(5),17-26

Pólya, G. (1932). Verleitung des Gauss' chen Fehlergesetzes aus einer Funktional- gleichung. *Mathematische Zeitschrift,* 18, 96-108

Pólya, G. (1945). Remarks on computing the probability integral in one and two dimensions. *Proceedings of the 1st Berkeley Symposium on Mathematical Statistics and Probability,* 63-78

Prakasa Rao, B.L.S. (1979). Characterizations of distributions through some identities, *Journal of Applied Probability, 16,* 903-909

Prasad, **A. (1955).** Bi-modal distributions derived from the normal distribution, *Sankltya,* 14,369-374

Quetelet, L. A. J. (1846). Lettres a *S. A. R. Le Due Régnant de Saxe-Cobourg et Gotha, sur la Théorie des Probabilités appliquée aux Sciences Morales et Politiques* (English trans., 1849). Hayez ,Brussels

Raab, D. H. and Green. E. H. (1961). A cosine approximation to the normal distribution, *Psychometrika, 26,* 447-450

Raff, M. S. (1956). On approximating the point binomial, Journal *of the American Statistical Association,* 51, 293-303

Ramasubramaniam, S. (1985). A characterisation of the normal distribution,. *Sankhyā A,* 410-414

RAND Corporation (1955). *A Million Random Digits with 100,000 Normal Deviates Glencoe, IL: Free Press*

Rao, C. R. (1967). On some characterizations of the normal law, *Sankya A*,29, 1-14

Rao, C. R. **(1973).** *Linear Statistical Inference and Its Applications* (2nd edn.).Wiley, New York

Rao, J. **N. K. (1958).** A characterization of the normal distribution. *Annals of Mathematical Statistics,* 29, 914-919

Read, C. B. (1985). Normal distribution, *Encyclopedia of Statistical Sciences,* 6. S. Kotz, N. L. Johnson, and C. B. Read (editors), 347-359. Wiley, New York:

Riffenburgh, R. H. (1967). Transformation for statistical distribution approximately normal but of finite sample range, Report NUWC-TP-19, Naval Undersea Warfare Center, San Diego, CA [Abbreviated version *Technometrics,* 11, (1969), 47-59]

Riordan, J. (1949). Inversion formulas in normal variable mapping, *Annals of Mathematical Statistics,* 20, 417-425

Rivest, L. P. (1981). On the sum of contaminated normals. *The American Statistician* ,35, 155-156

Rouncefield, M. (1990). Is it normal. How to use probability (or normal) graph paper, *Teaching Statistics,* 12, 6-8

Ruben, H. (1974). A new characterization of the normal distribution through the sample variance, *Sankya A,*36, 379-388

Ruben, H. (1975). A further characterization of normality through the sample variance, *Sankya AS,*7, 72-81

Ruben, H. (1976). A characterization of normality through the general linear model, *Sankya A,*38, 186-189

Sadikova, S. M.(1966). On two-dimensional analogs of an inequality of C.G. Essen and their application to the central limit theorem, *Teoriya Veroyatnostei i ee Primeneniya,*11,370-380.(In Russian)

Savage, L. J. (1976). On rereading R. A. Fisher (with discussion), *Annuls of Statistics,* 4, 441-500

Sazanov, V. V.(1967). On the rate of convergence in the multidimensional central limit theorem, *Teoriya Veroyatnostei i ee Primeneniya,*12,82-95.(In Russian)

Schmeiser, B. W. (1979). Approximations to the inverse cumulative normal function for use on hand calculators, *Applied Statistics,* 28, 175-176.

Schols, C. M. (1875). Over de theorie des fouten in de ruimte en in het platte vlak, *Verhandelungen der koninklijke Akademie van Wetenschappen,* Amsterdam, 15, 1-75

Schonfelder, J. L. (1978). Chebyshev expansions for the error and related functions, *Mathematics of Computation,* 32, 1232-1240

Schucany, W. R. and Gray, H. L. (1968). A new approximation related to the error function, *Mathematics of Computation,* 22, 201-202

Seal, H. L. (1967). Studies in the history of probability and statistics, XV: The historical development of the Gauss linear model, Biometrika, 54, 1-24; reprinted (1970) in *Studies in the History of Probability and Statistics* (E. S. Pearson and M. G. Kendall, eds.), 207-230. Hafner, New York

Sengupta, J. M. and Bhattacharya, N. (1958). Tables of random normal deviates, *Sankhya*, 20, 250-286

Seshadri, V. (1969). A characterization of the normal and Weibull distributions, *Canadian Mathematical Bulletin* ,12, 257-260

Shah, A. K. (1985). A simpler approximation for areas under the standard normal curve, *The American Statistician,* 39, 80. Correction, *Ibid.*, 39, 327

Sheynin, 0. B. (1966). Origin of the theory of errors, Nature 211, 1003-1004

Sheynin, 0. B. (1968). Studies in the history of probability and statistics, XXI: On the early history of the law of large numbers, Biometrika, 55, 459-467; reprinted (1970) in Studies in the *History of Statistics and Probability* (E. S. Pearson and M. G. Kendall, eds.), 231-239. Hafner, New York

Sheynin, 0. B. (1970). Studies in the history of probability and statistics, XXIII: Daniel Bernoulli on the normal law, Biometrika, 57, 199-202

Shimizu, R. (1961). A characterization of the normal distribution, *Annals of the Institute of Statistical Mathematics* , 13, 53-56

Shimizu, R. (1962). Characterization of the normal distribution, II, *Annals of the Institute of Statistical Mathematics,* 14, 173-178

Shore, H. (1982). Simple approximations for the inverse cumulative function, the density function and the loss integral of the normal distribution. *Applied Statistics,* 31, 108-114

Skitovitch, V. P. (1953). On a property of the normal distribution, *Doklady Akademii Nauk SSSR,* 89, 217-219 (in Russian)

Smirnov, N. V. (ed.) (1960, 1965). *Tables of the Normal Probability Integral, Normal Density, and Its Normalized Derivatives.* Macmillan, New York (Translation)

Spiegel, M.R. and Stephens L.J.(1999,1988,1961). *Στατιστική.* (3η έκδοση), McGraw-Hill: New York. Μετάφραση:Εκδόσεις ΤΖΙΟΛΑ, Θεσσαλονίκη

Spiegel, M. R. (1975). *Πιθανότητες και Στατιστική*. McGraw-Hill: New York Μετάφραση: Περσίδης Σ. Κ.(1977).Εκδόσεις ΕΣΠΙ, SCHAUM 3,Αθήνα

Sprott, D. A. (1978). Gauss's contributions to statistics, *Historia Mathematica*, 5, 183-203

Steffensen, J. F. (1937). On the semi-normal distribution, *Skandinarisk Akttiarietidskrift*, 20, 60-74

Stepniak, C. (1991). On characterization of the normal law in the Gauss Markov model, *Sankya* A, 115-117

Stigler, S. M. (1975a). Studies in the history of probability and Statistics, XXXIV: Napoleonic statistics: The work of Laplace, *Biometrika*, 62, 503-517

Stigler, S. M. (1975b). The transition from point to distribution estimation, Bulletin *of the International Statistical Institute* 46 (Proceedings of the 40th Session, Book 2), 332-340

Stigler, S. M. (1977). An attack on Gauss, published by Legendre in 1820, *Historia* Mathematics ,4, 31-35

Stigler, S. M. (1978). Francis Ysidro Edgeworth, statistician, Journal *of the Royal Statistical Society* A,141, 287-313, followed by discussion

Stigler, S. M. (1982). A modest proposal: A new standard for the normal, *The American Statistician*, 36, 137-138

Stigler, S. M. (1973). Laplace, Fisher, and the discovery of the concept of sufficiency, *Biometrika*, 60, 439-445

Strecock, A. J. (1968). On the calculation of the inverse of the error function. *Mathematics of Computation*, 22, 144-158

Stuart, A. and Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics, Vol. 1* (5th edn.). Oxford University Press, New York

Student (1908). The probable error of a mean, *Biometrika*, 6, 1-25

Tamhankar, M. V. (1967). A characterization of normality, *Annals of Mathematical Statistics*, 38, 1924-1927

Taylor, B. J. R. (1965). The analysis of polymodal frequency distributions, *Journal of Animal Ecology*, 34, 445-452

Tchebyshev, P. L. (1890). Sur deux théorèmes relatifs aux probabilités, Acta

Mathematica, 14, 305-315; reprinted (1962) in Oeuvres, Vol. 2. Chelsea, New York

Teicher, H. (1961). Maximum likelihood characterization of distributions, *Annals of Mathematical Statistics,* 32, 1214-1222

Teichroew, D. (1957). The mixture of normal distributions with different variances, *Annals of Mathematical Statistics,* 28, 510-512

Tippett, L. H. C. (1927). *Random Sampling Numbers* (Tracts for Computers XV).Cambridge University Press, Cambridge

Tranquilli, G. B. (1966). Sul teorema di Basu-Darmois, *Giornale dell'Istituto Italiano degliAttuari,29,135-152*

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics,* 5, 99-114

Uspensky, J. V. (1937). Introduction to Mathematical Probability. McGraw-Hill, New York

Wallace, D.(1959).*A corrected computation of Berry's bound for the central limit theorem error.* Statistics Research Center, University of Chicago

Wesolowski, J. (1987). A regressional characterization of the normal law, *Statistics and Probability Letters,* 6, 11-12

Wesolowski, J. (1990). Personal communication (to W. Bryc)

Wessels, J. (1964). Multimodality in a family of probability densities, with application to a linear mixture of two normal densities, *Statistica Neerlandica,* 18, 267-282

White, J. S. (1970). Tables of normal percentile points. *Journal of the American Statistical Association,* 65, 635-638

Whittaker, E. T., and Robinson, G. (1926). *The Calculus of Observations.* Blackie, London

Wilf, H. S. (1988). The quest for normality, *Educational and Psychological Measurement,* 48, 711-712

Wilson, E. B., and Hilferty. M. M. (1931). The distribution of chi-square, *Proceedings of the National Academy of Science* ,17, 684-688

Wold, H. (1948). *Random Normal Deviates* (Tracts for Computers, XXV). Cambridge University Press, Cambridge (Also *Statistica Uppsala,* 3)

Yanushkyavichyus, R. V. (1989). Doctoral thesis. *Abstracts, Vilnius,* *Lithuaniya,* Vilnius University

Zahl, S(1966). Bounds for the central limit theorem error, SIAM Journal of Applied Mathematics,14,1225-1245

Zelen, M., and Severo, N. C. (1964). Probability functions, *Handbook of Mathematical Functions,* M. Abramowitz and I. A. Stegun (editors), 925-995, U. S. Department of Commerce, *Applied Mathematics Series,* 55

Ziegler, R. K. (1965). A uniqueness theorem concerning moment distributions, *Journal of the American Statistical Association,* 60, 1203-1206

Zolotarev,V.M.(1967).A sharpening of the inequality of Berry -Esseen, *Zeitschrift für Wahrscheinlichkeistheorie und verwandte Gebiete,* 8,332-342

Δωρεά