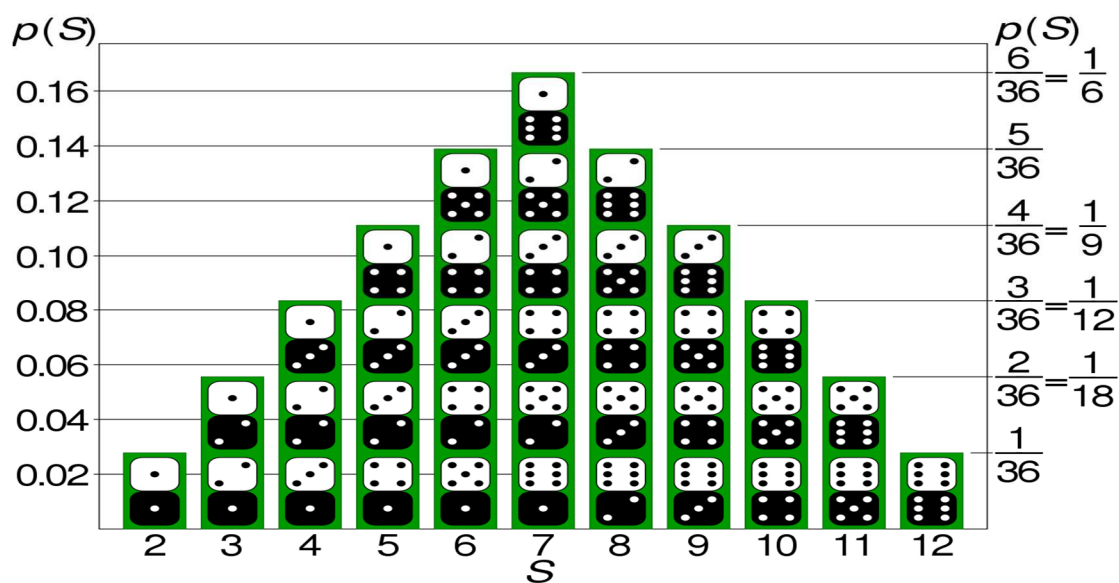




Ακαδημία Εμπορικού Ναυτικού Ασπροπύργου
Σχολή Μηχανικών



Πτυχιακή εργασία
Ειδικές κατανομές πιθανότητας

Σπουδαστές: **Καρακωνσταντής Κωνσταντίνος** (AM 9046)
Καναβάκης Γεώργιος (AM 9044)

Επιβλέπων Καθηγητής: **Στέφανος Ι. Καρναβάς**, Μαθηματικός (M.Ed.)
Επίκουρος Καθηγητής

Ακαδημαϊκό έτος: **2024–2025**

Ημερομηνία ανάθεσης: **15.12.2023**
 Ημερομηνία κατάθεσης: **....12.2024**
 Ημερομηνία εξέτασης: **....01.2025**

A/A	Ονοματεπώνυμο	Χαρακτηρισμός	Υπογραφή
1	Στέφανος Ι. Καρναβάς Μαθηματικός (M.Ed.) Επίκουρος Καθηγητής	Άριστα 10	
2	Δρ. Λάλου Παναγιώτα Μαθηματικός Καθηγήτρια		
3	Τσαγκανός Γεώργιος Μηχανικός Αυτοματισμού (M.sc.) ΕΔΙΠ		
Τελικός χαρακτηρισμός			

Περίληψη

Οι ειδικές κατανομές πιθανότητας, είναι μαθηματικές κατανομές που περιγράφουν τη συμπεριφορά τυχαίων μεταβλητών σε διάφορες καταστάσεις. Αυτές οι κατανομές, χρησιμοποιούνται για να μοντελοποιήσουν πολλές φυσικές, κοινωνικές και οικονομικές διαδικασίες. Σε αυτή την εργασία, παρουσιάζονται μερικές από τις πιο γνωστές ειδικές κατανομές, οι οποίες έχουν πολλές εφαρμογές σε διάφορους τομείς της επιστήμης, της μηχανικής και της καθημερινής ζωής.

Λέξεις κλειδιά

Διωνυμικό πείραμα, δοκιμή Bernoulli, διωνυμική κατανομή, υπεργεωμετρική κατανομή, γεωμετρική κατανομή, κατανομή Pascal, αρνητική διωνυμική κατανομή, κατανομή Poisson, διαδικασία Poisson, ομοιόμορφη κατανομή (ή ορθογώνια κατανομή), εκθετική κατανομή, αμνησία εκθετικής κατανομής, κατανομή B, κατανομή Γ, κανονική κατανομή, κατανομή Gauss, συνάρτηση πιθανότητας, αθροιστική συνάρτηση κατανομής, συνάρτηση πιθανότητας, κατανομή Cauchy, κατανομή Laplace, διπλή εκθετική κατανομή, λογαριθμοκανονική κατανομή, κατανομή Weibull, κατανομή Maxwell, κατανομή Cumbel, κατανομή Pareto, Polya, λογαριθμική κατανομή, αρνητική υπεργεωμετρική κατανομή

Περιεχόμενα

Περίληψη	2
Λέξεις κλειδιά	2
Εισαγωγή	4
1. Στατιστική και ανάλυση δεδομένων	4
2. Μηχανική και φυσική	4
3. Οικονομία και χρηματοοικονομικά	4
4. Βιοστατιστική και ιατρική	4
5. Ασφάλεια και βιομηχανία	5
6. Μοντελοποίηση και προσομοιώσεις	5
7. Μηχανική μάθηση και τεχνητή νοημοσύνη	5
8. Αστρονομία και κοσμολογία	5
9. Κοινωνικές επιστήμες	5
10. Μάρκετινγκ και επιχειρήσεις	5
1. Ειδικές διακριτές κατανομές	6
1.1 Διωνυμικό πείραμα – δοκιμές Bernoulli	6
1.1.1 Διωνυμικό πείραμα	6
1.1.2 Διωνυμική κατανομή	6
1.1.3 Εφαρμογές	9
1.2 Υπεργεωμετρική κατανομή	10
1.2.1 Εφαρμογή	12
1.3 Γεωμετρική κατανομή	13
1.4 Αρνητική διωνυμική κατανομή ή κατανομή Pascal	14
1.4.1 Εφαρμογή	16
1.5 Κατανομή Poisson	16
1.5.1 Διαδικασία Poisson	18
1.5.2 Εφαρμογές	20
2. Ειδικές συνεχείς κατανομές	22
2.1 Ομοιόμορφη κατανομή	22
2.2 Εκθετική κατανομή	23
2.2.1 Εφαρμογές	24
2.2.2 Ιδιότητα της αμνησίας της εκθετικής κατανομής	26
2.3 Κατανομή Γ	26
2.3.1 Σχέση με Poisson	27
2.3.2 Εφαρμογές	28
2.4 Κατανομή Β	28
2.4.1 Εφαρμογές	29
2.5 Κανονική κατανομή ή κατανομή Gauss	29
2.5.1 Θεώρημα	31
2.5.2 Εφαρμογές	33
2.6 Άλλες συνεχείς κατανομές	34
2.6.1 Κατανομή Cauchy	34
2.6.2 Κατανομή Laplace ή διπλή εκθετική κατανομή	34
2.6.3 Λογαριθμοκανονική (long-normal) κατανομή	35
2.6.4 Κατανομή Weibull	35
2.6.5 Κατανομή Maxwell	36
2.6.6 Κατανομή ακροτάτων τιμών ή του Gumbel	36
2.6.7 Κατανομή Pareto	36
2.7 Άλλες κατανομές	37
2.7.1 Αρνητική υπεργεωμετρική κατανομή	37

2.7.2 Κατανομή Polya	37
2.7.3 Κατανομή λογαριθμικής σειράς ή λογαριθμική κατανομή	38
2.7.4 Μίξη κατανομών (mixture distributions) ή σύνθετες κατανομές (compound distributions)	38
Ελληνόγλωσση βιβλιογραφία	39

Εισαγωγή

Θα παρουσιάσουμε διάφορες ειδικές κατανομές και εφαρμογές τους. Αυτές οι κατανομές, αποτελούν πρότυπα (μοντέλα) πιθανοτήτων. Με τη βοήθειά τους, μπορούμε να υπολογίζουμε τις πιθανότητες διάφορων ενδεχομένων, χωρίς να χρειάζεται να ανατρέχουμε κάθε φορά στις βασικές αρχές της θεωρίας πιθανοτήτων.

Πρώτα παρουσιάζουμε τις ειδικές διακριτές κατανομές και στη συνέχεια τις ειδικές συνεχείς κατανομές. Οι ειδικές κατανομές πιθανότητας, είναι θεμελιώδεις για την κατανόηση και την πρόβλεψη τυχαίων φαινομένων σε πολλές εφαρμογές και επιστήμες. Ακολουθούν μερικές από τις κύριες εφαρμογές τους:

1. Στατιστική και ανάλυση δεδομένων

- **Περιγραφή δεδομένων.** Χρησιμοποιούνται για να περιγράψουν και να αναλύσουν τα δεδομένα που συλλέγονται από πειράματα, έρευνες ή παρατηρήσεις. Παράδειγμα: η κανονική κατανομή (ή κανονική κατανομή του Gauss) χρησιμοποιείται για να περιγράψει πολλές φυσικές και κοινωνικές διεργασίες.

- **Αναγνώριση προτύπων.** Εφαρμόζεται στην αναγνώριση προτύπων (pattern recognition) και στη μηχανική μάθηση, για να βοηθήσει στον προσδιορισμό αν τα δεδομένα ταιριάζουν σε μια συγκεκριμένη κατανομή.

2. Μηχανική και φυσική

- **Θερμοδυναμική και στατιστική φυσική.** Στη στατιστική φυσική, οι κατανομές όπως η κατανομή Maxwell–Boltzmann περιγράφουν τη διανομή της ταχύτητας των μορίων σε ένα αέριο.

- **Μηχανική σφαλμάτων.** Στην τεχνική και στην κατασκευή, οι κατανομές πιθανότητας χρησιμοποιούνται για τον υπολογισμό της πιθανότητας σφαλμάτων σε συστήματα και την πρόβλεψη της αξιοπιστίας των μηχανών και των συσκευών.

- **Ελαστικότητα και ροή ρευστών.** Ορισμένες κατανομές, χρησιμοποιούνται για να περιγράψουν τη συμπεριφορά υλικών ή ρευστών σε διάφορες συνθήκες (π.χ., κατανομή Weibull).

3. Οικονομία και χρηματοοικονομικά

- **Ανάλυση κινδύνου:** Η κανονική κατανομή, χρησιμοποιείται για να μοντελοποιήσει την απόδοση των μετοχών ή των αγορών γενικότερα. Οι κατανομές όπως η κατανομή lognormal, χρησιμοποιούνται για να μοντελοποιήσουν τις τιμές των περιουσιακών στοιχείων.

- **Θεωρία παιγνίων.** Κατανομές πιθανοτήτων χρησιμοποιούνται για να μοντελοποιήσουν στρατηγικές, σε παιχνίδια και αποφάσεις με αβεβαιότητα.

4. Βιοστατιστική και ιατρική

- **Εκτίμηση κινδύνου και υγείας:** Στη βιοστατιστική, οι κατανομές όπως η κατανομή Poisson ή η εκθετική κατανομή, χρησιμοποιούνται για να περιγράψουν τον χρόνο αναμονής μέχρι να συμβεί κάποιο γεγονός, όπως η εμφάνιση μιας ασθένειας ή η διάρκεια ζωής ενός ασθενούς.

- **Δοκιμές και έρευνες:** Στη μελέτη της αποτελεσματικότητας θεραπειών, οι κατανομές χρησιμοποιούνται για την αξιολόγηση της πιθανότητας ότι ένα φάρμακο ή μια θεραπεία έχει σημαντικά αποτελέσματα.

5. Ασφάλεια και βιομηχανία

- **Αξιολόγηση κινδύνου:** Στην ασφάλεια και στη βιομηχανία, κατανομές όπως η Weibull ή η Γκαουσιανή, χρησιμοποιούνται για την πρόβλεψη της αποτυχίας των μηχανημάτων και τη διαχείριση των κινδύνων σε βιομηχανικές διαδικασίες.

- **Μοντελοποίηση αποτυχίας συστημάτων:** Οι κατανομές χρησιμοποιούνται για να μοντελοποιήσουν τη διάρκεια ζωής και την αποτυχία των προϊόντων και των εξαρτημάτων.

6. Μοντελοποίηση και προσομοιώσεις

- **Μοντελοποίηση τυχαίων φαινομένων.** Η προσομοίωση τυχαίων φαινομένων μέσω υπολογιστικών μεθόδων (π.χ., Monte Carlo simulations), βασίζεται σε κατανομές πιθανότητας, για να μιμηθεί τη συμπεριφορά του συστήματος.

- **Σενάρια τύχης και αβεβαιότητας:** Χρησιμοποιούνται για να μοντελοποιήσουν την αβεβαιότητα σε οικονομικά ή φυσικά συστήματα και να κάνουν προβλέψεις ή αποφάσεις κάτω από αβέβαιες συνθήκες.

7. Μηχανική μάθηση και τεχνητή νοημοσύνη

- **Κατηγοριοποίηση και ανάλυση.** Σε αλγορίθμους μηχανικής μάθησης, οι κατανομές χρησιμοποιούνται για να καθορίσουν την κατανομή πιθανότητας σε δεδομένα, ώστε να προβλέψουν κατηγορίες ή να εκτιμήσουν παραμέτρους.

- **Μοντελοποίηση αστάθειας και αβεβαιότητας.** Στη μηχανική μάθηση, οι κατανομές πιθανότητας χρησιμοποιούνται για να κατανοήσουμε και να αναλύσουμε την αβεβαιότητα στις προβλέψεις.

8. Αστρονομία και κοσμολογία

- **Κατανομή αστεριών και γαλαξιών.** Στην αστρονομία, οι κατανομές πιθανότητας χρησιμοποιούνται για να μελετηθούν οι διανομές των αστεριών, των γαλαξιών ή άλλων αστρονομικών φαινομένων στο διάστημα.

- **Αναλύσεις παρατηρήσεων.** Χρησιμοποιούνται για να αναλυθούν δεδομένα από τηλεσκόπια ή από άλλες αστρονομικές παρατηρήσεις.

9. Κοινωνικές επιστήμες

- **Μοντελοποίηση συμπεριφοράς:** Στην κοινωνιολογία και στην ψυχολογία, οι κατανομές χρησιμοποιούνται για να κατανοήσουν και να προβλέψουν τη συμπεριφορά ατόμων ή ομάδων, κάτω από διάφορες συνθήκες.

- **Ανάλυση δημογραφικών δεδομένων:** Κατανομές χρησιμοποιούνται για να μοντελοποιήσουν πληθυσμούς και τις δυναμικές τους, όπως την αύξηση ή μείωση του πληθυσμού.

10. Μάρκετινγκ και επιχειρήσεις

- **Προβλέψεις πωλήσεων:** Οι κατανομές, χρησιμοποιούνται για να αναλύσουν και να προβλέψουν τις πωλήσεις προϊόντων ή τις προτιμήσεις των καταναλωτών.

- **Διαχείριση αποθεμάτων:** Χρησιμοποιούνται για την πρόβλεψη της ζήτησης και της διαχείρισης αποθεμάτων με βάση τυχαία φαινόμενα, όπως οι καταναλωτικές τάσεις.

1. Ειδικές διακριτές κατανομές

1.1 Διωνυμικό πείραμα – δοκιμές Bernoulli

Πολλά από τα τυχαία πειράματα, έχουν το χαρακτηριστικό ότι αποτελούνται από μια πεπερασμένη ή άπειρη σειρά δοκιμών (προσπαθειών) ή επαναλήψεων ενός μερικού πειράματος, του οποίου τα αποτελέσματα είναι μόνον δυο: ένα ενδεχόμενο ή το αντίθετο αυτού. Για ευκολία, ας ονομάσω τα αποτελέσματα αυτά επιτυχία και αποτυχία. Θα συμβολίζω την επιτυχία με το E και την αποτυχία με το A . Στη σειρά αυτή των δοκιμών, ενδιαφερόμε να μελετήσω τον αριθμό των επιτυχιών και ιδιαίτερα να βρω την κατανομή πιθανότητας του αριθμού αυτού.

Ας αναφέρω μερικά παραδείγματα. Σε μια σφυγμομέτρηση της κοινής γνώμης, ενδιαφερόμαι να μελετήσω την κατανομή του αριθμού των ατόμων που υποστηρίζουν τη θέση ενός πολιτικού κόμματος, μεταξύ εκείνων που λαμβάνουν μέρος στη σφυγμομέτρηση. Οι δοκιμές του πειράματος, είναι η εξέταση των ατόμων που λαμβάνουν μέρος στη σφυγμομέτρηση. Επιτυχία είναι, η υποστήριξη από ένα άτομο της θέσης του πολιτικού κόμματος και αποτυχία το αντίθετο. Εδώ, ο πληθυσμός θεωρείται άπειρος. Σε μια μελέτη για τον ηλεκτροφωτισμό μιας αγροτικής περιοχής, ενδιαφερόμαστε να βρούμε το ποσοστό των σπιτιών που έχουν ηλεκτρικό ρεύμα.

Τριακόσια σπίτια εκλέγονται στην τύχη με επανατοποθέτηση και το κάθε ένα εξετάζεται αν έχει ηλεκτρικό ρεύμα. Τα τριακόσια σπίτια, αποτελούν τον αριθμό δοκιμών του πειράματος. Επιτυχία είναι το να έχει ένα σπίτι ηλεκτρικό ρεύμα και αποτυχία το να μην έχει. Ποια είναι η πιθανότητα ο αριθμός των σπιτιών που διαλέγουμε στο πείραμα και που έχουν ρεύμα, να είναι μεγαλύτερος του οκτώ; Μια απάντηση στο ερώτημα αυτό θα δοθεί πιο κάτω.

Χάριν ευκολίας, θα συμβολίζω με n τον αριθμό των δοκιμών του πειράματος. Έστω X ο αριθμός των επιτυχιών, σε n δοκιμές του πειράματος. Σκοπός είναι να βρω την κατανομή πιθανότητας της διακριτής τυχαίας μεταβλητής X . Είναι φανερό ότι για να πετύχω το σκοπό αυτό, πρέπει να κάνω ορισμένες υποθέσεις – εξιδανικεύσεις. Έτσι, δεχόμαστε ότι ισχύουν τα εξής:

- Η πιθανότητα επιτυχίας ή αποτυχίας παραμένει σταθερή σε όλες τις δοκιμές του πειράματος
- Οι δοκιμές του πειράματος είναι στατιστικώς ανεξάρτητες

Θα συμβολίζω την πιθανότητα επιτυχίας με το p , $0 \leq p \leq 1$, δηλαδή $P(E) = p$. Η πιθανότητα αποτυχίας θα συμβολίζεται με το q , δηλαδή $P(A) = q$. Έτσι $p + q = 1$.

1.1.1 Διωνυμικό πείραμα

Ένα πείραμα αποτελούμενο από δοκιμές, η κάθε μία από τις οποίες έχει μόνο δυο δυνατά αποτελέσματα E ή A και για το οποίο ισχύουν οι δύο παραπάνω υποθέσεις, λέγεται διωνυμικό πείραμα και κάθε επιμέρους γραμμή λέγεται δοκιμή του Bernoulli. Ο δειγματικός χώρος των n δοκιμών του Bernoulli. Ο δειγματικός χώρος των n δοκιμών του Bernoulli περιλαμβάνει 2^n σημεία ή σειρές από n σύμβολα E και A . Κάθε τέτοιο σημείο αντιπροσωπεύει ένα δυνατό αποτέλεσμα του διωνυμικού πειράματος.

1.1.2 Διωνυμική κατανομή

Έστω ένα διωνυμικό πείραμα στο οποίο εκτελώ n δοκιμές του Bernoulli, όπου n δεδομένος αριθμός. Έστω X η τυχαία μεταβλητή της οποίας οι τιμές x παριστάνουν τον αριθμό των επιτυχιών που θα σημειωθούν στην ακολουθία των n δοκιμών. Είναι φανερό ότι η X είναι μια διακριτή τυχαία μεταβλητή με τιμές $0, 1, 2, \dots, n$. Ας βρω την κατανομή πιθανότητας της X , δηλαδή τη συνάρτηση $P_X(x) = P(X = x)$. Έστω ότι το

x είναι σταθερό. Τα αποτελέσματα του διωνυμικού πειράματος είναι της μορφής ΕΕΑΕΑ...Ε ή ΑΑΕΑ...ΑΕ, κλπ (1)

όπου ο συνολικός αριθμός των Ε και Α είναι n . Όταν $X = x$, τα αποτελέσματα περιέχουν x φορές το Ε και $(n - x)$ φορές το Α. Τότε, λόγω των δύο συνθηκών του διωνυμικού πειράματος, η πιθανότητα κάθε αποτελέσματος με x φορές το Ε και $(n - x)$ φορές το Α, είναι $p^x (1 - p)^{n-x}$. Απομένει να υπολογίσω τον αριθμό των αποτελεσμάτων της μορφής (1) όταν $X = x$. Ο αριθμός αυτός είναι ο αριθμός των συνδυασμών των n θέσεων ανά x στις οποίες x θέσεις εμφανίζεται το Ε, δεδομένου ότι η σειρά με την οποία γίνονται οι επιτυχίες Ε και οι αποτυχίες Α δεν παίζει κανένα ρόλο.

Ο αριθμός αυτός, ισούται ακόμη με τον αριθμό των μεταθέσεων n πραγμάτων, από τα οποία x είναι ίδια, (Ε) και $(n - x)$ φορές είναι επίσης ίδια, αλλά άλλου τύπου, (Α), δηλαδή $n! / [x!(n - x)!]$. Συνεπώς, έχω

$$p_x(x) = P(X = x) = P\{EEAE...EA \text{ ή } AAEE...AE \text{ ή } \dots\}$$

$$= p^x (1 - p)^{n-x} + p^x (1 - p)^{n-x} + \dots = \binom{n}{x} p^x q^{n-x}$$

$$\text{Έτσι, } p_x(x) = \binom{n}{x} p^x p^{n-x}, \quad x = 0, 1, 2, \dots, n \quad (2)$$

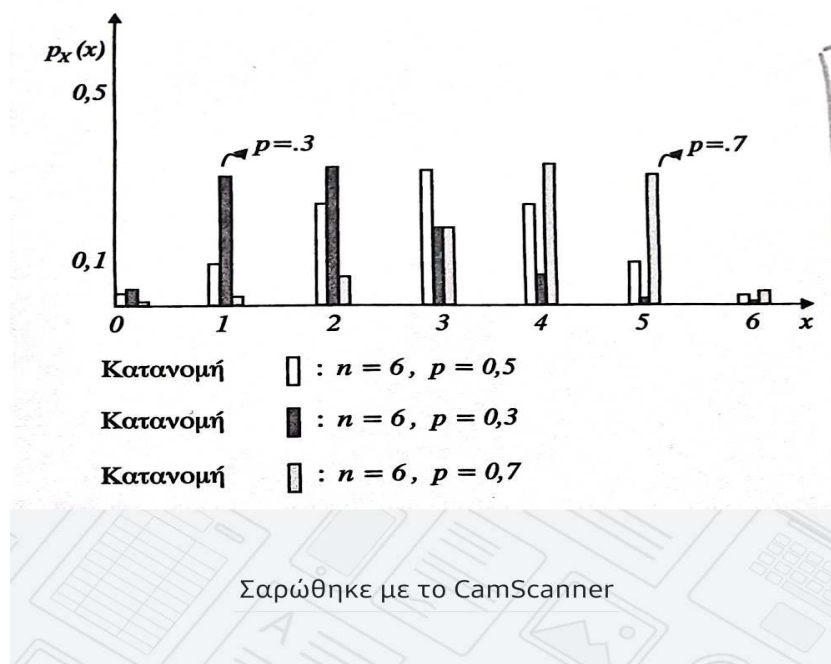
Η παραπάνω σχέση, δίνει τη λεγόμενη διωνυμική κατανομή. Από το θεώρημα του

$$\text{Νεύτωνα έχω, } \sum_{x=0}^n p_x(x) = \sum_{x=0}^n \binom{n}{x} p^x p^{n-x} = (p + q)^n = 1$$

Άρα, η $p_x(x)$ δίνει μια αληθή συνάρτηση πιθανότητας. Μια τυχαία μεταβλητή X η οποία έχει την κατανομή (2) λέγεται διωνυμική τυχαία μεταβλητή. Για συντομία αυτό συμβολίζεται με $X \sim B(n, p)$. Τα n, p λέγονται **παράμετροι** της διωνυμικής κατανομής. Αν $n=1$ τότε η X λέγεται και τυχαία μεταβλητή του Bernoulli και η κατανομή αυτής κατανομή Bernoulli. Η αθροιστική συνάρτηση κατανομής της X ,

$$\text{δίδεται από την σχέση } Fx(x) = \begin{cases} 0 & , x < 0 \\ \sum_{y=0}^{[x]} \binom{n}{y} p^y q^{n-y} & , 0 \leq x \leq n \\ 1 & , x > n \end{cases}, \text{ όπου } [x] \text{ παριστάνει το}$$

ακέραιο μέρος του θετικού αριθμού x . Το σχήμα 1 δίνει γραφικές παραστάσεις της (2) για $(n, p) = (6, 0.5), (6, 0.3), (6, 0.7)$. Παρατηρώ ότι για $p=0.5$ η γραφική παράσταση είναι συμμετρική. Για $p=0.3$, η γραφική παράσταση ανεβαίνει απότομα και μετά κατεβαίνει πιο ομαλά, ενώ αντίθετη είναι η συμπεριφορά της για $p=0.7$. Η παρατήρηση αυτή είναι γενικότερη και ισχύει για κάθε $p < 0.5$ και $p > 0.5$ αντίστοιχα.



Σχήμα 1: Διωνυμικές κατανομές

Ο υπολογισμός των διωνυμικών πιθανοτήτων βάσει της (2), είναι σχετικά απλός για μικρές τιμές του n ($n \leq 20$). Για μεγαλύτερες τιμές του n , η υπολογιστική προσπάθεια δεν είναι ασήμαντη και γίνεται υπερβολική για μεγάλες τιμές του n . Ο αναγωγικός τύπος $p_x(x) = \frac{(n-x+1)p}{xq} p_x(x-1)$, $x = 1, 2, \dots, n$ είναι χρήσιμος αν και συνήθως για μεγάλες τιμές του n προτιμάτε ο κατά προσέγγιση υπολογισμός του $p_x(x)$ με την κατανομή Poisson ή την κανονική κατανομή.

Πίνακες που δίνουν τις διωνυμικές πιθανότητες για διάφορες τιμές του n και p χρησιμοποιούνται ευρύτατα στη στατιστική. Επίσης υπάρχουν πίνακες που δίνουν την αθροιστική διωνυμική συνάρτηση κατανομής. Διωνυμικές πιθανότητες, απλές και αθροιστικές, μπορούν να υπολογιστούν εύκολα και με ειδικές συναρτήσεις προγραμμάτων ηλεκτρονικού υπολογιστή, όπως, π.χ., το MS-Excel. Η επόμενη σχέση είναι χρήσιμη και συνδέει μια διωνυμική τυχαία μεταβλητή με επιμέρους τυχαία μεταβλητή Bernoulli. Έστω X ο αριθμός των επιτυχιών ενός διωνυμικού πειράματος με n δοκιμές και $P(E) = p$. Τότε $X \sim B(n, p)$. Έστω επίσης οι τυχαίες μεταβλητές

$$X_i = \begin{cases} 1, & \text{αν το αποτέλεσμα της } i \text{ δοκιμής είναι } E \\ 0, & \text{αν το αποτέλεσμα της } i \text{ δοκιμής είναι } A \end{cases}$$

$$\text{Τότε } X = X_1 + X_2 + \dots + X_n \quad (3)$$

Δηλαδή, το άθροισμα n ανεξάρτητων τυχαίων μεταβλητών Bernoulli με παράμετρο p , είναι μια διωνυμική κατανομή με παραμέτρους n και p . Επίσης έχω $P\{x \text{ επιτυχίες } E \text{ σε } n \text{ δοκιμές}\} = P\{(n-x) \text{ αποτυχίες } A \text{ σε } n \text{ δοκιμές}\}$

$$\text{ή} \quad \binom{n}{x} p^x q^{n-x} = \binom{n}{n-x} q^{n-x} p^x. \text{ Η σχέση αυτή, είναι χρήσιμη για τον υπολογισμό}$$

διωνυμικών πιθανοτήτων, από διωνυμικούς πίνακες, στους οποίους το p περιορίζεται από 0 μέχρι 1/2.

1.1.3 Εφαρμογές

1. Το πιο γνωστό και απλό παράδειγμα δοκιμών του Bernoulli δίνεται από τις διαδοχικές ρίψεις ενός αληθινού ή συμμετρικού νομίσματος. Εδώ $p = q = 1/2$. Αν το νόμισμα δεν είναι συμμετρικό, πάλι υποθέτω ότι οι διαδοχικές ρίψεις είναι ανεξάρτητες, ώστε να έχω το πρότυπο των δοκιμών Bernoulli στις οποίες η πιθανότητα p επιτυχίας μπορεί να πάρει μία αυθαίρετη τιμή. Ομοίως, η διαδοχική ρίψη ενός αληθινού ζαριού και η διάκριση μεταξύ των ενδεχομένων άσος δηλαδή (Ε) και μη άσος (Α) οδηγεί σε δοκιμές Bernoulli με $p = 1/6$, ενώ η διάκριση μεταξύ περιττού ή άρτιου αποτελέσματος οδηγεί σε διωνυμικό πείραμα με $p = 1/2$. Άλλη εφαρμογή προσφέρεται στο **βιομηχανικό έλεγχο ποιότητας**, όπου αντικείμενα-είδη (π.χ. βίδες, λαμπτήρες, κουζίνες κ.λπ.) παράγονται συνεχώς και κατά τον έλεγχο ποιότητας ταξινομούνται ως Ε, στην περίπτωση που εκπληρώνουν τις προδιαγραφές και ως Α, στην περίπτωση που δεν εκπληρώνουν τις προδιαγραφές. Δοκιμές Bernoulli έχω και στις δειγματοληψίες με επανατοποθέτηση, από πεπερασμένους πληθυσμούς.

Το σχήμα των δοκιμών Bernoulli, είναι ένα θεωρητικό πρότυπο (μοντέλο) που μόνο η εμπειρία μπορεί να δείξει αν είναι κατάλληλο για την περιγραφή ενός συγκεκριμένου πειράματος. Έτσι, η γνώση μας ότι οι διαδοχικές ρίψεις νομίσματος ακολουθούν το πρότυπο των δοκιμών Bernoulli πηγάζει από εμπειρική μαρτυρία. Ο απλός άνθρωπος πιστεύει ότι αν φέρει στη σειρά 17 κορώνες, τα γράμματα γίνονται κατόπιν πιο πιθανά. Αυτό δε σημαίνει κατ' ανάγκη ότι το νόμισμα είναι ελαττωματικό. Είναι πολύ πιθανό να μην υπάρχει ανεξαρτησία των δοκιμών. Στις **δειγματοληψίες**, το βιομηχανικό ποιοτικό έλεγχο κ.λπ., το σχήμα των δοκιμών Bernoulli δίνει ένα ιδανικό πρότυπο, άσχετα και αν ποτέ δεν επιτυγχάνεται 100%. Πολλές φορές με την πάροδο του χρόνου, η $P(E) = p$ δεν παραμένει σταθερή διότι π.χ. επέρχονται αλλοιώσεις στο προϊόν της βιομηχανικής γραμμής. Από πλευράς ποιοτικού ελέγχου, είναι επιθυμητό η διαδικασία παραγωγής να ακολουθεί το σχήμα Bernoulli. Ο συνεχής έλεγχος, αποσκοπεί στο να αποκαλύψει βασικές αλλαγές από το ιδανικό σχήμα και να χρησιμοποιήσει αυτές ως ένδειξη επερχόμενης ανωμαλίας.

2. Έστω ότι κατά μέσο όρο 30% των σπιτιών μιας περιοχής διαθέτουν ηλεκτρικό ρεύμα. Ποια είναι η πιθανότητα σε μια δειγματοληψία με επανατοποθέτηση 20 σπιτιών,

(α) Το πολύ 3 από αυτά να διαθέτουν ηλεκτρικό ρεύμα,

(β) Τουλάχιστον 8 από αυτά να διαθέτουν ηλεκτρικό ρεύμα,

(γ) Περισσότερα από 3 και λιγότερα από 8 να διαθέτουν ηλεκτρικό ρεύμα;

Εδώ έχω διωνυμικό πείραμα με $n = 20$ και $p = 0.3$. Συνεπώς,

(α)

$$\begin{aligned} P(X \leq 3) &= \sum_{x=0}^3 \binom{20}{x} (0.3)^x (0.7)^{20-x} \\ &= 0.0008 + 0.0068 + 0.0278 + 0.0716 = 0.1070 \end{aligned}$$

(β)

$$\begin{aligned} P(X \geq 8) &= 1 - P(X \leq 7) \\ &= 1 - [P(X \leq 3) + P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7)] \\ &= 1 - [0.1070 + 0.1304 + 0.1789 + 0.1916 + 0.1643] = 0.2278 \end{aligned}$$

(γ)

$$P(3 \leq X < 8) = P(X=3) + P(X=4) + P(X=5) + P(X=6) + P(X=7) \\ = 0.0716 + 0.1304 + 0.1789 + 0.1916 + 0.1643 = 0.7368$$

3. Έστω ότι 10 εργάτες ενός εργοστασίου χρησιμοποιούν παροχές ηλεκτρικού ρεύματος σε διάφορες φάσεις της δουλειάς τους. Το πρόβλημα που απασχολεί το εργοστάσιο είναι πόσες παροχές χρειάζονται για να εξυπηρετηθούν οι ανάγκες των εργατών καλύτερα, δηλαδή κατά τρόπο ώστε να μην περιμένει πολύ καιρό ένας εργάτης για να χρησιμοποιήσει την παροχή. Εξυπακούεται, ότι λόγω υψηλού κόστους, το εργοστάσιο θέλει να αποφύγει την εγκατάσταση 10 παροχών. Σαν μια χονδρική προσέγγιση, μπορώ να υποθέσω ότι για κάθε χρονική στιγμή κάθε εργάτης έχει την ίδια πιθανότητα p να χρειασθεί μια παροχή. Αν οι εργάτες εργάζονται ο ένας ανεξάρτητα από τον άλλο, τότε ο αριθμός X των εργατών μπορεί να χρειαστεί παροχές την ίδια στιγμή, έχει την κατανομή $B(10, p)$. Το εργοστάσιο γνωρίζει ακόμη ότι κατά μέσο όρο ο κάθε εργάτης χρησιμοποιεί την παροχή του 12 λεπτά την ώρα. Αυτό δίνει την τιμή $p = 12/60 = 1/5$ για την παράμετρο p . Ας δώ την πιθανότητα ώστε 7 ή περισσότεροι εργάτες να χρειαστούν παροχή την ίδια στιγμή:

$$P(X \geq 7) = \sum_{x=7}^{10} \binom{10}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{10-x} = 0.0008 + 0.0001 + 0.0000 + 0.0000 = 0.0009$$

Αυτό σημαίνει ότι αν το εργοστάσιο εγκαταστήσει μόνο 6 παροχές, η πιθανότητα για υπερζήτηση είναι $0,0009 = \frac{1}{10.000} = \frac{1}{1.111,11}$, δηλαδή για 1 λεπτό περίπου στα 1.111

λεπτά ή για ένα λεπτό στις 18,5 ώρες περίπου.

1.2 Υπεργεωμετρική κατανομή

Η κατανομή αυτή, εμφανίζεται όταν κάνω δειγματοληψία χωρίς επανατοποθέτηση από ένα πεπερασμένο πληθυσμό. Έστω ότι ο πληθυσμός αποτελείται από N στοιχεία κάθε ένα από τα οποία είναι δυνατό να ανήκει σε μια από δυο κατηγορίες C και C' . Έστω Np ο αριθμός των στοιχείων που ανήκουν στην κατηγορία C και Nq ο αριθμός εκείνων που ανήκουν στη κατηγορία C' , όπου p και q , $0 \leq p, q \leq 1$, είναι τα ποσοστά των στοιχείων των κατηγοριών C , C' αντίστοιχα. Προφανώς $Np + Nq = N$ και $p + q = 1$. Έστω τώρα ότι n στοιχεία εκλέγονται στην τύχη από τον πληθυσμό, χωρίς επανατοποθέτηση. Έστω X η τυχαία μεταβλητή που παριστάνει τον αριθμό των στοιχείων της κατηγορίας C , που περιλαμβάνονται στο δείγμα των n στοιχείων. Προφανώς, η X είναι μια διακριτή τυχαία μεταβλητή. Δεδομένου ότι δεν ενδιαφέρομαι για τη σειρά με την οποία εμφανίζονται τα στοιχεία, ο αριθμός των δυνατών δειγμάτων είναι $\binom{N}{n}$. Από αυτά, ο αριθμός των δειγμάτων που περιέχουν x

στοιχεία της κατηγορίας C και $(n-x)$ της κατηγορίας C' είναι, $\binom{Np}{x} \binom{Nq}{n-x}$ βάσει της αρχής της απαρίθμησης. Άρα, $P\{X=x\} = \frac{\binom{Np}{x} \binom{Nq}{n-x}}{\binom{N}{n}}$. (4)

Η (4) δίνει τη συνάρτηση πιθανότητας $p_x(x)$ της τυχαίας μεταβλητής X . Το ότι η συνάρτηση αποτελεί μια συνάρτηση πιθανότητας, αποδεικνύεται από τις

ιδιότητες $\sum_{r=0}^k \binom{m}{r} \binom{n}{k-r} = \binom{m+n}{k}$ και $\binom{n}{r} = 0$ για n, r θετικά,

ακέραια με $r > n$ των διωνυμικών συντελεστών. Η κατανομή (4) λέγεται υπεργεωμετρική και κατά τον ίδιο τρόπο η X λέγεται υπεργεωμετρική τυχαία μεταβλητή και συμβολίζεται με $X \sim Hg(N, n, p)$. Δοθέντων των N, n , το p , $0 \leq p \leq 1$ αποτελεί την παράμετρο της κατανομής. Οι τιμές που μπορεί να πάρει μια υπεργεωμετρική τυχαία μεταβλητή X με θετική πιθανότητα δεν είναι $0, 1, \dots, n$ αλλά εξαρτώνται από τη σχέση που υπάρχει μεταξύ των n, Np, Nq . Οι διωνυμικοί συντελεστές $\binom{n}{r}$ είναι μη μηδενικοί, όταν το r είναι ακέραιος και $0 \leq r \leq n$. Άρα, η

$p(x)$ είναι θετική όταν $x \leq Np$, $n-x \leq Nq$, $x \geq 0$, $n-x \geq 0$.

Από αυτές τις ανισότητες, βλέπω ότι η μεγίστη δυνατή τιμή του x είναι το $\min(n, Np)$ και η ελάχιστη δυνατή τιμή είναι το $\max(n - Np, 0)$. Άρα, το πεδίο των τιμών μιας υπεργεωμετρικής τυχαίας μεταβλητής είναι $\max(0, n - Np), \dots, \min(n, Np)$. Η αθροιστική συνάρτηση κατανομής της υπεργεωμετρικής τυχαίας μεταβλητής δίνεται από τη σχέση

$$Fx(x) = \begin{cases} 0 & , x < \max(n - Nq, 0) \\ \sum_{y=\max(n-Nq,0)}^{[x]} \binom{Np}{n} \binom{Nq}{n-y} / \binom{N}{n} & , \max(0, n - Nq) \leq x \leq \min(n, Np) \\ 1 & , x > \min(n, Np) \end{cases} \quad (5)$$

Για τη διευκόλυνση των υπολογισμών, υπάρχουν πίνακες οι οποίοι δίνουν τη συνάρτηση πιθανότητας ή την αθροιστική συνάρτηση κατανομής της υπεργεωμετρικής κατανομής για μικρές τιμές του $N (\leq 100)$. Συνήθως, αυτοί οι πίνακες χρησιμοποιούνται στη στατιστική, για την εύρεση ακριβών διαστημάτων εμπιστοσύνης της παραμέτρου p . Υπεργεωμετρικές πιθανότητες, μπορούν να υπολογιστούν εύκολα και με ειδικές συναρτήσεις προγραμμάτων ηλεκτρονικού υπολογιστή, όπως π.χ. το MS-Excel. Για τους υπολογισμούς, μπορώ να χρησιμοποιήσω τον επαναληπτικό τύπο $p_x(x+1) = \frac{(n-x)(Np-x)}{(x+1)(Nq-n+x+1)} p_x(x)$,

ξεκινώντας από την τιμή του $p_x(0)$. Το σχήμα 2 δίνει την γραφική παράσταση της (4) για διάφορες τιμές των παραμέτρων της.

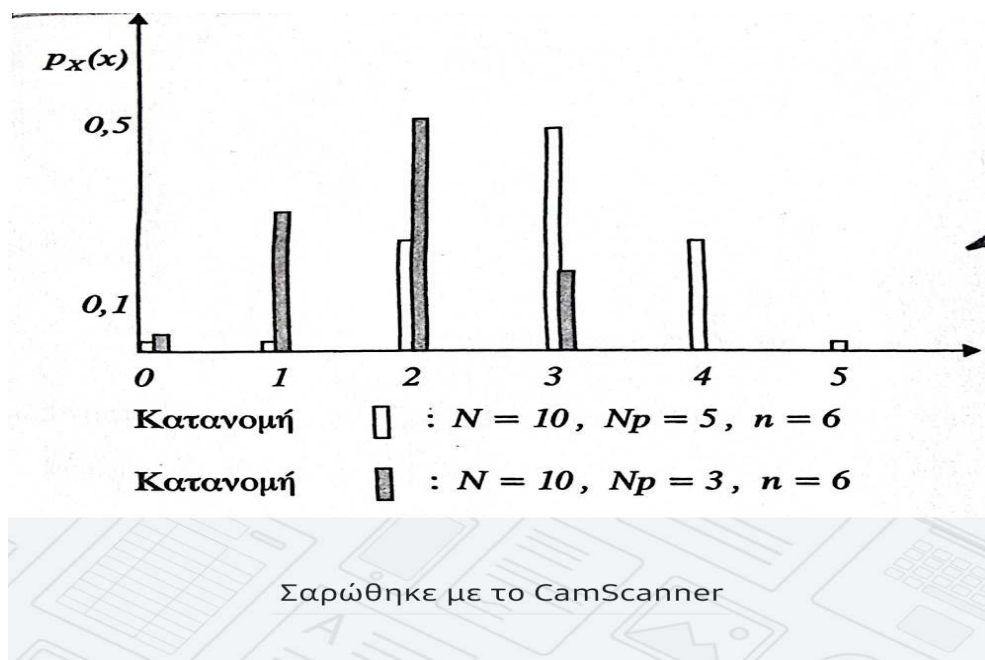
Ας θεωρήσω τώρα τη συνάρτηση πιθανότητας της υπεργεωμετρικής κατανομής ως μια ακολουθία συναρτήσεων του x και πάρουμε το όριο καθώς το $N \rightarrow \infty$ με τα n, p σταθερά. Τότε ισχύει ότι $\lim \binom{Np}{x} \binom{Nq}{n-x} / \binom{N}{x} = \binom{n}{x} p^x q^{n-x}$,

δηλαδή για μεγάλες τιμές του N , η υπεργεωμετρική κατανομή προσεγγίζει τη διωνυμική. Πράγματι, συχνά χρησιμοποιούμε τη διωνυμική κατανομή με το ίδιο p για να προσεγγίσουμε τις υπεργεωμετρικές πιθανότητες εφ' όσον το n δεν υπερβαίνει το 5% του N . Οι εφαρμογές της υπεργεωμετρικής κατανομής είναι πολύπλευρες.

Στο βιομηχανικό ποιοτικό έλεγχο, σύνολα από N αντικείμενα υπόκεινται σε έλεγχο δειγματοληψίας. Κάθε ελεγχόμενο αντικείμενο, χαρακτηρίζεται ως ελαττωματικό ή μη. Η δειγματοληψία, γίνεται συνήθως (και κατά ρεαλιστικό τρόπο)

χωρίς επανάθεση. Τα ελαττωματικά αντικείμενα, ανήκουν στην κατηγορία C και τα μη ελαττωματικά στη κατηγορία C'. Τότε, το υπεργεωμετρικό πρότυπο είναι εφαρμόσιμο.

Ο αριθμός Np των ελαττωματικών αντικείμενων, είναι άγνωστος. Ένα δείγμα μεγέθους n λαμβάνεται και ο αριθμός των ελαττωματικών αντικείμενων προσδιορίζεται. Το υπεργεωμετρικό πρότυπο, επιτρέπει να βγάλω συμπεράσματα για το πιθανό μέγεθος του Np , χρησιμοποιώντας μεθόδους της στατιστικής. Χρησιμοποιώντας το υπεργεωμετρικό πρότυπο πιθανότητας και μεθόδους της στατιστικής, μπορώ επίσης να εκτιμήσω το μέγεθος ενός πληθυσμού ζώων, εφαρμόζοντας διπλή δειγματοληψία χωρίς επανατοποθέτηση. Η κεντρική ιδέα είναι ως εξής: Έστω ότι θέλω να εκτιμήσω τον άγνωστο αριθμό N των ψαριών σε μία λίμνη. Για το σκοπό αυτό ψαρεύω Np ψάρια, τα σημαδεύω με μια κόκκινη ανεξίτηλη κουκκίδα και τα αφήνω πάλι ελεύθερα. Μετά από μια ή δυο ημέρες ξαναψαρεύω n ψάρια και μετρώ τον αριθμό των x ψαριών που έχουν την κόκκινη κουκκίδα. Το υπεργεωμετρικό πρότυπο είναι τώρα εφαρμόσιμο και χρησιμοποιώντας μεθόδους της στατιστικής μπορώ να εκτιμήσω το N .



Σχήμα 2: Υπεργεωμετρικές κατανομές

1.2.1 Εφαρμογή

Έστω ότι σε τμήμα 20 σπουδαστών, το 40% ναυτολογείται. Αν διαλέξω στη τύχη και χωρίς επανατοποθέτηση 6 σπουδαστές, ποια είναι η πιθανότητα

- 2 ακριβώς από αυτούς να ναυτολογούνται και
- τουλάχιστον 2 από αυτούς να ναυτολογούνται;

Επειδή ξέρω ότι το 40% των σπουδαστών του τμήματος ναυτολογείται και το 60% δε ναυτολογείται, οι 20 σπουδαστές αποτελούν δυο ομάδες με $Np = 20 \times 0.4 = 8$ άτομα η πρώτη και $Nq = 20 \times 0.6 = 12$ άτομα η δεύτερη. Έτσι, η κατανομή του αριθμού X των σπουδαστών που ναυτολογούνται στο δείγμα των 6 ατόμων είναι

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

Από αυτή τη σχέση έχω

(α)

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

(β)

$$P(\text{τουλάχιστον } 2 \text{ ναυτολογούνται}) = 1 - P(\text{το πολύ ένας ναυτολογείται})$$

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

1.3 Γεωμετρική κατανομή

Η κατανομή αυτή αναφέρεται στο διωνυμικό πείραμα ή τις δοκιμές Bernoulli, όπου η τυχαία μεταβλητή X τώρα αντιπροσωπεύει τον αριθμό των δοκιμών που εκτελούνται μέχρι να εμφανιστεί η πρώτη επιτυχία E . Δηλαδή, η γεωμετρική κατανομή είναι η κατανομή του αριθμού των «δοκιμών αναμονής» ή «χρόνου αναμονής» μέχρι να εμφανιστεί η $1^{\text{η}}$ επιτυχία σε ένα διωνυμικό πείραμα. Η συνάρτηση πιθανότητας της γεωμετρικής κατανομής δίνεται από τον τύπο $p_x(x) = pq^{x-1}$, $x=1,2,3,\dots$ (6)

Το p , $0 \leq p \leq 1$, είναι η πιθανότητα επιτυχίας των δοκιμών Bernoulli και αποτελεί την παράμετρο της κατανομής. Μια τυχαία μεταβολή X , με κατανομή την (6) όπως πριν, λέγεται γεωμετρική ή τυχαία μεταβλητή του Pascal και συμβολίζεται με $X \sim \text{Geo}(p)$. Το όνομα απορρέει από το γεγονός ότι για τις διαδοχικές τιμές του x , οι πιθανότητες (6) αποτελούν μια φθίνουσα γεωμετρική πρόοδο. Ισχύει ότι

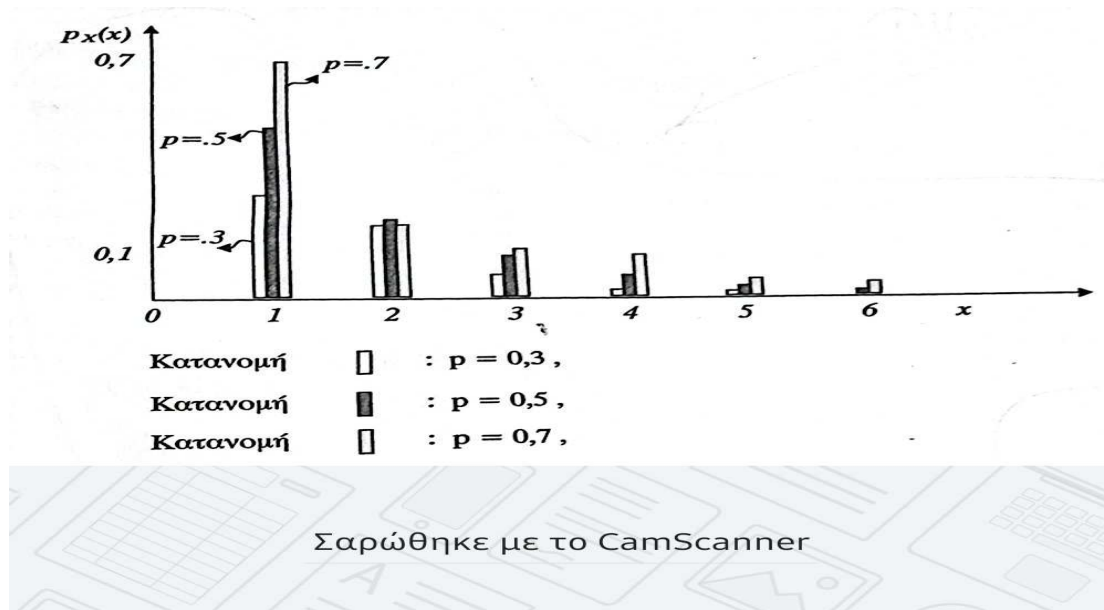
$$\sum_{x=1}^{\infty} pq^{x-1} = \frac{p}{1-q} = 1, \text{ γεγονός που μαζί με το μη αρνητικό των τιμών της κατανομής (6)}$$

αποδεικνύει ότι η ανωτέρω $p_x(x)$ αποτελεί μια συνάρτηση πιθανότητας. Η αθροιστική συνάρτηση κατανομής της γεωμετρικής τυχαίας μεταβλητής, είναι

$$F(x) = \begin{cases} 0, & x < 1 \\ \sum_{y=1}^{[x]} pq^{y-1}, & x \geq 1 \end{cases} \quad (7). \text{ Το σχήμα 3 δίνει τις τιμές της κατανομής (6) για}$$

$p=0.3$, $p=0.5$, $p=0.7$. Οι εφαρμογές της γεωμετρικής κατανομής είναι ευρύτατες.

Σε κάθε φαινόμενο που μπορεί να παρασταθεί ή να προσεγγιστεί με μια σειρά δοκιμών Bernoulli, η γεωμετρική κατανομή αποτελεί το μοντέλο πιθανοτήτων που μπορεί να υιοθετηθεί για το χρόνο αναμονής. Αναφέρω μερικές τυπικές περιπτώσεις: Σε ένα παιχνίδι κορώνα-γράμματα όπου ενδιαφέρομαι για το χρόνο αναμονής μέχρι να εμφανιστεί η $1^{\text{η}}$ κορώνα. Ένας γιατρός ενδιαφέρεται για την πιθανότητα ο $5^{\text{ος}}$ άρρωστος που θα τον επισκεφτεί στο γραφείο του, μια ημέρα, να είναι ο $1^{\text{ος}}$ που πάσχει από κάποιο ιό γρίπης. Ο βιομήχανος ενδιαφέρεται για την πιθανότητα το $8^{\text{ο}}$ κατασκευαζόμενο αντικείμενο να είναι το $1^{\text{ο}}$ ελαττωματικό κ.ο.κ.



Σχήμα 3: Γεωμετρικές κατανομές

1.4 Αρνητική διωνυμική κατανομή ή κατανομή Pascal

Αυτή η κατανομή, αποτελεί γενίκευση της γεωμετρικής κατανομής. Θεωρώ πάλι μια σειρά δοκιμών Bernoulli και αντί να ενδιαφέρομαι για τον αριθμό των δοκιμών αναμονής μέχρι την 1^η επιτυχία, ενδιαφέρομαι για τον αριθμό X των δοκιμών αναμονής (χρόνος αναμονής) μέχρι να εμφανιστεί η k επιτυχία με $(k \geq 1)$ και ακέραιο αριθμό. Οι δυνατές τιμές x της τυχαίας μεταβλητής X είναι $k, k+1, k+2, \dots$

Έστω x σταθερό ($x \geq k$). Τότε, το ενδεχόμενο $\{X = x\}$ σημαίνει ότι $(k-1)$ επιτυχίες έγιναν στις $(x-1)$ πρώτες δοκιμές και ότι η τελευταία δοκιμή, δηλαδή η δοκιμή x τάξης, κατέληξε σε επιτυχία. Οι $(k-1)$ επιτυχίες έγιναν στις $(x-1)$ πρώτες δοκιμές και ακολουθούν τη διωνυμική κατανομή. Άρα, βάσει της ανεξαρτησίας ισχύει ότι

$$P(X = x) = P\{(k-1) \text{ E στις πρώτες } (x-1) \text{ δοκιμές}\} P\{E \text{ στη } x \text{ δοκιμή}\}$$

$$= \binom{x-1}{k-1} p^{k-1} q^{x-1-(k-1)} p$$

Έτσι, η συνάρτηση πιθανότητας της αρνητικής διωνυμική κατανομής, δίνεται από τον τύπο $p_x(x) = \binom{x-1}{k-1} p^k q^{x-k}$, $x = k, k+1, \dots$ (8)

Τα k και p , $0 \leq p \leq 1$ είναι και πάλι οι παράμετροι της κατανομής. Μια τυχαία μεταβλητή X με κατανομή την (8) λέγεται αρνητική διωνυμική ή τυχαία μεταβλητή του Pascal και συμβολίζεται με $X \sim NB(k, p)$. Θα αποδείξω ότι η συνάρτηση $p_x(x)$ της (8) είναι πράγματι μια συνάρτηση πιθανότητας. Παράλληλα θα δικαιολογήσω και το όνομα αρνητική διωνυμική. Στηρίζομαι στην επέκταση του ορισμού και του διωνυμικού συντελεστή $\binom{x}{r}$ για κάθε πραγματικό αριθμό x και κάθε ακέραιο r και στο διωνυμικό τύπο του Νεύτωνα. Έχω

$$\sum_{x=k}^{\infty} p_x(x) = \sum_{x=k}^{\infty} \binom{x-1}{k-1} p^k q^{x-k} = p^k \sum_{x=k}^{\infty} \binom{x-1}{k-1} q^{x-k} =$$

$$= p^k \left\{ 1 + \binom{k}{1} q + \binom{k+1}{2} q^2 + \binom{k+2}{3} q^3 + \dots \right\} = p^k (1-p)^{-k} = p^k p^{-k} = 1 \quad (9)$$

$$\text{Επίσης ισχύει } \binom{-k}{y} (-1)^y = \binom{y+k-1}{y} \quad (10)$$

Άρα, ο διωνυμικός τύπος του Νεύτωνα γίνεται

$$(1-q)^{-k} = \sum_{y=0}^{\infty} \binom{-k}{y} (-q)^y = \sum_{y=0}^{\infty} \binom{y+k-1}{y} q^y = 1 + \binom{k}{1} q + \binom{k+1}{2} q^2 + \binom{k+2}{3} q^3 + \dots$$

Επειδή $p_x(x) \geq 0$ η (9) αποδεικνύει ότι η συνάρτηση $p_x(x)$ είναι πράγματι μια συνάρτηση πιθανότητας. Το όνομα αρνητική διωνυμική κατανομή απορρέει από το γεγονός ότι βάσει της (9) οι τιμές της $p_x(x)$ για $x = k, k+1, k+2, \dots$ είναι διαδοχικοί όροι του αρνητικού διωνυμικού αναπτύγματος του $p^k (1-p)^{-k}$. Επίσης και από το γεγονός ότι η $p_x(x)$ μπορεί να γραφεί ισοδύναμα

$$p_x(x) = \binom{-k}{x-k} p^k (-q)^{x-k}, \quad x = k, k+1, \dots \quad (11)$$

Η αθροιστική συνάρτηση κατανομής της αρνητικής διωνυμικής τυχαίας μεταβλητής,

$$\text{είναι } Fx(x) = \begin{cases} 0, & x < k \\ \sum_{n=k}^x \binom{-k}{n-k} p^k q^{n-k}, & x \geq k \end{cases} \quad (12)$$

Το σχήμα 4 δίνει γραφικές παραστάσεις της (8) για $p = 0,9$ και $k = 2, 4, 6$.

Είναι εύκολο ναδειχτεί ότι η αρνητική διωνυμική κατανομή είναι ένα άθροισμα k ανεξαρτήτων γεωμετρικών κατανομών. Ας θεωρήσουμε τις δοκιμές Bernoulli. Έστω Y_1 ο αριθμός των δοκιμών μέχρι να εμφανιστεί η 1^η επιτυχία, Y_2 ο αριθμός των δοκιμών από την 1^η επιτυχία μέχρι να εμφανιστεί η 2^η επιτυχία κ.ο.κ. και τελικά έστω Y_k ο αριθμός των δοκιμών από την $(k-1)$ επιτυχία μέχρι την k επιτυχία. Τότε αν $X \sim NB(k, p)$ εξ' ορισμού είναι $X = Y_1 + Y_2 + \dots + Y_k$ (13)

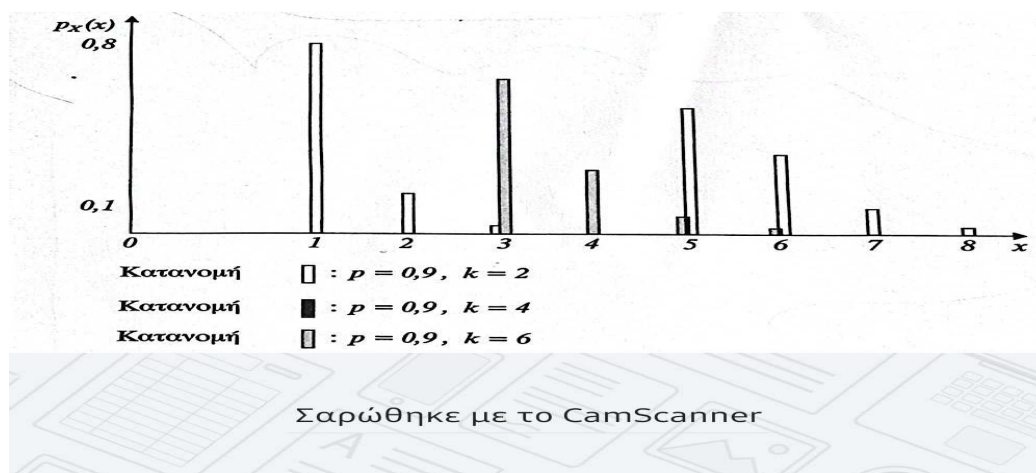
Ο υπολογισμός των τιμών των πιθανοτήτων μιας αρνητικής διωνυμικής κατανομής, μπορεί να γίνει εύκολα με τη βοήθεια της διωνυμικής κατανομής και της σχέσης

$$p_x(x) = \frac{k}{x} p_y(k) \quad (14) \text{ όπου } X \sim NB(k, p) \text{ και } Y \sim B(x, p). \text{ Για } k=1 \text{ η αρνητική}$$

διωνυμική κατανομή γίνεται η γεωμετρική κατανομή. Οι εφαρμογές αυτής είναι ανάλογες με τις εφαρμογές της γεωμετρικής κατανομής. Η αρνητική διωνυμική κατανομή, μερικές φορές παρουσιάζεται με ένα διαφορετικό άλλο ισοδύναμο τρόπο ως εξής: έστω Y ο αριθμός των αποτυχιών που χρειάζονται σε μια ακολουθία δοκιμών Bernoulli μέχρι να εμφανιστεί η k επιτυχία. Τότε, η τυχαία μεταβλητή Y παίρνει τις τιμές $0, 1, 2, \dots$ και αν $Y=y$, ο συνολικός αριθμός των δοκιμών είναι $(y+k)$. Έτσι, η συνάρτηση πιθανότητας της Y είναι

$$p_y(y) = \binom{y+k-1}{y} q^y p^k = \binom{y+k-1}{y} p^k q^y, \quad y = 0, 1, 2, \dots \quad (11^*)$$

Η διαφορά μεταξύ της κατανομής (11) και της (11*) είναι ότι έχουν διαφορετικό στήριγμα. Η (11) έχει στήριγμα τους φυσικούς ακέραιους αριθμούς $\{k, k+1, \dots\}$, ενώ η (11*) τους μη αρνητικούς ακέραιους $\{0, 1, 2, \dots\}$. Θα λέω την κατανομή (11*) εναλλακτική αρνητική διωνυμική.



Σχήμα 4: Αρνητικές διωνυμικές κατανομές

1.4.1 Εφαρμογή

Έστω ότι 20% είναι η πιθανότητα ένα νήπιο εκτιθέμενο σε μια μεταδοτική νόσο να προσβληθεί από αυτή. Η πιθανότητα το 12^ο νήπιο που εκτίθεται στη νόσο να είναι το 3^ο που θα προσβληθεί από αυτή, βρίσκεται υπολογίζοντας το $p_x(12)$ όπου

$$X \sim NB(3, 0.2), \text{ δηλαδή } p_x(12) = \binom{11}{2} 0.2^3 (1-0.2)^9 \approx 0.059$$

1.5 Κατανομή Poisson

Έστω μια ραδιενεργή ουσία η όποια με τη πάροδο του χρόνου εκπέμπει σωματίδια, ενός ορισμένου τύπου. Θέτω την ουσία αυτή υπό παρακολούθηση για ένα μοναδιαίο χρονικό διάστημα π.χ. ένα λεπτό, μια ώρα, κλπ. και ενδιαφέρομαι για τον αριθμό X των σωματιδίων που εκπέμπονται στο διάστημα αυτό. Ο αριθμός X είναι προφανώς μια τυχαία μεταβλητή. Θέλω να βρω την κατανομή της X .

Οι δυνατές τιμές της τυχαίας μεταβλητής X είναι $0, 1, 2, \dots, \infty$. Για να βρω την κατανομή της X , θα χρειαστεί να κάνω ορισμένες υποθέσεις οι οποίες προδικάζουν την ορθότητα του μαθηματικού μου συλλογισμού. Καταρχήν δέχομαι ότι οι συνθήκες του πειράματος παραμένουν σταθερές με την πάροδο του χρόνου. Στη συνέχεια υποδιαιρώ το μοναδιαίο χρονικό διάστημα σε ένα μεγάλο αριθμό n μικρών ίσων υποδιαστημάτων, κατά τέτοιο τρόπο ώστε το πολύ ένα σωματίδιο να δύναται να εκπεμφθεί σε κάθε υποδιαστολή με θετική πιθανότητα. Έστω p η πιθανότητα αυτή και έστω επιπλέον ότι είναι σταθερή για κάθε υποδιάστημα. Όσο μικρότερο είναι το υποδιάστημα, τόσο μικρότερο είναι το p . Έστω ότι η εκπομπή των σωματιδίων είναι ανεξάρτητη από υποδιάστημα σε υποδιάστημα. Τότε, για πρακτικούς λόγους λέω ότι η εκπομπή μοιάζει με μια σειρά δοκιμών Bernoulli με επιτυχία E την εκπομπή ενός σωματιδίου και σταθερό p . Έτσι, μια κατά προσέγγιση τιμή της $P(X=x)$ είναι η

$$\binom{n}{x} p^x q^{n-x}. \text{ Έστω τώρα ότι ο αριθμός } n \text{ των υποδιαστημάτων τείνει στο άπειρο. Καθώς}$$

το n τείνει στο άπειρο, το p τείνει στο μηδέν. Έστω ότι η ποσότητα $\lambda = np$ παραμένει σταθερή. Τότε

$$P(X=x) = \lim_{n \rightarrow \infty} \binom{n}{x} p^x q^{n-x} = \lim_{n \rightarrow \infty} \frac{n(n-1)\dots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{x!} \lim_{n \rightarrow \infty} \frac{n(n-1)\dots(n-x+1)}{nn\dots n} \left(1 - \frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x}{x!} e^{-\lambda}$$

Διότι $\lim_{n \rightarrow \infty} (1 - \lambda/n)^{-x} = 1$ και $\lim_{n \rightarrow \infty} (1 - \lambda/n)^n = e^{-\lambda}$. Η παραπάνω σύγκλιση είναι

ομοιόμορφη ως προς x . Η συνάρτηση $p_x(x) = \frac{e^{-\lambda} \lambda^x}{x!}$, $x=0,1,2,\dots$ και $\lambda > 0$ είναι

πράγματι μια συνάρτηση πιθανότητας, διότι είναι μη αρνητική και επιπλέον

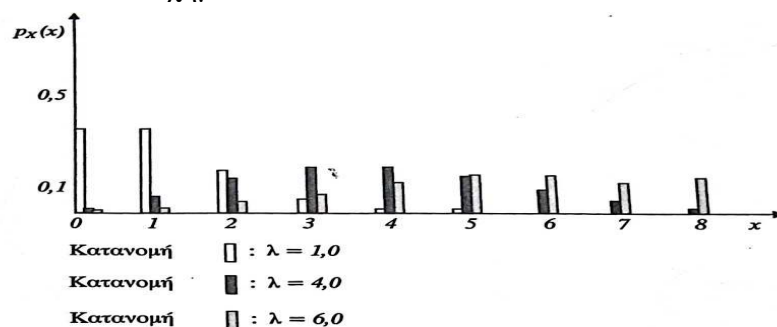
$$\sum_{x=0}^{\infty} p_x(x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

Η (15), δίνει την κατανομή της τυχαίας μεταβλητής X του παραπάνω παραδείγματος και αποτελεί τη περίφημη κατανομή Poisson με παράμετρο λ . Ο Poisson (1781–1840) ήταν Γάλλος μαθηματικός, που ανακάλυψε την παραπάνω κατανομή ως το όριο της διωνυμικής κατανομής, καθώς το $n \rightarrow \infty$ με $\lambda = np$ σταθερό. Αργότερα, η κατανομή αυτή μελετήθηκε μέσα στο γενικότερο πλαίσιο της θεωρίας των στοχαστικών διαδικασιών. Η τυχαία μεταβλητή X που έχει την κατανομή (15), λέγεται τυχαία μεταβλητή Poisson με παράμετρο λ και συμβολίζεται με $X \sim P(\lambda)$. Η αθροιστική συνάρτηση κατανομής της (15), δίνεται από την

$$F(x) = \begin{cases} 0 & , \text{αν } x < 0 \\ \sum_{y=0}^{[x]} \frac{e^{-\lambda} \lambda^y}{y!} & , \text{αν } x \geq 0 \end{cases}.$$

Αντιπροσωπευτικές γραφικές παραστάσεις της

συνάρτηση πιθανότητας της κατανομής του Poisson με $\lambda = 1,0$, $\lambda = 4,0$ και $\lambda = 6,0$ δίνεται στο ακόλουθο σχήμα.



Σαρώθηκε με το CamScanner

Σχήμα 5: Κατανομές του Poisson

Συνήα, η κατανομή Poisson χρησιμοποιείται για τον κατά προσέγγιση υπολογισμό των πιθανοτήτων της διωνυμικής κατανομής. Ο παρακάτω πίνακας συγκρίνει τις τιμές της διωνυμικής κατανομής με $n = 20$ και $p = 0.1$ με τις τιμές της κατανομής του Poisson με $\lambda = np = 20 \cdot 0.1 = 2$.

Πιθανότητες					
X	Διωνυμικές	Poisson	X	Διωνυμικές	Poisson
0	0,1216	0,1353	7	0,0020	0,0034
1	0,2702	0,2707	8	0,0004	0,0009
2	0,2852	0,2707	9	0,0001	0,0002
3	0,1901	0,1804	10	0,0000	0,0000
4	0,0898	0,0902	11	0,0000	0,0000
5	0,0319	0,0361	12	0,0000	0,0000
6	0,0089	0,0120			

Πίνακας 1. Σύγκριση διωνυμικής B (n = 20, p = 0.1) και κατανομής Poisson P (λ = 2).

Ο πίνακας δείχνει ότι η προσέγγιση είναι ικανοποιητική. Η προσέγγιση είναι ιδιαίτερα ικανοποιητική όταν το n είναι μεγάλο και το p μικρό. Κατά κανόνα, η κατανομή Poisson δίνει ικανοποιητική προσέγγιση στις διωνυμικές πιθανότητες όταν $n \geq 20$ και $p \leq 0,05$ ή όταν $n \geq 100$ και $p \leq 0,1$ (ή $\lambda \leq 10$). Η κατανομή Poisson, είναι μια από τις βασικότερες κατανομές της θεωρίας πιθανοτήτων και της στατιστικής. Έχει μεγάλη γενικότητα και τεράστιο εύρος εφαρμογών. Παρόμοιες κατανομές είναι η διωνυμική και η κανονική. Η κατανομή Poisson, εφαρμόζεται όταν έχω μια σειρά τυχαίων ενδεχομένων, τα οποία γίνονται με τη πάροδο του χρόνου και για τα όποια ισχύουν οι υποθέσεις που προαναφέραμε. Συγκεκριμένα, είναι απαραίτητο οι συνθήκες του πειράματος ή του φαινομένου να παραμένουν σταθερές με την πάροδο του χρόνου και τα χρονικά διαστήματα τα οποία δεν επικαλύπτονται, να είναι στατιστικώς ανεξάρτητα με την έννοια ότι η πληροφορία σχετικά με τον αριθμό των τυχαίων ενδεχομένων σε ένα χρονικό διάστημα, να μην αποκαλύπτει τίποτε για το άλλο διάστημα. Επίσης, πρέπει η πιθανότητα να συμβεί ένα τυχαίο γεγονός σε ένα μικρό διάστημα να είναι ανάλογη του μήκους του διαστήματος και η πιθανότητα να συμβούν δυο ή περισσότερα τυχαία ενδεχόμενα σε ένα πολύ μικρό διάστημα, να είναι αμελητέα.

Η τελευταία αυτή υπόθεση, συμφωνεί με τη διαισθητική εικόνα που έχω σχετικά με μεμονωμένα τυχαία ενδεχόμενα. Αποκλείει όμως την περίπτωση τα ενδεχόμενα μου να εμφανίζονται ανά ζεύγη. Για παράδειγμα, αναφέρω την περίπτωση των τροχαίων ατυχημάτων. Η πιθανότητα να συμβούν δυο ατυχήματα, σε ένα μικρό χρονικό διάστημα, είναι αμελητέα σε σύγκριση με την πιθανότητα να συμβεί ένα ατύχημα. Ένα ατύχημα όμως, ενδέχεται να περιλαμβάνει δυο αυτοκίνητα και έτσι αν με τυχαίο ενδεχόμενο εννοώ «καταστροφή ενός αυτοκινήτου», τότε τα τυχαία ενδεχόμενα ενδέχεται να εμφανιστούν ανά ζεύγη και οι βασικές υποθέσεις δεν ισχύουν.

1.5.1 Διαδικασία Poisson

Ας επανέλθουμε στο αρχικό παράδειγμα της ραδιενεργού ουσίας και ας θεωρήσω αντί του μοναδιαίου χρονικού διαστήματος, ένα οποιοδήποτε χρονικό διάστημα μήκους t. Έστω $X(t)$ ο αριθμός των σωματιδίων που εκπέμπονται στο διάστημα $(0, t)$. Ζητάω να προσδιορίσω τη κατανομή $p(x, t)$ της τυχαίας μεταβλητής $X(t)$, για σταθερό t. Σκεπτόμενος όπως προηγουμένως, υποδιαιρώ το διάστημα $(0, t)$ σε $n = t / \Delta t$ υποδιαστήματα μήκους Δt και δέχομαι τις ίδιες υποθέσεις. Δέχομαι ότι έχω πάλι δόκιμες Bernoulli με πιθανότητα επιτυχίας

$$p = P\{\text{μία E σε ένα οποιοδήποτε υποδιάστημα μήκους } \Delta t\} = \lambda \Delta t$$

Τότε, παίρνοντας το όριο της διωνυμικής κατανομής με $\Delta t \rightarrow 0$ (ή $n \rightarrow \infty$) βρίσκω, κατά τον ίδιο τρόπο, την κατανομή Poisson με παράμετρο λt .

$$p(x, t) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}, \quad x = 0, 1, 2, \dots, \quad (15)$$

Έστω τώρα ότι το t μεταβάλλεται και ότι για κάθε t οι ίδιες αρχικές υποθέσεις ισχύουν. Τότε, η (15) δίνει τη συνάρτηση πιθανότητας της τυχαίας μεταβλητής $X(t)$ για κάθε t . Η οικογένεια των τυχαίων μεταβλητών (ή η τυχαία συνάρτηση) $X(t)$ με $t > 0$ και συνάρτηση πιθανότητας την $p(x, t)$ της (15), λέγεται **στοχαστική διαδικασία Poisson** ή απλώς **διαδικασία Poisson**. Οι υποθέσεις που έχουμε δεχτεί είναι γνωστές και ως αξιώματα της διαδικασίας Poisson και μπορούν να διατυπωθούν ως εξής:

Αξίωμα 1

$P\{1 \text{ E στο διάστημα } (t, t + \Delta t)\} = \lambda \Delta t$ για κάθε t (ιδιότητα στατικότητας)

Αξίωμα 2 $P\{2 \text{ ή περισσότερες E στο διάστημα } (t, t + \Delta t)\} = \text{αμελητέα}$

Αξίωμα 3 (Ιδιότητα ανεξαρτησίας)

$P\{n_1 \text{ E στο διάστημα } (t_1, t_1 + \Delta t_1) \text{ και } n_2 \text{ E στο διάστημα } (t_2, t_2 + \Delta t_2), t_1 < t_2\} =$
 $P\{n_1 \text{ E στο } (t_1, t_1 + \Delta t_1) \} n_2 \text{ E στο } (t_2, t_2 + \Delta t_2)\}$

Με βάση τα παραπάνω αξιώματα, μπορώ να ξαναβρώ την $p(x, t)$ της (15) με τρόπο διαφορετικό από τον προηγούμενο. Ο τρόπος αυτός, χρησιμοποιείται συχνά στη θεωρία των στοχαστικών διαδικασιών. Έστω ότι x τυχαία ενδεχόμενα παρατηρούνται στο διάστημα $(0, t + \Delta t)$. Η αντίστοιχη πιθανότητα είναι $p(x, t + \Delta t)$. Βάσει του 2^{ου} αξιώματος, το παραπάνω σύνθετο ενδεχόμενο μπορεί να γίνει με δυο μόνο ασυμβίβαστους δυνατούς τρόπους:

- x τυχαία ενδεχόμενα παρατηρούνται στο διάστημα $(0, t)$ και κανένα στο διάστημα $(t, t + \Delta t)$ ή
- $x-1$ τυχαία ενδεχόμενα στο διάστημα $(0, t)$ και 1 στο διάστημα $(t, t + \Delta t)$.

Κατά συνέπεια έχω:

$$\begin{aligned} p(x, t + \Delta t) &= P[X(t + \Delta t) = x] \\ &= P\{x \text{ E στο } (0, t) \text{ και } 0 \text{ E στο } (t, t + \Delta t)\} + P\{(x-1) \text{ E στο } (0, t) \text{ και } 1 \text{ E στο } (t, t + \Delta t)\} \\ &= P\{x \text{ E στο } (0, t)\} P\{0 \text{ E στο } (t, t + \Delta t)\} + P\{(x-1) \text{ E στο } (0, t)\} P\{1 \text{ E στο } (t, t + \Delta t)\} \\ &\quad (\text{Λόγω ανεξαρτησίας}) \\ &= p(x, t)(1 - \lambda \Delta t) + p(x-1, t) \lambda \Delta t \end{aligned}$$

Διότι $P\{0 \text{ E στο } (t, t + \Delta t)\} = 1 - \lambda \Delta t$ και

$$P\{1 \text{ E στο } (t, t + \Delta t)\} = \lambda \Delta t$$

Μετά από πράξεις, προκύπτει $\frac{p(x, t + \Delta t) - p(x, t)}{\Delta t} = -\lambda p(x, t) + \lambda p(x-1, t)$

ή παίρνοντας το όριο για $\Delta t \rightarrow 0$ είναι $\frac{\partial p(x, t)}{\partial t} = -\lambda p(x, t) + \lambda p(x-1, t)$

Η παραπάνω διαφορική εξίσωση ισχύει για $x = 0, 1, 2, \dots$ και επιλύεται ως ακολούθως:

- Για $x=0$ έχω την ομογενή διαφορική εξίσωση $\frac{\partial p(0,t)}{\partial t} = -\lambda p(0,t)$

δεδομένου ότι $p(-1,t) = 0$ για κάθε t . Με τη βοήθεια της αρχικής συνθήκης $p(0,0) = 1$ βρίσκω τη λύση $p(0,t) = e^{-\lambda t}$.

- Για $x=1$ παίρνω τη γραμμική διαφορική εξίσωση $\frac{\partial p(1,t)}{\partial t} + \lambda p(1,t) = \lambda e^{-\lambda t}$

με λύση την $p(1,t) = \lambda t e^{-\lambda t}$, δεδομένου ότι $p(1,0) = 0$.

- Για $x=2$ προκύπτει η γραμμική διαφορική εξίσωση $\frac{\partial p(2,t)}{\partial t} + \lambda p(2,t) = \lambda^2 t e^{-\lambda t} / 2$

Τελικά με τη χρήση της επαγωγής βρίσκω $p(x,t) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}$, $x=0,1,2,\dots$, $t > 0$, $\lambda > 0$

Τυχαία ενδεχόμενα, τα οποία ακολουθούν τις υποθέσεις της κατανομής Poisson λέγονται συχνά «**σπάνια ενδεχόμενα**». Ο όρος «σπάνια» ενδεχόμενα δεν είναι πετυχημένος και είναι άσχετος με τη χρήση ή το θεμελιώδη ρόλο της κατανομής Poisson. Εκτός από την περίπτωση των **ραδιενεργών διασπάσεων** και των **τροχαίων ατυχημάτων**, η κατανομή Poisson εφαρμόζεται στις τηλεφωνικές κλήσεις που φτάνουν σε ένα τηλεφωνικό κέντρο κατά τη διάρκεια μίας χρονικής περιόδου. Στη **βιολογία**, ο αριθμός $X(k)$ των κυττάρων με k ακριβώς χρωματοσωμικές αλλαγές ακολουθεί την κατανομή Poisson. Χρωματοσωμικές αλλαγές, είναι οι διαδικασίες που διενεργούν σε οργανικά κύτταρα οι ακτίνες X. Ο αριθμός των αυτοκινήτων (ή πελατών) που φτάνουν σε ένα σταθμό διοδίων (ή σε ένα super market) σε μια χρονική περίοδο, ακολουθεί την κατανομή Poisson. Οι διάφορες εφαρμογές της κατανομής (διαδικασίας) Poisson έχουν επηρεάσει και την ορολογία της. Συχνά, τα τυχαία ενδεχόμενα αναφέρονται και ως «**αφίξεις ενδεχομένων**» και το λ ορίζεται ως ο «**ρυθμός των αφίξεων**». Έστω, σε μια διαδικασία Poisson, T_1 ο χρόνος μέχρι την άφιξη του 1^{ου} ενδεχομένου, T_2 ο χρόνος μεταξύ του 1^{ου} και του 2^{ου} ενδεχομένου, κ.ο.κ. Η ακολουθία $\{T_n, n=1,2,\dots\}$ λέγεται **ακολουθία ενδοαφιξιακών χρόνων** και οι τυχαίες μεταβλητές $T_1, T_2, \dots$ είναι ανεξάρτητες εκθετικές με παράμετρο λ .

1.5.2 Εφαρμογές

1. Το τηλεφωνικό κέντρο ενός νοσοκομείου, δέχεται τηλεφωνήματα που ακολουθούν τη κατανομή Poisson με ρυθμό αφίξεων 2 τηλεφωνήματα ανά 30 λεπτά κατά μέσον όρο. Να βρεθεί η πιθανότητα, το τηλεφωνικό κέντρο να δεχτεί τουλάχιστον 3 τέτοια τηλεφωνήματα από τις 12 το μεσημέρι μέχρι τις 2 μ.μ. Έστω X ο αριθμός των τηλεφωνημάτων από 12–2 μ.μ. Τότε, η τυχαία μεταβλητή X ακολουθεί την κατανομή Poisson. Αν ως μονάδα του χρόνου ληφθεί η ώρα, τότε $\lambda = 4$ άρα $\lambda t = 4 \times 2 = 8$. Η ζητούμενη πιθανότητα είναι:

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - P(X=0) - P(X=1) - P(X=2)$$

$$= 1 - \frac{e^{-8} 8^0}{0!} - \frac{e^{-8} 8^1}{1!} - \frac{e^{-8} 8^2}{2!} \approx 0,9862$$

Η κατανομή (ή η διαδικασία Poisson), δίνει την κατανομή των τυχαίων ενδεχομένων ή σημείων κατά μήκος του άξονα του χρόνου. Η ίδια κατανομή, μπορεί

να εφαρμοστεί και για να περιγράψω την κατανομή τυχαίων σημείων στο επίπεδο ή στο χώρο των 3 διαστάσεων και γενικότερα στο χώρο των k διαστάσεων. Αντί για διαστήματα μήκους t , έχω περιοχές με εμβαδόν ή όγκο t και η βασική υπόθεση είναι ότι η πιθανότητα να βρω x σημεία σε μια οποιαδήποτε περιοχή, εξαρτάται μόνο από το εμβαδόν ή τον όγκο της περιοχής και όχι από το σχήμα της. Οι ίδιες υποθέσεις ισχύουν όπως πριν:

- Αν το t είναι μικρό, η πιθανότητα να βρω περισσότερα από ένα σημεία της περιοχής εμβαδού ή όγκου t , είναι μικρή σε σύγκριση με το t .
- Μη επικαλυπτόμενες περιοχές, είναι στατιστικώς ανεξάρτητες.

Ακολουθώντας την ίδια μέθοδο, βρίσκω και πάλι την κατανομή Poisson με παράμετρο λt . Αστέρρες στο διάστημα, σταφίδες σε ένα γλυκό, σπόροι ζιζανίων αναμειγμένοι με σπόρους χλόης, ελαττώματα σε υλικά, βακτηρίδια σε ένα υγρό ή ένα δίσκο μικροσκοπίου, κατανέμονται σύμφωνα με το μοντέλο πιθανότητας Poisson.

2. Ο μέσος όρος των κηλίδων που εμφανίζονται σε κάποιο τύπο ταινίας είναι μία ανά 2.000 μέτρα. Με την υπόθεση ότι ο αριθμός των κηλίδων ακολουθεί τη κατανομή Poisson, να βρεθεί η πιθανότητα σε 5.000 μέτρα ταινίας

(α) να μην υπάρχουν κηλίδες

(β) να υπάρχουν το πολύ δύο κηλίδες και

(γ) να υπάρχουν ακριβώς δυο κηλίδες.

Όταν ως μονάδα μήκους ληφθεί το μέτρο, τότε ο μέσος όρος λ των κηλίδων ανά μέτρο είναι $\lambda = 1 / 2000 = 0,0005$. Για τα 5000 μέτρα έχω $\lambda t = 0,0005 \times 5000 = 2,5$

Και έτσι $P(X = x) = \frac{2,5^x e^{-2,5}}{x!}, x = 0, 1, 2, \dots$

(α)

$$P(X = 0) = \frac{2,5^0 e^{-2,5}}{0!} \approx 0,0821$$

(β)

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{2,5^0 e^{-2,5}}{0!} + \frac{2,5^1 e^{-2,5}}{1!} + \frac{2,5^2 e^{-2,5}}{2!} \approx 0,5438$$

(γ)

$$P(\text{τουλάχιστον } 2 \text{ κηλίδες}) = P(X \geq 2) = 1 - P(X \leq 1) = 1 - \frac{2,5^0 e^{-2,5}}{0!} - \frac{2,5^1 e^{-2,5}}{1!} = 0,7125$$

3. Στο παρακάτω πίνακα δίνω τα αποτελέσματα των φημισμένων πειραμάτων των Rutherford και Geiger, οι οποίοι παρατήρησαν τους αριθμούς των σωματιδίων α που εκπέμπονται από μια ραδιενεργό ουσία σε $n = 2608$ χρονικές περιόδους των 7.5 δευτερολέπτων η κάθε μία. Στον πίνακα αυτό f_x παριστάνει τον αριθμό των περιόδων (συχνότητα) στις οποίες ο αριθμός των εκπεμπόμενων σωματιδίων σε μια χρονική

$$\text{περίοδο των } 7.5 \text{ sec είναι } \lambda = \frac{\sum_{x=0}^{10} f_x x}{n} = \frac{10.086}{2.608} = 3,87$$

Αν η κατανομή του αριθμού των εκπεμπόμενων σωματιδίων είναι Poisson, τότε

$$P(X = x) = e^{-3,87} \frac{3,87^x}{x!}. \text{ Οι Rutherford και Geiger, υπολόγισαν τις πιθανότητες}$$

$P(X = x)$ εκπομπής x σωματιδίων σε μία χρονική περίοδο και στη συνέχεια τις

αναμενόμενες συχνότητες $nP(X=x)$ που δίνονται στην 3^η στήλη του παρακάτω πίνακα βλέπω ότι οι θεωρητικές τιμές (συχνότητες) $nP(X=x)$ συμπίπτουν κατά πολύ με τις συχνότητες που πραγματοποιήθηκαν (2^η στήλη πίνακα). Έτσι μπορώ να πω ότι το μοντέλο Poisson περιγράφει ικανοποιητικά το φαινόμενο της εκπομπής σωματιδίων α από μία ραδιενεργό ουσία.

x	f_x	$nP(X=x)$
0	57	54,399
1	203	210,523
2	383	407,361
3	525	525,496
4	532	508,418
5	408	393,515
6	273	253,817
7	139	140,325
8	45	67,882
9	27	29,189
	16	17,075

Πίνακας 2:

Πρέπει να εκτιμήσω το μέγεθος των διακυμάνσεων που οφείλεται στην τύχη για να κρίνουμε την καταλληλότητα της προσαρμογής (goodness of fit). Ο στατιστικός, κρίνει την καταλληλότητα της προσαρμογής με το τεστ χ^2 που μελετάται στη στατιστική.

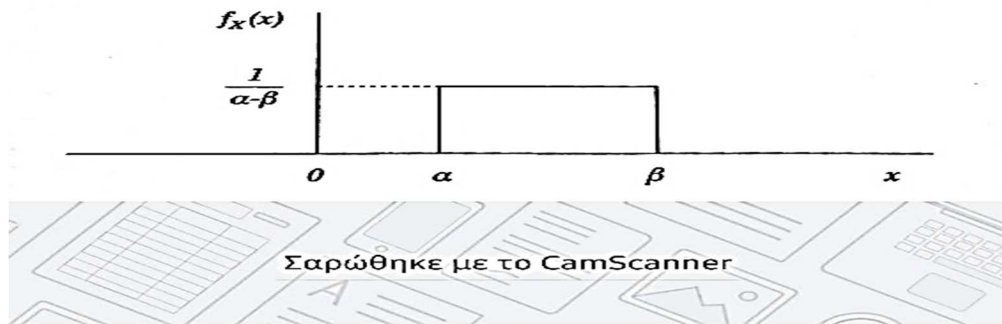
2. Ειδικές συνεχείς κατανομές

2.1 Ομοιόμορφη κατανομή

Η απλούστερη από τις συνεχείς κατανομές, η ομοιόμορφη κατανομή είναι η κατανομή που εκφράζει το ισοπίθανο (ομοιόμορφο) των αποτελεσμάτων ενός πειράματος. Εμφανίζεται στη διακριτή και στη συνεχή περίπτωση. Έστω X μία τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha}, & \alpha \leq x \leq \beta \\ 0, & \text{αλλού} \end{cases}$$

όπου α και β είναι σταθεροί πραγματικοί αριθμοί. Τότε η X λέγεται ομοιόμορφη τυχαία μεταβλητή στο διάστημα (α, β) και συμβολίζεται με $X \sim U(\alpha, \beta)$. Η κατανομή της X λέγεται ομοιόμορφη στο διάστημα (α, β) . Καταχρηστικά, λέμε ότι η τυχαία μεταβλητή παίρνει τιμές στο διάστημα (α, β) . Αν τα σημεία α, β είναι άγνωστα (μεταβλητά) τότε τα α, β είναι οι παράμετροι της κατανομής. Η γραφική παράσταση της παραπάνω συνάρτησης δίνεται στο παρακάτω σχήμα

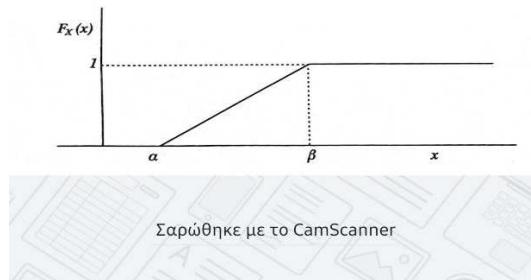


Σχήμα 6: Ομοιόμορφη κατανομή

Η παραπάνω συνάρτηση είναι μια γνήσια συνάρτηση πυκνότητας πιθανότητας. Το εμβαδόν του ορθογωνίου του παραπάνω σχήματος, ισούται με τη μονάδα. Εξ αιτίας του σχήματος αυτού, η ομοιόμορφη κατανομή λέγεται και **ορθογώνια κατανομή**. Η αθροιστική συνάρτηση κατανομής της ομοιόμορφης τυχαίας μεταβλητής είναι

$$f(x) = \begin{cases} 0, & x \leq \alpha \\ \frac{x - \alpha}{\beta - \alpha}, & \alpha \leq x \leq \beta \\ 1, & x \geq \beta \end{cases}$$

και η γραφική της παράσταση δίνεται στο παρακάτω σχήμα



Σχήμα 7: Αθροιστική συνάρτηση ομοιόμορφης κατανομής

Η χαρακτηριστική ιδιότητα της ομοιόμορφης κατανομής είναι ότι ίσα υποδιαστήματα του (α, β) έχουν την ίδια πιθανότητα. Οι εφαρμογές της $U(\alpha, \beta)$ στηρίζονται σε αυτή ακριβώς την ιδιότητα. Χρησιμοποιείται ευρύτατα για την επεξήγηση διαφόρων πτυχών της στατιστικής θεωρίας.

2.2 Εκθετική κατανομή

Έστω μια διαδικασία Poisson $X(t)$ με παράμετρο λ . Τα τυχαία ενδεχόμενα αυτής της διαδικασίας παρατηρούνται με την πάροδο του χρόνου. Έστω T ο χρόνος αναμονής από την αρχή των μετρήσεων μέχρι την εμφάνιση (άφιξη) του πρώτου τυχαίου ενδεχομένου. Το T είναι μια συνεχής τυχαία μεταβλητή. Για να βρω την κατανομή πιθανότητας της τυχαίας μεταβλητής T σκέπτομαι ότι ο χρόνος αναμονής T είναι μεγαλύτερος μίας τιμής t , $t \geq 0$ όταν δεν έχω καμία άφιξη τυχαίου ενδεχομένου στο χρονικό διάστημα $(0, t)$, δηλαδή $\{T \geq t\} = \{X(t) = 0\}$

Άρα, $P(T > t) = P(X(t) = 0) = e^{-\lambda t}$,

Δεδομένου ότι η συνάρτηση πυκνότητας πιθανότητας της $X(t)$ είναι $e^{-\lambda t}(\lambda t)^x/x!$

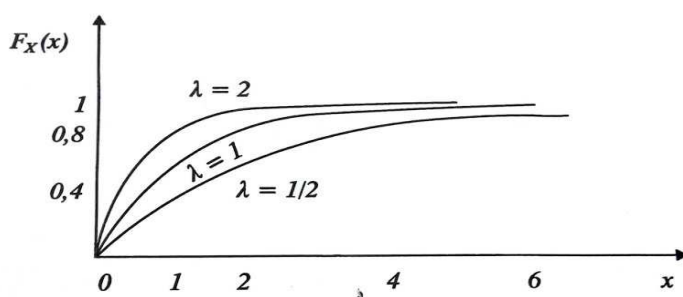
Συνεπώς, $P(T \leq t) = 1 - P(T > t) = 1 - e^{-\lambda t}$,

Και η αθροιστική συνάρτηση κατανομής της T είναι $F_T(t) = \begin{cases} 0, & t \leq 0 \\ 1 - e^{-\lambda t}, & t > 0 \end{cases}$

Παραγωγίζοντας την $F_T(t)$ βρίσκω τη συνάρτηση πυκνότητας πιθανότητας της T

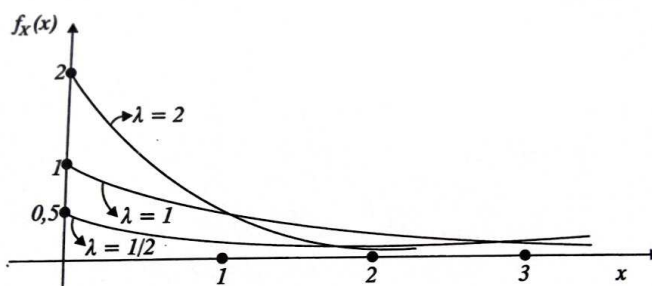
$$f_T(t) = \begin{cases} 0, & t \leq 0 \\ \lambda e^{-\lambda t}, & t > 0 \end{cases}, \lambda > 0$$

Από τον ορισμό της διαδικασίας Poisson είναι φανερό ότι η παραπάνω κατανομή δίνει την κατανομή του χρόνου αναμονής μεταξύ δύο οιασδήποτε αφίξεων τυχαίων ενδεχομένων. Μια τυχαία μεταβλητή X με συνάρτηση πυκνότητας πιθανότητας την και αθροιστική συνάρτηση κατανομής την (όπου θέτω x αντί t) λέγεται εκθετική (ή αρνητική εκθετική) τυχαία μεταβλητή με παράμετρο λ και συμβολίζεται με $X \sim \text{Εκθ}(\lambda)$. Η αντίστοιχη κατανομή λέγεται εκθετική. Είναι εύκολο ναδειχθεί ότι το ολοκλήρωμα της από $-\infty$ έως $+\infty$ ισούται με τη μονάδα. Ο υπολογισμός πιθανοτήτων διαστημάτων βάσει της εκθετικής κατανομής γίνεται με τη βοήθεια της αθροιστικής συνάρτησης κατανομής ή με απευθείας ολοκλήρωση της συνάρτησης πυκνότητας πιθανότητας. Η γραφική παράσταση της αθροιστικής συνάρτησης κατανομής της εκθετικής τυχαίας μεταβλητής δίνεται στο σχήμα και της συνάρτησης πυκνότητας πιθανότητας στο σχήμα



Σαρώθηκε με το CamScanner

Σχήμα 8: Αθροιστικές συναρτήσεις εκθετικών κατανομών



Σαρώθηκε με το CamScanner

Σχήμα 9: Συναρτήσεις πυκνότητας πιθανότητας εκθετικών κατανομών

2.2.1 Εφαρμογές

1. Έστω ότι οι πέστροφες που πιάνει ένας ψαράς ακολουθούν τη διαδικασία Poisson με $\lambda=1,8$ και μονάδα χρόνου την ώρα. Να βρεθεί η πιθανότητα μετά ένα

ψάρεμα, ο ψαράς να περιμένει λιγότερο από 20 λεπτά μέχρι το νέο ψάρεμα. Ο χρόνος αναμονής μεταξύ δύο ψαρεμάτων ακολουθεί την εκθετική κατανομή με παράμετρο $\lambda=1,8$. Τα 20 λεπτά είναι $1/3$ της ώρας. Έτσι, η πιθανότητα που ζητώ είναι:

$$P(T \leq 1/3) = \int_0^{1/3} 1,8e^{-1,8t} dt = 1 - e^{-1,8/3} = 0,4511$$

Ο χρόνος ή η διάρκεια της ζωής ηλεκτρικών ή ηλεκτρικών συσκευών ή άλλων μηχανημάτων (ψυγεία, πλυντήρια κλπ.) συνήθως ακολουθεί την εκθετική κατανομή. Το ίδιο ισχύει και για πολλές τυχαίες μεταβλητές εμπορικοοικονομικής φύσης όπως π.χ. οι ημερήσιες πωλήσεις, ημερήσια κατανάλωση ενός είδους κλπ.

2. Μετά την πρόσληψη ενός νέου υπαλλήλου, η αύξηση των ημερήσιων πωλήσεων ενός καταστήματος ακολουθεί μια εκθετική κατανομή με $\lambda=1/2$. Αν οι ημερήσιες πωλήσεις είναι ανεξάρτητες μεταξύ τους και διαλέξω 3 ημέρες στην τύχη, ποια είναι η πιθανότητα:

(i) και τις 3 ημέρες, η αύξηση να είναι μεγαλύτερη από 10 μονάδες και

(ii) η αύξηση να είναι μεγαλύτερη από 8 μονάδες τουλάχιστον σε 2 από τις 3 ημέρες.

(i) Η πιθανότητα η αύξηση στις πωλήσεις μία οποιαδήποτε τυχαία ημέρα να είναι μεγαλύτερη από 10 μονάδες είναι

$$P(X \geq 10) = \int_{10}^{\infty} e^{-x/2} dx = e^{-5}$$

Λόγω της ανεξαρτησίας, για τις τρεις ημέρες έχω διωνυμικό πείραμα με $p = e^{-5}$. Άρα, η πιθανότητα η ίδια αύξηση να παρατηρηθεί κατά τις 3 ημέρες είναι

$$\binom{3}{3}(e^{-5})^3(1 - e^{-5})^0 = e^{-15} = 3,06 \times 10^{-7}$$

(ii) Η πιθανότητα η αύξηση να είναι μεγαλύτερη από 8 μονάδες την ημέρα είναι

$$P(X \geq 8) = \int_8^{\infty} 1/2 e^{-x/2} dx = e^{-4}$$

Βάσει της διωνυμικής κατανομής, η πιθανότητα είναι

$$\binom{3}{2}(e^{-4})^2(1 - e^{-4}) + \binom{3}{3}(e^{-4})^3(1 - e^{-4})^0 = 3e^{-8} - 2e^{-12} = 9,94 \cdot 10^{-4}$$

3. Έστω ότι ο χρόνος ζωής X ενός εξαρτήματος μιας συσκευής ακολουθεί την εκθετική κατανομή με παράμετρο $\lambda=7$. Έστω επιπλέον ότι το εξάρτημα έχει ζήσει x_0 ώρες. Να βρεθεί η πιθανότητα το εξάρτημα να διαρκέσει τουλάχιστον x πλέον ώρες. Η πιθανότητα που ζητάω είναι

$$\begin{aligned} P(X > x_0 + x | X > x_0) &= \frac{P(X > x_0 + x, X > x_0)}{P(X > x_0)} \\ &= \frac{P(X > x_0 + x)}{P(X \leq x_0)} \\ &= \frac{1 - P(X \leq x_0 + x)}{1 - P(X \leq x_0)} \end{aligned}$$

Παρατηρώ ότι, η πιθανότητα το εξάρτημα να διαρκέσει τουλάχιστον x αρχικές ώρες είναι επίσης e^{-7x} . Το εξάρτημα δεν κουράζεται ή φθείρεται.

4. Έστω ότι οι αφίξεις λεωφορείων σε μία στάση ακολουθούν τη διαδικασία Poisson με $\lambda=6$ αφίξεις ανά ώρα ή 0,1 αφίξεις ανά λεπτό. Να βρεθεί η πιθανότητα κάποιος να περιμένει στη στάση λιγότερο από επιπλέον 5 λεπτά, δοθέντος ότι έχει ήδη περιμένει 15 λεπτά. Βάσει της συνάρτησης η πιθανότητα που ζητάω είναι

$$\begin{aligned}
 P(T \leq 20 | T > 15) &= \int_{15}^{20} \lambda e^{-\lambda(t-15)} dt \\
 &= \int_0^5 \lambda e^{-\lambda t} dt = 1 - e^{-0,1-5} \approx 0,39
 \end{aligned}$$

2.2.2 Ιδιότητα της αμνησίας της εκθετικής κατανομής

Μια χαρακτηριστική ιδιότητα της διαδικασίας του Poisson και της εκθετικής κατανομής είναι γνωστή ως ιδιότητα της αμνησίας. Έστω ότι θέλω να βρω την κατανομή του χρόνου αναμονής T μεταξύ δύο αφίξεων, δοθέντος ότι ο χρόνος αυτός είναι μεγαλύτερος δεδομένου χρόνου t_0 . Δηλαδή, ζητώ την κατανομή της τυχαίας μεταβλητής $T | T > t_0$. Για $t > t_0$ έχω

$$\begin{aligned}
 P(T \leq t | T > t_0) &= \frac{P(T \leq t, T > t_0)}{P(T > t_0)} = \frac{P(t_0 < T \leq t)}{1 - P(T \leq t_0)} \\
 &= \frac{F_t(t) - F_t(t_0)}{1 - F_t(t_0)} = \frac{1 - e^{-\lambda t} - (1 - e^{-\lambda t_0})}{1 - (1 - e^{-\lambda t_0})} \\
 &= 1 - e^{-\lambda(t-t_0)}
 \end{aligned}$$

Παραγωγίζοντας ως προς t , βρίσκω τη συνάρτηση πυκνότητας πιθανότητας της T με $T > t_0$

$$f_{T>t_0}(t) = \begin{cases} 0, & t \leq t_0 \\ \lambda e^{-\lambda(t-t_0)}, & t > t_0 \end{cases}$$

Ο επιπρόσθετος χρόνος αναμονής έχει την ίδια κατανομή με τον αρχικό χρόνο αναμονής, δηλαδή εκθετική κατανομή με παράμετρο λ . Είναι εύκολο τώρα να δικαιολογήσω την ακόλουθη σχέση

$$P(T > t + t_0 | T > t_0) = P(T > t)$$

Δηλαδή, η πιθανότητα αναμονής που συνεχίζεται σε χρόνο t_0 να συνεχιστεί πέρα από το χρόνο $t + t_0$ είναι ανεξάρτητη της προηγούμενης διάρκειας της αναμονής. Αυτό λέγεται αμνησία της εκθετικής κατανομής. Η διαδικασία Poisson, κάθε στιγμή ξεχνά το παρελθόν και αρχίζει εκ νέου. Η εκθετική κατανομή, είναι η μόνη από τις συνεχείς κατανομές με την ιδιότητα της αμνησίας. Η μόνη από τις διακριτές κατανομές που έχει την ίδια ιδιότητα είναι η γεωμετρική κατανομή. Αν ο χρόνος ζωής ενός ατόμου ή ενός εξαρτήματος μηχανής ακολουθεί την εκθετική κατανομή (ή γεωμετρική στην περίπτωση που ο χρόνος μετριέται με διακριτό τρόπο), τότε γέραςμα δε γίνεται εφόσον το άτομο ζει, έχει την ίδια πιθανότητα να αποσυντεθεί την επόμενη στιγμή.

Αντίστροφα, αν είναι γνωστό ότι ένα φαινόμενο έχει την ιδιότητα της αμνησίας ή έλλειψη γεράσματος, τότε η κατανομή της διάρκειας του πρέπει να είναι γεωμετρική ή εκθετική.

2.3 Κατανομή Γ

Η κατανομή Γ , αποτελεί γενίκευση της εκθετικής και χρησιμοποιείται για την περιγραφή διαφόρων φυσικών φαινομένων. Μπορεί να αποδειχθεί ότι ο χρόνος αναμονής μέχρι της εμφάνισης του α στη σειρά (α φυσικός αριθμός) τυχαίου ενδεχομένου σε μια διαδικασία Poisson με παράμετρο $\lambda=1/\beta$, ακολουθεί την κατανομή Γ . Η συνάρτηση πυκνότητας πιθανότητας της κατανομής αυτής είναι

$$f(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, & x > 0 \end{cases} \quad \text{όπου } \alpha > 0, \beta > 0 \text{ είναι οι παράμετροι της}$$

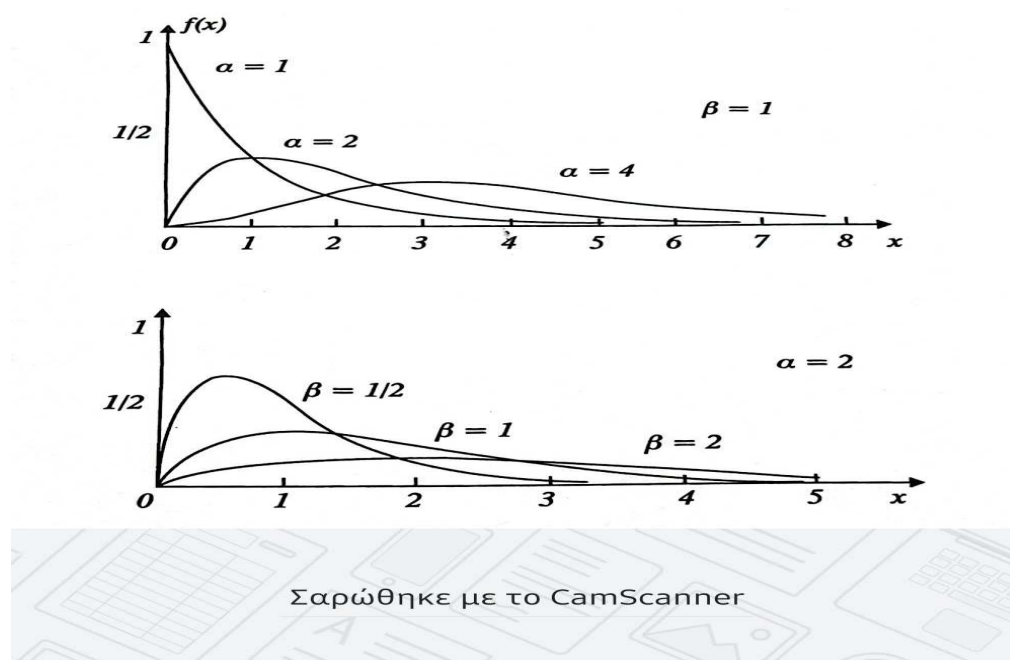
κατανομής. Χρησιμοποιώντας τον ορισμό της συνάρτησης Γ είναι εύκολο να δειχθεί

ότι η συνάρτηση είναι μια γνήσια συνάρτηση πυκνότητας πιθανότητας ως μια αρνητική με ολοκλήρωμα ίσο προς τη μονάδα. Μια τυχαία μεταβλητή X με συνάρτηση πυκνότητας πιθανότητας λέγεται Γ τυχαία μεταβλητή και συμβολίζεται με $X \sim G(\alpha, \beta)$.

Η γραφική παράσταση της Γ συνάρτηση πυκνότητας πιθανότητας δίνεται στο σχήμα για διάφορες τιμές των παραμέτρων α και β .

Αν $\alpha=1$ και $\beta=1/\lambda$, η κατανομή γίνεται η εκθετική κατανομή με παράμετρο λ .

Αν $\alpha=\nu/2$ και $\beta=2$ η κατανομή Γ γίνεται μια άλλη ειδική κατανομή, γνωστή ως x^2 με ν βαθμούς ελευθερίας χ^2_ν .



Σχήμα 10: Κατανομές Γ

Ιδιαίτερα η συνάρτηση πυκνότητας πιθανότητας της κατανομής χ^2_ν είναι

$$f(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}, & x > 0 \end{cases}$$

Αν εξαιρέσει κανείς μερικές απλές περιπτώσεις, πχ. $\alpha = 2$ οποιοδήποτε β , η αθροιστική συνάρτηση κατανομής της Γ κατανομής δε μπορεί να υπολογιστεί με τύπο κλειστής μορφής. Αν το α είναι φυσικός αριθμός, το ολοκλήρωμα

$$F_x(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x y^{\alpha-1} e^{-y/\beta} dy$$

υπολογίζεται επακριβώς, διαφορετικά χρειάζεται αριθμητική ολοκλήρωση. Για τους λόγους αυτούς ο υπολογισμός πιθανοτήτων βάσει της κατανομής Γ , παρουσιάζει δυσκολίες και συχνά γίνεται χρήση πινάκων της συνάρτησης Γ ή, σε ειδικές περιπτώσεις, της κατανομής χ^2_ν .

2.3.1 Σχέση με Poisson

Αν $X \sim G(\alpha, \beta)$ και α φυσικός αριθμός, τότε η αθροιστική συνάρτηση κατανομής της X δίνεται από τον τύπο

$$F_x(x) = 1 - \sum_{j=0}^{a-1} \frac{e^{-x/\beta} (x/\beta)^j}{j!}$$

Ισοδύναμα: Αν $X \sim G(a - \beta)$, α ακαίρεος θετικός και $Y \sim \text{Poisson}(x/\beta)$ τότε

$$P(X \leq x) = P(Y \geq a)$$

Για $\beta = 1$ η $F_x(x)$ λέγεται μη πλήρης συνάρτηση Γ. Εκτεταμένοι πίνακες που δίνουν τις τιμές της συνάρτησης αυτής, υπάρχουν στη βιβλιογραφία. Μέχρι τώρα έχω δει δύο είδη χρόνου αναμονής, διακριτός και συνεχής. Ο διακριτός χρόνος αναμονής εμφανίζεται με τη γεωμετρική και την αρνητική διωνυμική κατανομή και ο συνεχής με την εκθετική και την κατανομή Γ. Η εκθετική κατανομή, αντιστοιχεί στη γεωμετρική και η Γ στην αρνητική διωνυμική κατανομή.

2.3.2 Εφαρμογές

Η περισσότερη από 1.000 λίτρα ημερήσια κατανάλωση βενζίνης, μετρούμενη σε χιλιάδες λίτρα, ενός πρατηρίου βενζίνης, μπορεί να θεωρηθεί ότι έχει κατανομή Γ με $\alpha = 3$ και $\beta = 2$. Αν οι δεξαμενές του πρατηρίου περιέχουν 5.000 λίτρα, ποια είναι η πιθανότητα το πρατήριο να μείνει χωρίς βενζίνη μια ημέρα. Αν Y είναι η ημερήσια κατανάλωση βενζίνης του πρατηρίου σε χιλιάδες λίτρα, τότε $X = Y - 1 \sim G(3, 2)$. Η

$$\text{συνάρτηση πυκνότητας πιθανότητας της } X \text{ είναι } f_X(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{2^3 2!} x^2 e^{-x/2}, & x > 0 \end{cases}$$

Το πρατήριο θα μείνει χωρίς βενζίνη αν $Y \geq 5$. Άρα, η πιθανότητα που ζητώ είναι

$$\begin{aligned} P(Y \geq 5) &= P(X + 1 \geq 5) = P(X \geq 4) \\ &= \int_4^\infty \frac{1}{2^3 \times 2!} x^2 e^{-x/2} dx \quad (z = x/2) \\ &= \int_2^\infty 1/2 z^2 e^{-z} dz = 5e^{-2} = 0,6767 \end{aligned}$$

2.4 Κατανομή B

Όταν η τυχαία μεταβλητή παίρνει τιμές μεταξύ 0 και 1 (π.χ. συνεχή ποσοστά) συχνά χρησιμοποιώ τη B κατανομή. Η ομοιόμορφη κατανομή στο διάστημα (0, 1), είναι ειδική περίπτωση της κατανομής. Η συνάρτηση πυκνότητας πιθανότητας της B

$$\text{κατανομής είναι } f_X(x) = \begin{cases} 0, & \text{αλλού} \\ \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)} = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, & 0 < x < 1 \end{cases}$$

όπου $\alpha > 0$, $\beta > 0$ είναι οι παράμετροι της κατανομής. Η ισότητα στο 1^ο σκέλος της, στηρίζεται στο θεώρημα. Βάσει του ορισμού της συνάρτησης B είναι εύκολο ναδειχθεί ότι η συνάρτηση είναι μια γνήσια συνάρτηση πυκνότητας πιθανότητας. Μία τυχαία μεταβλητή X με συνάρτηση πυκνότητας πιθανότητας τη συνάρτηση λέγεται B τυχαία μεταβλητή και συμβολίζεται με $X \sim Be(\alpha, \beta)$.

Η γραφική της παράσταση της B δίνεται στο παρακάτω σχήμα. Βλέπω ότι όταν:

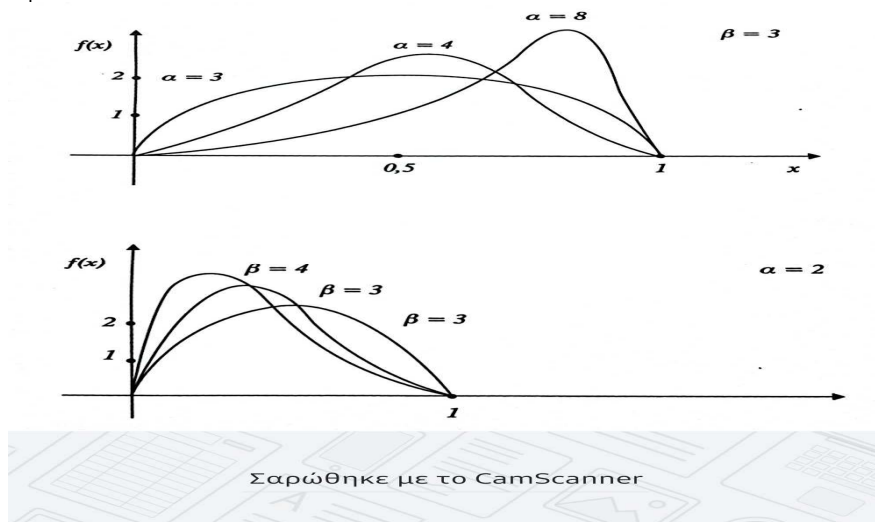
- $\alpha < \beta$ η συνάρτηση πυκνότητας πιθανότητας είναι λοξή προς τα δεξιά
- $\beta < \alpha$ είναι λοξή προς τα αριστερά
- $\alpha = \beta$ η κατανομή είναι συμμετρική.

Η γραφική παράσταση της $Be(\alpha, \beta)$ έχει σχήμα U όταν $\alpha < 1$ και $\beta < 1$ και σχήμα J όταν $\alpha < 1$ ή $\beta < 1$. Όταν $\alpha = \beta = 1$ η κατανομή γίνεται U (0,1).

Η αθροιστική συνάρτηση κατανομής της B κατανομής είναι

$$f_X(x) = \begin{cases} 0, & x \leq 0 \\ \int_0^x \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha, \beta)} dy, & 0 < x < 1 \\ 1, & x \geq 1 \end{cases}$$

Όπως στην περίπτωση της Γ κατανομής έτσι και εδώ, το ολοκλήρωμα της συνάρτησης δεν υπολογίζεται με κλειστό τύπο εκτός αν τα α, β έχουν ορισμένες ακέραιες τιμές. Έτσι, ο υπολογισμός των πιθανοτήτων της B κατανομής παρουσιάζει δυσκολίες. Το ολοκλήρωμα της συνάρτησης λέγεται μη πλήρης συνάρτηση B και εκτεταμένοι πίνακες δίνοντάς τις τιμές του για διάφορα x, α, β υπάρχουν στη βιβλιογραφία.



Σχήμα 11: Κατανομές B

2.4.1 Εφαρμογές

Το ημερήσιο ποσοστό των αρρένων που δέχονται πρώτες βοήθειες σε ένα νοσοκομείο μπορεί να θεωρηθεί ως τυχαία μεταβλητή με κατανομή Be ($\alpha=6, \beta=4$). Να βρεθεί πιθανότητα μια ημέρα το ποσοστό αυτό, να κυμαίνεται μεταξύ 60% και 80%. Η πιθανότητα που ζητώ είναι.

$$P(0,6 \times 0,8) = \int_{0,6}^{0,8} \frac{\Gamma(10)}{\Gamma(4)\Gamma(6)} x^5 (1-x)^3 dx = 0,4032$$

2.5 Κανονική κατανομή ή κατανομή Gauss

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 των κλάδων αυτών. Έχει μοναδικές μαθηματικές, πιθανοθεωρητικές και στατιστικές ιδιότητες και υπό ορισμένες συνθήκες, μπορεί να εφαρμοστεί σε κάθε κοινωνικό πρόβλημα.

Ανακαλύφθηκε το 1733 από τον De Moivre, ως το όριο της διωνυμικής κατανομής. Το 1774 ο Laplace μελέτησε τις μαθηματικές της ιδιότητες. Παρόλα αυτά αποδίδεται στον Gauss, ο οποίος την αναφέρει για πρώτη φορά σε μια εργασία του το 1809. Μελετήθηκε εκτεταμένα τον 18^ο και 19^ο αιώνα από επιστήμονες, οι οποίοι παρατήρησαν ότι τα μοντέλα (κατανομές) των σφαλμάτων των μετρήσεων τους μπορούσαν να προσεγγισθούν ικανοποιητικά από μια συνεχή καμπύλη. Η καμπύλη αυτή λεγόταν κανονική καμπύλη των σφαλμάτων και αποδίδονταν στους νόμους της

τύχης. Προς τιμήν του Gauss, αναφέρεται ως κατανομή Gauss. Σήμερα, ο όρος κανονική κατανομή είναι ο πιο διαδεδομένος. Διακρίνω δύο περιπτώσεις:

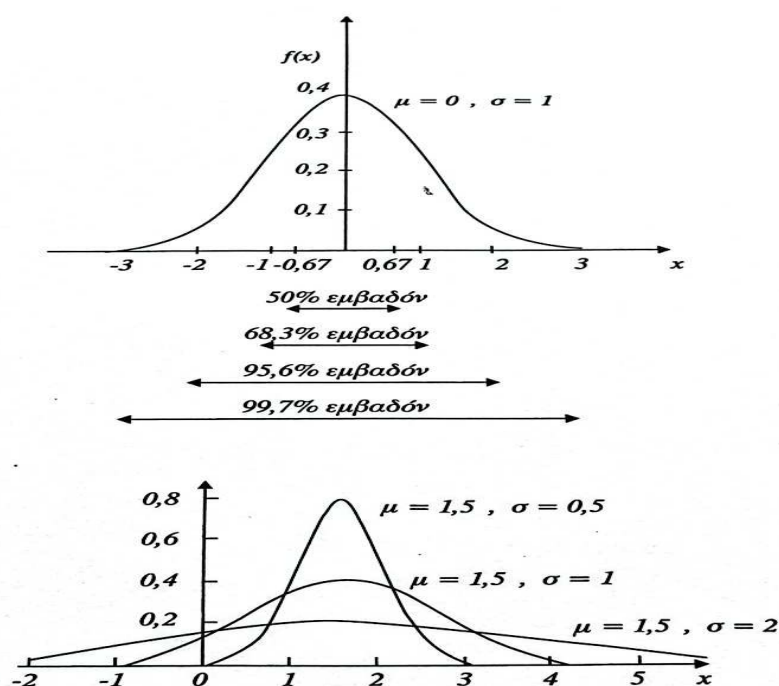
(α) την τυπική (ή τυποποιημένη) κατανομή, με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-1/2x^2}, \quad -\infty < x < +\infty$$

(β) τη γενική κανονική κατανομή, με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty, \quad \begin{matrix} \mu < +\infty \\ \sigma > 0 \end{matrix}$$

όπου μ και σ^2 ή (μ και σ) είναι οι παράμετροι της κατανομής. Οι συναρτήσεις είναι μη αρνητικές. Το θεώρημα αποδεικνύει ότι το ολοκλήρωμα από $-\infty$ έως $+\infty$ ισούται με τη μονάδα και [με τη βοήθεια του μετασχηματισμού $y = (x - \mu)/\sigma$] το ίδιο ισχύει για τη συνάρτηση. Έτσι και οι δύο προηγούμενες συναρτήσεις είναι συνάρτηση πυκνότητας πιθανότητας. Μια τυχαία μεταβλητή X με συνάρτηση πυκνότητας πιθανότητας λέγεται κανονική τυχαία μεταβλητή και συμβολίζεται με $X \sim N(0,1)$. Η τυπική κατανομή προκύπτει από τη γενική θέτοντας $\mu=0$ και $\sigma^2 = 1$. Οι γραφικές παραστάσεις των συναρτήσεων δίνονται στο παρακάτω σχήμα. Το μ αντιπροσωπεύει τη μέση τιμή και το σ^2 τη διακύμανση ή διασπορά της κατανομής. Η συνάρτηση πυκνότητας πιθανότητας της κανονικής είναι συμμετρική γύρω από το σημείο $x=\mu$, δηλαδή $f(\mu+x) = f(\mu-x)$ και το μέγιστο της παρατηρείται στο σημείο μ με τιμή $1/(\sigma\sqrt{2\pi})$. Από αυτό και από το γεγονός ότι το εμβαδό μεταξύ της καμπύλης της $f(x)$ και του άξονα των x ισούται με τη μονάδα συμπεραίνω ότι όσο μεγαλύτερο είναι το σ τόσο περισσότερο απλωμένη είναι η $f(x)$ και αντίστροφα. Η συνάρτηση $f(x)$ έχει σημεία ανάκαμψης τα $x = \mu - \sigma$ και $x = \mu + \sigma$.



Σαρώθηκε με το CamScanner

Σχήμα 12: Κανονική κατανομή

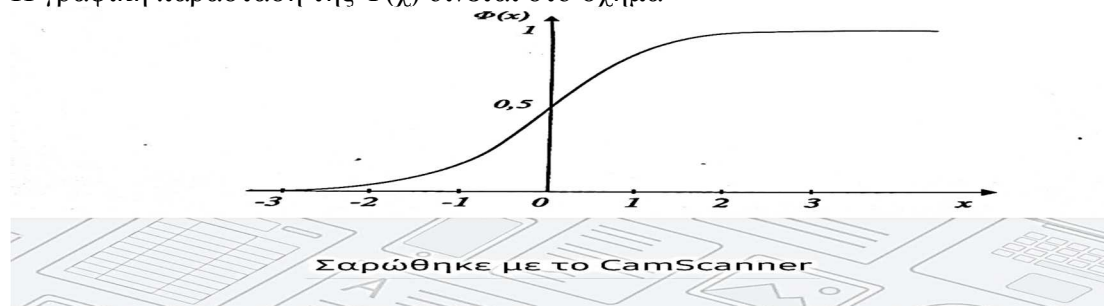
Η αθροιστική συνάρτηση κατανομής, μίας κανονικής τυχαίας μεταβλητής με παραμέτρους μ και σ^2 είναι

$$F_x(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \quad -\infty < x < +\infty$$

Ιδιαίτερα η αθροιστική συνάρτηση κατανομής της τυπικής κανονικής κατανομής συμβολίζεται με $\Phi(x)$, δηλαδή

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-1/2 t^2} dt, \quad -\infty < x < +\infty$$

Η γραφική παράσταση της $\Phi(x)$ δίνεται στο σχήμα



Σχήμα 13: Αθροιστική συνάρτηση τυπικής κανονικής κατανομής

2.5.1 Θεώρημα

Αν $X \sim N(0,1)$ τότε $Y = \sigma X + \mu \sim N(\mu, \sigma^2)$

Αν $X \sim N(\mu, \sigma^2)$ τότε $Y = (X - \mu)/\sigma \sim N(0,1)$

Απόδειξη

Θα αποδείξω μόνο τη 2^η πρόταση με τη μέθοδο της αθροιστικής συνάρτησης κατανομής. Η 1^η πρόταση αποδεικνύεται ανάλογα. Έχω

$$\begin{aligned} F_Y(y) &= (Y \leq y) = P\left(\frac{X - \mu}{\sigma} \leq y\right) \\ &= P(X \leq \sigma y + \mu) = F_X(\sigma y + \mu) \end{aligned}$$

Αλλά η $F_X(x)$ δίνεται από το τύπο. Έτσι αλλάζοντας μεταβλητές έχω

$$F_Y(y) = \int_{-\infty}^{\sigma y + \mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

Η αθροιστική συνάρτηση κατανομή της τυχαίας μεταβλητής Y είναι η $\Phi(y)$ η οποία ορίζει μονοσήμαντα τη κατανομή. Άρα $Y \sim N(0,1)$. Το ολοκλήρωμα με το οποίο ορίζεται η $\Phi(x)$ και κατά συνέπεια και το ολοκλήρωμα της, δε μπορούν να υπολογιστούν με κλειστό τύπο παρά μόνο αριθμητικά. Εκτεταμένοι πίνακες της $f(x)$ και της $\Phi(x)$, υπάρχουν στη βιβλιογραφία. Οι πίνακες της $\Phi(x)$ χρησιμοποιούνται για τον υπολογισμό των πιθανοτήτων της κανονικής κατανομής. Οι πίνακες αυτοί είτε δίνουν την $\Phi(x)$ είτε μόνο το ολοκλήρωμα

$$\int_0^x \frac{1}{\sqrt{2\pi}} e^{-1/2 t^2} dt$$

για θετικές τιμές του x . Λόγω της συμμετρίας της κατανομής, από τις νέες αυτές μπορώ να υπολογίσω την $\Phi(x)$ για κάθε x . Ισχύουν οι ακόλουθες σχέσεις

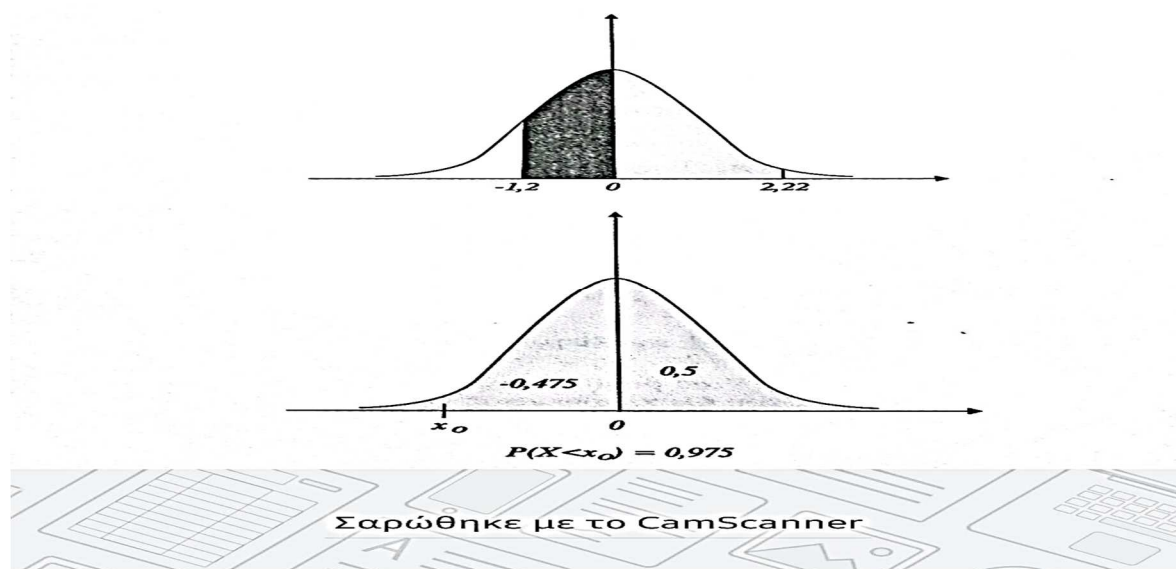
$$\Phi(x) = 1 - \Phi(-x) \quad \text{για κάθε } x$$

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

Δίνω τώρα απλά αριθμητικά παραδείγματα υπολογισμού πιθανοτήτων:

• Έστω $X \sim N(0,1)$. Τότε $P(-1,2 < X < 2,22) = P(-1,2 < X < 0) + P(0 < X < 2,22) = P(0 < X < 1,2) = P(0 < X < 2,22) = 0,3849 + 0,4864 = 0,8713$ Εάν

θέλω να βρω το σημείο x_0 τέτοιο ώστε $P(X > x_0) = 0,975$. Το x_0 πρέπει να είναι αρνητικό. Ισοδύναμα έχω $P(0 < X < -x_0) = 0,475$ και βρίσκω $-x_0 = 1,96$ ή $x_0 = -1,96$. Οι υπολογισμοί αυτοί γραφικά φαίνονται στο παρακάτω σχήμα



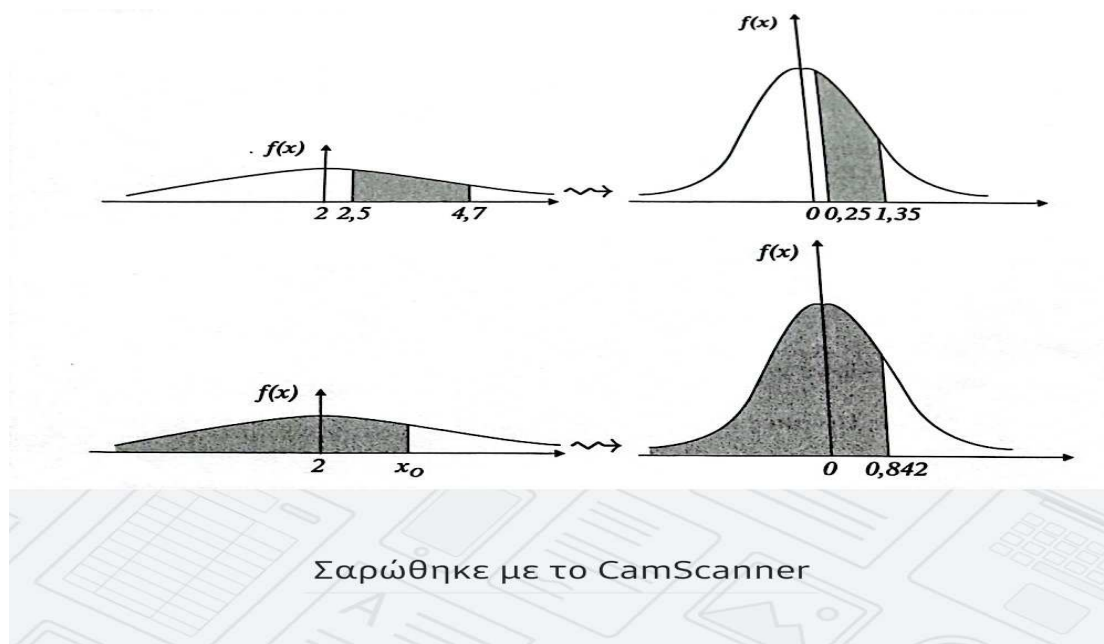
Σχήμα 14: Υπολογισμοί πιθανοτήτων κανονικής κατανομής

• Έστω τώρα $X \sim N(2,4)$, δηλαδή $\mu=2$ και $\sigma^2=4$. Τότε με τη βοήθεια του θεωρήματος έχω $P(2,5 < X < 4,7) = P\left(\frac{2,5-2}{2} < \frac{x-\mu}{\sigma} < \frac{4,7-2}{2}\right) = P(0,25 < Z < 1,35)$, όπου $Z = (X - \mu)/\sigma \sim N(0,1)$. Έτσι $P(2,5 < X < 4,7) = P(0 < Z < 1,35) - P(0 < Z < 0,25) = 0,4115 - 0,0987 = 0,3128$. Για να βρω σημείο x_0 έτσι ώστε $P(X < x_0) = 0,8$ έχω τις ακόλουθες ισοδύναμες σχέσεις $P\left(\frac{x-\mu}{\sigma} < \frac{x_0-2}{2}\right) = 0,8$

$$P\left(0 < Z < \frac{x_0 - 2}{2}\right) = 0,3$$

Από το σχετικό πίνακα παίρνω $\frac{x_0-2}{2} = 0,842$ ή $x_0 = 3,684$

Το παρακάτω σχήμα παριστάνει γραφικά τους παραπάνω υπολογισμούς. Οι πράξεις $Z = (X - \mu)/\sigma$ λέγονται τυποποίηση της τυχαίας μεταβλητής X .



Σχήμα 15: Υπολογισμοί πιθανοτήτων κανονικής κατανομής

Η κανονική κατανομή, δίνει το κατάλληλο μοντέλο πιθανοτήτων για την περιγραφή πολλών τυχαίων φαινομένων. Ποσότητες όπως το μήκος, το βάρος, ο όγκος, το ύψος, η επίδοση, οι βαθμοί κλπ. ακολουθούν την κανονική κατανομή. Ο μέσος όρος μεγάλου αριθμού μετρήσεων, έχει κατανομή που μπορεί να προσεγγιστεί από την κανονική κατανομή. Αυτό συνδέεται με το κεντρικό οριακό θεώρημα, που μελετάται στη θεωρία πιθανοτήτων και στη στατιστική και από το οποίο απορρέει η βασική σημασία της κανονικής κατανομής. Επίσης, η κανονική κατανομή είναι το όριο πολλών κατανομών πιθανότητας. Τυχαία (μη συστηματικά) σφάλματα, ακολουθούν την κατανομή του Gauss. Υπό ορισμένες προϋποθέσεις, η κατανομή των ταχυτήτων των μορίων είναι κανονική (αρχή του Maxwell).

2.5.2 Εφαρμογές

Μία αυτόματη μηχανή πώλησης πορτοκαλάδας, μπορεί να ρυθμιστεί έτσι ώστε να βγάζει κατά μέσο όρο μ λίτρα ανά κύπελλο. Εν το ποσό της πορτοκαλάδας που βγαίνει κάθε φορά έχει κανονική κατανομή με $\sigma=0.008$ λίτρα, να βρεθεί η τιμή (θέση) του μ ώστε κύπελλα των 0,230 λίτρων να ξεχειλίζουν μόνο μία στις εκατό φορές. Έστω X η τυχαία μεταβλητή που μετρά το ποσό της πορτοκαλάδας. Τότε

$$X \sim N(\mu, \sigma^2 = 0,008^2)$$

Για να ξεχειλίζει το κύπελλο, πρέπει να συμβεί το ενδεχόμενο $\{X > 0,230\}$.

Έτσι έχω, $P(X > 0,230) = 0,01$ ή $P\left(\frac{X-\mu}{\sigma} > \frac{0,23-\mu}{0,08}\right) = 0,01$ ή

$$P\left(Z > \frac{0,23 - \mu}{0,008}\right) = 0,01$$

Από τους σχετικούς πίνακες έχω, $\frac{0,23-\mu}{0,008} = 2,329$, άρα $\mu = 0,2281$

Οι βαθμοί σε ένα μάθημα, υποτίθεται ότι ακολουθούν κατά προσέγγιση την κανονική κατανομή με $\mu = 6,5$, $\sigma = 0,5$ και άριστα το δέκα. Αν 3 μαθητές εκλεγούν στην τύχη, να βρεθεί η πιθανότητα ακριβώς 2 από τους 3 μαθητές να έχουν βαθμό μεγαλύτερο από 7. Έστω X η τυχαία μεταβλητή που παριστάνει το βαθμό στο μάθημα. Τότε

$$X \sim N(\mu = 6,5, \sigma^2 = 0,5^2)$$

Έστω p η πιθανότητα ο μαθητής που εκλέγεται στην τύχη να έχει βαθμό μεγαλύτερο από 7. Τότε με τυποποίηση και τη βοήθεια του σχετικού πίνακα έχω

$$p = P(X > 7) = P\left(\frac{X - \mu}{\sigma} > \frac{7 - 6,5}{0,5}\right) \\ = P(Z > 1) = 0,1587$$

Η πιθανότητα αυτή παραμένει η ίδια για κάθε μαθητή τάξης. Υποθέτω δηλαδή ότι το διωνυμικό πρότυπο μπορεί να υιοθετηθεί. Άρα η πιθανότητα που ζητώ είναι

$$\binom{3}{2} p^2 (1 - p) = \binom{3}{2} (0,1586)^2 (0,8414) = 0,0635$$

2.6 Άλλες συνεχείς κατανομές

Ακολουθεί ονομαστική αναφορά άλλων χρήσιμων συνεχών κατανομών

2.6.1 Κατανομή Cauchy

Έχει συνάρτηση πυκνότητας πιθανότητας $f(x) = \frac{1}{\pi} \frac{\alpha}{\alpha^2 + (x - \mu)^2}$ με

$-\infty < x < +\infty$, $\alpha > 0$, $-\infty < \mu < +\infty$. Η κατανομή Cauchy συναντάται κυρίως σε φυσικές επιστήμες και μηχανική. Στη φυσική, σχετίζεται με τη συνάρτηση Lorentz για το εύρος φάσματος. Στη στατιστική, χρησιμοποιείται ως παράδειγμα κατανομής με κακή συμπεριφορά, καθώς παραβιάζει τη βασική υπόθεση για ύπαρξη μέσου όρου και διακύμανσης. Η κατανομή Cauchy μοιάζει με την κανονική κατανομή, αλλά με βαρύτερες ουρές. Αυτές οι ουρές ευθύνονται για την απουσία του μέσου και της διακύμανσης.

2.6.1.1 Χαρακτηριστικά της κατανομής Cauchy

- **Μη ύπαρξη μέσου και διακύμανσης.** Ο μέσος όρος και η διακύμανση δεν ορίζονται για την κατανομή Cauchy, διότι οι αντίστοιχες ολοκληρώσεις δε συγκλίνουν.
- **Βαριά ουρά.** Η κατανομή Cauchy έχει «βαριές ουρές», δηλαδή η πιθανότητα για πολύ μεγάλες (ή μικρές) τιμές x φθίνει σχετικά αργά.
- **Συμμετρία.** Η κατανομή είναι συμμετρική γύρω από το σημείο x_0 .
- **Ταχύτητα σύγκλισης.** Δεν συγκλίνει τόσο γρήγορα όσο η κανονική κατανομή.
- **Αθροιστική ιδιότητα.** Το άθροισμα δύο ανεξάρτητων τυχαίων μεταβλητών X και Y , οι οποίες ακολουθούν κατανομή Cauchy, είναι και πάλι κατανομή Cauchy.

2.6.2 Κατανομή Laplace ή διπλή εκθετική κατανομή

Έχει συνάρτηση πυκνότητας πιθανότητας $f(x) = \frac{1}{2\sigma} e^{-\frac{|x-\mu|}{\sigma}}$ με $-\infty < x < +\infty$,

$\sigma > 0$, $-\infty < \mu < +\infty$. Χρησιμοποιείται σε μηχανολογικά προβλήματα. Στη στατιστική και μοντελοποίηση σφαλμάτων, χρησιμοποιείται για να περιγράψει σφάλματα με βαρύτερες ουρές, τα οποία δεν ταιριάζουν στην κανονική κατανομή. Στη μηχανική και επεξεργασία σήματος και στην επεξεργασία εικόνας, χρησιμοποιείται για τη μοντελοποίηση θορύβου και τη συμπίεση δεδομένων. Στην οικονομία, χρησιμοποιείται για να περιγράψει αλλαγές τιμών ή returns που έχουν συχνά ακραίες τιμές. Στη μοντελοποίηση δεδομένων με ακραίες τιμές, όταν τα δεδομένα περιέχουν ακραίες τιμές (outliers), η κατανομή Laplace είναι πιο κατάλληλη σε σχέση με την κανονική κατανομή.

2.6.2.1 Σχέση της κατανομής Laplace με άλλες κατανομές

- **Κανονική κατανομή.** Η κατανομή Laplace είναι πιο αιχμηρή στο κέντρο και έχει βαρύτερες ουρές από την κανονική κατανομή.

- **Εκθετική κατανομή.** Η κατανομή Laplace μπορεί να θεωρηθεί ως ένωση δύο εκθετικών κατανομών που συμμετρικά τοποθετούνται γύρω από το μ .

- **Gaussian θόρυβος vs Laplacian θόρυβος.** Στην επεξεργασία σήματος, ο θόρυβος Laplace είναι πιο κατάλληλος για δεδομένα με outliers σε σχέση με τον κανονικό θόρυβο.

2.6.3 Λογαριθμοκανονική (long-normal) κατανομή

Έχει συνάρτηση πυκνότητας πιθανότητας $f(x) = \begin{cases} \frac{1}{x\beta\sqrt{2\pi}} e^{-\frac{(\ln x - \ln a)^2}{2\beta^2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$

με $\alpha, \beta > 0$. Αν X έχει τη λογαριθμοκανονική τότε η τυχαία μεταβλητή $\ln X$ έχει την κανονική κατανομή. Έχει χρησιμοποιηθεί σε προβλήματα φυσικής, βιομηχανίας, οικονομικών επιστημών, κοινωνιολογίας, βιολογίας και ανθρωπομετρίας. Η λογαριθμοκανονική κατανομή είναι ιδιαίτερα χρήσιμη σε περιπτώσεις όπου οι τιμές είναι θετικές και οι ακραίες τιμές (μεγάλες τιμές) εμφανίζονται πιο συχνά από ό,τι προβλέπει η κανονική κατανομή. Βρίσκει εφαρμογές στους παρακάτω τομείς:

- **Οικονομικά και χρηματοοικονομικά.** Περιγραφή τιμών μετοχών και αποδόσεων, μοντέλα αύξησης εισοδημάτων.

- **Φυσικές επιστήμες.** Περιγραφή μεγεθών που προκύπτουν από πολλαπλασιαστικές διαδικασίες, όπως το μέγεθος σωματιδίων.

- **Βιολογία και ιατρική.** Χρόνος επιβίωσης μικροοργανισμών, ανάπτυξη όγκων ή οργανισμών.

- **Μηχανική και ανάλυση αξιοπιστίας.** Κατανομή διάρκειας ζωής εξαρτημάτων.

- **Περιβαλλοντικές επιστήμες.** Κατανομή συγκεντρώσεων ρύπων σε φυσικά περιβάλλοντα.

2.6.3.1 Σχέση της λογαριθμοκανονικής κατανομής με άλλες κατανομές

- **Κανονική κατανομή:** Ο λογάριθμος μιας λογαριθμοκανονικής κατανομής ακολουθεί κανονική κατανομή.

- **Εκθετική κατανομή:** Μπορεί να χρησιμοποιηθεί για την περιγραφή θετικών δεδομένων με διαφορετική κλίση.

- **Κατανομή Weibull:** Μια άλλη πιθανή επιλογή για δεδομένα με δεξιά ασυμμετρία.

2.6.4 Κατανομή Weibull

Έχει συνάρτηση πυκνότητας πιθανότητας $f(x) = \begin{cases} \frac{\beta}{\alpha} (x - \gamma)^{\beta-1} e^{-\frac{(x-\gamma)^\beta}{\alpha}}, & x > \gamma \\ 0, & x \leq \gamma \end{cases}$

με $\alpha, \beta, \gamma > 0$. Περιγράφει ικανοποιητικά τη διάρκεια ζωής διαφόρων αντικειμένων, όπως εξαρτήματα, συσκευές κλπ. Χρησιμοποιείται ευρέως στη στατιστική ανάλυση αξιοπιστίας, στη θεωρία διάρκειας ζωής και στις φυσικές επιστήμες και τη μηχανική. Οφείλει το όνομά της στον Σουηδό μαθηματικό Waloddi Weibull, ο οποίος την πρότεινε το 1951. Βρίσκει εφαρμογές στους παρακάτω τομείς:

- **Ανάλυση αξιοπιστίας.** Χρησιμοποιείται για να περιγράψει το χρόνο μέχρι την αποτυχία (time-to-failure) εξαρτημάτων και συστημάτων. Ο ρυθμός αποτυχίας μπορεί να αυξάνεται, μειώνεται ή να είναι σταθερός ανάλογα με το k .

- **Μηχανική και δοκιμές ανθεκτικότητας.** Ανάλυση διάρκειας ζωής μηχανικών εξαρτημάτων.
- **Κλιματολογία.** Μοντελοποίηση ταχυτήτων ανέμου.
- **Οικονομικά.** Μοντελοποίηση κινδύνων και ακραίων γεγονότων.
- **Ιατρική.** Ανάλυση επιβίωσης (survival analysis).

2.6.5 Κατανομή Maxwell

Έχει συνάρτηση πυκνότητας πιθανότητας
$$f(x) = \begin{cases} \frac{2}{\pi} x^2 e^{\frac{-1}{2x^2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Εφαρμόζεται σε προβλήματα της φυσικής (ταχύτητα μορίων). Βρίσκει εφαρμογές στους παρακάτω τομείς:

- **Κινητική θεωρία των αερίων.** Περιγράφει την κατανομή των ταχυτήτων των μορίων σε ένα ιδανικό αέριο που βρίσκεται σε θερμική ισορροπία.
- **Φυσική και στατιστική μηχανική.** Χρησιμοποιείται για την ανάλυση της κατανομής κινητικής ενέργειας σωματιδίων.
- **Μηχανική ρευστών.** Περιγράφει τις ταχύτητες μικροσωματιδίων σε ροές ρευστών.
- **Ακουστική και δονήσεις.** Χρησιμοποιείται στην ανάλυση πλάτους ταλαντώσεων.

2.6.6 Κατανομή ακροτάτων τιμών ή του Gumbel

Έχει συνάρτηση πυκνότητας πιθανότητας $f(x) = e^{-x-e^{-x}}$ με $-\infty < x < +\infty$.

Αποτελεί μέρος των κατανομών ακραίων τιμών και χρησιμοποιείται ευρέως σε στατιστική, μηχανική αξιοπιστίας και ανάλυση κινδύνου και κυρίως για τη μοντελοποίηση ακραίων τιμών (όπως οι μεγαλύτερες ή μικρότερες τιμές σε ένα σύνολο δεδομένων). Βρίσκει εφαρμογές στους παρακάτω τομείς:

- **Ανάλυση ακραίων φαινομένων.** Περιγραφή ακραίων καιρικών φαινομένων, όπως η μέγιστη βροχόπτωση ή οι ισχυροί άνεμοι.
- **Μηχανική αξιοπιστίας.** Πρόβλεψη της μέγιστης αντοχής υλικών μέχρι τη θραύση τους.
- **Χρηματοοικονομικά.** Μοντελοποίηση ακραίων αποδόσεων και κινδύνων (Value at risk, VaR).
- **Φυσικές επιστήμες.** Μελέτη ακραίων τιμών σε φαινόμενα όπως σεισμοί ή πλημμύρες.
- **Ανάλυση χρονικών σειρών.** Προσδιορισμός τανώτατων ορίων τιμών σε δεδομένα.

2.6.7 Κατανομή Pareto

Έχει συνάρτηση πυκνότητας πιθανότητας
$$f(x) = \begin{cases} \frac{ak^a}{x^{a+1}}, & x \geq k \\ 0, & x < k \end{cases} \text{ με } a, k \geq 0$$

Εφαρμόζεται σε οικονομετρικά και ηλεκτρομηχανολογικά προβλήματα, για τη μοντελοποίηση φαινομένων όπου παρατηρούνται **βαριές ουρές**. Βρίσκει εφαρμογές στους παρακάτω τομείς:

- **Οικονομία.** Κατανομή εισοδημάτων ή πλούτου (λίγοι άνθρωποι έχουν πολύ μεγάλο πλούτο, ενώ οι περισσότεροι έχουν χαμηλά εισοδήματα). Παράδειγμα: Ο κανόνας 80/20, όπου το 80% του πλούτου ανήκει στο 20% των ανθρώπων.

- **Ανάλυση διαδικτύου.** Κατανομή επισκεψιμότητας ιστοσελίδων (λίγες ιστοσελίδες έχουν πολλούς επισκέπτες).
- **Ασφαλιστικά ρίσκα.** Μοντελοποίηση ακραίων ασφαλιστικών απαιτήσεων (μεγάλες ζημιές σε σπάνιες περιπτώσεις).
- **Φυσικά φαινόμενα.** Κατανομή μεγεθών σεισμών ή ποταμών. Κατανομή μεγεθών σωματιδίων σε υλικά.
- **Επιχειρηματική ανάλυση.** Κατανομή πωλήσεων προϊόντων: λίγα προϊόντα συνεισφέρουν στο μεγαλύτερο ποσοστό των πωλήσεων.

2.6.7.1 Σχέση κατανομής Pareto με άλλες κατανομές

- Η κατανομή Pareto σχετίζεται με την εκθετική κατανομή, όταν εφαρμόζεται σε λογαριθμική κλίμακα.
- Η γενικευμένη Pareto κατανομή, είναι μια επέκταση που περιλαμβάνει και άλλες παραμέτρους, ώστε να προσαρμόζεται καλύτερα σε δεδομένα.

2.7 Άλλες κατανομές

2.7.1 Αρνητική υπεργεωμετρική κατανομή

Είναι η κατανομή της αντιστροφής δειγματοληψίας (inverse sampling). Θεωρώ το υπεργεωμετρικό σχήμα με N , N_p , N_q στοιχεία για το σύνολο του πληθυσμού και τις δύο κατηγορίες C και $\bar{C}$. Στοιχεία εκλέγονται από τον πληθυσμό το ένα μετά το άλλο, χωρίς επανάθεση, όπως στην απλή τυχαία δειγματοληψία μέχρι να εξαχθούν k στοιχεία της κατηγορίας C , $k \leq N_p$. Έστω X ο αριθμός των επιλογών (εκλογών) αυτών. Η τυχαία μεταβλητή X είναι διακριτική και έχει την αρνητική υπεργεωμετρική κατανομή με

$$\text{συνάρτηση πιθανότητας} \quad P_X(x) = \frac{\binom{N_p}{k-1} \binom{N_q}{x-k}}{\binom{N}{x-1}} \frac{N_p - (k-1)}{N - (x-1)} \quad \text{με}$$

$$x = k, k+1, \dots, N_q + k.$$

2.7.2 Κατανομή Polya

Η κατανομή αυτή είναι διακριτή και προκύπτει από το υπεργεωμετρικό σχήμα ή από την απλή τυχαία δειγματοληψία, με τη διαφορά ότι η εκλογή των n στοιχείων γίνεται με επανάθεση και κάθε φορά που εκλέγω ένα στοιχείο προσθέτω στον πληθυσμό k στοιχεία της ίδιας κατηγορίας με εκείνη που εκλέχτηκε. Αν X είναι ο αριθμός των επιτυχιών, δηλαδή των στοιχείων της κατηγορίας C που εκλέγονται, τότε το X έχει κατανομή Polya με συνάρτηση πιθανότητας

$$p_X(x) = \binom{n}{x} \frac{[a]_x [b]_{n-x}}{[a+b]_n}, \quad x = 0, 1, 2, 3, \dots, n, \quad \text{όπου } a = N_p/k, \quad b = N_q/k \text{ και } [a]_x = a(a+1) \dots (a+x-1).$$

Η κατανομή Polya είναι χρήσιμη για τη μοντελοποίηση φαινομένων. Βρίσκει εφαρμογές στους παρακάτω τομείς:

- **Βιολογία:** Ανάλυση αναπαραγωγής βακτηρίων ή εξάπλωσης ειδών σε ένα οικοσύστημα.
- **Κοινωνικά δίκτυα:** Οι άνθρωποι τείνουν να αλληλοεπιδρούν περισσότερο με δημοφιλή άτομα ή περιεχόμενο.
- **Διάδοση πληροφοριών:** Όσο περισσότερο μιλάμε για κάτι, τόσο πιο πιθανό είναι να συνεχίσουμε να το διαδίδουμε.
- **Συσσώρευση πλούτου:** Τα άτομα που έχουν ήδη πλούτο, τείνουν να κερδίζουν περισσότερο, ενισχύοντας τη θέση τους.

- **Ανάλυση κειμένου:** Μοντελοποίηση λέξεων σε κείμενα (π.χ., στο πλαίσιο της θεματικής ανάλυσης – Topic modeling).
- **Προσομοίωση.** Χρήση του μοντέλου κληρωτίδας για προσομοιώσεις σε στατιστικά προβλήματα και αλγορίθμους.

2.7.3 Κατανομή λογαριθμικής σειράς ή λογαριθμική κατανομή

Η λογαριθμική κατανομή χρησιμοποιείται για να μοντελοποιήσει τυχαία μεγέθη, των οποίων η πιθανότητα εμφάνισης μειώνεται με τον λογαριθμικό ρυθμό, όπως φαινόμενα που περιλαμβάνουν υπερβολικές ή μεγάλες τιμές που εμφανίζονται με χαμηλότερη συχνότητα. Μπορεί να περιγραφεί ως κατανομή για την οποία το φαινόμενο μειώνεται, καθώς το μέγεθός του αυξάνεται και η πιθανότητα εμφάνισης μεγάλων τιμών (π.χ., μεγεθών σε καταστάσεις σαν σεισμούς ή καταστροφές) ακολουθεί ένα λογαριθμικό μοτίβο. Είναι διακριτή και έχει χρησιμοποιηθεί για να μοντελοποιήσει το μέγεθος/ πλήθος βιολογικών ειδών (species). Η συνάρτηση πιθανότητας είναι

$$p_x(x) = -q^x / (x \ln p), \quad x = 1, 2, \dots, \quad 0 < p < 1, \quad p + q = 1.$$

Η λογαριθμική κατανομή (Logarithmic distribution) σχετίζεται με κατανομές που προκύπτουν από τη διαδικασία πολλαπλασιασμού και στη συνέχεια λήψης του λογαρίθμου. Βρίσκει εφαρμογές στους παρακάτω τομείς:

- **Κοινωνικές επιστήμες και οικονομία:** Μοντελοποίηση των εισοδημάτων και του πλούτου, όπου οι περισσότερες παρατηρήσεις είναι κοντά σε χαμηλά επίπεδα, ενώ οι εξαιρετικά υψηλές τιμές (πλούσιοι άνθρωποι) είναι σπάνιες, αλλά υπαρκτές.
- **Βιολογία:** Αριθμός των ειδών σε μια περιοχή ή η κατανομή των μεγεθών των σωματιδίων (π.χ. μεγέθη κυττάρων ή άλλων βιολογικών παραμέτρων).
- **Πληροφορική:** Χρήση στη συσχέτιση δεδομένων ή στην ανάλυση της συμπεριφοράς των χρηστών, όπου ένα μικρό ποσοστό των χρηστών μπορεί να παράγει τις περισσότερες ενέργειες ή αλληλεπιδράσεις.
- **Επιστήμη Δεδομένων:** Ανάλυση των αναζητήσεων ή των άλλων δραστηριοτήτων, όπου ορισμένες λέξεις ή αντικείμενα αναζητούνται πολύ συχνά, ενώ άλλες μόνο σπάνια.

2.7.4 Μίξη κατανομών (mixture distributions) ή σύνθετες κατανομές (compound distributions)

Αν f_1, f_2 είναι δύο κατανομές, τότε η συνάρτηση πιθανότητας $f(x) = [f_1(x) + f_2(x)]/2$ είναι επίσης μια κατανομή. Ομοίως αν $f_1, f_2, \dots, f_n$ είναι κατανομές, $\lambda_1, \lambda_2, \dots, \lambda_n$ θετικοί αριθμοί με $\lambda_1 + \dots + \lambda_n = 1$, τότε η

$$f(x) = \lambda_1 f_1(x) + \lambda_2 f_2(x) + \dots + \lambda_n f_n(x)$$

είναι επίσης μια κατανομή. Οι κατανομές αυτές, λέγονται **μεικτές** ή **σύνθετες κατανομές**. Η ιδέα έχει επεκταθεί περαιτέρω όπως π.χ. στην περίπτωση που η παράμετρος μιας κατανομής έχει κατανομή πιθανότητας. Έτσι, αν η τυχαία μεταβλητή X έχει κατανομή $f(x, \theta)$ και θ είναι τυχαία μεταβλητή με κατανομή $g(\theta)$, τότε η μικτή ή σύνθετη κατανομή που προκύπτει είναι

$$f(x) = \int_{\theta} f(x, \theta) g(\theta) d\theta$$

όπου Θ ο παραμετρικός χώρος. Αν το θ έχει διακριτή κατανομή θα χρησιμοποιήσω άθροισμα αντί ολοκλήρωμα. Ένα κλασσικό παράδειγμα μίξης κατανομών είναι το ακόλουθο: $X \sim \text{Διωνυμική}(n, p)$ και $n \sim \text{Poisson}(\lambda)$. Έχω

$$P(X = x) = \sum_{n=x}^{\infty} \binom{n}{x} p^x q^{n-x} \left[\frac{e^{-\lambda} \lambda^n}{n!} \right] \quad (\text{οι όροι για } n < x \text{ είναι μηδέν})$$

$= e^{-\lambda p} (\lambda p)^x / x!$ (μετά από απλοποίηση) Έτσι η μικτή κατανομή είναι Poisson (λp). Η κατανομή αυτή λέγεται μίξη διωνυμικής–Poisson.

Ελληνόγλωσση βιβλιογραφία

- Αθανασόπουλος, Δ. (1983), Θεωρία πιθανοτήτων, τόμοι I και II, Πειραιάς
- Δρακάτος, Κ. (1984). Στατιστική, 2^η έκδοση, Αθήνα, εκδόσεις Παπαζήση
- Κάκουλλος, Θ. (1972). Στατιστική, θεωρία και εφαρμογές, Αθήνα
- Κάκουλλος, Θ. (1969). Μαθήματα θεωρίας πιθανοτήτων, Αθήνα
- Κουνιάς, Σ. (1977). Μαθήματα θεωρίας πιθανοτήτων και στατιστικής, Μέρος I Πιθανότητες, Θεσσαλονίκη
- Κουνιάς, Σ. και Μωυσιάδης, Χ. (1995). Θεωρία πιθανοτήτων I, Θεσσαλονίκη, εκδόσεις Ζήτη
- Κεβόρκ, Κ. (1972). Στατιστική I, II, III, Αθήνα
- Λαμπράκη, Δ. (1980). Στατιστική, 2^η έκδοση, Αθήνα
- Παπαϊωάννου, Τ. (1997). Θεωρία πιθανοτήτων και στατιστικής, Αθήνα, εκδόσεις Σταμούλη
- Παπαϊωάννου, Τ. (2000). Εισαγωγή στις πιθανότητες, Αθήνα, εκδόσεις Σταμούλη
- Ρούσσας, Γ. (1973). Στοιχεία πιθανοθεωρίας με εφαρμογές, Πάτρα
- Χαραλαμπίδης, Χ. (1990). Θεωρία πιθανοτήτων και εφαρμογές, τόμοι I, II, Αθήνα, εκδόσεις Συμμετρία