

Ποσοτικές Μέθοδοι στην Διαχείριση Κινδύνου

ΕΝΟΤΗΤΑ 2 -α

Βασικές μονοδιάστατες κατανομές

Α. Ν. Γιαννακόπουλος

Ο.Π.Α

Χειμερινό Εξάμηνο 2025-2026

- 1 Βασικές διακριτές κατανομές
 - Η κατανομή *Bernoulli*
 - Η διωνυμική κατανομή
 - Η υπεργεωμετρική κατανομή
 - Η γεωμετρική κατανομή
 - Η αρνητική διωνυμική
 - Η κατανομή *Poisson*
- 2 Συνεχείς κατανομές
 - Η ομοιόμορφη κατανομή
 - Η εκθετική κατανομή
 - Η κανονική κατανομή
 - Η κατανομή *Gamma*
 - Η κατανομή *Beta*
 - Η κατανομή *Cauchy*
 - Η κατανομή *Frechet*
 - Η κατανομή *Gumbel*
 - Η κατανομή *Student*

Διακριτές κατανομές

Παράδειγμα (Πεπερασμένα σύνολα)

Αν το Ω είναι ένα πεπερασμένο σύνολο π.χ. $\Omega = \{\omega_1, \dots, \omega_n\}$ και $\mathcal{F}$ είναι μια σ -άλγεβρα επάνω σε αυτό ένα μετρο πιθανότητας P μπορεί να αναπαρασταθεί με την χρήση ενός διανύσματος $(p_1, \dots, p_n)$ με την ιδιότητα $p_i > 0$ και $\sum_{i=1}^n p_i = 1$.

Σε οποιοδήποτε γεγονός $A \subset \Omega$, το οποίο και μπορεί να γραφεί ως $A = \{\omega_j : j \in J\}$, όπου $J \subset \{1, \dots, n\}$, αντιστοιχούμε την πιθανότητα $P(A) = \sum_{j \in J} P(\omega_j) = \sum_{j \in J} p_j$.

Βασικές διακριτές κατανομές

Η κατανομή *Bernoulli*

Ένα πείραμα μπορεί να έχει 2 εκβάσεις: Επιτυχία ή αποτυχία με πιθανότητα p και $1 - p$ αντίστοιχα.

Ο δειγματικός χώρος είναι το σύνολο $\Omega = \{E, A\} = \{1, 0\}$. Αν $X : \Omega \rightarrow \mathbb{R}$ η τυχαια μεταβλητή που παίρνει την τιμή 1 στην επιτυχία και την τιμή 0 στην αποτυχία, η κατανομή της τυχαιας αυτής μεταβλητής θα είναι

$$f(x) = P(X = x) = \begin{cases} p & x = 1, \\ 1 - p & x = 0 \end{cases}$$

και η αθροιστική συνάρτηση κατανομής θα είναι

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < 0, \\ 1 - p & 0 \leq x < 1, \\ 1 & x \geq 1 \end{cases}$$

Η κατανομή αυτή ονομάζεται κατανομή *Bernoulli*.

Η διωνυμική κατανομή

Ας υποθέσουμε ότι έχουμε ανεξάρτητα επαναλαμβανόμενα πειράματα τα οποία μπορεί να έχουν 2 εκβάσεις: επιτυχία 1 ή αποτυχία 0.

Αν η πιθανότητα επιτυχίας είναι p και η πιθανότητα αποτυχίας $1 - p$, ποιά θα είναι η πιθανότητα σε n πειράματα να έχουμε m επιτυχίες $m = 0, 1, \dots, n$;

Ο δειγματικός χώρος αποτελείται από 'λέξεις' μήκους n οι οποίες αποτελούνται από 2 σύμβολα, το σύμβολο 1 και το σύμβολο 0.

Εμείς μας ενδιαφέρει το πλήθος των λέξεων αυτών που περιέχουν ακριβώς m μονάδες.

Αυτές γνωρίζουμε πως είναι ίσες προς $\binom{n}{m}$.

Μπορούμε λοιπόν να ορίσουμε την τυχαία μεταβλητή $X : \Omega \rightarrow \mathbb{N}$, ως X = αριθμός επιτυχιών σε n πειράματα.

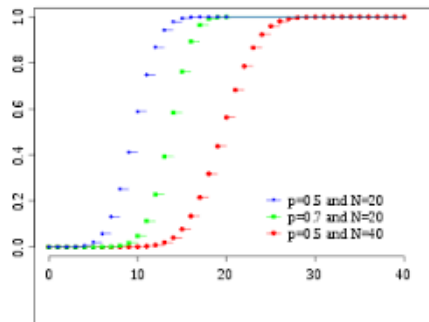
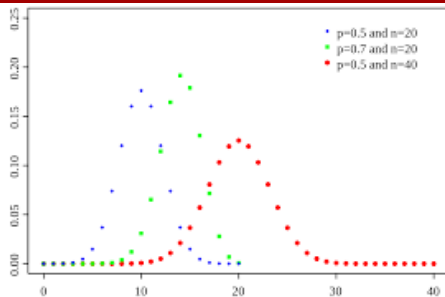
Ας πάρουμε τώρα μια οποιαδήποτε λέξη αυτής της μορφής.

Η πιθανότητα να έρθει μια τέτοια λέξη θα είναι ίση προς $p^m(1-p)^{n-m}$.

Έχουμε $\binom{n}{m}$ ταυτόσημες τέτοιες λέξεις, άρα η πιθανότητα να έρθουν m επιτυχίες σε n πειράματα θα είναι ίση προς

$$P(X = m) = \binom{n}{m} p^m(1-p)^{n-m}.$$

Η κατανομή αυτή ονομάζεται διωνυμική κατανομή και συμβολίζεται $\text{Bin}(n, p)$.



Μπορεί κανείς να δει ότι η διωνυμική κατανομή μπορεί να εκφραστεί μέσω της κατανομής Bernoulli ως εξής:

Αν X_i είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κατανομή Bernoulli (δηλαδή τυχαίες μεταβλητές οι οποίες παίρνουν την τιμή 1 με πιθανότητα p και την τιμή 0 με πιθανότητα $1 - p$) το άθροισμα $X = X_1 + \dots + X_n$ είναι μια τυχαία μεταβλητή η οποία ακολουθεί την κατανομή $Bin(n, p)$.

$$X_i \sim Ber(p) \text{ i.i.d.} \implies X = \sum_{i=1}^n X_i \sim Bin(n, p).$$

Επίσης

$$X_k \sim Bin(N_k, p), \quad k = 1, \dots, K \implies \sum_{k=1}^K X_k \sim Bin\left(\sum_{k=1}^K N_k, p\right)$$

Η αθροιστική συνάρτηση κατανομής για την διωνυμική δίνεται από τον τύπο

$$F(x) = \begin{cases} 0 & x < 0 \\ \sum_{r=0}^{[x]} \binom{n}{r} p^r (1-p)^{n-r} & 0 \leq x < n \\ 1 & x \geq n \end{cases}$$

Η μέση τιμή της διωνυμικής κατανομής δίνεται από το άθροισμα

$$\mathbb{E}[X] = \sum_{r=0}^n r \binom{n}{r} p^r (1-p)^{n-r} = np.$$

Αυτό μπορούμε να το δούμε είτε από το διωνυμικό ανάπτυγμα ή πιο απλά από την αναπαράσταση $X = X_1 + \dots + X_n$ όπου X_i ανεξαρτητες τυχαίες μεταβλητές τύπου Bernoulli.

Έχουμε λοιπόν ότι

$$\mathbb{E}[X] = \mathbb{E}[X_1 + \dots + X_n] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_n] = p + \dots + p = np.$$

Μπορούμε επίσης να υπολογίσουμε και την διασπορά της X .
Αυτό μπορεί να εκφραστεί σαν το άθροισμα

$$\text{Var}(X) = \sum_{r=0}^n (r - np)^2 \binom{n}{r} p^r (1-p)^{n-r} = np(1-p).$$

Εναλλακτικά, από την αναπαράσταση $X = X_1 + \dots + X_n$ και την ανεξαρτησία των X_i , έχουμε ότι

$$\begin{aligned}\text{Var}(X) &= \text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n) \\ &= p(1-p) + \dots + p(1-p) = np(1-p).\end{aligned}$$

Παράδειγμα

Μια τράπεζα έχει δώσει 100 δάνεια (σε 100 ανεξάρτητους πελάτες οι οποίοι θεωρούνται ότι έχουν την ίδια πιθανότητα να μην αποπληρώσουν το δάνειο τους).

Το δάνειο αποπληρώνεται σε μια και μοναδική δόση. Αν το κάθε δάνειο είναι ύψους 10000 ευρώ και ο κάθε πελάτης έχει πιθανότητα 10% να μην αποπληρώσει το δάνειο του ποια θα είναι η αναμενόμενη ζημιά της τράπεζας από το χαρτοφυλάκιο αυτό.

Πόσο μπορούμε να εμπιστευτούμε την αναμενόμενη ζημιά σαν μια πρόβλεψη της πραγματικής ζημιάς;

Η πιθανότητα να μην αποπληρώσουν m πελάτες το δάνειο τους θα δίνεται από την κατανομή $Bin(100, p)$ με $p = 0.1$, δηλαδή αν X ο αριθμός των πελατών που δεν αποπλήρωσαν τότε

$$P(X = m) = \binom{100}{m} (0.1)^m (0.9)^{100-m}$$

Κάθε πελάτης που δεν αποπληρώνει αντιστοιχεί σε ζημιά στην τράπεζα ίση προς 10000 ευρώ, άρα αν Y η τυχαία μεταβλητή που δίνει την ζημιά της τράπεζας θα έχουμε ότι $Y = 10000X$.

Η αναμενόμενη ζημιά θα είναι ίση προς

$$\mathbb{E}[Y] = \mathbb{E}[10000X] = 10000\mathbb{E}[X] = 10000 np = 10000 \times 100 \times 0.1 = 1000000$$

Το πόσο καλή είναι η $\mathbb{E}[Y]$ σαν πρόβλεψη της αναμενόμενης ζημιάς θα μας το δώσει η διασπορά,

$$\begin{aligned} Var(Y) &= Var(10000X) = (10000)^2 Var(X) = (10000)^2 np(1-p) \\ &= (10000)^2 100 \times 0.1 \times 0.9 \end{aligned}$$

Παράδειγμα

Στο παραπάνω παράδειγμα ποια θα είναι η πιθανότητα περισσότεροι από 95 πελάτες να μην αποπληρώσουν;

Ζητάμε την πιθανότητα

$$\begin{aligned} &P(X > 95) \\ &= P(X = 96) + P(X = 97) + P(X = 98) + P(X = 99) + P(X = 100) \\ &= \sum_{r=96}^{100} \binom{100}{r} (0.1)^r (0.9)^{100-r} \end{aligned}$$

Παράδειγμα (Το διωνυμικό μοντέλο)

Ένα τυπικό υπόδειγμα για την εξέλιξη της τιμής μια μετοχής είναι το διωνυμικό υπόδειγμα σύμφωνα με το οποίο η τιμή μιας μετοχής την χρονική στιγμή t δίνεται από την σχέση

$$S(t+1) = S(t)H(t),$$

$$H(t) \text{ i.i.d}$$

$$P(H(t) = u) = p, \quad P(H(t) = d) = 1 - p.$$

Παράδειγμα (Multinomial distribution)

Ας υποθέσουμε ότι ένα πείραμα δεν έχει μόνο 2 αποτελέσματα αλλά k διαφορετικά αποτελέσματα, o_i , $i = 1, \dots, k$. καθε ένα με πιθανότητα p_i .

Each experiment has k different outcomes o_k , each with probability p_i .

Ποιά είναι η πιθανότητα σε N ανεξαρτητα πειράματα να παρατηρήσουμε

n_1 times outcome o_1 ,

n_2 times outcome o_2 ,

...

n_k times outcome o_k ,

$n_1 + n_2 + \dots + n_k = N$.

What is the probability of the above event?

Ορίζουμε τις τυχαίες μεταβλητές - we define the random variables

$$Y_i = \begin{cases} 1 & \text{with probability } p_1 \\ 2 & \text{with probability } p_2 \\ \dots & \dots \\ k & \text{with probability } p_k \end{cases}$$

και

$$X_1 = \sum_{i=1}^N \mathbf{1}_{Y_i=1},$$

$$X_2 = \sum_{i=1}^N \mathbf{1}_{Y_i=2},$$

...

$$X_k = \sum_{i=1}^N \mathbf{1}_{Y_i=k},$$

όπου

$$\sum_{j=1}^k X_j = \sum_{j=1}^k \sum_{i=1}^N \mathbf{1}_{Y_i=j} = \sum_{i=1}^N \left(\sum_{j=1}^k \mathbf{1}_{Y_i=j} \right) = \sum_{i=1}^N 1 = N$$

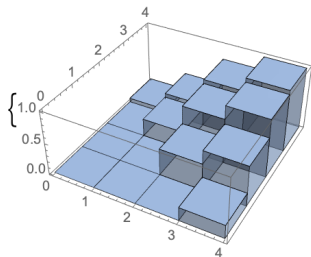
X_j counts the number of times that outcome o_j has arrived in N experiments

$$P(X_1 = n_1, \dots, X_k = n_k) = \frac{N!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k} = N! \prod_{i=1}^k \frac{p_i^{n_i}}{n_i!}$$
$$n_1 + \dots + n_k = N$$

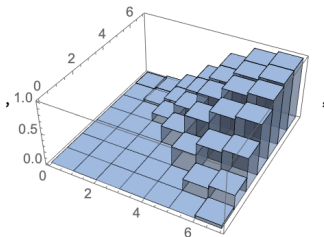
Η κατανομή αυτή μπορεί να περιγράψει π.χ. την περίπτωση που N μπάλες τοποθετούνται σε k κάλπες και η πιθανότητα να τοποθετηθεί μια μπάλα στην κάλπη i είναι p_i , $i = 1, \dots, k$.

This distribution models the situation where N balls are to be placed in k urns and the probability of a ball being placed in urn i is p_i .

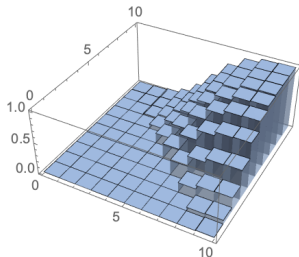
$n = 3$



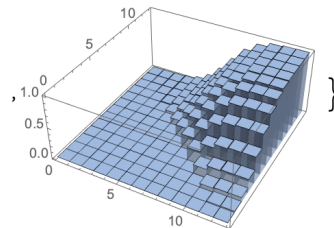
$n = 6$



$n = 9$



$n = 12$



Source: Wikicommons

Η υπεργεωμετρική κατανομή

Έχουμε μια συλλογή από n_1 αντικείμενα τύπου Α και n_2 αντικείμενα τύπου Β, και επιλέγουμε το ένα μετά το άλλο χωρίς επανάθεση n αντικείμενα.

Ποιά η πιθανότητα στο δείγμα των n αντικειμένων να περιέχονται m αντικείμενα τύπου Α;

Αν X η τυχαία μεταβλητή που μας δίνει το πλήθος των αντικειμένων τύπου Α στο δείγμα μεγέθους n θα λεμε ότι η X ακολουθεί την υπεργεωμετρική κατανομή με παραμέτρους n , n_1 , n_2 , και θα συμβολίζουμε $X \sim H(n, n_1, n_2)$.

Για να καθορίσουμε αυτή την κατανομή θα εργαστούμε ως εξής:

Το συνολικό πλήθος των τρόπων με τους οποίους μπορούμε να επιλέξουμε n αντικείμενα από $n_1 + n_2$ αντικείμενα είναι $\binom{n_1 + n_2}{n}$.

Με πόσους τρόπους θα μπορούσαμε να διαλέξουμε m αντικείμενα τύπου A ;

Για να έχουμε $X = m$ θα πρέπει m αντικείμενα να είναι τυπου A και $n - m$ τύπου B .

Αυτό μπορεί να γίνει με $\binom{n_1}{m} \binom{n_2}{n - m}$ τρόπους.

Συνεπώς

$$P(X = m) = \frac{\binom{n_1}{m} \binom{n_2}{n - m}}{\binom{n_1 + n_2}{n}}$$

Αν ρωτούσαμε την ίδια ερώτηση αλλά με επανάθεση τότε η τυχαία μεταβλητή X θα ακολουθεί την διωνυμική κατανομή.

$$X \sim H(n, n_1, n_2),$$

$$P(X = m) = \frac{\binom{n_1}{m} \binom{n_2}{n-m}}{\binom{n_1 + n_2}{n}}$$

Παράδειγμα (Υπεργεωμετρική κατανομή - Hypergeometric)

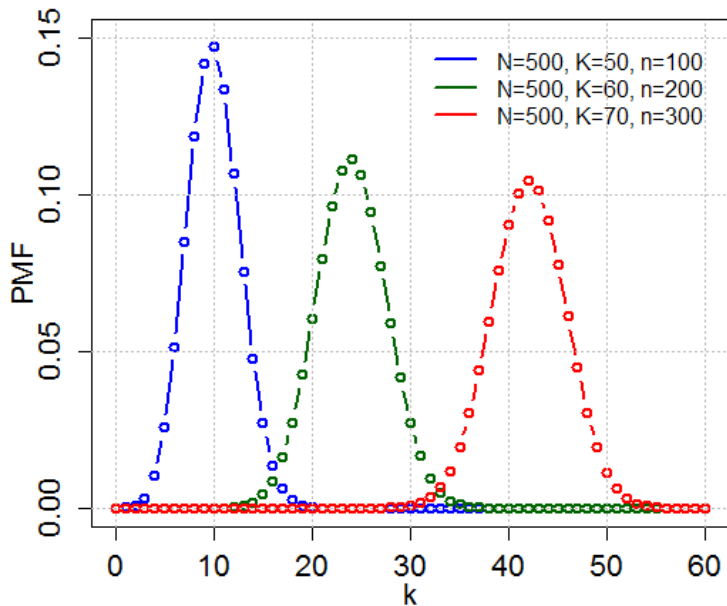
Ας υποθέσουμε μια κάλη που έχει N μπάλες εκ των οποίων οι K είναι κόκκινες.

Επιλέγουμε n μπάλες απο την καλη (χωρίς επανάθεση)

Ποια η πιθανότητα να επιλέξουμε k κόκκινες μπάλες σε αυτές;

$X = \# \text{ κόκκινες μπάλες, } X \sim \text{Hyp}(n, N, K),$

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$



Source: Wikicommons

Η μέση τιμή της $X \sim H(n, n_1, n_2)$ δίνεται από τον τύπο

$$\mathbb{E}[X] = \sum_{m=0}^n m P(X = m) = \sum_{m=0}^n m \frac{\binom{n_1}{m} \binom{n_2}{n-m}}{\binom{n_1+n_2}{n}} = \frac{nn_1}{n_1+n_2},$$

ενώ η διασπορά

$$\begin{aligned} \text{Var}(X) &= \sum_{m=0}^n \left(m - \frac{nn_1}{n_1+n_2} \right)^2 P(X = m) \\ &= \sum_{m=0}^n \left(m - \frac{nn_1}{n_1+n_2} \right)^2 \frac{\binom{n_1}{m} \binom{n_2}{n-m}}{\binom{n_1+n_2}{n}} \\ &= n \frac{n_1}{n_1+n_2} \frac{n_2}{n_1+n_2} \left(1 - \frac{n-1}{n_1+n_2-1} \right) \end{aligned}$$

Παράδειγμα

Μια τράπεζα για λόγους πολιτικής στα 1000 δάνεια που ενέκρινε το 2020 τα 10 είναι επισφαλής (κατηγορίας B και τα 990 όχι (κατηγορίας A).

Στον τυπικό έλεγχο απο την αρχη διαχειρισης η επιτροπή επιλέγει τυχαία 10 δάνεια απο το χαρτοφυλάκιο του 2020.

Ποιά η πιθανοτητα να είναι τα 5 επισφαλής και ποιά η καλύτερη εκτίμηση για το πλήθος των επισφαλών δανείων;

Η πιθανότητα αυτή δίνεται απο την υπεργεωμετρική κατανομή με παραμέτρους $n = 10$, $n_1 = 10$, $n_2 = 990$. 'ρα

$$P(X = 5) = \frac{\binom{10}{5} \binom{990}{5}}{\binom{1000}{10}}$$

Η αναμενόμενη τιμή της X είναι

$$\mathbb{E}[X] = \frac{10 \cdot 10}{10 + 990} = \frac{100}{1000} = 0.1.$$

Παράδειγμα (Κατανομές σε αριθμήσιμα σύνολα - Distributions on countable sets)

Στις περισσότερες περιπτώσεις πρακτικού ενδιαφέροντος ο δειγματικός χώρος δεν είναι πεπερασμένος.

Ένα τέτοιο παράδειγμα είναι όταν ο δειγματικός χώρος είναι αριθμήσιμο σύνολο π.χ. το $\mathbb{N}$.

Στην περίπτωση αυτή $\Omega = \{\omega_i : i \in \mathbb{N}\}$ και μπορούμε να ορίσουμε ένα μέτρο πιθανότητας στο Ω , χρησιμοποιώντας μια ακολουθία $p = \{p_i : i \in \mathbb{N}\}$ τέτοια ώστε $p_i \in (0, 1)$ και $\sum_{i=1}^{\infty} p_i = 1$.

Μετά για οποιοδήποτε $A \subset \Omega$, το οποίο και θα είναι αναγκαστικά της μορφής $A = \{\omega_j : j \in J \subset \mathbb{N}\}$ θα έχουμε $P(A) = \sum_{j \in J} p_j$.

Πολύ εύκολα μπορούμε να δούμε ότι ένα αντικείμενο που ορίστηκε με αυτό τον τρόπο είναι μετρο.

Εναλλακτικά μπορεί να θεωρήσουμε μια τυχαία μεταβλητή $X : \Omega' \rightarrow \mathbb{N}$, όπου $(\Omega', \mathcal{F}', P')$ ένας αυθαίρετος χώρος πιθανότητας και να ορίσουμε το επαγόμενο μετρο επάνω στον $\mathbb{N}$ από την τυχαία μεταβλητή δηλ. το $P(A) = P'(X^{-1}(A))$ για κάθε $A \subset \mathbb{N}$, το οποίο και είναι μετρο πιθανότητας

Η γεωμετρική κατανομή

Ας υποθέσουμε ότι επαναλαμβάνουμε ανεξαρτητα πειράματα τα οποία έχουν πιθανότητα επιτυχίας p και αποτυχίας $1 - p$.

Πόσες φορές θα πρέπει να περιμένουμε μέχρις ότου εμφανιστεί η πρώτη επιτυχία;

Ας ονομάσουμε X την τυχαία μεταβλητή η οποία μας δίνει τον αριθμό των επαναλήψεων μέχρι να εμφανιστεί η πρώτη επιτυχία.

Η κατανομή που ακολουθεί η X ονομάζεται **γεωμετρική κατανομή**.

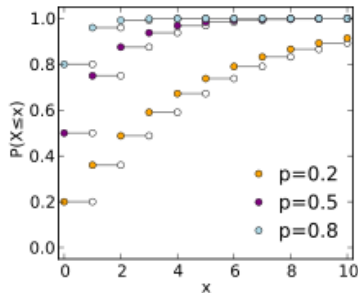
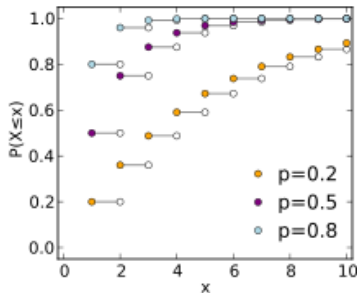
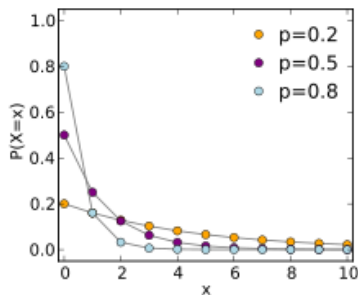
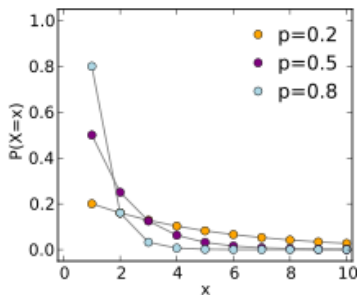
Για να πάρει η X την τιμή $X = m$ θα πρέπει να έχει εμφανιστεί το γεγονός

$$\underbrace{AA \cdots A}_{(m-1)\text{-φορές}} E$$

το οποίο θα εμφανιστεί με πιθανότητα $(1-p)^{m-1}p$.

Άρα $P(X = m) = (1-p)^{m-1}p$. Θα συμβολίζουμε ότι η X ακολουθεί την γεωμετρική κατανομή με πιθανότητα p , με το $X \sim G(p)$.

$$X \sim G(p), \quad P(X = m) = (1-p)^{1-m}p$$



Source: Wikicommons

Αν

$$X \sim \text{Geom}(p) \implies P(X > n) = (1 - p)^n.$$

Επίσης

$$Y_0, Y_1, \dots \sim_{i.i.d.} \text{Geom}(p) \implies X := \min\{k : \sum_{i=0}^k Y_i > n\} \sim \text{Bin}(n, p).$$

Ας υποθέσουμε ότι $X \sim G(p)$.

Η αναμενόμενη τιμή της X είναι

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} n(1-p)^{n-1}p = p \sum_{n=1}^{\infty} n(1-p)^{n-1} = \frac{1}{p},$$

η δε διασπορά είναι

$$\begin{aligned} \text{Var}(X) &= \sum_{n=1}^{\infty} \left(n - \frac{1}{p}\right)^2 (1-p)^{n-1}p = p \sum_{n=1}^{\infty} \left(n - \frac{1}{p}\right)^2 (1-p)^{n-1} \\ &= \frac{1-p}{p^2}. \end{aligned}$$

Παράδειγμα

Μια μετοχή σε κάθε περίοδο μπορεί να έχει ανοδικές αποδόσεις με πιθανότητα p και καθοδικές αποδόσεις με πιθανότητα $1 - p$.

Ένας επενδυτής ακολουθεί την στρατηγική να μην αγοράσει την μετοχή μέχρις ότου η μετοχή φέρει την πρώτη ανοδική απόδοση.

Ποιός είναι ο αναμενόμενος χρόνος που θα χρειαστεί να περιμένει ο επενδυτής μέχρι να αγοράσει την μετοχή

(α) αν $p = 1/2$ και (β) αν $p = 1/4$;

Είναι το ίδιο σαν να έχουμε επαναλαμβανόμενα πειράματα με πιθανότητα επιτυχίας p και αποτυχίας $1 - p$, και να ζητάμε τον αναμενόμενο αριθμό επαναλήψεων των πειραμάτων μέχρι να εμφανιστεί η πρώτη επιτυχία.

(α) Η τυχαία μεταβλητή που μας δίνει τον αριθμο των επαναλήψεων μεχρι την εμφάνιση της πρώτης επιτυχίας θα ακολουθεί την γεωμετρική κατανομή $G(\frac{1}{2})$ οπότε $\mathbb{E}[X] = \frac{1}{1/2} = 2$ και $\text{Var}(X) = \frac{1-1/2}{(1/2)^2} = \frac{1/2}{1/4} = 2$.

(β) Η τυχαία μεταβλητή που μας δίνει τον αριθμο των επαναλήψεων μεχρι την εμφάνιση της πρώτης επιτυχίας θα ακολουθεί την γεωμετρική κατανομή $G(\frac{1}{4})$ οπότε $\mathbb{E}[X] = \frac{1}{1/4} = 4$ και $\text{Var}(X) = \frac{1-1/4}{(1/4)^2} = \frac{3/4}{1/16} = 12$.

Παράδειγμα

Αν $X \sim G(p)$ υπολογίστε την $P(X > n)$.

Έχουμε ότι

$$\begin{aligned} P(X > n) &= 1 - P(X \leq n) = 1 - \sum_{r=1}^n (1-p)^{r-1} p = 1 - p \sum_{r=1}^n (1-p)^{r-1} \\ &= 1 - p \frac{1 - (1-p)^n}{1 - (1-p)} = (1-p)^n \end{aligned}$$

Πρόταση (Η ιδιότητα έλλειψης μνήμης)

Ας υποθέσουμε ότι η τυχαία μεταβλητή X ακολουθεί την γεωμετρική κατανομή $G(p)$ για κάποιο p . Τότε

$$P(X > n_1 + n_2) = P(X > n_1)P(X > n_2),$$

ή ισοδύναμα

$$P(X > n_1 + n_2 | X > n_1) = P(X > n_2).$$

Απόδειξη.

Έχουμε ότι

$$P(X > n_1 + n_2) = 1 - P(X \leq n_1 + n_2) = (1 - p)^{n_1 + n_2} = (1 - p)^{n_1} (1 - p)^{n_2}$$

ή ισοδύναμα

$$P(X > n_1 + n_2 | X > n_1) = \frac{P(X > n_1 + n_2)}{P(X > n_1)} = P(X > n_2).$$



Παράδειγμα

Το διωνυμικό μοντέλο για την εξέλιξη της τιμής μιας μετοχής υποθέτει ότι σε κάθε χρονική περίοδο μπορούμε να έχουμε ανοδικές ή καθοδικές αποδόσεις με πιθανότητα p και $1 - p$ αντίστοιχα.

Οι αποδόσεις κάθε χρονική στιγμή είναι ανεξάρτητες.

Η επενδύτική στρατηγική ενός hedge fund είναι να μην τοποθετηθεί στην αγορά μέχρι να παρατηρηθεί η πρώτη καθοδική απόδοση για την μετοχή.

Δεδομένου το hedge fund δεν έχει τοποθετηθεί μέχρι την χρονική στιγμή t ποια η πιθανότητα να μην έχει τοποθετηθεί μέχρι την χρονική στιγμή $T > t$;

Είναι το ίδιο σαν να μοντελοποιούμε ανεξάρτητα επαναλαμβανόμενα πειράματα με πιθανότητα επιτυχίας (καθοδική απόδοση) p και πιθανότητα αποτυχίας (ανοδική απόδοση) $1 - p$.

X είναι η τυχαία μεταβλητή που μας δίνει τον χρόνο που θα κάνει το hedge fund να τοποθετηθεί στην αγορά και

$$X \sim \text{Geo}(p).$$

Ζητάμε την δεσμευμένη πιθανότητα

$$\begin{aligned} P(X > T \mid X > t) &= P(X > t + (T - t) \mid X > t) = P(X > T - t) \\ &= (1 - p)^{T-t}. \end{aligned}$$

Η αρνητική διωνυμική

Ας υποθέσουμε τώρα ότι επαναλαμβάνουμε πειράματα Bernoulli με πιθανότητα επιτυχίας p και μας ενδιαφέρει ο αριθμός των δοκιμών μέχρι την εμφάνιση της m επιτυχίας.

Έστω X η τυχαία μεταβλητή που μας δίνει τον αριθμό των επαναλήψεων μέχρι την εμφάνιση της m επιτυχίας.

Η κατανομή που ακολουθεί η X ονομάζεται αρνητική διωνυμική $Nb(m, p)$, και θα συμβολίζουμε $X \sim Nb(m, p)$.

Ας δούμε τα γεγονότα τα οποία αντιστοιχούν στο $X = n$.

Για να έχουμε $X = n$ σημαίνει ότι στην δοκιμή n έχουμε επιτυχία και στις δοκιμές 1 έως $n - 1$ έχουμε ακριβώς $m - 1$ επιτυχίες.

Το γεγονός αυτό μπορούμε να το γράψουμε σαν την τομή $A \cap B$ όπου

$$A = \{\text{επιτυχία στην δοκιμή } n\},$$

$$B = \{\text{ακριβώς } m - 1 \text{ επιτυχίες στις πρώτες } n - 1 \text{ δοκιμές}\}$$

Η πιθανότητα του γεγονότος A είναι $P(A) = p$. Η πιθανότητα του γεγονότος B δίνεται από την διωνυμική κατανομή με παραμέτρους p και $n - 1$, δηλαδή $P(B) = \binom{n-1}{m-1} p^{m-1} (1-p)^{n-m}$.

$$P(B) = \binom{n-1}{m-1} p^{m-1} (1-p)^{n-m}.$$

Λόγω της ανεξαρτησίας των δοκιμών

$$\begin{aligned} P(X = n) &= P(A \cap B) = P(A)P(B) = p \binom{n-1}{m-1} p^{m-1} (1-p)^{n-m} \\ &= \binom{n-1}{r-1} p^m (1-p)^{n-m}. \end{aligned}$$

Κατά συνέπεια, αν $X \sim Nb(m, p)$ τότε

$$P(X = n) = \binom{n-1}{m-1} p^m (1-p)^{n-m}.$$

Μπορούμε να υπολογίσουμε την αναμενόμενη τιμή της X ,

$$\mathbb{E}[X] = \sum_{r=m}^{\infty} r P(X = r) = \sum_{r=m}^{\infty} r \binom{n-1}{r-1} p^r (1-p)^{n-r} = \frac{m}{p}$$

και την διασπορά

$$\begin{aligned} \text{Var}(X) &= \sum_{r=m}^{\infty} \left(r - \frac{m}{p}\right)^2 P(X = r) \\ &= \sum_{r=m}^{\infty} \left(r - \frac{m}{p}\right)^2 \binom{n-1}{r-1} p^r (1-p)^{n-r} = \frac{m(1-p)}{p^2}. \end{aligned}$$

Ενας εναλλακτικός τρόπος να περιγράψουμε την αρνητική διωνυμική κατανομή είναι σαν την κατανομή που ακολουθεί το αθροισμα m ανεξάρτητων τυχαίων μεταβλητών X_i που ακολουθούν την γεωμετρική κατανομή $Geo(p)$.

Άρα, αν $X_i \sim G(p)$ τότε $X = X_1 + \dots + X_m \sim Nb(m, p)$.

Απο αυτή την αναπαράσταση μπορούμε να πάρουμε εύκολα τους τύπους για την μέση τιμή και την διασπορά.

Επίσης

$$X_k \sim NegBin(r_k, p), \text{ ανεξάρτητες} \implies \sum_{k=1}^n X_k \sim NegBin\left(\sum_{k=1}^n r_k, p\right)$$

Παράδειγμα

Ένα *hedge fund* ακολουθεί την επενδυτική στρατηγική να μην υίοθετει μια θέση αγοράς σε μια μετοχή αν αυτή δεν έχει σημειώσει m ανοδικές κινήσεις στις αποδόσεις της.

Αν υποθέσουμε ότι η εξέλιξη της τιμής της μετοχής περιγράφεται από το διωνυμικό μοντέλο με πιθανότητα p ανοδικών αποδόσεων, ποια θα είναι η πιθανότητα το *hedge fund* να τοποθετηθεί στην μετοχή αυτή την χρονική περίοδο T ;

Η τυχαία μεταβλητή X που μετράει το πλήθος των χρονικών περιόδων που πρέπει να περάσουν πριν σημειωθούν m ανοδικές κινήσεις στην αγορά ακολουθεί την αρνητική διωνυμική κατανομή

$$X \sim \text{NegBin}(m, p).$$

Η πιθανότητα το τοποθετηθεί το hedge fund στην μετοχή την χρονική περίοδο T είναι

$$P(X = T - 1) = \binom{(T - 1) - 1}{m - 1} p^m (1 - p)^{(T - 1) - m}.$$

Παράδειγμα

Ένας καθηγητής μαθηματικών έχει σε κάθε μία απο τις 2 τσέπες του σακακιού του απο ένα πακέτο με αντισηπτικά μαντηλάκια το καθένα απο τα οποία περιέχει N μαντήλια.

Κάθε φορά που χρειάζεται να απολυμανει τα χέρια του επιλέγει τυχαία (με πιθανότητα $1/2$) ένα απο τα πακέτα αυτά και παίρνει ένα μαντήλι χωρίς να κοιτάει το πακέτο.

Ποια είναι η πιθανοτητα την πρώτη φορά που ο μαθηματικός ανακαλύπτει ότι το ένα κουτί είναι κενό το άλλο κουτί να περιέχει η μαντηλια;

Ας υποθέσουμε ότι ο καθηγητής ανακαλύπτει πρώτα ότι έχει αδειάσει το πακέτο στην αριστερή του τσέπη και έστω ότι την στιγμή αυτή το πακέτο στην δεξιά του τσέπη έχει n μαντηλία.

Αυτό σημαίνει ότι έχει επιλέξει το πακέτο στην αριστερή του τσέπη $N + 1$ φορές και το πακέτο στην δεξιά του τσέπη $N - n$ φορές.

Είναι λοιπόν σαν να έχει κάνει συνολικά $N + 1 + (N - n) = 2N + 1 - n$ ανεξάρτητες δοκιμες Bernoulli (επιλογή αριστερής ή δεξιάς τσέπης).

Για να έχει n μαντήλια το πακέτο στην δεξιά του τσέπη όταν ανακαλύψει ότι το πακέτο στην αριστερή έχει αδειάσει θα πρέπει η $N + 1$ επιλογή της αριστερής τσέπης να γίνει ακριβώς στην $2N + 1 - n$ επανάληψη των δοκιμών.

Αν ονομασούμε επιτυχία την επιλογή της αριστερής τσέπης ζητάμε την πιθανότητα να έχουμε $N + 1$ επιτυχίες σε $2N + 1 - n$ δοκιμές η οποία δίνεται απο την αρνητική διωνυμική κατανομή ως

$$\begin{aligned} P(X = N + 1) &= \binom{2N + 1 - n - 1}{N + 1 - 1} \left(\frac{1}{2}\right)^{N+1} \left(\frac{1}{2}\right)^{(2N+1-n)-(N+1)} \\ &= \frac{1}{2} \binom{2N + 1 - n - 1}{N + 1 - 1} \left(\frac{1}{2}\right)^{2N-n} \end{aligned}$$

Το πρόβλημα θα ήταν εντελώς συμμετρικό αν ανακαλύψει οτι πρώτα έχει αδειάσει το πακέτο στην δεξιά του τσέπη και αυτη την στιγμή το πακετο στην αριστερή τσέπη έχει n μαντήλια.

Συνεπώς η πιθανότητα που ζητάμε είναι

$$P = \binom{2N + 1 - n - 1}{N + 1 - 1} \left(\frac{1}{2}\right)^{2N-n}$$

Η κατανομή *Poisson*

Η τυχαία μεταβλητή X θα λέμε ότι ακολουθεί την κατανομή *Poisson* με παράμετρο $\lambda > 0$ αν

$$P(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}$$

Συμβολίζουμε $X \sim \text{Pois}(\lambda)$.

Η κατανομή *Poisson* είναι ένα καλό μοντέλο για την καταμετρηση πλήθους γεγονότων π.χ. αιτήσεις αποζημίωσης σε μια ασφαλιστική εταιρεία, πλήθος κλήσεων σε ένα τηλεφωνικό κέντρο, κυβερνοεπιθέσεις σε ένα πληροφοριακό σύστημα, κρούσματα κάποιας επιδημίας κλπ.

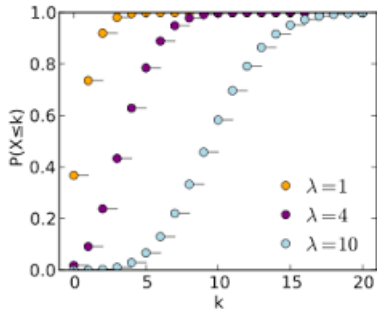
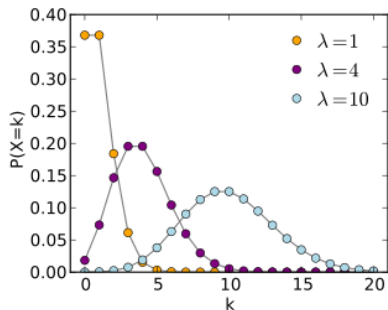
Η κατανομή Poisson μοντελοποιεί το αριθμο σπανίων γεγονότων που μπορεί να συμβούν σε μια χρονική περίοδο.

Αυτό μπορεί κανείς να το δει απο το γεγονός ότι

$$\binom{n}{m} p^m (1-p)^{n-m} \simeq e^{-np} \frac{(np)^m}{m!}, \quad m = 0, 1, 2, \dots,$$

αρκεί $\lambda := np$ να είναι αρκετά μικρό.

Άρα αν ο αριθμός των επαναλήψεων n είναι μεγάλος και η πιθανότητα επιτυχίας p μικρή, αλλά κατά τρόπο ώστε το γινόμενο $\lambda := np$ να είναι μικρό η διωνυμική κατανομή $Bin(n, p)$ προσεγγίζεται απο την $Pois(np)$.



Source: Wikicommons

$$X_1 \sim Poi(\lambda_1), \dots, X_n \sim Poi(\lambda_n) \text{ independent} \implies \sum_{i=1}^n X_i \sim Poi\left(\sum_{i=1}^n \lambda_i\right).$$

Η κατανομή Poisson μπορεί να προσεγγίσει πολλές διακριτές κατανομές στο κατάλληλο όριο

$$X_n \sim Bin\left(n, \frac{\lambda}{n}\right) \implies X_n \xrightarrow{d} Y \sim Poi(\lambda) \text{ as } n \rightarrow \infty.$$

$$X_n \sim NegBin\left(n, 1 - \frac{\lambda}{n}\right) \implies X_n \xrightarrow{d} Y \sim Poi(\lambda) \text{ as } n \rightarrow \infty.$$

Η μεση τιμή της X δίνεται απο τον τύπο

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} nP(X = n) = \sum_{n=0}^{\infty} ne^{-\lambda} \frac{\lambda^n}{n!} = \lambda,$$

και η διασπορά

$$\text{Var}(X) = \sum_{n=0}^{\infty} (n - \lambda)^2 P(X = n) = \sum_{n=0}^{\infty} (n - \lambda)^2 e^{-\lambda} \frac{\lambda^n}{n!} = \lambda$$

Παράδειγμα

Μία ασφαλιστική εταιρεία δραστηριοποιείται σε 4 διαφορετικούς κλάδους ασφαλειών και τα χαρτοφυλάκια της είναι ανεξάρτητα.

Αν το πλήθος των αιτήσεων για αποζημίωση X_i από τον κλάδο i περιγράφεται από την κατανομή Poisson $X_i \sim \text{Poi}(\lambda_i)$ βρείτε τον αναμενόμενο συνολικό αριθμο αποζημιώσεων που θα αντιμετωπίσει η εταιρεία και από τους 4 κλάδους;

Μπορείτε να εκτιμήσετε την πιθανότητα η απόκλιση του πραγματοποιούμενου πλήθους αιτήσεων $X(\omega)$ από την εκτίμηση που δίνει η αναμενόμενη τιμή να είναι μεγαλύτερη από ϵ ;

Οι X_i , $i = 1, \dots, 4$ είναι ανεξάρτητες οπότε

$$X = X_1 + \dots + X_4 \sim Poi(\lambda_1 + \dots + \lambda_4).$$

Το αναμενόμενο πλήθος των συνολικών αποζημιώσεων είναι

$$\mathbb{E}[X] = \lambda_1 + \dots + \lambda_4.$$

Η πιθανότητα που μας ζητείται είναι η

$$P(|X(\omega) - (\lambda_1 + \dots + \lambda_4)| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2} \leq \frac{(\lambda_1 + \dots + \lambda_4)}{\epsilon^2}$$

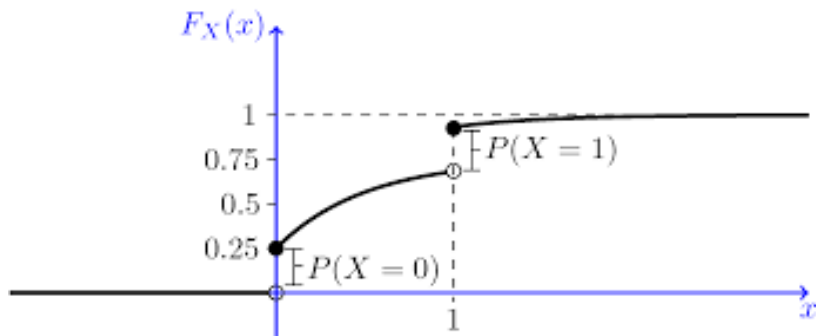
Συνεχείς κατανομές

Ένα μετρο πιθανότητας P επάνω στο $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ μπορεί να χαρακτηριστεί απο μια συνάρτηση $F : \mathbb{R} \rightarrow [0, 1]$ η οποία ονομάζεται συναρτηση κατανομής και ορίζεται ως

$$F(x) := P((-\infty, x]), \quad \forall x \in \mathbb{R}.$$

Η συνάρτηση κατανομής χαρακτηρίζει το μετρο πιθανοτητας. Η συναρτηση αυτή απο κατασκευής είναι

- ① αύξουσα - **increasing**
- ② συνεχής απο τα δεξιά - **continuous to the right**
- ③ $\lim_{x \rightarrow -\infty} F(x) = 0$, και $\lim_{x \rightarrow \infty} F(x) = 1$.



Source: Wikicommons

Θεώρημα

Η $F : \mathbb{R} \rightarrow [0, 1]$ είναι μια συνάρτηση κατανομής ενός (μοναδικού) μέτρου πιθανότητας στο $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ αν και μόνο αν ικανοποιεί τις ιδιότητες 1,2,3 παραπάνω.

Ένα ενδιαφέρον ερώτημα είναι πως θα μετασχηματιστεί η συναρτηση κατανομής μιας τυχαίας μεταβλητής $X : \Omega \rightarrow \mathbb{R}$ αν την συνθέσουμε με μια κατάλληλη συνάρτηση $\varphi : \mathbb{R} \rightarrow \mathbb{R}$.

Η επόμενη πρόταση είναι πολύ ενδιαφέρουσα τόσο απο αναλυτικής όσο και απο υπολογιστικής άποψης.

Πρόταση

Έστω $X : \Omega \rightarrow \mathbb{R}$ μια τυχαία μεταβλητή (*random variable*) και $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ μια κατάλληλη συνάρτηση. Ας θεωρήσουμε την τυχαία μεταβλητή $Y := \varphi(X)$, $Y : \Omega \rightarrow \mathbb{R}$.

- ❶ Αν F_X, F_Y είναι οι συναρτήσεις κατανομής των X, Y αντίστοιχα ισχύει ότι

$$F_Y(y) = F_X(\varphi^{-1}(y)).$$

- ❷ Αν επιπλέον η F_X έχει πυκνότητα πιθανότητας και η φ είναι συνεχώς διαφορίσιμη με μη μηδενιζόμενη παράγωγο τότε έχει πυκνότητα και Y και μάλιστα

$$f_Y(y) = f_X(\varphi^{-1}(y))|(\varphi^{-1})'(y)|.$$

if φ is continuously differentiable with non vanishing derivative that Y also has a density

- ❸ Αν η φ είναι κατα τμήματα μονότονη, δηλαδή υπάρχουν διαστήματα I_i , $i = 1, \dots, m$ τα οποία διαμερίζουν το $\mathbb{R}$ και σε καθένα από τα οποία η φ είναι μονότονη, δηλαδή έχει αντίστροφο $h_i = \varphi_i^{-1}$, τότε

$$f_Y(y) = \sum_{i=1}^m f_X(\varphi_i^{-1}(y))|(\varphi_i^{-1})'(y)|\mathbf{1}_{Rg(\varphi)}(y),$$

όπου $Rg(\varphi)$ είναι το πεδίο ορισμού (*range*) της φ .

Απόδειξη.

1. $P(Y \leq y) = P(\varphi(X) \leq y) = P(X \leq \varphi^{-1}(y))$ και το αποτέλεσμα προκύπτει απο εκεί.
2. Προκύπτει με παραγωγή χρησιμοποιώντας τον κανόνα για την παραγωγή της σύνθετης συνάρτησης και της αντίστροφης συνάρτησης.



Αν γνωρίσουμε την συνάρτηση κατανομής μπορούμε να συναγουμε πολλά ενδιαφέροντα συμπεράσματα για κάποιο δείγμα

Παράδειγμα (Order statistics)

Έστω

$$X_1, X_2, \dots, X_n \sim F, \text{ independent}$$

Τότε

$$X_{(1)} = \min\{X_1, \dots, X_n\} \sim F_{(1)} := 1 - (1 - F)^n.$$

Εχουμε ότι

$$\{X_{(1)} > x\} = \bigcap_{i=1}^n \{X_i > x_i\}$$

οπότε

$$\begin{aligned} F_{(1)}(x) &= P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x) = 1 - P\left(\bigcap_{i=1}^n \{X_i > x_i\}\right) \\ &= 1 - (1 - F(x))^n \end{aligned}$$

Παράδειγμα (Order statistics)

Έστω

$$X_1, X_2, \dots, X_n \sim F, \text{ independent}$$

Τότε

$$X_{(n)} = \max\{X_1, \dots, X_n\} \sim F_{(n)} := (F)^n.$$

Εχουμε ότι

$$\{X_{(n)} \leq x\} = \bigcap_{i=1}^n \{X_i \leq x_i\},$$

οπότε

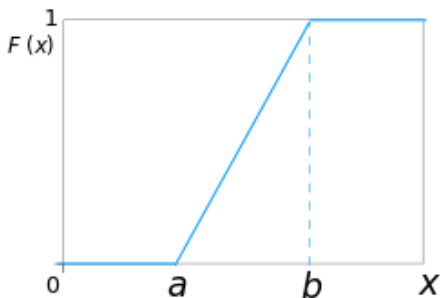
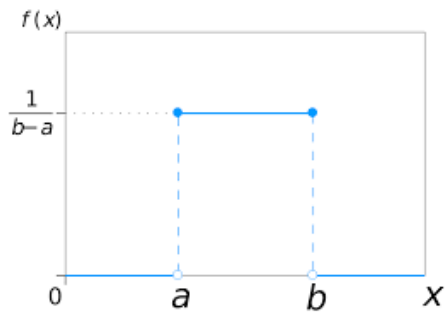
$$F_{(n)}(x) = P(X_{(n)} \leq x) = P\left(\bigcap_{i=1}^n \{X_i \leq x_i\}\right) = (F(x))^n$$

Παράδειγμα (Η ομοιόμορφη κατανομή - Uniform distribution)

Η ομοιόμορφη κατανομή αντιστοιχεί σε ένα μέτρο πιθανότητας στο $[a, b] \subset \mathbb{R}$, και έχει πυκνότητα

$$f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}, \text{ density function of uniform distribution.}$$

$$P([a, x]) = F(x) = \int_a^x f(x) dx = \frac{1}{b-a}(x-a), \text{ distribution function}$$



Source: Wikicommons

Η μέση τιμή της ομοιόμορφης κατανομής στο $[0, 1]$ δίνεται ως

$$\mathbb{E}[X] = \int_0^1 x dx = \frac{1}{2},$$

ενώ η διασπορά ως

$$\text{Var}(X) = \int_0^1 \left(x - \frac{1}{2}\right)^2 dx = \frac{1}{12}.$$

Η ομοιόμορφη κατανομή είναι πολύ σημαντική γιατί ξεκινώντας από αυτή μπορούμε να παράγουμε οποιαδήποτε άλλη κατανομή, διακριτή ή συνεχή: για την ακρίβεια έχοντας ένα δείγμα $X_1, \dots, X_n$ από την ομοιόμορφη κατανομή μπορούμε να κατασκευάσουμε ένα δείγμα $Y_1, \dots, Y_n$ από οποιαδήποτε κατανομή με συνάρτηση κατανομής F .

Αυτό μπορούμε να το δούμε ως εξής

$$F(y) = P(X \leq F(y)) = P(F^{-1}(X) \leq y),$$

επίσης

$$P(Y \leq y) = F(y),$$

συνεπώς συγκρίνοντας τις δυο παραπάνω σχέσεις καταλήγουμε ότι αν $X \sim U([0, 1])$ τότε $Y = F^{-1}(X) \sim F$.

Η απλή αυτή παρατήρηση μας οδηγεί σε μια πολύ βασική μέθοδο στην θεωρία της προσομοίωσης, την μέθοδο της αντιστροφής.

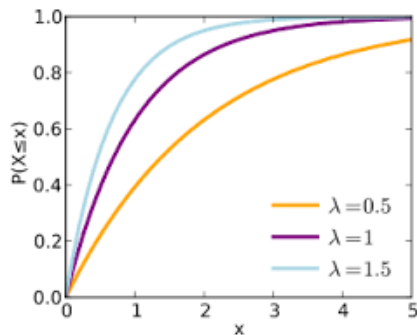
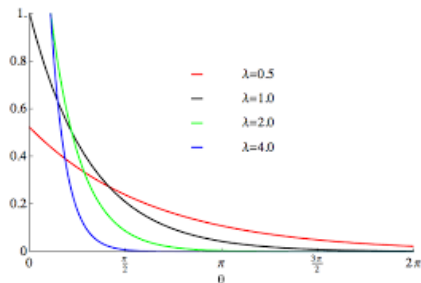
Παράδειγμα (Η εκθετική κατανομή - Exponential distribution)

Η εκθετική κατανομή αντιστοιχεί σε ένα μέτρο πιθανότητας στο $\mathbb{R}_+ = [0, \infty) \subset \mathbb{R}$ και έχει πυκνότητα

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0, \infty)}, \quad \lambda > 0$$

$$F(x) = P([0, x]) = \int_0^x f(x) dx = 1 - e^{-\lambda x},$$

Η κατανομή αυτή χρησιμοποιείται για να μοντελοποιούμε τυχαίους χρόνους, π.χ. τον χρόνο αναμονής σε μια ουρά, ή μεταξύ της εμφάνισης διαφόρων συμβάντων.



Source: Wikicommons

Σύνδεση εκθετικής και Poisson

$$\{Y_i\} \sim_{i.i.d} \text{Exp}(\lambda) \implies X = \max\{n : \sum_{j=1}^n Y_j \leq 1\} \sim \text{Poi}(\lambda)$$

Η αναμενόμενη τιμή της X αν $X \sim \text{Exp}(\lambda)$ δίνεται απο την σχέση

$$\mathbb{E}[X] = \int_0^{\infty} \lambda e^{-\lambda x} x dx = \frac{1}{\lambda},$$

ενώ η διασπορά δίνεται απο την σχέση

$$\text{Var}(X) = \int_0^{\infty} (x - \frac{1}{\lambda})^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2}.$$

Οι σχέσεις αυτές μπορεί να επαληθευθούν χρησιμοποιώντας ολοκλήρωση κατά παράγοντες (άσκηση!).

Η εκθετική κατανομή έχει την ιδιότητα της απώλειας μνήμης όπως και η γεωμετρική.

Πρόταση

Ας υποθέσουμε ότι $X \sim \text{Exp}(\lambda)$. Τότε,

$$P(X > t_1 + t_2 \mid X > t_1) = P(X > t_2)$$

Απόδειξη.

Έχουμε ότι

$$\begin{aligned}P(X > t_1 + t_2) &= 1 - P(X \leq t_1 + t_2) = 1 - F(t_1 + t_2) \\&= e^{-\lambda(t_1+t_2)} = e^{-\lambda t_1} e^{-\lambda t_2} \\&= (1 - P(X \leq t_1))(1 - P(X \leq t_2)) = P(X > t_1)P(X > t_2).\end{aligned}$$

Ας ορίσουμε τώρα τα γεγονότα

$$A = \{X > t_1 + t_2\},$$

$$B = \{X > t_1\},$$

και απο τον ορισμο της δεσμευμένης πιθανότητας,

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{P(X > t_1)P(X > t_2)}{P(X > t_1)} = P(X > t_2).$$

Χρησιμοποιήσαμε την παρατήρηση $A \subset B$ οπότε $A \cap B = A$. □

Παράδειγμα

Ας υποθέσουμε ότι ο αριθμός των τυχαίων συμβάντων (π.χ. κλήσεις σε ένα τηλεφωνικό κέντρο, ατυχήματα, επιθέσεις σε κάποιο server, αιτήσεις για αποζημίωση σε μια ασφαλιστική εταιρεία κλπ) δίνεται από την κατανομή *Poisson* με τον ακόλουθο τρόπο: Αν $N(t)$ είναι ο αριθμός των συμβάντων στο χρονικό διάστημα $[0, t]$ τότε

$$P(N(t) = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n = 0, 1, 2, \dots$$

Ποια θα είναι η κατανομή του χρόνου αναμονής μεταξύ 2 διαδοχικών συμβάντων;

Ας ονομάσουμε T_1 το (τυχαίο) χρόνο στον οποίο πραγματοποιήθηκε το πρώτο συμβάν, T_2 τον τυχαίο χρόνο στον οποίο πραγματοποιήθηκε το δεύτερο συμβάν, κλπ.

Αν $T_1 \geq t$ τότε $N(t) = 0$. Αν $T_1 < t$ και $T_2 \geq t$ τότε $N(t) = 1$. Αν $T_1 < T_2 < t$ και $T_3 > t$ τότε $N(t) = 2$ κ.ο.κ.

Ο χρόνος $\tau_1 := T_1 - 0$ είναι ο χρόνος αναμονής για το πρώτο συμβάν.
Η πιθανότητα

$$P(\tau_1 > t) = P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t} \frac{(\lambda t)^0}{0!} = e^{-\lambda t},$$

άρα

$$P(\tau_1 \leq t) = 1 - P(\tau_1 > t) = 1 - e^{-\lambda t},$$

συνεπώς $\tau_1 \sim \text{Exp}(\lambda t)$.

Μπορεί κανείς να δείξει ότι ο χρόνος αναμονής για διαδοχικά συμβάντα ακολουθεί γενικά την κατανομή $\text{Exp}(\lambda t)$

Παράδειγμα

Ο χρόνος μεταξύ διαδοχικών αιτήσεων αποζημίωσης σε μια ασφαλιστική εταιρεία ακολουθεί την εκθετική κατανομή με μέση τιμή 0.5 min .

Ποιά η πιθανότητα ο χρόνος μεταξύ δύο διαδοχικών αιτήσεων αποζημίωσης να ξεπεράσει τα 2 λεπτά;

Αν εδώ και 2 λεπτά δεν έχει έρθει αιτηση ποια θα είναι η πιθανότητα να μην έρθει αιτηση ούτε στα επόμενα 2;

Ας πάρουμε σαν μονάδα μετρησης του χρόνου το λεπτό.

Η πληροφορία ότι η μεση αναμονή μεταξύ των αιτήσεων είναι 0.5 min μας λέει ότι η παράμετρος της κατανομής θα είναι $\frac{1}{\lambda} = 0.5 \text{ min}$ δηλαδή $\lambda = 2$ και η μονάδα του λ θα είναι min^{-1} .

Συνεπώς η συνάρτηση κατανομής θα είναι η $F(x) = 1 - e^{-2x}$ όπου το μετράμε σε λεπτά.

Η πιθανότητα λοιπόν να ξεπεράσει ο χρόνος μεταξύ 2 διαδοχικών αιτησεων τα 2 λεπτά θα είναι

$$P(T > 2 \text{ min}) = 1 - P(T \leq 2 \text{ min}) = 1 - F(2) = e^{-4}.$$

Για το δεύτερο ερώτημα ζητάμε την πιθανότητα

$$P(T > 2 + 2 \mid T > 2) = P(T > 2) = e^{-4},$$

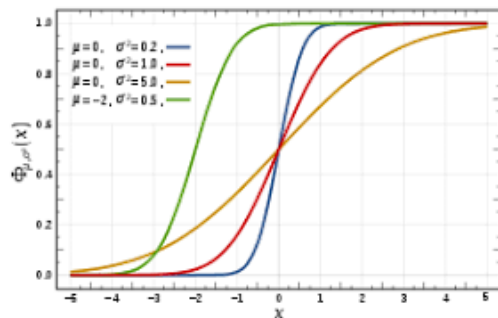
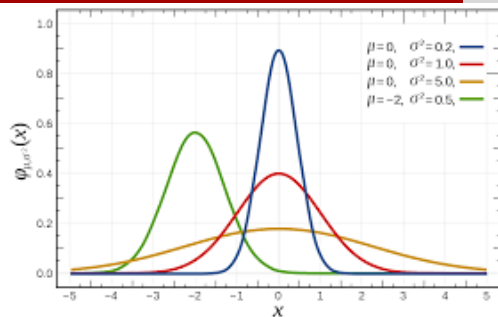
όπου χρησιμοποιήσαμε την ιδιότητα έλλειψης μνήμης.

Παράδειγμα (Η κανονική κατανομή - Normal distribution)

Η κανονική κατανομή αντιστοιχεί σε ένα μετρο πιθανότητας στο $\mathbb{R}$ και έχει πυκνότητα

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \mu \in \mathbb{R}, \quad \sigma > 0.$$

$$P((-\infty, x]) = F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz,$$



Source: Wikicommons

$$X_1, \dots, X_n, \text{ independent } X_k \sim N(\mu_k, \sigma_k^2) \implies \\ a + \sum_{k=1}^n b_k X_k \sim N\left(a + \sum_{k=1}^n b_k \mu_k, \sum_{k=1}^n b_k^2 \sigma_k^2\right).$$

Θεώρημα (Central Limit Theorem)

$X_1, X_2, \dots, X_n$ όμοια κατανεμημένες και ανεξαρτητες με μεσο και διασπορά $\mu, \sigma^2 < \infty$.

Τότε

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z \sim N(0, 1), \quad S_n = \sum_{k=1}^n X_k.$$

Η κανονική κατανομή μπορεί να περιγράψει στο όριο παρα πολλές κατανομες

$$X_n \sim \text{Bin}(n, p) \implies \frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} Y \sim N(0, 1), \quad \text{CLT},$$

$$X \sim \text{Poi}(\lambda), \implies X \sim N(\lambda, \lambda), \quad \text{approximately for large } \lambda$$

Μέση τιμή και διασπορά για την κανονική κατανομή

Αν $X \sim N(\mu, \sigma^2)$ τότε

$$\begin{aligned}\mathbb{E}[X] &= \mu, \\ \text{Var}(X) &= \sigma^2\end{aligned}$$

Παράδειγμα (Το υπόδειγμα Black-Scholes)

Ένα πολύ δημοφιλές υπόδειγμα για την εξέλιξη της τιμής των μετοχών σε συνεχή χρόνο είναι το υπόδειγμα Black-Scholes σύμφωνα με το οποίο η λογαριθμική απόδοση μιας μετοχής κάθε χρονική στιγμή t ακολουθεί την κανονική κατανομή,

$$\ln \left(\frac{S(t)}{S(0)} \right) \sim N(\mu t, \sigma^2 t).$$

Παράδειγμα (Η κατανομή Gamma)

Η κατανομή Γ αντιστοιχεί σε ένα μέτρο πιθανότητας στο $\mathbb{R}_+ = [0, \infty)$ και έχει πυκνότητα

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbf{1}_{[0, \infty)}, \quad \alpha, \lambda \in (0, \infty),$$

$$\Gamma(\alpha) := \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \alpha > 0,$$

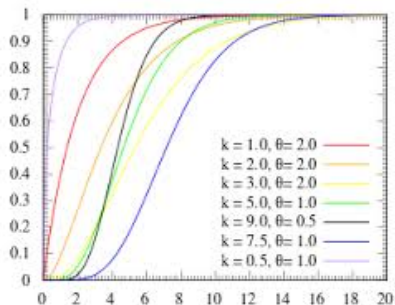
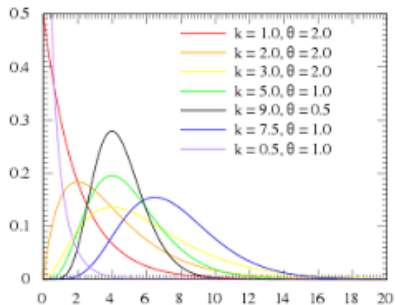
density of Gamma distribution

Εναλλακτική παραμετροποίηση

$$\theta = \frac{1}{\lambda} \text{ inverse scale parameter - rate parameter }, \quad k = \alpha, \text{ shape parameter}$$

Η συνάρτηση κατανομής δεν μπορεί εν γένει να γραφεί σε κλειστή μορφή.

Η κατανομή αυτή εμφανίζεται στην θεωρία αξιοπιστίας και μοντελοποιεί τον χρόνο αναμονής για την εμφάνιση πολλαπλών συμβάντων.



Source: Wikicommons

$$X_1, \dots, X_n \sim_{iid} \text{Exp}(\lambda) \implies \sum_{i=1}^n X_i \sim \text{Gamma}(n, \lambda)$$

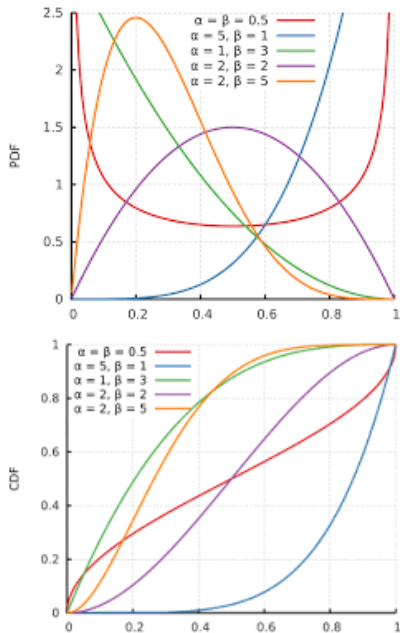
$$X_1, \dots, X_n \sim_{iid} \text{Gamma}(\alpha, \lambda) \implies \sum_{i=1}^n X_i \sim \text{Gamma}(n\alpha, \lambda)$$

Παράδειγμα (Η κατανομή B - Beta distribution)

Η κατανομή B αντιστοιχεί σε ένα μέτρο πιθανότητας στο $[0, 1] \subset \mathbb{R}$, και έχει πυκνότητα

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbf{1}_{[0,1]}, \quad \alpha, \beta > 0, \text{ shape parameters}$$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$



Source: Wikicommons

$X \sim \text{Beta}(a, b)$ αν η συνάρτηση πυκνότητας πιθανότητας της X είναι η

$$f(x) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}, \quad x \in [0, 1],$$

όπου η σταθερά $B(a, b)$ δίνεται με την βοήθεια του ολοκλήρωματος

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Η μεση τιμή και η διασπορά της X δίνεται από τους τύπους

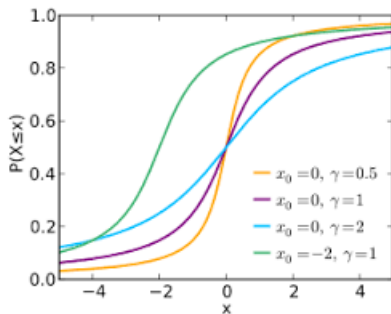
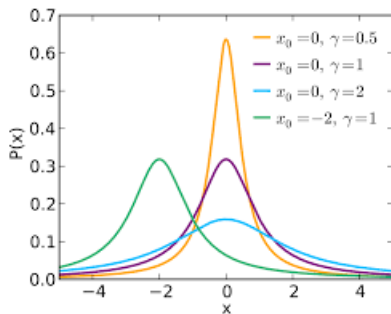
$$\mathbb{E}[X] = \frac{a}{a+b},$$

$$\text{Var}(X) = \frac{ab}{(a+b+1)(a+b)^2}.$$

Παράδειγμα (Η κατανομή *Cauchy*)

Η κατανομή *Cauchy* αντιστοιχεί σε ένα μέτρο πιθανότητας στο $\mathbb{R}$ και έχει πυκνότητα

$$f(x) = \frac{1}{\pi\gamma} \frac{\gamma^2}{\gamma^2 + (x - x_0)^2}$$



Source: Wikicommons

$$X_i \sim \text{Cauchy}(\mu_i, \sigma), \text{ independent} \implies \sum_{i=1}^n X_i \sim \text{Cauchy}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i\right),$$

$$X \sim \text{Cauchy}(\mu, \sigma) \implies \frac{1}{X} \sim \text{Cauchy}\left(\frac{\mu}{\mu^2 + \sigma^2}, \frac{\sigma}{\mu^2 + \sigma^2}\right).$$

Επίσης

$$Y_1, Y_2 \sim_{i.i.d} N(0, 1) \implies \frac{Y_1}{Y_2} \sim \text{Cauchy}(0, 1)$$

$$Y_1, Y_2 \sim_{i.i.d} N(0, 1) \implies \frac{Y_1 - Y_2}{2\sqrt{Y_1 Y_2}} \sim \text{Cauchy}(0, 1)$$

Παράδειγμα (Η κατανομή Fréchet)

Η κατανομή Fréchet αντιστοιχεί σε ένα μέτρο πιθανότητας στο $\mathbb{R}_{++} = (0, \infty)$ και έχει πυκνότητα

$$f(x, \alpha) = \alpha x^{-\alpha-1} e^{-x^{-\alpha}} \mathbf{1}_{(0, \infty)}, \quad \alpha > 0, \text{ shape parameter}$$

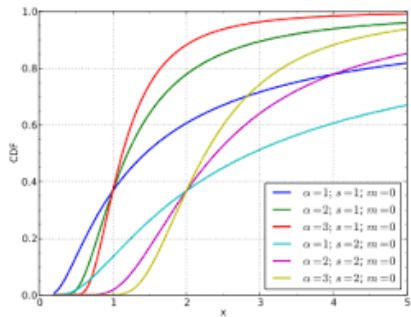
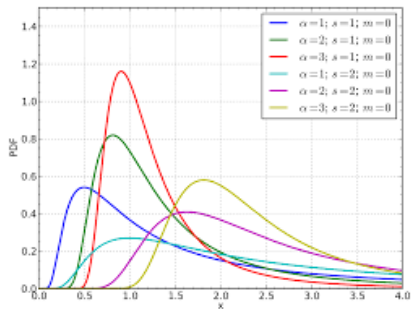
Η συνάρτηση κατανομής είναι

$$F(x) = e^{-x^{-\alpha}}, \quad x > 0.$$

Εναλλακτική παραμετροποίηση

$$F(x) = e^{-\left(\frac{x-\mu}{s}\right)^{-\alpha}}, \quad x > m.$$

Η κατανομή αυτή είναι μια απο τις κατανομές ακραίων τιμών (τύπου II).



Source: Wikicommons

Παράδειγμα (Η κατανομή Gumbel)

Η κατανομή Gumbel αντιστοιχεί σε ένα μετρο πιθανότητας στο $\mathbb{R}$ και έχει πυκνότητα

$$f(x) = e^{-x-e^{-x}}.$$

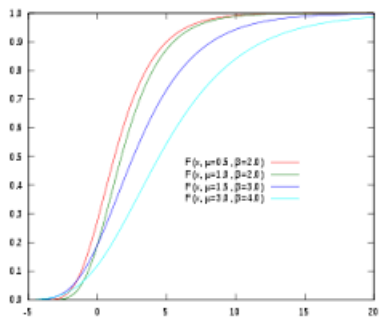
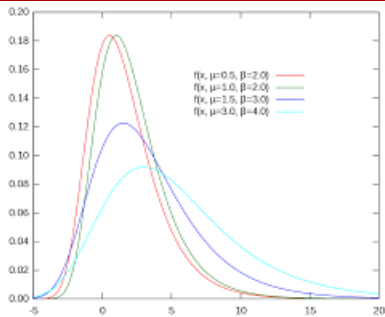
Η συνάρτηση κατανομής είναι η

$$F(x) = e^{e^{-x}}$$

Εναλλακτική παραμετροποίηση

$$F(x) = e^{e^{-\frac{x-\mu}{\beta}}}$$

Είναι και αυτή μια απο τις κατανομές ακραίων τιμών (τύπου I).



Source: Wikicommons

Παράδειγμα (Η κατανομή Student)

Η κατανομή Student αντιστοιχεί σε ένα μετρο πιθανότητας στο $\mathbb{R}$ και έχει πυκνότητα $f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$, όπου $\nu > 0$ είναι μια παράμετρος που αντιστοιχεί στους βαθμούς ελευθερίας.

Η κατανομή αυτή συμβολίζεται με t_ν .

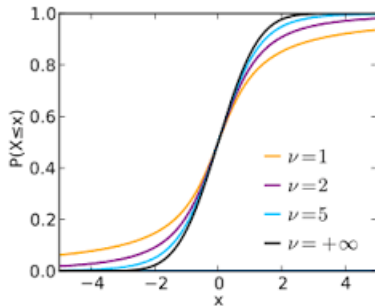
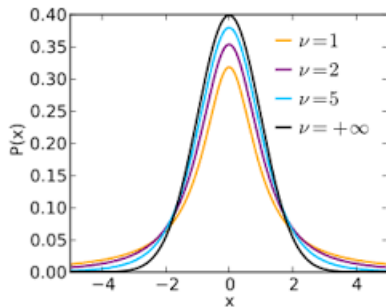
Μπορούμε να έχουμε και μια παραμετρική μορφή (location-scale) ως $t_\nu(\mu, \sigma^2) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$.

Η κατανομή του Student εμφανίζεται στην εκτίμηση του μεσου της κανονικής κατανομής, όταν επίσης η διασπορά πρέπει να εκτιμηθεί και το μέγεθος του δείγματος είναι μικρό.

Για παράδειγμα η εκτιμήτρια για τον δειγματικό μέσο $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, και η δειγματική διασπορά $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ έχουν την ιδιότητα

$$T := \frac{\bar{X}_n - \mu}{\frac{S_n}{\sqrt{n}}} \sim t_{n-1},$$

αν $X_1, \dots, X_n \sim_{iid} N(\mu, \sigma^2)$.



Source: Wikicommons

Παράδειγμα

Αν $X \sim N(0, 1)$ ποιά είναι η κατανομή της $Y = X^2$;

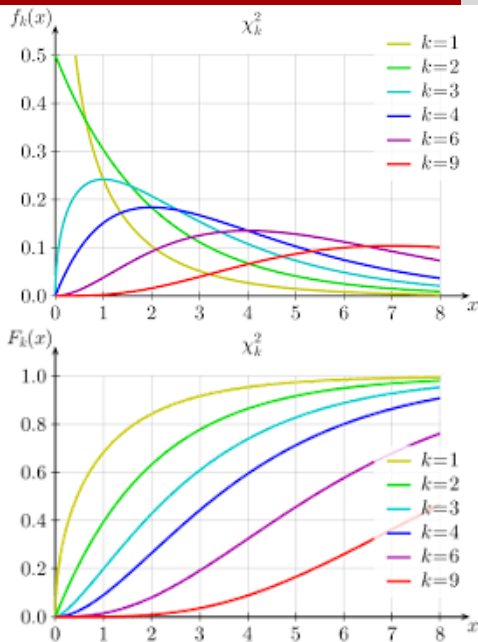
Η συνάρτηση $\varphi(s) = s^2$ μπορεί να χωριστεί σε 2 κλάδους στον καθένα απο τους οποίους έχουμε τις αναγκαίες συνθήκες μονοτονίας, $\varphi_1 : (-\infty, 0) \rightarrow \mathbb{R}_+$ και $\varphi_2 : [0, \infty) \rightarrow \mathbb{R}^+$, η κάθεμία με τύπο $\varphi_i(x) = x^2$.

Εύκολα μπορούμε να δούμε ότι $(\varphi_1)^{-1} = -\sqrt{y}$, και $(\varphi_2)^{-1}(y) = \sqrt{y}$.

Επίσης εύκολα βλέπουμε ότι $|(\varphi_i^{-1})'(y)| = \frac{1}{2\sqrt{y}}$, $y > 0$, $i = 1, 2$, οπότε η εφαρμογή του τύπου για τον μετασχηματισμο δίνει

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \frac{1}{\sqrt{y}}, \quad y > 0.$$

Η κατανομή αυτή ονομάζεται κατανομή χ^2 με ένα βαθμό ελευθερίας.



Source: Wikicommons