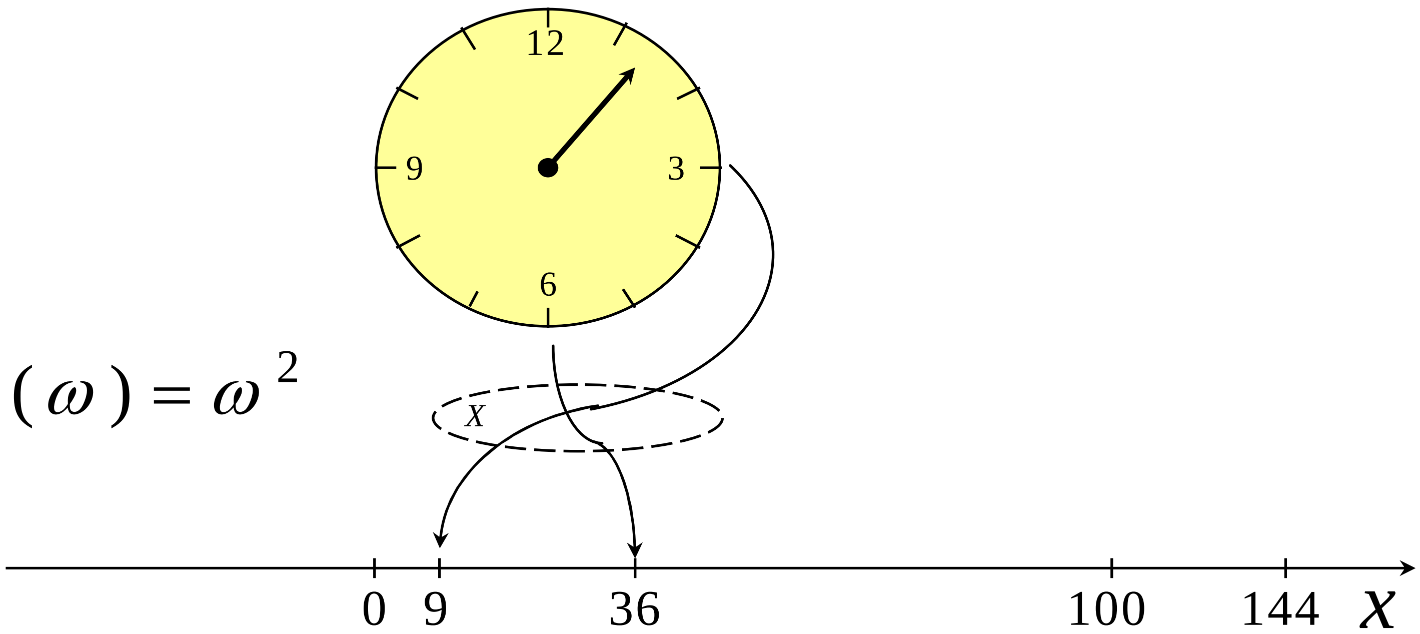
Η ΕΝΝΟΙΑ ΤΗΣ ΤΥΧΑΙΑΣ ΜΕΤΑΒΛΗΤΗΣ

Η απεικόνιση των εκβάσεων ενός πειράματος τύχης στην ευθεία των πραγματικών αριθμών οδηγεί στην ***τυχαία μεταβλητή***.

$$X = X(\omega) = \omega^2$$



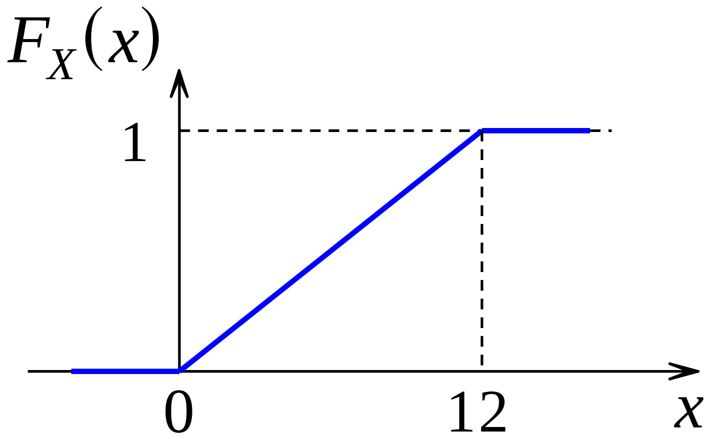
Τα αποτελέσματα ενός πειράματος τύχης ορίζουν μια ***τυχαία μεταβλητή*** (*random variable*).

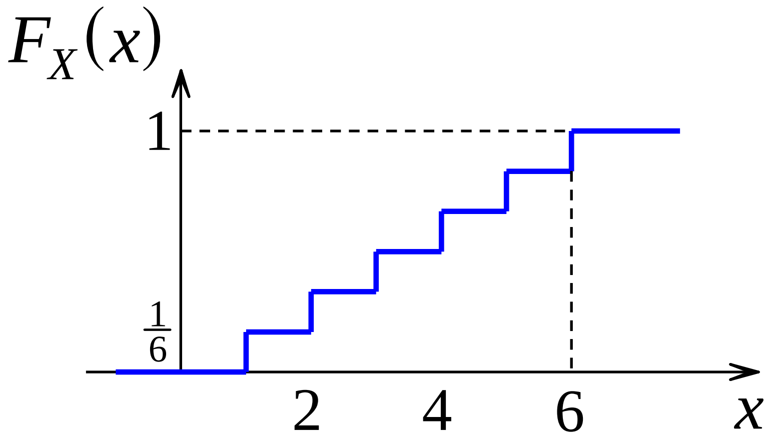
Συνεχείς - ***Διακριτές*** τυχαίες μεταβλητές

ΑΘΡΟΙΣΤΙΚΗ ΣΥΝΑΡΤΗΣΗ ΚΑΤΑΝΟΜΗΣ

Η ***αθροιστική συνάρτηση κατανομής*** *(cumulative distribution function (CDF))* μίας τυχαίας μεταβλητής *X* ορίζεται ως

$$F_X(x) = P(\omega \in \Omega : X(\omega) \leq x)$$



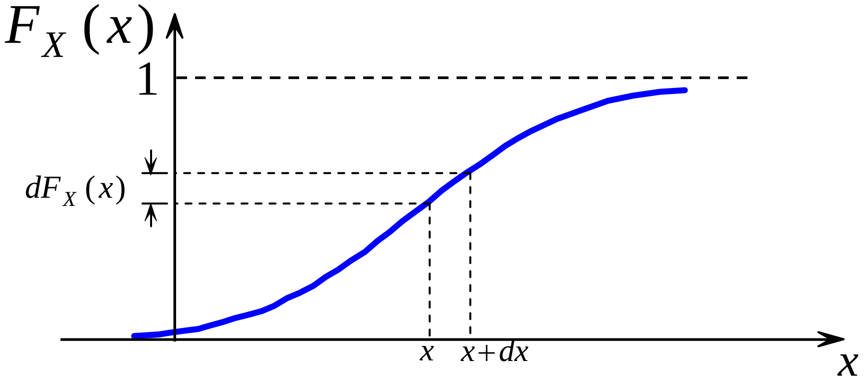


$$F_X(x) = \sum_{k=1}^{N} P(x_k)\, u(x - x_k)$$

Ιδιότητες της Αθροιστικής Συνάρτησης Κατανομής

$$0 \le F_X(x) \le 1$$

Η $F_X(x)$ είναι μη φθίνουσα



$$dF_X(x) = P[x \le X < x + dx]$$

$$\lim_{x \to -\infty} F_X(x) = 0 \quad \text{και} \quad \lim_{x \to +\infty} F_X(x) = 1$$

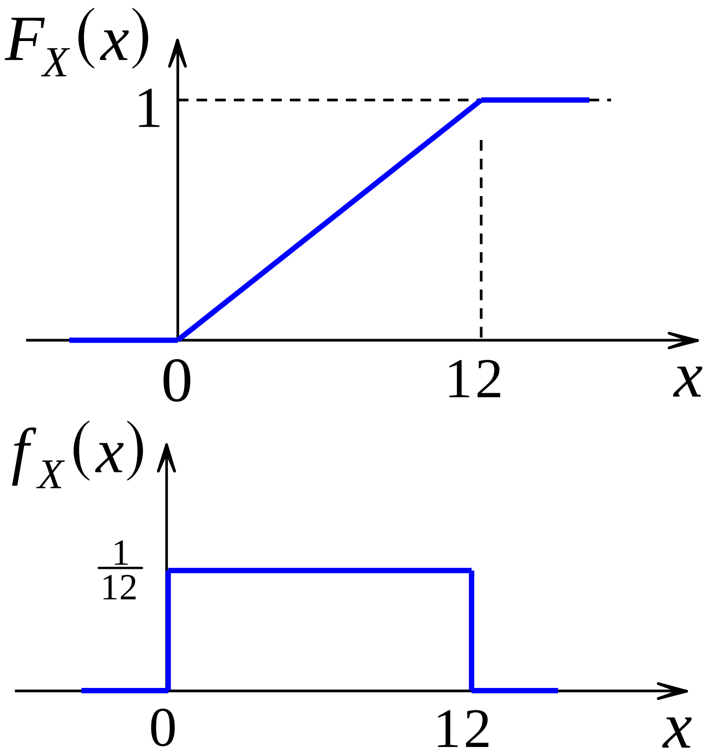
$$F_X(x_1) \le F_X(x_2) \quad \text{αν} \quad x_1 < x_2$$

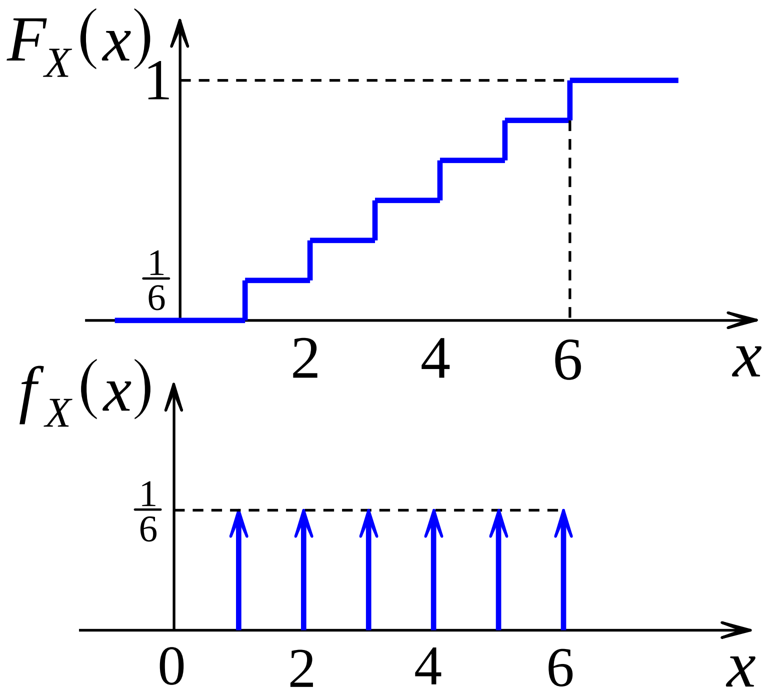
$$P(x_1 < X \le x_2) = F_X(x_2) - F_X(x_1)$$

ΣΥΝΑΡΤΗΣΗ ΠΥΚΝΟΤΗΤΑΣ ΠΙΘΑΝΟΤΗΤΑΣ

Η ***συνάρτηση πυκνότητας πιθανότητας*** (*probability density function* (PDF)) μίας τυχαίας μεταβλητής X ορίζεται ως η παράγωγος της $F_X(x)$, δηλαδή,

$$f_X(x) = \frac{dF_X(x)}{dx}$$



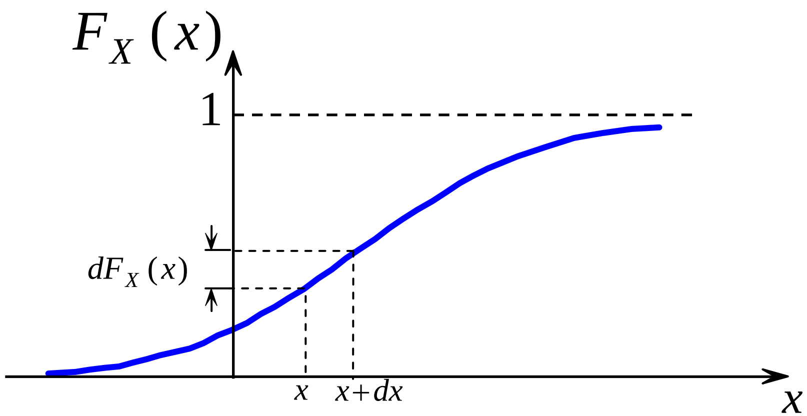


$$f_X(x) = \sum_{k=1}^{N} P(x_k)\delta(x - x_k)$$

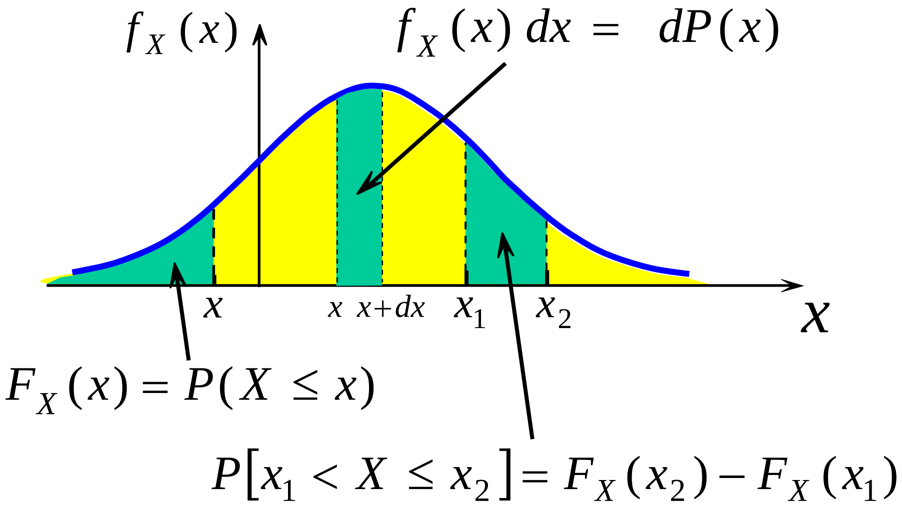
Για διακριτές τυχαίες μεταβλητές, συνηθίζεται να ορίζουμε τη ***συνάρτηση πιθανότητας μάζας*** (*probability mass function* (*PMF*)), η οποία ορίζεται ως $\{p_i\}$ όπου $p_i = P(X = x_i)$.

Προφανώς για κάθε *i* ισχύει $p_i \geq 0$ και $\sum_i p_i = 1$

Ιδιότητες της Συνάρτησης Πυκνότητας Πιθανότητας



$$f_X(x) = \frac{dF_X(x)}{dx}$$



$$0 \le f_X(x) \quad \text{για κάθε } x$$

$$\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$$

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

$$P(x_1 < X \le x_2) = \int_{x_1}^{x_2} f_X(x)\,dx$$

$$dF_X(x) = P[x \le X < x + dx]$$

$$f_X(x)\,dx = dP(x)$$

$$f_X(x) = \frac{dP(x)}{dx}$$

Συνάρτηση πυκνότητας πιθανότητας

Συναρτήσεις Τυχαίων Μεταβλητών

Μία συνάρτηση μίας τυχαίας μεταβλητής *X*, έστω $Y=g(X)$, είναι επίσης μία τυχαία μεταβλητή.

$$\underset{f_X(x)}{X} \longrightarrow \boxed{Y = g(X)} \longrightarrow \underset{f_Y(y)}{Y}$$

Για την οποία η αθροιστική συνάρτηση κατανομής είναι

$$F_Y(y) = P\big(\omega \in \Omega : g(X(\omega)) \le y\big)$$

Στην ειδική περίπτωση για την οποία, για κάθε *y*, η εξίσωση $g(x) = y$ έχει αριθμήσιμο σύνολο λύσεων $\{x_i\}$ και, για όλες αυτές τις λύσεις υπάρχει η παράγωγος $g'(x_i)$ και είναι μη μηδενική τότε

$$f_Y(y) = \sum_i \frac{f_X(x_i)}{|g'(x_i)|}$$

Παράμετροι κατανομής τυχαίας μεταβλητής

Η κατανομή πιθανότητας μιας τυχαίας μεταβλητής, όπως έχουμε ήδη αναφέρει, δύναται να εκφραστεί είτε από την αθροιστική συνάρτηση κατανομής είτε από τη συνάρτηση πυκνότητας πιθανότητας.

Στη συνέχεια θα περιγραφούν οι βασικές παράμετροι της κατανομής μιας τυχαίας μεταβλητής. Οι βασικές αυτές παράμετροι της κατανομής δίδουν μία περιληπτική περιγραφή της πιθανοθεωρητικής συμπεριφοράς μιας τυχαίας μεταβλητής.

Στατιστικοί Μέσοι Όροι

Η ***αναμενόμενη τιμή*** ή ***μαθηματική ελπίδα*** ή απλά ***μέση τιμή***, μιας διακριτής τυχαίας μεταβλητής X ορίζεται ως

$$E[X] = \sum_i x_i \, P(X = x_i)$$

όπου $p_i = P(X = x_i)$ είναι η συνάρτηση πιθανότητα μάζας της διακριτής τυχαίας μεταβλητής. Η μέση τιμή συνήθως δηλώνεται με m_X ή X.

Η ***αναμενόμενη τιμή*** ή ***μαθηματική ελπίδα*** ή απλά ***μέση τιμή,*** μιας συνεχούς τυχαίας μεταβλητής X ορίζεται ως

$$E[X] = \int_{-\infty}^{\infty} x \, f_X(x) \, dx$$

όπου $f_X(x)$ είναι η συνάρτηση πυκνότητας πιθανότητας. Η μέση τιμή συνήθως δηλώνεται με m_X ή $\overline{X}$.

Μέση τιμή συνάρτησης τυχαίας μεταβλητής

Ας θεωρήσουμε τη τυχαία μεταβλητή Y οι τιμές της οποίας ορίζονται από την πραγματική συνάρτηση $y = g(x)$, όπου x είναι οι τιμές μιας τυχαίας μεταβλητής X.

Η μέση τιμή της διακριτής τυχαίας μεταβλητής $Y = g(X)$ δίνεται από τη

$$E(Y) \equiv E\left[g(X)\right] = \sum_{k=0}^{\infty} g(x_k)\, P(X = x_k)$$

Η μέση τιμή της συνεχούς τυχαίας μεταβλητής $Y = g(X)$ δίνεται από τη

$$E(Y) \equiv E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx$$

Διασπορά και Τυπική Απόκλιση

Η ύπαρξη κατανομών οι οποίες έχουν την ίδια μέση τιμή και των οποίων οι τιμές είναι περισσότερο ή λιγότερο διασπαρμένες και απομακρυσμένες από αυτήν καθιστά αναγκαία την εισαγωγή της διακύμανσης η οποία είναι ένα μέτρο του βαθμού συγκέντρωσης ή του βαθμού μεταβολής της κατανομής μιας τυχαίας μεταβλητής.

Η ***διακύμανση*** (*variance*) ή η ***διασπορά*** της τυχαίας μεταβλητής X είναι η μέση τιμή του τετραγώνου της απόκλισης $g(X) = (X - m_X)^2$ της τυχαίας μεταβλητής X από τη μέση της τιμής $m_X = E(X)$.

Η διακύμανση ή η διασπορά της X συμβολίζεται με VAR(X) ή σ^2 ή απλά σ_X^2, και ορίζεται από τη σχέση

$$\sigma_X^2 = E\left[(X - m_X)^2\right]$$

Η θετική τετραγωνική ρίζα της διασποράς σ_X καλείται ***τυπική απόκλιση*** (*standard deviation*) της τυχαίας μεταβλητής X.

Για τη διακύμανση έχουμε

$$\sigma_X^2 = E\left[\left(X - E[X]\right)^2\right] = E\left[X^2 - 2X \cdot E[X] + \left(E[X]\right)^2\right]$$
$$= E\left[X^2\right] - \left(E[X]\right)^2$$

Για κάθε σταθερά c έχουμε τις ακόλουθες σχέσεις για τη μέση τιμή

1. $E[c\,X] = c\,E[X]$
2. $E[c] = c$
3. $E[X + c] = E[X] + c$

και τις ακόλουθες σχέσεις για τη διακύμανση

1. $\mathrm{VAR}(cX) = c^2\,\mathrm{VAR}(X)$
2. $\mathrm{VAR}(c) = 0$
3. $\mathrm{VAR}(X + c) = \mathrm{VAR}(X)$

Χαρακτηριστική συνάρτηση τυχαίας μεταβλητής

Η χαρακτηριστική συνάρτηση μίας τυχαίας μεταβλητής X δηλώνεται ως $\Psi_X(\nu)$ και ορίζεται ως

$$\Psi_X(\nu) \overset{\text{ορισ.}}{=} \int_{-\infty}^{\infty} f_X(x)\, e^{j\nu x}\, dx$$

Η χαρακτηριστική συνάρτηση τυχαίας μεταβλητής X συνδέεται με το μετασχηματισμό Fourier (διαφέρει στο πρόσημο στον εκθετικό όρο) της συνάρτησης πυκνότητας πιθανότητάς της.

Η χαρακτηριστική συνάρτηση τυχαίας μεταβλητής X παρέχει ένα εύκολο τρόπο υπολογοσμού των ροπών της τυχαίας μεταβλητής. Για να προσδιορίσουμε την ***n-στη ροπή*** της τυχαίας μεταβλητής X, χρησιμοποιούμε τη σχέση

$$m_X^{(n)} \overset{\text{ορισ.}}{=} \int_{-\infty}^{\infty} x^n\, f_X(x)\, dx = \frac{1}{j^n} \frac{d^n}{d\nu^n} \Psi_X(\nu)\Big|_{\nu=0}$$

Η χαρακτηριστική συνάρτηση τυχαίας μεταβλητής X είναι φραγμένη από την τιμή της στο μηδέν

$$\left|\Psi_X(\nu)\right| \le \Psi_X(0) = 1$$

Πολλαπλές Τυχαίες Μεταβλητές

Έστω X και Y δύο τυχαίες μεταβλητές. Για τις δύο αυτές τυχαίες μεταβλητές ορίζουμε τη ***συνδυασμένη αθροιστική συνάρτηση κατανομής*** (*joint* (CDF))

$$F_{X,Y}(x, y) = P(\omega \in \Omega : X(\omega) \leq x, Y(\omega) \leq y)$$

ή πιο απλά

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

και τη ***συνδυασμένη συνάρτηση πυκνότητας πιθανότητας*** (*joint* (PDF))

$$f_{X,Y}(x, y) = \frac{\vartheta^2}{\vartheta x \, \vartheta y} F_{X,Y}(x, y)$$

Βασικές ιδιότητες των συνδυασμένων και ***περιθώριων*** (*marginal*) CDF και PDF

$$F_X(x) = F_{XY}(x, \infty)$$

$$F_Y(y) = F_{XY}(\infty, y)$$

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

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

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)\, dx\, dy = 1$$

$$F_{XY}(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(u, \upsilon)\, du\, d\upsilon$$

Δεσμευμένη ή *υποσυνθήκη συνάρτηση πυκνότητας πιθανότητας*

Η ***δεσμευμένη*** ή ***υποσυνθήκη*** PDF της τυχαίας μεταβλητής X, υπό την προϋπόθεση ότι η τιμή της τυχαίας μεταβλητής Y είναι ίση με y ορίζεται ως

$$f_{X|Y}(x \mid y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_Y(y)}, & f_Y(y) \neq 0 \\ 0, & \text{αλλιώς} \end{cases}$$

Η συνάρτηση $f_{X|Y}(x|y)$ μπορεί να θεωρηθεί ως συνάρτηση της μεταβλητής x με τη μεταβλητή y αυθαίρετη αλλά σταθερή, συνεπώς ικανοποιεί όλες τις ιδιότητες της συνάρτησης πυκνότητας πιθανότητας

$$f_{X|Y}(x \mid y) \geq 0$$

$$\int_{-\infty}^{\infty} f_{X|Y}(x \mid y)\, dx = 1$$

$$F_{X|Y}(x \mid y) = \int_{-\infty}^{x} f_{X|Y}(\xi \mid y)\, d\xi$$

$$P\left(x_1 < X \leq x_2 \mid y\right) = \int_{x_1}^{x_2} f_{X|Y}(\xi \mid y)\, d\xi$$

Αν συμβαίνει η συνάρτηση πυκνότητας πιθανότητας μετά την γνωστοποίηση της Y να είναι η ίδια με τη συνάρτηση πυκνότητας πιθανότητας πριν τη γνωστοποίηση της Y τότε οι τυχαίες μεταβλητές ονομάζονται ***στατιστικά ανεξάρτητες*** (*statistical independent*) και ισχύει

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

Αν οι τυχαίες διαδικασίες X και Y είναι στατιστικά ανεξάρτητες, τότε η έκβαση της Y δεν επηρεάζει την κατανομή της X. Η υποσυνθήκη πιθανότητα $f_{X|Y}(x|y)$ απλοποιείται στην περιθώρια συνάρτηση πυκνότητας πιθανότητας

$$f_{X|Y}(x|y) = f_X(x)$$

Μία συνάρτηση δύο τυχαίων μεταβλητών X και Y, $g(X,Y)$, είναι επίσης τυχαία μεταβλητή.

Η αναμενόμενη τιμή της g(X, Y) δίνεται από την

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

Στην ειδική περίπτωση όπου g(X, Y) = $X\cdot Y$, λαμβάνουμε την $E[X\cdot Y]$ που ονομάζεται ***συσχέτιση*** (*correlation*) των X και Y και συμβολίζεται με $R_{X,Y}$

$$R_{X,Y}=E\left[X\cdot Y\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, y\ f_{XY}\left(x,y\right) dx\, dy$$

Αν η συσχέτιση γράφεται ως $R_{X,Y} = E[X]\cdot E[Y]$ τότε οι τυχαίες μεταβλητές λέγονται ***ασυσχέτιστες*** (*uncorrelated*). Αν οι τυχαίες μεταβλητές X και Y είναι στατιστικά ανεξάρτητες τότε είναι και ασυσχέτιστες. Το αντίθετο δεν ισχύει γενικά.

Αν η συσχέτιση γράφεται είναι ίση με μηδέν $R_{XY} = 0$ τότε οι τυχαίες μεταβλητές λέγονται ***ορθογώνιες*** (*orthogonal*).

Στην περίπτωση κατά την οποία $g(X,Y) = (X - m_X)\cdot(Y - m_Y)$ λαμβάνουμε την

$$E[(X - m_X)\cdot(Y - m_Y)]$$

η οποία καλείται ***συμμεταβολή*** (*covariance* - *COV*) των X και Y , η οποία συμβολίζεται με $C_{X,Y}$ ή με $COV(X,Y)$ και είναι

$$C_{X,Y} = E\left[\left(X - m_X\right)\cdot\left(Y - m_Y\right)\right]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x - m_X\right)\cdot\left(y - m_Y\right)\, f_{XY}(x,y)\,dx\,dy$$

αποδεικνύεται ότι $C_{XY} = R_{XY} - E[X]\cdot E[Y]$)

Αν οι X και Y είναι ανεξάρτητες ή ασυσχέτιστες $\left(R_{XY} = E[X]\cdot E[Y]\right)$ τότε η συμμεταβολή τους είναι ίση με μηδέν, $C_{XY} = 0$.

Αν οι X και Y είναι ορθογώνιες $\left(R_{XY} = 0\right)$ τότε η συμμεταβολή τους είναι $C_{XY} = - E[X]\cdot E[Y]$.

Αν επιπλέον μία τουλάχιστον από τις X και Y έχει μέση τιμή ίση με μηδέν τότε η συμμεταβολή τους είναι ίση με μηδέν $C_{XY} = 0$.

Η κανονικοποιημένη μορφή της συμμεταβολής καλείται ***συντελεστής συσχέτισης*** (*correlation coefficient*), συμβολίζεται με $\rho_{X,Y}$ και ορίζεται ως

$$\rho_{X,Y} = \frac{C_{XY}}{\sigma_X \, \sigma_Y} = E\left[\frac{X - m_X}{\sigma_X} \frac{Y - m_Y}{\sigma_Y}\right]$$

αποδεικνύεται ότι $-1 \leq \rho_{XY} \leq 1$

Σημαντικές Τυχαίες Μεταβλητές

Οι συχνότερα χρησιμοποιούμενες τυχαίες μεταβλητές είναι

Ομοιόμορφη τυχαία μεταβλητή

Η Gaussian τυχαία μεταβλητή

Διωνυμική (Binomial) Τυχαία μεταβλητή

Bernoulli τυχαία μεταβλητή

Poisson τυχαία μεταβλητή

Εκθετική τυχαία μεταβλητή

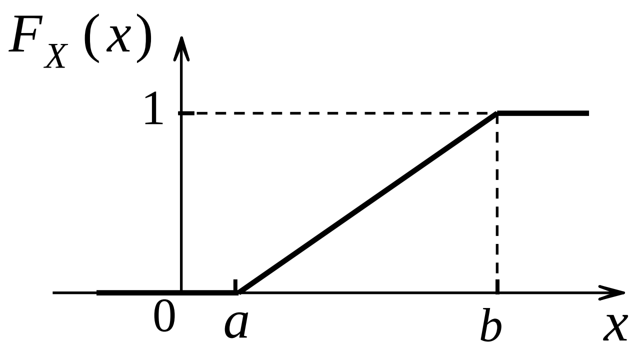
Rayleigh τυχαία μεταβλητή

Ομοιόμορφη Τυχαία Μεταβλητή

Ομοιόμορφη τυχαία μεταβλητή είναι μία συνεχής τυχαία μεταβλητή που λαμβάνει τιμές μεταξύ a και b με ίσες πιθανότητες σε διαστήματα τιμών που έχουν ίσα μήκη.

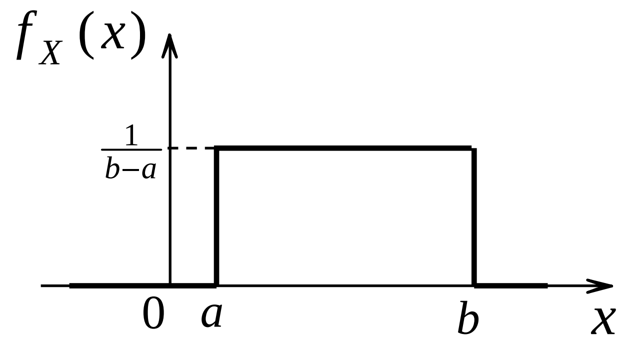
Η αθροιστική συνάρτηση κατανομής

$$F_X(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}x - \frac{a}{b-a} & a \le x < b \\ 1, & b \le x \end{cases}$$



Η συνάρτηση πυκνότητας πιθανότητας

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{αλλιώς} \end{cases}$$

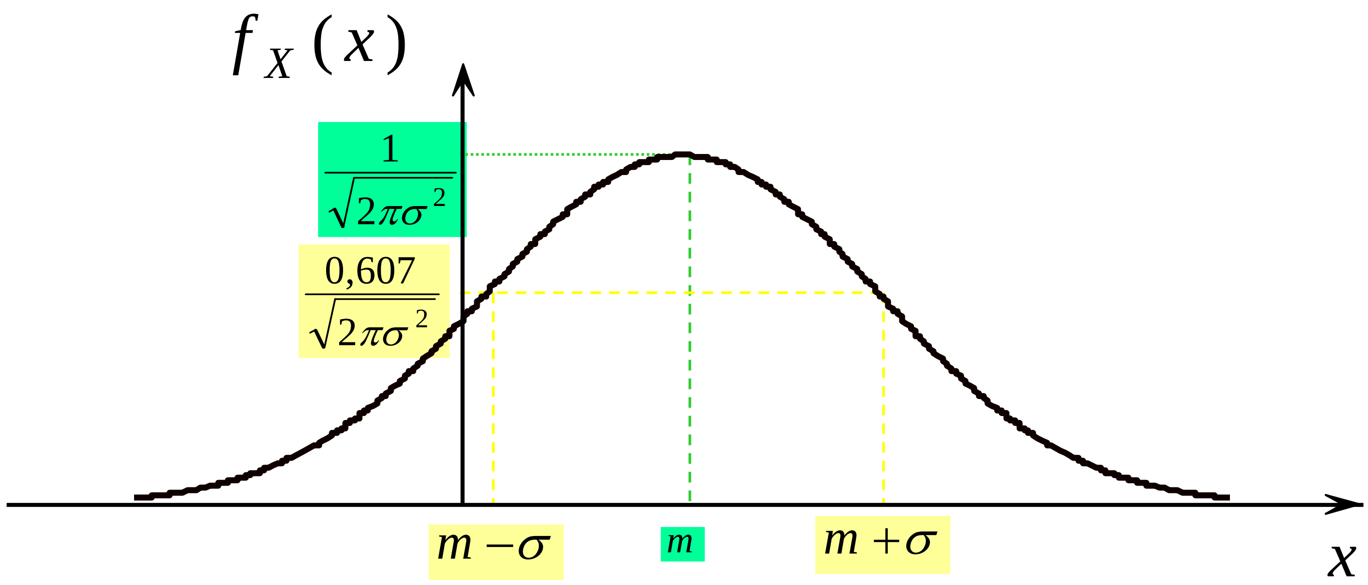


Η Gaussian Τυχαία Μεταβλητή

Η συνάρτηση πυκνότητας πιθανότητας είναι

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

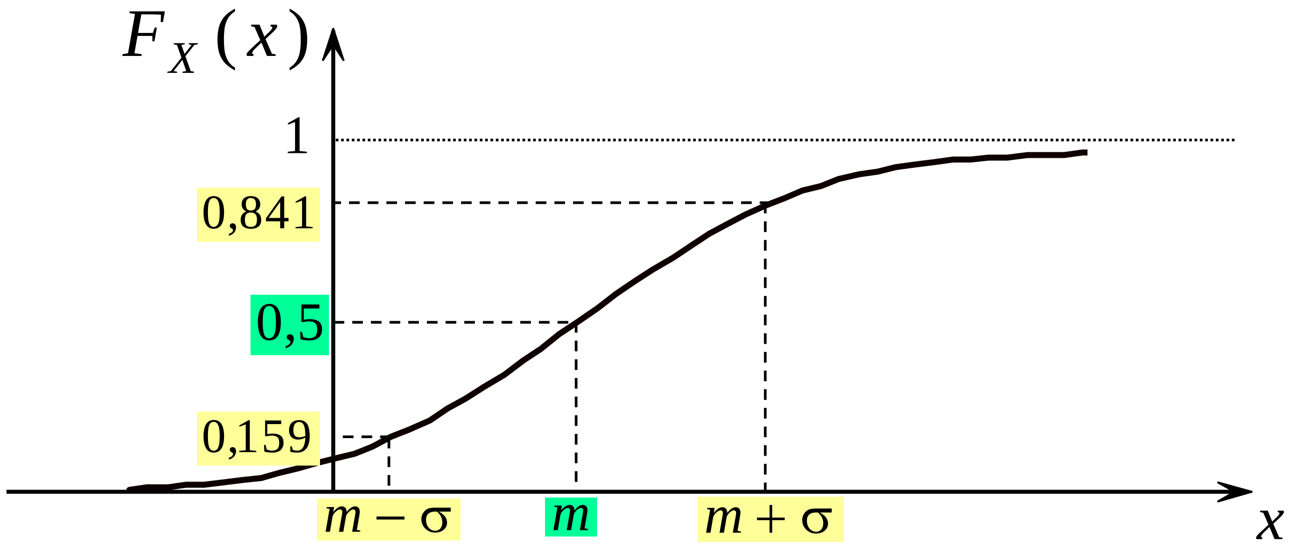
όπου m είναι η μέση τιμή και σ η τυπική απόκλιση



Η συνάρτηση πυκνότητας πιθανότητας της Gaussian τυχαίας μεταβλητής

Η συνάρτηση κατανομής είναι

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



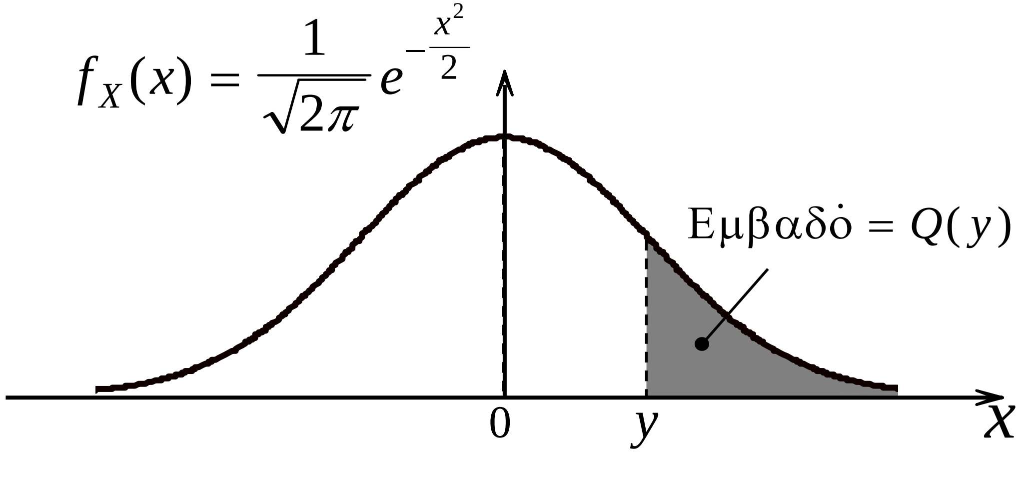
Η συνάρτηση κατανομής της Gaussian τυχαίας μεταβλητής

Η συνάρτηση κατανομής της Gaussian τυχαίας μεταβλητής για $m = 0$ και $\sigma = 1$ δηλώνεται με $\Phi(x)$ και δίνεται από τη σχέση

$$\Phi(x) = P(X \le x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{\xi^2}{2}} d\xi$$

Η συνάρτηση Q του Marcum ορίζεται ως

$$Q(y) = P(X > y) = \frac{1}{\sqrt{2\pi}} \int_{y}^{\infty} e^{-\frac{x^2}{2}} dx$$



Διάφορες παρουσιάσεις του ολοκληρώματος αυτού δίνονται σε μορφή εύχρηστων πινάκων ή διαγραμμάτων.

y	$Q(y)$	y	$Q(y)$	y	$Q(y)$
0,0	5,0000e-01	2,4	8,1975e-03	4,8	7,9332e-07
0,1	4,6017e-01	2,5	6,2096e-03	4,9	4,7918e-07
0,2	4,2074e-01	2,6	4,6611e-03	5,0	2,8665e-07
0,3	3,8208e-01	2,7	3,4669e-03	5,1	1,6982e-07
0,4	3,4458e-01	2,8	2,5551e-03	5,2	9,9644e-08
0,5	3,0853e-01	2,9	1,8658e-03	5,3	5,7901e-08
0,6	2,7425e-01	3,0	1,3498e-03	5,4	3,3320e-08
0,7	2,4196e-01	3,1	9,6760e-04	5,5	1,8989e-08
0,8	2,1185e-01	3,2	6,8713e-04	5,6	1,0717e-08
0,9	1,8406e-01	3,3	4,8342e-04	5,7	5,9903e-09
1,0	1,5865e-01	3,4	3,3692e-04	5,8	3,3157e-09
1,1	1,3566e-01	3,5	2,3262e-04	5,9	1,8175e-09
1,2	1,1506e-01	3,6	1,5910e-04	6,0	9,8658e-10
1,3	9,6800e-02	3,7	1,0779e-04	6,1	5,3034e-10
1,4	8,0756e-02	3,8	7,2348e-05	6,2	2,8231e-10
1,5	6,6807e-02	3,9	4,8096e-05	6,3	1,4882e-10
1,6	5,4799e-02	4,0	3,1671e-05	6,4	7,7688e-11
1,7	4,4565e-02	4,1	2,0657e-05	6,5	4,0160e-11
1,8	3,5930e-02	4,2	1,3345e-05	6,6	2,0557e-11
1,9	2,8716e-02	4,3	8,5398e-06	6,7	1,0420e-11
2,0	2,2750e-02	4,4	5,4125e-06	6,8	5,2309e-12
2,1	1,7864e-02	4,5	3,3976e-06	6,9	2,6001e-12
2,2	1,3903e-02	4,6	2,1124e-06	7,0	1,2798e-12
2,3	1,0724e-02	4,7	1,3008e-06		

Παρατηρούμε ότι

$$Q(x) = 1 - \Phi(x) \qquad \text{και} \qquad P(X > x) = Q\left(\frac{x-m}{\sigma}\right)$$

Bernoulli Τυχαία Μεταβλητή

Αυτή είναι μια διακριτή τυχαία μεταβλητή που παίρνει δύο τιμές, το ένα και το μηδέν, με πιθανότητες p και 1 - p. Η τυχαία μεταβλητή Bernoulli είναι ένα καλό μοντέλο για μια γεννήτρια δυαδικών δεδομένων.

Όταν δυαδικά δεδομένα μεταβιβάζονται μέσω ενός καναλιού επικοινωνίας, και μερικά bits λαμβάνονται εσφαλμένα μπορούμε να μοντελοποιήσουμε ένα σφάλμα με μία πρόσθεση modulo-2 του 1 στο bit εισόδου, μεταβάλλοντας έτσι το 0 σε 1 και το 1 σε 0.

Η τυχαία μεταβλητή Bernoulli χρησιμοποιείται για τη μοντελοποίηση των σφαλμάτων καναλιού.

Bernoulli trials

Σύνολο έχει n στοιχεία. Ο ολικός αριθμός των υποσυνόλων με k στοιχεία είναι

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{1\cdot 2\cdots k} = \frac{n!}{k!(n-k)!}$$

Παράδειγμα. Τα στοιχεία ενός συνόλου είναι a, b, c και d. Ποιος ο αριθμός των υποσυνόλων με δύο στοιχεία;

Έχουμε $n = 4$ και $k = 2$, επομένως το πλήθος των υποσυνόλων με 2 στοιχεία είναι

$$\binom{4}{2} = \frac{4\cdot 3}{1\cdot 2} = 6$$

Πράγματι έχουμε τα σύνολα *ab*, *ac*, *ad*, *bc*, *bd* και *cd*

Το παραπάνω συμπέρασμα χρησιμοποιείται για να βρεθεί η πιθανότητα να πραγματοποιηθεί ένα γεγονός k φορές σε n ανεξάρτητες πραγματοποιήσεις ενός πειράματος τύχης. Το πρόβλημα αυτό είναι βασικά το ίδιο με το πρόβλημα να έχουμε k φορές κεφάλι σε n στριψίματα ενός νομίσματος.

Παράδειγμα.

Αν στρίψουμε ένα νόμισμα η πιθανότητα να έχουμε κεφάλι είναι $P(K) = p$.
Σχηματίζεται μία ακολουθία από N στριψίματα του νομίσματος.
Η πιθανότητα στην ακολουθία να έχουμε k φορές κεφάλι και $N - k$ φορές γράμματα θα είναι λόγω της ανεξαρτησίας

$$P(K)\cdot P(K)\cdots P(K)P(\Gamma)\cdot P(\Gamma)\cdots P(\Gamma) = p^k(1-p)^{N-k}$$

Δεδομένου ότι οι πιθανές ακολουθίες μήκους N στις οποίες παρουσιάζονται k κεφάλια ανεξαρτήτου θέσης είναι

$$\binom{N}{k} = \frac{N!}{k!(N-k)!}$$

έχουμε τη πιθανότητα

$$P\left(K \text{ να συμβεί ακριβώς } k \text{ φορές}\right) = \binom{N}{k} p^k (1-p)^{N-k}$$

Διωνυμική (Binomial) Τυχαία Μεταβλητή

Αυτή είναι μία διακριτή τυχαία μεταβλητή που δίνει το πλήθος των 1 σε μία ακολουθία από n ανεξάρτητα πειράματα Bernoulli. Η PMF δίνεται από

$$P(X=k)=\begin{cases}\binom{n}{k}p^k(1-p)^{n-k}, & 0\le k\le n\\ 0, & \text{αλλιώς}\end{cases}$$

όπου $0 < p < 1$, και $n = 1, 2, \ldots$.

Η διωνυμική συνάρτηση πυκνότητα πιθανότητας είναι

$$f_X(x)=\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}\,\delta(x-k)$$

και η αθροιστική διωνυμική συνάρτηση κατανομής

$$F_X(x)=\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}\,u(x-k)$$

Αυτή η τυχαία μεταβλητή μοντελοποιεί, για παράδειγμα, τον ολικό αριθμό των bits τα οποία λαμβάνονται εσφαλμένα όταν μια ακολουθία από n bits μεταβιβάζεται μέσα από ένα κανάλι με πιθανότητα σφάλματος-bit ίση με p.

Poisson Τυχαία Μεταβλητή

Αυτή είναι μία διακριτή τυχαία μεταβλητή της οποίας η συνάρτηση πυκνότητας πιθανότητας είναι

$$f_X(x) = e^{-b} \sum_{k=0}^{\infty} \frac{b^k}{k!} \delta(x-k)$$

όπου $b > 0$ πραγματική σταθερά.

Η αθροιστική συνάρτηση κατανομής

$$F_X(x) = e^{-b} \sum_{k=0}^{\infty} \frac{b^k}{k!} u(x-k)$$

Αυτή η τυχαία μεταβλητή εφαρμόζεται σε πολλές εφαρμογές συνεχούς χρόνου. Μοντελοποιεί, για παράδειγμα τον αριθμό των τηλεφωνικών κλήσεων που πραγματοποιούνται σε ένα τηλεφωνικό κέντρο κατά τη διάρκεια μιας χρονικής περιόδου.

Αν το χρονικό διάστημα ενδιαφέροντος έχει διάρκεια T, και τα γεγονότα τα οποία καταμετρώνται πραγματοποιούνται με μέσο ρυθμό λ και ακολουθούν κατανομή Poisson τότε

$$b = \lambda T$$

επομένως

$$f_X(x) = e^{-\lambda T} \sum_{k=0}^{\infty} \frac{(\lambda T)^k}{k!} \delta(x-k)$$

Σε ένα τηλεφωνικό κέντρο δημιουργείται ουρά όταν φτάνουν τηλεφωνικές κλήσεις ενώ όλες οι γραμμές είναι απασχολημένες.

Παράδειγμα: Υποθέτοντας ότι οι τηλεφωνικές κλήσεις σε ένα τηλεφωνικό κέντρο ακολουθούν κατανομή Poisson και πραγματοποιούνται με μέσο ρυθμό 50 κλήσεις/h. Το τηλεφωνικό κέντρο έχει τη δυνατότητα να εξυπηρετήσει μόνο ένα συνδρομητή και κάθε τηλεφωνική κλήση διαρκεί ένα λεπτό. Ποια είναι η πιθανότητα να πραγματοποιηθεί ουρά αναμονής στο τηλεφωνικό κέντρο;

Η ουρά αναμονής θα πραγματοποιηθεί αν δύο ή περισσότερη συνδρομητές εκδηλώσουν επιθυμία να πραγματοποιήσουν κλήση σε χρονικό διάστημα ενός λεπτού.

Η πιθανότητα πραγματοποιηθεί ουρά αναμονής στο τηλεφωνικό κέντρο είναι ένα μείον την πιθανότητα να έχουμε περισσότερες από μία τηλεφωνικές κλήσεις. Έχουμε λ = 50/60 τηλεφωνικές κλήσεις / min και T = 1 min έτσι b =5/6 . Έχουμε λοιπόν

$$\begin{aligned} P(\text{πραγματοποιηθεί ουρά}) &= 1 - F_X(1) \\ &= 1 - e^{-\frac{5}{6}}\left(1 + \frac{5}{6}\right) \\ &= 0{,}2032 \end{aligned}$$

Αναμένεται λοιπόν να υπάρχει ουρά στο τηλεφωνικό κέντρο περίπου κατά το 20,32% του χρόνου.

Εκθετική τυχαία μεταβλητή

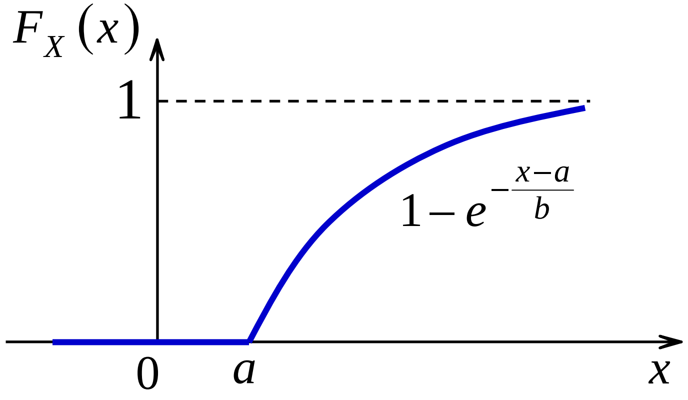
Η αθροιστική συνάρτηση κατανομής

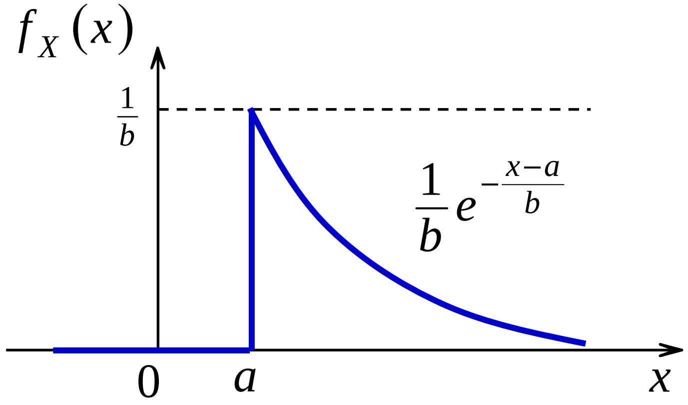
$$F_X(x) = \begin{cases} 1 - e^{-\frac{x-a}{b}}, & x > a \\ 0, & x < a \end{cases}$$

Η συνάρτηση πυκνότητας πιθανότητας

$$f_X(x) = \begin{cases} \frac{1}{b} e^{-\frac{x-a}{b}}, & x > a \\ 0, & x < a \end{cases}$$

για πραγματικούς αριθμούς a και b για του οποίους ισχύει $-\infty < a < \infty$ και $b > 0$.





Για α = 0 και λ = 1/b έχουμε

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x < 0 \end{cases}$$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x < 0 \end{cases}$$

Αποδεικνύεται ότι

$$E[X] = \frac{1}{\lambda} \qquad \mathrm{Var}[X] = \frac{1}{\lambda^2}$$

Η εκθετική συνάρτηση είναι η μοναδική συνάρτηση με έλλειψη μνήμης, δηλαδή, ικανοποιεί τη

$$P(X > s + t \mid X > t) = P(X > s) \qquad \text{για κάθε } s \text{ και } t > 0$$

Αυτή η τυχαία μεταβλητή μοντελοποιεί, για παράδειγμα τους χρόνου εξυπηρέτησης σε ένα τηλεφωνικό κέντρο.

Αν τυχαία μεταβλητή X παριστάνει το χρόνο μεταξύ δύο διαδοχικών τηλεφωνικών κλήσεων, ακολουθεί εκθετική κατανομή και θεωρήσουμε ότι μία κλήση έφτασε τη χρονική στιγμή t = 0 τότε η συνάρτηση πυκνότητας πιθανότητας του χρόνου μέχρι την επόμενη κλήση θα είναι

$$f_X(t) = \lambda\, e^{-\lambda t}$$

Επίσης η τυχαία μεταβλητή μοντελοποιεί, τη διακύμανση της ισχύος του σήματος που λαμβάνει το radar για τους διάφορους τύπους αεροσκαφών.

Παράδειγμα:
Η ισχύς που ανακλάται από αεροσκάφος σύνθετης μορφής και η οποία λαμβάνεται από διάταξη radar μοντελοποιείται από εκθετική τυχαία μεταβλητή P. Η συνάρτηση πυκνότητας πιθανότητας είναι,

$$f_P(p) = \begin{cases} \frac{1}{P_0} e^{-\frac{p}{P_0}}, & p > 0 \\ 0, & p < 0 \end{cases}$$

όπου P_0 είναι η μέση τιμή της λαμβανομένης ισχύος.

Σε κάποια δεδομένη χρονική στιγμή P έχει τιμή διάφορη από τη μέση της τιμή. Ποια είναι η πιθανότητα η λαμβανόμενη ισχύς να είναι μεγαλύτερη από τη ισχύ που λαμβάνει κατά μέσο όρο;

$$\begin{aligned} P(P > P_0) = 1 - P(P \le P_0) = 1 - F_P(P_0) \\ = 1 - \left(1 - e^{-P_0/P_0}\right) \\ = 0{,}368 \end{aligned}$$

Αναμένεται λοιπόν η μέση ισχύς να είναι μεγαλύτερη της μέσης τιμής κατά το 36,8% του χρόνου.

Rayleigh τυχαία μεταβλητή

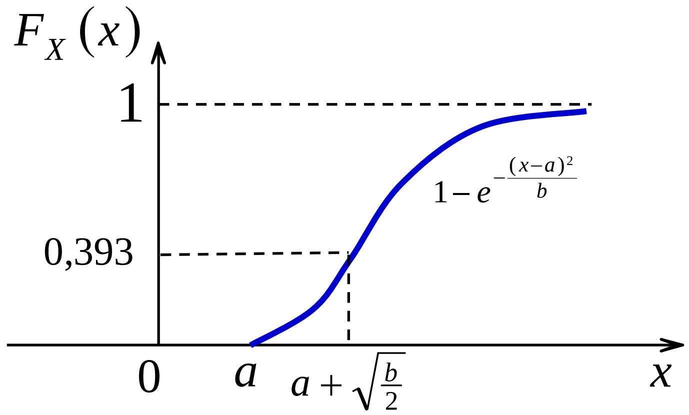
Η αθροιστική συνάρτηση κατανομής

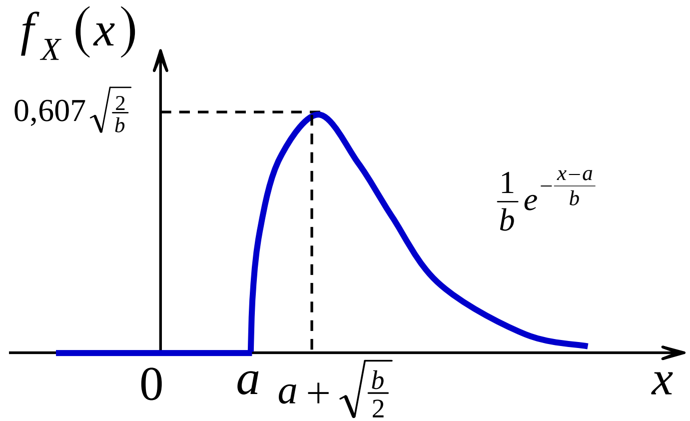
$$F_X(x) = \begin{cases} 1 - e^{-\frac{(x-a)^2}{b}}, & x \geq a \\ 0, & x < a \end{cases}$$

Η συνάρτηση πυκνότητας πιθανότητας

$$f_X(x) = \begin{cases} \frac{2}{b}(x-a)e^{-\frac{(x-a)^2}{b}}, & x \geq a \\ 0, & x < a \end{cases}$$

για πραγματικούς αριθμούς a και b για του οποίους ισχύει $-\infty < a < \infty$ και $b > 0$.





Αυτή η τυχαία μεταβλητή έχει πολλές εφαρμογές στα κανάλια επικοινωνίας τα οποία παρουσιάζουν διαλείψεις. Επίσης περιγράφει την περιβάλλουσα του θορύβου όταν αυτός διέρχεται από bandpass φίλτρα. Επίσης βρίσκει εφαρμογή στην ανάλυση των λαθών στα συστήματα μέτρησης

Poisson Τυχαία Διαδικασία

$$P\big(X(t)=k\big)=\frac{(\lambda t)^k\ e^{-\lambda t}}{k!},\ \ k=0,1,2,\ldots$$

$$f_X(x)=\sum_{k=0}^{\infty}\frac{(\lambda t)^k\ e^{-\lambda t}}{k!}\delta(x-k)$$

όπου λ ονομάζεται ***ρυθμός της διαδικασίας*** (*rate of the process*).

Η αθροιστική συνάρτηση κατανομής της τυχαίας διαδικασίας είναι

$$F_X(x)=\sum_{k=0}^{\infty}\frac{(\lambda t)^k\ e^{-\lambda t}}{k!}\ u(x-k)$$

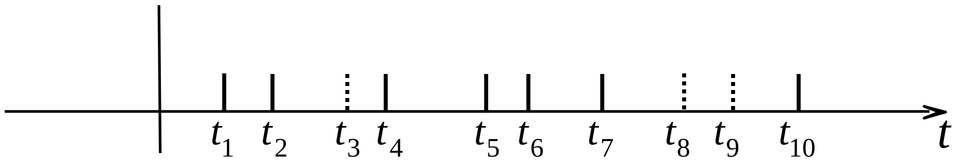
Η μέση τιμή και η διασπορά της τυχαίας μεταβλητής είναι

$$E\big[X(t)\big]=\mathrm{Var}[X(t)]=\int_{-\infty}^{\infty}x\ f_X(x)\,dx=\lambda t$$

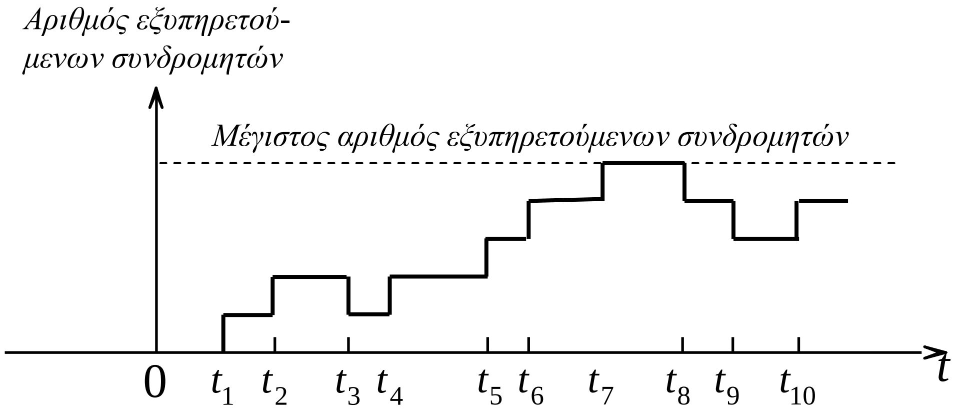
και η μέση τιμή του τετραγώνου της τυχαίας διαδικασίας είναι

$$E\big[X^2(t)\big]=\int_{-\infty}^{\infty}x^2\ f_X(x)\,dx=\lambda t\big(1+\lambda t\big)$$

Η χρονική στιγμή που πραγματοποιείται ένα γεγονός αντιστοιχεί σε ένα σημείο στον άξονα των χρόνων. Έχουμε λοιπόν τη γραφική παράσταση



Οι τυχαίες χρονικές στιγμές κατά τις οποίες πραγματοποιούνται και διακόπτονται τηλεφωνικές κλήσεις οι οποίες δημιουργούν την τυχαία διαδικασία



Ένα δείγμα συνάρτησης της διακριτής τυχαίας διαδικασίας Poisson -

Παρατηρούμε ότι η συνάρτηση $x(t)$ που παριστάνει τον αριθμό των εξυπηρετούμενων συνδρομητών αυξάνεται κατά ένα για κάθε νέα τηλεφωνική κλήση και ελαττώνεται επίσης κατά ένα όταν μία κλήση τελειώνει.

Η ένταση της κίνησης για χρονικό διάστημα T είναι

$$a = \frac{1}{T}\int_0^T x(t)\,dt = \frac{\lambda}{\mu} = \lambda\, s = \quad \text{μέσος ρυθμός άφιξης} \times \text{μέσος χρόνος κατάληψης}$$

Πολλαπλές Συναρτήσεις Πολλαπλών Τυχαίων Μεταβλητών

$$Z = g(X, Y)$$
$$W = h(X, Y)$$

η συνδυασμένη συνάρτηση πυκνότητας πιθανότητας είναι

$$f_{ZW}(z, w) = \sum_i \frac{f_{XY}(x_i, y_i)}{|\det \mathbf{J}(x_i, y_i)|}$$

όπου $\{x_i, y_i\}$ είναι ένα αριθμήσιμο πλήθος λύσεων του συστήματος

$$g(x, y) = z$$
$$h(x, y) = w$$

και στα σημεία αυτά η ορίζουσα του Jacobian πίνακα

$$\mathbf{J}(x, y) = \begin{bmatrix} \frac{\vartheta z}{\vartheta x} & \frac{\vartheta z}{\vartheta y} \\ \frac{\vartheta w}{\vartheta x} & \frac{\vartheta w}{\vartheta y} \end{bmatrix}$$

να είναι μη μηδενική

Συνδυασμένες Gaussian Τυχαίες Μεταβλητές

Οι ***συνδυασμένες Gaussian*** ή ***δικανονικές τυχαίες*** μεταβλητές X και Y ορίζονται από τη συνδυασμένη συνάρτηση πυκνότητας πιθανότητας

$$f_{X,Y}(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\cdot\left[\frac{(x-m_1)^2}{\sigma_1^2}-\frac{2\rho(x-m_1)(y-m_2)}{\sigma_1\sigma_2}+\frac{(y-m_2)^2}{\sigma_2^2}\right]\right\}$$

Όταν οι δύο τυχαίες μεταβλητές X και Y κατανέμονται σύμφωνα με την δικανονική κατανομή τότε η κατανομή κάθε μίας από αυτές $f_X(x)$ και $f_Y(y)$ και οι υποσυνθήκη πυκνότητες πιθανότητας $f_{X|Y}(x|y)$ και $f_{Y|X}(y|x)$ είναι Gaussian κατανομές.

Η ιδιότητα αυτή είναι η κύρια διαφορά μεταξύ δύο συνδυασμένα Gaussian τυχαίων μεταβλητών και δύο τυχαίων μεταβλητών η κάθε μία από τις οποίες έχει Gaussian κατανομή.

Αν οι X και Y είναι συνδυασμένες Gaussian της πιο πάνω μορφής, τότε, η X είναι Gaussian με μέση τιμή m_1 και διακύμανση σ_1^2, και η Y είναι Gaussian με μέση τιμή m_2 και διακύμανση σ_2^2, ο δε συντελεστής συσχέτισης μεταξύ των X και Y είναι ρ.

Ο ορισμός των δύο συνδυασμένων Gaussian τυχαίων μεταβλητών μπορεί να επεκταθεί σε n τυχαίες μεταβλητές X_1, X_2, …, Xn.

$$f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)=\frac{1}{\sqrt{(2\pi)^n\det(\mathbf{C})}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})^t\right\}$$

όπου **X**, **x**, **m** και **x** – **m** είναι τα διανύσματα

$$\mathbf{X}=\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \qquad \mathbf{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \mathbf{m}=\begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix} \qquad \mathbf{x}-\mathbf{m}=\begin{bmatrix} x_1-m_1 \\ x_2-m_2 \\ \vdots \\ x_n-m_n \end{bmatrix}$$

και **C** είναι ο πίνακας συμμεταβολής

$$\mathbf{C}=\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}$$

και **C** είναι ο πίνακας συμμεταβολής

$$\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{bmatrix}$$

με στοιχεία

$$C_{ij} = E\left[\left(\mathbf{x}_i - \mathbf{m}_i\right)\left(\mathbf{x}_j - \mathbf{m}_j\right)\right] = \begin{cases} \sigma_{x_i}^2, & i = j \\ \rho\sigma_{x_i}\sigma_{x_j}, & i \neq j \end{cases}$$

Για την περίπτωση στην οποία $n = 2$ είναι

$$\mathbf{C} = \begin{bmatrix} \sigma_{x_1}^2 & \rho\sigma_{x_1}\sigma_{x_2} \\ \rho\sigma_{x_1}\sigma_{x_2} & \sigma_{x_2}^2 \end{bmatrix}$$

και ο αντίστροφος πίνακας είναι

$$\mathbf{C}^{-1} = \frac{1}{\left(1-\rho^2\right)} \begin{bmatrix} 1/\sigma_{x_1}^2 & -\rho/\sigma_{x_1}\sigma_{x_2} \\ -\rho/\sigma_{x_1}\sigma_{x_2} & 1/\sigma_{x_2}^2 \end{bmatrix} \quad \text{τελικά} \quad \mathbf{C}^{-1} = \frac{1}{\sigma_{x_1}^2\sigma_{x_2}^2\left(1-\rho^2\right)}$$

Αθροίσματα Τυχαίων Μεταβλητών

Αν έχουμε μια ακολουθία τυχαίων μεταβλητών (X_1, X_2, …, X_n), που έχουν βασικά τις ίδιες ιδιότητες, τότε αναμένεται ότι η μέση τιμή

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

να συμπεριφέρεται "λιγότερο τυχαία", από κάθε X_i.

Ο ***νόμος των μεγάλων αριθμών*** και το θεώρημα του ***κεντρικού ορίου*** αποτελούν μια ακριβή διατύπωση του διαισθητικού αυτού γεγονότος.

Ο ***ασθενής νόμος των μεγάλων αριθμών*** δηλώνει ότι αν οι τυχαίες μεταβλητές X_1, X_2, … ,X_n είναι ασυσχέτιστες με την ίδια μέση τιμή m_X και διακύμανση $\sigma_X^2 < \infty$, τότε για κάθε $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(| Y - m_X | > \varepsilon\right) = 0$$

Αυτό σημαίνει ότι η μέση τιμή του αθροίσματος συγκλίνει (ως προς την πιθανότητα) στην αναμενόμενη τιμή.

Το ***θεώρημα του κεντρικού ορίου*** όχι μόνο διατυπώνει τη σύγκλιση της αναμενόμενης τιμής προς τη μέση τιμή αλλά επίσης δίνει κάποια πληροφορία για την κατανομή του μέσου αυτού όρου.

Το θεώρημα αυτό δηλώνει ότι αν είναι ανεξάρτητες τυχαίες μεταβλητές $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες τυχαίες μεταβλητές με αναμενόμενες τιμές $m_1, m_2, \ldots, m_n$ και διακυμάνσεις σ_1^2, $\sigma_2^2, \ldots, \sigma_n^2$, τότε η αθροιστική συνάρτηση κατανομής της τυχαίας μεταβλητής

$$Y = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - m_i}{\sigma_i}$$

συγκλίνει στην αθροιστική συνάρτηση κατανομής μίας Gaussian τυχαίας μεταβλητής με μέση τιμή 0 και διακύμανση 1, δηλαδή συμπεριφέρεται ασυμπτωτικά σαν κανονική μεταβλητή. Με άλλα λόγια ισχύει

$$\lim_{n \to \infty} P\left(a \le \frac{S_n - m_n}{\sigma\sqrt{n}} \le b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{u^2}{2}} du$$

όπου $S_n = X_1 + X_2 + \ldots + X_n$, $(n = 1, 2, 3, \ldots)$ τυχαία διαδικασία με μέση τιμή $m_n = m_1 + m_2 + \ldots + m_n$

Στην ειδική περίπτωση στην οποία οι X_i είναι ανεξάρτητες και όμοια κατανεμημένες (*independent and identically distributed-i.i.d.*) το θεώρημα του κεντρικού ορίου λέει ότι η αθροιστική συνάρτηση κατανομής της

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

συγκλίνει στην αθροιστική συνάρτηση κατανομής της μιας Gaussian τυχαίας μεταβλητής με μέση τιμή m και διακύμανση σ_2/n, $N(m, \sigma_2/n)$.

Για να ισχύει το θεώρημα του κεντρικού ορίου, πρέπει οι τυχαίες μεταβλητές να είναι ανεξάρτητες, ενώ ο νόμος των μεγάλων αριθμών ισχύει κάτω από λιγότερο περιοριστικές συνθήκες, δηλαδή, της μη συσχέτισης των τυχαίων μεταβλητών.