

Náhodná premenná

Pojmom náhodná premenná (veličina) označujeme funkciu, ktorej hodnota je určená výsledkom náhodného pokusu.

Priraduje číselnú hodnotu každému elementárnemu javu.

Hodnota funkcie sa nedá pred prevedením pokusu určiť, ale dokážeme ju popísať rozdelením pravdepodobnosti.

Distribučná funkcia

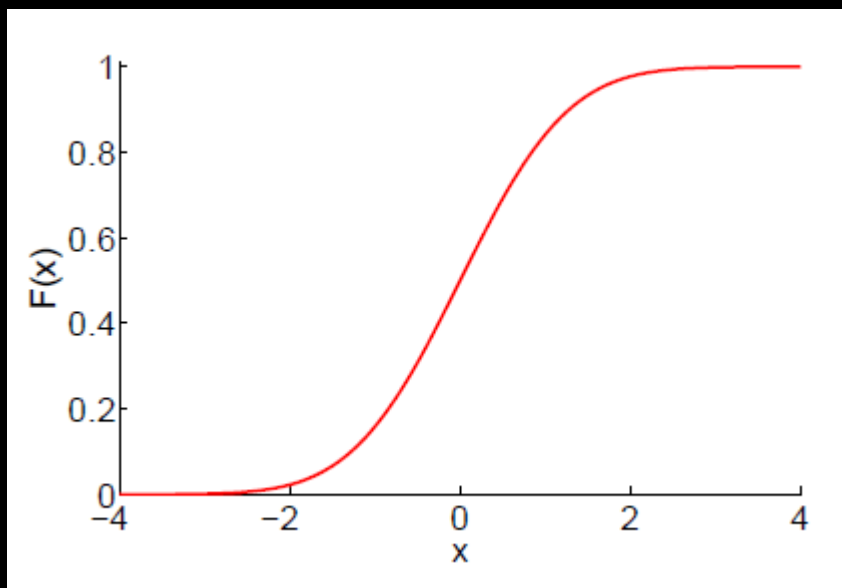
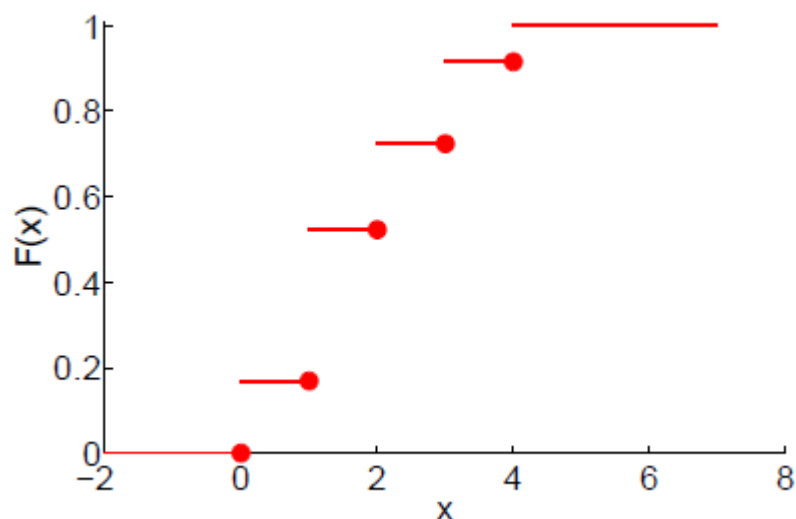
Rozdelenie pravdepodobnosti náhodnej premennej definovanej na pravdepodobnostnom priestore $(\Omega, \mathcal{A}, P)$ opisujeme pomocou distribučnej funkcie $F: \mathbb{R} \rightarrow [0,1]$

$$F(x) = P(X \leq x)$$

Distribučná funkcia je neklesajúca a zľava spojitá a platí

$$\lim_{x \rightarrow -\infty} F(x) = 0$$

$$\lim_{x \rightarrow \infty} F(x) = 1$$



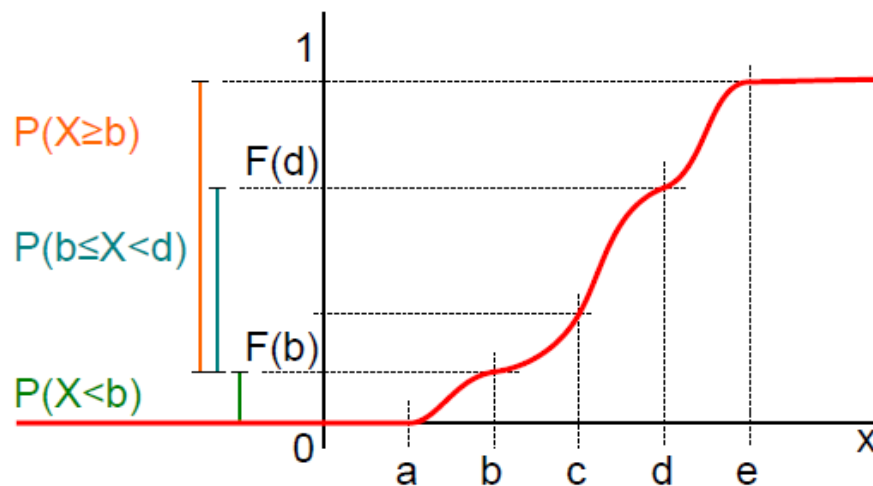
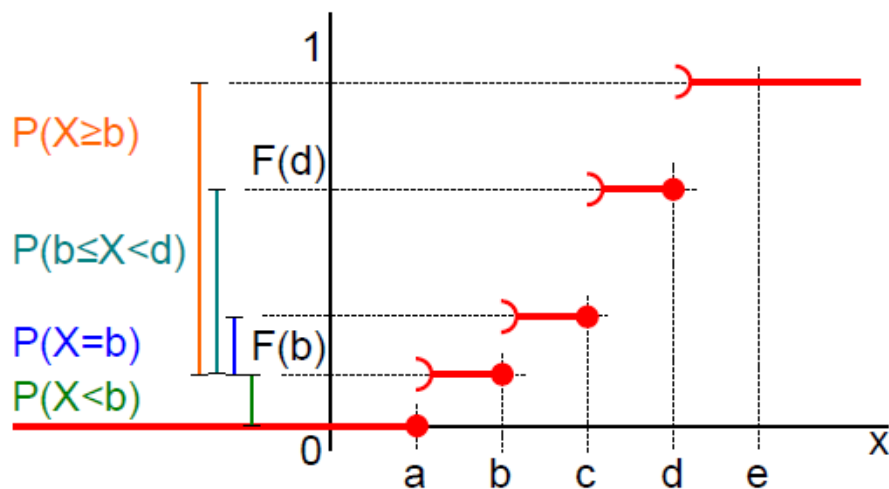
Vzťahy medzi pravdepodobnosťou a distribučnou funkciou sú nasledovné

$$\forall a < b; a, b \in \mathbb{R}$$

$$P(X \geq a) = 1 - F(a) = \int_a^{\infty} f(x) dx$$

$$P(a \leq X < b) = F(b) - F(a) = \int_a^b f(x) dx$$

$$P(X = a) = \begin{cases} \lim_{x \rightarrow a^+} F(x) - F(a), & \text{diskrétne premenné} \\ 0, & \text{spojité premenné.} \end{cases}$$



Distribučnú funkciu určíme ako

$$F(x) = \sum_{x_i < x} P(x_i)$$

$P(x)$ - pravdepodobnostná funkcia

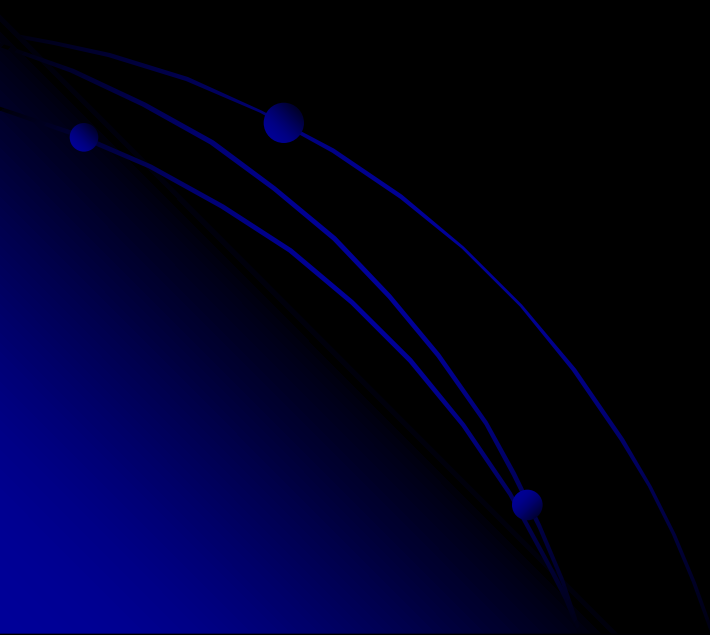
$$F(x) = \int_{-\infty}^x f(t) dt$$

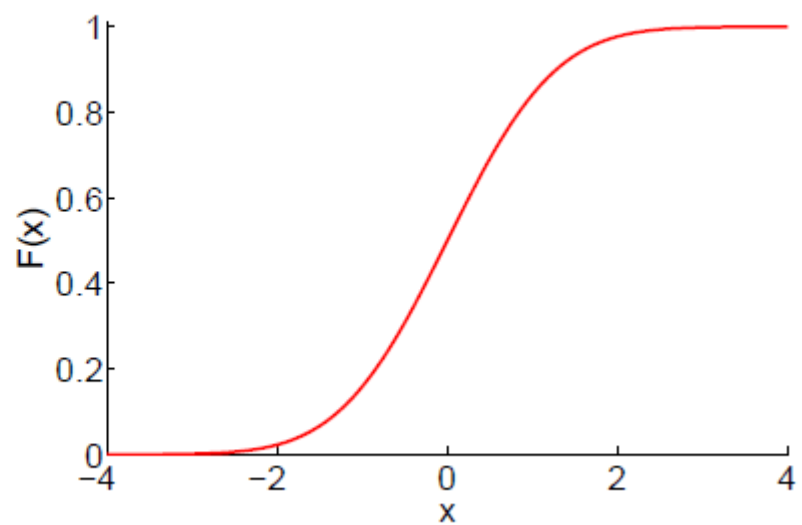
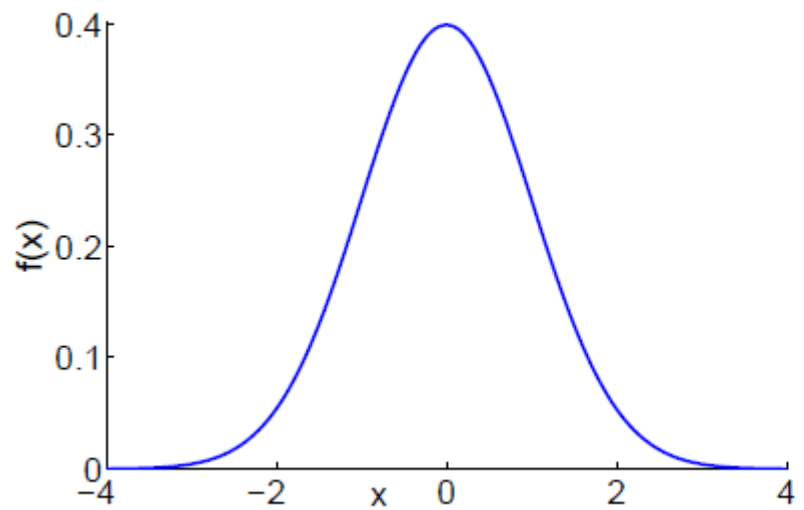
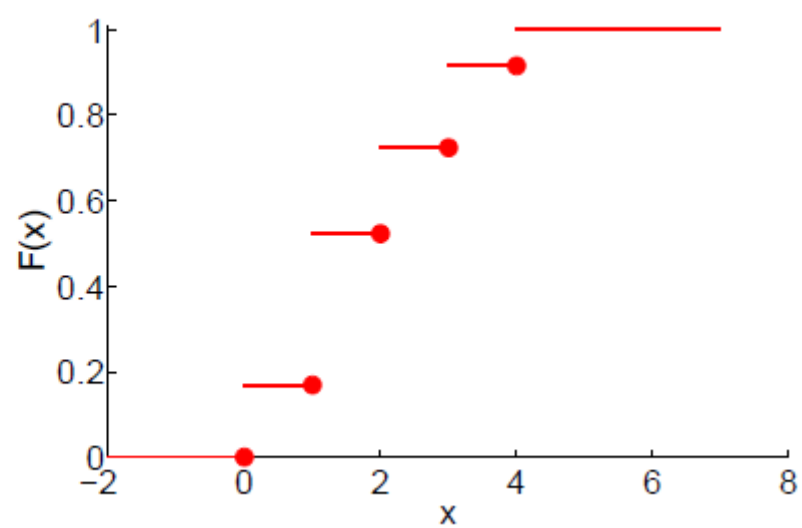
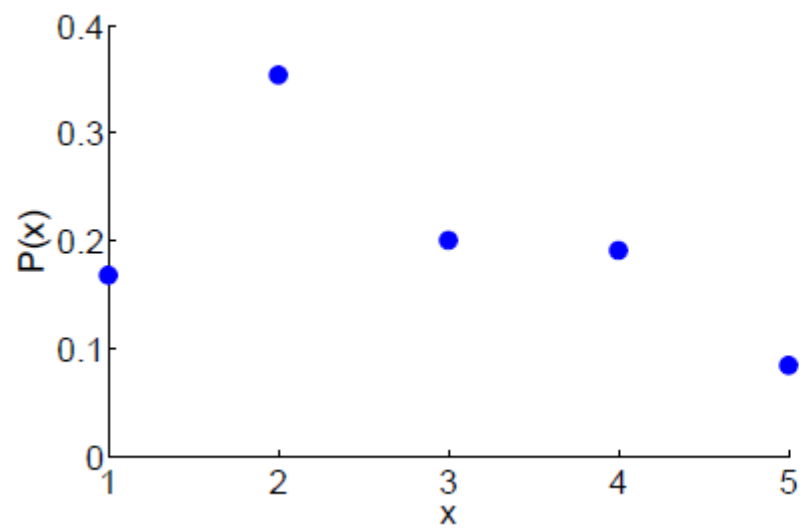
$f(t)$ - hustota rozdelenia pravdepodobnosti

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

$$\lim_{x \rightarrow -\infty} f(x) = 0$$

$$\lim_{x \rightarrow \infty} f(x) = 0$$





Diskrétny príklad

Máme dva stroje (nezávislé na sebe)

Jeden sa pokazí s pravdepodobnosťou 20%

Druhý s pravdepodobnosťou 30%

Náhodná premenná X označuje počet pokazených strojov.

Popíšte túto náhodnú premennú

$$\Omega = \{0,1,2\}$$

$$P(S1)=0,2$$

$$P(S2)=0,3$$

$$P(X=0) = P(S1^c \cap S2^c) = 0,8*0,7 = 0,56$$

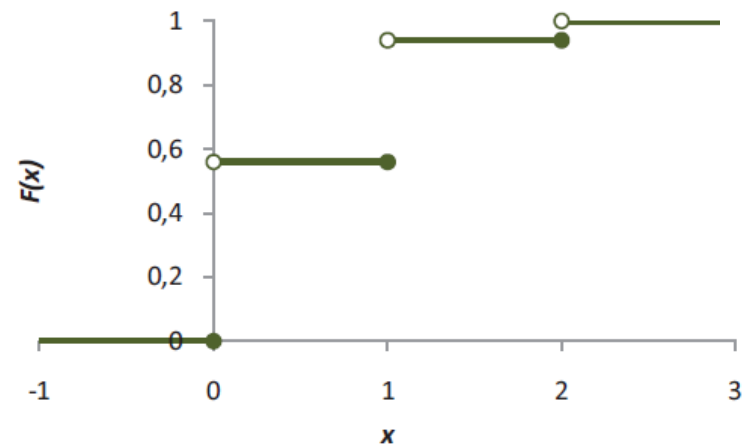
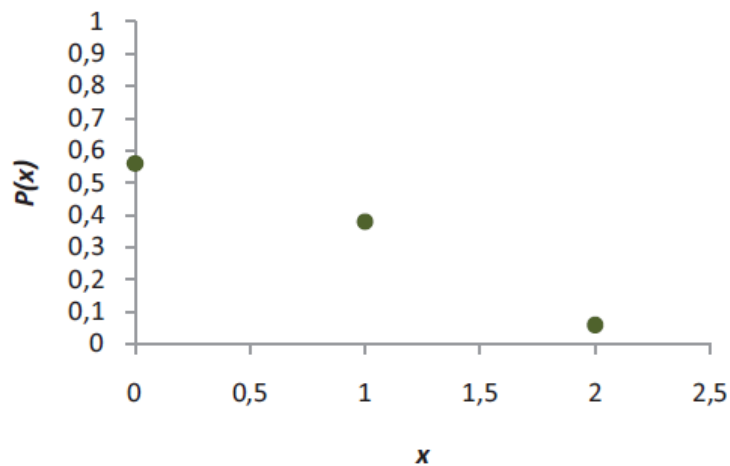
$$P(X=1) = P((S1 \cap S2^c) \cup (S1^c \cap S2)) = 0,2*0,7 + 0,8*0,3 = 0,38$$

$$P(X=2) = P(S1 \cap S2) = 0,2*0,3 = 0,06$$

$$F(x) = \sum_{x_i < x} P(x_i)$$

x_i	$P(X = x_i)$
0	0,56
1	0,38
2	0,06
Σ	1,00

x	$F(x)$
$(-\infty; 0)$	0
$(0; 1)$	0,56
$(1; 2)$	0,94
$(2; \infty)$	1



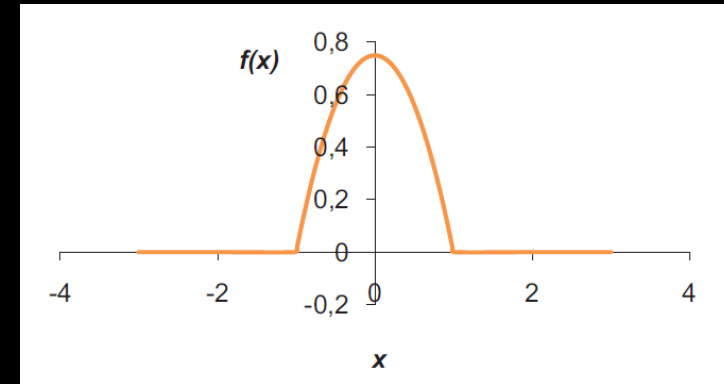
Distribučná funkcia je kumulatívna

Spojitéý prípad

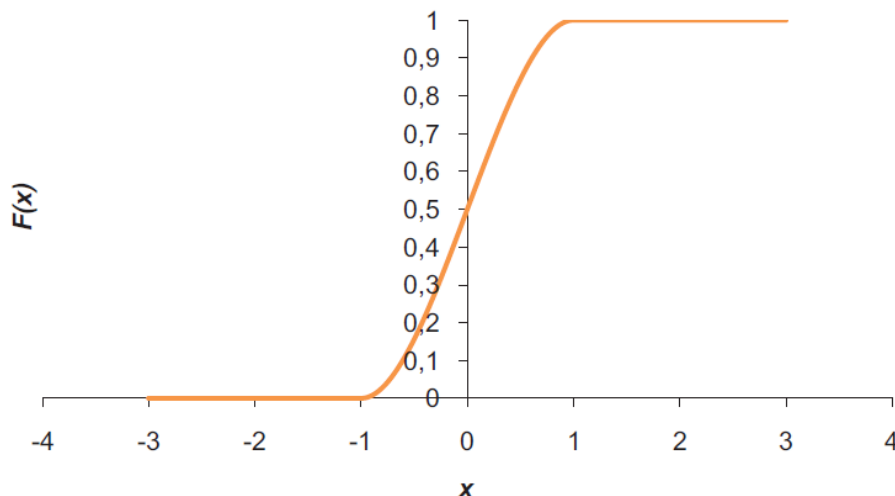
$$F(x) = \int_{-\infty}^x f(t) dt$$

Nech X je spojitá náhodná premenná definovaná hustotou pravdepodobnosti

$$f(x) = \begin{cases} 0,75(1-x)(1+x) = 0,75(1-x^2) & -1 < x < 1 \\ 0 & \text{inde.} \end{cases}$$



$$F(x) = \begin{cases} 0 & x \in (-\infty; -1) \\ \frac{1}{4}(-x^3 + 3x + 2) & x \in \langle -1; 1 \rangle \\ 1 & x \in \langle 1; \infty \rangle \end{cases}$$



Rozdelenia pravdepodobnosti

Diskrétne:

Alternatívne, Binomické, Hypergeometrické, Geometrické, Negatívne binomické (Pascalove), Poissonove

Spojité:

Rovnomerné, Exponenciálne, Weibullove, Gama - Erlangove, Triangulárne - Simpsonove, Cauchyho, Beta, Rayeleighove,
• Normálne, Normované normálne, Logaritmicko-normálne

Špecifické (výberové):

χ^2 (chí- kvadrát) rozdelenie, Studentovo t – rozdelenie, Fischerove F- rozdelenie

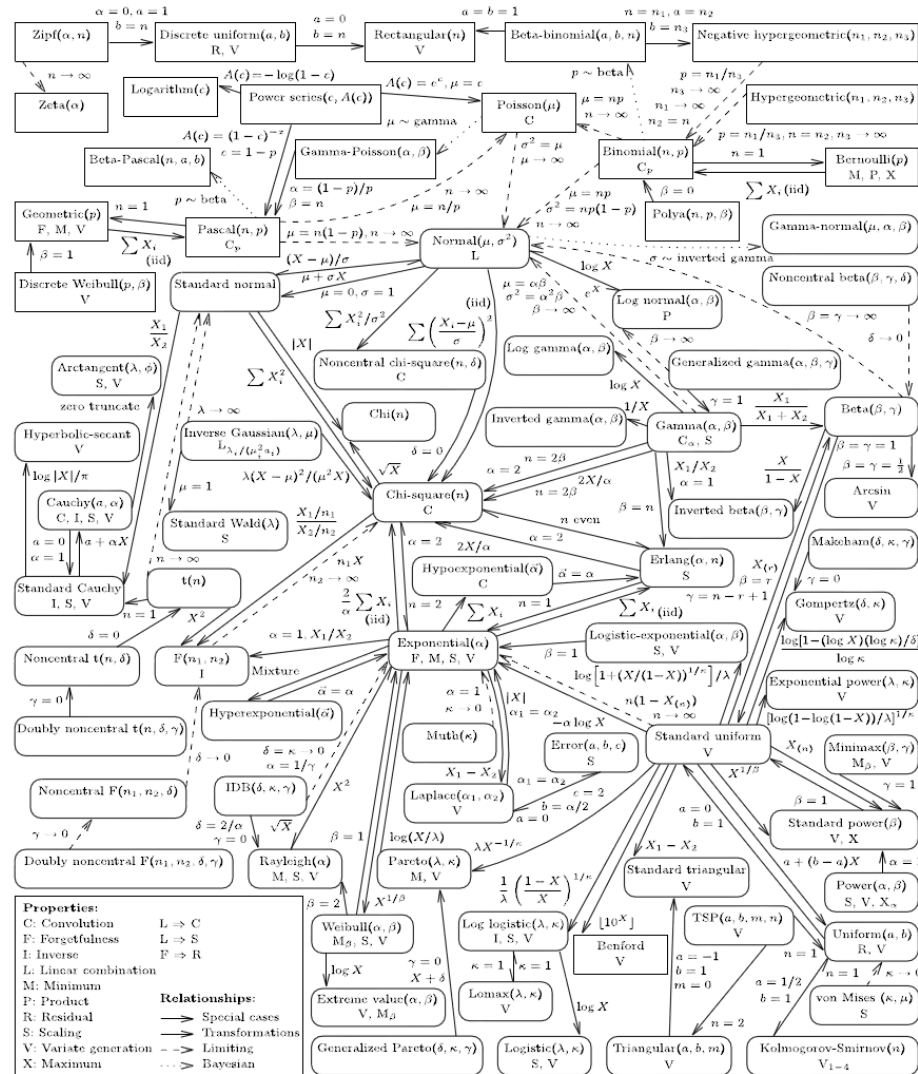


Figure 1. Univariate distribution relationships.

Binomické rozdelenie

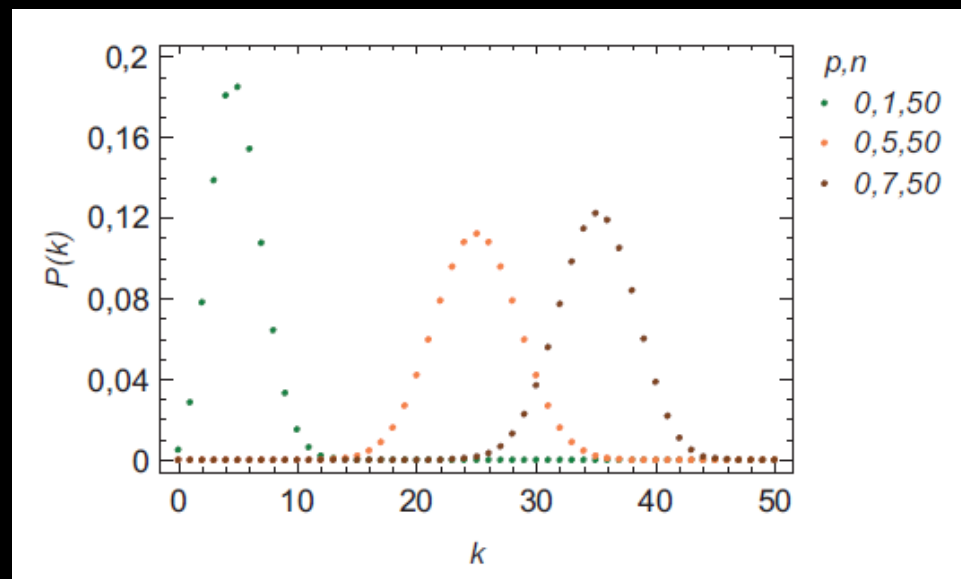
Bernoulliho pokusy – postupnosť **nezávislých** pokusov, kde pravdepodobnosť úspechu je p

n = celkový počet Bernoulliho pokusov

p = pravdepodobnosť úspechu v každom z pokusov

$$Bi(n, p)$$

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



Diskrétne rozdelenia

Alternatívne - binomické pre $n=1$

Hypergeometrické – modeluje závislé pokusy

Geometrické – počet Bernoulliho pokusov do výskytu prvého úspechu
(niekedy vrátane)

- Pascalove - počet Bernoulliho pokusov do výskytu k -teho úspechu
(niekedy vrátane)

- Poissonove - počet náhodných udalostí v nejakom pevnom časovom intervale

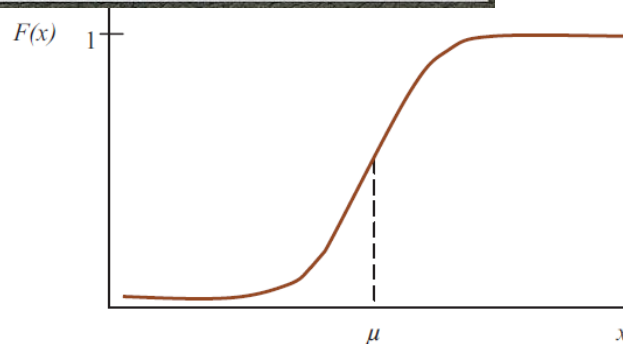
Normálne rozdelenie

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

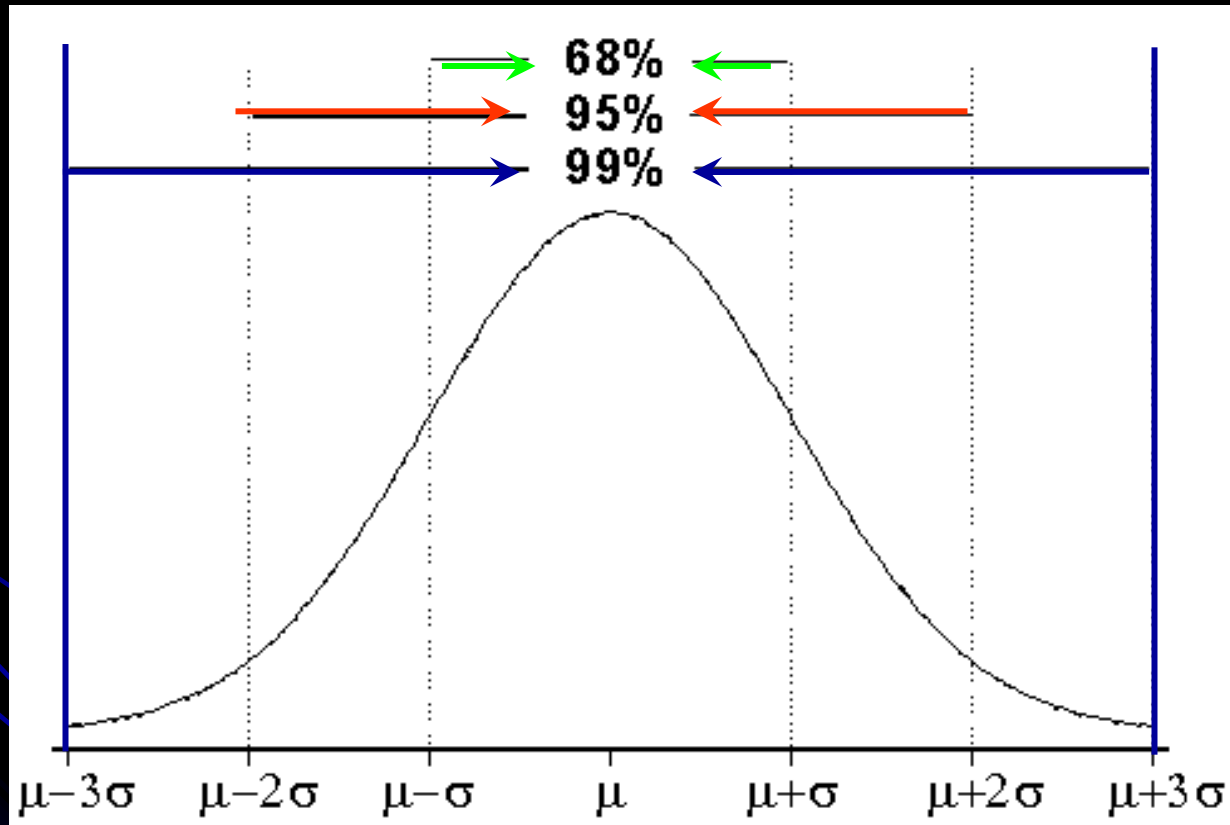
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



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



pravidlo 3σ



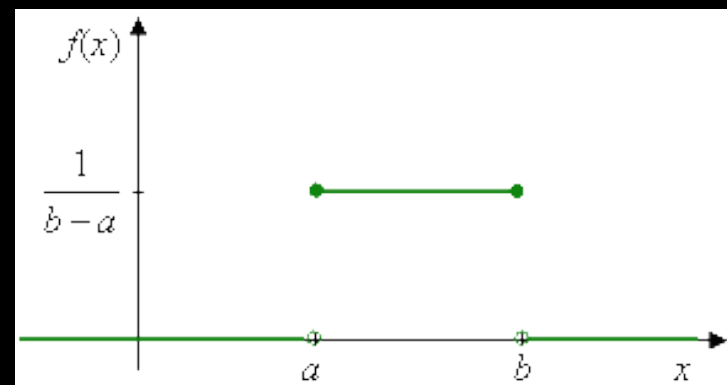
(68.26%, 95.45%, 99.73%)

Rovnomerné rozdelenie

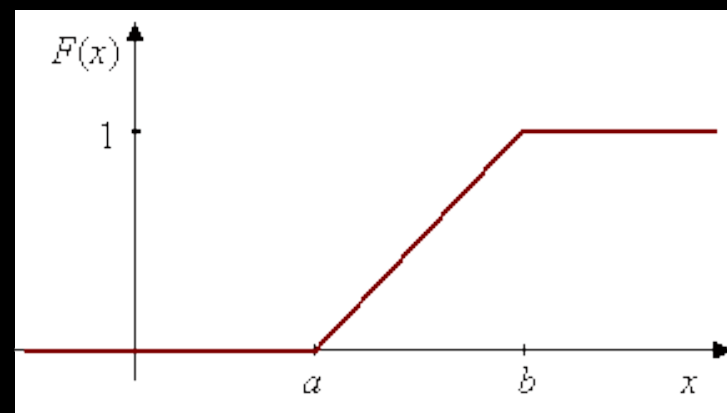
$R(a; b)$

interval (a, b)

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a; b) \\ 0 & x \notin (a; b) \end{cases}$$



$$F(x) = \begin{cases} 0 & x \in (-\infty; a) \\ \frac{x-a}{b-a}, & x \in (a; b) \\ 1 & x \in (b; \infty) \end{cases}$$



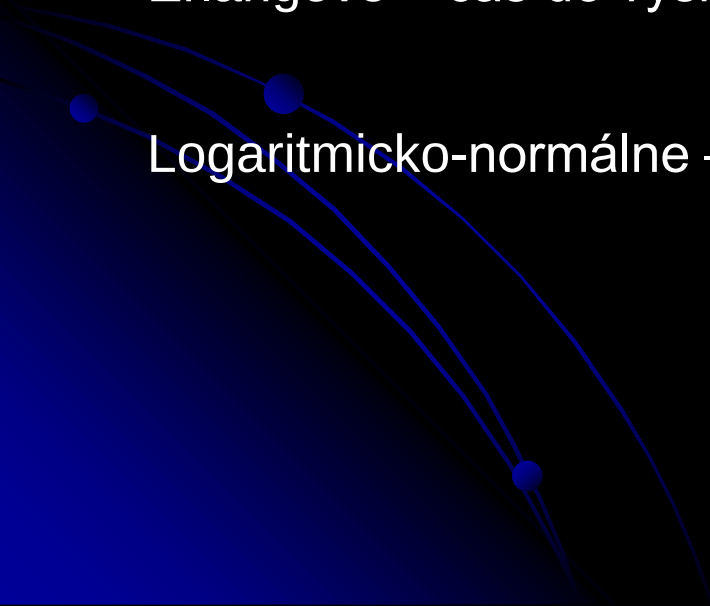
Spojité rozdelenia

Exponenciálne – čas do prvého výskytu udalosti v Poissonovom procese (čas medzi udalosťami)

Weibullove – čas do prvého výskytu udalosti, keď nevyhovuje exp.r.

Erlangove – čas do výskytu k-tej udalosti v Poissonovom procese

Logaritmicko-normálne – pre jednostranne ohraničené údaje



Špeciálne rozdelenia

$$\chi_n^2 - \quad X = X_1^2 + X_2^2 + \dots + X_n^2 \quad X_i \sim N(0,1)$$

$$\text{Studentove} - \frac{\frac{X}{\sqrt{Y}}}{\sqrt{n}}$$

$$X \sim N(0, 1) \text{ a } Y \sim \chi_n^2$$

$$\text{Fisherove} - \frac{\frac{X}{n}}{\frac{Y}{m}}$$

$$X \sim \chi_n^2, Y \sim \chi_m^2$$

Číselné charakteristiky náhodnej premennej

Charakterizujú vlastnosti náhodnej
premennej a umožňujú porovnávať
premenné:



- Momenty

- Kvantily

- ...

Momenty

Všeobecný moment r-tého rádu

$$\mu_r = E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx$$

Stredná hodnota $E(X)$ (r=1)

$$\mu = \int_{-\infty}^{\infty} x f(x) dx$$

Centrálny moment r-tého rádu

$$\mu'_r = E((X - E(X))^r) = \int_{-\infty}^{\infty} (x - E(X))^r f(x) dx$$

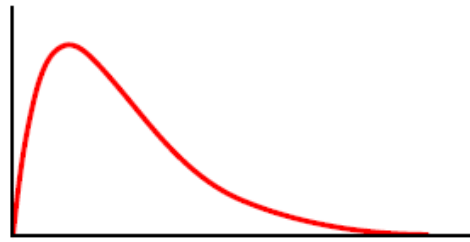
Rozptyl (r=2)

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$$

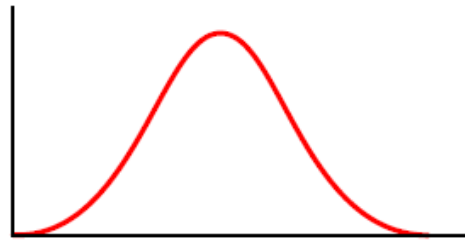
$$\text{Var}(X) = E(X^2) - (E(X))^2$$

Šikmosť

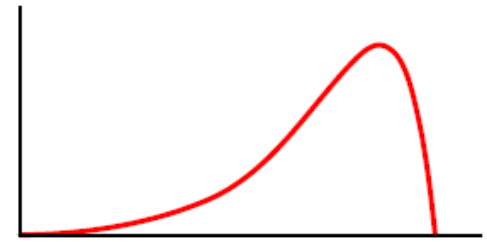
$$\alpha_3 = \frac{\mu'_3}{\sigma^3}$$



(a) Pozitívne zošikmené rozdelenie.



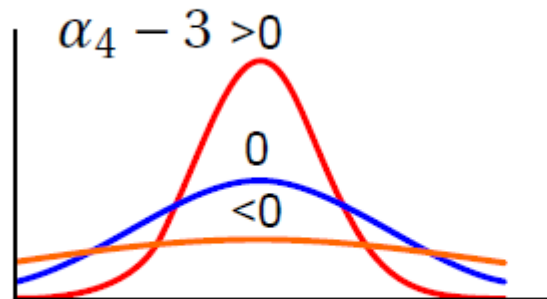
(b) Symetrické rozdelenie.



(c) Negatívne zošikmené rozdelenie.

Špicatosť

$$\alpha_4 = \frac{\mu'_4}{\sigma^4}$$



Normálna (0), menšia (< 0) a väčšia (> 0) špicatosť rozdelenia.

Kvantily

Kvantil Q_p je hodnota, pre ktorú platí, že pravdepodobnosť, že náhodná veličina X má hodnotu menšiu ako Q_p , je $100p\%$

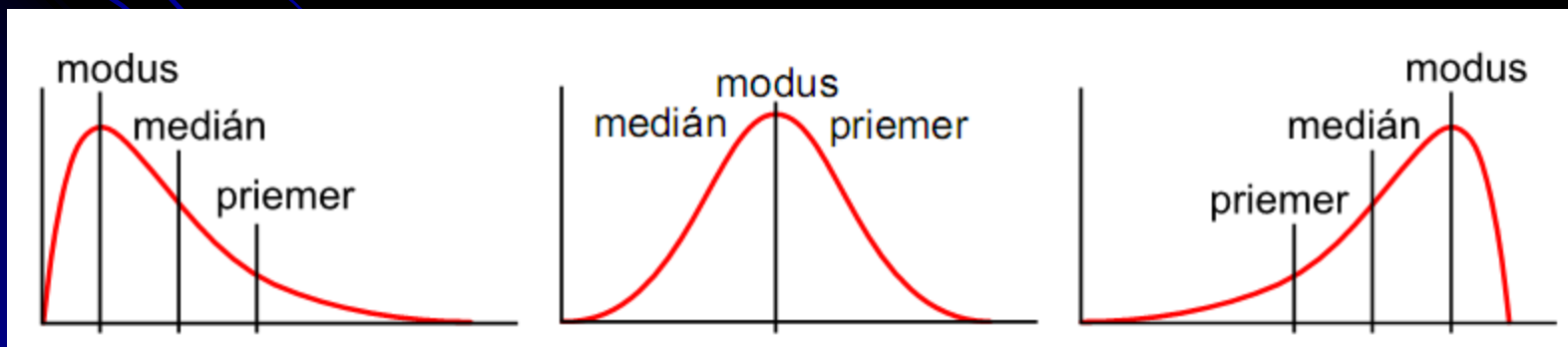
$$P(X < Q_p) = p$$

$$F(Q_p) = p$$

Medián

kvantil deliaci súbor na dve polovice

$$Q_{0,5}$$



Kvartil

$Q_{0,25}$ (dolný kvartil), $Q_{0,5}$ a $Q_{0,75}$ (horný kvartil)

Medzikvartilové rozpätie

Decil

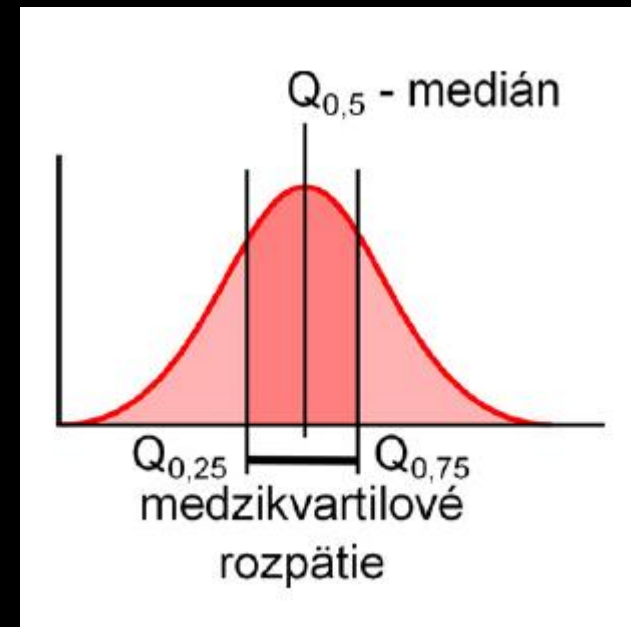
$Q_{0,1}, Q_{0,2}, \dots, Q_{0,9}$

Medzidecilové rozpätie

Percentil

$Q_{0,01}, Q_{0,02}, \dots, Q_{0,99}$

Medzipercentilové rozpätie



Viacrozmerný prípad

(X, Y) - dvojrozmerná náhodná premenná

Združená distribučná funkcia $F(x, y) = P(X < x, Y < y)$

pravdepodobnosť súčasného nastatia dvoch udalostí

Marginálna distribučná funkcia

$$F_1(x) = P(X < x) = P(X < x, Y < \infty) = F(x, \infty)$$

Podmienená distribučná funkcia

$$F(x|y) = P(X < x|Y = y)$$

Nezávislosť premenných

$$F(x, y) = F_1(x) \cdot F_2(y)$$

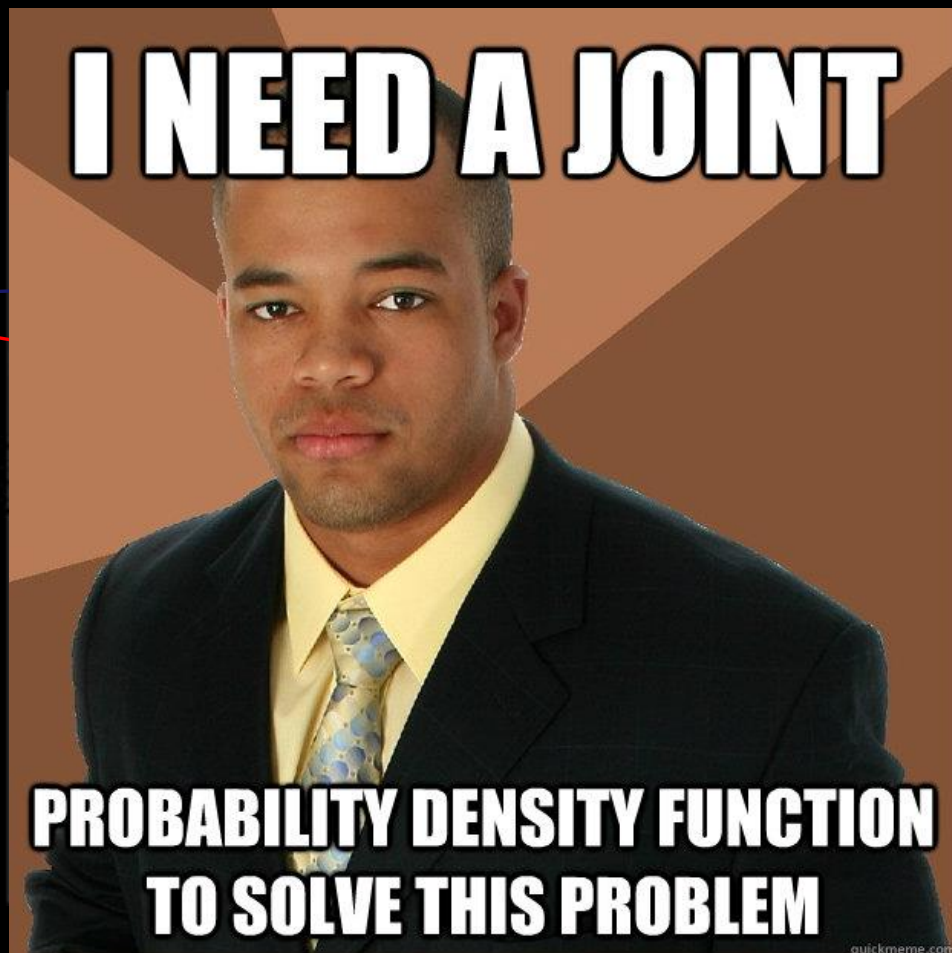
$f(x, y)$ – združená hustota rozdelenia pravdepodobnosti

podmienená

marginálna

$$f(x, y) = f(x|y)f(y) = f(y|x)f(x)$$

$$f(x, y) = f(x)f(y)$$



Charakteristiky

Vektor stredných hodnôt $E(\mathbf{X}) = (E(X_1), \dots, E(X_N))$

Vektor rozptylov $\text{Var}(\mathbf{X}) = (\text{Var}(X_1) \dots \text{Var}(X_N))$

Obečný združený moment $(r_1, \dots, r_N)$ -tého rádu

$$\begin{aligned}\mu_{r_1, \dots, r_N} &= E(X_1^{r_1} X_2^{r_2} \dots X_N^{r_N}) = \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1^{r_1} \dots x_N^{r_N} f(x_1, \dots, x_N) dx_1 \dots dx_N\end{aligned}$$

Centrálny združený moment $(r_1, \dots, r_N)$ -tého rádu

$$\begin{aligned}\mu'_{r_1, \dots, r_N} &= E((X_1 - E(X_1))^{r_1} (X_2 - E(X_2))^{r_2} \dots (X_N - E(X_N))^{r_N}) = \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_1 - E(X_1))^{r_1} \dots (x_N - E(X_N))^{r_N} f(x_1, \dots, x_N) dx_1 \dots dx_N\end{aligned}$$

Momenty používané v 2D počítačovom videní

škálovo a translačne invariantné momenty

$$\eta_{i,j} = \frac{\mu'_{i,j}}{\mu'_{0,0} \left(1 + \frac{i+j}{2}\right)}$$

škálovo, translačne aj rotačne invariantné

$$I_1 = \eta_{2,0} + \eta_{0,2}$$

$$I_2 = (\eta_{2,0} - \eta_{0,2})^2 + 4\eta_{1,1}^2$$

....

Kovariancia

$$\text{Cov}(X_i, X_j) = E((X_i - E(X_i))(X_j - E(X_j)))$$

$$\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$$

$$\text{Cov}(X_i, X_i) = E(X_i^2) - (E(X_i))^2 = \text{Var}(X_i)$$

$$\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i).$$

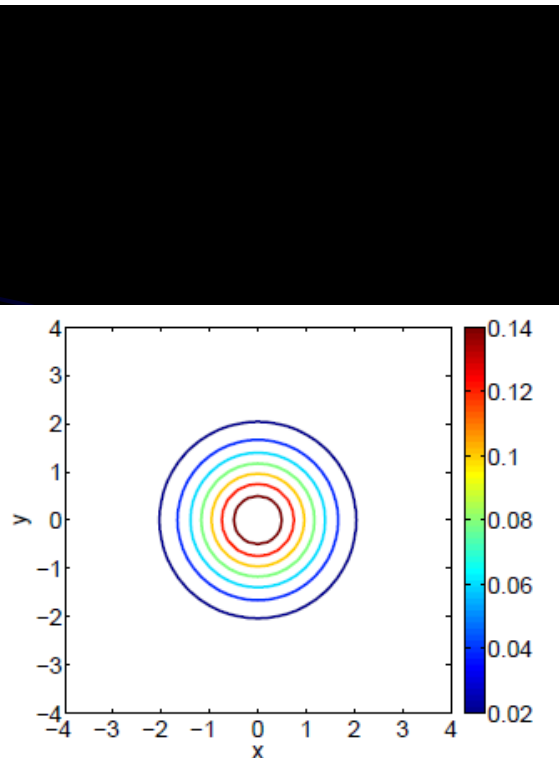
Kovariančná matica

$$\Sigma = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_N) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_N) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_N, X_1) & \text{Cov}(X_N, X_2) & \cdots & \text{Var}(X_N) \end{bmatrix}$$

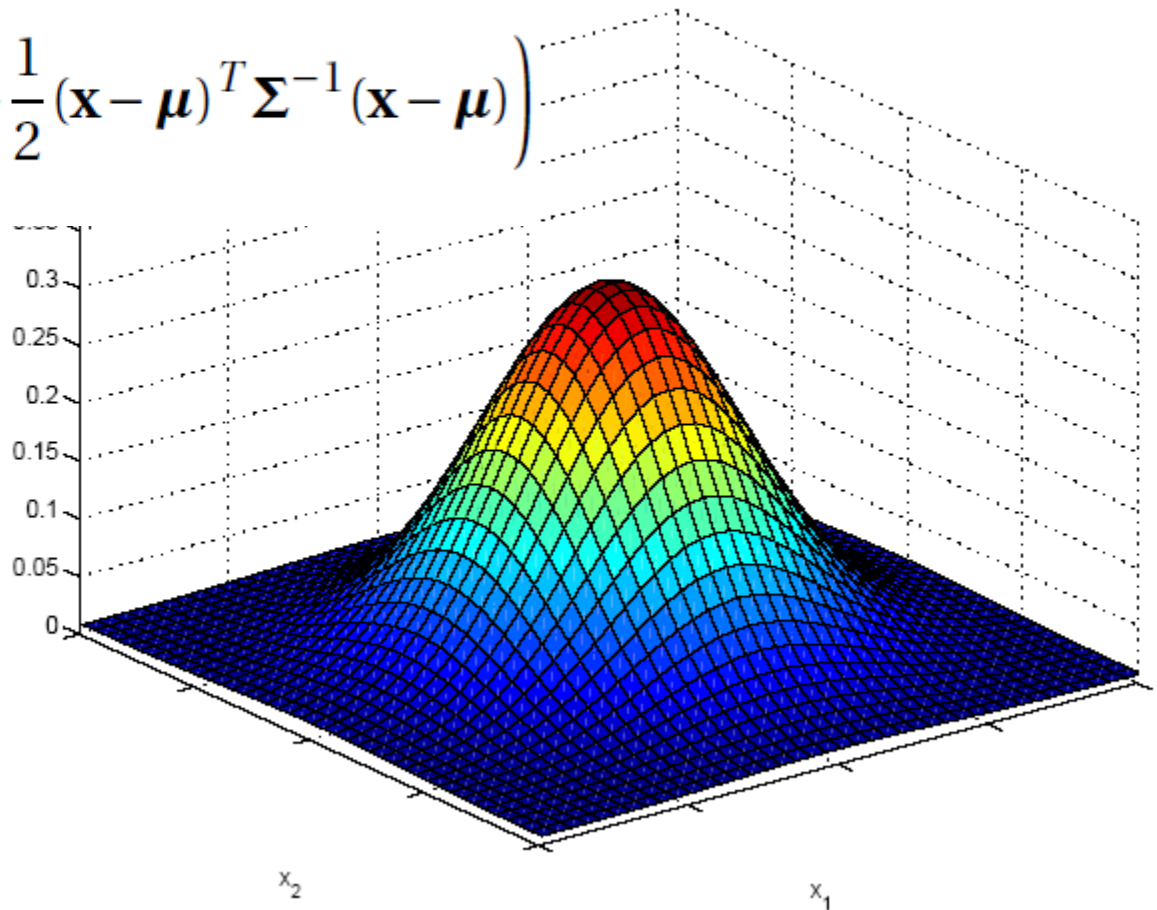
Normálne rozdelenie

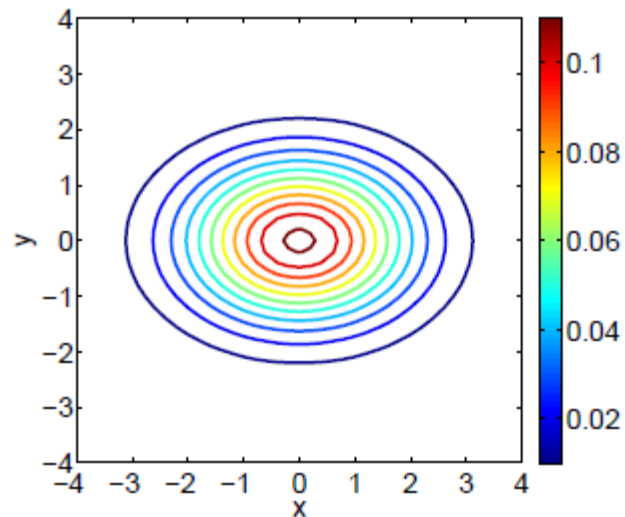
$$\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$$

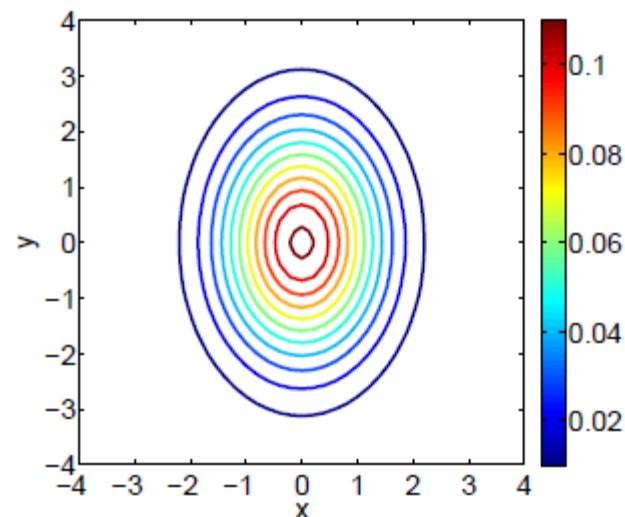


$$\text{Var}(X) = \text{Var}(Y) = 1, \text{Cov}(X, Y) = 0$$

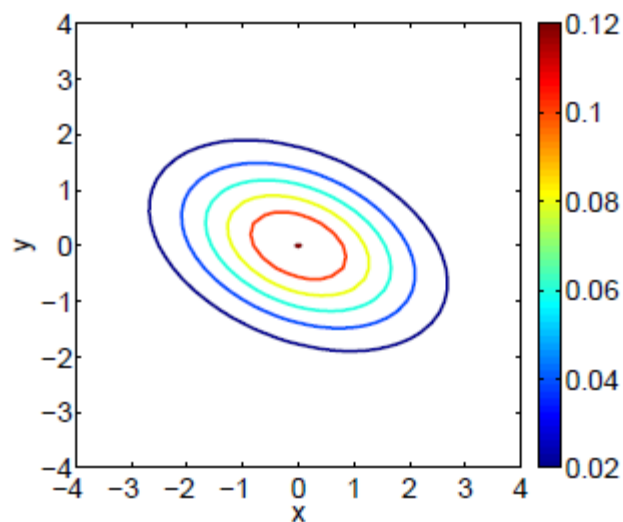




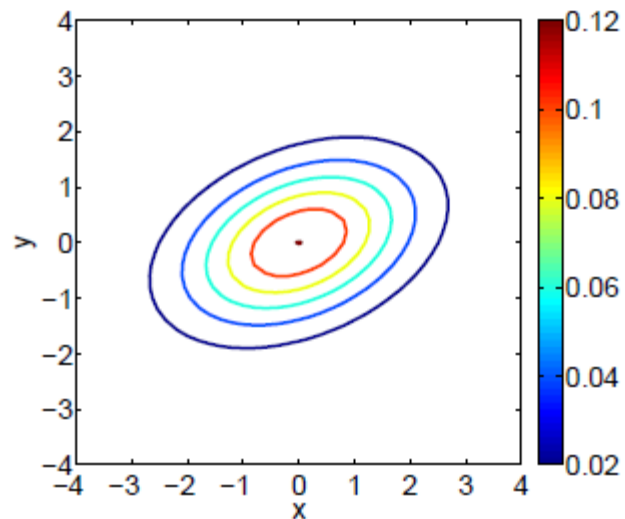
$\text{Var}(X) = 2, \text{Var}(Y) = 1, \text{Cov}(X, Y) = 0$



$\text{Var}(X) = 1, \text{Var}(Y) = 2, \text{Cov}(X, Y) = 0$



$\text{Var}(X) = 2, \text{Var}(Y) = 1, \text{Cov}(X, Y) = -0,5$



$\text{Var}(X) = 2, \text{Var}(Y) = 1, \text{Cov}(X, Y) = 0,5$

Štatistika

Odhad parametrov rozdelení

Testovanie hypotéz

Náhodný výber je náhodný vektor $(X_1, X_2, \dots, X_n)$, pre ktorý platí, že náhodné premenné X_i majú rovnaké rozdelenie pravdepodobnosti a sú navzájom nezávislé

Súbor všetkých jednotiek, ktoré skúmame, nazývame základný súbor

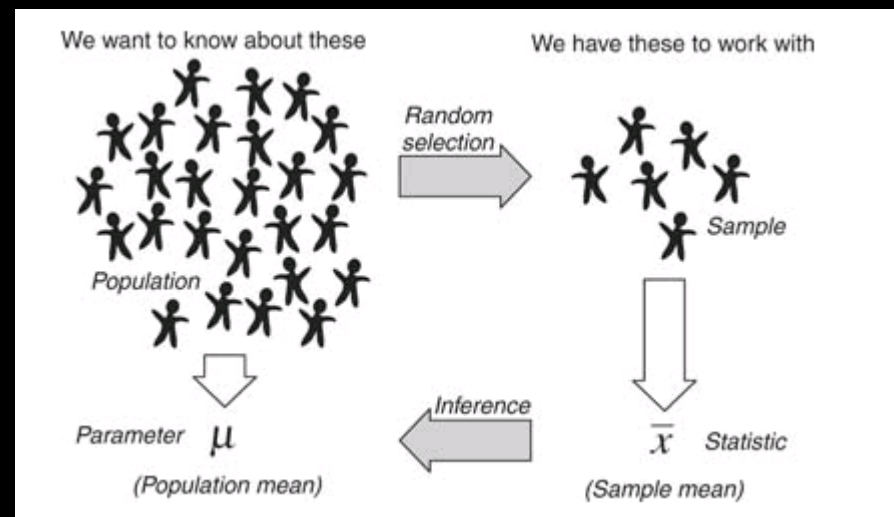
Súbor vybraných jednotiek nazývame výberový súbor

Parameter rozdelenia – popisuje populáciu –
nepoznáme hodnotu

Štatistika (odhad parametra, výberová
charakteristika) – vypočítame z náhodného
výberu

$E(X)$

$\bar{X}$



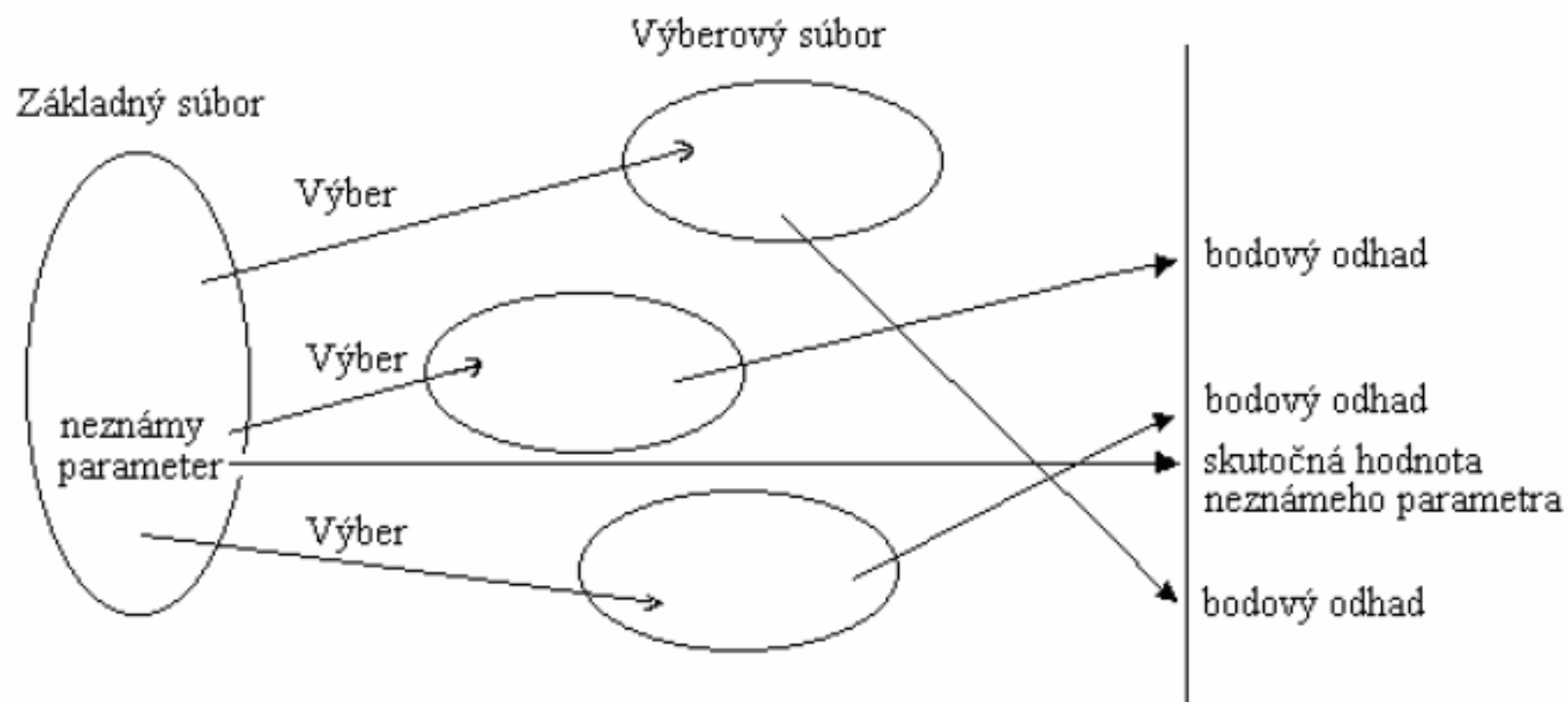
Bodové odhady

Výberový priemer $\bar{X}$: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

Výberový rozptyl S^2 : $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

Výberová smerodajná odchýlka S : $S = \sqrt{S^2}$

Výberová kovariancia S_{XY} : $S_{xy} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n(n-1)}$



Intervalový odhad

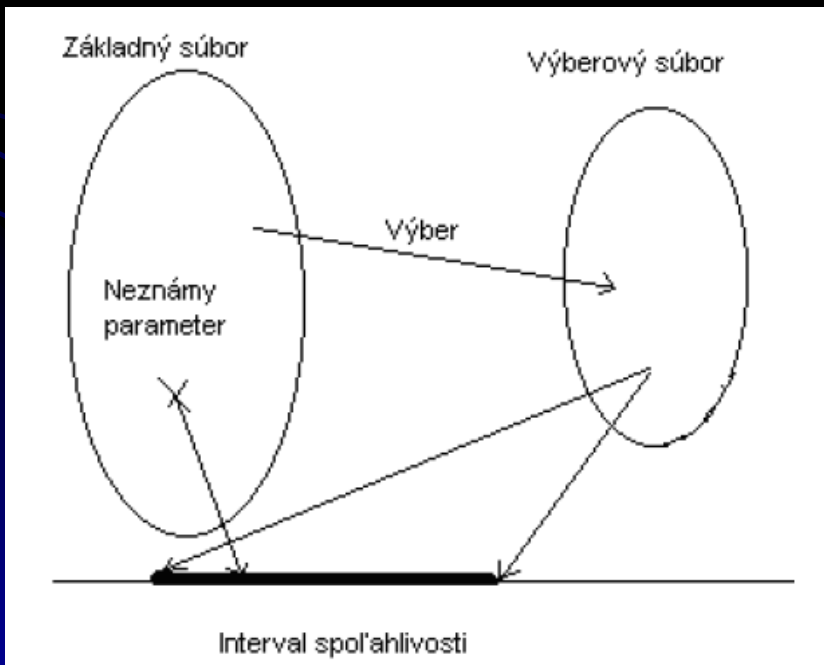
Interval spolehlivosti

$$P(G_D < \theta < G_H) = 1 - \alpha$$

(G_D, G_H) – interval spolehlivosti

$1 - \alpha$ – koeficient spolehlivosti

θ – skutečná hodnota parametra

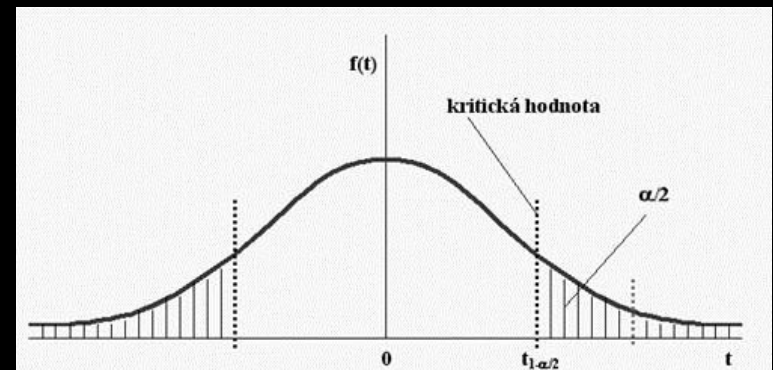


$$P(G_D < \theta < G_H) = 1 - \alpha$$

Príklad – intervalový odhad strednej hodnoty normálneho rozdelenia, keď poznáme hodnotu σ

Štatistika $U = \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} \sim N(0,1)$

Kritická hodnota rozdelenia $N(0,1)$



$$\begin{aligned} 1 - \alpha &= P(-u_\alpha \leq U \leq u_\alpha) = P\left(-u_\alpha \leq \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} \leq u_\alpha\right) = \\ &= P\left(\bar{X} - u_\alpha \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + u_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right). \end{aligned}$$

G_D

G_H

príklad

náhodný výber 100 výrobkov

priemerná váha 2,2 kg

smerodajná odchýlka váhy výrobkov $\sigma = 0,6$ kg

Určite 95-percentný obojstranný interval spoľahlivosti strednej hodnoty váhy

$$\left\langle \bar{x} - u_{\alpha} \cdot \frac{\sigma}{\sqrt{n}}; \bar{x} + u_{\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right\rangle = \left\langle 2,2 - 1,96 \cdot \frac{0,6}{10}; 2,2 + 1,96 \cdot \frac{0,6}{10} \right\rangle = \langle 2,08; 2,31 \rangle$$

$$X \sim N(0,1), P(|X| > u_{\alpha}) = \alpha$$

α	0,01	0,02	0,05	0,1	0,2
u_{α}	2,5758	2,3263	1,9599	1,6448	1,299