MV011 Statistika I

3. Náhodná veličina

Jan Koláček (kolacek@math.muni.cz)

Ústav matematiky a statistiky, Přírodovědecká fakulta, Masarykova univerzita, Brno

Motivační příklad

Příklad 1

Házíme opakovaně mincí. Ze 100 náhodných pokusů: 56 × „hlava" a 44 × „orel".
Otázka: *je tato mince „spravedlivá"?*

prostor elementárních jevů: $\Omega = \{$„hlava";„orel"$\}$
elementární jevy: $\omega_1 =$„hlava", $\omega_2 =$„orel"
jevová $\sigma-$ algebra: $\mathcal{A} = \{\emptyset, \omega_1, \omega_2, \Omega\}$

Nás však zajímá např. **počet hlav ve 100 pokusech**

$$\{\text{„hlava"}\} \in \Omega \overset{X}{\mapsto} 1 \in \mathbb{R}$$

$$\{\text{„orel"}\} \in \Omega \overset{X}{\mapsto} 0 \in \mathbb{R}$$

X je **zobrazení**, číslo 1 se nazývá jeho **realizace**
Otázka: Jaké jsou pravděpodobnosti jednotlivých hodnot zobrazení X?

Náhodná veličina

Definice 1

Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, $X : \Omega \to \mathbb{R}$ je takové zobrazení, že pro $\forall\, x \in \mathbb{R}$ platí

$$\{\omega \in \Omega : X(\omega) \leq x\} = X^{-1}((-\infty, x\rangle) \in \mathcal{A}$$

Pak X nazýváme **náhodnou veličinou** (**random variable**) (vzhledem k jevovému poli $(\Omega, \mathcal{A})$).

Pravděpodobnostní chování náhodné veličiny lze popsat pomocí distribuční funkce.

Definice 2

Nechť X je náhodná veličina definovaná na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$. Pak funkci $F(x) = P(X \leq x)$, kde $x \in \mathbb{R}$, nazýváme **distribuční funkcí** (**cumulative distribution function**) náhodné veličiny X.

Poznámka 3

Zjednodušené značení: $P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\})$

Distribuční funkce

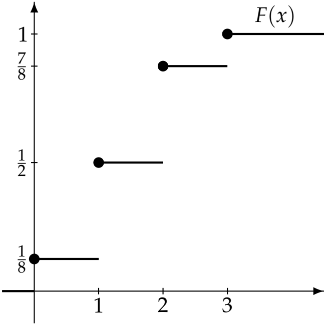
Příklad 2 (3 nezávislé hody mincí)

$\Omega = \{\omega_1 = (H,H,H), \omega_2 = (H,H,O), \omega_3 = (H,O,H), \omega_4 = (O,H,H), \omega_5 = (O,O,H), \omega_6 = (O,H,O), \omega_7 = (H,O,O), \omega_8 = (O,O,O)\}$.

Jevová σ-algebra: $\mathcal{A} = 2^{\Omega}$.

X . . . „*počet hlav ve třech hodech*“ $\Rightarrow X \in \{0,1,2,3\}$.

Distribuční funkce:



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

$$\begin{aligned}
F(x) &= P(\emptyset) = 0 && x < 0\\
&= P(X=0) = \tfrac{1}{8} && 0 \le x < 1\\
&= P(X=0 \vee 1) = \tfrac{1}{8} + \tfrac{3}{8} = \tfrac{1}{2} && 1 \le x < 2\\
&= P(X=0 \vee 1 \vee 2) = \tfrac{1}{8} + \tfrac{3}{8} + \tfrac{3}{8} = +\tfrac{7}{8} && 2 \le x < 3\\
&= P(X=0 \vee 1 \vee 2 \vee 3) = \tfrac{1}{8} + \tfrac{3}{8} + \tfrac{3}{8} + \tfrac{1}{8} = 1 && x \ge 3
\end{aligned}$$

Distribuční funkce

Věta 4 (Vlastnosti distribuční funkce)

Nechť $F(x)$ *je distribuční funkce náhodné veličiny* X *definované na* $(\Omega, \mathcal{A}, P)$. *Pak*

- F *je neklesající.*
- F *je zprava spojitá.*
- $\lim\limits_{x\to\infty} F(x) = 1$ *a* $\lim\limits_{x\to-\infty} F(x) = 0$.
- $0 \leq F(x) \leq 1$ *pro* $x \in \mathbb{R}$.
- $P(X = x) = F(x) - \lim\limits_{y\to x^-} F(y)$.
- F *má nejvýše spočetně mnoho bodů nespojitosti.*
- $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$ *pro* $x_1, x_2 \in \mathbb{R}$, $x_1 < x_2$.

Náhodná veličina

Příklad 3

Házíme opakovaně mincí.
X_1 ...*„počet hlav v 10 pokusech"*
X_2 ...*„čas, který stráví mince ve vzduchu, než spadne"*

Obecně

$$X \begin{cases} \textbf{diskrétní} \text{ náhodná veličina – max. } \textbf{spočetně} \text{ mnoho hodnot} \\ \qquad \text{např. } X \in \mathbb{N}_0 \\ \\ \textbf{spojitá} \text{ náhodná veličina – } \textbf{nespočetně} \text{ mnoho hodnot} \\ \qquad \text{např. } X \in (0, \infty) \end{cases}$$

Diskrétní náhodná veličina

Definice 5

Řekneme, že náhodná veličina X je **diskrétního typu** (**discrete**), pokud existuje nejvýše spočetná množina $M \subset \mathbb{R}$ taková, že platí $P(X \in M) = 1$.

Definice 6

Nechť X je diskrétní náhodná veličina. Pak funkci $p(x) = P(X = x)$, $x \in M$, nazýváme **pravděpodobnostní funkcí** (**probability distribution function**) diskrétní náhodné veličiny X a množinu M **oborem hodnot** X.

Poznámka 7

Pravděpodobnostní funkci lze definovat pro všechna reálná čísla, když položíme $p(x) = 0$ *pro* $x \notin M$.

Značení: Fakt, že jde o diskrétní náhodnou veličinu budeme značit $X \sim (M, p)$.

Diskrétní náhodná veličina

Věta 8 (Vlastnosti pravděpodobnostní funkce)

Nechť $X \sim (M, p)$. *Pak*

- $p(x) \geq 0$ *pro* $\forall\, x \in \mathbb{R}$ *a* $\sum\limits_{x \in M} p(x) = 1$.
- $P(X \in B) = \sum\limits_{x \in M \cap B} p(x)$ *pro libovolné* $B \in \mathcal{B}$.
- $F(x) = \sum\limits_{t \in M, t \leq x} p(t)$ *pro* $\forall\, x \in \mathbb{R}$.
- $p(x) = F(x) - \lim\limits_{y \to x^-} F(y)$ *pro* $\forall\, x \in \mathbb{R}$.

Příklady

Příklad 4 (Alternativní rozdělení (Alternative distribution))

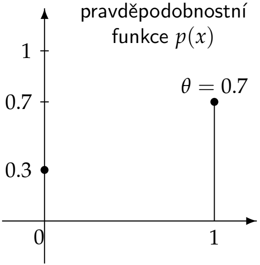
Uvažujme náhodný pokus, který může skončit s pravděpodobností $\theta \in (0,1)$ „úspěchem“ a s pravděpodobností $1-\theta$ „neúspěchem“.

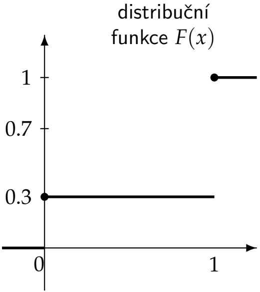
Prostor elementárních jevů: $\Omega = \{\omega_1, \omega_2\}$
σ-algebra náhodných jevů: $\mathcal{A} = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \Omega\}$
PST: $P(\emptyset) = 0$, $P(\omega_1) = 1-\theta$, $P(\omega_2) = \theta$ a $P(\Omega) = 1$.

Náhodná veličina: $X(\omega_1) = 0$ (neúspěch),

$X(\omega_2) = 1$ (úspěch).

Příklad





Diskrétní náhodná veličina s definičním oborem $M = \{0, 1\}$ a pravděpodobnostní funkcí

$$p(x) = \begin{cases} \theta & x = 1 \\ 1 - \theta & x = 0 \\ 0 & \textit{jinak} \end{cases} = \begin{cases} \theta^x (1-\theta)^{1-x} & x = 0, 1 \\ 0 & \textit{jinak}. \end{cases}$$

Náhodnou veličinu značíme $X \sim A(\theta)$.

Příklad

Příklad 5 (Binomické rozdělení (Binomial distribution))

Uvažujme posloupnost n nezávislých alternativních pokusů typu úspěch/neúspěch s pravděpodobností úspěchu $\theta \in (0,1)$ pro každý pokus.
X je náhodná veličina udávající **počet úspěchů v n pokusech**.

Obor hodnot náhodné veličiny X:

$$M = \{0, 1, \dots, n\}$$

a pravděpodobnostní funkce

$$p(x) = \begin{cases} \binom{n}{x}\theta^x(1-\theta)^{n-x} & x = 0, 1, \dots, n,\ n \in \mathbb{N},\ \theta \in (0,1) \\ 0 & \text{jinak}. \end{cases}$$

Náhodnou veličinu značíme $\boxed{X \sim Bi(n, \theta)}$.

Příklad

Příklad 6

Basketbalista hází trestný hod (šestku) s pravděpodobností úspěchu 0,9. Určete pravděpodobnost, že z pěti hodů:
a) dá 5 košů
b) dá alespoň dva koše
c) dá nejvýše dva koše

X ... počet vstřelených košů z pěti pokusů, $X \sim Bi(5; 0,9)$

$$\text{a)} \quad P(X = 5) = \binom{5}{5} 0,9^5 0,1^0 = 0,59$$

$$\text{b)} \quad P(X \geq 2) = 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1))$$

$$= 1 - (\binom{5}{0} 0,9^0 0,1^5 + \binom{5}{1} 0,9^1 0,1^4) = 0,99954$$

$$\text{c)} \quad P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= \binom{5}{0} 0,9^0 0,1^5 + \binom{5}{1} 0,9^1 0,1^4 + \binom{5}{2} 0,9^2 0,1^3 = 0,00856$$

Příklad

Příklad 7 (Poissonovo rozdělení)

Jestliže $M = \{0, 1, 2, \dots\}$ *a pravděpodobnostní funkce je tvaru*
$$p(x) = \begin{cases} e^{-\lambda}\frac{\lambda^x}{x!} & x = 0, 1, 2, \dots, \quad \lambda > 0 \\ 0 & jinak \end{cases}$$
, pak značíme $X \sim Po(\lambda)$.

Poissonovo rozdělení popisuje výskyt řídkých jevů za určitou jednotku času, prostoru apod. Parametr λ *značí očekávaný (průměrný) počet výskytů za jednotku. Jako příklad můžeme uvést*

- *počet organismů v jednotce půdy*
- *počet listí na stromech*
- *počet havárií za časovou jednotku (den, týden, měsíc, rok, ...)*
- *počet hovorů v telefonní síti za časovou jednotku*

Příklad

Příklad 8

Na server přijde během hodiny průměrně 120 požadavků.
Jaká je pravděpodobnost, že během 2 minut, po které je server restartován:
a) nepřijde žádný požadavek,
b) přijdou více jak 3 požadavky,
c) přijdou více jak 3 požadavky, ale méně než 7 požadavků.

X ... počet požadavků během 2 minut, 120 požadavků za hodinu $\Rightarrow$ 4 požadavky za 2 minuty $\Rightarrow X \sim Po(4)$

$$\begin{aligned}
\text{a)} \qquad P(X=0) &= e^{-4}\tfrac{4^0}{0!} = 0,0183 \\
\text{b)} \qquad P(X>3) &= 1 - P(X \leq 3) \\
&= 1 - e^{-4}\left(\tfrac{4^0}{0!} + \tfrac{4^1}{1!} + \tfrac{4^2}{2!} + \tfrac{4^3}{3!}\right) = 0,5665 \\
\text{c)} \quad P(3<X<7) &= P(X=4) + P(X=5) + P(X=6) \\
&= e^{-4}\left(\tfrac{4^4}{4!} + \tfrac{4^5}{5!} + \tfrac{4^6}{6!}\right) = 0,4558
\end{aligned}$$

Příklad

Příklad 9 (Geometrické rozdělení)

Uvažujme nekonečnou posloupnost nezávislých alternativních pokusů typu úspěch/neúspěch s pravděpodobností úspěchu $\theta \in (0,1)$ pro každý pokus. Náhodná veličina X udává **počet neúspěchů před prvním úspěchem***.*
Definiční obor náhodné veličiny: $M = \{0, 1, 2, \ldots\}$.
Pravděpodobnostní funkce je tvaru

$$p(x) = \begin{cases} (1-\theta)^x \theta & x = 0, 1, 2, \ldots, \quad \theta \in (0,1) \\ 0 & jinak \end{cases}$$ *a značíme* $\boxed{X \sim Ge(\theta)}$

Příklad

Příklad 10

Pravděpodobnost narození dívky je 0,49. Jaká je pravděpodobnost, že dívka se narodí až jako třetí?

X ... počet narozených chlapců před první dívkou, $X \sim Ge(0,49)$

$$P(X = 2) = (1 - 0,49)^2 \cdot 0,49 = 0,127$$

Spojitá náhodná veličina

Definice 9

Řekneme, že náhodná veličina X definovaná na $(\Omega, \mathcal{A}, P)$ je **absolutně spojitého typu** (**continuous**), jestliže existuje nezáporná integrovatelná funkce f taková, že rozdělení pravděpodobností

$$P_X(B) = \int_B f(x)dx \qquad \text{pro každé} \qquad B \in \mathcal{B}.$$

Funkci f nazýváme **hustotou rozdělení pravděpodobností** (**density**) náhodné veličiny X absolutně spojitého typu, stručněji f je hustotou X.

Značení: Fakt, že jde o spojitou náhodnou veličinu budeme značit $X \sim f$.

Spojitá náhodná veličina

Věta 10 (Vlastnosti hustoty)

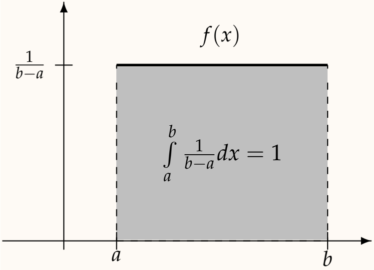
Nechť X je náhodná veličina absolutně spojitého typu, f její hustota a F její distribuční funkce. Pak

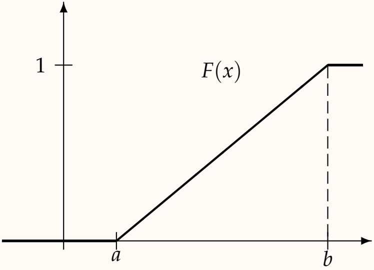
- $\int\limits_{-\infty}^{\infty} f(x)dx = 1$
- $F(x) = \int\limits_{-\infty}^{x} f(t)dt$
- *F je absolutně spojitá funkce.*
- *Hustota f je určena skoro všude jednoznačně vzhledem k Lebesgueově míře, tj. jsou-li f a g hustoty náhodné veličiny X, pak $\mu\left(\{x : f(x) \neq g(x)\}\right) = 0$.*
- *Existuje F' skoro všude vzhledem k Lebesgueově míře a funkce $f(x) = F'(x)$ je hustotou náhodné veličiny X.*
- *Pro každé $a < b$ platí $F(b) - F(a) = \int\limits_{a}^{b} f(x)dx$*

  a také $P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = \int\limits_{a}^{b} f(x)dx$.

Příklady

Příklad 11 (Rovnoměrné rozdělení (Uniform distribution))





$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b),\ a < b, \\ 0 & \textit{jinak.} \end{cases} \qquad F(x) = \begin{cases} 0 & x \le a, \\ \frac{x-a}{b-a} & x \in (a,b),\ a < b \\ 1 & x \ge b \end{cases}$$

Značíme $\boxed{X \sim Ro(a,b)}$.

Příklad

Příklad 12

Tramvaj jezdí v 10 minutových intervalech. Náhodně přijdeme na zastávku a měříme čas čekání na tramvaj. Určete pravděpodobnost, že budu čekat méně než 2 minuty.

X ... čas (v minutách) do příjezdu tramvaje $\Rightarrow X \sim Ro(0,10)$, $f(x) = \frac{1}{10}$ pro $x \in (0,10)$

$$P(X < 2) = \int_{-\infty}^{2} f(x)dx = \int_{0}^{2} \frac{1}{10}dx = \left[\frac{x}{10}\right]_0^2 = F(2) - F(0) = 0{,}2$$

Příklad 13

Náhodnou veličinou s rovnoměrným rozdělením je např. chyba při zaokrouhlování. Například zaokrouhlujeme-li na k desetinných míst, pak chyba $X \sim Ro\left(-5 \cdot 10^{-k-1}, 5 \cdot 10^{-k-1}\right)$.

Příklad

Příklad 14 (Normální (Gaussovo) rozdělení (Normal, Gaussian distribution))

$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $\quad x \in \mathbb{R}, ; \mu \in \mathbb{R},\ \sigma > 0$ *značíme* $\boxed{X \sim N(\mu, \sigma^2)}$.

$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$ $\quad u \in \mathbb{R}$ *značíme* $\boxed{U \sim N(0,1)}$.

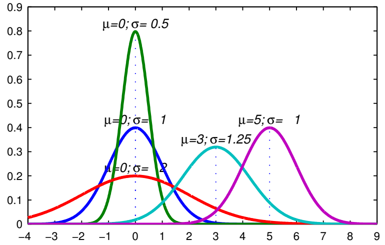
Standardizace: $U = \dfrac{X-\mu}{\sigma}$

Hustota $\varphi(u)$ *je hustotou tzv.* **standardizovaného normálního rozdělení**. *Bývá zvykem značit její distribuční funkci jako* $\Phi(u) = \int\limits_{-\infty}^{u} \varphi(t)dt$.

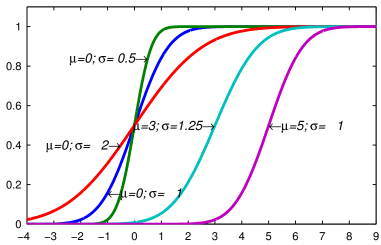
Distribuční funkci normálního rozdělení $F(x) = \int\limits_{-\infty}^{x} f(t)dt$ *nelze vyjádřit pomocí elementárních funkcí, lze ji však zapsat pomocí mocninných řad.*

Příklad

Hustoty



Distribuční funkce



Příklad

Příklad 15

Při prodeji vánočních kaprů má hmotnost kapra v jedné z kádí přibližně normální rozdělení s parametry $\mu = 2,3$ *a* $\sigma^2 = 0,3^2$*. Jaký podíl kaprů přesáhne svou hmotností 2,6 kg?*

X ... hmotnost kapra $\Rightarrow$ $\boxed{X \sim N(2,3; 0,3^2)}$

$$P(X > 2,6) = 1 - P(X \leq 2,6) = 1 - P\left(\frac{X-\mu}{\sigma} \leq \frac{2,6-2,3}{0,3}\right)$$
$$= 1 - P(U \leq 1) = 1 - \Phi(1) = 1 - 0,84 = 0,16$$

Příklad

Příklad 16 (Exponenciální rozdělení)

Nechť jev A se vyskytuje v náhodných okamžicích a předpokládáme, že výskyty tohoto jevu v nepřekrývajících se intervalech jsou nezávislé.
Označme

X ... *náhodná veličina udávající čas, kdy poprvé nastane sledovaný jev A.*

Distribuční funkce

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

Hustota

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Řekneme, že X má **exponenciální** *rozdělení s parametrem λ a značíme* $\boxed{X \sim Ex(\lambda)}$.

Příklad

Příklad 17

V porodnici se narodí v průměru každé 2 hodiny dítě. Určete pravděpodobnost, že se v daném dni nenarodí žádné dítě.

$X \dots$ **čas** do narození prvního dítěte (jednotka $=$ 1 den), 1 dítě za 2 hodiny $\Rightarrow$ 12 dětí za den $\Rightarrow$ $X \sim Ex(12)$

$$P(X > 1) = 1 - P(X \leq 1) = 1 - F(1) = 1 - (1 - e^{-12}) = e^{-12} = 6,14 \cdot 10^{-6}$$

nebo

$X \dots$ **počet** narozených dětí za 1 den $\Rightarrow$ $X \sim Po(12)$

$$P(X = 0) = p(0) = e^{-12} \frac{12^0}{0!} = e^{-12} = 6,14 \cdot 10^{-6}$$

Příklad

Příklad 18 (Gamma rozdělení)

Jestliže náhodná veličina X *má hustotu*

$$f(x) = \begin{cases} \frac{1}{\mu^a \Gamma(a)} x^{a-1} e^{-\frac{x}{\mu}} & a > 0, x \geq 0, \\ 0 & x < 0. \end{cases}$$

značíme $X \sim Gamma(a, \mu)$

Speciální případy: $a = 1$ exponenciální rozdělení

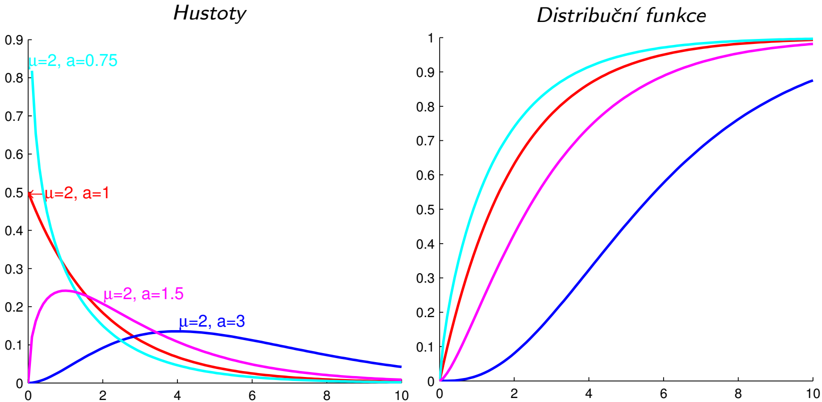
$a = n \in N$ Erlangovo rozdělení

Funkce Γ je pro $a > 0$ definována předpisem $\Gamma(a) = \int\limits_0^\infty x^{a-1} e^{-x} dx$.

Její nejčastěji používané vlastnosti jsou

$$\Gamma(a+1) = a\Gamma(a),$$
$$\Gamma(1/2) = \sqrt{\pi}$$
$$\Gamma(n) = (n-1)! \text{ pro } n \in \mathbb{N}$$

Příklad



Gamma rozdělení se používá především v teorii spolehlivosti, kdy například exponenciální rozdělení modeluje dobu do poruchy u komponent, které nejsou trvale namáhány, Erlangovo rozdělení se využívá pro popis doby života do n-té poruchy apod.

Příklad

Příklad 19 (Beta rozdělení)

Jestliže náhodná veličina X *má hustotu*

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & a,b > 0, x \in \langle 0,1\rangle \\ 0 & \textit{jinak} \end{cases}$$

značíme $\boxed{X \sim Beta(a,b)}$

Speciální případy: $a = 1,\ b = 1$ ROVNOMĚRNÉ ROZDĚLENÍ $Ro(0,1)$

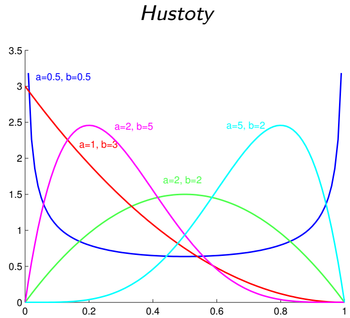
Funkce $B(a,b)$ je pro $a,b > 0$ definována předpisem

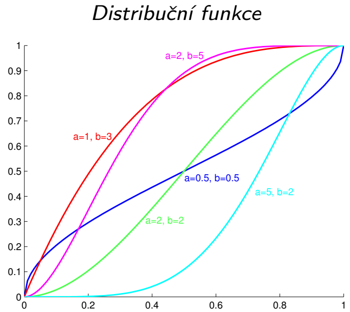
$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx.$$

Platí vztah mezi beta a gamma funkcí

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Příklad





V souvislosti s předchozími rozděleními se dají ukázat vztahy

$$\lim_{n\to\infty} nBeta(1,n) = Exp(1),$$
$$\lim_{n\to\infty} nBeta(k,n) = Gamma(k,1).$$