Náhodné veličiny

Zavedení náhodné veličiny a transformované náhodné veličiny

Náhodná veličina slouží k tomu, aby výsledky náhodného pokusu popsala reálnými čísly (resp. reálnými vektory):

zobrazení $X : \Omega \to R$ je (skalární) náhodná veličina, tj. zobrazení, které možnému výsledku ω přiřadí číslo $X(\omega)$,

zobrazení $\mathbf{X} = (X_1, \ldots, X_n) : \Omega \to R^n$ je náhodný vektor, tj. zobrazení, které možnému výsledku ω přiřadí čísla $X_1(\omega), \ldots, X_n(\omega)$.

Např. při hodu kostkou lze poloze kostky číslem i nahoru (tj. možnému výsledku ω_i) přiřadit číslo i, i = 1, 2, …, 6.

Číslo $X(\omega)$ se nazývá číselná realizace náhodné veličiny X příslušná možnému výsledku ω. (Nehrozí-li nebezpečí nedorozumění, náhodnou veličinu i její číselnou realizaci značíme týmž symbolem X.)

V některých situacích potřebujeme náhodnou veličinu X transformovat pomocí funkce g na transformovanou náhodnou veličinu Y = g(X). Např. X – průměr náhodně vybrané kuličky do kuličkového ložiska, $Y = \frac{4}{3}\pi\left(\frac{X}{2}\right)^3$ - objem kuličky.

Vztah mezi znakem a náhodnou veličinou

Pojem „znak“, který jsme zavedli v popisné statistice, je sice blízký pojmu „náhodná veličina“, ale není s ním totožný. Znak může být považován za náhodnou veličinu, jestliže jeho hodnoty zjišťujeme na objektech, které byly vybrány ze základního souboru náhodně.

Simultánní a marginální náhodný vektor

Jestliže z náhodného vektoru (X_1, …, X_n) vybereme některé složky, např. X_k, …, X_l, dostaneme marginální náhodný vektor (X_k, …, X_l). Původní náhodný vektor se v této souvislosti nazývá simultánní náhodný vektor.

Např. výsledky pěti zkoušek na konci semestru považujeme za náhodné veličiny X_1, …, X_5, přičemž X_3, X_4 jsou známky ze dvou nejdůležitějších předmětů. Náhodný vektor (X_1, …, X_5) je simultánní, náhodný vektor (X_3, X_4) je marginální.

Zapisování jevů pomocí náhodných veličin

Zápis $\{\omega \in \Omega; X(\omega) \in B\}$ znamená jev, že náhodná veličina X se realizovala v číselné množině B. Zkráceně píšeme $\{X \in B\}$.

Je-li $B = (-\infty, x\rangle$ nebo $B = \{x\}$, jev $\{X \le x\}$ znamená, že náhodná veličina X se realizovala hodnotou nejvýše x a jev $\{X = x\}$ znamená, že náhodná veličina X se realizovala hodnotou x.

Popis pravděpodobnostního chování náhodné veličiny

Při pozorování realizací náhodné veličiny si povšimneme, že některé její hodnoty se vyskytují s větší pravděpodobností, jiné s menší. Pravděpodobnostní chování náhodné veličiny X budeme popisovat pomocí distribuční funkce, která udává pravděpodobnost jevu, že náhodná veličina X se realizuje hodnotou nejvýše x:

$$\forall x \in R : \Phi(x) = P(X \leq x)$$

Je to zidealizovaný protějšek empirické distribuční funkce zavedené v popisné statistice:

$$\forall x \in R : F(x) = \frac{N(X \leq x)}{n}$$

Lze očekávat, že s rostoucím rozsahem výběrového souboru se budou hodnoty empirické distribuční funkce F(x) ustalovat kolem hodnot distribuční funkce Φ(x). Vlastnosti empirické distribuční funkce se přenášejí i na distribuční funkci – je to funkce neklesající, zprava spojitá a normovaná v tom smyslu, že

$\lim_{x\to-\infty} \Phi(x) = 0$, $\lim_{x\to\infty} \Phi(x) = 1$

Podobně se definuje i simultánní distribuční funkce náhodného vektoru $(X_1, \ldots, X_n)$:

$$\forall (x_1, \ldots, x_n) \in R^n : \Phi(x_1, \ldots, x_n) = P(X_1 \leq x_1 \wedge \ldots \wedge X_n \leq x_n)$$

Distribuční funkce libovolného marginálního náhodného vektoru se nazývá marginální distribuční funkce. Největší význam mají jednorozměrné marginální distribuční funkce $\Phi_1(x_1), \ldots, \Phi_n(x_n)$ jednotlivých složek náhodného vektoru.

Diskrétní náhodné veličiny

V praxi mají značný význam náhodné veličiny, které nabývají pouze konečně nebo spočetně mnoha hodnot – jsou to diskrétní náhodné veličiny.

Příklady diskrétních náhodných veličin: počet chyb, jichž se dopustí nějaké zařízení za určitou dobu, počet zákazníků ve frontě, počet zmetků v denní produkci apod.

Pravděpodobnostní chování diskrétní náhodné veličiny popisujeme pravděpodobnostní funkcí:

$$\forall x \in R : \pi(x) = P(X = x).$$

Je to zidealizovaný protějšek četnostní funkce zavedené v popisné statistice v souvislosti s bodovým rozložením četností:

$$\forall x \in R : p(x) = \frac{N(X = x)}{n}.$$

S rostoucím rozsahem výběrového souboru se budou hodnoty četnostní funkce ustalovat kolem hodnot pravděpodobnostní funkce.

Vlastnosti četnostní funkce se přenášejí i na pravděpodobnostní funkci, tedy pravděpodobnostní funkce

je nezáporná $\forall x \in R : \pi(x) \geq 0$,

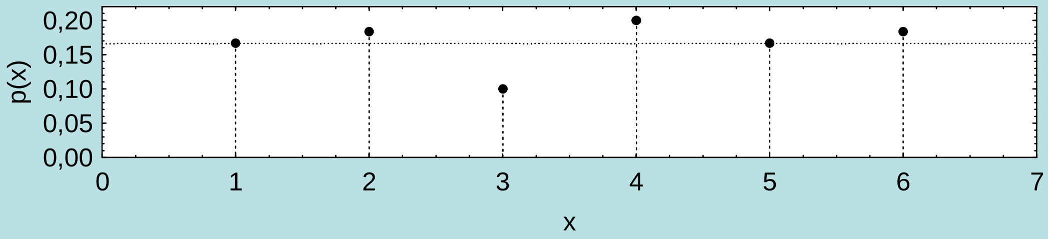
je normovaná $\sum_{x=-\infty}^{\infty} \pi(x) = 1$,

s distribuční funkcí je spjata součtovým vztahem $\forall x \in R : \Phi(x) = \sum_{t \leq x} \pi(t)$

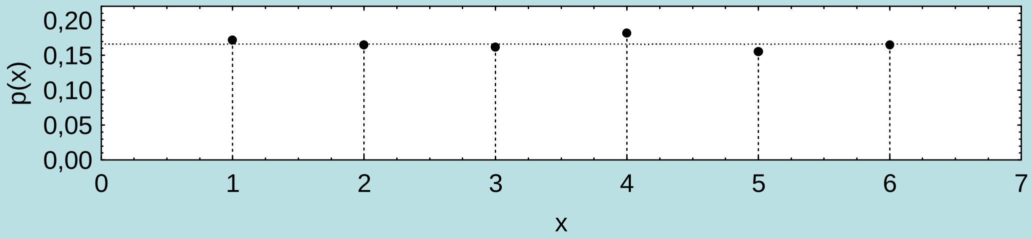
Ilustrace vztahu mezi četnostní funkcí a pravděpodobnostní funkcí

Provedeme n hodů kostkou. Zajímáme se o četnostní funkci počtu ok.

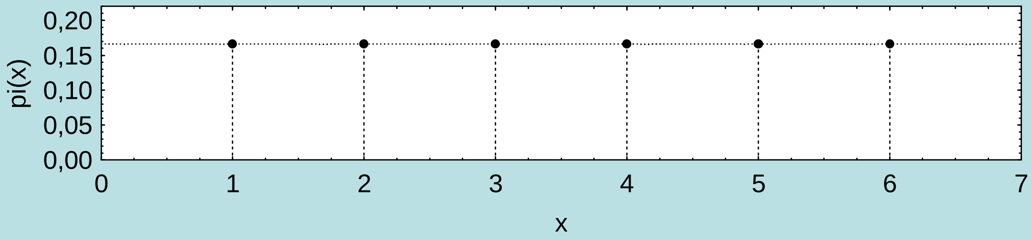
n = 60



n = 600



$n \longrightarrow \infty$



Diskrétní náhodný vektor

Jestliže všechny složky náhodného vektoru (X_1, …, X_n) jsou diskrétní náhodné veličiny, hovoříme o diskrétním náhodném vektoru. Jeho pravděpodobnostní chování je popsáno simultánní pravděpodobnostní funkcí:

$$\forall(x_1,\ldots,x_n)\in R^n : \pi(x_1,\ldots,x_n) = P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$$

Pravděpodobnostní funkce libovolného diskrétního marginálního náhodného vektoru se nazývá marginální pravděpodobnostní funkce. Největší význam mají jednorozměrné marginální pravděpodobnostní funkce $\pi_1(x_1)$, …, $\pi_n(x_n)$ jednotlivých složek náhodného vektoru. Získáme je tak, že hodnoty simultánní pravděpodobnostní funkce sečteme přes všech n-1 přebývajících proměnných.

Spojité náhodné veličiny

Dalším velmi důležitým typem veličin jsou spojité náhodné veličiny – ty nabývají všech hodnot z nějakého intervalu.

Příklady spojitých náhodných veličin: rozměry sériově vyráběných výrobků, hektarový výnos nějaké zemědělské plodiny, výsledky různých fyzikálních a chemických měření apod.

Pravděpodobnostní chování spojité náhodné veličiny popisujeme hustotou pravděpodobnosti $\varphi(x)$, což je zidealizovaný protějšek hustoty četnosti f(x) zavedené v popisné statistice v souvislosti s intervalovým rozložením četností. S rostoucím rozsahem výběrového souboru a klesajícími šířkami třídicích intervalů se budou hodnoty hustoty četnosti ustalovat kolem hodnot hustoty pravděpodobnosti.

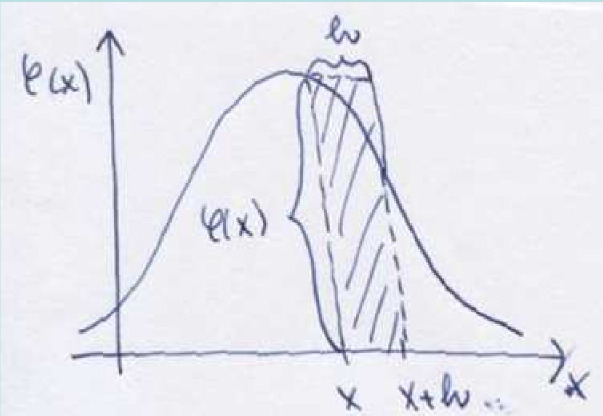
Vlastnosti hustoty četnosti se přenášejí i na hustotu pravděpodobnosti, tedy hustota pravděpodobnosti

je nezáporná $\forall x \in R : \varphi(x) \geq 0$,

je normovaná $\int_{-\infty}^{\infty} \varphi(x)dx = 1$,

s distribuční funkcí je spjata integrálním vztahem $\forall x \in R : \Phi(x) = \int_{-\infty}^{x} \varphi(t)dt$.

Pozor – na rozdíl od pravděpodobnostní funkce diskrétní náhodné veličiny nemá hustota pravděpodobnosti spojité náhodné veličiny význam pravděpodobnosti! Její význam lze odvodit z integrálního vztahu mezi distribuční funkcí a hustotou pravděpodobnosti.



Pravděpodobnost, že náhodná veličina se bude realizovat v intervalu (x, x+h], je:

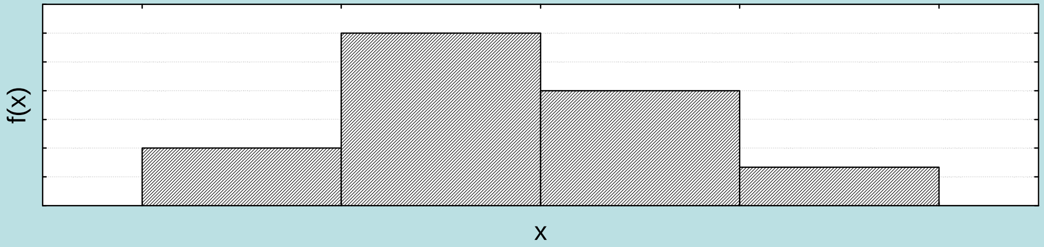
$$P(x < X \leq x+h) = \Phi(x+h) - \Phi(x) = \int_{-\infty}^{x+h} \varphi(t)dt - \int_{-\infty}^{x} \varphi(t)dt = \int_{x}^{x+h} \varphi(t)dt$$

Bude-li h dostatečně malé číslo, lze plochu pod grafem hustoty nahradit obsahem obdélníka o stranách $\varphi(x)$ a h, tj. $P(x < X \leq x+h) \approx \varphi(x) \cdot h$

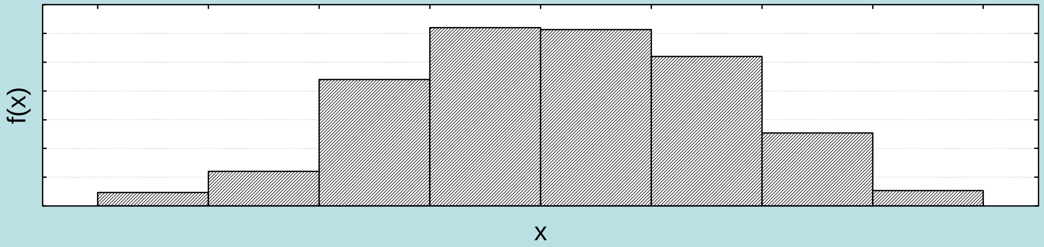
Ilustrace vztahu mezi hustotou četnosti a hustotou pravděpodobnosti

Náhodně vybereme n sériově vyráběných součástek, změříme jejich délku a budeme se zajímat o hustotu četnosti odchylek těchto měření od deklarované délky součástky.

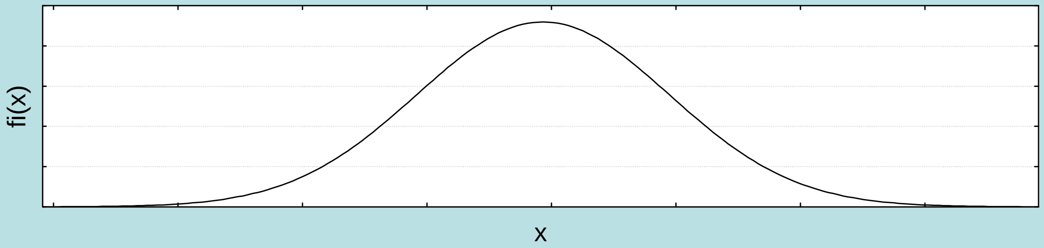
$n = 40$, $r = 4$



$n = 400$, $r = 8$



$n \longrightarrow \infty$, $r \longrightarrow \infty$



Spojitý náhodný vektor

Jestliže všechny složky náhodného vektoru $(X_1, \ldots, X_n)$ jsou spojité náhodné veličiny, hovoříme o spojitém náhodném vektoru. Jeho pravděpodobnostní chování je popsáno simultánní hustotou pravděpodobnosti $\varphi(x_1, \ldots, x_n)$.

Hustota pravděpodobnosti libovolného spojitého marginálního náhodného vektoru se nazývá marginální hustota pravděpodobnosti. Největší význam mají jednorozměrné marginální hustoty pravděpodobnosti $\varphi_1(x_1), \ldots, \varphi_n(x_n)$ jednotlivých složek náhodného vektoru. Získáme je tak, že simultánní hustotu pravděpodobnosti integrujeme přes všech n-1 přebývajících proměnných.

Stochasticky nezávislé náhodné veličiny

Při provedení pokusu se může stát, že se realizace jedné náhodné veličiny Y dají jednoznačně určit ze známé realizace druhé náhodné veličiny X, tedy je mezi nimi funkční vztah Y = g(X). Takové náhodné veličiny se nazývají deterministicky závislé.

Jejich protipólem jsou náhodné veličiny stochasticky nezávislé: informace o realizaci jedné z nich nijak nemění šance, s nimiž při témž pokusu očekáváme realizaci druhé.

Např. náhodný pokus spočívá v hodu dvěma kostkami. Náhodná veličina X udává počet ok, která padla na 1. kostce a náhodná veličina Y udává počet ok, která padla na druhé kostce. Náhodné veličiny X, Y jsou stochasticky nezávislé.

Stochastickou nezávislost náhodných veličin zavádíme na základě analogie s četnostní nezávislostí znaků v daném výběrovém souboru, která se používá v popisné statistice. Musí platit multiplikativní vztah:

$\forall (x, y) \in R^2 : p(x, y) = p_1(x) p_2(y)$ pro bodové rozložení četností,

$\forall (x, y) \in R^2 : f(x, y) = f_1(x) f_2(y)$ pro intervalové rozložení četností.

V počtu pravděpodobnosti nahradíme četnostní funkci pravděpodobnostní funkcí resp. hustotu četnosti nahradíme hustotou pravděpodobnosti. Místo dvou náhodných veličin X, Y můžeme uvažovat n náhodných veličin:

Náhodné veličiny $X_1, \dots, X_n$ jsou stochasticky nezávislé, když platí:

$\forall (x_1, \dots, x_n) \in R^n : \pi(x_1, \dots, x_n) = \pi_1(x_1) \cdot \ldots \cdot \pi_n(x_n)$ v diskrétním případě,

$\forall (x_1, \dots, x_n) \in R^n : \varphi(x_1, \dots, x_n) = \varphi_1(x_1) \cdot \ldots \cdot \varphi_n(x_n)$ ve spojitém případě,

$\forall (x_1, \dots, x_n) \in R^n : \Phi(x_1, \dots, x_n) = \Phi_1(x_1) \cdot \ldots \cdot \Phi_n(x_n)$ v obecném případě.

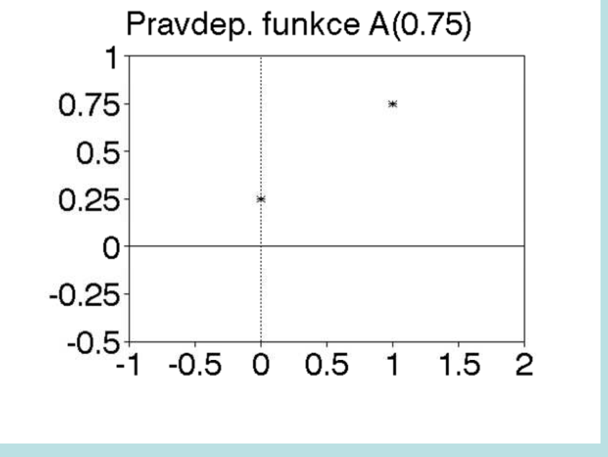
Alternativní a binomické rozložení, normální rozložení a rozložení z něj odvozená

Označení

Známe-li distribuční funkci $\Phi(x)$ náhodné veličiny X (resp. pravděpodobnostní funkci $\pi(x)$ v diskrétním případě resp. hustotu pravděpodobnosti $\varphi(x)$ ve spojitém případě), pak řekneme, že známe rozložení pravděpodobností (zkráceně rozložení) náhodné veličiny X. Toto rozložení závisí na nějakém parametru ϑ, což nejčastěji je reálné číslo nebo reálný vektor.
Zápis X ~ L(ϑ) čteme: náhodná veličina X má rozložení L s parametrem ϑ.

Alternativní rozložení: Náhodná veličina X udává počet úspěchů v jednom pokusu, přičemž pravděpodobnost úspěchu je ϑ. Píšeme X ~ A(ϑ).

$$\pi(x) = \begin{cases} 1-\vartheta \text{ pro } x=0 \\ \vartheta \text{ pro } x=1 \\ 0 \text{ jinak} \end{cases} \quad \text{neboli } \pi(x) = \begin{cases} \vartheta^x (1-\vartheta)^{1-x} \text{ pro } x=0,1 \\ 0 \text{ jinak} \end{cases}$$

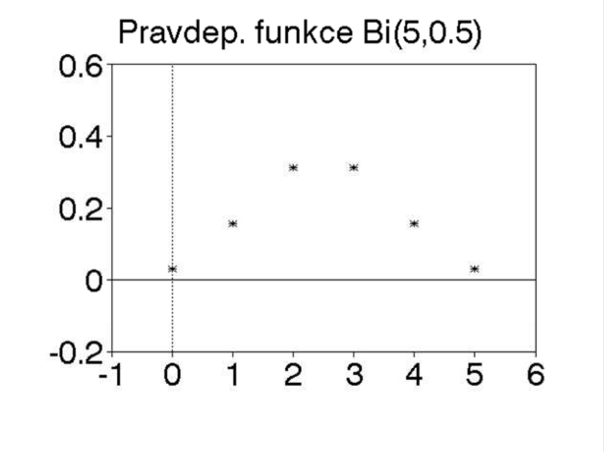


Binomické rozložení: Náhodná veličina X udává počet úspěchů v posloupnosti n nezávislých opakovaných pokusů, přičemž pravděpodobnost úspěchu je v každém pokusu ϑ. Píšeme X ~ Bi(n, ϑ).

$$\pi(x) = \begin{cases} \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x} & \text{pro } x = 0, \dots, n \\ 0 & \text{jinak} \end{cases}$$

(Alternativní rozložení je speciálním případem binomického rozložení pro n = 1.

Jsou-li $X_1, \dots, X_n$ stochasticky nezávislé náhodné veličiny, $X_i \sim A(\vartheta)$, i = 1, ..., n, pak $X = \sum_{i=1}^{n} X_i \sim Bi(n, \vartheta)$.)



Příklad na binomické rozložení pravděpodobností: V rodině je 10 dětí. Za předpokladu, že chlapci i dívky se rodí s pravděpodobností 0,5 a pohlaví se formuje nezávisle na sobě, určete pravděpodobnost, že v této rodině je

a) právě 5 chlapců

b) nejméně 3 a nejvýše 8 chlapců.

Řešení: X … počet chlapců, X ~ Bi(10; 0,5)

ad a) $P(X=5)=\pi(5)=\binom{10}{2}0{,}5^5(1-0{,}5)^{10-5}=0{,}2461$

V systému STATISTICA: =Binom(5;0,5;10)

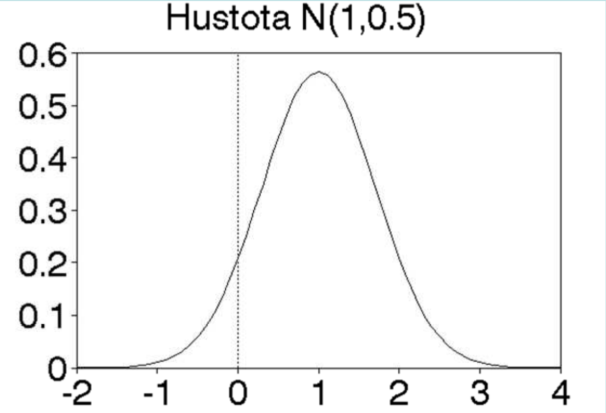
ad b) $P(3\le X\le 8)=\pi(3)+\pi(4)+\ldots+\pi(8)=\binom{10}{3}0{,}5^3 0{,}5^7+\binom{10}{4}0{,}5^4 0{,}5^6+\ldots+\binom{10}{8}0{,}5^8 0{,}5^2=0{,}9346$

V systému STATISTICA: = IBinom(8;0,5;10) - IBinom(2;0,5;10) = 0,934570

Normální rozložení: Tato náhodná veličina vzniká např. tak, že ke konstantě μ se přičítá velké množství nezávislých náhodných vlivů mírně kolísajících kolem nuly. Proměnlivost těchto vlivů je vyjádřena konstantou $\sigma > 0$.

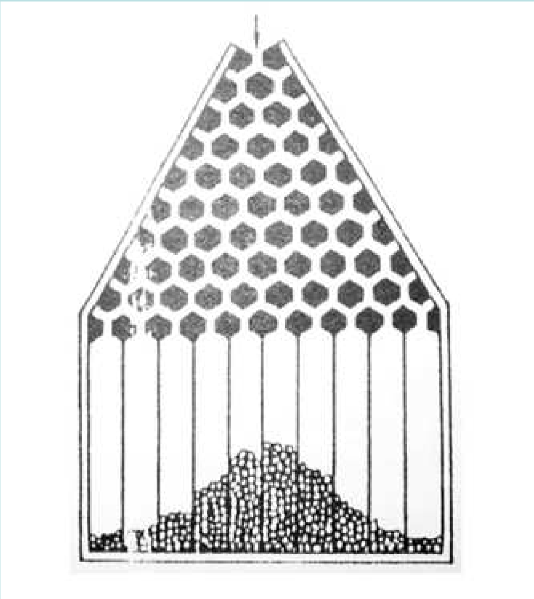
Píšeme $X \sim N(\mu, \sigma^2)$, hustota $\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Grafem této hustoty je tzv.

Gaussova křivka.



Ilustrace vzniku normálního rozložení pomocí Galtonovy desky:
Deska obsahuje n řad pravidelně uspořádaných klínů, a to tak, že v k-té řadě je právě k klínů. Do otvoru nahoře padají kuličky, které jsou v každé řadě se stejnou pravděpodobností 1/2 vychylovány vlevo nebo vpravo. Pod poslední řadou je n - 1 přihrádek, ve kterých se kuličky shromaždují. Nasypeme-li do tohoto systému velké množství kuliček, vytvoří v přihrádkách jakýsi ”kopec”, jehož tvar je velmi podobný tvaru grafu hustoty náhodné veličiny s normálním rozložením.
Náhodné vychylování kuliček jednotlivými řadami překážek je možno chápat jako speciální případ velkého množství chybových faktorů, náhodně působících na nějaký proces, jako působení mnoha blíže nespecifikovatelných vlivů, které ovlivňují zcela náhodně rozložení jeho výsledku.

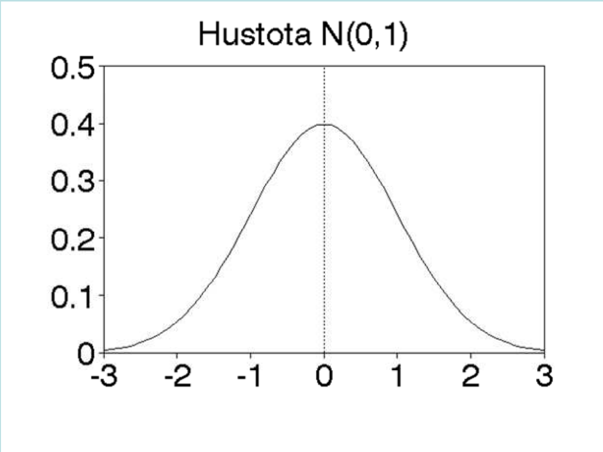
Obrázek

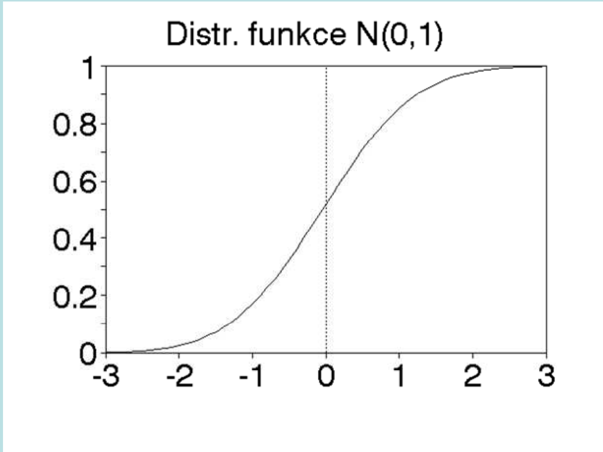


Standardizované normální rozložení:

Pro $\mu = 0$, $\sigma^2 = 1$ se jedná o standardizované normální rozložení, píšeme

$U \sim N(0, 1)$. Hustota pravděpodobnosti má v tomto případě tvar $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$.





$\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ je tabelována pro $u \geq 0$, pro $u < 0$ se používá přepočtový vzorec $\Phi(-u) = 1 - \Phi(u)$.

Příklad na normální rozložení: Výsledky u přijímacích zkoušek na jistou VŠ jsou normálně rozloženy s parametry μ = 550 bodů, σ = 100 bodů. S jakou pravděpodobností bude mít náhodně vybraný uchazeč aspoň 600 bodů?

Řešení:

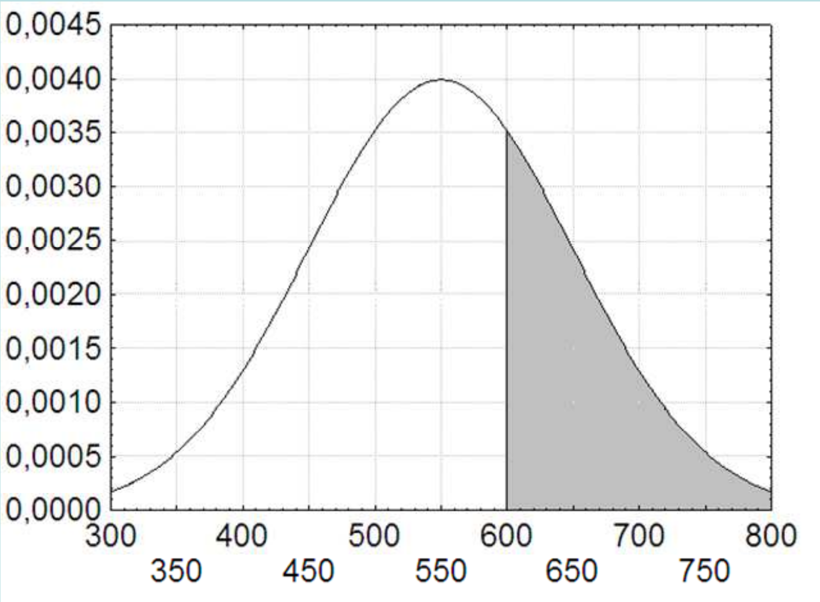
X – výsledek náhodně vybraného uchazeče, $X \sim N(550, 100^2)$,

$P(X \geq 600) = 1 - P(X \leq 600) + P(X = 600) = 1 - P(X \leq 600) =$

$=1 - P\left(\frac{X-\mu}{\sigma} \leq \frac{600-\mu}{\sigma}\right) = 1 - P\left(U \leq \frac{600-550}{100}\right) = 1 - \Phi(0,5) = 1 - 0,69146 = 0,30854.$

Ve STATISTICE: výpočet pomocí funkce 1 - INormal(600;550;100)

Náhodně vybraný uchazeč bude mít u zkoušek aspoň 600 bodů s pravděpodobností 0,31.



Některé vlastnosti normálního rozložení:

Jestliže $X \sim N(\mu, \sigma^2)$, pak $U = \frac{X - \mu}{\sigma} \sim N(0,1)$.

Jestliže $X \sim N(\mu, \sigma^2)$, a $Y = a + bX$, pak $Y \sim N(a + b\mu, b^2\sigma^2)$.

Jestliže $X_1, \ldots, X_n$ jsou stochasticky nezávislé náhodné veličiny, $X_i \sim N(\mu_i, \sigma_i^{\,2})$, $i = 1, \ldots, n$, $Y = \sum_{i=1}^{n} X_i$, pak

$$Y \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^{\,2}\right).$$

Význam normálního rozložení:

Normální rozložení hraje ústřední roli v počtu pravděpodobnosti i matematické statistice. Jeho význam spočívá jednak v tom, že normálním rozložením se řídí pravděpodobnostní chování mnoha náhodných veličin a jednak v tom, že za určitých podmínek konverguje k normálnímu rozložení součet nezávislých náhodných veličin s týmž rozložením (viz centrální limitní věta).

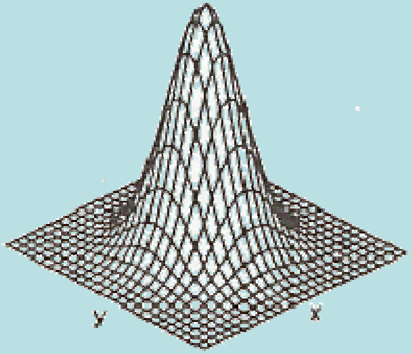
Dvourozměrné normální rozložení: $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right)$

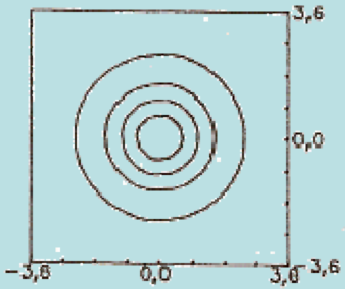
Náhodný vektor $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ vzniká ve dvourozměrných situacích podobně jako skalární náhodná veličina v bodě (e).

$$\varphi(x_1, x_2) = \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{q(x_1,x_2)}{2}}, \text{ kde } q(x_1, x_2) = \frac{1}{1-\rho^2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho \cdot \frac{x_1-\mu_1}{\sigma_1} \cdot \frac{x_2-\mu_2}{\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right].$$

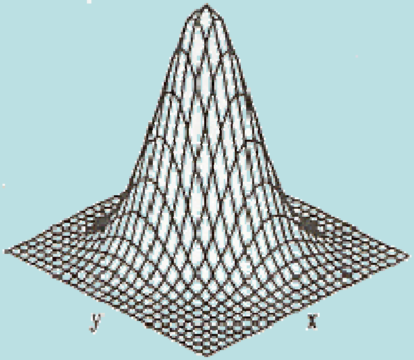
Pro $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\rho = 0$ se jedná o **standardizované dvourozměrné normální rozložení**.

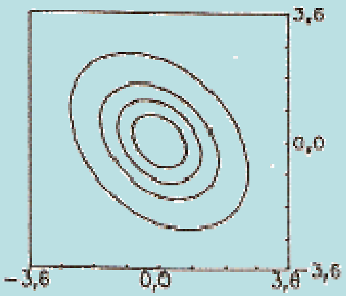
Vrstevnice a graf hustoty standardizovaného dvourozměrného normálního rozložení





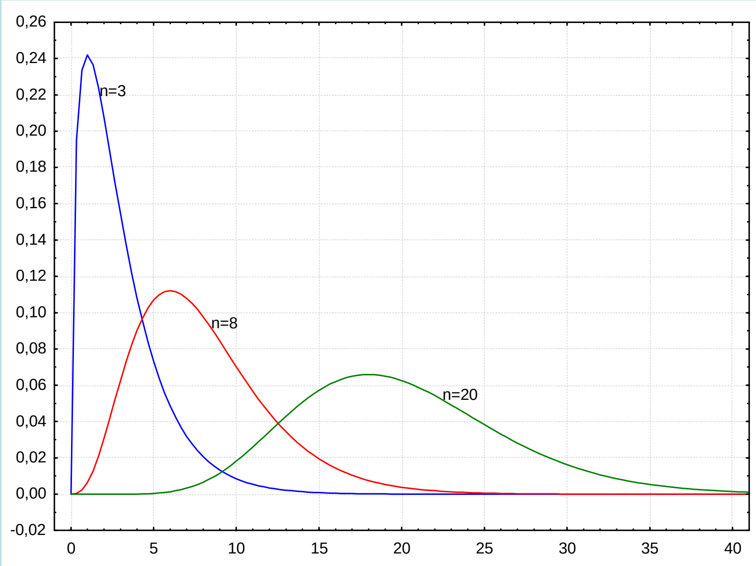
Vrstevnice a graf hustoty dvourozměrného normálního rozložení s parametry $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\rho = -0{,}75$





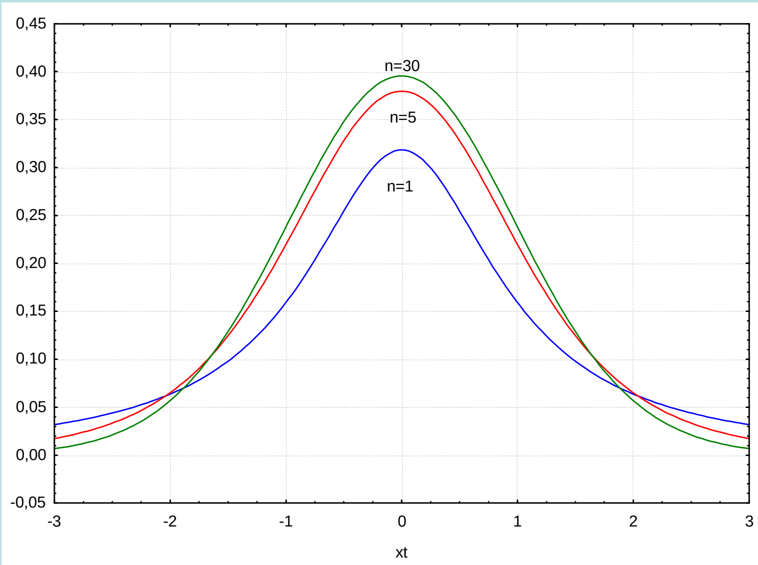
Upozornění: Následující tři rozložení – Pearsonovo, Studentovo a Fisherovo-Snedecorovo – jsou odvozena ze standardizovaného normálního rozložení. Mají velký význam především v matematické statistice při konstrukci intervalů spolehlivosti a testování hypotéz. Vyjádření hustot těchto rozložení neuvádíme, je příliš složité

Pearsonovo rozložení chí-kvadrát s n stupni volnosti: Nechť $X_1, ..., X_n$ jsou stochasticky nezávislé náhodné veličiny, $X_i \sim N(0, 1)$, i = 1, ..., n.
Pak náhodná veličina $X = X_1^2 + \ ... + X_n^2 \sim \chi^2(n)$.



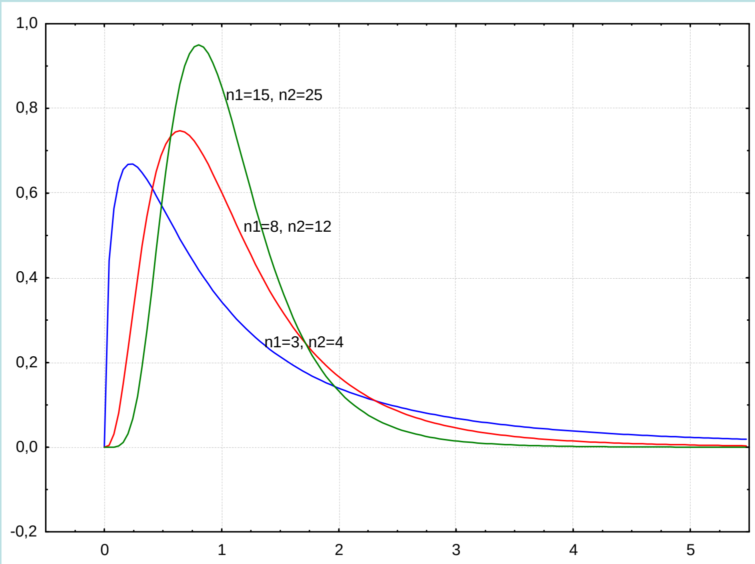
Studentovo rozložení s n stupni volnosti: Nechť X_1, X_2 jsou stochasticky nezávislé náhodné veličiny, $X_1 \sim N(0, 1)$, $X_2 \sim \chi^2(n)$.

Pak náhodná veličina $X = \dfrac{X_1}{\sqrt{\dfrac{X_2}{n}}} \sim t(n)$.



Fisherovo-Snedecorovo rozložení s n_1 a n_2 stupni volnosti: Nechť $X_1, ..., X_n$ jsou stochasticky nezávislé náhodné veličiny, $X_i \sim \chi^2(n_i)$, i = 1, 2.

Pak náhodná veličina $X = \dfrac{X_1 / n_1}{X_2 / n_2} \sim F(n_1, n_2)$.



Číselné charakteristiky náhodných veličin

Motivace

Doposud jsme poznali funkcionální charakteristiky náhodných veličin (např. distribuční funkce, pravděpodobnostní funkce, hustota pravděpodobnosti), které plně popisují pravděpodobnostní chování náhodné veličiny. Číselné charakteristiky vystihují pouze některé rysy tohoto chování, např. popisují polohu realizací náhodné veličiny na číselné ose či jejich proměnlivost (variabilitu). Jsou jednodušší než číselné charakteristiky, ale nesou jen částečnou informaci.
Podobně jako v popisné statistice volíme vhodnou číselnou charakteristiku podle toho, jakého typu je daná náhodná veličina - zda je ordinální nebo intervalová či poměrová. Číselné charakteristiky znaků mají své teoretické protějšky v číselných charakteristikách náhodných veličin.

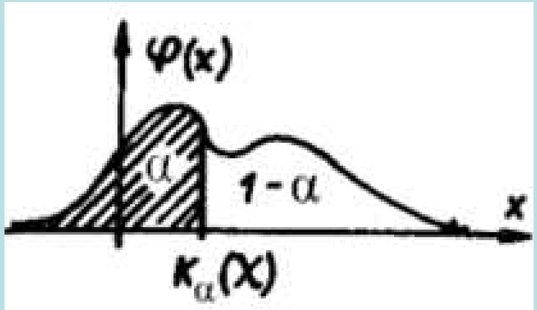
Číselné charakteristiky spojité náhodné veličiny aspoň ordinálního typu

Charakteristika polohy : α-kvantil
Nechť X je spojitá náhodná veličina aspoň ordinálního typu s distribuční funkcí $\Phi(x)$ a hustotou pravděpodobnosti $\varphi(x)$.
Nechť $\alpha \in (0, 1)$.
Číslo $K_\alpha(X)$, které splňuje podmínku

$$\alpha = \Phi(K_\alpha(X)) = \int_{-\infty}^{K_\alpha(X)} \varphi(x)\,dx,$$

se nazývá α-kvantil náhodné veličiny X.
$K_{0,50}(X)$ - medián,
$K_{0,25}(X)$ - dolní kvartil,
$K_{0,75}(X)$ - horní kvartil,
$K_{0,10}(X)$, ..., $K_{0,90}(X)$ - 1. až 9. decil,
$K_{0,01}(X)$, ..., $K_{0,99}(X)$ - 1. až 99. percentil.
Kterýkoliv α-kvantil je charakteristikou polohy číselných realizací náhodné veličiny na číselné ose.
Charakteristika variability: kvartilová odchylka $q = K_{0,75}(X) - K_{0,25}(X)$.

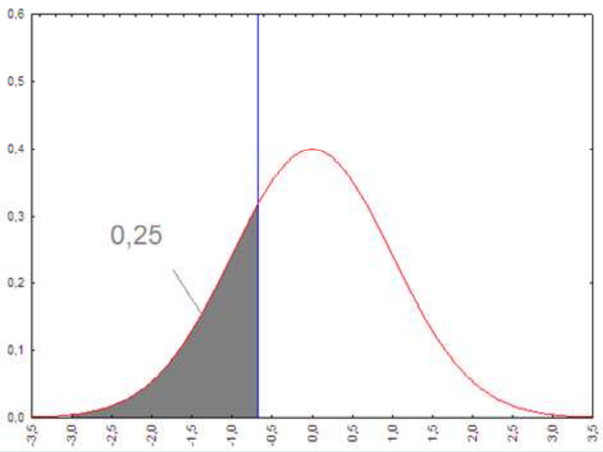
Ilustrace



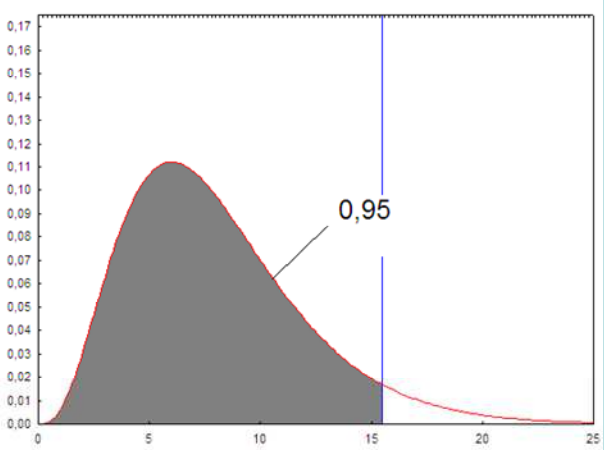
Označení pro kvantily speciálních rozložení

$X \sim N(0, 1) \Rightarrow K_\alpha(X) = u_\alpha$,
$X \sim \chi^2(n) \Rightarrow K_\alpha(X) = \chi^2_\alpha(n)$,
$X \sim t(n) \Rightarrow K_\alpha(X) = t_\alpha(n)$,
$X \sim F(n_1, n_2) \Rightarrow K_\alpha(X) = F_\alpha(n_1, n_2)$.
Tyto kvantily najdeme ve statistických tabulkách.
Používáme vztahy:
$u_\alpha = - u_{1-\alpha}$,
$t_\alpha(n) = - t_{1-\alpha}(n)$,
$$F_\alpha(n_1, n_2) = \frac{1}{F_{1-\alpha}(n_2, n_1)}.$$

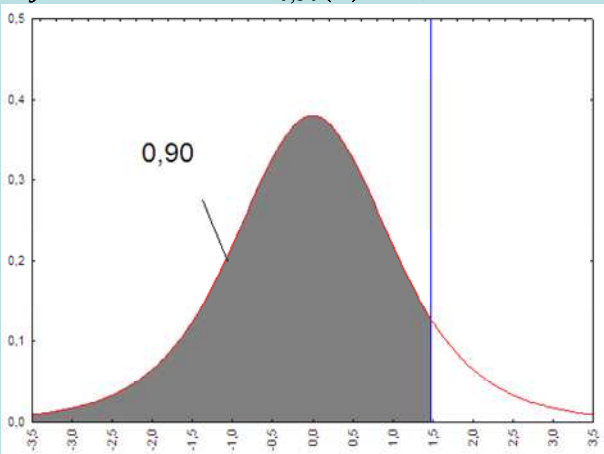
Význam kvantilu $u_{0,25} = -0,6745$



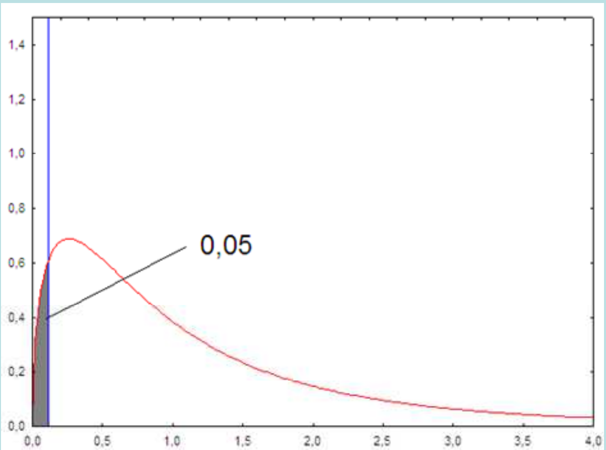
Význam kvantilu $\chi^2_{0,95}(8) = 15,5073$



Význam kvantilu $t_{0,90}(5) = 1,4759$



Význam kvantilu $F_{0,05}(3,7) = \dfrac{1}{F_{0,95}(7,3)} = \dfrac{1}{8,8867} = 0,1125$



Příklad: Nechť U ~ N(0, 1). Pomocí systému STATISTICA najděte 2. decil a první a poslední percentil.

První možnost: Použijeme Pravděpodobnostní kalkulátor. Do okénka průměr napíšeme 0, do okénka Sm. Odch. napíšeme 1, do okénka p napíšeme pro 2. decil 0,2, pro první percentil 0,01 a pro poslední percentil 0,99. V okénku X se objeví -0,841621 pro 2. decil, -2,326348 pro první percentil a 2,326348 pro poslední percentil.
Ilustrace pro poslední percentil:

Šedá plocha pod grafem hustoty má velikost 0,99 a hodnota distribuční funkce v bodě 2,326348 je 0,99 (značeno šrafovaně).

Druhá možnost: Otevřeme nový datový soubor o třech proměnné a jednom případu.
Do dlouhého jména první proměnné napíšeme =VNormal(0,2;0;1). Dostaneme -0,841621.
Do dlouhého jména druhé proměnné napíšeme =VNormal(0,01;0;1). Dostaneme -2,326348.
Do dlouhého jména třetí proměnné napíšeme =VNormal(0,99;0;1). Dostaneme 2,326348.

Příklad: Nechť X ~ N(-1, 4). Pomocí systému STATISTICA najděte horní kvartil.

První možnost: Spustíme Pravděpodobnostní kalkulátor, vybereme Rozdělení Normální. Do okénka průměr napíšeme -1, do okénka Sm. Odch. napíšeme 2, do okénka p napíšeme 0,75 a v okénku X se objeví 0,34898.
Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a jednom případu.
Do dlouhého jména této proměnné napíšeme =VNormal(0,75;-1;2). Dostaneme 0,34898.

Příklad: Pomocí systému STATISTICA určete $\chi^2_{0,05}(5)$.

První možnost: Spustíme Pravděpodobnostní kalkulátor, vybereme Rozdělení Chi 2. Do okénka sv. napíšeme 5 a do okénka p napíšeme 0,05. V okénku Chi 2 se objeví 1,145476.
Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a jednom případu.
Do dlouhého jména této proměnné napíšeme =VChi2(0,05;5). Dostaneme 1,145476.

Příklad: Pomocí systému STATISTICA určete $t_{0,975}(18)$ a $t_{0,01}(4)$.

První možnost: Spustíme Pravděpodobnostní kalkulátor, vybereme Rozdělení t (Studentovo). Do okénka sv. napíšeme 18 (resp. 4) a do okénka p napíšeme 0,975 (resp. 0,01). V okénku t se objeví 2,100922 (resp. -3,746947).

Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a jednom případu.
Do Dlouhého jména této proměnné napíšeme =VStudent(0,975;18) (resp. VStudent(0,01;4)). Dostaneme 2,100922 (resp. -3,746947).

Příklad: Pomocí systému STATISTICA určete $F_{0,975}(3, 12)$ a $F_{0,05}(18, 20)$.

První možnost: Spustíme Pravděpodobnostní kalkulátor, vybereme Rozdělení F (Fisherovo). Do okénka sv1 napíšeme 3 (resp. 18), do okénka sv2 napíšeme 12 (resp. 20) a do okénka p napíšeme 0,975 (resp. 0,05).
V okénku F se objeví 4,474185 (resp. 0,456486).

Druhá možnost: Otevřeme nový datový soubor o jedné proměnné a dvou případech
Do Dlouhého jména první proměnné napíšeme =VF(0,975;3;12), do Dlouhého jména druhé proměnné napíšeme =VF(0,05;18;20).Dostaneme 4,474185 (resp. 0,456486).

Číselné charakteristiky diskrétních a spojitých náhodných veličin aspoň intervalového typu

Charakteristika polohy: střední hodnota E(X) – číslo, které charakterizuje polohu realizací náhodné veličiny na číselné ose s přihlédnutím k jejich pravděpodobnostem.

Diskrétní případ: náhodná veličina X má pravděpodobnostní funkci $\pi(x)$.

Střední hodnota $E(X) = \sum_{x=-\infty}^{\infty} x\pi(x)$, pokud je suma vpravo konečná.

Fyzikální význam: střední hodnota je těžiště soustavy hmotných bodů, jejichž celková hmotnost je 1 a bod o souřadnici x má hmotnost $\pi(x)$.

Spojitý případ: náhodná veličina X má hustotu pravděpodobnosti $\varphi(x)$.

Střední hodnota $E(X) = \int_{-\infty}^{\infty} x\varphi(x)dx$, pokud je integrál vpravo konečný.

Fyzikální význam: střední hodnota je těžiště hmotné přímky, jejíž celková hmotnost je 1 a hmota je na přímce rozprostřena podle předpisu φ (x).

Centrovaná náhodná veličina: Y = X - E(X).

(Pro náhodnou veličinu Y platí: E(Y) = 0.)

Střední hodnota transformované náhodné veličiny Y = g(X)

$$E(Y) = \left\langle \begin{array}{l} \sum\limits_{x=-\infty}^{\infty} g(x)\pi(x) \\ \int\limits_{-\infty}^{\infty} g(x)\varphi(x)dx \end{array} \right|$$

Střední hodnota transformované náhodné veličiny Y = g(X_1, X_2)

$$E(Y) = \left\langle \begin{array}{l} \sum\limits_{x_1=-\infty}^{\infty} \sum\limits_{x_2=-\infty}^{\infty} g(x_1, x_2)\pi(x_1, x_2) \\ \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} g(x_1, x_2)\varphi(x_1, x_2)dx_1dx_2 \end{array} \right|$$

Charakteristika variability: rozptyl D(X) - číslo, které charakterizuje proměnlivost realizací náhodné veličiny kolem její střední hodnoty s přihlédnutím k jejich pravděpodobnostem.

Definiční vzorec: $D(X) = E\left([X - E(X)]^2\right)$ (rozptyl je střední hodnota kvadrátu centrované náhodné veličiny).

Výpočetní vzorec: $D(X) = E\left(X^2\right) - [E(X)]^2$ (rozptyl je střední hodnota kvadrátu mínus kvadrát středních hodnot).

$$D(X) = \left\langle \begin{array}{l} \displaystyle\sum_{x=-\infty}^{\infty} x^2 \pi(x) - \left[\sum_{x=-\infty}^{\infty} x\pi(x)\right]^2 \\ \displaystyle\int_{-\infty}^{\infty} x^2 \varphi(x)dx - \left[\int_{-\infty}^{\infty} x\varphi(x)dx\right]^2 \end{array} \right|$$

Směrodatná odchylka $\sqrt{D(X)}$ - vyjadřuje průměrnou variabilitu realizací náhodné veličiny X kolem její střední hodnoty.

Standardizovaná náhodná veličina: $Z = \dfrac{X - E(X)}{\sqrt{D(X)}}$

(Pro náhodnou veličinu Z platí: E(Z) = 0, D(Z) = 1.)

Příklad na výpočet střední hodnoty a rozptylu diskrétní náhodné veličiny: Střelec střílí do terče až do prvního zásahu. Má v zásobě 4 náboje. Pravděpodobnost zásahu je při každém výstřelu 0,6. Náhodná veličina X udává počet nespotřebovaných nábojů. Vypočtěte střední hodnotu a rozptyl náhodné veličiny X.

Řešení:

X nabývá hodnot 0, 1, 2, 3 a její pravděpodobnostní funkce je

$\pi(0) = P(X=0) = 0{,}4^4 + 0{,}4^3 * 0{,}6 = 0{,}064$,

$\pi(1) = P(X=1) = 0{,}4^2 * 0{,}6 = 0{,}096$,

$\pi(2) = P(X=2) = 0{,}4 * 0{,}6 = 0{,}24$,

$\pi(3) = P(X=3) = 0{,}6$,

$\pi(0) = 0$ jinak

$E(X) = 0 * 0{,}064 + 1 * 0{,}096 + 2 * 0{,}24 + 3 * 0{,}6 = 2{,}376$

$D(X) = 0^2 * 0{,}064 + 1^2 * 0{,}096 + 2^2 * 0{,}24 + 3^2 * 0{,}6 - 2{,}376^2 = 0{,}8106$

Kovariancí náhodných veličin X_1, X_2, které mají střední hodnoty $E(X_1)$, $E(X_2)$, rozumíme číslo

$C(X_1, X_2) = E([X_1 - E(X_1)]\,[X_2 - E(X_2)])$ (pokud střední hodnoty vpravo existují).

(Kovariance je číslo, které charakterizuje proměnlivost realizací náhodných veličin X_1, X_2 kolem jejich středních hodnot s přihlédnutím k jejich pravděpodobnostem. Je-li kovariance kladná (záporná), pak to svědčí o existenci jistého stupně přímé (nepřímé) lineární závislosti mezi realizacemi náhodných veličin X_1, X_2. Je-li kovariance nulová, pak říkáme, že náhodné veličiny X_1, X_2 jsou nekorelované a znamená to, že mezi jejich realizacemi není žádný lineární vztah. Pozor – z nekorelovanosti nevyplývá stochastická nezávislost, zatímco ze stochastické nezávislosti plyne nekorelovanost. Kovariance je teoretickým protějškem vážené kovariance. Je vhodnější počítat kovarianci podle vzorce $C(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$.)

Koeficientem korelace náhodných veličin X_1, X_2 rozumíme číslo

$$R(X_1, X_2) = \begin{cases} E\left(\dfrac{X_1 - E(X_1)}{\sqrt{D(X_1)}} \cdot \dfrac{X_2 - E(X_2)}{\sqrt{D(X_2)}}\right) \text{pro } \sqrt{D(X_1)}\sqrt{D(X_2)} > 0 \\ 0 \text{ jinak} \end{cases}$$

, pokud střední hodnoty vpravo existují.

(Koeficient korelace je číslo, které charakterizuje těsnost lineární závislosti realizací náhodných veličin X_1, X_2. Čím bližší je 1, tím těsnější je přímá lineární závislost, čím bližší je -1, tím těsnější je nepřímá lineární závislost. Je vhodnější počítat koeficient korelace podle vzorce $R(X_1, X_2) = \dfrac{C(X_1, X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}}$.)

V diskrétním případě je kovariance dána vzorcem

$$C(X_1, X_2) = \sum_{x_1=-\infty}^{\infty} \sum_{x_2=-\infty}^{\infty} [x_1 - E(X_1)] \cdot [x_2 - E(X_2)] \pi(x_1, x_2) =$$

$$= \sum_{x_1=-\infty}^{\infty} \sum_{x_2=-\infty}^{\infty} x_1 x_2 \pi(x_1, x_2) - E(X_1)E(X_2)$$

a ve spojitém případě vzorcem

$$C(X_1, X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x_1 - E(X_1)] \cdot [x_2 - E(X_1)] \varphi(x_1, x_2) dx_1 dx_2 =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 \varphi(x_1, x_2) dx_1 dx_2 - E(X_1)E(X_2)$$

Příklad na výpočet koeficientu korelace: Diskrétní náhodný vektor (X_1, X_2) má simultánní pravděpodobnostní funkci s hodnotami: $\pi(0,-1) = c$, $\pi(0,0) = \pi(0,1) = \pi(1,-1) = \pi(2,-1) = 0$, $\pi(1,0) = \pi(0,1) = \pi(2,1) = 2c$, $\pi(2,0) = 3c$, $\pi(x_1,x_2) = 0$ jinak. Určete konstantu c a vypočtěte $R(X_1, X_2)$.

Řešení: Hodnoty simultánní pravděpodobnostní funkce a obou marginálních pravděpodobnostních funkcí uspořádáme do kontingenční tabulky.

x_1	x_2			$\pi_1(x_1)$
	-1	0	1	
0	c	0	0	c
1	0	2c	2c	4c
2	0	3c	2c	5c
$\pi_2(x_2)$	c	5c	4c	1

Z normovanosti pravděpodobnostní funkce diskrétního náhodného vektoru dostáváme: 10 c = 1, tedy c = 0,1. Po dosazení za c vznikne tabulka:

x_1	x_2			$\pi_1(x_1)$
	-1	0	1	
0	0,1	0	0	0,1
1	0	0,2	0,2	0,4
2	0	0,3	0,2	0,5
$\pi_2(x_2)$	0,1	0,5	0,4	1

$$E(X_1) = \sum_{x_1=0}^{2} x_1 \pi_1(x_1) = 0.0{,}1 + 1.0{,}4 + 2.0{,}5 = 1{,}4,\quad E(X_2) = \sum_{x_2=-1}^{1} x_2 \pi_2(x_2) = -1.0{,}1 + 0.0{,}5 + 1.0{,}4 = 0{,}3$$

$$D(X_1) = \sum_{x_1=0}^{2} x_1^{\,2} \pi_1(x_1) - [E(X_1)]^2 = 1^2 \cdot 0{,}4 + 2^2 \cdot 0{,}5 - 1{,}4^2 = 0{,}44,\quad D(X_2) = \sum_{x_1=-1}^{1} x_2^{\,2} \pi_2(x_2) - [E(X_2)]^2 = (-1)^2 \cdot 0{,}1 + 1^2 \cdot 0{,}4 - 0{,}3^2 = 0{,}41$$

$$C(X_1, X_2) = \sum_{x_1=-0}^{2} \sum_{x_2=-1}^{1} x_1 x_2 \pi(x_1, x_2) - E(X_1)E(X_2) = 1 \cdot 1 \cdot 0{,}2 + 2 \cdot 1 \cdot 0{,}2 - 1{,}4 \cdot 0{,}3 = 0{,}18$$

$$R(X_1, X_2) = \frac{C(X_1, X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}} = \frac{0{,}18}{\sqrt{0{,}44}\sqrt{0{,}41}} = 0{,}42$$

Střední hodnoty a rozptyly vybraných diskrétních a spojitých rozložení

$X \sim A(\vartheta) \Rightarrow E(X) = \vartheta$, $D(X) = \vartheta(1-\vartheta)$

$X \sim Bi(n, \vartheta) \Rightarrow E(X) = n\vartheta$, $D(X) = n\vartheta(1-\vartheta)$

$X \sim N(\mu, \sigma^2) \Rightarrow E(X) = \mu$, $D(X) = \sigma^2$

$X \sim \chi^2(n) \Rightarrow E(X) = n$, $D(X) = 2n$

$X \sim t(n) \Rightarrow E(X) = 0$ pro $n \geq 2$ (pro $n = 1$ střední hodnota neexistuje), $D(X) = \frac{n}{n-2}$ pro $n \geq 3$ (pro $n = 1,2$ rozptyl neexistuje).

$X \sim F(n_1, n_2) \Rightarrow E(X) = \frac{n_2}{n_2 - 2}$ pro $n_2 \geq 3$ (pro $n_2 = 1,2$ střední hodnota neexistuje), $D(X) = \frac{2n_2^{\,2}(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}$ pro $n_2 \geq 5$ (pro $n_2 = 1,2,3,4$ rozptyl neexistuje).

Centrální limitní věta:

Jsou-li náhodné veličiny $X_1, \ldots, X_n$ stochasticky nezávislé a všechny mají stejné rozložení se střední hodnotou μ a rozptylem σ^2, pak pro velká n ($n \geq 30$) lze rozložení součtu $\sum_{i=1}^{n} X_i$ aproximovat normálním rozložením $N(n\mu, n\sigma^2)$.

Zkráceně píšeme $\sum_{i=1}^{n} X_i \approx N\left(n\mu, n\sigma^2\right)$.

Pokud součet $\sum_{i=1}^{n} X_i$ standardizujeme, tj. vytvoříme náhodnou veličinu $U_n = \dfrac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$, pak rozložení této náhodné veličiny lze aproximovat standardizovaným normálním rozložením. Zkráceně píšeme $U_n \approx N(0,1)$

Normální rozložení je tedy rozložením limitním, k němuž se blíží všechna rozložení, proto hraje velmi důležitou roli v počtu pravděpodobnosti a matematické statistice.

Ilustrace centrální limitní věty – opakované hody kostkou