

Náhodné veličiny

Zavedení náhodné veličiny a transformované náhodné veličiny

Výsledky náhodného pokusu lze popsat reálnými čísly (resp. reálnými vektory) pomocí nějakého zobrazení $X: \Omega \rightarrow \mathbb{R}$ resp. $\mathbf{X} = (X_1, \dots, X_n): \Omega \rightarrow \mathbb{R}^n$, tj. zobrazení, které možnému výsledku ω přiřadí číslo $X(\omega)$ nebo několik čísel $X_1(\omega), \dots, X_n(\omega)$. Např. při hodu kostkou lze poloze kostky číslem i nahoru (tj. možnému výsledku ω_i) přiřadit číslo i , $i = 1, 2, \dots, 6$.

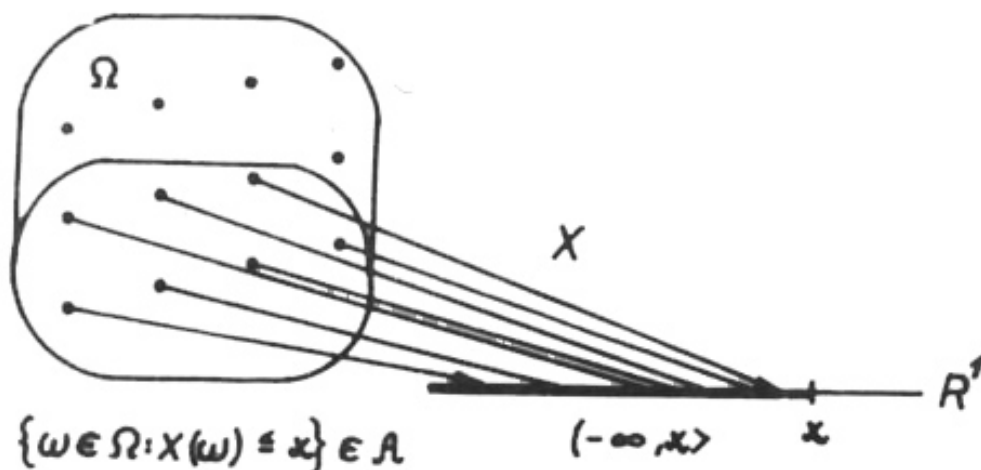
Pokud bude toto zobrazení splňovat podmínku tzv. měřitelnosti, tj. pro všechna reálná čísla x je množina $\{\omega \in \Omega; X(\omega) \leq x\}$ jev vzhledem k jevovému poli \mathcal{A} , pak se X nazývá **náhodná veličina** a $\mathbf{X} = (X_1, \dots, X_n)$ se nazývá **náhodný vektor** (se složkami X_1, \dots, X_n). 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.

Ilustrace náhodné veličiny



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, \dots, X_n) vybereme některé složky, např. X_k, \dots, X_l , dostaneme **marginální náhodný vektor** (X_k, \dots, 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, \dots, X_5 , přičemž X_3, X_4 jsou známky ze dvou nejdůležitějších předmětů. Náhodný vektor (X_1, \dots, 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 $B = (-\infty, x)$ nebo $B = \{x\}$, jev $\{X \leq x\}$ 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 :

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

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

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

Lze očekávat, že s rostoucím rozsahem výběrového souboru se budou hodnoty empirické distribuční funkce $F(x)$ ustalovat kolem hodnot distribuční funkce $\Phi(x)$. Vlastnosti empirické distribuční funkce se přenášejí i na distribuční funkci – je to funkce neklesající, zprava spojitá a normovaná v tom smyslu, že $\lim_{x \rightarrow -\infty} \Phi(x) = 0$, $\lim_{x \rightarrow \infty} \Phi(x) = 1$

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

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

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

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

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

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

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 $\forall x \in \mathbb{R} : \pi(x) \geq 0$,

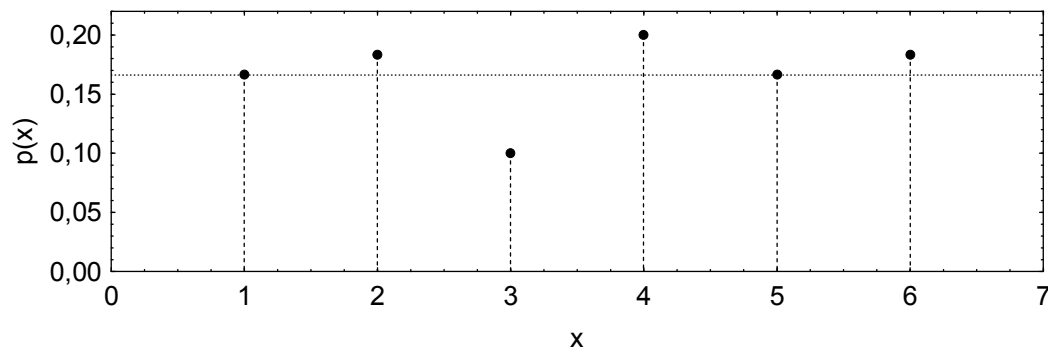
je normovaná $\sum_{x=-\infty}^{\infty} \pi(x) = 1$,

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

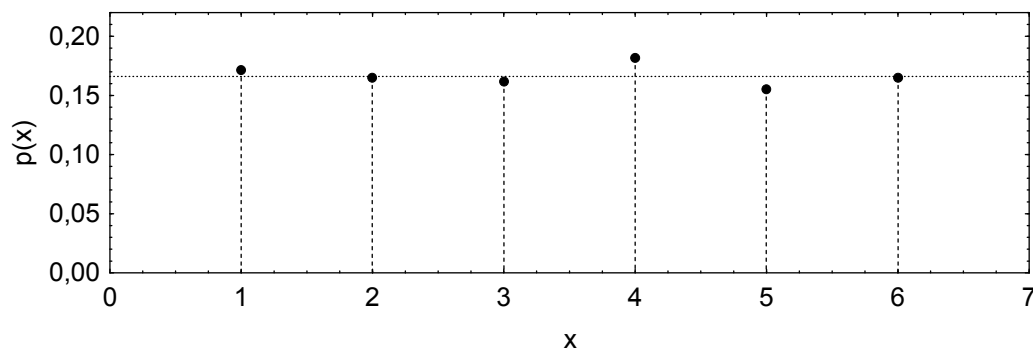
Ilustrace vztahu mezi četnostní funkcí a pravděpodobnostní funkcí

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

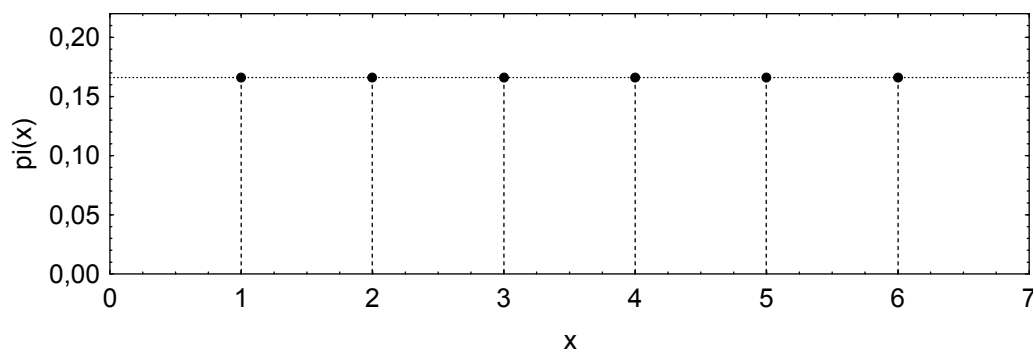
$n = 60$



$n = 600$



$n \rightarrow \infty$



Diskrétní náhodný vektor

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

$$\forall (x_1, \dots, x_n) \in \mathbb{R}^n : \pi(x_1, \dots, x_n) = P(X_1 = x_1 \wedge \dots \wedge X_n = x_n)$$

Pravděpodobnostní funkce libovolného diskrétního marginálního náhodného vektoru se nazývá marginální pravděpodobnostní funkce. Největší význam mají jednorozměrné marginální pravděpodobnostní funkce $\pi_1(x_1), \dots, \pi_n(x_n)$ jednotlivých složek náhodného vektoru.

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 spojitě 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 třídících intervalů se budou hodnoty hustoty četnosti ustalovat kolem hodnot hustoty pravděpodobnosti.

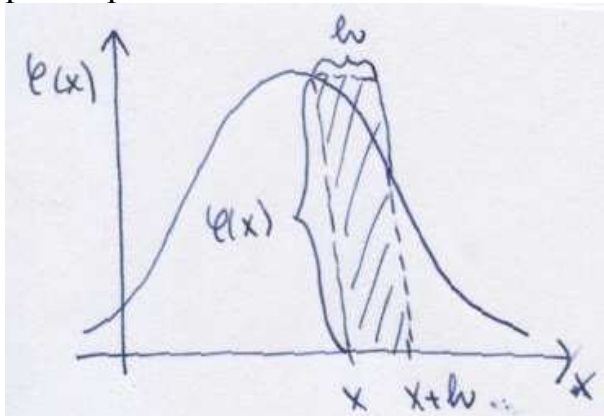
Vlastnosti hustoty četnosti se přenášejí i na hustotu pravděpodobnosti, tedy hustota pravděpodobnosti

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

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

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

Pozor – na rozdíl od pravděpodobnostní funkce diskrétní náhodné veličiny nemá hustota pravděpodobnosti spojitě 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:

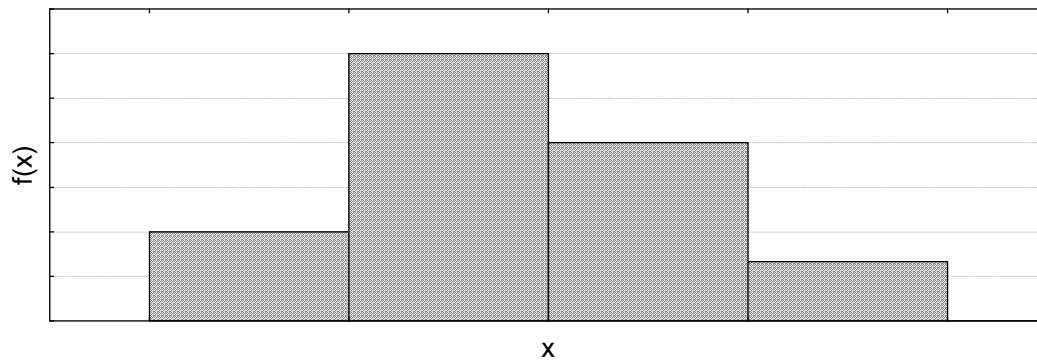
$$P(x < X \leq x+h) = \Phi(x+h) - \Phi(x) = \int_{-\infty}^{x+h} \varphi(t) dt - \int_{-\infty}^x \varphi(t) dt = \int_x^{x+h} \varphi(t) dt$$

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

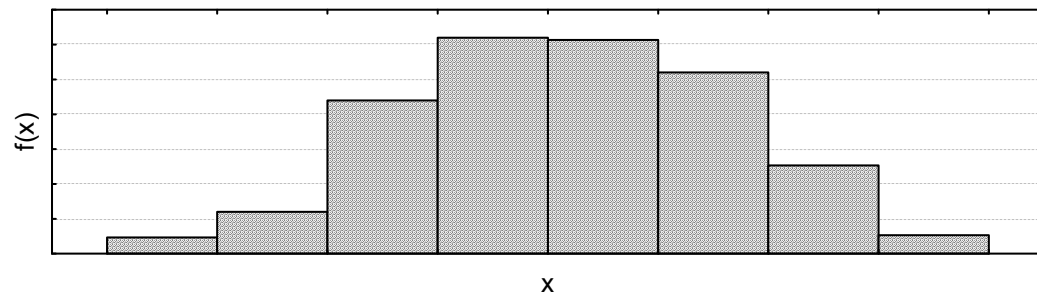
Ilustrace vztahu mezi hustotou četnosti a hustotou pravděpodobnosti

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

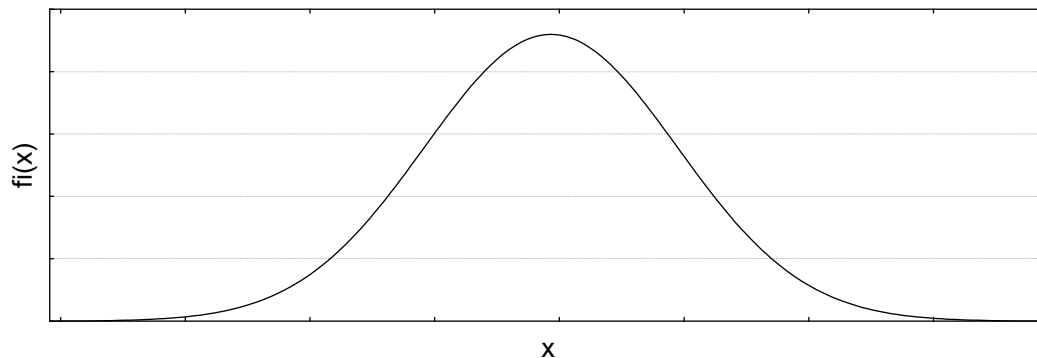
$$n = 40, r = 4$$



$$n = 400, r = 8$$



$$n \rightarrow \infty, r \rightarrow 0$$



Spojité náhodný vektor

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

Hustota pravděpodobnosti libovolného spojitého marginálního náhodného vektoru se nazývá marginální hustota pravděpodobnosti. Největší význam mají jednorozměrné marginální hustoty pravděpodobnosti $\varphi_1(x_1), \dots, \varphi_n(x_n)$ jednotlivých složek náhodného vektoru.

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:

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

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

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:

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

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

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

Vybraná rozložení diskrétních a spojitých náhodných veličin

Motivace

Nyní se seznámíme s přehledem důležitých pravděpodobnostních funkcí a hustot pravděpodobnosti. Uvedeme nejenom analytické vyjádření těchto funkcí, ale též jejich grafy. Vysvětlíme rovněž, v jakých situacích se lze s uvedenými rozloženími pravděpodobností setkat. Zvláštní pozornost budeme věnovat normálnímu rozložení, které hraje velkou roli v celé řadě praktických aplikací počtu pravděpodobnosti a jak uvidíme později, i v matematické statistice.

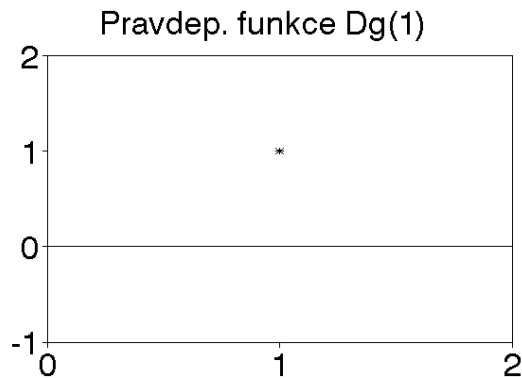
Označení

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 $X \sim L(\vartheta)$ čteme: náhodná veličina X má rozložení L s parametrem ϑ .

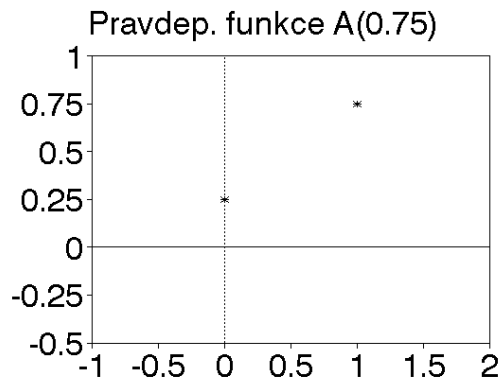
Degenerované rozložení: Náhodná veličina X nabývá pouze konstantní hodnoty μ , píšeme $X \sim Dg(\mu)$.

$$\pi(x) = \begin{cases} 1 & \text{pro } x = \mu \\ 0 & \text{jinak} \end{cases}$$



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

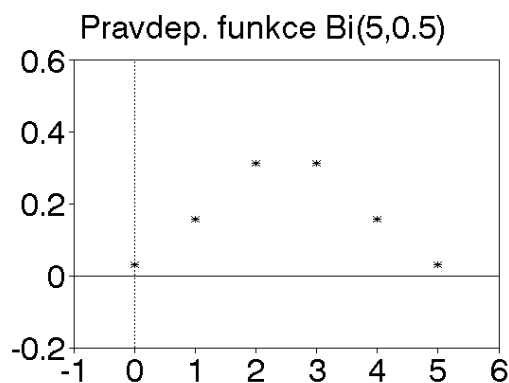
$$\pi(x) = \begin{cases} 1 - \vartheta & \text{pro } x = 0 \\ \vartheta & \text{pro } x = 1 \\ 0 & \text{jinak} \end{cases} \quad \text{neboli } \pi(x) = \begin{cases} \vartheta^x (1 - \vartheta)^{1-x} & \text{pro } x = 0, 1 \\ 0 & \text{jinak} \end{cases}$$



Binomické rozložení: Náhodná veličina X udává počet úspěchů v posloupnosti n nezávislých opakovaných pokusů, přičemž pravděpodobnost úspěchu je v každém pokusu ϑ . Píšeme $X \sim \text{Bi}(n, \vartheta)$.

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

(Alternativní rozložení je speciálním případem binomického rozložení pro $n = 1$. Jsou-li X_1, \dots, X_n stochasticky nezávislé náhodné veličiny, $X_i \sim A(\vartheta)$, $i = 1, \dots, n$, pak $X = \sum_{i=1}^n X_i \sim \text{Bi}(n, \vartheta)$.)



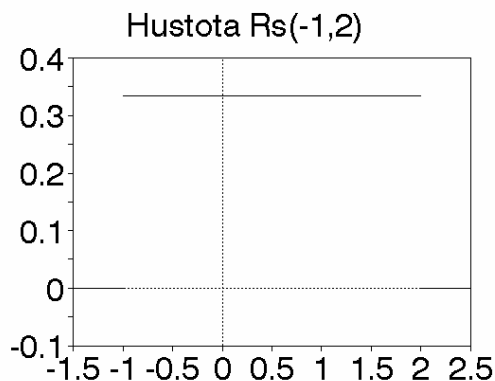
Rovnoměrné spojité rozložení: Předpokládejme, že veličina X

- může nabýt jakékoliv hodnoty mezi čísly a a b
- pravděpodobnost, že nabude hodnoty z jakéhokoliv intervalu v tomto rozmezí je stejná jako pravděpodobnost, že nabude hodnoty z jakéhokoliv jiného intervalu stejné délky.

Jsou-li tyto podmínky splněny, pak X má rovnoměrné spojité rozložení na intervalu (a, b) . Hustota pravděpodobnosti náhodné veličiny X je konstantní na intervalu (a, b) a plocha pod křivkou hustoty tvoří obdélník.

Píšeme $X \sim \text{Rs}(a, b)$.

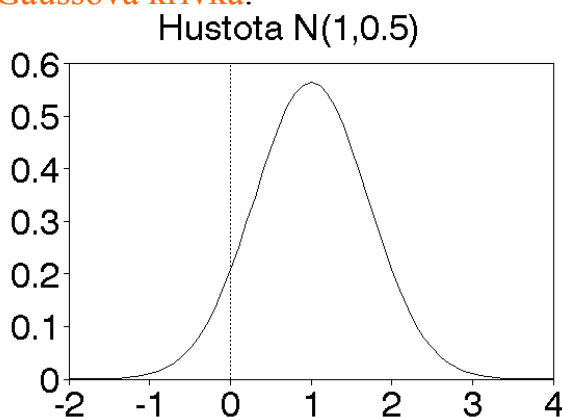
$$\varphi(x) = \begin{cases} \frac{1}{b-a} & \text{pro } x \in (a, b) \\ 0 & \text{jinak} \end{cases}$$



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.

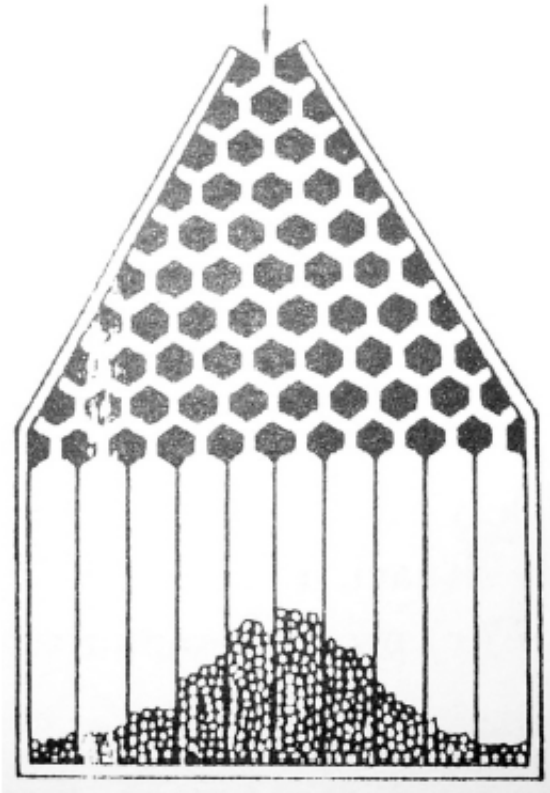


Ilustrace vzniku normálního rozložení pomocí Galtonovy desky:

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 radou je $n - 1$ přihrádek, ve kterých se kuličky shromažďují. 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 nespécifikovatelných vlivů, které ovlivňují zcela náhodně rozložení jeho výsledku.

Obrázek

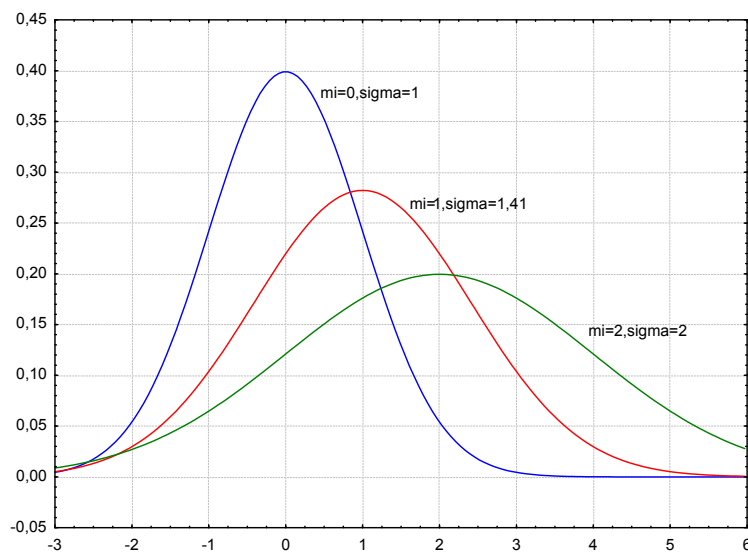


Standardizované normální rozložení:

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

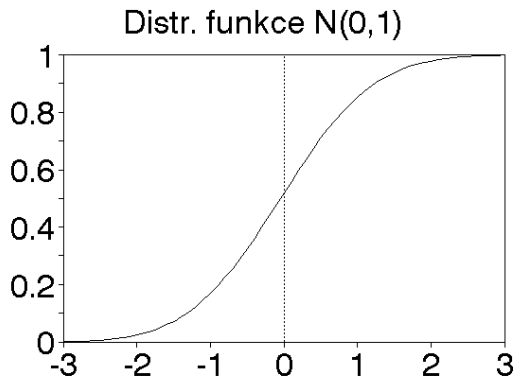
$U \sim N(0, 1)$. Hustota pravděpodobnosti má v tomto případě tvar $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$.

Vliv parametrů μ σ^2 na graf hustoty



Distribuční funkce standardizovaného normálního rozložení

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



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, \dots, X_n jsou stochasticky nezávislé náhodné veličiny, $X_i \sim N(\mu_i, \sigma_i^2)$,

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

Dvourozměrné normální rozložení: Náhodný vektor $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ vzniká ve dvou-

rozměrných situacích podobně jako skalární náhodná veličina v předešlém případě. Parametry μ_1, μ_2 popisují polohu realizací náhodných veličin X_1, X_2 na číselné ose, parametry σ_1^2, σ_2^2 popisují variabilitu realizací náhodných veličin X_1, X_2 kolem čísel μ_1, μ_2 a parametr ρ vystihuje sílu lineární závislosti mezi náhodnými veličinami X_1, X_2 .

Simultánní hustota má tvar:

$$\varphi(x_1, x_2) = \frac{1}{\sigma_1 \sigma_2 \sqrt{1 - \rho^2}} e^{-\frac{q(x_1, x_2)}{2}}, \text{ kde}$$

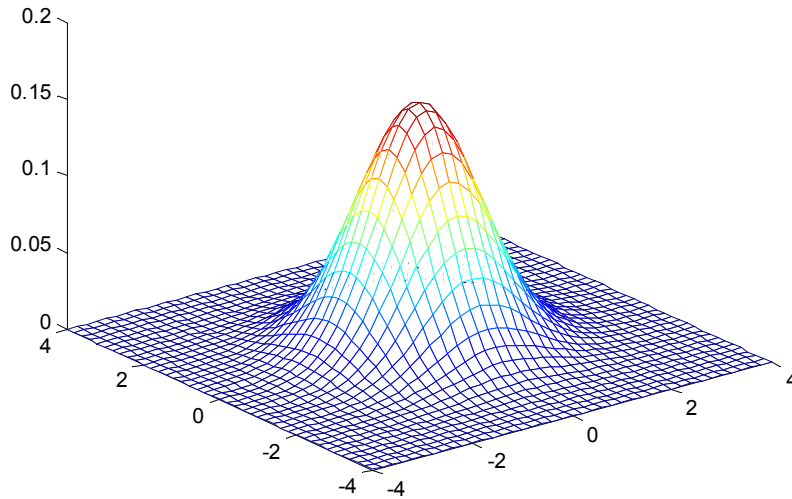
$$q(x_1, x_2) = \frac{1}{1 - \rho^2} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \cdot \frac{x_1 - \mu_1}{\sigma_1} \cdot \frac{x_2 - \mu_2}{\sigma_2} + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right].$$

$$\text{Píšeme } \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right).$$

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

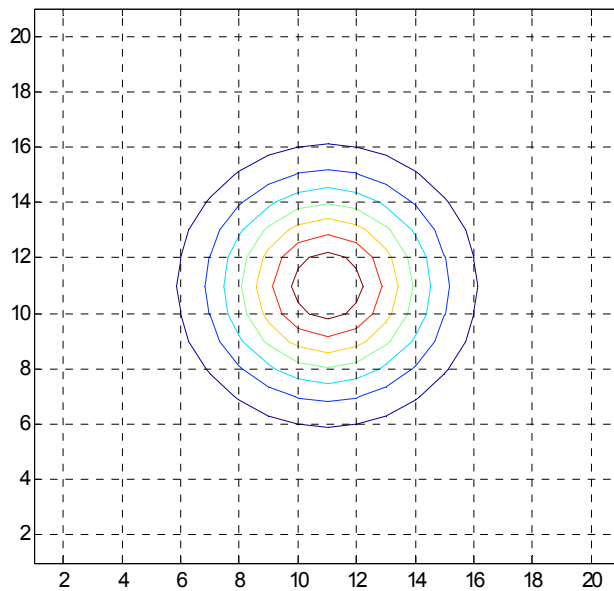
Graf hustoty rozložení $N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$

Graf dvourozrnmé hustoty



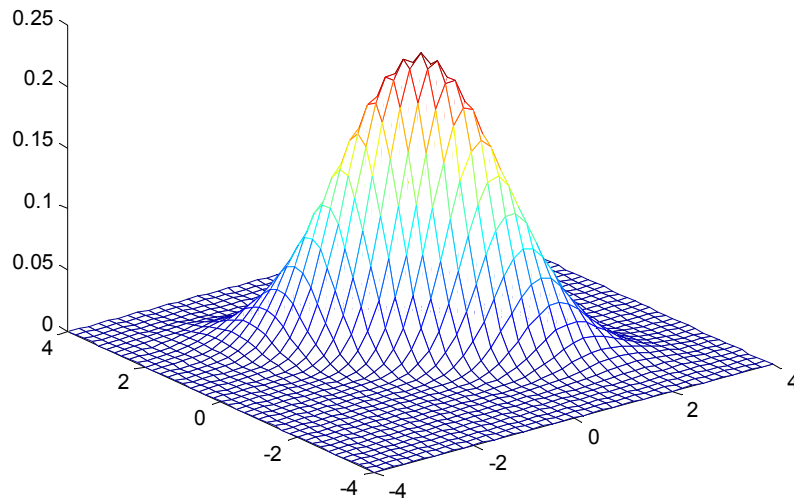
Vrstevnice hustoty rozložení $N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$

Vrstevnice normalní hustoty



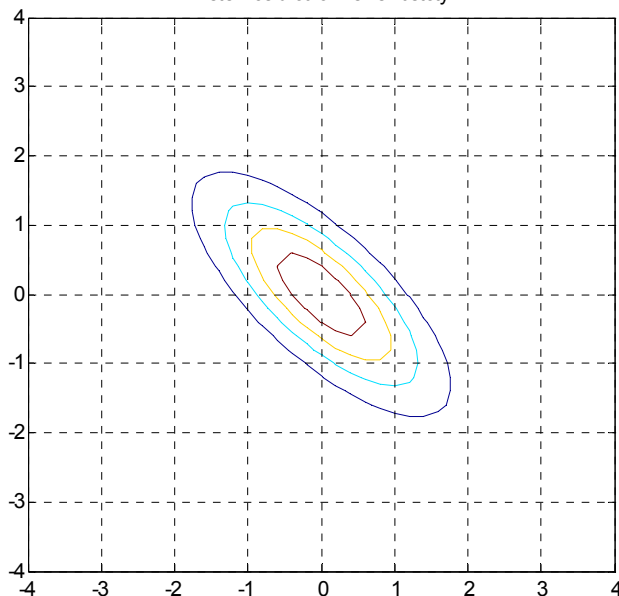
Graf hustoty rozložení $N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0,75 \\ -0,75 & 1 \end{pmatrix}\right)$

Graf dvourozmerne hustoty



Vrstevnice hustoty rozložení $N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0,75 \\ -0,75 & 1 \end{pmatrix}\right)$

Vrstevnice dvourozmerne hustoty



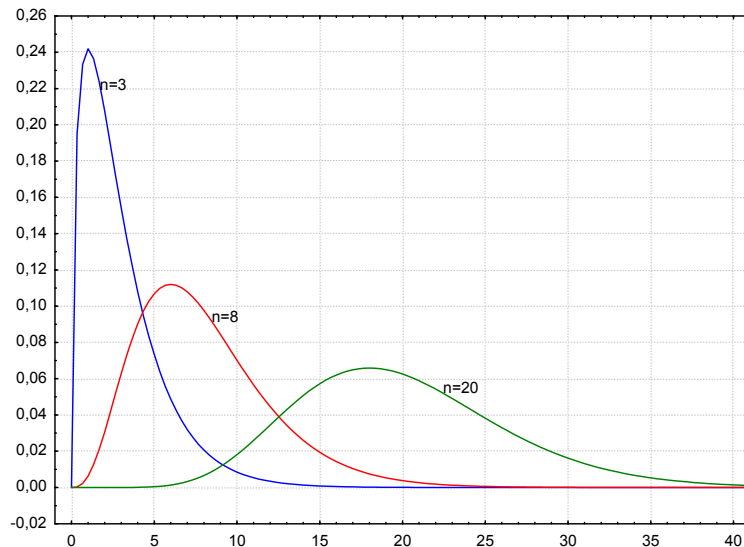
Některé vlastnosti dvourozměrného normálního rozložení:

Jestliže $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right)$, pak $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$

Jestliže $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right)$, pak X_1, X_2 jsou stochasticky nezávislé.

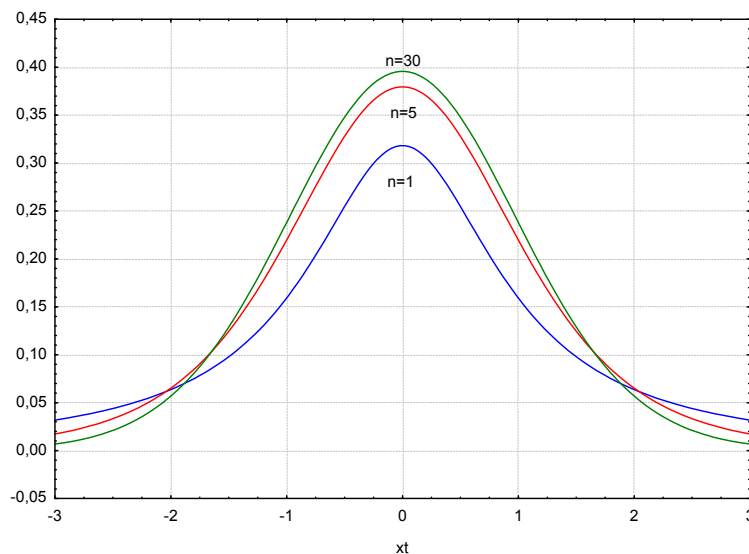
Upozornění: Následující tři rozložení – Pearsonovo, Studentovo a Fisherovo-Snedecorovo – jsou odvozena ze standardizovaného normálního rozložení. Mají velký význam především v matematické statistice při konstrukci intervalů spolehlivosti a testování hypotéz. Vyjádření hustot těchto rozložení neuvádíme,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

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



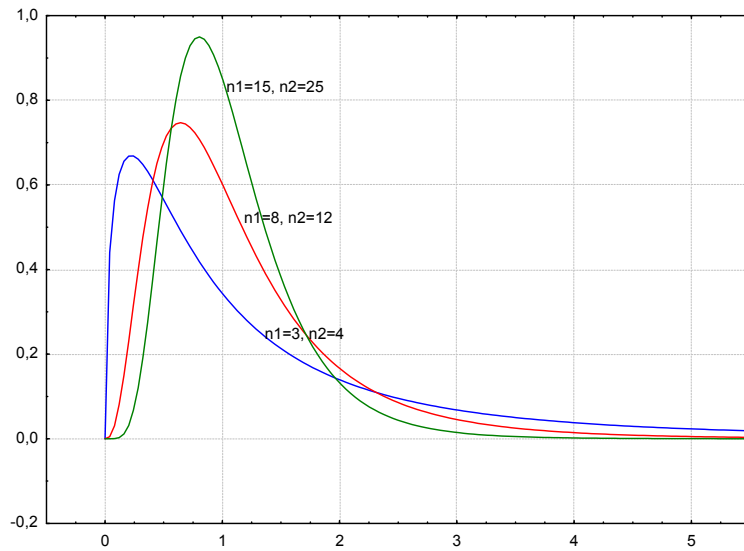
Studentovo rozložení s n stupni volnosti: Necht' X_1, X_2 jsou stochasticky nezávislé náhodné veličiny, $X_1 \sim N(0, 1)$, $X_2 \sim \chi^2(n)$.

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



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

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



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,

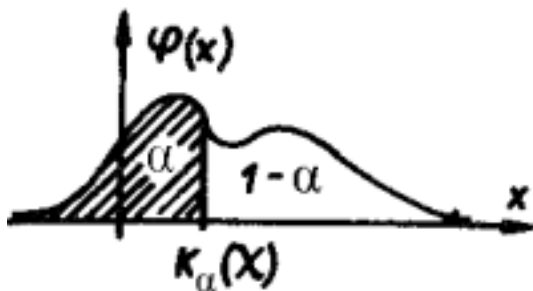
$K_{0,10}(X), \dots, K_{0,90}(X)$ - 1. až 9. decil,

$K_{0,01}(X), \dots, K_{0,99}(X)$ - 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.

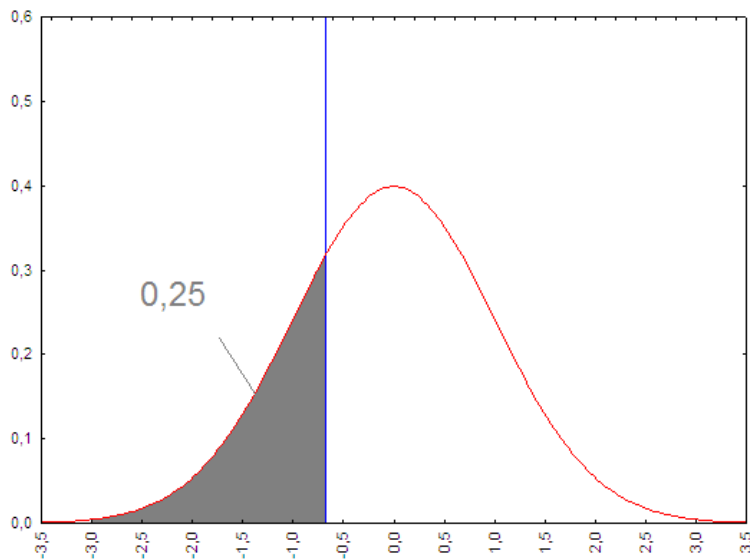
Používáme vztahy:

$$u_\alpha = -u_{1-\alpha},$$

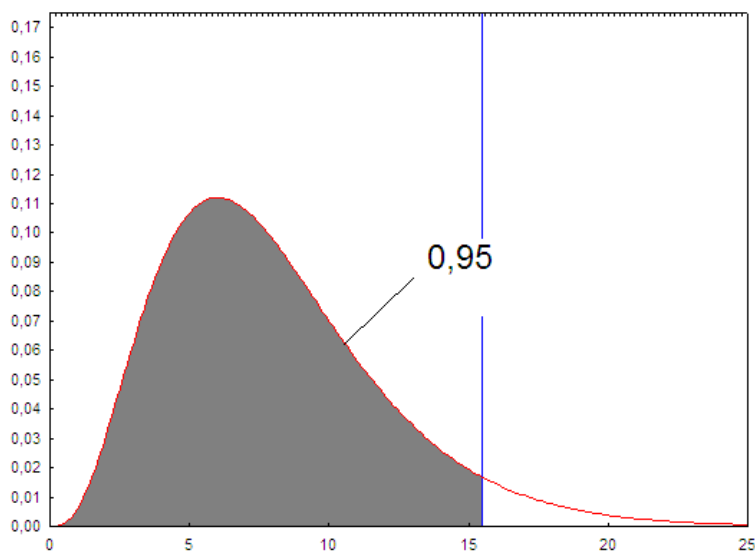
$$t_\alpha(n) = -t_{1-\alpha}(n),$$

$$F_\alpha(n_1, n_2) = \frac{1}{F_{1-\alpha}(n_2, n_1)}.$$

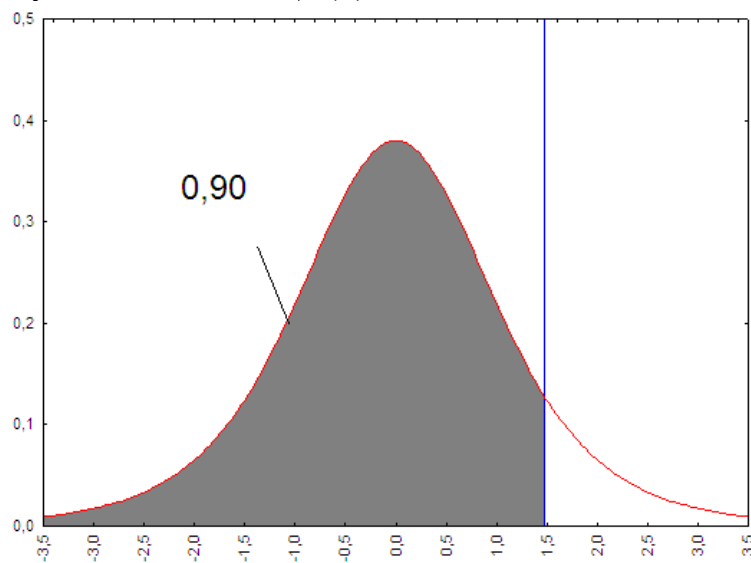
Význam kvantilu $u_{0,25} = -0,6745$



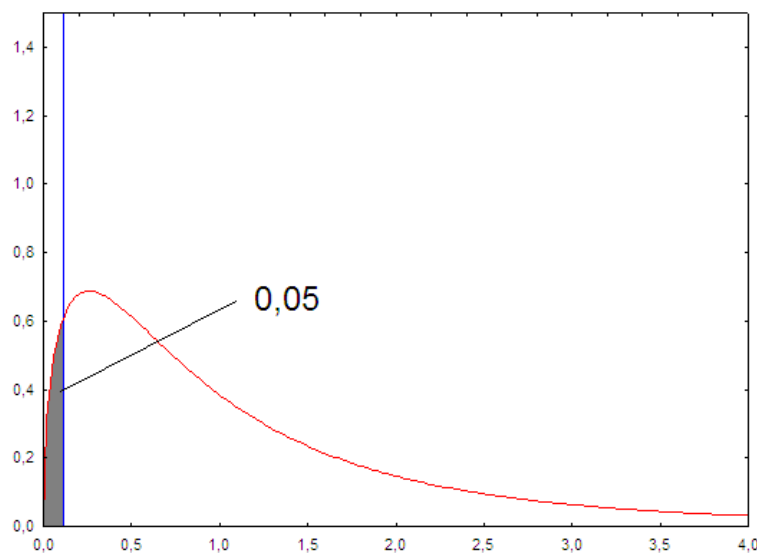
Význam kvantilu $\chi^2_{0,95}(8) = 15,5073$



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



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



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 $\pi(x)$.

Spojité 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 $\varphi(x)$.

Centrovaná náhodná veličina: $Y = X - E(X)$.
(Pro náhodnou veličinu Y platí: $E(Y) = 0$.)

Střední hodnota transformované náhodné veličiny $Y = g(X)$

$$E(Y) = \left\langle \begin{array}{l} \sum_{x=-\infty}^{\infty} g(x)\pi(x) \\ \int_{-\infty}^{\infty} g(x)\varphi(x)dx \end{array} \right\rangle$$

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

$$E(Y) = \left\langle \begin{array}{l} \sum_{x_1=-\infty}^{\infty} \sum_{x_2=-\infty}^{\infty} g(x_1, x_2)\pi(x_1, x_2) \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2)\varphi(x_1, x_2)dx_1dx_2 \end{array} \right\rangle$$

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

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

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

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

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

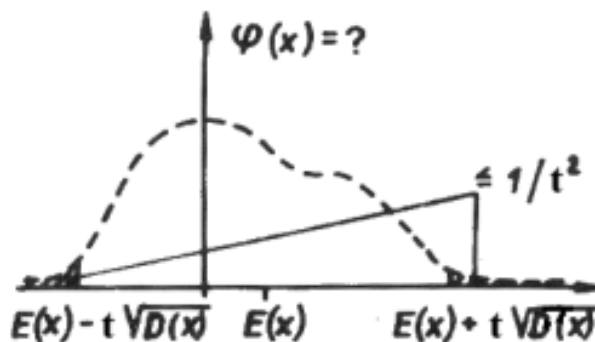
Čebyševova nerovnost:

Jestliže náhodná veličin X má střední hodnotu $E(X)$ a rozptyl $D(X)$, pak

$$\forall t > 0 : P(|X - E(X)| > t\sqrt{D(X)}) \leq \frac{1}{t^2}.$$

(Význam: pokud neznáme rozložení náhodné veličiny, ale známe její střední hodnotu a rozptyl, pak můžeme odhadnout pravděpodobnost, s jakou se od své střední hodnoty odchýlí o více než t -násobek své směrodatné odchylky.)

Ilustrace:



Pravidlo 3 σ : Necht' $E(X) = \mu$, $D(X) = \sigma^2$. Pak $P(|X - \mu| > 3\sigma) \leq \frac{1}{3^2} = \frac{1}{9} = 0,1$. (Tento výsledek je znám jako pravidlo 3 σ a říká, že nejvýše 11,1% realizací náhodné veličiny leží vně intervalu $(\mu - 3\sigma, \mu + 3\sigma)$.)

Střední hodnoty a rozptyly vybraných diskrétních a spojitých rozložení

$$X \sim Dg(\mu) \Rightarrow E(X) = \mu, D(X) = 0$$

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

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

$$X \sim Rs(a, b) \Rightarrow E(X) = \frac{a+b}{2}, D(X) = \frac{(b-a)^2}{12}$$

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

Charakteristika společné variability: **kovariance** $C(X_1, X_2)$ – číslo, které charakterizuje variabilitu realizací dvou náhodných veličin X_1, X_2 kolem jejich středních hodnot s přihlédnutím k pravděpodobnostem těchto realizací.

Definiční vzorec: $C(X_1, X_2) = E([X_1 - E(X_1)][X_2 - E(X_2)])$ (kovariance je střední hodnota součinu centrovaných náhodných veličin).

Výpočetní vzorec: $C(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$ (kovariance je střední hodnota součinu mínus součin středních hodnot).

$$C(X_1, X_2) = \left\{ \begin{array}{l} \sum_{x_1=-\infty}^{\infty} \sum_{x_2=-\infty}^{\infty} x_1 x_2 \pi(x_1, x_2) - \sum_{x_1=-\infty}^{\infty} x_1 \pi_1(x_1) \sum_{x_2=-\infty}^{\infty} x_2 \pi_2(x_2) \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 \varphi(x_1, x_2) dx_1 dx_2 - \int_{-\infty}^{\infty} x_1 \varphi_1(x_1) dx_1 \int_{-\infty}^{\infty} x_2 \varphi_2(x_2) dx_2 \end{array} \right.$$

Význam kovariance: 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.

Charakteristika těsnosti lineárního vztahu: **koeficient korelace** $R(X_1, X_2)$ - 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.

Definiční vzorec: $R(X_1, X_2) = E\left(\frac{X_1 - E(X_1)}{\sqrt{D(X_1)}} \cdot \frac{X_2 - E(X_2)}{\sqrt{D(X_2)}}\right)$ pro

kladné směrodatné odchylky, jinak klademe $R(X_1, X_2) = 0$ (koeficient korelace je střední hodnota součinu standardizovaných náhodných veličin).

Výpočetní vzorec: $R(X_1, X_2) = \frac{C(X_1, X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}}$ (koeficient korelace je

podíl kovariance a součinu směrodatných odchylek).

Cauchyova – Schwarzova – Buňakovského nerovnost:

$|R(X_1, X_2)| \leq 1$ a rovnost nastane tehdy a jen tehdy, když mezi veličinami X_1, X_2 existuje s pravděpodobností 1 úplná lineární závislost, tj. existují konstanty a_1, a_2 tak, že $P(X_2 = a_1 + a_2 X_1) = 1$.

Centrální limitní věta:

Jsou-li náhodné veličiny X_1, \dots, 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 roz-

ložení součtu $\sum_{i=1}^n X_i$ aproximovat normálním rozložením $N(n\mu, n\sigma^2)$. Zkráceně

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

Pokud součet $\sum_{i=1}^n X_i$ standardizujeme, tj. vytvoříme náhodnou veličinu

$U_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$, 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.

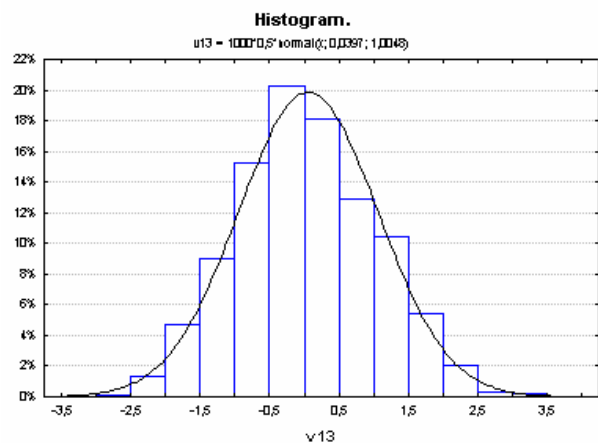
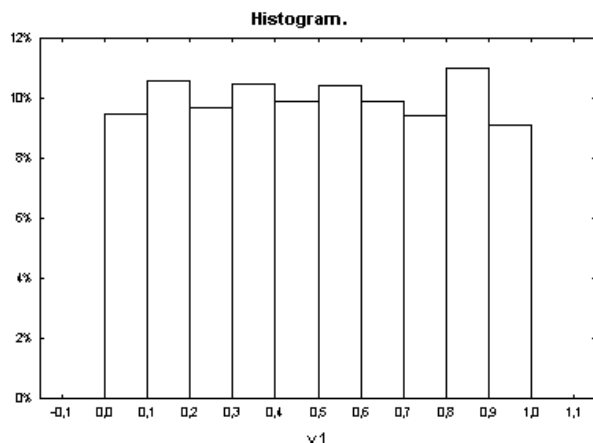
Ilustrace centrální limitní věty:

Uvažme n stochasticky nezávislých náhodných veličin X_1, \dots, X_n , přičemž každá z nich má rovnoměrné spojité rozložení na intervalu $(0,1)$, tj. $X_i \sim Rs(0,1)$, $i = 1, \dots, n$. Protože $E(X_i) = \frac{1}{2}$ a $D(X_i) = \frac{1}{12}$, podle centrální limitní věty náhodná

veličina $U_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^n X_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \approx N(0,1)$.

Položíme-li $n = 12$, pak $U_{12} = \sum_{i=1}^{12} X_i - 6 \approx N(0,1)$.

Při ilustraci působení centrální limitní věty na počítači postupujeme tak, že pro každou z 12 veličin $X_i \sim Rs(0,1)$, $i = 1, \dots, 12$ vygenerujeme dostatečně velký počet realizací, např. 1000 a uložíme je do proměnných v_1 až v_{12} . Do proměnné v_{13} uložíme součet proměnných v_1 až v_{12} zmenšený o 6. Histogram kterékoliv proměnné v_1 až v_{12} se svým tvarem bude blížit obdélníku, zatímco histogram proměnné v_{13} se svým tvarem bude blížit Gaussově křivce.



Důsledkem centrální limitní věty je **Moivreova – Laplaceova věta**:

Nechť X_1, \dots, X_n jsou stochasticky nezávislé náhodné veličiny, všechny se řídí alternativním rozložením $A(\vartheta)$. Pak jejich součet $Y_n = \sum_{i=1}^n X_i$ má binomické roz-

ložení $Bi(n, \vartheta)$. Střední hodnota veličiny Y_n je $E(Y_n) = n\vartheta$, rozptyl je $D(Y_n) = n\vartheta(1-\vartheta)$. Podle centrální limitní věty se standardizovaná náhodná veličina

$\frac{Y_n - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}}$ asymptoticky řídí standardizovaným normálním rozložením $N(0,1)$.

Aproximace se považuje za vyhovující, když jsou splněny podmínky:

$$\frac{1}{n+1} < \vartheta < \frac{n}{n+1} \text{ a } n\vartheta(1-\vartheta) > 9.$$

Na základě Moivreovy – Laplaceovy věty se používá aproximační vzorec, který složitý výpočet distribuční funkce binomického rozložení nahrazuje jednoduchým hledáním v tabulkách hodnot distribuční funkce standardizovaného normálního rozložení.

Máme náhodnou veličinu $Y_n \sim Bi(n, \vartheta)$. Pak

pravděpodobnostní funkce $P(Y_n = y) = \binom{n}{y} \vartheta^y (1-\vartheta)^{n-y}$ pro $y = 0, 1, \dots, n$,

distribuční funkce $P(Y \leq y) = \sum_{t=0}^y P(Y = t) = \sum_{t=0}^y \binom{n}{t} \vartheta^t (1-\vartheta)^{n-t}$ - složitý výpočet

Aproximační vzorec: $P(Y_n \leq y) \approx \Phi\left(\frac{y - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}}\right)$.

Příklad na aplikaci Moivreovy – Laplaceovy věty: 100 x nezávisle na sobě hodíme hrací kostkou. Jaká je pravděpodobnost, že šestka padne aspoň 20 x?

Řešení:

Y_{100} – počet šestek ve 100 hodech, $Y_{100} \sim Bi(100, \frac{1}{6})$.

Ověření podmínek dobré aproximace: $\frac{1}{n+1} < \vartheta < \frac{n}{n+1}$ a $n\vartheta(1-\vartheta) > 9$

$$\frac{1}{101} < \frac{1}{6} < \frac{100}{101} \text{ a } 100 \cdot \frac{1}{6} \cdot \left(1 - \frac{1}{6}\right) = 13,8\bar{8} > 9.$$

Aproximativní výpočet:

$$\begin{aligned} P(Y_{100} \geq 20) &= 1 - P(Y_{100} \leq 19) = 1 - P\left(\frac{Y_{100} - \frac{100}{6}}{\sqrt{100 \cdot \frac{1}{6} \cdot \frac{5}{6}}} \leq \frac{19 - \frac{100}{6}}{\sqrt{100 \cdot \frac{1}{6} \cdot \frac{5}{6}}}\right) = 1 - P(U_{100} \leq 0,626) \approx \\ &\approx 1 - \Phi(0,626) = 1 - 0,73565 = 0,26435 \end{aligned}$$

Přesný výpočet:

$$P(Y_{100} \geq 20) = 1 - P(Y_{100} \leq 19) = \sum_{t=0}^{19} \binom{100}{t} \left(\frac{1}{6}\right)^t \left(\frac{5}{6}\right)^{100-t} = 0,219751$$