

Alternativní rozdělení

Definice

Náhodná veličina X má *alternativní rozdělení* s parametrem p , jestliže nabývá hodnot 0 a 1 s pravděpodobnostmi

$$P[X = 1] = p \quad \text{a} \quad P[X = 0] = 1 - p$$

pro nějaké $p \in (0, 1)$. Značíme $X \sim \text{Alt}(p)$.

- dva možné výsledky: 1 (úspěch) a 0 (neúspěch)
- alternativním rozdělením modelujeme indikátor náhodného jevu (úspěch/neúspěch) apod.
- parametr p často nazýváme pravděpodobnost úspěchu

Alternativní rozdělení

Platí:

- $EX = p, \quad \text{var } X = p(1 - p).$
- distribuční funkce má skoky v bodech 0 a 1 o velikostech $1 - p$ a p , jinak je konstantní

Binomické rozdělení

Motivace:

- n nezávislých pokusů, v každém z nich nastane úspěch s pravděpodobností p
- zajímáme se o počet úspěchů (náhodná veličina)
- příklad: n hodů kostkou, zajímá nás počet šestek, které padly
- odvození *rozdělení*:
počet úspěchů je celé číslo mezi 0 a n ,
pravděpodobnost, že zaznamenáme právě k úspěchů je rovna

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

Binomické rozdělení

Definice

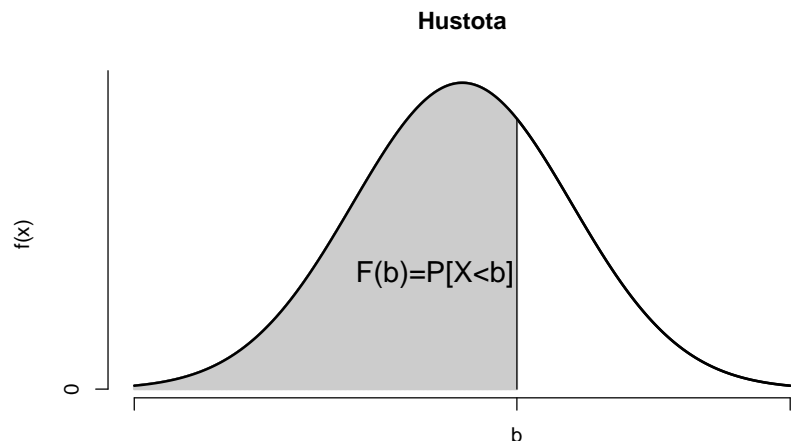
Náhodná veličina X má *binomické rozdělení* s parametry $n \geq 1$ a $p \in (0, 1)$, jestliže nabývá hodnot $0, \dots, n$ s pravděpodobnostmi

$$P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, \dots, n.$$

Značíme $X \sim \text{Bi}(n, p)$.

Poznámka: Platí $\sum_{k=0}^n P[X = k] = 1$ (z binomické věty).

Pravděpodobnosti binomického rozdělení



Příklad

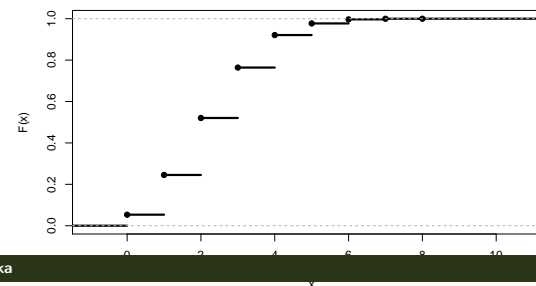
Příklad: Test obsahuje 10 otázek, ke každé z nich jsou uvedeny 4 možnosti: a, b, c, d. U každé otázky je právě jedna odpověď správná. Předpokládejme, že student zaškrtnává odpovědi zcela náhodně. Označme X počet správně zodpovězených otázek.

- 1 Jaké rozdělení veličiny X ?
- 2 Jaký je očekávaný počet správně zodpovězených otázek?
- 3 Na úspěšné napsání testu je nutné správně zodpovědět alespoň 8 otázek. S jakou pravděpodobností se to studentovi povede?
- 4 Jaká je pravděpodobnost, že student zodpoví alespoň jednu otázku správně?

Binomické rozdělení

Platí

- $EX = np \quad \text{var } X = np(1 - p)$
- distribuční funkce je po částech konstantní a má skoky v bodech $0, 1, \dots, n$,
v bodě k je skok o velikosti $\binom{n}{k} p^k (1 - p)^{n-k}$

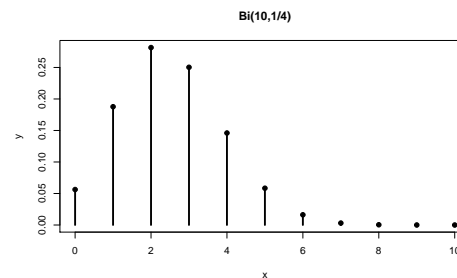


Příklad — řešení

1. Pro $k \in \{0, 1, \dots, n\}$ máme

$$P(X = k) = \binom{10}{k} \left(\frac{1}{4}\right)^k \left(1 - \frac{1}{4}\right)^{n-k}.$$

Tj. X má binomické rozdělení $\text{Bi}(10, 1/4)$.



Příklad— pokrač.

$$2. EX = \frac{10}{4} = 2.5.$$

Očekáváme, že student zodpoví čtvrtinu všech otázek správně, tj. v průměru dvě a půl otázky.

$$3. P(X \geq 8) = P(X = 8) + P(X = 9) + P(X = 10) =$$

(po dosazení) = 0.0004

$$4. P(X \geq 1) = P(X = 1) + \dots + P(X = 10) = \dots$$

Lepší postup:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \left(\frac{3}{4}\right)^{10} = 0.9437$$

Poissonovo rozdělení

Platí

$$EX = \lambda, \quad \text{var } X = \lambda.$$

Použití:

- Poissonovým rozdělením modelujeme počty událostí \iff kolikrát nastal určitý jev během jednotkového časového intervalu, na jednotkové ploše apod.
- parametr λ představuje intenzitu výskytu událostí
- **příklady:** počet autonehod na dálnici D1 za jeden den, počet srážkových dnů v měsíci, počet vadných pixelů na obrazovce, apod.
počet částic emitovaných radioaktivní látkou určité hmotnosti za jednotku času

Poissonovo rozdělení

Definice

Náhodná veličina X má **Poissonovo rozdělení** s parametrem λ , jestliže nabývá hodnot $0, 1, 2, \dots$ s pravděpodobnostmi

$$P[X = k] = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots,$$

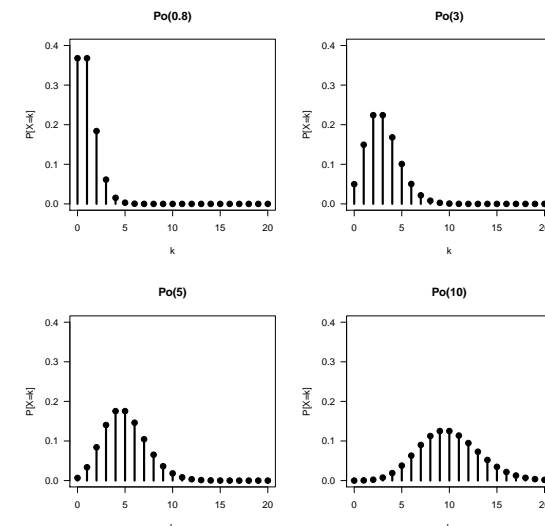
pro nějaké $\lambda > 0$. Značíme $X \sim \text{Po}(\lambda)$.

Jelikož

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda},$$

musí platit $\sum_{k=0}^{\infty} P[X = k] = 1$.

Pravděpodobnosti Poissonova rozdělení



Souvislost s binomickým rozdělením

- pro n velké lze $\text{Bi}(n, \frac{\lambda}{n})$ aproximovat pomocí $\text{Po}(\lambda)$ neboli
- pro velké n a malé p lze $\text{Bi}(n, p)$ aproximovat $\text{Po}(np)$
- příklad: počet nehod, ...
- výhoda: numerický výpočet je snazší pro Poissonovo rozdělení

Rovnoměrné rozdělení

Definice

Náhodná veličina X má *rovnoměrné rozdělení* na intervalu (a, b) , jestliže její hustota je

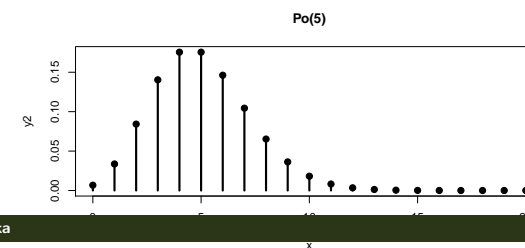
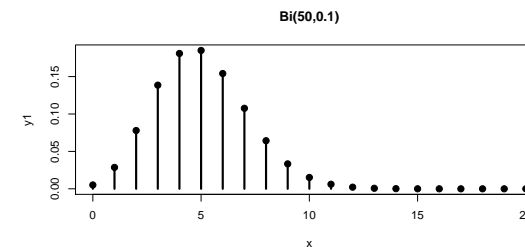
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{pro } x \in (a, b), \\ 0 & \text{jinak.} \end{cases}$$

Značíme $X \sim R(a, b)$.

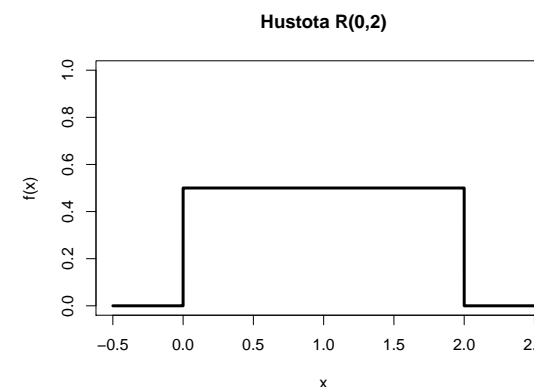
- hodnoty pouze z intervalu (a, b)
- hustota je na tomto intervalu konstantní $\rightsquigarrow X$ nabývá všech hodnot „rovnoměrně“, resp. „stejně pravděpodobně“

Vztah $\text{Bi}(n, \lambda/n)$ a $\text{Po}(\lambda)$

Porovnání pravděpodobností pro $n = 50$ a $\lambda = 5$



Hustota rovnoměrného rozdělení



- platí $\int_{-\infty}^{\infty} f(x) dx = 1$
- pravděpodobnost $P[c < X < d]$ závisí pouze na délce intervalu (c, d)

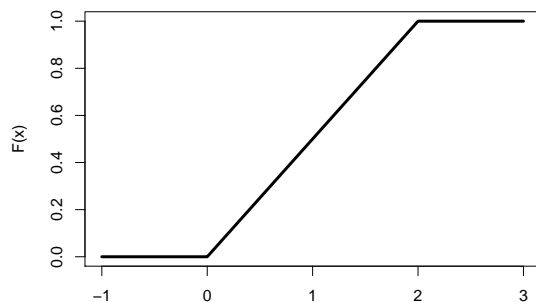
Rovnoměrné rozdělení

Distribuční funkce rovnoměrného rozdělení je

$$F(x) = \frac{x-a}{b-a} \quad \text{pro } x \in (a, b),$$

$$F(x) = 0 \quad \text{pro } x \leq a \quad \text{a} \quad F(x) = 1 \quad \text{pro } x \geq b.$$

Distr. funkce R(0,2)



Příklady

Příklady využití

- modelování doby čekání na událost, která nastává v pravidelných intervalech délky d
- např. čekání na příjezd autobusu (přicházíme-li v náhodný okamžik)

Charakteristiky rovnoměrného rozdělení

- Střední hodnota

$$EX = \frac{a+b}{2}$$

- Rozptyl

$$\text{var } X = \frac{(b-a)^2}{12}$$

- Medián je stejný jako střední hodnota, tj. $\frac{a+b}{2}$ (plyne ze symetrie rozdělení)
- Kvantil na hladině α

$$F(q_\alpha) = \frac{q_\alpha - a}{b - a} = \alpha$$

$$q_\alpha = \alpha(b - a) + a$$

Normální rozdělení

Definice

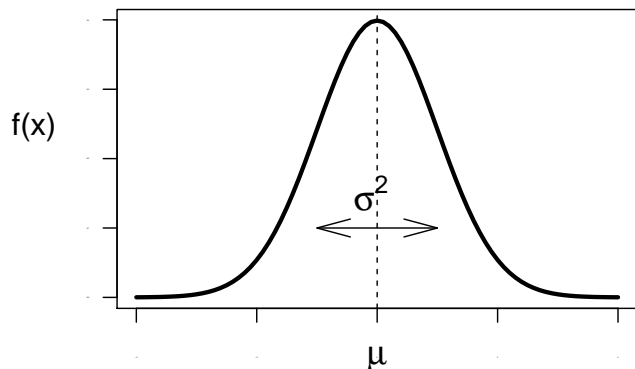
Náhodná veličina X má *normální rozdělení* s parametry $\mu \in \mathbb{R}$ a $\sigma^2 > 0$, jestliže má hustotu

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{pro } x \in \mathbb{R}.$$

Značíme $X \sim N(\mu, \sigma^2)$.

- nejdůležitější rozdělení ve statistice
- může nabývat jakýchkoliv reálných hodnot
- hustota je tzv. Gaussova křivka, je symetrická kolem μ a má v tomto bodě maximum
- parametr σ ovlivňuje „zploštělost“ hustoty

Hustota normálního rozdělení



Charakteristiky

Střední hodnota

$$EX = \mu$$

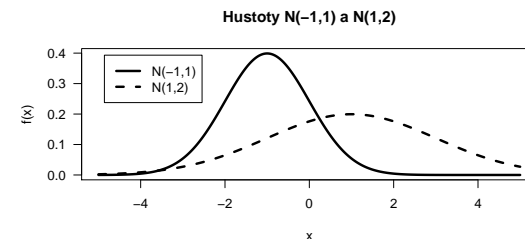
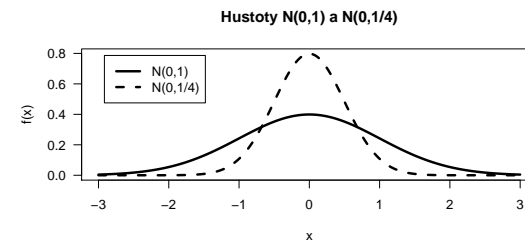
Rozptyl

$$\text{var } X = \sigma^2$$

Distribuční funkci je nutné počítat numericky.

Medián je stejný jako střední hodnota, tj. μ (ze symetrie).

Hustota normálního rozdělení



Využití

- výsledky experimentálního měření, chyby měření
- rozdělení kvantitativních znaků v populaci (výška, hmotnost, IQ) apod.
- má hezké teoretické vlastnosti
- má **výsadní postavení ve statistice**, neboť průměry většího počtu veličin mají rozdělení blízké normálnímu (přesněji viz **Centrální limitní věta**, která bude později).
- normální rozdělení je předpokladem řady statistických metod

Normované normální rozdělení

Normální rozdělení $N(0, 1)$ (s nulovou střední hodnotou a jednotkovým rozptylem) nazýváme **normované**. Jeho hustotu značíme φ , tj.

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Distribuční funkce $N(0, 1)$ se značí $\Phi(x)$ a počítá se jako

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

Nelze ji vyjádřit explicitně, její hodnoty lze najít v tabulkách.

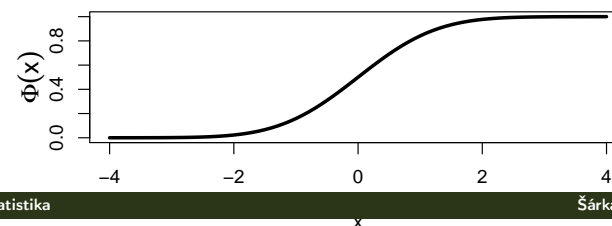
Distribuční funkce normálního rozdělení

Tabulka: Hodnoty distribuční funkce $N(0, 1)$

x	0	0.5	1	1.5	2	2.5	3
$\Phi(x)$	0.5	0.691	0.841	0.933	0.977	0.994	0.999

Přičemž platí

$$\Phi(-x) = 1 - \Phi(x).$$



Distribuční funkce normálního rozdělení

Důležité vztahy:

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

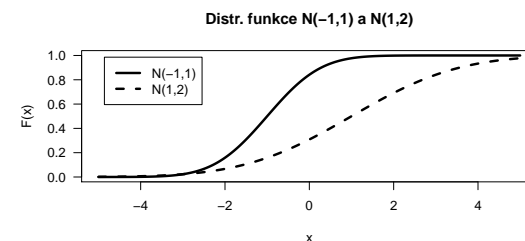
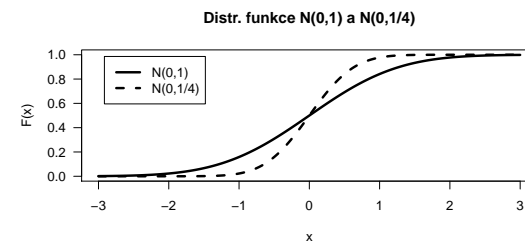
Distribuční funkci $N(\mu, \sigma^2)$ lze proto vyjádřit jako

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

a proto platí

$$P(a < X < b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

Distribuční funkce normálního rozdělení



Příklad

Příklad: Obsah vápníku v krevním séru (v mmol/l) se řídí normálním rozdělením $N(2.30, 0.18^2)$ u zdravých lidí a rozdělením $N(2.10, 0.18^2)$ u nemocných lidí. Diagnostický test zjistí obsah krevního séra v krvi a na základě něj je člověk diagnostikován jako nemocný nebo jako zdravý (nízké hodnoty signalizují nemoc), přičemž za hraniční hodnotu je bráno 2.15 mmol/l. (Pacienti označení jako nemocní jsou následně podrobeni detailnějším diagnostickým testům.)

- 1 Kolik procent zdravých osob bude falešně označeno za nemocné?
- 2 Kolik procent nemocných nebude odhaleno?

Kolik procent zdravých osob bude falešně označeno za nemocné?

Vezmeme $X \sim N(2.30, 0.18^2)$. Zajímá nás

$$\begin{aligned} P(X < 2.15) &= F(2.15) = \Phi\left(\frac{2.15 - 2.30}{0.18}\right) = \Phi(-0.833) \\ &= 1 - \Phi(0.833) = (\text{tabulky}) = 0.202. \end{aligned}$$

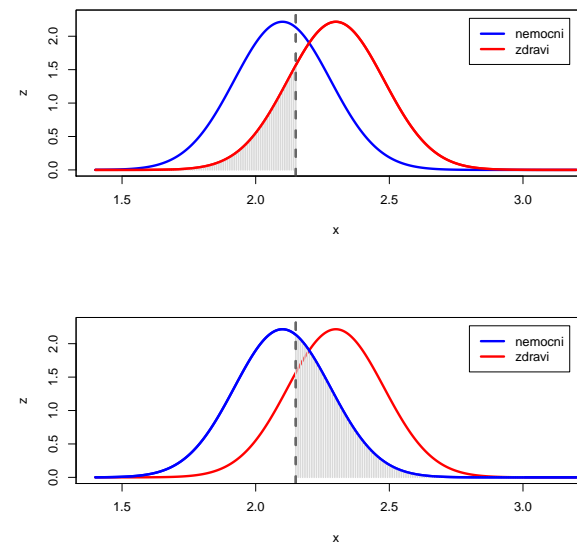
Tedy asi 20% zdravých bude zbytečně podrobněji testováno.

Kolik procent nemocných nebude odhaleno?

Vezmeme $Y \sim N(2.10, 0.18^2)$. Zajímá nás

$$\begin{aligned} P(Y > 2.15) &= 1 - F(2.15) = 1 - \Phi\left(\frac{2.15 - 2.10}{0.18}\right) = 1 - \Phi(0.278) \\ &= (\text{tabulky}) = 0.391. \end{aligned}$$

Tedy téměř 40% nemocných nebude odhaleno.



Pravidlo dvou (tří) sigma

Z tabelovaných hodnot $\Phi(x)$ plyne, že pro náhodnou veličinu $X \sim N(\mu, \sigma^2)$ platí

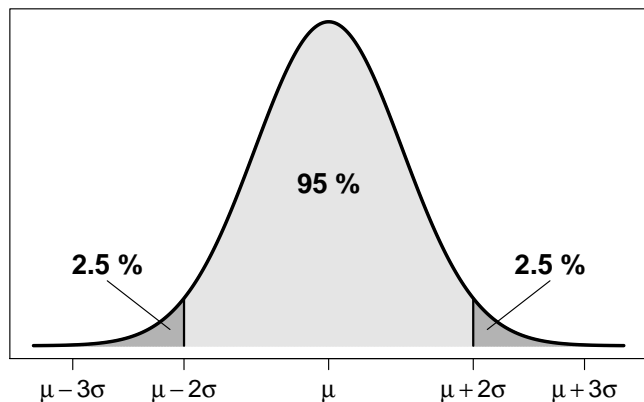
$$P[\mu - \sigma \leq X \leq \mu + \sigma] \approx 0.68$$

$$P[\mu - 2\sigma \leq X \leq \mu + 2\sigma] \approx 0.95$$

$$P[\mu - 3\sigma \leq X \leq \mu + 3\sigma] \approx 0.997$$

Náhodná veličina X tedy zřídka padá dále než $2\sigma-3\sigma$ od středu μ její hustoty.

Ilustrace



Příklad

Příklad krevní sérum:

Obsah vápníku v krvi zdravých lidí je s pravděpodobností přibližně 95 % v intervalu

$$(2.30 - 2 \cdot 0.18, 2.30 + 2 \cdot 0.18) = (1.94, 2.66)$$

(v mmol/l) a pouze 0.3% zdravých osob má obsah vápníku mimo interval

$$(2.30 - 3 \cdot 0.18, 2.30 + 3 \cdot 0.18) = (1.76, 2.84)$$

(v mmol/l).

Kvantily normovaného normálního rozdělení

- důležité ve statistice
- α -kvantily $N(0, 1)$ budeme značit jako u_α
- platí

$$P(X < u_\alpha) = \Phi(u_\alpha) = \alpha$$

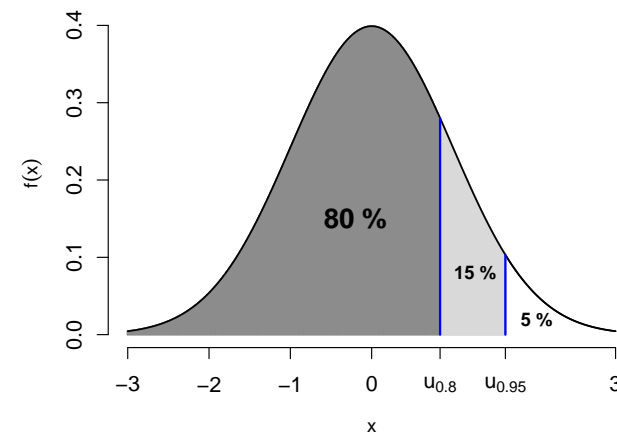
hodnoty jsou uvedeny v tabulkách a implementovány ve statistických (matematických) programech

- hodnoty $z_\alpha = u_{1-\alpha}$ splňují

$$P(X > z_\alpha) = \alpha$$

a nazývají se **kritické hodnoty** $N(0, 1)$

Kvantily normálního rozdělení



Tabulka: Vybrané kvantily rozdělení $N(0, 1)$

α	0.5	0.8	0.9	0.95	0.975	0.99	0.995
u_α	0	0.842	1.282	1.645	1.960	2.326	2.576

Jelikož platí $\Phi(-x) = 1 - \Phi(x)$, máme

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

Jakou je třeba zvolit hranici, aby pouze 5 % nemocných nebylo testem odhaleno?

Hledáme kvantil rozdělení $N(2.10, 0.18^2)$ na hladině 95 %.

Vezmeme $X \sim N(2.10, 0.18^2)$, pak

$$0.95 = P(X < q) = P\left(\frac{X - 2.1}{0.18} < \frac{q - 2.1}{0.18}\right) = \Phi\left(\frac{q - 2.1}{0.18}\right)$$

a odtud

$$1.645 = u_{0.95} = \frac{q - 2.1}{0.18}$$

a tedy $q = 2.1 + 0.18 \cdot 1.645 = 2.396$. Je tedy potřeba vzít za hraniční hodnotu 2.396 mmol/l.

Příklad

Příklad: Obsah vápníku v krevním séru (v mmol/l) se řídí normálním rozdělením $N(2.30, 0.18^2)$ u zdravých lidí a rozdělením $N(2.10, 0.18^2)$ u nemocných lidí.

- 1 Jakou je třeba zvolit hranici, aby pouze 5% nemocných nebylo testem odhaleno?
- 2 Jakou je třeba zvolit hranici, aby se maximálně 10% zdravých lidí muselo zbytečně podrobovat dalším testům?

Jakou je třeba zvolit hranici, aby se maximálně 10% zdravých lidí muselo zbytečně podrobovat dalším testům?

Hledáme kvantil rozdělení $N(2.30, 0.18^2)$ na hladině 10 %.

Vezmeme $Y \sim N(2.30, 0.18^2)$, pak

$$0.10 = P(Y < q) = P\left(\frac{Y - 2.3}{0.18} < \frac{q - 2.3}{0.18}\right) = \Phi\left(\frac{q - 2.3}{0.18}\right)$$

a odtud

$$-1.282 = u_{0.10} = \frac{q - 2.3}{0.18}$$

a tedy $q = 2.3 - 0.18 \cdot 1.282 = 2.069$. Je tedy potřeba vzít za hraniční hodnotu 2.069 mmol/l.

Exponenciální rozdělení

Definice

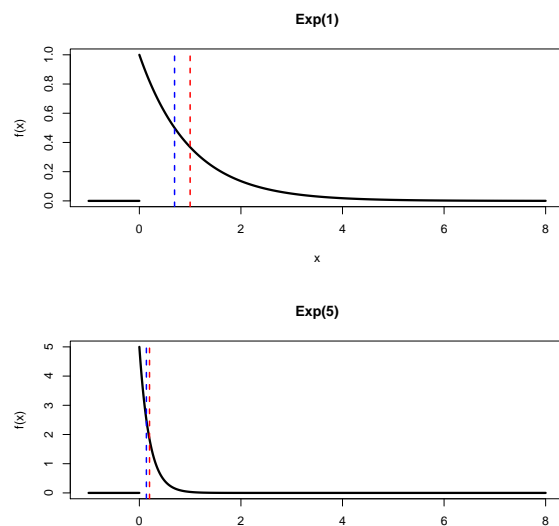
Náhodná veličina X má *exponenciální rozdělení* s parametrem $\lambda > 0$, jestliže má hustotu

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

Značíme $X \sim \text{Exp}(\lambda)$.

- nabývá pouze kladných hodnot,
- její hustota exponenciálně klesá \iff „nejpravděpodobnější“ jsou malé hodnoty

Hustota exponenciálního rozdělení



Charakteristiky

Střední hodnota

$$EX = \frac{1}{\lambda}$$

Čím vyšší je parametr λ , tím nižší je střední hodnota a naopak.

Distribuční funkce

$$F(x) = \int_{-\infty}^x f(t) dt = [-e^{-\lambda t}]_0^x = 1 - e^{-\lambda x} \quad \text{pro } x > 0$$

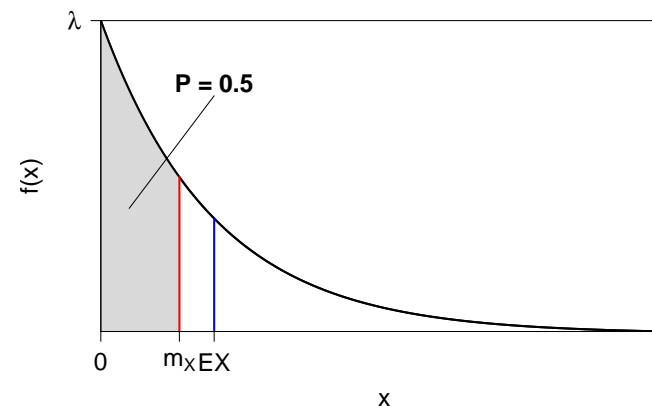
a $F_X(x) = 0$ pro $x \leq 0$.

Odtud pak vypočteme **medián**

$$\text{med}X = \frac{\log 2}{\lambda}$$

a platí $\text{med}X < EX$.

Medián a střední hodnota



Příklady

Příklady využití

- modelování doby čekání na událost
- délka telefonního hovoru, doba obsluhy u pokladny, apod.
- rozdělení kinetické energie molekuly ideálního plynu ve dvourozměrném prostoru je exponenciální

Příklady

Využití

- Maxwellovo rozdělení se používá pro modelování rychlosti molekul ideálního plynu. V tom případě je jeho parametr roven

$$a = \sqrt{\frac{kT}{m}},$$

kde k je Boltzmannova konstanta, T je teplota [K] a m je hmotnost částice [kg].

Maxwellovo rozdělení

Definice

Náhodná veličina X má *Maxwellovo rozdělení* s parametrem $a > 0$, jestliže má hustotu

$$f(x) = \begin{cases} \frac{2}{a^3\sqrt{2\pi}}x^2e^{-\frac{x^2}{2a^2}} & \text{pro } x > 0, \\ 0 & \text{jinak.} \end{cases}$$

- Maxwellova veličina X nabývá pouze kladných hodnot
- její hustota je mírně asymetrická a nabývá maximální hodnoty v bodě $x = \sqrt{2}a$

Charakteristiky

Distribuční funkci nelze vyjádřit explicitně, je nutné ji počítat numericky jako

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_0^{x^2/a^2} \sqrt{z} e^{-z/2} dz \quad \text{pro } x > 0.$$

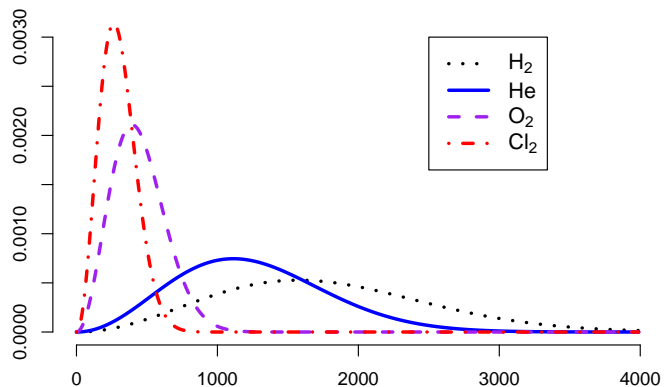
(Tj. i medián a kvantily lze spočítat jen numericky.)

Střední hodnota

$$EX = \sqrt{\frac{8}{\pi}}a = \sqrt{\frac{8}{\pi}} \frac{kT}{m}$$

Hustota Maxwellova rozdělení

Hustota Maxwellova rozdělení pro různé plyny při 25°C



Hustota Maxwellova rozdělení pro různé teploty

Hustoty rychlostí molekul kyslíku při teplotách 100, 300 a 500 K.

