NMAI059 Pravděpodobnost a statistika 1
6. přednáška

Robert Šámal

Přehled

Spojité náhodné veličiny

Konkrétní spojitá rozdělení a jejich parametry

Náhodné vektory

Obecná náhodná veličina – co už víme

- N.v. je zobrazení $X : \Omega \to \mathbb{R}$, které pro každé $x \in \mathbb{R}$ splňuje $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$.
- Diskrétní n.v. je n.v.
- Distribuční funkce n.v. X je funkce $F_X(x) := P(X \leq x)$.
- Distr. fce je F_X je neklesající, zprava spojitá, s definovanými limitami v $\pm\infty$.

Kvantilová funkce

$Q(\frac{1}{4})$ = kvantil, $Q(\frac{5}{100})$ = pátý percentil

Pro náhodnou veličinu X definujeme *kvantilovou funkci* $Q_X : [0,1] \to \mathbb{R}$ pomocí

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

inf

$$Q_X(p) := \min\{x \in \mathbb{R} : p \le F_X(x)\}$$

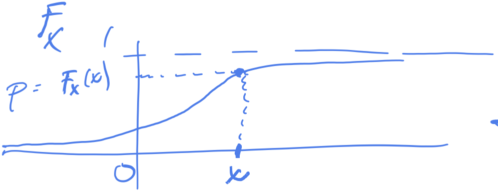
$\boxed{\text{a rostoucí}}$

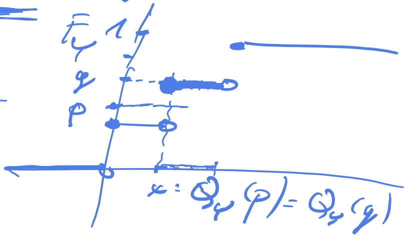
$Q_X(\frac{1}{2}) = x$

$F(x) = \frac{1}{2}$

n.v. X

- Pokud F_X je spojitá, tak $Q_X = F_X^{-1}$.
- Obecně platí: $Q_X(p) \le x \Leftrightarrow p \le F_X(x)$.
- $Q_X(1/2)$ = medián (pozor, když F_X není rostoucí)
- Pokud F_X je spojitá, tak $Q_X = F_X^{-1}$. def. na $(0,1)$





Spojitá náhodná veličina

$\lim_{x\to\infty} F_X(x) = 1$

$\boxed{\int_{-\infty}^{\infty} f_X(t)\,dt = 1}$ → f nemusí být spojitá

Definice

N.v. X se nazývá spojitá (continuous), pokud existuje nezáporná reálná funkce f_X tak, že

distrib. fce

$$\boxed{F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt.}$$

→ F je spojitá (dk. – věta z MA2?)

$\boxed{f = F'}$ tam kde je f spoj.

(Někdy se též používá pojem absolutně spojitá veličina.)
Funkce f_X se nazývá hustota (probability density function, pdf) náhodné veličiny X.

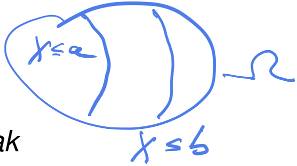
- Alternativně: máme zadanou funkci $f \geq 0$ s $\boxed{\int_{-\infty}^{\infty} f = 1.}$
- Vybereme náhodný bod pod grafem f.
- Označíme jeho souřadnice (X, Y).
- Pak X je n.v. s hustotou f.

$P(\text{bod} \in A) = \dfrac{\lambda(A)}{\lambda(\text{celé oblasti pod } f)} = \lambda(A)$ velikost

ověření: $P(X \leq x) = P((X,Y) \in M_x) = \lambda(M_x)$

Práce s hustotou $\frac{\text{velká}}{b-a} = \frac{1}{b-a}\int_a^b f$

$a \quad b$



Věta

Nechť spojitá n.v. X *má hustotu* f_X*. Pak*

1. $P(X = x) = 0$ *pro každé* $x \in \mathbb{R}$.
2. $P(a \leq X \leq b) = \int_a^b f_X(t)dt$ *pro každé* $a, b \in \mathbb{R}$.

- V důsledku taky platí (pro "rozumnou množinu A")

$$P(X \in A) = \int_A f_X(t)dt$$

„velká" $f_X(a) = \frac{P(a \leq X \leq a+\varepsilon)}{\varepsilon}$

Dk $\boxed{2 \Rightarrow 1}$: $P(x \leq X \leq x) = \int_x^x f = 0$

$\boxed{2}$ $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a) = \int_{-\infty}^b f - \int_{-\infty}^a f =$

$P(a \leq X \leq b) = \lim_{n\to\infty} P(a - \frac{1}{n} < X \leq b) = \lim_{n\to\infty} \int_{a-\frac{1}{n}}^b f = \int_a^b f$

$\| \quad \bigcap \{a - \frac{1}{n} < X \leq b\}$

Střední hodnota spojité n.v.

Definice

Nechť spojitá n.v. X má hustotu f_X. Pak její střední hodnota (expectation, expected value, mean) je označována $\mathbb{E}(X)$ a definována

$$\boxed{\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f_X(x)dx,}$$

pokud integrál má smysl, tj. pokud se „nejedná o typ $\infty - \infty$“.

$\sum_x x \cdot P(X=x)$ pro diskrétní n.v.

- Analogie s výpočtem těžiště tyče ze znalosti hustoty.
- Diskretizace.

$\Delta > 0$, $\Delta \to 0$

$n\Delta$, $(n+1)\Delta$

$$E(X) = \sum_{n\in\mathbb{Z}} \int_{n\Delta}^{(n+1)\Delta} x\, f_X(x)\, dx \doteq \sum_{n\in\mathbb{Z}} n\Delta \int_{n\Delta}^{(n+1)\Delta} f_X(x) = \sum n\Delta\, P(n\Delta \le X \le (n+1)\Delta) = E(X')$$

„zaokrouhlené X“

Vlastnosti střední hodnoty

podle def. bychom museli zjistit $f_{g(X)}$

Věta (LOTUS)

Pokud X je spojitá n.v. s hustotou f_X a g reálná funkce, tak

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx,$$

pokud integrál má smysl.

(Důkaz vynecháme, dělal by se pomocí substituce v integrálu.)

Věta (Linearita střední hodnoty)

Pro $X_1, \ldots, X_n$ diskrétní nebo spojité n.v. platí

$$\boxed{\mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n).}$$

(Důkaz bude později.)

Rozptyl spojité n.v.

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f_X(x)dx$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2\, f_X(x)dx$$

Označíme-li $\mu = \mathbb{E}(X)$, tak $\mathbb{E}((X-\mathbb{E}(X))^3)$

$$var(X) := \mathbb{E}((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2\, f_X(x)dx.$$

Věta

I pro spojité n.v. platí $var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

(Důkaz jako pro diskréní n.v.)

Rozptyl součtu

Věta (Rozptyl součtu)

Pro $X_1, \ldots, X_n$ nezávislé diskrétní nebo spojité n.v. platí

$$\boxed{var(X_1 + \cdots + X_n) = var(X_1) + \cdots + var(X_n).}$$

(X_1 lib., $X_2 = -X_1$ --- $var(X_1 + X_2) = var(0) = 0$

$var(X_1) + var(X_2)$ je „cokoli")

($X \sim Bin(n,p)$

$X_1 + \ldots + X_n$

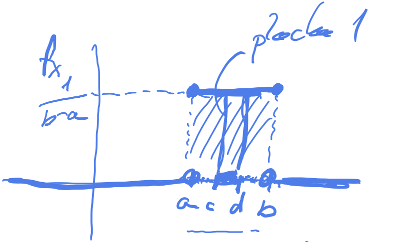
$var(X) = var X_1 + \ldots + var X_n$

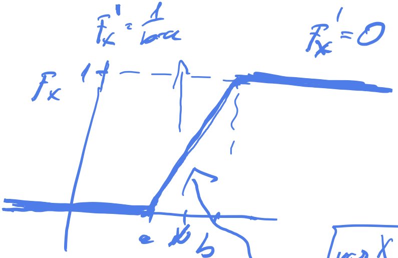
$= n \cdot var X_1 = np(1-p)$)

Přehled

Uniformní rozdělení

- N.v. X má uniformní rozdělení na intervalu $[a, b]$, píšeme $X \sim U(a, b)$, pokud $f_X(x) = 1/(b-a)$ pro $x \in [a, b]$ a $f_X(x) = 0$ jinak.





$$P(c \le X \le d) = \int_c^d f_X(t)\,dt$$

$$(a < c < d < b) \quad = \frac{d-c}{b-a}$$

$$F_X(x) = \frac{x-a}{b-a}$$

$(c = a,\ d = x)$

$$\boxed{\text{var}\,X \text{ — cvič.}}$$

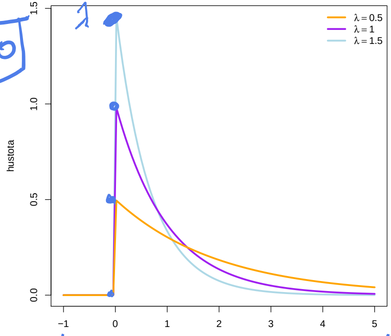
$$E(X) = \int_{-\infty}^{\infty} x f(x) = \int_a^b x \cdot \frac{1}{b-a} = \left[\frac{x^2}{2} \cdot \frac{1}{b-a}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \boxed{\frac{a+b}{2}}$$

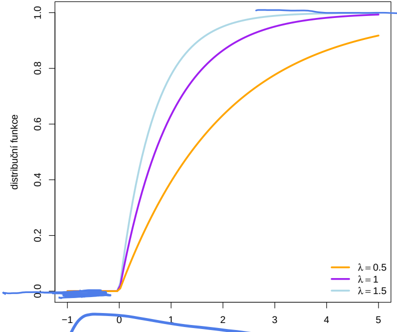
Exponenciální rozdělení

$$F_X(x) = \begin{cases} 0 & \text{pro } x \le 0 \\ 1 - e^{-\lambda x} & \text{pro } x \ge 0 \end{cases}$$

$$f_X(x) = \begin{cases} 0 & x < 0 \\ -e^{-\lambda x} \cdot (-\lambda) & x \ge 0 \end{cases} \quad \boxed{= \lambda e^{-\lambda x}}$$

$\boxed{\lambda > 0}$





$\int_{-\infty}^{\infty} f_X = 1 \Longleftarrow F_X(x) \to 1$

$\boxed{EX, \operatorname{var} X \text{ ... cvic.}}$

- X modeluje např. čas před příchodem dalšího telefonního hovoru do call-centra/dotazu na web-server/čas do dalšího blesku v bouřce/…

Souvislost $X \sim Exp(\lambda)$ a $Y \sim Geom(p)$

$X \doteq Y.\Delta$

- $P(X > x) = e^{-\lambda x}$ pro $x > 0$
- $P(Y > n) = (1-p)^n$ pro $n \in \mathbb{N}$

$P(Y.\Delta > n\Delta)$

$P(X > x) \doteq P(Y > x)$

$x = n\Delta$

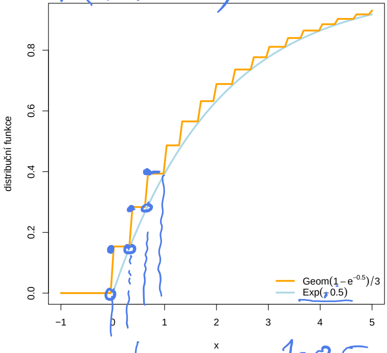
$e^{-\lambda n \Delta} \quad (1-p)^n$

$e^{-\lambda\Delta} \doteq 1-p$

$p \doteq \underbrace{1 - e^{-\lambda\Delta}}$

$1 - (1 - \lambda\Delta)$

$= \lambda\Delta$



$Y.\Delta$

$p = 1 - e^{-\lambda\Delta}$

$\Delta = \frac{1}{3}$ $\quad \lambda = 0.5$

$\lambda = \frac{p}{\Delta}$ ... intenzita

$\boxed{p \doteq \lambda\Delta}$

Standardní normální rozdělení

- $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
- $\Phi(x)$ – primitivní funkce k φ
- *Standardní normální rozdělení* $N(0,1)$ má hustotu φ a distribuční funkci Φ.
- Pokud $Z \sim N(0,1)$, tak $\mathbb{E}(Z) = 0$, $var(Z) = 1$

díky $\sqrt{2\pi}$

$\int_{-\infty}^{\infty} \varphi(x) = 1$ --- hezký výpočet, vynecháme

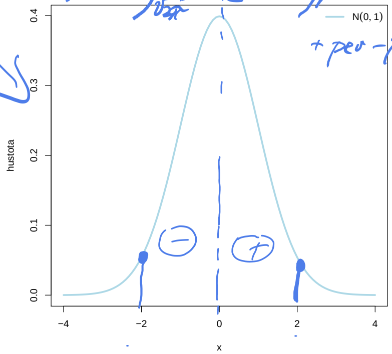
$\int_{-\infty}^{x} \varphi(t)\,dt$ distr. fce

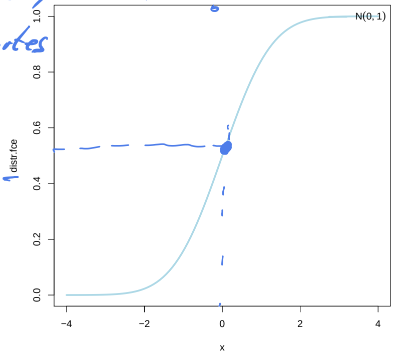
$\mathbb{E}(Z) = \int_{-\infty}^{\infty} t\varphi(t) = \int_{-\infty}^{0} t\varphi(t) + \int_{0}^{\infty} t\varphi(t) = 0$

$\int_0^{\infty} \frac{1}{\sqrt{2\pi}} t e^{-t^2/2} = \left[\frac{-1}{\sqrt{2\pi}} e^{-t^2/2}\right]_0^{\infty} = \frac{1}{\sqrt{2\pi}}$

$var\, Z = \mathbb{E}Z^2 = \int \frac{1}{\sqrt{2\pi}} t^2 e^{-t^2/2} = \int \frac{1}{\sqrt{2\pi}} t\cdot(t e^{-t^2/2}) = \dots = 1$

+ per-partes





Obecné normální rozdělení

- Pro $\mu, \sigma \in \mathbb{R}$, $\sigma > 0$ položme $\boxed{X = \mu + \sigma \cdot Z}$, kde $Z \sim N(0,1)$.
- Píšeme $X \sim N(\mu, \sigma^2)$ – obecné normální rozdělení.
- Normální rozdělení $N(\mu, \sigma^2)$ má hustotu $\boxed{\frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma})}$.

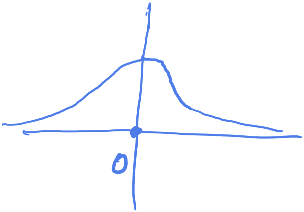
kolísá znač.

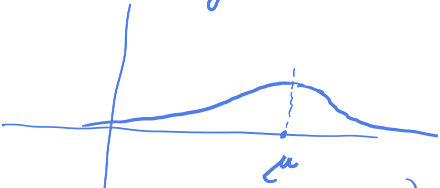
$N(\mu, \sigma^2)$ vs. $N(\mu, \sigma)$

z

$Z \longrightarrow X$

$f_X(x) = \frac{1}{\sigma}\varphi(z)$





$\Phi(z) = P(Z \le z) = P(X \le \mu + \sigma z) = F_X(\mu + \sigma z)$ der.

$\varphi(z)$ $\qquad x = \mu + \sigma z \qquad f_X(\mu + \sigma z)\sigma$

Odolnost vůči součtu

- Pokud $X_1, \dots, X_k$ jsou n.n.v., kde $X_i \sim N(\mu_i, \sigma_i^2)$, pak

$$X_1 + \cdots + X_k \sim N(\mu, \sigma^2),$$

kde $\mu = \mu_1 + \dots + \mu_k$ lin. stř.

$\sigma^2 = \sigma_1^2 + \dots + \sigma_k^2$ rozptyl součtu n.n.v.

Normální rozdělení – klíčové vlastnosti

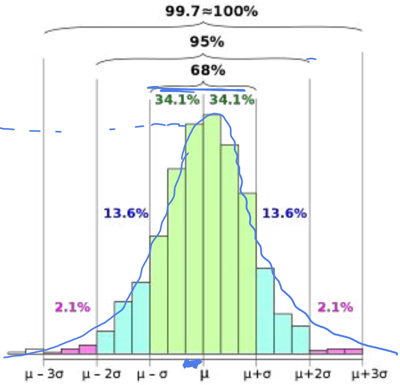
$X \sim N(\mu, \sigma^2)$

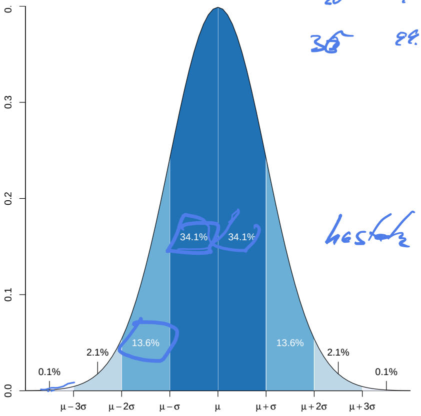
- Pravidlo 3σ (68–95–99.7 rule)
- Centrální limitní věta

$P(\mu-\sigma \le X \le \mu+\sigma) = 68\%$

2σ 95%

3σ 99.7%





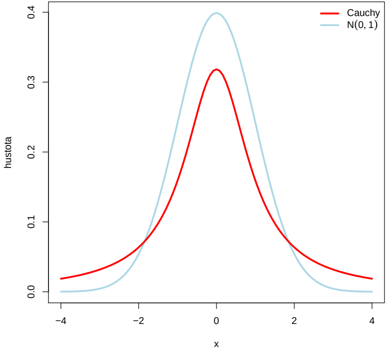
hustota

(Obrázek vlevo z Wikipedie, autor Melikamp.)

samplují hodnoty X, nakreslím histogram

Cauchyho rozdělení

- *Cauchyho rozdělení:* hustota $f(x) = \frac{1}{\pi(1+x^2)}$
- nemá střední hodnotu!!!



Gamma rozdělení

- $Gamma(w, \lambda)$, *gamma rozdělení s parametry* $w > 0$ *a* $\lambda > 0$ má hustotu

$$f(x) = \begin{cases} 0 & \text{pro } x \leq 0 \\ \frac{1}{\Gamma(w)} \lambda^w x^{w-1} e^{-\lambda x} & \text{pro } x \geq 0 \end{cases}$$

  kde $\Gamma(w) = (w-1)! = \int_0^\infty x^{w-1} e^{-x} dx$.
- Pro $w = 1$ dostáváme znovu exponenciální rozdělení.
- Pokud $X_1, \ldots, X_n$ jsou n.n.v. s rozdělením $Exp(\lambda)$, tak $X_1 + \cdots + X_n \sim Gamma(n, \lambda)$.
- Modeluje mj. životnost součástky, souhrn dešťových srážek za rok, latenci webového serveru.

A mnoho dalších

- $Beta(s,t)$ – beta rozdělení
- χ^2 rozdělení s k stupni volnosti = chí-kvadrát (χ^2_k) je jiné jméno pro $Gamma(\frac{1}{2}k, \frac{1}{2})$. Je to rozdělení $Z_1^2 + \cdots + Z_k^2$, kde $Z_i \sim N(0,1)$ jsou n.n.v.
- Studentova t-distribuce
- atd. atd.

Uniformní rozdělení

- N.v. X má uniformní rozdělení na intervalu $[a, b]$, píšeme $X \sim U(a, b)$, pokud $f_X(x) = 1/(b - a)$ pro $x \in [a, b]$ a $f_X(x) = 0$ jinak.

Universalita unif.

Věta

Nechť X je n.v. s distribuční funkcí $F_X = F$, nechť F je spojitá a rostoucí. Pak $F(X) \sim U(0,1)$.

Věta

Nechť F je funkce „typu distribuční funkce“: neklesající zprava spojitá funkce s $\lim_{x\to-\infty} F(x) = 0$ a $\lim_{x\to+\infty} F(x) = 1$. Nechť Q je odpovídající kvantilová funkce.
Nechť $U \sim U(0,1)$ a $X = Q(U)$. Pak X má distribuční funkci F.