NMAI059 Pravděpodobnost a statistika 1
6. přednáška

Robert Šámal

Přehled

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

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

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

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

Spojitá náhodná veličina

Definice

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

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

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

Práce s hustotou

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”)

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

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*

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

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

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

Vlastnosti střední hodnoty

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

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

(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

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

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*

$$var(X_1 + \cdots + X_n) = var(X_1) + \cdots + var(X_n).$$

Přehled

Spojité náhodné veličiny

Konkrétní spojitá rozdělení a jejich parametry

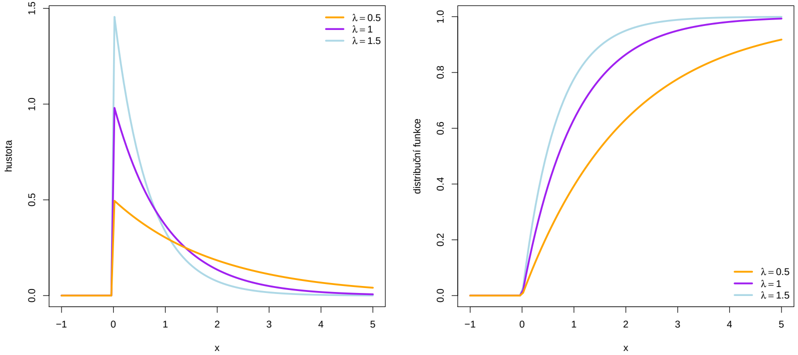
Náhodné vektory

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.

Exponenciální rozdělení

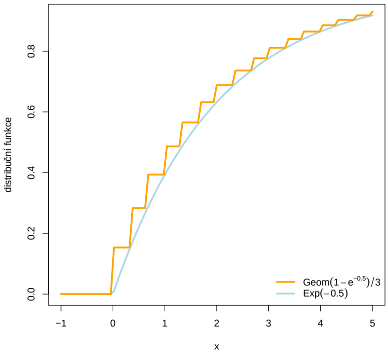
$$F_X(x) = \begin{cases} 0 & \text{pro } x \leq 0 \\ 1 - e^{-\lambda x} & \text{pro } x \geq 0 \end{cases}$$



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

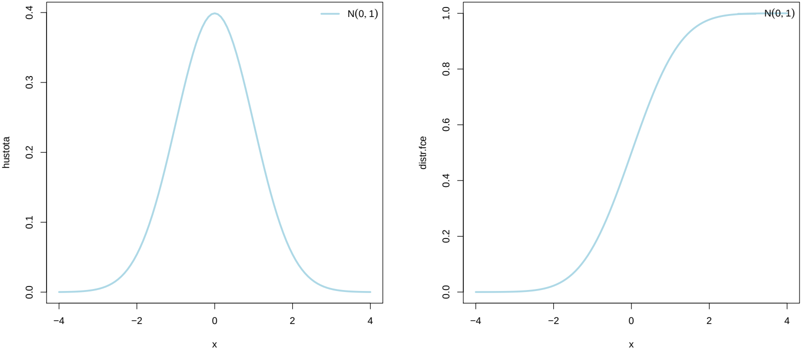
Souvislost $X \sim Exp(\lambda)$ a $Y \sim Geom(p)$

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



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 $\mathbb{E}(Z) = 0$, $var(Z) = 1$



Obecné normální rozdělení

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

Odolnost vůči součtu

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

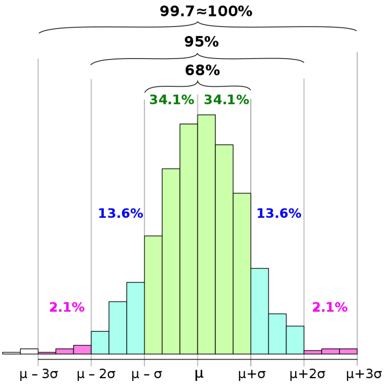
$$X_1 + \cdots + X_k \sim N(\mu, \sigma^2),$$

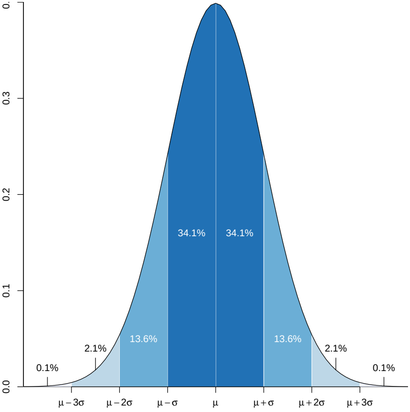
kde $\mu =$

$\sigma =$

Normální rozdělení – klíčové vlastnosti

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

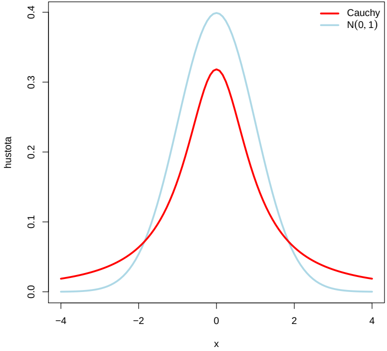




(Obrázek vlevo z Wikipedie, autor Melikamp.)

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 $X_1, \dots, X_n$ 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_k^2) 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.

Přehled

Sdružená distribuční funkce (Joint cdf)

Definice

Pro n.v. X, Y na pravděpodobnostním prostoru $(\Omega, \mathcal{F}, P)$ definujeme jejich sdruženou distribuční funkci (joint cdf) $F_{X,Y} : \mathbb{R}^2 \to [0,1]$ předpisem

$$F_{X,Y}(x,y) = P(\{\omega \in \Omega : X(\omega) \leq x \,\&\, Y(\omega) \leq y).$$

- Mohli bychom definovat i pro více než dvě n.v. $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$.
- Můžeme odsud odvodit pravděpodobnost obdélníku: $P(X \in (a,b] \,\&\, Y \in (c,d]) =$

Sdružená hustota (Joint pdf)

- Často můžeme sdruženou distribuční funkci psát jako integrál pomocí nezáporné funkce $f_{X,Y}$

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x,y)dxdy.$$

- Pak nazýváme n.v. X, Y sdruženě spojité. Funkce $f_{X,Y}$ je jejich *sdružená hustota.*
- Jako u jednorozměrného případu může být $f_{X,Y} > 1$.
- Stejně jako u jednorozměrného případu můžeme pak pomocí hustoty vyjádřit i další pravděpodobnosti:

$$P((X,Y) \in S) = \int_{S} f_{X,Y}(x,y)dxdy.$$

- $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$

LOTUS

- Stejně jako v diskrétním případu platí pro střední hodnotu funkce dvou n.v.

$$\mathbb{E}(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy.$$

- A stejně jako v diskrétním případu odsud odvodíme $\mathbb{E}(aX + bY + c) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y) + c$.

Vícerozměrné normální rozdělení