NMAI059 Pravděpodobnost a statistika 1
7. přednáška

Robert Šámal

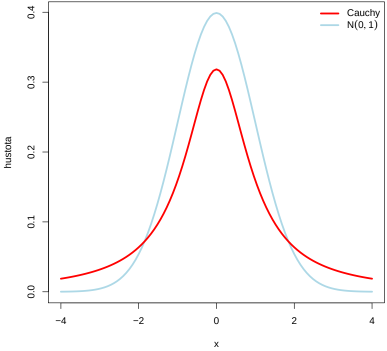
Přehled

Jaká rozdělení jsme už potkali

- $U(a, b)$ – uniformní – rovnoměrné na intervalu $[a, b]$
- $Exp(\lambda)$ – exponenciální – za jak dlouho se utrhne ucho u džbánu
- $N(\mu, \sigma^2)$ – normální – kolik váží chleba

Cauchyho rozdělení

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

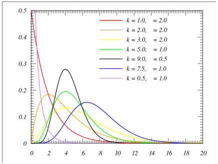


Gamma rozdělení

- $Gamma(w, \lambda)$, *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 $X_1 + \dots + X_n \sim Gamma(n, \lambda)$.
- 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.

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

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

- Formální podmínka: potřebujeme $\{X \leq x \,\&\, Y \leq y\} \in \mathcal{F}$, jinak (X,Y) není náhodný vektor.
- Mohli bychom definovat i pro více než dvě n.v. $F_{X_1,\dots,X_n}(x_1,\dots,x_n) =$
- 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 $f_{X,Y}$

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

- 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, pro „rozumnou množinu A“.

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

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

- $f_{X,Y}(x,y) \doteq \frac{P(x \leq X \leq x+\Delta_x \;\&\; y \leq Y \leq y+\Delta_y)}{\Delta_x \Delta_y}$

LOTUS

- Analogicky 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 tak jako v diskrétním případu odsud odvodíme

$$\mathbb{E}\big(aX + bY + c\big) = a \cdot \mathbb{E}\big(X\big) + b \cdot \mathbb{E}\big(Y\big) + c.$$

Nezávislost spojitých náhodných veličin

Definice

Libovolné náhodné veličiny nazveme nezávislé (independent), pokud jevy $\{X \leq x\}$ *a* $\{Y \leq y\}$ *jsou nezávislé pro libovolná* $x, y \in \mathbb{R}$*. Ekvivalentně,*

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y),$$
$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

Věta

Nechť X, Y *mají sdruženou hustotu* $f_{X,Y}$ *(a hustoty* f_X*,* f_Y*). Následující tvrzení jsou ekvivalentní:*

- X, Y *jsou nezávislé*
- $f_{X,Y}(x, y) = f_X(x)f_Y(y)$

Vícerozměrné normální rozdělení

- $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$
- $f(t_1, \dots, t_n) = \varphi(t_1)\varphi(t_2)\cdots\varphi(t_n) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{t_1^2+\cdots+t_n^2}{2}}$
- $f(t_1, \dots, t_n) = (2\pi)^{-n/2} e^{-r^2/2}$, kde $r^2 = t_1^2 + \cdots + t_n^2$
  radiálně symetrická funkce

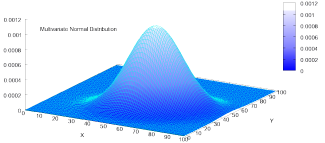


Image by Wikipedia editor Piotrg.

- Nechť $Z = (Z_1, \dots, Z_n)$ má hustotu f.
- $Z_1, \dots, Z_n$ jsou n.n.v., $Z_i \sim N(0,1)$
- $Z/\|Z\|$ je uniformně náhodný bod na n-rozměrné sféře.
- tudíž skal. součin Z s libovolným jednotkovým vektorem je $N(0,1)$
- $\langle u, Z\rangle = \sum_{i=1}^n u_i Z_i$ má také rozdělení $N(0,1)$

Vícerozměrné normální rozdělení obecné

- Obecněji můžeme vzít náhodný vektor s hustotou $c \cdot e^{Q(t)}$, kde $c > 0$ je vhodná konstanta a $Q(t)$ je obecná kvadratická funkce.
- Používá se ve strojovém učení.
- Souřadnice nejsou nezávislé!

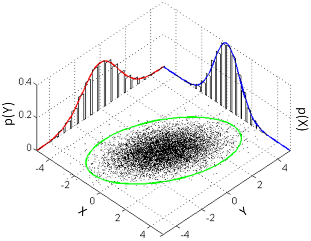


Image by Wikipedia editor Bscan.

Součet spojitých n.v.

Věta

Nechť spojité X, Y jsou n.n.v. Pak $Z = X + Y$ je také spojitá n.v. a její hustotu dostaneme jako konvoluci funkcí f_X, f_Y, neboli

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx.$$

Podmiňování

Definice

X je n.v. na $(\Omega, \mathcal{F}, P)$, $B \in \mathcal{F}$.

$$F_{X|B}(x) := P(X \leq x \mid B)$$

K tomu přísluší hustotní funkce $f_{X|B}$.

Věta

Nechť B_1, B_2, ... je rozklad Ω. Pak

$$F_X(x) = \sum_i F_{X|B_i} P(B_i) \quad \textit{a}$$

$$f_X(x) = \sum_i f_{X|B_i} P(B_i).$$

Důkaz: věta o úplné pravděpodobnosti. (Spec. případ byl na cvičení – dva algoritmy.)

Přehled

Kovariance

Definice

Pro n.v. X, Y definujeme jejich kovarianci (covariance) předpisem

$$cov(X, Y) = \mathbb{E}\big((X - \mathbb{E}X)(Y - \mathbb{E}Y)\big).$$

Věta

$$cov(X, Y) = \mathbb{E}\big(XY\big) - \mathbb{E}\big(X\big)\mathbb{E}\big(Y\big)$$

- $var(X) = cov(X, X)$
- $cov(X, aY + bZ + c) = a\, cov(X, Y) + b\, cov(X, Z)$
- $cov(X, Y) = 0$ pokud X,Y jsou nezávislé
- ale nejen tehdy

Korelace

Definice

Korelace náhodných veličin X, Y je definována předpisem

$$\varrho(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)\,var(Y)}}.$$

- je to přenormovaná kovariance
- $-1 \leq \varrho(X,Y) \leq 1$ (cvič.)
- Korelace neznamená příčinnou souvislost! (Např., korelace je symetrická, kauzalita nikoli!)
- Naopak, nekorelace neznamená nezávislost. (Př: X libovolná, $Y = +X$ nebo $Y = -X$, obojí se stejnou pravděpodobností.)

Rozptyl součtu

Věta

Nechť $X = \sum_{i=1}^{n} X_i$. *Pak*

$$var(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} cov(X_i, X_j) = \sum_{i=1}^{n} var(X_i) + \sum_{i \neq j} cov(X_i, X_j).$$

Spec. jsou-li $X_1, \ldots, X_n$ *nezávislé, pak*

$$var(X) = \sum_{i=1}^{n} var(X_i).$$

Přehled

Podmíněná hustota

Definice

Pro spojité n.v. X, Y definujeme podmíněnou hustotu (conditional pdf) předpisem

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

pokud je $f_Y(y) > 0$, jinak ji nedefinujeme.

- připomeňme, že $f_Y(y) = \int_{x \in \mathbb{R}} f_{X,Y}(x,y)dx$
- pro fixované y je $f_{X|Y}(x|y)$ hustota

Podmíněná, sdružená a marginální hustota

Věta

$$
\begin{aligned}
f_{X,Y}(x,y) &= f_Y(y) f_{X|Y}(x|y) \\
f_X(x) &= \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y)\,dy
\end{aligned}
$$

Podmíněná hustota a střední hodnota

- $F_{X|B}(x) := P(X \leq x \mid B)$ ... distr. funkce n.v. X zúžené na $B \subseteq \Omega$, potřebujeme $P(B) > 0$.
- $f_{X|B}$ odpovídající hustota
- pokud $B = \{X \in S\}$, tak

$$f_{X|B}(x) = \begin{cases} \frac{f_X(x)}{P(X \in S)} & \text{pokud } x \in S \\ 0 & \text{jinak} \end{cases}$$

- $\mathbb{E}(X \mid B) = \int_{-\infty}^{\infty} x \cdot f_{X|B}(x) dx$
- $\mathbb{E}(g(X)|B) = \int_{-\infty}^{\infty} g(x) f_{X|B}(x) dx$
- Pokud $B_1, B_2, \ldots$ je rozklad, tak

$$\mathbb{E}(X) = \sum_i \mathbb{E}(X \mid B_i) P(B_i).$$

Podmíněná hustota a střední hodnota

- $f_{X|Y}(x|y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}$ je hustota n.v. X, pokud $Y = y$
- $\mathbb{E}\big(X \mid Y = y\big) := \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x,y)dx$ je střední hodnota této veličiny
- $\mathbb{E}\big(g(X)\big|Y = y) = \int_{-\infty}^{\infty} g(x) \cdot f_{X|Y}(x,y)dx$
- $$\mathbb{E}\big(X\big) = \int_{-\infty}^{\infty} \mathbb{E}\big(X \mid Y = y\big) f_Y(y)dy$$
- $\mathbb{E}\big(X\big) = \mathbb{E}(\mathbb{E}\big(X \mid Y\big))$

Přehled

Cauchyho nerovnost

Věta

Nechť X, Y mají konečnou střední hodnotu a rozptyl. Pak

$$\mathbb{E}(XY) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$$

- Důsledek pro korelaci: $-1 \leq \varrho(X, Y) \leq 1$

Jensenova nerovnost

Věta

Nechť X má konečnou střední hodnotu a nechť g je konvexní reálná funkce. Pak

$$\mathbb{E}\big(g(X)\big) \geq g(\mathbb{E}\big(X\big)).$$

(Pro konkávní platí opačná nerovnost.)

Markovova nerovnost

Věta

Nechť náhodná veličina X *splňuje* $X \geq 0$*. Pak*

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Čebyševova (Chebyshev) nerovnost

Věta

Nechť X má konečnou střední hodnotu μ a rozptyl σ^2. Pak

$$P(|X-\mu| \geq a \cdot \sigma) \leq \frac{1}{a^2}.$$

Chernoffova (Černovova) nerovnost

Věta

Nechť $X = \sum_{i=1}^{n} X_i$*, kde* X_i *jsou n.n.v. nabývající hodnot* ± 1 *s pravděpodobností* $1/2$*. Pak pro* $t > 0$ *platí*

$$P(X \le -t) = P(X \ge t) \le e^{-t^2/2\sigma^2},$$

kde $\sigma = \sigma_X = \sqrt{n}$*.*

Bez dk.

Přehled

Silný zákon velkých čísel (strong law of large numbers)

Věta

Nechť $X_1, \dots, X_n$ *jsou n.n.v. se stř. hodnotou* μ *a rozptylem* σ^2. *Označme* $S_n = (X_1 + \cdots + X_n)/n$ *tzv. výběrový průměr (sample mean). Pak platí*

$$\lim_{n\to\infty} S_n = \mu \quad \textit{skoro jistě (tj. s pravděpodobností 1).}$$

Říkáme, že posloupnost S_n *konverguje k* μ *skoro jistě (almost surely).*

Monte Carlo integration

Jak spočítat $\int_{x \in A} g(x)dx$?

Slabý zákon velkých čísel (weak law of large numbers)

Věta

Nechť $X_1, \ldots, X_n$ *jsou n.n.v. se stř. hodnotou* μ *a rozptylem* σ^2*. Označme* $S_n = (X_1 + \cdots + X_n)/n$*. Pak pro každé* $\varepsilon > 0$ *platí*

$$\lim_{n\to\infty} P(|S_n - \mu| > \varepsilon) = 0.$$

Říkáme, že posloupnost S_n *konverguje k* μ *v pravděpodobnosti (in probability).*

Centrální Limitní věta

Centrální Limitní věta

Věta

Nechť $X_1, \ldots, X_n$ *jsou n.n.v. se střední hodnotou* μ *a rozptylem* σ^2*. Označme* $Y_n = ((X_1 + \cdots + X_n) - n\mu)/(\sqrt{n} \cdot \sigma)$*. Pak* $Y_n \xrightarrow{d} N(0,1)$*. Neboli, pokud* F_n *je distribuční funkce* Y_n*, tak*

$$\lim_{n\to\infty} F_n(x) = \Phi(x) \quad \text{for every } x \in \mathbb{R}.$$

Říkáme, že posloupnost Y_n *konverguje k* $N(0,1)$ *v distribuci (in distribution).*

Momentová vytvořující funkce

Definice

Pro náhodnou veličinu X označíme

$$M_X(t) = \mathbb{E}\big(e^{tX}\big).$$

Funkci $M_X(t)$ nazýváme momentová vytvořující funkce (moment generating function).

- $M_{Bern(p)}(t) = p \cdot e^t + (1-p)$.
- $M_X(t) = \sum_{n=0}^{\infty} \mathbb{E}\big(X^n\big)\frac{t^n}{n!}$.
- $M_{X+Y}(t) = M_X(t)M_Y(t)$, jsou-li X, Y n.n.v.
- $M_{Bin(n,p)} = (pe^t + 1 - p)^n$
- $M_{N(0,1)} = e^{t^2/2}$
- $M_{Exp(\lambda)} = \frac{1}{1-t/\lambda}$
- Pokud $M_X(t) = M_Y(t)$ na intervalu $(-a, a)$ pro nějaké $a > 0$, tak je $X = Y$ s.j.