NMAI059 Pravděpodobnost a statistika 1
7. přednáška

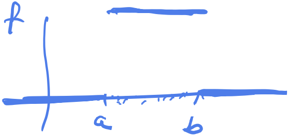
Robert Šámal

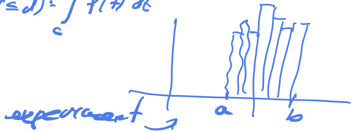
Přehled

Jaká rozdělení jsme už potkali

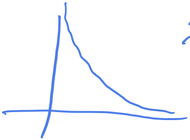
- $U(a, b)$ – uniformní – rovnoměrné na intervalu $[a, b]$

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





- $Exp(\lambda)$ – exponenciální – za jak dlouho se utrhne ucho u džbánu



$\lambda e^{-\lambda x}$

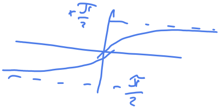
- $N(\mu, \sigma^2)$ – normální – kolik váží chleba

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

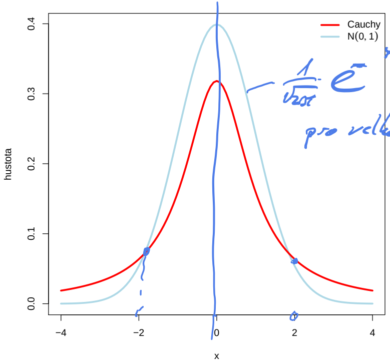
$\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$

Cauchyho rozdělení

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



$$\int_{-\infty}^{\infty} f = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{1+x^2} = \frac{1}{\pi}\left[\operatorname{arctg} x\right]_{-\infty}^{\infty} = 1$$



$\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \ll f(x)$ pro velké x

$$EX = \int_{-\infty}^{\infty} x f(x)$$

$$= \int_0^{\infty} \frac{2x}{2\pi(1+x^2)} + \int_{-\infty}^{0} \frac{x}{\pi(1+x^2)}$$

$$\left[\frac{1}{2\pi}\log(1+x^2)\right]_0^{\infty} + \left[\frac{1}{2\pi}\log(1+x^2)\right]_{-\infty}^{0}$$

$$\infty - 0 + 0 - \infty$$

$$\infty - \infty \; !$$

Gamma rozdělení

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

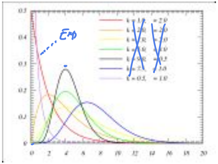
pro $w \in \mathbb{N}$

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

$\frac{1}{0!} \lambda^1 e^{-\lambda x}$

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.



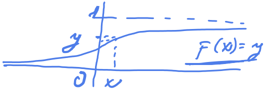
A mnoho dalších

- $Beta(s,t)$ – beta rozdělení
- χ^2 rozdělení s k stupni volnosti = chí-kvadrát (χ_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.



$Q(p)=x \Rightarrow p=F(x)$

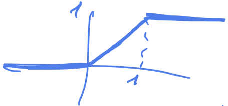
$Q=F^{-1}$

Věta

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

$\mathbb{R}\to\mathbb{R}$ n.v. = zobr. $\Omega\to\mathbb{R}$

F je rostoucí



Dk: $F_Y(y) = P(F(X) \le y) = \begin{cases} 0 & \text{pro } y \le 0 \\ 1 & \text{pro } y \ge 1 \\ (y\in(0,1)) & = P(X \le x) = F(x) = y \end{cases}$ $\to Y \sim U(0,1)$

spojité fce

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.

Dk $F_X(x) = P(Q(U) \le x) = P(U \le \underbrace{F(x)}_{\in[0,1]}) = F(x)$

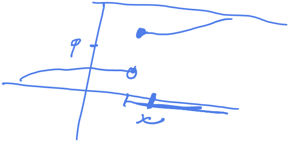
!

i pro diskr. n.v.!

$\forall F, Q: \; Q(p) \le x \Leftrightarrow F(x) \ge p$

$Q(p) = \min/\inf \{x : F(x) \geq p\}$

$Q(p) \leq x \iff F(x) \geq p$



$F(x) = 1 - e^{-\lambda x}$ $\quad Exp(\lambda)$

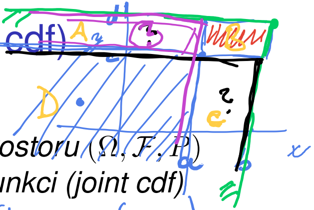
p

$$Q(p) = \frac{\lg(1-p)}{-\lambda} > 0$$

$$U \sim U(0,1) \qquad \boxed{\frac{\lg(1-U)}{-\lambda} \sim Exp(\lambda)}$$

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

$P(X \leq x \,\&\, Y \leq y)$

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

- 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. $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) =$ $P(X_1 \leq x_1 \,\&\, X_2 \leq x_2 \,\&\, \ldots \,\&\, X_n \leq x_n)$
- Můžeme odsud odvodit pravděpodobnost obdélníku: $P(X \in (a,b] \,\&\, Y \in (c,d]) =$ ~~$F(b,d) - F(a,c)$~~

$F(b,d) - F(b,c) - F(a,d) + F(a,c)$ ✓

$P(A \cup B \cup C \cup D)$ $\qquad$ $P(D)$

Sdružená hustota (Joint pdf)

f je hustota pravd. $\iint_A f$ je pravděpodobnost A

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

Fubiniho

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

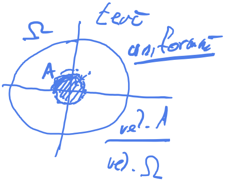
$F_X(x) = \int_{-\infty}^{x} f_X(s)\, ds$

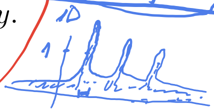
- Pak nazýváme n.v. X, Y *sdruženě spojité*. Funkce $f_{X,Y}$ je jejich *sdružená hustota.*
- Jako u jednorozměrného případu může být $f_{X,Y} > 1$.
- Stejně jako u jednorozměrného případu můžeme pak pomocí hustoty vyjádřit i další pravděpodobnosti, pro „rozumnou množinu A“.

$\iint_{\mathbb{R}^2} f_{X,Y} = P((X,Y) \in \mathbb{R}^2) = 1$

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

„pravd. tady blízko x, y“



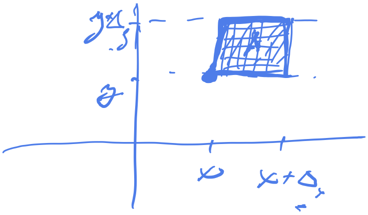


- $$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$$

$= F_{X,Y}(x,y)$ zderivuj podle x, pak podle y

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

$\frac{P(A)}{\text{velikost}(A)}$



$P((X,Y) \in A) = \int_A f$

$= \int_x^{x+\Delta_x} \int_y^{y+\Delta_y} f_{X,Y}(s,t)\, ds\, dt$

$\approx$

$f_{X,Y}(x,y) \cdot \Delta_x \cdot \Delta_y$

$\Delta_x, \Delta_y \doteq 0$, f spojitá

LOTUS

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

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

- A tak jako v diskrétním případu odsud odvodíme

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

$$g(x,y) = ax + by + c$$

$$E(g(X,Y)) = \iint g(x,y) f_{X,Y}(x,y) = \iint (ax+by+c) f(x,y) = \iint ax \cdot f(x,y) + \iint by\, f(x,y) + c \underbrace{\iint f(x,y)}_{1}$$

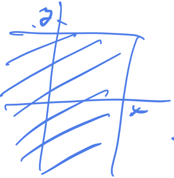
$$a \int_{-\infty}^{\infty} x \underbrace{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy}_{f_X(x)}\, dx = a \int x f_X(x)\, dx = a\, E(X)$$

Nezávislost spojitých náhodných veličin

$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t) = \iint f_X(s) f_Y(t) = \int_{-\infty}^{x} f_X(s) \cdot \int_{-\infty}^{y} f_Y(t) = F_X(x) F_Y(y)$

Definice

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



$$\boxed{\begin{aligned} P(X \le x, Y \le y) &= P(X \le x)P(Y \le y), \\ F_{X,Y}(x,y) &= F_X(x)F_Y(y) \end{aligned}}$$

Věta

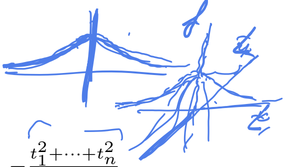
Nechť X, Y mají sdruženou hustotu $\cancel{f_X, f_Y}$ $f_{X,Y}$. *Následující tvrzení jsou ekvivalentní:*

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

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

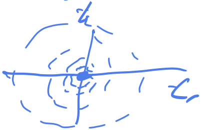
$\frac{\partial^2}{\partial x \partial y}\left(F_X(x) \cdot F_Y(y)\right) = F_X' \cdot F_Y' = f_X(x) f_Y(y)$

Vícerozměrné normální rozdělení



- $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ ... hustota $N(0,1)$
- $f(t_1, \ldots, t_n) = \varphi(t_1)\varphi(t_2)\cdots\varphi(t_n) = \frac{1}{\sqrt{2\pi}^n} e^{-\frac{t_1^2+\cdots+t_n^2}{2}}$
- $f(t_1, \ldots, t_n) = (2\pi)^{-n/2} e^{-r^2/2}$, kde $r^2 = t_1^2 + \cdots + t_n^2$
  radiálně symetrická funkce

  $r: \sqrt{t_1^2 + \cdots + t_n^2}$ = vzd. od $\vec{0}$ do $(t_1, \ldots, t_n)$



- Nechť $Z = (Z_1, \ldots, Z_n)$ má hustotu f.
- $Z_1, \ldots, Z_n$ jsou n.n.v., $Z_i \sim N(0, 1)$

- $Z/\|Z\|$ je uniformně náhodný bod na n-rozměrné sféře.
- tudíž skal. součin 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

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

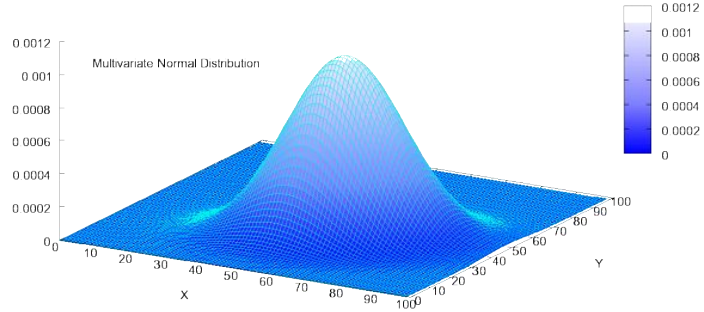


Image by Wikipedia editor Piotrg.

- 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~~ neg. definitní kvadratická funkce.
- Používá se ve strojovém učení.
- Souřadnice nejsou nezávislé!

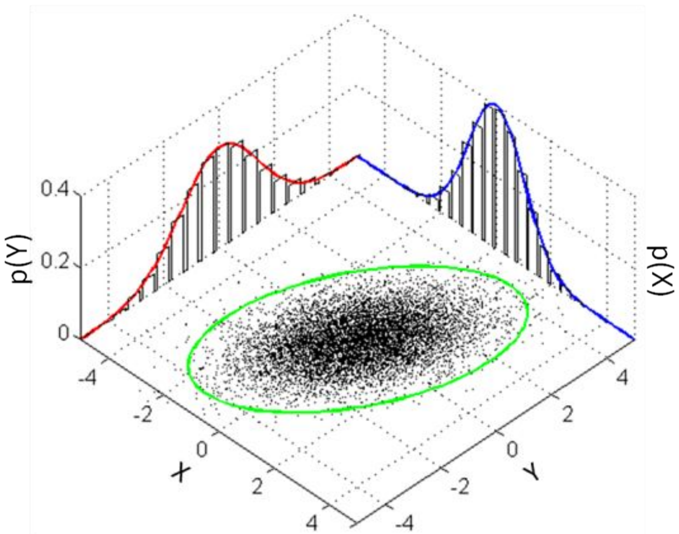


Image by Wikipedia editor Bscan.

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.)