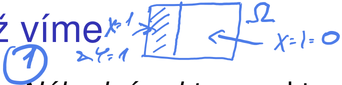
NMAI059 Pravděpodobnost a statistika 1
5. přednáška

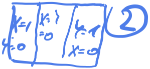
Robert Šámal

9

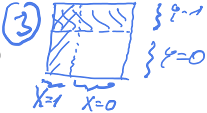
Přehled

Co už víme





$X, Y \sim Bern(p)$



- *Náhodný vektor* = vektor, kde každá souřadnice je náhodná veličina, **stále jen diskrétní.**
- *Sdružená pravděpodobnostní funkce (joint pmf)* $p_{X,Y} : \mathbb{R}^2 \to [0,1]$ je definována předpisem

$$p_{X,Y}(x,y) = P(X = x \,\&\, Y = y)$$

p_X, p_Y

- *Marginální rozdělení* = rozdělení jednotlivých souřadnic
- Sdružené rozdělení má víc informace, obecně nejde z jednotlivých složek rekonstruovat.
- Jde to pro *nezávislé n.v.* – tam platí (podle definice)

$\sum p_{X,Y}(\cdot,\cdot) = 1$

$$p_{X,Y}(x,y) = p_X(x)p_Y(y)$$

$\sum p_X(x) p_Y(y)$

a taky (cvičení)

$F_X(x) \cdot F_Y(y)$ $= \sum p_X(x) \cdot \sum p_Y = 1 \cdot 1 = 1$

$$P(X \le x \,\&\, Y \le y) = P(X \le x)P(Y \le y)$$

Marginální rozdělení ze sdruženého (opak.)

Věta

Nechť X, Y jsou diskrétní n.v. Pak

$$p_X(x) = \sum_{y\,\cancel{Y}\in Im(Y)} P(X = x \,\&\, Y = y) = \sum_{y\,\cancel{Y}\in Im(Y)} p_{X,Y}(x,y)$$

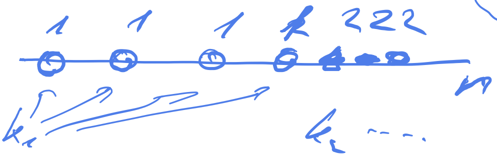
$$p_Y(y) = \sum_{x\,\cancel{X}\in Im(X)} P(X = x \,\&\, Y = y) = \sum_{x\,\cancel{X}\in Im(X)} p_{X,Y}(x,y)$$

Příklad: Multinomické rozdělení

$p_1 + \dots + p_6 = 1$, $n \in \mathbb{N}$

- Na kostce padne číslo i s pravděpodobností p_i pro $i = 1, \dots, 6$. Hodíme n-krát a označíme X_i počet hodů, kdy padlo i.

$$P(X_1 = k_1, \dots, X_6 = k_6) = \binom{n}{k_1, \dots, k_6} p_1^{k_1} \cdots p_6^{k_6}$$



rozdělení $\{1, \dots, n\}$ na k_1-prvk., k_2-prvk., ...

$n = k_1 + k_2 + \dots + k_6$

$$\frac{n!}{k_1! \dots k_6!}$$

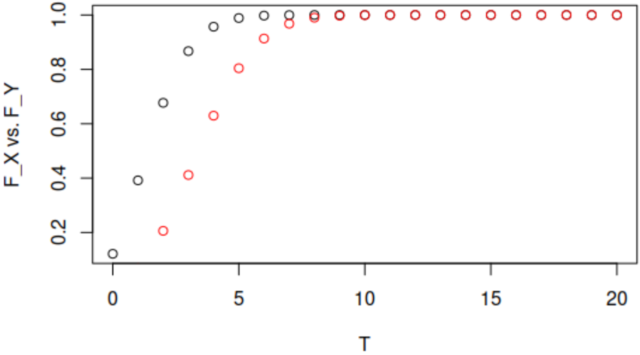
- $X_1 + X_2 + \dots + X_6 = n$
- $X_1 \sim Bin(n, p_1)$

Sdružování/Coupling – netriviální využití sdružených rozdělení

- $X \sim Bin(n,p)$ a $Y \sim Bin(n,q)$ a pro $p < q$
- Co můžeme říct o F_X a F_Y?
- $\underbrace{\sum_{i=0}^{k} \binom{n}{k} p^{k} (1-p)^{n-k}}_{}$ je rostoucí funkce p – ale proč?

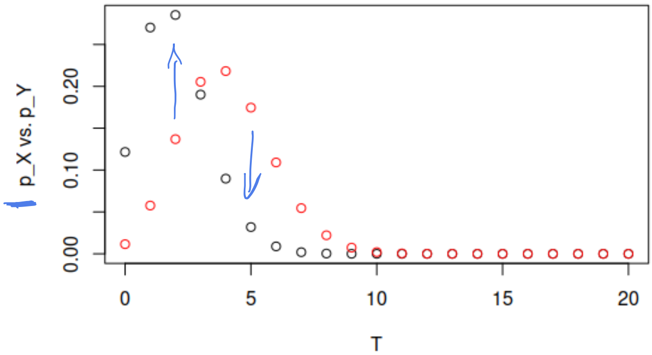
$F_X(k)$ $=$

$F_X \sim Bin(20, 0.1)$ $F_Y \sim Bin(20, 0.2)$



Sdružování/Coupling – netriviální využití sdružených rozdělení

- $X \sim Bin(n, p)$ a $Y \sim Bin(n, q)$ a pro $p < q$
- Co můžeme říct o F_X a F_Y?
- $\sum_{i=0}^{k} \binom{n}{k} p^k (1-p)^{n-k}$ je rostoucí funkce p – ale proč?



Coupling

$X \sim Bin(n,p)$

- $X = \sum_{i=1}^{n} X_i$, kde $X_1, \ldots, X_n$ jsou n.n.v. $\sim Bern(p)$
- $Y = \sum_{i=1}^{n} Y_i$, kde $Y_1, \ldots, Y_n$ jsou n.n.v $\sim Bern(q)$ $\quad p \le q$
- Vztah X a Y je není určen – můžou být jakékoliv.
- Zařídíme, že nebudou nezávislé, dokonce bude vždy $X \le Y$.
- Stačí definovat $Y_i = \begin{cases} \text{pokud } X_i = 1 \text{ tak } Y_i = 1 & p \\ \text{pokud } X_i = 0 \text{ tak } Y_i = \begin{cases} 1 & \frac{q-p}{1-p} \\ 0 & 1-\left(\quad\right) \end{cases} & 1-p \end{cases}$

Y_i nezávislé

$X_1, \ldots, X_i, \ldots, X_n$

$P(Y_i = 1) = q$

$\Rightarrow Y_1 \ldots Y_n$ jsou n.n.v. $\Rightarrow Y \sim Bin(n,q)$

$\forall i: X_i \le Y_i$

$\Rightarrow X \le Y$ vždy $\Rightarrow P(X \le k) \ge P(Y \le k)$

$Y \le k \Rightarrow X \le k$

Funkce náhodného vektoru

Věta (LOTUS)

Nechť X, Y jsou diskrétní *n.v. na $(\Omega, \mathcal{F}, P)$, nechť $g : \mathbb{R}^2 \to \mathbb{R}$ je funkce.*

- *Pak $Z = g(X, Y)$ je n.v. na $(\Omega, \mathcal{F}, P)$*
- *a platí pro ni*

$$\mathbb{E}(g(X,Y)) = \sum_{x \in ImX} \sum_{y \in ImY} g(x,y) P(X = x, Y = y).$$

$$\sum_{y \in Im(Y)} P(X=x, Y=y) = P(X=x)$$

$\boxed{\text{v2()}}$

Věta LINEARITA STŘ. HODNOTY

Pro X, Y n.v. a $a, b \in \mathbb{R}$ platí

NEMUSÍ BÝT NEZÁVISLÉ

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

$ax + by$

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

$$\mathbb{E}(aX+bY) = \mathbb{E}(g(X,Y)) = \sum_{x,y} g(x,y) P(X=x, Y=y) = \sum_{x,y} ax P(X=x, Y=y) + \sum_{x,y} by P(X=x, Y=y)$$

$$= \sum_x ax P(X=x) + \sum_y by P(Y=y)$$

$Z = g(X, Y)$ je diskr. n.v.

Aksiom
$P(\bigcup A_n) = \sum P(A_n)$
disj. jerg

$\forall z : Z^{-1}(z) = \{\omega : Z(\omega) = z\} \in \mathcal{F}$

$= \bigcup_{\substack{x \in \text{Im}(X) \\ y \in \text{Im}(Y) \\ g(x,y) = z}} \{\omega : X = x \,\&\, Y = y\} \in \mathcal{F}$ (*)

$Y(\omega)$
$X(\omega)$

spoč. sjednocenje

$EZ = \sum_z z \cdot P(Z = z) \stackrel{?}{=} \sum_z z \cdot \sum_{\substack{x,y \\ (*)}} P(X = x, Y = y)$

$\sum_z \sum_{(*)} g(x,y) P(X = x, Y = y) = \sum_{\substack{x \in \text{Im } X \\ y \in \text{Im } Y}} g(x,y) P(X = x, Y = y)$

Součin nezávislých n.v.

Věta

Pro nezávislé diskrétní n.v. X, Y platí

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Důkaz byl minule, využili jsme tehdy nejasný krok

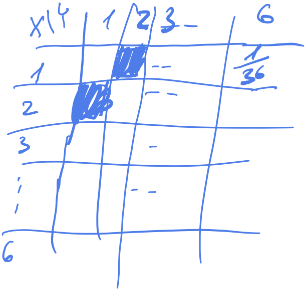
$$\mathbb{E}(XY) = \sum_{x \in Im(X), y \in Im(Y)} xy \cdot P(X = x \,\&\, Y = y)$$

$h(X,Y)$

$h(x,y)$

Součet nezávislých n.v.

- Máme-li dáno $p_{X,Y}$, jak zjistit rozdělení součtu, $Z = X + Y$? (obecněji: $Z = g(X,Y)$)



$P(Z=3) = P(X=1 \,\&\, Y=2)$

$+ P(X=2 \,\&\, Y=1)$

Součet nezávislých n.v. – konvoluce

Ω $Z = z$ $X = x$ $Y = z - x$

Věta

Pokud X, Y *jsou diskrétní náhodné veličiny, tak pro* $Z = X + Y$ *platí*

$$P(Z = z) = \sum_{x \in Im(X)} P(X = x, Y = z - x).$$

Pokud X, Y *jsou navíc nezávislé, tak* // pro nez. n.v.

$$P(Z = z) = \sum_{x \in ImX} P(X = x)P(Y = z - x).$$

Dk snadno

$$p_Z = \sum_{x} p_X(x)\, p_Y(z - x)$$

konvoluce

Ukázka konvoluce

$$X \sim Bin(m,p), \quad Y \sim Bin(n,p) \implies X+Y \sim Bin(m+n,p)$$

$$Z = X+Y$$

$$P(Z=z) = \sum_k P(X=k \,\&\, Y=z-k)$$

$$= \sum_{k=0}^{m} P(X=k)\cdot P(Y=z-k)$$

$$= \sum \binom{m}{k} p^k (1-p)^{m-k} \cdot \binom{n}{z-k} p^{z-k} (1-p)^{\overset{n-z+k}{n-(z-k)}}$$

$$= \sum_{k=0}^{m} \underbrace{p^z (1-p)^{m+n-z}}_{\text{nezávisí na } k} \binom{m}{k}\binom{n}{z-k} = p^z (1-p)^{m+n-z} \overbrace{\sum \binom{m}{k}\binom{n}{z-k}}^{\binom{m+n}{z}}$$

Přehled

Náhodné vektory

Podmíněné rozdělení

Spojité náhodné veličiny

Spojité náhodné veličiny

Podmíněné rozdělení

x	1	2	3	4	5	6
p_X	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

x	1	2	3	4	5	6
$p_{X\mid A}$	0	$\frac{1}{3}$	0	$\frac{1}{3}$	0	$\frac{1}{3}$

X, Y – diskrétní náhodné veličiny na $(\Omega, \mathcal{F}, P)$, $A \in \mathcal{F}$

- $p_{X|A}(x) := P(X = x \mid A)$ $\qquad E(X|A) := \sum x \cdot P(X=x|A)$
  příklad: X je výsledek hodu kostkou, $A =$ padlo sudé číslo
- $p_{X|Y}(x|y) = P(X = x \mid Y = y)$ příklad: X, Z jsou výsledky dvou nezávislých hodů kostkou, $Y = X + Z$.

  $PS = \{X=6\}$
  $SD = \{X+Z=10\}$

$$p_{X|Y}(6|10) = P(PS|SD)$$

- $p_{X|Y}$ z $p_{X,Y}$:

$$p_{X|Y}(x|y) = \frac{P(X=x \,\&\, Y=y)}{P(Y=y)} = \boxed{\frac{p_{X,Y}(x,y)}{p_Y(y)}} = \frac{p_{X,Y}(x,y)}{\sum_{x'} p_{X,Y}(x',y)}$$

jak dostat marg. r. ze sdruž.:

Sdružené vs. podmíněné rozdělení

$Y = X + Z$ součet dvou nez. kostek
$(Y \le X+6)$

Y, X

$p_{X,Y}$	...	10	11	12
1		0	0	0
2		0	0	0
3		0	0	0
4		$\frac{1}{36}$	0	0
5		$\frac{1}{36}$	$\boxed{\frac{1}{36}}$	0
6		$\frac{1}{36}$	$\frac{1}{36}$	$\boxed{\frac{1}{36}}$
p_Y		$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

přeškálujte sloupce aby měly součet 1

$y = 11$

$p_{X\|Y}$	...	10	11	12
1		0	0	0
2		0	0	0
3		0	0	0
4		$1/3$	0	0
5		$1/3$	$1/2$	0
6		$1/3$	$1/2$	1
		1	1	1

Přehled

Obecná náhodná veličina

Definice

Náhodná veličina (random variable) na $(\Omega, \mathcal{F}, P)$ *je zobrazení* $X : \Omega \to \mathbb{R}$*, které pro každé* $x \in \mathbb{R}$ *splňuje*

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}.$$

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

- diskrétní n.v. je n.v.

$$\{\omega : X \leq x\} = \bigcup_{\substack{x' \in \mathrm{Im}(X) \\ x' \leq x}} \{\omega : X = x'\}$$

pro diskr. n.v.

spoč. sjedn.

Distribuční funkce

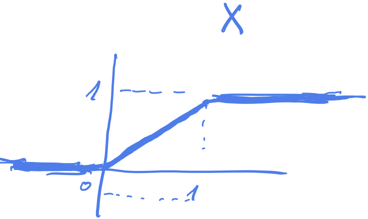
Definice

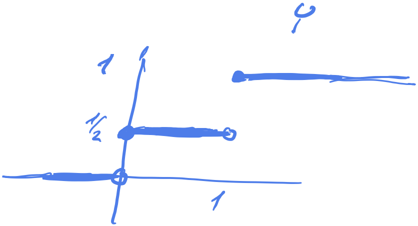
Distribuční funkce (cumulative distribution function, CDF) n.v. X je funkce

$$F_X(x) := P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}).$$

- F_X je neklesající funkce
- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to +\infty} F_X(x) = 1$
- F_X je zprava spojitá

Distribuční funkce – další ukázky





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

$P(X \in [0,1]) = F(1) - F(0) = 1$

$Y = \begin{cases} 1 & \text{s pst } 1/2 \\ 0 & 1/2 \end{cases}$

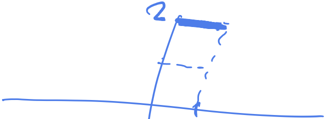
Kvantilová funkce

Pro nádhnou 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 $Q_X = F_X^{-1}$.
- $Q_X(1/2) =$ medián (pozor, když F_X není rostoucí)
- $Q_X(10/100) =$ desátý percentil, atd.

Spojitá náhodná veličina

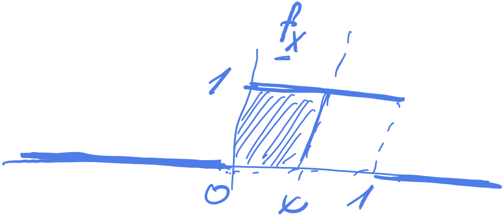


Definice

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

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

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



podmínka na hustotu:

$$\int_{-\infty}^{\infty} f_X(t) = 1$$

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

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* $a, b \in \mathbb{R}$.

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 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.

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

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.

Spojitý LOTUS

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.)