NMAI059 Pravděpodobnost a statistika 1
5. přednáška

Robert Šámal

Přehled

Co už víme

- *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

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

- *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)

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

a taky (cvičení)

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

Marginální rozdělení ze sdruženého

Věta

Nechť X, Y jsou diskrétní n.v. Pak

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

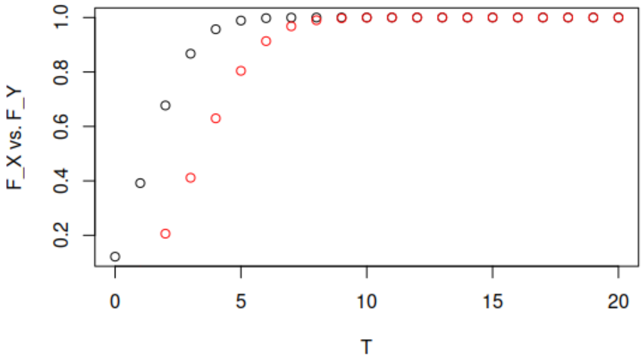
$$p_Y(y) = \sum_{X \in Im(X)} P(X = x \,\&\, Y = y) = \sum_{X \in Im(X)} p_{X,Y}(x, y)$$

Příklad: Multinomické rozdělení

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

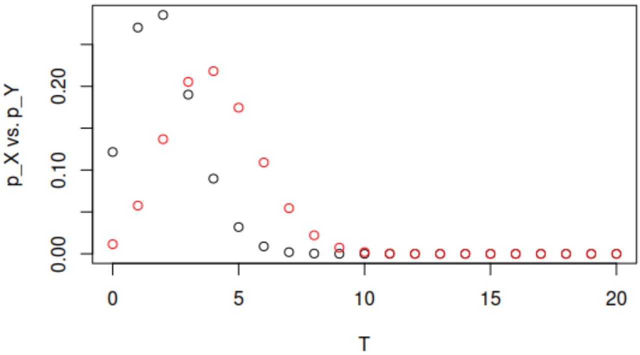
Sdružování/Coupling – netriviální využití sdružených rozdělení

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



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

- $X = \sum_{i=1}^{n} X_i$, kde $X_1, \dots, X_n$ jsou n.n.v
- $Y = \sum_{i=1}^{n} Y_i$, kde $Y_1, \dots, Y_n$ jsou n.n.v
- Vztah X a Y je není určen – můžou být jakékoliv.
- Zařídíme, že nebudou nezávislé, dokonce bude vždy $X \leq Y$.
- Stačí definovat $Y_i =$

Funkce náhodného vektoru

Věta

Nechť X, Y jsou 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*

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

Věta

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

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

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

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

Součet nezávislých n.v.

- Máme-li dáno $p_{X,Y}$, jak zjistit rozdělení součtu, $Z = X + Y$?

Součet nezávislých n.v. – konvoluce

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

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

Ukázka konvoluce

Přehled

Podmíněné rozdělení

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

- $p_{X|A}(x) := P(X = x \mid A)$
  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, $Y = X + Z$.

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

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

Sdružené vs. podmíněné rozdělení

$p_{X,Y}$	...	10	11	12
1				
2				
3				
4				
5				
6				

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

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*

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

- diskrétní n.v. je n.v.

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

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.

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.)

Přehled

Shrnutí, co už víme

- $F_X(x) := P(X \leq x)$ distribuční funkce, existuje vždy, je neklesající, zprava spojitá, $F_X(-\infty) = 0$, $F_X(+\infty) = 1$
- $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ pro tzv. spojité n.v. Pro ty dále platí:
- $P(a \leq X \leq b) = \int_a^b f_X(t)dt$, spec. tedy $P(X = x) = 0$ pro každé $x \in \mathbb{R}$
- obecněji: $P(X \in A) = \int_A f_X(t)dt$, kdykoli umíme přes množinu A integrovat
- spec. tedy $\int_{-\infty}^{\infty} f_X(t)dt = 1$
- $\mathbb{E}(X) = \int_{-\infty}^{\infty} t\, f_X(t)dt$
- $\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(t) f_X(t)dt$
- Pokud je hustota spojitá, tak navíc platí: $f_X = F'_X$ (základní věta kalkulu).

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.

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

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