NMAI059 Pravděpodobnost a statistika 1

4 ~~3~~. přednáška

Robert Šámal

Přehled

Shrnutí k střední hodnotě a rozptylu

Podmíněné rozdělení

Spojité náhodné veličiny

Střední hodnota

diskrétní n.v.

- $\mathbb{E}(X) = \sum_{x \in Im(X)} x \cdot P(X = x)$
- $\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\})$

X, Y mohou být závislé

- linearita: $\mathbb{E}(aX + bY + c) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y) + c$
- LOTUS: $\mathbb{E}(g(X)) = \sum_{x \in Im(X)} g(x) \cdot P(X = x)$
- $\mathbb{E}(X \mid A)$, rozbor možností
- „Kolik čekáme, že průměrně dostaneme, když budeme opakovat nezávislé pokusy s výsledkem popsaným X“ ... bude tzv. zákon velkých čísel

Rozptyl

- $var(X) = \mathbb{E}((X - \mathbb{E}X)^2)$
- $var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$
- *Směrodatná odchylka (standard deviation)* $\sigma_X = \sqrt{var(X)}$ – „stejné jednotky jako X“.
- Měří, jak je daleko „typicky“ je X od $\mathbb{E}(X)$. Mohli bychom to měřit i jinak (např. $\mathbb{E}(|X - \mathbb{E}(X)|)$, ale rozptyl je výhodnější).

$$P(|X - EX| > 3\sigma_X) < 0.01$$

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

$p_{X\|A}$:	1	2	3	4	5	6	ostatní
	0	1/3	0	1/3	0	1/3	0

- $p_{X|Y}(x|y) = P(X = x \mid Y = y)$ příklad: X, Z jsou výsledky dvou nezávislých hodů kostkou, $Y = X + Z$.

X, Y

$x, y \in \mathbb{R}$

$$p_{X|Y}(6|10) = \frac{P(X=6 \,\&\, Y=10)}{P(Y=10)} = \frac{1/36}{3/36} = \frac{1}{3}$$

PS: $X=6$
SD: $Y=10$
$P(PS \mid SD)$

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

podm.

sdružená

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

Sdružené vs. podmíněné rozdělení

$Y = X + Z$ součet dvou hodů

$p_{X,Y}$	...	10	11	12
1		0	0	0
2		0	0	0
3		0	0	0
4		1/36	0	0
5		1/36	1/36	0
6		1/36	1/36	1/36

$p_{X\|Y}$	...	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
$\sum$		1	1	1

$\sum_{y \text{ pevné } y} P_{X|Y}(x|y) \neq 1$

fix $x = 6$

$\sum_x P_{X|Y}(x|y) = \sum_{x \text{ v/m } X} \frac{P(X=x|Y=y)}{}$

$P_{X|Y}(x|11)$

$P_{X|Y}(x|12)$

Přehled

Shrnutí k střední hodnotě a rozptylu

Podmíněné rozdělení

Spojité náhodné veličiny

Obecná náhodná veličina

$\Omega \xrightarrow{X} \mathbb{R}$

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*

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

$\Omega = [0,1]$
$X(\omega) = \omega$
$\{X = x\} = \{x\}$ jednob.
$\{X \le x\} = [0,x]$

- diskrétní n.v. je n.v.

$\forall x \; \{X = x\} \in \mathcal{F} \longrightarrow \forall x \; \{X \le x\} \in \mathcal{F}$

$$\{\omega : X(\omega) \le x\} = \bigcup_{\substack{y \in \operatorname{Im} X \\ y \le x}} \underbrace{\{\omega : X(\omega) = y\}}_{\in \mathcal{F}}$$

spoč. sjedn.

Distribuční funkce

Definice

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

$\le$

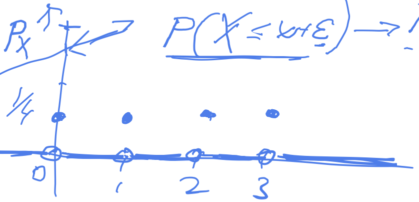
$$F_X(x) := P(X \le x) = P(\{\omega \in \Omega : X(\omega) \le x\}).$$

$\varepsilon \to 0^+$

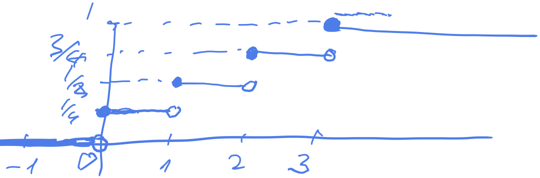
$P(X \le x+\varepsilon) \to P(X \le x)$

$\boxed{\text{neklesající}}$

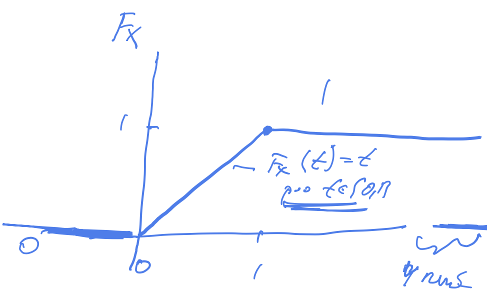
- F_X je ~~nerostoucí~~ funkce
- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to +\infty} F_X(x) = 1$
- F_X je zprava spojitá

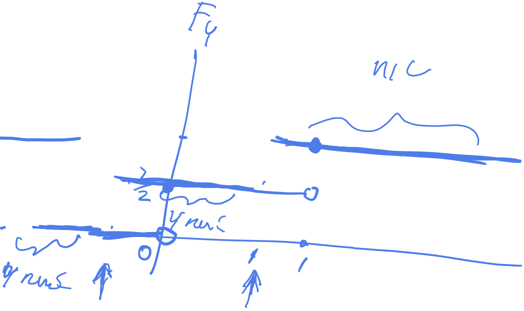


$X \sim Unif(\{0,1,2,3\})$



Distribuční funkce – další ukázky





X unif. na $[0,1]$

$P(X \leq t) = t$

$P(X \in [0,t])$

$P(\cancel{X}Y \leq 0) = \frac{1}{2}$

$P(Y = 0) = \frac{1}{2}$

$P(Y \leq 1) = 1$

$P(Y = 1) = \frac{1}{2}$

$\mathrm{Im}\, Y = \{0,1\}$

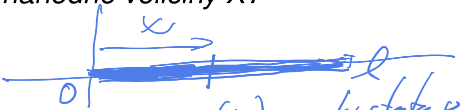
Spojitá náhodná veličina

Definice

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

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

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



$\rho(x)$ ... hustota v bodě x

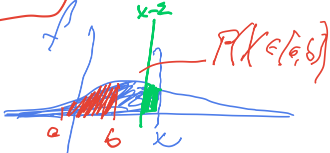
hmotnost $m = \int_0^l \rho(x) = 1$

hmotnost úseku od 0 do x je $\int_0^x \rho(t)\,dt$

těžiště má souř.: $\dfrac{\int_0^l t\,\rho(t)}{m}$ } EX

Práce s hustotou

Rozhodněte, zda f_X musí být v $[0,1]$



Věta

Nechť spojitá n.v. X má hustotu f_X. Pak

1. $P(X = x) = 0$ *pro každé* $x \in \mathbb{R}$.
2. $P(a \le X \le b) = \boxed{\int_a^b f_X(t)\,dt}$ *pro každé* $a, b \in \mathbb{R}$.

$\boxed{a \le b}$ $\qquad P(X \le x)$

Dk $(2 \Rightarrow 1)$ $(a = b = x)$

(1) $P(X=x) \overset{?}{=} \boxed{\lim_{\varepsilon\to 0+} P(x-\varepsilon < X \le x)}$

$= \lim \; P(X \le x) - P(X \le x-\varepsilon)$

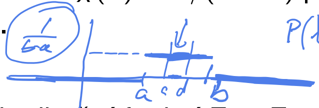
$= \lim \int_{-\infty}^{x} f_X - \int_{-\infty}^{x-\varepsilon} f_X = \lim_{\varepsilon\to 0+} \int_{x-\varepsilon}^{x} f_X = 0$

$\{\omega : X(\omega) = x\} = \bigcap_{\varepsilon>0} \{\omega : x-\varepsilon < X(\omega) \le x\}$

(2) $P(a \le X \le b) = P(X \le b) - P(X \le a) + \underbrace{P(X=a)}_{0} = F_X(b) - F_X(a) = \int_{-\infty}^{b} f - \int_{-\infty}^{a} f =$

Uniformní rozdělení

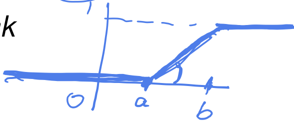
- N.v. X má uniformní rozdělení na intervalu $[a, b]$, píšeme $X \sim U(a, b)$, pokud $f_X(x) = 1/(b-a)$ pro $x \in [a, b]$ a $f_X(x) = 0$ jinak.



$P(X \in [c,d]) = \int_c^d f_X = \frac{d-c}{b-a} \leq 1$

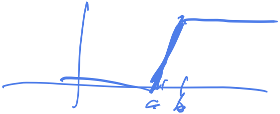
Věta

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



Věta

Nechť $U \sim U(0, 1)$ a F je rostoucí spojitá funkce s $\lim_{x\to-\infty} F(x) = 0$ a $\lim_{x\to+\infty} F(x) = 1$. Pak $F^{-1}(U)$ je n.v. s distribuční funkcí F.



Universalita unif.

Střední hodnota spojité n.v.

průměr

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

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

$$\sum x_i \cdot P(X = x_i)$$

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

- Analogie s výpočtem těžiště tyče ze znalosti hustoty.

$$\mathbb{E}X = \sum_n \int_n^{n+1} x f_X(x)$$

$$\doteq \sum_n n \int_n^{n+1} f_X = \sum_n n \, P(n \leq X \leq n+1)$$

Spojitý LOTUS

Věta (LOTUS)

Pokud X je spojitá n.v. s hustotou f_X a g reálná funkce, tak

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

pokud integrál má smysl.

(Důkaz vynecháme.)

$$\mathbb{E}(g(X)) = \sum g(x)\, p_X(x)$$

pro diskr.

$$\int_t^{t+\varepsilon} f_X = P(t \le X < t+\varepsilon)$$