Pravděpodobnost a Statistika 1

Poznámky z přednášek

Letní semestr 2020/2021

Viktor Soukup, Lukáš Salak

Verze 0.1
29. března 2021

Obsah

1 První přednáška

Modely náhody → Pravděpodobnost → Pozorovaná data → Modely náhody
Model náhody např. kostka 1,...,6,
Pozorovaná data : 1,5,4,3,3
otázka na pravděpodobnost: jaká je pravděpodobnost... hodně pozorovaných dat → statistika na model náhody.

Příklad (Schwartz-Zippel algoritmus)**:** Máme dány dva polynomy $f(x), g(x)$ stupně d. Chceme zjistit, zda jsou stejné, a to co nejrychleji.
Problém: $g(x)$ je součin několik polynomů stupně $\leq \frac{d}{4}$, dostávame víc než lineární čas.

Řešení: Algoritmus: zvolíme náhodně $x \in \{1, 2 \ldots, 100d\}$, ověříme, zda $f(x_1) = g(x_1)$. Když $f \neq g$, tak x_1 je kořen polynomu $f - g$. ... takových x_1 je $\leq d$.

$$P(f(x_1) = g(x_1) : f \neq g) \leq \frac{1}{100}$$

Pokud jsme spokojeni s 1%, končíme, když ne, volíme $x_2, x_3 ... \in \{1, 2, \ldots, 100d\}$, pak

$$P(Prox_1, x_2, x_3 \ldots f(x_i) = g(x_i) : f \neq g) \leq \left(\frac{1}{3}\right)^3 = 10^{-6}$$

... aproximační algoritmy

Některé jevy neumíme/nechceme popsat kauzálně

1. hod kostkou
2. tři hody kostkou, nekonečně mnoho hodů kostkou
3. hod šipkou na terč
4. počet emailů za den
5. dobu běhu programu (v reálnem počítači)...

Důvody:

- fyzikální vlastnost přírody
- komplikovaný proces (počasí, medicína, molekuly plynu...)
- neznáme vlivy (působení dalších lidí, programů...)
- randomizované algoritmy (test prvočíselnosti, quicksort)
- náhodné grafy (Ramseyovy čísla)

Pro popis pomocí teorie pravděpodobnosti napřed vybereme množinu elementárních jevů Ω (sample space)

$$\Omega = \{1, 2 \ldots, 6\} = [6] \implies \text{hod kostkou}$$
$$\Omega = [6]^3 \implies \text{hod třemi kostkami}$$

Definice (Prostor jevů)**:** $\mathcal{F} \subseteq \mathcal{P}(\Omega)$
$\mathbb{F} \subseteq \mathbb{P}(\Omega)$ je prostor jevů (též σ-algebra), pokud

1. $\emptyset \in \mathcal{F}$ a $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \implies \Omega \backslash A \in \mathcal{F}$
3. $A_1, A_2, ... \in \mathcal{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Často $\mathcal{F} = \mathcal{P}(\Omega)$, to je možné vždy, když je Ω spočetná, např. pro $\Omega = \mathbb{R}$ to již nejde.

Definice (Pravděpodobnost)**:** $P : \mathcal{F} \to [0, 1]$ se nazývá pravděpodobnost (probability), pokud:

1. $P(\emptyset) = 0, P(\Omega) = 1$, a
2. $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, pro libovolnou posloupnost po dvou *disjunktních* jevů **TODOOT**

Šance (odds) jevu A je $O(A) = \frac{P(A)}{P(A^c)}$. Např. šance na výhru je 1 ku 2 znamená, že pravděpodobnost výhry je $\frac{1}{3}$; šance, že na kostce padne šestka je 1 ku 5.

Konvence:

- „A je jistý jev" znamená $P(A) = 1$. Také se říká, že A nastáva skoro jistě (almost surely), zkráceně s.j. (a.s.).
- „A je nemožný jev" znamená P(A) = 0.

$$P(A) = 0 \Rightarrow^{?} A = \emptyset$$

$$\leftarrow \text{axiom}$$

$$\rightarrow \text{platí často, ne vždy}$$

- Např. A = střed kruhu (házení šipek na terč) $\implies$ $P(A) = 0$ B spočetná (konečná, velká jako $\mathbb{N}$) množina:

$$P(B) = 0 + 0 + 0 + \cdots = 0$$

  B_i je i-tý bod, $B = \bigcup B_i$

Věta (Vlastnosti pravděpodobnostního prostoru)**:** *V pravděpodobnostním prostoru* $(\Omega, \mathcal{F}, P)$ *platí pro* $A,B \in \mathcal{F}$*:*

1. $P(A) + P(A^c) = 1$
2. $A \subseteq B \implies P(A) \leq P(B)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. $P(A_1 \cup A_2 \cup \dots) \leq \sum_i P(A_i)$ *(subaditiva, Booleova nerovnost) (nevyžadujeme disjunktnost, pak by platila rovnost)*

Důkaz:

1. $\Omega = A \cup A^c$; A, A^c disj.,
   $1 = P(\Omega) = P(A) + P(A^c)$
2. $P(B) = P(A) + P(B \backslash A) \geq P(A)$
3. cvičení **TODOOT**

4. trik zdisjunktnění: z $A_1, A_2...$ uděláme disjunktní množiny

$$B_1 = A_1, B_2 = A_2 \backslash A_1, B_3 = A_3 \backslash A_1 \cup A_2 \dots$$

$$B_i \subseteq A_i \implies P(B_i) \leq P(A_i)$$

$$B_i \cap B_j = \emptyset : j < i \dots B_i \cap B_j \subseteq B_i \cap A_j = \emptyset$$

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$$

$$\subseteq ok$$

opačná inkluze **TODOOT**

$$P(\bigcup A_i = P(\bigcup B_i) = \sum P(B_i) \leq \sum P(A_i)).$$

□

Příklad (Pravděpodobnostní prostory)**:**

1. Konečný s uniformní pravděpodobností
   Ω je libovolná konečná množina, $\mathcal{F} = \mathcal{P}(\Omega)$,
   $P(A) = \frac{|A|}{|\Omega|}$.
2. Diskrétní
   $\Omega = \{\omega_1, \omega_2 \dots\}$ je libovolná spočetná množina. Jsou dány $p_1, p_2 \dots \in [0, 1]$ se součtem 1.
   $P(A) = \sum_{i:\omega_i \in A} p_i$ (cinknutá loterie, nějaké možnosti mají jiné procenta)
3. Spojitý
   $\Omega \subseteq \mathbb{R}^d$ pro vhodné d (Ω např. uzavřená nebo otevřená)
   $\mathcal{F}$ vhodná (obsahuje např. všechny otevřené množiny)
   $f : \Omega \to [0, 1]$ je funkcne taková, že $\int_\Omega f(x)dx = 1$.
   $P(A) = \int_A f(x)dx$
   Speciální případ: $f(x) = 1/V_d(\Omega)$
   $P(A) = \frac{V_d(A)}{V_d(\Omega)}$, kde $V_d(A) = \int_A 1$ je d-rozměrný objem A.
4. Bernoulliho krychle - nekonečné opakování
   $\Omega = S^{\mathbb{N}}$, kde S je diskrétní s pravděpodobností Q,
   $\mathcal{F}$ vhodná (obsahuje např. všechny množiny tvaru
   $A = A_1 \times \dots \times A_k \times S \times S \times \dots$)
   $P(A) = Q(A_1) \dots Q(A_k)$

Příklad (Nepříklady)**:**

1. Náhodné přirozené číslo: můžeme si vybrat mnoha způsoby, Ale všechna přirozená čísla nemají stejnou pravděpodobnost.
   není možné, aby měly všechny stejnou nenulovou pravděpodobnost, protože pokud $P(0) = P(1) = P(2) \dots = P$ tak $P(\mathbb{N}) = p + p + p \dots = \infty$.
2. Náhodné reálne číslo
3. Betranův paradox

Definice (Podmíněná pravděpodobnost)**:** Pokud $A, B \in \mathcal{F}$ a $P(B) > 0$, pak definujeme podmíněnou pravděpodobnost A při B (probability of A given B) jako

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

$Q(A) := P(A|B)$. Pak $(\Omega, \mathcal{F}, Q)$ je pravděpodobností prostor.

Definice (Zřetězené podmíňování)**:** $P(A \cap B) = P(B)P(A|B)$

Věta: *Pokud* $A_1, \dots A_n \in \mathcal{F}$ *a* $P(A_1, \cap \dots \cap A_n) > 0$*, tak*

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)...TODOOT$$

2 Druhá přednáška

2.1 Opakování

1. definice pravděpodobnostního prostoru $(\Omega, \mathcal{F}, P)$: dva axiomy,
2. **naivní** pravděpodobnostní prostor: Ω konečná, $\mathcal{F} = \mathcal{P}(\Omega)$
   $P(A) := |A|/|\Omega|$
3. **diskrétní** pravděpodobnostní prostor: $\Omega = \omega_1, \omega_2, ...,$
   $\mathcal{F} = \mathcal{P}(\Omega), \sum_i p_i = 1$
   $P(A) := \sum_{i:\omega_i \in A} p_i$
4. **geometrický** pravděpodobnostní prostor:
   $\omega \subseteq \mathbb{R}^d$ s konečným objemem,
   $P(A) := V_d(A)/V_d(\Omega)$
5. pravděpodobnostní prostor **spojitý s hustotou**:
   $\Omega \subseteq \mathbb{R}^d$ s funkcí f, kde $\int_\Omega f = 1$,
   $P(A) := \int_A f$

V pravděpodobnostním prostoru $(\Omega, \mathcal{F}, P)$ platí pro $A, B \in \mathcal{F}$

1. $P(A^c) = 1 - P(A)$ ... $(A^c = \Omega \backslash A)$
2. $A \subseteq B \implies P(A) \leq P(B)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ ... PIE
4. $P(A_1 \cup A_2 \cup \dots) \leq \sum_i P(A_i)$ (subaditivita, Booleova nerovnost)
5. Definujeme podmíněnou pravděpodobnost (pro $P(B) > 0$).
   $$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$
6. $Q(A) = P(A|B)$ splňuje axiomy pro pravděpodobnost.
   $$P(\emptyset|B) = 0$$
   $$P(\Omega|B) = \frac{P(B)}{P(B)} = 1$$
   $$P(A_1 \circ A_2|B) = \frac{P((A_1 \circ A_2) \cup B)}{P(B)} = \frac{P((A_1 \cap B) \cup (A_2 \cap B))}{P(B)} = \frac{P(A_1 \cap B)}{P(B) + \frac{P(A_2 \cap B)}{P(B)}} = P(A_1|B) + P(A_2|B)$$

Definice (Zřetězené podmíňování)**:**

$$P(A \cup B) = P(B)P(A|B)$$

Věta: *Pokud* $A_1, \dots, A_n \in \mathcal{F}$ *a* $P(A_1 \cap \dots \cap A_n) > 0$*, tak*

$$P(A_1 \cap A_2 \cap \dots \cap A_n) =$$

$$P(A_1)P(A_2|A_1)P(A_3)|A_1 \cap A_2) \dots P(A_n|\bigcap_{i=1}^{n-1} A_i)$$

Důkaz: indukcí □

Příklad: Příklad: vytáhneme 3 karty z balíčku 52 karet. Jaká je P(žádne srdce)?
A_i = i-tá karta není srdce

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \times P(A_2|A_1)P(A_3|A_2 \cap A_1) = \frac{13 * 3}{52} \times \frac{13 * 3 - 1}{51} \times \frac{13 * 3 - 2}{50}$$

$$\frac{\#\text{dobrých}}{\#\text{všech}} = \frac{\binom{39}{3}}{\binom{52}{3}}$$

Definice: Spočetný systém množin $B_1, B_2, .. \in \mathcal{F}$ je rozklad (partition) Ω, Pokud

1. $B_i \cap B_j = \emptyset$ pro $i \neq j$ a
2. $\bigcup_i B_i = \Omega$.

Věta: *Věta o úplné pravd. = Rozbor všech možností*
Pokud $B_1, B_2, \dots$ *je rozklad* Ω *a* $A \in \mathcal{F}$*, tak*

$$P(A) = \sum_i P(B_i)P(A|B_i)$$

(sčítance s $P(B_i) = 0$ *považujeme za* 0*).*

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \dots$$

(sjednocení disjunktních množin)

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(B_i)P(A|B_i)$$

Příklad: Máme tři mince: P+O, P+P, O+O. Jaká je pravděpodobnost, že padne orel?
Označíme M_1, M_2, M_3 pro P+O, P+P, O+O.

$$P(O) = P(M_1)P(O|M_1) + P(M_2)P(O|M_2) + P(M_3)P(O|M_3)$$
$$= \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times 0 + \frac{1}{3} \times 1 = \frac{1}{2}$$

Rychlejší je vypsat si strom a pak poscítat výsledné jevy

Příklad (Gambler's ruin - zbankrotování hazardního hráče.)**:** Máme a korun, náš protihráč b korun. Hrajeme opakovaně spravedlivou hru o 1kč,
dokud někdo nepřijde o všechny peníze. Jaká je pravděpodobnost, že vyhrajeme?

Důkaz:

$$P_a = P(\text{z této pozice vyhrajeme})$$
$$P_0 = 0, P_n = 1 \dots (a + b = n)$$
$$P(\text{výhra}|1.\text{ kolo výhra})P(1.\text{ kolo výhra})$$
$$+P(\text{výhra}|1.\text{ kolo prohra})P(1.\text{ kolo prohra})$$
$$\text{výhra} \implies P_{a+1}, \text{prohra} \implies P_{a-1}$$
$$P_a = \frac{P_{a+1}}{2} + \frac{P_{a-1}}{2}$$
$$\text{ekvivalencia} <=> \text{pls}$$
$$P_a - P_{a-1} = P_{a+1} - P_a = \Delta$$
$$1 = P_n = P_0 + n * \Delta \implies \Delta = \frac{1}{n}$$
$$P_a = \frac{a}{a+b} = \frac{a}{n}$$

□

Věta (Bayesova Věta)**:** *Pokud $B_1, B_2, \dots$ je rozklad $\Omega, A \in \mathcal{F}, P(A) > 0$ a $P(B_j) > 0$, tak*

$$P(B_j)|A) = \frac{P(B_j)P(A|B_j)}{P(A)} = \frac{P(B_j)P(A|B_j)}{\sum_i P(B_i)P(A|B_i)}.$$

(sčítance s $P(B_i) = 0$ považujeme za 0).

Důkaz:

$$P(B_j|A)P(A) = P(B_j)P(A|B_j)$$
$$P(A \cap B_j) = P(B_j \cap A)$$

□

Příklad: N = nemocný, T = testovaný, specif. $P(N|T)$, sens. $P(T|N)$.

$$P(N|T) = \frac{P(N)P(T|N)}{P(N)P(T|N) + P(N^c)P(T|N^c)} = \frac{p * 0.8}{p * 0.8 + (1-p) * 0.01}$$
$$p = 0.001 \dots 7\%$$
$$p = 0.0016 \dots 56\% \dots \text{momentální stav testování}$$
$$p = 0.05 \dots 80\%$$

Definice: Jevy $A, B \in \mathcal{F}$ jsou nezávislé (independenet) pokud $P(A \cap B) = P(A)P(B)$.
Pak také platí $P(A|B) = P(A)$, pokud $P(B) > 0$.

Definice: Jevy $\{A_i : i \in I\}$ jsou (vzájemně) nezávislé, pokud pro každou konečnou množinu $J \subseteq I$

$$P(\bigcap_{i \in J} A_i) = \prod_{i \in J} P(A_i).$$

Pokud podmínka platí jen pro dvouprvkové množiny J, nazýváme jevy $\{A_i\}$ po dvou nezávislé (pairwise independenet).

Definice: Nechť pro množiny z prostoru jevů platí

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$$

a $A = \bigcup_{i=1}^{\infty} A_i$. Pak platí

$$P(A) = lim_{i\to\infty} P(A_i).$$

Důkaz:

$$A = A_1 \cup (A_2 \backslash A_1) \cup (A_3 \backslash A_2) \cup \dots$$

$$P(A) = P(A_1) + P(A_2 \backslash A_1) + P(A_3 \backslash A_2) + \dots$$

$$lim_{i\to\infty}(P(A_1) + \dots + P(A_i \backslash A_{i-1})) = lim_{i\to\infty P(A_i)}.$$

□

$A_n \subset P, O^{\mathbb{N}}, A_n =$ mezi prvními n hody padl aspoň jednou orel.

$$P(A) = P(\geq 1 orelv \infty hodech) = lim_{n\to\infty} \dots = 1$$

Definice (Náhodná veličina)**:** Mějme pravděpodobnostní prostor $(\Omega, \mathcal{F}, P)$. Funkci $X : \Omega \to \mathbb{R}$ nazveme diskrétní náhodná veličina, pokud $I_m(X)$ je spočetná množina a pokud pro všechna reálna x platí

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

Definice: Pravděpodobnostní funkce diskrétní náhodné veličiny X je funkce $p_X : \mathbb{R} \to [0,1]$ taková, že

$$p_X(x) = P(X = x) = P(\{x \in \Omega : X(\omega) = x\})$$

Definice: $\sum_{x \in I_m(X)} p_X(x) = 1$

Definice: $S := I_m(X)1, Q(A) := \sum_{x \in A} p_X(x)$
$(S, \mathcal{P}(S), Q)$ je diskrétní pravděpodobnostní prostor.

Definice: Pro $S = \{s_i : i \in I\}$ spočetnou množinu reálných čísel a $c_i \in [0,1]$ $\sum_{i \in I} c_i = 1$ existuje pravděpodobnostní prostor a diskrétní n.v. X na něm taková, že $p_X(s_i) = c_i$ pro $i \in I$.

3 Třetí přednáška

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

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

1. F_X je neklesajíci funkce
2. $lim_{x\to-\infty} F_X(x) = 0$
3. $lim_{x\to+\infty} F_X(x) = 1$
4. F_X je zprava spojitá

Příklad: $X = \{0 \text{ s pravděpodobností } \frac{1}{2}, 1 \text{ s psravděpodobností } \frac{1}{2}\}$

Důkaz: F_X je neklesajíci funkce

$x < y \implies P(X \leq x) \leq P(X \leq y)$
protože $A = \omega : X(\omega) \leq x$,
$B = \omega : X(\omega) \leq y$, pak

$A \subseteq B \implies P(A) \leq P(B)$ □

Důkaz: $\lim_{x \to +\infty} F_X(x) = 1$

$A_n = X \leq n$; platí $A_1 \subseteq A_2 \subseteq \dots$
Takže $\bigcup_{n=1}^{\infty} A_n = \Omega$, podle věty o spojitosti pak

$$P(\Omega) = \lim_{n \to \infty} P(A_n) = \lim_{n \to \infty} F_X(n)$$

□

druhá limita obdobně

Definice (Bernoulliho/alternativní rozdělení)**:** white x

1. $X =$ počet orlů při jednom hodu nespravedlivou mincí.
2. Značíme $X \sim Bern(p)$. Někdy se značí $Alt(p)$.

1. Dáno $p \in [0, 1]$.
2. $p_X(1) = p$
3. $p_X(0) = 1 - p$
4. $p_X(k) = 0$ pro $k \neq 0, 1$

1. Pro libovolný jev $A \in \mathcal{F}$ definujeme *indikátorovou* n.v. I_A:
2. $I_A(\omega) = 1$ pokud $\omega \in A, I_A(\omega) = 0$ jinak.
3. $I_A \sim Bern(P(A))$.

Definice: white x

1. $X =$ počet orlů při n hodech nespravedlivou mincí.
2. Dáno $p \in [0, 1]-$ pravděpodobnost orla při jednom hodu.
3. Značíme $X \sim Bin(n, p)$.

1. $X = \sum_{i=1}^{n} X_i$ pro nezávislé n.v. $X_1, \dots X_n \sim Bern(p)$.
2. $p_X(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ pro $k \in 0, 1, \dots n$.

$$\sum_{k=1}^{n} \binom{n}{k} p^k (1 - p)^{n-k} = 1$$

$$(p + (1 - p))^n = 1^n = 1$$

Definice (Hypergeometrické rozdělení)**:** white x

1. $X =$ počet vytažených červených míčku při n tazích, v osudí je K červených z N celkových míčků
2. Dáno n, N, K.
3. Značíme $X \sim Hyper(N, K, n)$.
4. $p_X(k) = P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$

Definice (Poissonovo rozdělení (poasón))**:** white x

1. Značíme $X \sim Pois(\lambda)$.
2. Dáno reálné $\lambda > 0$.
3. $p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}$
4. $Pois(\lambda)$ je limitou $Bin(n, \lambda/n) \ldots \sim X_n \ldots \lambda$ pevné
5. X popisuje např. počet emailů, které dostaneme za jednu hodinu.

chceme $\sum \frac{\lambda^k}{k!} e^{-1} = 1$

$$e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$$

$$P(X_n = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} =$$

$$= \frac{n(n-1)\ldots(n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{1}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} =$$

$$= \frac{\lambda^k}{k!} e^{-\lambda}$$

Poznámka (Poissonovo paradigma)**:** $A_1, \ldots A_n$ jsou (skoro-)nezávislé jevy s $P(A_i) = p_i$, $\lambda = \sum_j p_j$. Nechť n je velké, každé z p_i malé. Pak přibližně platí

$$\sum_{i=1}^{n} I_{A_i} \sim Pois(\lambda)$$

Definice (Geometrické rozdělení)**:** white x

1. $X =$ kolikátým hodem mincí padl první orel.
2. Značíme $X \sim Geom(p)$.
3. Dáno $p \in [0, 1]$.
4. $p_X(k) = (1-p)^{k-1} p$, pro $k = 1, 2, \ldots$
5. Někdy se tomuto rozdělení říká posunuté geometrické, a za normální geometrické se považuje rozdělení $X - 1$, t.j. počet neúspěšných hodů.

Důkaz: chceme $\sum (1-p)^{k-1} p = 1$

$$= \frac{(1-p)^0 p}{1 - (1-p)} = \frac{p}{p} = 1$$

□

Definice (Střední hodnota)**:** Pokud X je diskrétní n.v., tak její střední hodnota (expectation) je označovaná $\mathbb{E}(X)$ a definovaná

$$\mathbb{E}(X) = \sum_{x \in Im(X)} x \times P(X = x),$$

pokud součet má smysl.

Nechť X je definovaná na diskrétním prostoru $(\Omega, \mathcal{F}, P)$. Pak střední hodnotu lze také definovat

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) P(\omega).$$

... vážený průměr **Důkaz:** dk., že obě definice souhlasí.

$$\sum_{x \in Im(X(\omega))} \sum_{\omega \in \Omega} X(\omega) P(\omega) = \sum_{x \in Im(X)} (x \times P(\omega \in \Omega : X(\omega) = x))$$

□

Definice (Rozptyl)**:** Rozptyl(variace) n.v. X nazveme číslo $\mathbb{E}((X - \mathbb{E}X)^2)$.
Značíme jej $var(X)$

Věta:

$$var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

Definice (LOTUS (Law of The Unconscious Statistist))**:** Pro reálnou funkci g a diskrtétni n.v. X je $Y = g(X)$ také diskrétní n.v.

Věta (LOTUS)**:** *Pokud X je diskrétní n.v. a g reálná funkce, tak*

$$\mathbb{E}(g(X)) = \sum_{x \in Im(X)} g(x) P(X = x)$$

pokud součet má smysl.

Důkaz:
$Y = g(X)$

$$\begin{aligned} \mathbb{E}Y &= \sum_{y \in Y} y \times P(Y = y) \text{... definice} \\ &= \sum_{y \in Y} \sum_{x \in Im(X)} g(x) P(X = x) \\ &= \sum_{x \in Im(X)} g(x) P(X = x) \end{aligned}$$

□

4 Čtvrtá přednáška

Věta: *Věta*

Nechť X, Y jsou diskrétní n.v. a $a, b \in \mathbb{R}$.

1. *Pokud $P(X \geq 0) = 1$ a $\mathbb{E}(X) = 0$, tak $P(X = 0) = 1$.*
2. *Pokud $\mathbb{E}(X) \geq 0$ tak $P(X \geq 0) > 0$.*
3. $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.
4. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

Důkaz: *1.*

$$\mathbb{E}(X) = \sum_{x \in X} xP(X = x) = 0P(X = 0) + \sum_{x > \& x \in X} xP(X = 0) = \sum_{x < 0 \& x \in X} xP(X = x) = 0$$

$$\implies \forall x > 0 : P(x = x) = 0 \implies P(X = 0) = 1$$

2.

$$\mathbb{E}(X) = \sum xP(X \geq x) = 0$$

$$\textit{kdyby ne: } P(x \geq 0) = 0, \textit{ všechny členy v sumě by byly záporné...spor}$$

3.

$$\mathbb{E}(aX + b) = \sum_{x \in X} (ax + b)P(X = x) = a \sum xP(X = x) + b \sum P(X = x)$$

4.

$$\mathbb{E}(X + Y) = \sum_{\omega} (X(\omega) + Y(\omega))P(\omega) = \sum X(\omega)P(\omega) + \sum Y(\omega)P(\omega) = \mathbb{E}(X) + \mathbb{E}(Y)$$

□

Věta: *Nechť X je diskrétní n.v. nabývajíci jen hodnot z $\mathbb{N} = 0, 1, 2, s$. Pak platí*

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} P(X > n).$$

Důkaz:

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} kP(X = k) = \sum_{k=0}^{\infty} \sum_{n=0}^{k-1} P(X = k)$$

$$= \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} P(X = k) = \sum \sum P(\omega \in \Omega : X(\omega) = k)$$

$$= \sum P(\omega \in \Omega : X(\omega) > n) = \sum_{n=0}^{\infty} P(X > n)$$

□

Definice: Rozptyl

Rozptyl (*variance*) n.v. X nazveme číslo $\mathbb{E}((X - \mathbb{E}(X))^2)$.
Značíme jej $var(X)$. … (kvadratické měření odchylky)

1. Směrodatná odchylka (standard deviation) $\sigma_X = \sqrt{var(X)}$

   Poznámka: "stejné jednotky jako X"

2. Měří, jak je daleko "typicky"X od $\mathbb{E}(X)$. Mohli bychom to měrit i jinak (např. $\mathbb{E}(|X - \mathbb{E}(X)|)$, ale rozptyl je výhodnější).

Věta: *Věta*

$$var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

Důkaz:

$$\mu = \mathbb{E}(X)$$
$$var(X) = \mathbb{E}((X-\mu)^2) = \mathbb{E}(X^2 - 2\mu + \mu^2) = \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2$$
$$=$$

□

Definice: Podmíněná střední hodnota

Pokud X je diskrétní n.v. a $P(B) > 0$, tak podmíněná střední hodnota X za předpokladu B (conditional expectation of X given by B)

Věta: *Věta o úplné střed. hodnotě*

Pokud $B_1, B_2, \ldots$ je rozklad Ω a X je d.n.v., tak

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

kdykoliv má součet smysl. (Sčítance s $P(B_i) = 0$ považujeme za 0.) ***Důkaz:***

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

$$= \sum_i P(B_i) \sum_x xP(X = x|B_i)$$
$$= \sum_x x(\sum_i P(B_i)P())$$

□

Rozbor všech možností $X \sim Geom(p)$
$B_1 = S \ldots$ *první pokus úspěšný*
$B_2 = B_1^C = F \ldots$ *první pokus neúspěšný*

$$\mathbb{E}(X) = P(S)\mathbb{E}(X|S) + P(F)\mathbb{E}(X|F)$$
$$= p1 + (1-p)(\mathbb{E}(X+1))$$
$$p\mathbb{E}(X) = p + (1-p) = 1$$
$$\mathbb{E}(X) = \frac{1}{p}$$

Věta: *Parametry rozdělení - Bernoulliho*

Pro $X \sim Bern(p)$ je

1. $\mathbb{E}(X) = p$
2. $var(X) = p(1-p)$

Důkaz: $\mathbb{E}(X) = 0P(X=0) + 1P(X=1) = P(X=1) = p$
$var(X) = \mathbb{E}(X-p)^2 = (0-p)^2P(X=0) + (1-p)^2P(X=1) = p(1-p)(p+(1-p)) = p(1-p)$ □

Věta: *Parametry rozdělení - binomické*

Pro $X \sim Bin(n, p)$ *je*

1. $\mathbb{E}(X) = np$
2. $var(X) = np(1-p)$

1. *První postup:* $X = \sum_{i=1}^{n} X_i$, *kde* $X_i =$ [*i-tý hod uspěl*]

$$\mathbb{E}(X_i) = P(X_i = 1) = p$$

2. *Druhý postup:*

$$\mathbb{E}(X) = \sum_{k=0}^{n} kP(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

$$\sum_{k=1}^{n} pn \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)}$$

$$= pn(p + (1-p))^{n-1} = np$$

Věta: *Parametry rozdělení - hypergeometrické*

Pro $X \sim Hyper(N, K, n)$ *je*

1. $\mathbb{E}(X) = n\frac{K}{N}$
2. $var(X) = n\frac{K}{N}(1 - \frac{K}{N})\frac{N-n}{N-1}$

1. *První postup:* $X = \sum_{i=1}^{n} X_i$, *kde* $X_i =$ [*i-tý míček červený*]

$$\mathbb{E}(X_i) = P(X_i = 1) = \frac{K}{N}$$

2. *Druhý postup:*

$$\mathbb{E}(X) = \sum_{j=1}^{K} Y_j, \text{ kde } Y_j = [\text{byl vytažen (z n tahů) míček s číslem j}]$$

$$\mathbb{E}(Y_j) = P(Y_j = 1) = \frac{n}{N}$$

$$= \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

SKONTROLOVAŤ, STRATIL SOM SA :Sadge:

Věta: *Parametry rozdělení - geometrické*

Pro $X \sim Geom(p)$ *je*

1. $\mathbb{E}(X) = n\frac{1}{p}$
2. $var(X) = \frac{1-p}{p^2}$

Věta: *Parametry rozdělení - hypergeometrické*

Pro $X \sim Hyper(N, K, n)$ je

1. $\mathbb{E}(X) = \lambda$
2. $var(X) = \lambda$

$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$$

$$\mathbb{E}(X) = \sum k\frac{\lambda^k}{k!}e^{-\lambda} = 1$$

$$\lambda \sum \frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda} = \lambda$$

Definice: Základní popis náhodných vektorů

1. X, Y - **nestihol som, lebo preskočil slide za 2 sekundy :) boli tam asi 4 odrážky**

Definice: Nejaká definícia, nesithol som, lebo som rantoval

Definice: Marginální rozdělení

1. Máme-li dáno $p_{X,Y}$, jak zjistit rozdělení jednotlivých složek, t.j. p_X a p_Y ?

Důkaz: Nechť X, Y jsou diskrétní n.v. Pak:

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

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

□

Definice: Nezávislost náhodných veličin

Diskrétní n.v. X, Y jsou nezávislé (independent) pokud pro každé $x, y \in \mathbb{R}$ jsou jevy $\{X = x\}$ a $\{Y = y\}$ nezávislé. To nastane právě když

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

Věta: *Součin nezávislých n.v.*

Pro nezávislé diskrétní n.v. X, Y platí

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

Důkaz:

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

$$= \sum_x xP(X = x) \sum_y yP(Y = y) = \mathbb{E}(X)\mathbb{E}(Y)$$

□

5 Pátá přednáška

Definice: Coupling

1. $X = \sum_{i=1}^{n} X_i$ kde $X_1, \ldots, X_n$ jsou n.n.v. $\cdots \sim Bern(p)$
2. $Y = \sum_{i=1}^{n} Y_i$ kde $Y_1, \ldots, Y_n$ jsou n.n.v. $\cdots \sim Bern(q) \ldots o < q$
3. vztah X, Y není určen, můžou být jakékoliv.
4. Zařídíme, že nebudou nezávislé, dokonce bude vždy $X \leq Y$.
5. Stačí definovat:

$$\text{pokud} X_i = 1 \text{tak} Y_i = 1$$
$$\text{pokud} X_i = 0 \text{tak} Y_i \text{ buď 1 nebo 0}$$
$$\implies Y_1, \ldots, Y_n \text{jsou n.n.v} \implies Y \sim Bin(n, q)$$
$$\implies X \leq Y \text{vždy} (Y \leq k \implies X \leq k) \implies P(X \leq k) \geq P(Y \leq k)$$

Věta: *Funkce náhodného vektoru*

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

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

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

Věta (linearita střední hodnoty)

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

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

The End