

10. PST - Způsoby popisu rozdělení náhodných veličin a vektorů. Odhady parametrů rozdělení. Základní statistické testy. Markovské řetězce a jejich asymptotické vlastnosti.

http://cmp.felk.cvut.cz/~navara/stat/PMS_ebook.pdf

<http://cmp.felk.cvut.cz/~navara/stat/>

Korbelar: <https://math.feld.cvut.cz/korbelar/PSI/PSI.htm>

https://cs.wikipedia.org/wiki/Rozd%C4%9Blen%C3%AD_pravd%C4%9Bpodobnosti
i - dobrý shrnutí popisu náhodných veličin a vektorů

Původní verze: [+10. Způsoby popisu rozdělení náhodných veličin a vektorů. Odhady parametrů rozdělení. Základní statistické testy. Markovské řet](#)

1 Základní pojmy pravděpodobnosti

1.1 Laplaceova (klasická) pravděpodobnost

- **Náhodný pokus** má $n \in \mathbb{N}$ různých, vzájemně se vylučujících výsledků, které jsou stejně možné.
- **Elementární jevy** = výsledky nahodného pokusu
- **Množina všech elementárních jevů:** Ω
- **Jev** je podmnožina všech elementárních jevů ($A \subseteq \Omega$)
- **Pravděpodobnost jevu** $A : P(A) = \frac{|A|}{|\Omega|}$
- **Jevové pole:** všechny jevy pozorovatelné v náhodném pokusu, zde $\exp \Omega$ (=množina všech podmnožin množiny Ω)

1.2 Kolmogorovova pravděpodobnost

- Elementárních jevů (=prvků množiny Ω) může být nekonečně mnoho, **nemusí být stejně pravděpodobné**

- **Jevy** jsou podmnožiny množiny Ω , ale ne nutně všechny. Tvoří podmnožinu $A \subseteq \exp \Omega$, která splňuje podmínky σ -algebry (viz. 1.3).
- **Pravděpodobnost** není určena strukturou jevů jako u Laplaceova modelu, je to funkce $P: A \rightarrow \{0,1\}$, splňující podmínky:
 - (P1) $P(\Omega) = 1$, (pst jistého jevu je 1)
 - (P2) $P(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} P(A_n)$, pokud jsou množiny (=jevy) A_n , $n \in \mathbb{N}$, **po dvou neslučitelné**
- **Pravděpodobnostní prostor** je trojice (Ω, A, P) , kde Ω je neprázdná množina, A je σ -algebra podmnožin množiny Ω a P je pravděpodobnost.

1.3 σ -algebra

σ -algebra je teoretický koncept výběru jistých podmnožin dané množiny, který splňuje pevně definované podmínky. Koncept σ -algebry umožňuje například zavést míru, čehož se dále využívá zejména v matematické analýze k budování pojmu integrál a právě v teorii pravděpodobnosti [wikipedia]. Systém podmnožin A nějaké množiny Ω musí splňovat podmínky:

1. $\emptyset \in A$
2. $A \in A \Rightarrow \bar{A} \in A$ (uzavřenost vůči doplňku) // TODO — divne, mrkni na to
3. $\forall n \in \mathbb{N} : A_n \in A \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in A$ (uzavřenost vůči sjednocení)

Nejmenší σ -algebra podmnožin \mathbb{R} , která obsahuje všechny intervaly, se nazývá **Borelova σ -algebra**. Obsahuje všechny intervaly otevřené, uzavřené i polouzavřené, i jejich spočetná sjednocení, a některé další množiny, ale je menší než $\exp \mathbb{R}$. Její prvky nazýváme borelovské množiny.

1.4 Nezávislost

Řekneme, že jevy A a B jsou nezávislé iff: $P(A \cap B) = P(A)P(B)$

Pak pro nezávislé jevy platí: $P(A \cup B) = P(A) + P(B) - P(A)P(B)$ (představte si množinový obrázek)

Jevy A_1, \dots, A_n se nazývají **po dvou nezávislé**, jestliže jsou každé dva z nich nezávislé.

1.5 Podmíněná pravděpodobnost

$P(A|B) = \frac{P(A \cap B)}{P(B)}$, pravděpodobnost jevu A za podmínky B .

$P(A|B) = P(A)$ iff jevy A a B jsou **nezávislé**.

Pro B a jev k němu doplňkový platí: \bar{A}

$$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$$

1.6 Úplný systém jevů:

Jestliže máme jevy B_i pro $i = 1, 2, \dots, n$, které jsou po dvou neslučitelné, tzn. $B_i \cap B_j = \emptyset$ pro $i \neq j$, a jev A , který lze vyjádřit jako $A = B_1 \cup B_2 \cup \dots \cup B_n$, pak říkáme, že jev A se rozpadá na částečné jevy B_1, B_2, \dots, B_n . Pokud navíc platí, že $B_1 \cup B_2 \cup \dots \cup B_n = \Omega$, kde Ω je jev jistý, pak jevy B_1, B_2, \dots, B_n tvoří **úplný systém jevů**. Při daném náhodném pokusu nastane alespoň jeden jev B_i z úplného systému jevů.

Věta o úplné psti: Necht' $\{B_i \mid P(B_i) \neq 0, \forall i = 1, \dots, n\}$ je úplný spočetný systém jevů. Pak pro každý jev A platí:

$$P(A) = \sum_{i=1}^n P(A \mid B_i)P(B_i)$$

1.7 Bayesova věta

Pokud máme úplný systém jevů, jako výše, pak platí:

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^n P(A \mid B_j)P(B_j)}$$

1.8 Kombinatorika

Permutace (zpřeházení): $n!$ možností

Vybíráme k objektů z množiny n objektů:

výběr	s vracením (opakováním)	bez vracení (opakování)
uspořádaný (variace)	n^k s p-stmi $\frac{1}{n^k}$	$\frac{n!}{(n-k)!}$ s p-stmi $\frac{(n-k)!}{n!}$
neuspořádaný (kombinace)	$\binom{n+k-1}{k}$ s různými p-stmi	$\frac{n!}{k!(n-k)!} = \binom{n}{k}$ s p-stmi $\frac{k!(n-k)!}{n!}$

2 Náhodná veličina a náhodný vektor

2.1 Náhodná veličina

Je na pravděpodobnostním prostoru (Ω, \mathcal{A}, P) měřitelná funkce $X: \Omega \rightarrow \mathbb{R}$ (přiřazuje každému jevu jevového pole reálné číslo [wikipedia]).

Náhodné veličiny lze rozdělit na nespojitě (diskrétní) a spojitě. Diskrétní veličiny mohou nabývat pouze spočetného počtu hodnot (konečného i nekonečného), zatímco spojitě veličiny nabývají hodnoty z nějakého intervalu (konečného nebo nekonečného) [wikipedia].

Příklad: Havárie aut označíme cenou škody a můžeme se ptát, jak je pravděpodobné, že havárie dosáhne určité škody.

Pro každý interval I platí:

$$X^{-1}(I) = \{\omega \in \Omega \mid X(\omega) \in I\} \in \mathcal{A}$$

Popisuje ji **Rozdělení pravděpodobnosti** náhodné veličiny X :

$$P_X(I) = P(X \in I) = P(\{\omega \in \Omega \mid X(\omega) \in I\})$$

to je funkce, která určuje pravděpodobnost toho, že náhodná veličina nabude určité hodnoty.

2.1.1 Distribuční funkce a hustota

Místo pravděpodobnostní funkce, můžeme použít úspornější **Distribuční funkci** (F_X), která se omezuje na intervaly tvaru $I = (-\infty, t]$, $t \in \mathbb{R}$

$$P(X \in (-\infty, t]) = P(X \leq t) = P_X((-\infty, t]) = F_X(t)$$

Různými kombinacemi distribuční funkce ($F_X : \mathbb{R} \rightarrow [0, 1]$) můžeme plně nahradit pravděpodobnostní funkci.

Vlastnosti distribuční funkce:

- neklesající
- zprava spojitá
- $\lim_{x \rightarrow -\infty} F_X(t) = 0$, $\lim_{x \rightarrow \infty} F_X(t) = 1$

Distribuční funkce pro absolutně spojitou veličinu:

$$F_X(t) = \int_{-\infty}^t f_X(u) du,$$

kde f_X je tzv. **hustota náhodné veličiny**. Je to nezáporná funkce a splňuje

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

2.1.2 Kvantilová funkce

Distribuční funkce "obráceně" - tedy pokud je spojitá, je to její inverze. Jinak:

Kvantilová funkce $q_X : (0, 1) \rightarrow \mathbb{R}$, $q_X(\alpha) = t$, kdy pro t platí:

$P(X < t) \leq \alpha \leq P(X \leq t)$. Tato t tvoří omezený interval, z něhož bereme obvykle prostředek.

$q_X(1/2)$ – **medián**

Náhodné veličiny X_1, \dots, X_n jsou **nezávislé**, pokud pro všechny intervaly I_1, \dots, I_n jsou jevy $X_1 \in I_1, \dots, X_n \in I_n$ nezávislé, tj.

$$P(X_1 \in I_1, \dots, X_n \in I_n) = \prod_{i=1}^n P(X_i \in I_i)$$

2.2 Náhodný vektor (n-rozměrná náhodná veličina)

Je na pravděpodobnostním prostoru (Ω, \mathcal{A}, P) měřitelná funkce $X : \Omega \rightarrow \mathbb{R}^n$.

Používáme ho v případech, kdy je k popisu výsledku náhodného pokusu nutné použít více čísel [wikipedia].

Příklad: Chceme popsat vztah např. mezi výškou a váhou osob. K tomu potřebujeme více informací než jen popis jednotlivých náhodných veličin.

Pro každý n -rozměrný interval I platí:

$$X^{-1}(I) = \{\omega \in \Omega \mid X(\omega) \in I\} \in \mathcal{A}$$

Lze psát $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$, kde zobrazení $X_k : \Omega \rightarrow \mathbb{R}$, $k = 1, \dots, n$ jsou náhodné veličiny.

Náhodný vektor lze považovat za vektor náhodných veličin $X = (X_1, \dots, X_n)$.

Je popsán **Sdruženým rozdělením pravděpodobnosti**:

$$P_X(I_1 \times \dots \times I_n) = P(X_1 \in I_1, \dots, X_n \in I_n)$$

$$= P(\{\omega \in \Omega \mid X_1(\omega) \in I_1, \dots, X_n(\omega) \in I_n\}), \text{ kde } I_1, \dots, I_n \text{ jsou intervaly v } \mathbb{R}.$$

Opět můžeme použít úspornější **Sdruženou distribuční funkci** (F_X)

$$P(X_1 \in (-\infty, t_1], \dots, X_n \in (-\infty, t_n]) =$$

$$P(X_1 \leq t_1, \dots, X_n \leq t_n) = P_X((-\infty, t_1] \times \dots \times (-\infty, t_n]) = F_X(t_1, \dots, t_n)$$

2.3 Obecné náhodné veličiny

Náhodné veličiny nemusí být reprezentovány pouze reálnými čísly, ale třeba i čísly komplexními. V některých případech se používají i jiné než numerické hodnoty, například "rub", "líc", "kámen", "papír" atp.

Směs náhodných veličin

Směs X náhodných veličin V_1, \dots, V_n s koeficienty c_1, \dots, c_n , $\sum_{i=1}^n c_i = 1$ má pravděpodobnostní míru:

$$P_X = \sum_{i=1}^n c_i P_{V_i} \text{ (distribuční funkce je definována podobně)}$$

Př: Učitel s nějakou psí vybere otázku ke zkoušení, náhodné veličiny V_1, \dots, V_n jsou odpovědi studenta na danou otázku. Znamka studenta je pak náhodná veličina modelována pomocí směsi veličin.

Směsi mohou být diskrétní (všechny veličiny jsou diskrétní), spojité, ale i smíšené.

3 Základní charakteristiky náhodných veličin a náhodných vektorů

3.1 Střední hodnota

Značení E nebo μ . Jedná se o tzv. charakteristiku polohy (angl. measures of central tendency)

3.1.1 Střední hodnota náhodné veličiny

Je definována zvlášť pro:

- **diskrétní** náhodnou veličinu U s oborem hodnot R : $EU = \mu_U = \sum_{t \in R} t \cdot p_U(t)$
- **spojitou** náhodnou veličinu V : $EV = \mu_V = \int_{-\infty}^{\infty} t \cdot f_V(t) dt$

Vlastnosti: $E(X \pm Y) = EX \pm EY$, $E(rX) = rE(X)$ — linearita střední hodnoty

Pokud X a Y jsou nezávislé: $E(XY) = EXEY$

3.1.2 Střední hodnota náhodného vektoru

$EX = (EX_1, \dots, EX_n)$

3.2 Rozptyl (disperze)

Značení σ^2 , D , var . Jedná se o charakteristiku variability (angl. ??).

3.2.1 Rozptyl náhodné veličiny

Je to vlastně střední hodnota kvadrátu odchylky od střední hodnoty.

$$DX = E((X - EX)^2) = E(X^2) - (EX)^2$$

3.2.2 Rozptyl náhodného vektoru

$$DX = (DX_1, \dots, DX_n)$$

Směrodatná odchylka - odmocnina z rozptylu — má stejný fyzikální rozměr jako veličina

3.3 Další číselné charakteristiky náhodného vektoru

Pro jednorozměrné náhodné veličiny střední hodnota a rozptyl dávají **dostatečnou** informaci pro výpočet rozptylu jeho lineárních funkcí:

$$E(X+Y) = EX+EY \quad E(X-Y) = EX-EY \quad D(X+Y) = DX+DY \quad D(X-Y) = DX+DY$$

Pro náhodný vektor to ale nestačí, a proto zavádíme další charakteristiky:

$$E(X+Y) = EX+EY$$

$$D(X+Y) = DX+DY+2\text{cov}(X,Y),$$

kde $\text{cov}(X,Y)$ je kovariance náhodných veličin X,Y .

3.3.1 Kovariance a korelace

Kovariance určuje míru statistické závislosti mezi náhodnými veličinami. Je definována jako střední hodnota součinu odchylek obou náhodných veličin X, Y od jejich středních hodnot (druhý vzorec není tak srozumitelný, ale je jednodušší pro výpočet)

$$\text{cov}(X,Y) = E((X-EX)(Y-EY)) = E(XY) - EXEY$$

Vlastnosti kovariance:

$$\text{cov}(X,X) = DX,$$

$$\text{cov}(Y,X) = \text{cov}(X,Y)$$

$$\text{cov}(aX+b, cY+d) = a \cdot c \cdot \text{cov}(X,Y)$$

$\text{cov}(X, Y) = 0$ - pro nezávislé veličiny.

Při výpočtech je místo kovariance výhodnější používat **korelaci** (což je kovariance pro normované náhodné veličiny).

$$\rho(X, Y) = \text{cov}(\text{norm}X, \text{norm}Y) = \text{cov}(X, Y) / (\sigma_X \sigma_Y) = E(\text{norm}X \cdot \text{norm}Y)$$

Korelace nabývá hodnot $\{-1, 1\}$

Pro $\rho = 1$ je mezi X, Y **přímá lineární závislost**.

Pro $\rho = -1$ **nepřímá lineární závislost**.

Pro $\rho = 0$ říkáme, že jsou veličiny **nekorelované**. Zároveň to znamená, že jsou lineárně nezávislé, nikoliv obecně nezávislé (je to nutná podmínka pro obecnou nezávislost, ale nikoliv postačující).

Pro náhodný vektor $X = (X_1, \dots, X_n)$ definujeme **kovariační matici**:

$$\begin{aligned} \Sigma_X &= \begin{bmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \cdots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \cdots & \text{cov}(X_n, X_n) \end{bmatrix} \\ &= \begin{bmatrix} DX_1 & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_n) \\ \text{cov}(X_1, X_2) & DX_2 & \cdots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_1, X_n) & \text{cov}(X_2, X_n) & \cdots & DX_n \end{bmatrix}. \end{aligned}$$

Je symetrická pozitivně (pozitivně if ur Velebil či Dostál) semidefinitní, na diagonále má rozptyly.

Definujeme také **korelační matici**:

$$\rho_X = \begin{bmatrix} 1 & \rho(X_1, X_2) & \cdots & \rho(X_1, X_n) \\ \rho(X_1, X_2) & 1 & \cdots & \rho(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(X_1, X_n) & \rho(X_2, X_n) & \cdots & 1 \end{bmatrix}.$$

Ta je také symetrická pozitivně definitní.

Důležitá rozdělení

Diracova - všude pst. 0, jen v jednom místě 1 → všechna diskrétní rozdělení se dají chápat jako směs Diracových rozdělení (useless poznámka 😊)

Rovnoměrná - každý výsledek má stejnou pst

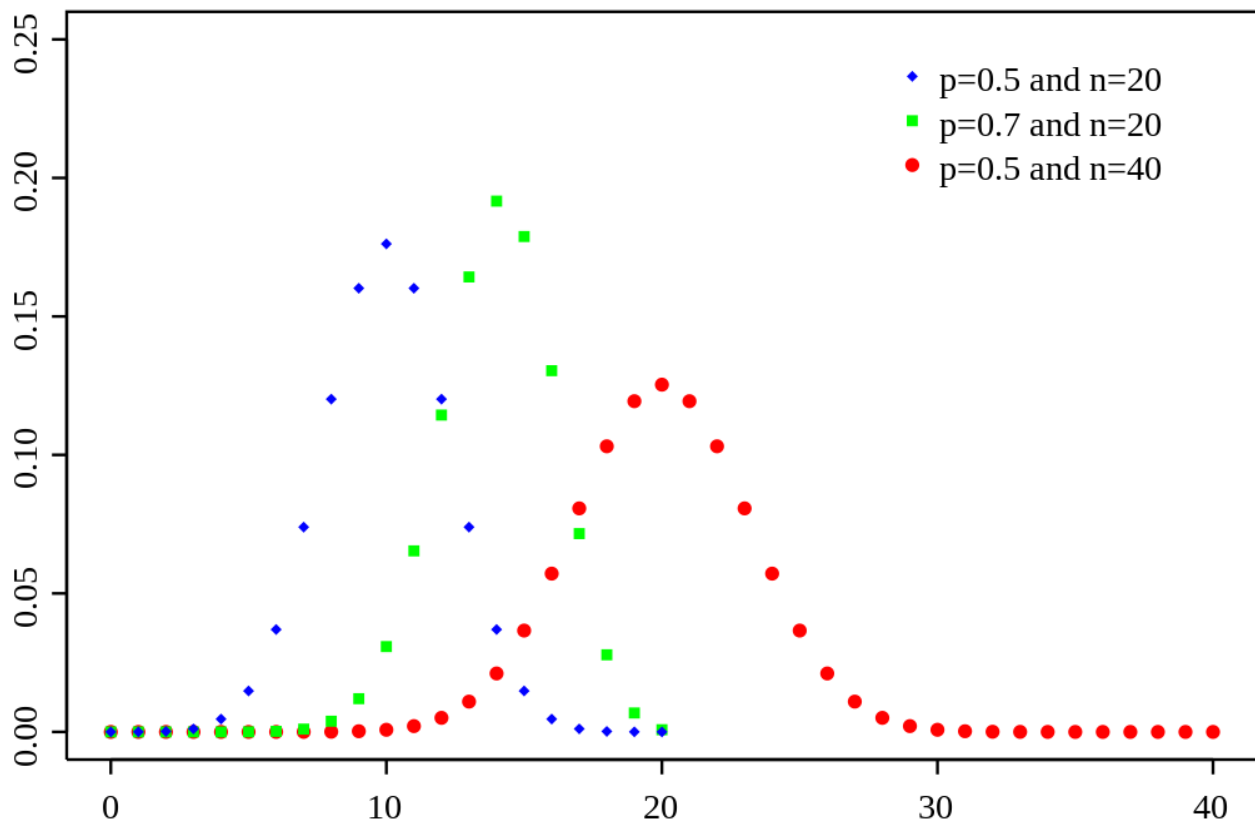
Bernoulliho (Alternativní) - Alt(q)

- 2 možné výsledky, q ... pst výsledku 1

Binomická - $Bi(n, p)$ - popisuje četnost výskytu náhodného jevu v n nezávislých pokusech, v nichž má jev stále stejnou pravděpodobnost p (pro $n = 1$ dostavame $Alt(p)$)

Pravděpodobnost, že jev nastane právě x -krát z n pokusů při pravděpodobnosti jevu p , je určena rozdělením

$$P[X = x] = \binom{n}{x} p^x (1 - p)^{n-x}$$



Geometrická - počet neúspěchů do prvního úspěchu, kde v každém pokusu je stejná pravděpodobnost úspěchu p

$$p_x(k) = p^k(1 - k), EX = \frac{q}{1-q}, DX = \frac{q}{(1-q)^2}$$

Rovnoměrné - všude 0, kromě intervalu a, b

Normální (Gaussovo) $N(\mu, \sigma^2)$

Hustota:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

4 Odhady základních číselných charakteristik

Typicky jsou číselné parametry rozdělení skryté a nebývají přímo měřitelné (kvůli velikosti celého souboru).

Výběrový soubor rozsahu n označujeme jako **náhodný výběr $X = (X_1, \dots, X_n)$** .

Statistika je každá měřitelná funkce, definovaná na náhodném výběru libovolného rozsahu.

Empirické rozdělení = rozdělení náhodného výběru.

4.1 Vlastnosti odhadů

Nestranný: $E(\hat{\theta}) = \theta^*$ (střední hodnota odhadu je skutečná hodnota)

Asymptoticky nestranný: pro velikost souboru blíží se limitně k nekonečnu bude nestranný

Eficientní - s malým rozptylem, posuzujeme podle kritéria: $E((\hat{\theta} - \theta^*)^2)$

Konzistentní - se zvětšováním souboru se zlepšuje, tj. asymptoticky nestranný a rozptyl se blíží 0

Robustní - odolný vůči šumu, těžce se definuje, ale je to v praxi velmi důležitá vlastnost

Následující statistiky se používají pro odhady číselných parametrů.

4.2 Výběrový průměr

z náhodného výběru $\mathbf{X} = (X_1, \dots, X_n)$ je **nestranný konzistentní odhad střední**

hodnoty: $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$

4.3 Výběrový rozptyl

náhodného výběru $\mathbf{X} = (X_1, \dots, X_n)$ je nestranný konzistentní odhad rozptylu:

$$S_x^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \hat{X}_n)^2$$

(pokud bychom dělili pouze n, byl by vychýlený, ne nestranný)

Ještě tam má Navara spousty věcí typu rozdělení “chí kvadrát” a podobné nebezpečné věci. Ty si stejně nikdy nezapamatujeme 😊

Centrální limitní věta

Nechť jsou X_1, \dots, X_n nezávislé stejně rozdělené náhodné veličiny, se stejnou střední hodnotou EX a směrodatnou odchylkou $\sigma_x \neq 0$, jejichž průměr je \bar{X}_n . Pak normovaná náhodná veličina:

$$Y_n = \frac{\sqrt{n}}{\sigma_X}(\bar{X}_n - EX)$$

se blíží normovanému normálnímu rozdělení $N(0, 1)$:

$\forall t \in \mathbb{R} : \lim_{n \rightarrow \infty} F_{Y_n}(t) = \Phi(t)$, kde Φ je distribuční funkce normovaného normálního rozdělení.

Tedy zvětšujeme počet veličin, které sčítáme.

https://cmp.felk.cvut.cz/~navara/stat/CLV_ebook.pdf — Co (ne)říká CLV

Čebyševova nerovnost

1. označujeme tvrzení, že pro libovolnou nezápornou náhodnou veličinu X se střední hodnotou $E(X)$ je pravděpodobnost, že veličina X nabude alespoň hodnoty ϵ dána podmínkou

$$P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon}$$

2. Pro libovolnou náhodnou veličinu X se střední hodnotou $E(X)$ a rozptylem $D(X)$ je pravděpodobnost že absolutní hodnota $|X - E(X)|$ nabude hodnoty menší než libovolné $\epsilon > 0$ omezená Čebyševovou nerovností 2. typu

$$P(|X - E(X)| < \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}$$

nebo také

$$P(|X - E(X)| < \varepsilon \cdot \sigma) \geq 1 - \frac{1}{\varepsilon^2}$$

$$\text{kde } \sigma = \sqrt{D(X)}$$

Nevím, proč je tady toho tolik o ML odhadu a metodě momentů — možná proto, že je má Navara rád? (a ML je všude)

Každopádně na mé ústní z PST jej to hodně zajímalo, jaké má která metoda výhody



5 Metoda maximální věrohodnosti

Definice:

Nechť náhodný výběr X_1, \dots, X_n má rozdělení určené sdruženou hustotu

$$f(\mathbf{x}; \theta) = \prod_{i=1}^n f_{\theta}(x_i) \quad \text{pro spojitě rozdělení, resp.}$$

$$p(\mathbf{x}; \theta) = \prod_{i=1}^n \mathbb{P}_{\theta}(X_i = x_i) \quad \text{pro diskrétní rozdělení.}$$

Při pevné hodnotě $\mathbf{x} = (x_1, \dots, x_n)$ se funkce $f(\mathbf{x}; \theta)$ resp. $p(\mathbf{x}; \theta)$ jakožto funkce θ nazývá **věrohodnostní funkce** a značí $L(\theta; \mathbf{x})$.

Obvykle je vhodné nehledat maximum věrohodnostní funkce přímo, ale zlogaritmovat ji. Součiny se nám změň na součty ale na výsledek to nebude mít vliv.

Často nechceme odhadnout jen střední hodnotu či rozptyl, ale i jiné parametry rozdělení. Rozdělení náhodné veličiny X závisí na vektoru parametrů $\vartheta = (\vartheta_1, \dots, \vartheta_k) \in \Pi$, kde $\Pi \subseteq \mathbb{R}^k$ je **parametrický prostor**, tj. množina všech přípustných hodnot parametrů. Myšlenkou této metody je, že hledáme takové hodnoty parametrů, které by nejlépe vysvětlovali realizaci náhodného výběru, tj. při kterých by pozorované výsledky byly "nejméně nepravděpodobné".

5.1 Metoda maximální věrohodnosti pro diskrétní rozdělení

Nechť $x = (x_1, \dots, x_n)$ je realizace náhodného výběru z diskrétního rozdělení s pravděpodobnostní funkcí $p_X(\cdot; \vartheta)$ závislou na vektoru parametrů $\vartheta \in \Pi$. Pak definujeme věrohodnost realizace diskrétního rozdělení (angl. likelihood) $L: \Pi \rightarrow (0, 1)$ vztahem

$$L(\vartheta) = p_X(x; \vartheta) = \prod_{j=1}^n p_X(x_j; \vartheta)$$

Je třeba rozlišovat mezi pravděpodobností a věrohodností! Pravděpodobnost je určena pro předpovídání výsledků budoucího pokusu při známém pravděpodobnostním modelu a tudíž má definiční obor jevy z nějaké σ -algebry. Naopak věrohodnost vyhodnocuje sérii pokusů již realizovaných a dovoluje jim přizpůsobit neznámé parametry pravděpodobnostního modelu, a proto je definována na prostoru Π všech možných hodnot parametrů rozdělení.

Metoda maximální věrohodnosti považuje za správný odhad parametrů takové hodnoty, které maximalizují věrohodnost. Protože výpočet skoro vždy vede na hledání nulové derivace, často se maximalizujeme logaritmus věrohodnosti (je to jednodušší na derivování).

5.2 Metoda maximální věrohodnosti pro spojitě rozdělení

Nechť $x = (x_1, \dots, x_n)$ je realizace náhodného výběru ze spojitě rozdělení se spojitou hustotou $f_X(\cdot; \vartheta)$ závislou na vektoru parametrů $\vartheta \in \Pi$. Pak definujeme věrohodnost realizace spojitě rozdělení $\Lambda: \Pi \rightarrow (0, \infty)$ vztahem

$$\Lambda(\vartheta) = f_X(x; \vartheta) = \prod_{j=1}^n f_X(x_j; \vartheta)$$

Metoda maximální věrohodnosti opět považuje za správný odhad parametrů takové hodnoty, které maximalizují věrohodnost.

Metoda momentů

Metoda momentů je založena na rovnosti výběrových momentů a momentů rozdělení. Nelze jednoznačně rozhodnout, která z metod dává lepší výsledky. Rozhodování provádíme podle konkrétní situace, nejčastěji rozhoduje jednoduchost získaných vzorců.

Metoda momentů zohledňuje všechna data z výběru a volíme ji v případech, kdy je soustava věrohodnostních rovnic obtížně řešitelná.

Pro základní rozdělení dávají obě metody shodné výsledky a v případě složitějších rozdělení můžeme jako další kritérium uvažovat, které vzorce jsou méně citlivé na zavlečené chyby do hodnot výběru.

Je-li $X_1 X_2 X_3$ náhodný výběr pak definujeme k -tý výběrový moment jako

$$M'_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad 1 \leq k$$

a k -tý centrální výběrový moment jako

$$M_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, \quad 1 \leq k.$$

Intervalové odhady

Dosud jsme skutečnou hodnotu parametru ϑ^* nahrazovali **bodovým odhadem** $\hat{\Theta}$ (což je náhodná veličina). Nyní místo toho hledáme **intervalový odhad**, tzv. **interval spolehlivosti** I , což je minimální interval takový, že

$$P(\vartheta^* \in I) \geq 1 - \alpha,$$

kde $\alpha \in (0, 1)$ je pravděpodobnost, že meze intervalu I budou překročeny; $1 - \alpha$ je **koeficient spolehlivosti**. Obvykle hledáme **horní**, resp. **dolní jednostranný** odhad,

$$I = (-\infty, q_{\hat{\Theta}}(1 - \alpha)), \text{ resp. } I = \langle q_{\hat{\Theta}}(\alpha), \infty \rangle,$$

nebo **(symetrický) oboustranný** odhad,

$$I = \langle q_{\hat{\Theta}}(\frac{\alpha}{2}), q_{\hat{\Theta}}(1 - \frac{\alpha}{2}) \rangle.$$

K tomu potřebujeme znát rozdělení odhadu $\hat{\Theta}$.

Pro představu, co musíme dělat (odhad střední hodnoty pro normální rozdělení se známém rozptylem):

σ^2 známe; X odhadneme z realizace x ; μ neznáme

S pravděpodobností $1 - \alpha$ je $q_{N(\mu, \sigma^2)}(\frac{\alpha}{2}) \leq X \leq q_{N(\mu, \sigma^2)}(1 - \frac{\alpha}{2})$,
po znormování

$$-\Phi^{-1}(1 - \frac{\alpha}{2}) = \Phi^{-1}(\frac{\alpha}{2}) \leq \text{norm } X = \frac{X - \mu}{\sigma} \leq \Phi^{-1}(1 - \frac{\alpha}{2}),$$

$$X - \sigma \Phi^{-1}(1 - \frac{\alpha}{2}) \leq \mu \leq X + \sigma \Phi^{-1}(1 - \frac{\alpha}{2}).$$

Obvykle místo jedné realizace použijeme realizaci výběrového průměru $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$;

po normalizaci $\text{norm } \bar{X}_n = \frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu) \sim N(0, 1)$.

S pravděpodobností $1 - \alpha$ je

$$\bar{X}_n - \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \frac{\alpha}{2}) \leq \mu \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \frac{\alpha}{2}),$$

pro jednostranné odhady

$$\mu \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \alpha),$$

$$\bar{X}_n - \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \alpha) \leq \mu,$$

Dostali jsme intervalové odhady pro μ

$$\left\langle \bar{X} - \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \frac{\alpha}{2}), \bar{X} + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \frac{\alpha}{2}) \right\rangle,$$

Ale když rozptyl neznáme, nevadí! 9 z 10 statistiků jej zvládne hravě nahradit výběrovým rozptylem! Ovšem pozor, pak nebudeme mít kvantilovou funkci normálního rozdělení, ale Studentova t-rozdělení s n stupni volnosti ❤

To si snad nemusíme pamatovat, protože pro velký počet dat (tedy i stupňů volnosti), se Studentovo rozdělení pěkně blíží normálnímu - a kdo dneska používá málo dat?

Testování hypotéz

Otravná záležitost plná tabulek a taak.

Jako asi to je užitečné a na VŠE to mají rádi, ale já si to zapamatuji jen stěží..

Tohle tady k tomu bylo:

Testování hypotéz - příklad

https://youtu.be/VMRZO_yYWrA


https://youtu.be/VMRZO_yYWrA

- ✓ chybi druhy rozdělení a co vyjadruji
- ✓ Taky chybi centralni limitni veta
- ✓ vlastnosti odhadu — eg vychylenost
- ✓ intervalove odhady take chybi
- ☐ testovani hypotez tu jsou jen ve videu..
- ✓ Tady potrebuju dodelat markovovy retezce

Často si studenti říkají: PST by se mělo rozdělit do více předmětů :(

Markovské řetězce a jejich asymptotické vlastnosti

https://cs.wikipedia.org/wiki/Markov%C5%AFv_%C5%99et%C4%Bzec

 Markovův řetězec • cs.wikipedia.org

https://cmp.felk.cvut.cz/~navara/stat/Markov_ebook.pdf

Markovův řetězec popisuje obvykle **diskrétní náhodný** (stochastický či pravděpodobnostní) **proces**, pro který platí, že **pravděpodobnosti přechodu do následujícího stavu závisí pouze na současném stavu**, ne na předchozích stavech. Tato tzv. **Markovova vlastnost** dovoluje proces znázornit stavovým diagramem, kde z každého stavu (uzlu grafu) vycházejí hrany možných přechodů do dalšího stavu s přiřazenou pravděpodobností.

Diagram můžeme také zapsat maticí přechodu P :

- stochastická = součty řádků a sloupců jsou 1

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix},$$

p_{ij} pravděpodobnost přechodu ze stavu i do stavu j .

Máme zadán počáteční **řádkový** vektor stavů $p(0)$, který říká, s jakou pravděpodobností jsme právě (v čase $t = 0$) v jakém stavu. Aplikací matice P na vektor získáme rozdělení v čase $t + 1$ atd.

Protože je $p(t)$ řádkový, aplikujeme takto:

$$p(t + 1) = p(t)P$$

$$p(t + m) = p(t)P^m$$

$$P^{t+u} = P^t P^u$$

Klasifikace stavů

- Nedosažitelné - do takového stavu se nedostaneme nikdy (opakem je dosažitelný) - jako MIT
- Trvalý - pokud z něj vylezeme, zase se do něj někdy určitě dostaneme - oblíbená hospoda
- Absorpční - z takového stavu se nelze dostat = smrt
- Periodické - vracíme se do něj s nějakou periodou (větší než 0) - Vánoce
- Ergodický stav - trvalý a neperiodický
- Přechodný - někdy v budoucnosti se do něj vrátíme s pravděpodobností menší než 1 - podnik, co nás nenadchnul
- Stavy komunikují, pokud z jednoho vlezeme do druhého a zase zpět — když je jeden z nich přechodný, i druhý je přechodný, když je jeden z nich trvalý, druhý je také

Množina **trvalých** stavů je **uzavřená**, jestliže ji nelze opustit.

Komponenta je (každá) neprázdná uzavřená množina (trvalých) stavů, která neobsahuje vlastní (=menší neprázdnou) uzavřenou podmnožinu stavů.

Řetězec je **nerozložitelný**, jestliže je tvořen jednou komponentou. Jinak jej můžeme rozložit na komponenty.

\Rightarrow Pokud jsou všechny stavy trvalé a řetězec je rozložitelný, lze matici přechodu (po vhodné permutaci stavů) vyjádřit jako *blokově diagonální*,

$$P = \begin{pmatrix} D_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & D_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & D_k \end{pmatrix},$$

kde každý blok odpovídá jedné komponentě a $\mathbf{0}$ je nulová matice odpovídajícího řádu. Pokud existují přechodné stavy a zařadíme je až za trvalé (pomocí permutace stavů), matice přechodu má tvar

$$P = \left(\begin{array}{c|c} D & \mathbf{0} \\ R & Q \end{array} \right),$$

kde D vyjadřuje pravděpodobnosti přechodů mezi trvalými stavy, Q mezi přechodnými a R vyjadřuje pravděpodobnosti přechodů z přechodných stavů do trvalých.

Po rozkladu množiny trvalých stavů na komponenty dostaneme tvar

$$P = \left(\begin{array}{cccc|c} D_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & D_2 & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & D_k & \mathbf{0} \\ \hline R_1 & R_2 & \cdots & R_k & Q \end{array} \right).$$

Věta. *Nechť stavy $1, \dots, i$ jsou absorpční, $i + 1, \dots, n$ přechodné. Pak matice přechodu má tvar*

$$P = \left(\begin{array}{c|c} I_i & 0 \\ \hline R & Q \end{array} \right),$$

kde I_i je jednotková matice řádu i . Pravděpodobnost, že z přechodného stavu $j > i$ skončíme v absorpčním stavu $k \leq i$, je prvek na pozici $(j - i, k)$ v matici

$$\underbrace{(I_{n-i} + Q + Q^2 + Q^3 + \dots)}_F R = F R,$$

kde I_{n-i} je jednotková matice řádu $n - i$ a

$$F = I_{n-i} + Q + Q^2 + Q^3 + \dots$$

*je **fundamentální matice** tohoto řetězce.*

Jak lépe počítat fundamentální matici?

$$F = I_{n-i} + Q + Q^2 + Q^3 + \dots = (I_{n-i} - Q)^{-1}.$$

K čemu to je? Chceme zjistit, s jakou pravděpodobností za různých podmínek skončíme v kterém absorpčním stavu, což se hodí při např. určování psti, že někdo vyhraje atd — příklady typu Alice a Bob hází míčem.

Nerozložitelné Markovské řetězce

Stacionární rozdělení pravděpodobností \mathbf{p} je takové, které se zachovává, tj.

$$\mathbf{p} P = \mathbf{p},$$

neboli levý vlastní vektor odpovídající vlastnímu číslu 1.

Lze je najít vyřešením této (homogenní) soustavy lineárních rovnic pro \mathbf{p} s dodatečnou podmínkou, že součet neznámých je 1.

Markovův řetězec je **ergodický**, je-li nerozložitelný a má všechny stavy ergodické.

Věta. ***Ergodický** Markovův řetězec má jediné stacionární rozdělení pravděpodobností; k tomu konverguje při libovolném počátečním rozdělení.*

Důsledek. *V tom případě $\lim_{t \rightarrow \infty} P^t$ existuje a je rovna matici, jejíž všechny řádky jsou rovné stacionárnímu rozdělení.*

Soustava rovnic zmíněná výše může vypadat třeba takto:

$$(a, b, 1 - a - b) P_Z = (a, b, 1 - a - b),$$

Rozložitelné řetězce bez přechodných stavů můžeme rozložit na nerozložitelné komponenty.