

Společná část: 19 - PSI

Náhodná veličina a náhodný vektor. Distribuční funkce, hustota a pravděpodobnostní funkce náhodné veličiny. Střední hodnota a rozptyl náhodné veličiny a jejich odhady. Sdružené charakteristiky náhodného vektoru. Korelace a nezávislost náhodných veličin. Metoda maximální věrohodnosti. Základní principy statistického testování hypotéz. Markovské řetězce, klasifikace stavů.

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 náhodného 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 $\mathcal{A} \subseteq \exp \Omega$, která splňuje podmínky σ -algebry (viz. 1.3).
- **Pravděpodobnost** není určena strukturou jevů jako u Laplaceova modelu, je to funkce $P : \mathcal{A} \rightarrow \langle 0, 1 \rangle$, splňující podmínky:

$$(P1) \quad P(\mathbf{1}) = 1,$$

$$(P2) \quad P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n), \text{ pokud jsou množiny (=jevy) } A_n, n \in \mathbb{N}, \text{ po dvou neslučitelné}$$

- **Pravděpodobnostní prostor** je trojice (Ω, \mathcal{A}, P) , kde Ω je neprázdná množina, \mathcal{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 \mathcal{A} nějaké množiny Ω musí splňovat podmínky:

1. $\emptyset \in \mathcal{A}$
2. $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$ (uzavřenost vůči doplňku)

3. $(\forall n \in \mathbb{N} : A_n \in \mathcal{A}) \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$ (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.

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.

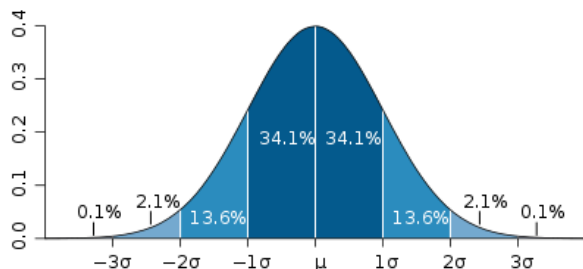
Pro každý interval I platí

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

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 udává pravděpodobnost toho, že náhodná veličina nabyde určité hodnoty.



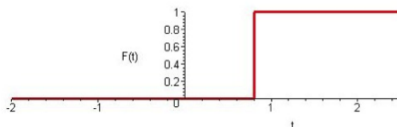
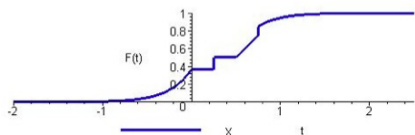
Místo pravděpodobnostní funkce, můžeme použít úspornější **Distribuční funkci** (F_X), která se omezuje na intervaly tvaru $I = (-\infty, t], t \in \mathbb{R}$

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

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

Vlastnosti distribuční funkce:

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



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 ($f_X : \mathbb{R} \rightarrow \langle 0, \infty \rangle$) a splňuje $\int_{-\infty}^{\infty} f_X(u) du = 1$. 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 $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$. 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

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

Lze psát

$$\mathbf{X}(\omega) = (X_1(\omega), \dots, X_n(\omega)),$$

kde zobrazení $X_k : \Omega \rightarrow \mathbb{R}, k = 1, \dots, n$ jsou náhodné veličiny.

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

Je popsán **Sdruženým rozdělením pravděpodobnosti**:

$$P_{\mathbf{X}}(I_1 \times \dots \times I_n) = P[X_1 \in I_1, \dots, X_n \in I_n] = P(\{\omega \in \Omega \mid X_1(\omega) \in I_1, \dots, X_n(\omega) \in I_n\}),$$

kde I_1, \dots, I_n jsou intervaly v \mathbb{R} . Z toho vyplývá pravděpodobnost pro libovolnou borelovskou množinu I v \mathbb{R}^n

$$P_{\mathbf{X}}(I) = P[\mathbf{X} \in I] = P(\{\omega \in \Omega \mid \mathbf{X}(\omega) \in I\})$$

Opět můžeme použít úspornější **Sdruženou distribuční funkci** ($F_{\mathbf{X}}$)

$$P[X_1 \in (-\infty, t_1], \dots, X_n \in (-\infty, t_n]] = P[X_1 \leq t_1, \dots, X_n \leq t_n] = P_{\mathbf{X}}((-\infty, t_1] \times \dots \times (-\infty, t_n]) = F_{\mathbf{X}}(t_1, \dots, t_n)$$

Vlastnosti distribuční funkce:

- neklesající (ve všech proměnných)
- zprava spojitá (ve všech proměnných)
- $\lim_{t_1 \rightarrow \infty, \dots, t_n \rightarrow \infty} F_{\mathbf{X}}(t_1, \dots, t_n) = 1$
- $\lim_{t_1 \rightarrow -\infty, \dots, t_n \rightarrow -\infty} F_{\mathbf{X}}(t_1, \dots, t_n) = 0$

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.

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 \mathbb{R} :

$$EU = \mu_U = \sum_{t \in \mathbb{R}} t \cdot p_U(t)$$

- **spojitou** náhodnou veličinu V :

$$EV = \mu_V = \int_{-\infty}^{\infty} t \cdot f_V(t) dt$$

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. measures of central tendency).

3.2.1 Rozptyl náhodné veličiny

Je to vlastně střední hodnota kvadrátu odchylky od střední hodnoty.

$$DX = E\left((X - EX)^2\right) = E(X^2) - (EX)^2$$

3.2.2 Rozptyl náhodného vektoru

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

3.3 Další číselné charakteristiky náhodného vektoru

Pro jednorozměrné náhodné veličiny střední hodnota a rozptyl dávají **dostatečnou** informaci pro výpočet rozptylu jeho lineárních funkcí (lineární kombinace různých náhodných veličin):

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

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

$$D(X + Y) = DX + DY$$

$$D(X - Y) = DX + DY$$

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

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

$$D(X + Y) = DX + DY + 2cov(X, Y),$$

kde $cov(X, Y)$ 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:

$$\begin{aligned}\text{cov}(X, X) &= DX, \\ \text{cov}(Y, X) &= \text{cov}(X, Y) \\ \text{cov}(aX + b, cY + d) &= ac \text{cov}(X, Y) \\ \text{cov}(X, Y) &= 0 - \text{pro nezávislé veličiny}\end{aligned}$$

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

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

Korelace nabývá hodnot $\langle -1, 1 \rangle$, pro $\varrho = 1$ je mezi X, Y přímá lineární závislost, pro $\varrho = -1$ nepřímá lineární závislost. Pro $\varrho = 0$ 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 $\mathbf{X} = (X_1, \dots, X_n)$ definujeme **kovariační matici**:

$$\Sigma_{\mathbf{X}} = \begin{bmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \dots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \dots & \text{cov}(X_n, X_n) \end{bmatrix} = \begin{bmatrix} DX_1 & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & DX_2 & \dots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \dots & DX_n \end{bmatrix}$$

a **korelační matici**:

$$\varrho_{\mathbf{X}} = \begin{bmatrix} 1 & \varrho(X_1, X_2) & \dots & \varrho(X_1, X_n) \\ \varrho(X_2, X_1) & 1 & \dots & \varrho(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \varrho(X_n, X_1) & \varrho(X_n, X_2) & \dots & 1 \end{bmatrix}$$

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** $\mathbf{X} = (X_1, \dots, X_n)$.

Statistika je každá měřitelná funkce, definovaná na náhodném výběru libovolného rozsahu. Následující statistiky se používají pro odhady číselných parametrů.

4.1 Výběrový průměr

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

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

4.2 Výběrový rozptyl

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

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

5 Metoda maximální věrohodnosti

Často nechceme odhadnout jen střední hodnotu či rozptyl, ale i jiné parametry rozdělení. Rozdělení náhodné veličiny X závisí na vektoru parametrů $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_k) \in \Pi$, 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 $\mathbf{x} = (x_1, \dots, x_n)$ je realizace náhodného výběru z diskrétního rozdělení s pravděpodobnostní funkcí $p_X(\cdot; \boldsymbol{\vartheta})$ závislou na vektoru parametrů $\boldsymbol{\vartheta} \in \Pi$. Pak definujeme věrohodnost realizace diskrétního rozdělení (angl. likelihood) $L : \Pi \rightarrow \langle 0, 1 \rangle$ vztahem

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

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

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

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