

Matematika 6F — Statistika

1. Distribuční funkce

Distribuční funkce $F(t)$, X je náhodná veličina, reálná funkce $F(t)$ je definovaná pro každé $t \in R$ vztahem

$$F(t) = P[X \in (-\infty, a)] = P[X < a]$$

$$P[X \geq a] = 1 - F(a)$$

$$P[X = a] = P[X \leq a] - P[X < a] = \lim_{t \rightarrow a^+} F(t) - F(a)$$

Směs $G = \alpha \cdot F_X + (1 - \alpha) \cdot F_Y \quad \forall \alpha \in (0, 1)$

Hustota

$$F(t) = \sum_i p_i, \quad F(t) = \int_{-\infty}^x f(t) dt$$

$$f(x) = F'(x) = \frac{dF(x)}{dx}$$

Nezávislé veličiny

$$F(x, y) = F_1(x) \cdot F_2(y), \quad f(x, y) = f_1(x) \cdot f_2(y)$$

Marginální hustoty

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

2. Číselné charakteristiky

2.1. Střední hodnota (EX), rozptyl (DX)

EX pro diskrétní rozdělení

$$EX = \sum_i x_i \cdot P[X = x_i] = \sum_i x_i \cdot p_i$$

EX pro spojitě rozdělení

$$EX = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

$$Ea = a, \quad E(EX) = EX, \quad E(aX) = a \cdot EX$$

$$E(X + Y) = EX + EY, \quad E(a + x) = a + EX$$

$$P[a \leq X \leq b] = 1 \Rightarrow a \leq X \leq b$$

Pro nezávislé X, Y : $E(X \cdot Y) = (EX)(EY)$

Rozptyl

$$DX = \text{var } X = \sigma^2 = EX^2 - (EX)^2$$

$$Da = 0, \quad D(aX) = a^2 DX, \quad D(a + X) = DX$$

Pro nezávislé X, Y : $D(X + Y) = DX + DY$

Směrodatná odchylka $\sigma = \sqrt{DX}$

Čebyševova nerovnost

$$P[|X - EX| > \varepsilon] < \frac{DX}{\varepsilon^2}, \quad P[|X - EX| \leq \varepsilon] \geq 1 - \frac{DX}{\varepsilon^2}$$

Obecný moment $m_k = EX^k$

Centrální moment $\mu_k = E(X - EX)^k$

3. Statistická rozdělení

3.1. Rovnoměrné – diskrétní

$$P[X = a_k] = \frac{1}{n}, k = 1, 2, \dots, n$$

Parametry: n – přirozené číslo, $a_1, a_2, \dots, a_n \in R$

$$EX = \frac{1}{n} \sum_{k=1}^n a_k$$

$$DX = \frac{1}{n} \sum_{k=1}^n a_k^2 - \left(\frac{1}{n} \sum_{k=1}^n a_k\right)^2$$

Použití: Klasická pravděpodobnost, teorie her

3.2. Rovnoměrné – spojitě

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{jinde} \end{cases} \quad \text{Parametry: } a, b \in R, a < b$$

$$EX = \frac{a+b}{2}, \quad DX = \frac{1}{12}(b-a)^2$$

Použití: Geometrická pravděpodobnost, generování pseudonáhodných čísel

3.3. Binomické rozdělení

$$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \quad p \in (0, 1), n - \text{přirozené číslo}$$

$$EX = np, \quad DX = np(1-p)$$

Použití: Počet výskytů X nějakého jevu A v n nezávislých pokusech, pokud je pravděpodobnost výskytu jevu A v každém jednotlivém pokusu roven p .

3.4. Poissonovo rozdělení

$$P[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}, k = 0, 1, 2, \dots \quad \text{Parametr: } \lambda > 0$$

$$EX = \lambda, \quad DX = \lambda$$

Použití: Počet částic v jednotce plochy (délky objemu), počet událostí za časovou jednotku.

3.5. Alternativní rozdělení

$$P[X = 1] = p \quad P[X = 0] = 1 - p$$

Parametr: $p \in (0, 1)$

$$EX = p, \quad DX = p(1-p)$$

Použití: Náhodné pokusy s dvěma možnými výsledky, **ano×ne** experimenty.

3.6. Geometrické rozdělení

$$P[X = k] = (1-p)p^k, k = 0, 1, 2, \dots \quad \text{Parametr: } p \in (0, 1)$$

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

Použití: Rozdělení počtu pokusů do prvního úspěchu.

3.7. Normální rozdělení

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{Parametry: } \mu \in R, \sigma^2 > 0$$
$$EX = \mu, \quad DX = \sigma^2$$

Normalizace:

$$f(x) = \frac{X - EX}{\sqrt{DX}} \sim N(0, 1)$$

Použití: Základní rozdělení pravděpodobnostních a statistických metod v přírodních vědách.

3.8. Logaritmicke-normální rozdělení

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} & \text{pro } x > 0 \\ 0 & \text{jinde} \end{cases} \quad \text{Par.: } \mu \in R, \sigma^2 > 0$$
$$EX = e^{\mu + \frac{\sigma^2}{2}}, \quad DX = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$$

Použití: Rozdělení délky života, velikosti materiálu, doby čekání apod.

3.9. Trojúhelníkové rozdělení

$$f(x) = \begin{cases} \frac{2}{b-a} \left(1 - \frac{2}{b-a} \cdot \left|x - \frac{a+b}{2}\right|\right) & \text{pro } a < x < b \\ 0 & \text{jinde} \end{cases}$$

Parametry: $a, b \in R, a < b$

$$EX = \frac{a+b}{2}, \quad DX = \frac{1}{24}(b-a)^2$$

Použití: Rozdělení součtu dvou náhodných veličin se spojitým rovnoměrným rozdělením.

3.10. Exponenciální rozdělení

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{pro } x \geq 0 \\ 0 & \text{pro } x < 0 \end{cases} \quad \text{Parametr: } \lambda \in R$$
$$EX = \frac{1}{\lambda}, \quad DX = \frac{1}{\lambda^2}$$

Použití: Teorie spolehlivosti.

3.11. χ^2 rozdělení

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{pro } x > 0 \\ 0 & \text{jinde} \end{cases}$$

Parametr: n – přirozené číslo (stupeň volnosti)

$$EX = n, \quad DX = 2n$$

Použití: Test dobré shody.

3.12. t-rozdělení

$$f(x) = \frac{\Gamma(\frac{n+1}{2})(1 + \frac{x^2}{n})^{-\frac{n+1}{2}}}{\Gamma(\frac{n}{2})\sqrt{n\pi}}$$

Parametr: n – přirozené číslo (stupeň volnosti)

$$EX = 0 \quad (\text{pro } n \geq 2), \quad DX = \frac{n}{n-2} \quad (\text{pro } n \geq 3)$$

Použití: Testování střední hodnoty při neznámém rozptylu.

3.13. F-rozdělení

$$f(x) = \begin{cases} \frac{(\frac{m}{n})^{m/2} x^{m/2-1} (1 + \frac{m}{n}x)^{-\frac{m+1}{2}}}{B(\frac{m}{2}, \frac{n}{2})} & \text{pro } x > 0 \\ 0 & \text{jinde} \end{cases}$$

Parametry: m, n – přirozená čísla (stupeň volnosti)

$$EX = \frac{n}{n-2} \quad (\text{pro } n \geq 3)$$

$$\mu = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \quad (\text{pro } n \geq 5)$$

Použití: Testování hypotéz o rozptylu.

3.14. Cauchyovo rozdělení

$$f(x) = \frac{1}{\pi} \frac{\lambda}{\lambda^2 + (x - \mu)^2} \quad \text{Parametry: } \lambda > 0, \mu \in R$$
$$EX \text{ neexistuje, } \quad DX \text{ neexistuje}$$

Použití: Rozdělení podílu dvou nezávislých normálních náhodných veličin s nulovými středními hodnotami.

4. Výběry

Výběrový průměr

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Výběrový rozptyl S^2

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Výběrová směrodatná odchylka S

$$S = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

Výběrový n -tý obecný moment

$$M_r = \frac{1}{n} \sum_{i=1}^n X_i^r$$

Výběrový r -tý centrální moment

$$M'_r = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^r$$

5. Odhady parametrů

Věta (Centrální limitní věta) Nechť $\{X_i\}_{i=1}^n$ je posloupnost vzájemně nezávislých náhodných veličin, které mají totéž rozdělení se střední hodnotou μ a s konečným rozptylem σ^2 . Potom:

$$\lim_{n \rightarrow \infty} P \left[\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq t \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-u^2/2} du = \Phi(u)$$

$$\Phi \left(\frac{x - \mu}{\sigma} \right) = F(x)$$

$$\bar{X} = \frac{1}{n} \sum X$$

$$E\bar{X} = E\left(\frac{1}{n} \sum X\right) = \frac{1}{n} \sum EX = n \cdot EX$$

$$D\bar{X} = D\left(\frac{1}{n} \sum X\right) = \frac{1}{n^2} \sum DX = \frac{1}{n^2} n \cdot DX = \frac{1}{n} DX$$

$$\mu = E\bar{X}, \quad \sigma^2 = D\bar{X}$$

Potom normováním dostaneme:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Odhad může být:

- **nestranný**, jestliže střední hodnota jeho výběrového rozdělení je rovna hledanému parametru (tozn. není vychýlený),
- **eficientní**, pokud má minimální možný rozptyl,
- **konzistentní**, jestliže $EX \rightarrow X$ pro $n \rightarrow \infty$ a $var X \rightarrow 0$ pro $n \rightarrow \infty$.

5.1. Metoda momentů

$$m_p = EY^p, \quad M_p = \hat{m}_p = \frac{1}{n} \sum_{i=1}^n X_i^p$$

$$m_p \equiv \hat{m}_p$$

5.2. Metoda maximální věrohodnosti

$$f(t, \Theta) = \prod_{i=1}^n f(t_i, \Theta)$$

5.3. Intervalové odhady

Dvojice statistik (T_D, T_H) se nazývá $100 \cdot (1 - \alpha)\%$ **interval spolehlivosti**. Statistika T_D se nazývá **dolní mez**, statistika T_H **horní mez** intervalu spolehlivosti. Číslo $(1 - \alpha)$ se nazývá **koefficient spolehlivosti**. Číslo α je **hladina významnosti**.

Věta Pro náhodný výběr z rozdělení $N(\mu, \sigma^2)$, **pokud známe hodnotu rozptylu σ^2** , platí:

$$P \left[\bar{X} - u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma^2}{n}} < \mu < \bar{X} + u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma^2}{n}} \right] = 1 - \alpha$$

kde u_β je β -kvantil $N(0, 1)$.

Věta Pro náhodný výběr z rozdělení $N(\mu, \sigma^2)$, **pokud neznáme hodnotu rozptylu σ^2** , platí:

$$P \left[\bar{X} - t_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} \right] = 1 - \alpha$$

kde $t_\beta(\nu)$ je β -kvantil t -rozdělení s ν stupni volnosti.

Věta Odhad intervalu spolehlivosti pro rozptyl σ^2 z náhodného výběru z rozdělení $N(\mu, \sigma^2)$:

$$P \left[\frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2(n-1)} \right] = 1 - \alpha$$

6. Testování hypotéz

Rozhodneme se pro:	Platí:	
	H_0	H_1
H_0	Ok	Chyba 1. druhu β
H_1	Chyba 2. druhu α	Ok

Věta Necht np_1, np_2, \dots, np_k a o_1, o_2, \dots, o_k jsou teoretické a napozorované četnosti k možným, navzájem se vylučujícím výsledkům pokusu, potom statistika

$$\sum_{i=1}^k \frac{(o_i - np_i)^2}{np_i}$$

má při $n \rightarrow \infty$ asymptoticky rozdělení χ^2 o $(k - 1)$ stupni volnosti.

7. Srovnání dvou výběrů

Nejsou-li párovatelné, potom testovaná statistika je

$$T = \frac{\bar{X} - \bar{Y}}{S \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

$$S^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{(m-1) + (n-1)}$$

$$|T| < t_{1-\frac{\alpha}{2}}(m+n-2)$$