Přednáška 12

Přehled probraných rozdělení

Standardní normální rozdělení $N(0,1)$

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

Normální rozdělení $N(\mu, \sigma^2)$

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

Γ - rozdělení $\Gamma(a,b)$, $a>0$, $b>0$

Hustota $$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$$

Momentová funkce
$$M(t) = \left(\frac{b}{b-t}\right)^a$$

$\chi^2(n)$ chí kvadrát s n stupni volnosti

$$= \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

$$f(x) = \frac{b^{1/2}}{\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}$$

Ukázali jsme:

je-li $X \sim N(0,1)$, je $X^2 \sim \chi^2(1)$

jsou-li $X_1, X_2, \dots, X_n \sim N(0,1)$ je
$$X_1^2 + X_2^2 + \dots + X_n^2 \sim \chi^2(n)$$

Studentovo t-rozdělení s n stupni volnosti

Máme náhodné veličiny $Z \sim N(0,1)$ a $X \sim \chi^2(n)$ a zajímá nás podíl náhodných veličin

$$\frac{Z}{\sqrt{\frac{X}{n}}}$$

Distribuční funkce $Y = \sqrt{X}$

$$F_Y(y) = P(\sqrt{X} \leq y) = P(X < y^2) =$$

$$= \int_0^{y^2} \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} \cdot e^{-\frac{x}{2}}\, dy =$$

$$= \int_0^{y} \frac{1}{2^{n/2-1}\Gamma(n/2)} t^{n-1} e^{-\frac{t^2}{2}}\, dt$$

Hustota distribuční funkce Y je

$$f_Y(t) = \frac{1}{2^{\frac{n}{2}-1}\Gamma(n/2)} t^{n-1} e^{-\frac{t^2}{2}}$$

Nyní počítáme distribuční funkci pro $T = \frac{Z}{\sqrt{X/n}} = \frac{\sqrt{n}\, Z}{\sqrt{X}}$

Dostaneme $f_T(t) = \frac{\Gamma((n+1)/2)}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$

Věta Nechť $X_1, X_2, \dots, X_n$ je náhodný výběr z normálního rozložení $N(\mu, \sigma^2)$. Potom pro statistiky M a S^2 platí následující

(1) M a S^2 jsou nezávislé náhodné veličiny

(2) $M \sim N(\mu, \frac{\sigma^2}{n})$ a proto

$$U = \frac{M-\mu}{\sigma \sqrt{n}} \sim N(0,1),$$

(3) $$T = \frac{M-\mu}{S\sqrt{n}} \sim t(n-1),$$

(4) $$K = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

(5) $$\sum_{i=1}^{n} (X_i - \mu)^2 / \sigma^2 \sim \chi^2(n).$$

Komentář (2) se používá k odhadu μ známe-li σ.

(3) se používá k odhadu μ, neznáme-li σ.

(4) slouží k odhadu σ^2, neznáme-li μ.

(5) slouží k odhadu σ^2, známe-li μ

Bodové a intervalové odhady

Nechť náhodný výběr $X_1, X_2, \dots, X_n$ závisí na parametru θ (např. střední hodnota nebo rozptyl, nebo obojí). Bodovým odhadem parametru θ je nějaká statistika $T(X_1, X_2, \dots, X_n)$, která je blízko parametru θ. Rozdíl $E(T) - \theta$ nazveme vychýlení, je-li $E(T) = \theta$, pak je odhad T nestranný.

M je nestranný odhad střední hodnoty

S^2 je nestranný odhad rozptylu

Intervalový odhad parametru θ rozumíme interval (T_L, T_U), kde $T_L(X_1, \dots X_n)$ a $T_U(X_1, \dots X_n)$ jsou statistiky výběru $X_1, X_2, \dots, X_n$. Platí-li

$$P(T_L \leq \theta \leq T_U) = 1 - \alpha$$

říkáme, že (T_L, T_U) je interval spolehlivosti $1-\alpha$ pro parametr θ.

Jsou-li T_1 a T_2 dva odhady parametru θ, říkáme, že T_1 je <u>lepší</u> než T_2, jestliže

$$\operatorname{var} T_1 < \operatorname{var} T_2 .$$

Posloupnost odhadů T_n parametru θ je <u>asymptoticky nestranná</u>, jestliže

$$\lim_{n\to\infty} E(T_n) = \theta .$$

<u>Příklad</u> $X_1, X_2, .. X_n$ je náhodný výběr z rozdělení se střední hodnotou μ a rozptylem σ^2. Statistiky

$$M = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad L = \frac{1}{2}(X_1 + X_n)$$

jsou nestrannými odhady střední hodnoty μ. M je lepší neboť

$$\operatorname{var} M = \frac{\sigma^2}{n}, \quad \operatorname{var} L = \frac{\sigma^2}{2}$$

<u>Příklad</u> Výběrová směrodatná odchylka

S není nestranným odhadem směrodatné odchylky σ. Kdyby $E(S)=\sigma$, pak by

$$\operatorname{var} S = E(S^2)-(E(S))^2=\sigma^2-\sigma^2=0,$$

což znamená, že rozptyl S je nulový, tedy S by byla konstantní, což není pravda.

<u>Příklad</u> Statistika $s_n^2=\frac{1}{n}\sum_{1}^{n}(X_i-M)^2$ není nestranným odhadem σ^2. Je totiž

$$E(s_n^2)=E\left(\frac{n-1}{n}S^2\right)=\frac{n-1}{n}\sigma^2.$$

s_n^2 je ale asymptoticky ~~odhadem~~ nestranným odhadem σ^2.

<u>Intervaly spolehlivosti</u>

- oboustranný
- levostranný a pravostranný

<u>Konstrukce intervalů spolehlivosti</u>

(1) Zvolíme statistiku V, která je nestranným bodovým odhadem parametru Θ.

(2) Najdeme tzv. pivotovou statistiku W, která je transformací V se známým

rozdělením, které nezávisí na θ.
(např. $N(0,1)$, ~~kvad~~ $\chi^2(n)$, $t(n)$).

(3) Najdeme příslušné kvantily statistiky W tak, že

$$P\left(w_{\frac{\alpha}{2}} \le W \le w_{1-\alpha}\right) = 1-\alpha$$

(4) Nerovnost $w_{\frac{\alpha}{2}} \le W \le w_{1-\alpha}$ převedeme ekvivalentními úpravami na nerovnost

$$T_V \le \theta \le T_H$$

(5) Z daného výběru zjistíme konkrétní číselné realizace statistik T_L, T_H a dostaneme intervalový odhad parametrické statistiky spolehlivosti $1-\alpha$.

Intervaly spolehlivosti pro parametry normálního rozdělení

μ (známe σ^2) $\quad \left(M - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}},\ M + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right)$

μ (neznáme σ^2) $\quad \left(M - \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1),\ M + \frac{S}{\sqrt{n}} t_{1-\alpha}(n-1)\right)$

σ^2 (neznáme μ) $$\left(\frac{(n-1)\,S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} , \frac{(n-1)\,S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \right)$$

σ^2 (známe μ) $$\left(\frac{\sum_1^n (X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)} , \frac{\sum_1^n (X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)} \right)$$

<u>Ukázka odvození</u> μ hledáme a σ^2 neznáme.

(1) Statistika pro μ je M.

(2) Pivotová statistika je v tomto případě

$$T = \frac{M-\mu}{S/\sqrt{n}}$$

(3) Nechť t je kvantilová funkce pro rozdělení $t(n-1)$. Hledáme

$$t_{\frac{\alpha}{2}}(n-1) \leq T \leq t_{1-\frac{\alpha}{2}}(n-1)$$

(4) V tomto případě je $t_{\frac{\alpha}{2}} = -t_{1-\frac{\alpha}{2}}$, tedy

$$-t_{1-\frac{\alpha}{2}}(n-1) \leq \frac{M-\mu}{S/\sqrt{n}} \leq t_{1-\frac{\alpha}{2}}(n-1)$$

$$\underbrace{M - \frac{S}{\sqrt{n}}\, t_{1-\frac{\alpha}{2}}(n-1)}_{\text{dolní odhad}} \leq \mu \leq \underbrace{M + \frac{S}{\sqrt{n}}\, t_{1-\frac{\alpha}{2}}(n-1)}_{\text{horní odhad}}$$

<u>Příklad</u> Náhodná veličina X má normální rozdělení $N(\mu, \sigma^2)$, kde μ a σ^2 nejsou známy. V následující tabulce jsou uvedeny četnosti realizací této náhodné veličiny

X_i	8	11	12	14	15	16	17	18	20	21
n_i	1	2	3	4	7	5	4	3	2	1

Tedy $n = \sum' n_i = 32$

Určete 99% interval spolehlivosti pro střední hodnotu μ.

Řešení

$$M = \frac{490}{32} \approx 15{,}3$$

Výběrový rozptyl je

$$S = \left(\sum' n_i \,(X_i - M)^2\right) / (n-1) \approx 7{,}6$$

Interval spolehlivosti 0,99 je

$$\left(M - \frac{S}{\sqrt{n}}\, t_{1-\frac{\alpha}{2}}(n-1),\ M + \frac{S}{\sqrt{n}}\, t_{1+\frac{\alpha}{2}}(n-1) \right)$$

$\alpha = 0{,}01$, ~~[illegible]~~ $t_{0{,}995}(31) \approx 2{,}745$

$$\left(15{,}3 - \frac{7{,}6}{\sqrt{32}}\, 2{,}745,\ 15{,}3 + \frac{7{,}6}{\sqrt{32}}\, 2{,}745 \right) \approx (14{,}0;\ 16{,}7)$$

<u>Příklad</u> Rozměry výrobků jsou náhodné veličiny s $N(10\,\mathrm{mm}, 0{,}0734\,\mathrm{mm}^2)$. Odběratel testuje 21 výrobků a pokud směrodatná odchylka nepřesahuje u 21 výrobků 0,2 mm, tak dodávku přijme. Zjistěte pravděpodobnost, že ji přijme.

<u>Řešení</u>

$$P(S \le 0{,}2) = P(S^2 \le (0{,}2)^2) =$$

$$= P\left(\frac{(n-1)S^2}{\sigma^2} \le \frac{0{,}04\cdot 20}{\sigma^2}\right) = \chi^2_{20}\left(\frac{0{,}04\cdot 20}{0{,}0734}\right)$$

$$\approx \cancel{0{,}05} = \chi^2_{20}(10{,}9) \approx 0{,}05.$$

Jak se změní pravděpodobnost přijetí dodávky, při testování pouze 4 výrobků

$$\chi^2_3\left(\frac{0{,}04\cdot \cancel{20}3}{0{,}0734}\right) = \chi^2_3(1{,}63) \approx 0{,}25.$$

<u>Příklad</u> Nechť $X_1, \dots, X_n$ je náhodný výběr z normálního rozdělení $N(\mu, 0{,}04)$. Určete nejmenší počet měření, aby šířka 95% intervalu spolehlivosti nepřesáhla 0,16.

<u>Řešení:</u> Šířka pásma pro spolehlivost $1-\alpha$ je

$$2\,\frac{\sigma}{n}\,z_{1-\frac{\alpha}{2}} = \frac{2\cdot 0{,}2}{\sqrt{n}}\,z(0{,}975) \le 0{,}16$$

Odtud ($z_{0{,}975} \approx 1{,}96$) dostáváme

$$n \ge 24{,}01 .$$

Je tedy potřeba udělat aspoň 25 pokusů. ∎