Statistický seminář

Krejčová, Vojtek

Sekvenční testy, základní pojmy a vlastnosti

4. března 2025

Klasický přístup

Neyman-Pearsonovo lemma.

Nechť $\Theta = \{\theta_0, \theta_1\}$ a P_0 a P_1 jsou pravděpodobnostní rozdělení odpovídající parametrům θ_0 a θ_1. Nechť P_j má hustotu p_j vůči σ-konečné míře μ, $j = 0, 1$.

1. Pro test $H_0 : \theta_X = \theta_0$ proti $H_0 : \theta_X = \theta_1$ a hladinu α existuje funkce ϕ a konstanta $k \geq 0$ tak, že

$$\mathbb{E}_{\theta_0} \varphi(X) = \alpha, \tag{1}$$

$$\varphi(x) = \begin{cases} 1, & \text{pro } p_1(x) \geq k\, p_0(x), \\ 0, & \text{pro } p_1(x) < k\, p_0(x). \end{cases} \tag{2}$$

2. Pokud ϕ splňuje (1) a (2), potom je to nejsilnější test H_0 proti H_1 na hladině α.
3. Pokud ϕ je nejsilnější testem H_0 proti H_1 na hladině α, potom pro nějaké $k \geq 0$ musí splňovat (2).

Waldův sekvenční přístup

Rozhodovací pravidlo.

- Přijmeme H_0 pokud $\prod_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)} \leq B$.
- Přijmeme H_1 pokud $\prod_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)} \geq A$.
- Pokračujeme ve výběru pokud $B < \prod_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)} < A$.

Čísla $B < A$ volíme tak, aby pravděpodobnosti chyb byly rovny předepsaným číslům.

Waldův sekvenční přístup

Rozhodovací pravidlo.

- Přijmeme H_0 pokud $\prod_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)} \leq B$.
- Přijmeme H_1 pokud $\prod_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)} \geq A$.
- Pokračujeme ve výběru pokud $B < \prod_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)} < A$.

Čísla $B < A$ volíme tak, aby pravděpodobnosti chyb byly rovny předepsaným číslům.

Aproximace A, B.

$$A^* = \frac{1-\beta}{\alpha}, \quad B^* = \frac{\beta}{1-\alpha}.$$

Obecné sekvenční testy

- Test s pevným rozsahem výběru
- Dvoustupňový test
- Useknutý sekvenční test
- Waldův sekvenční test

Vlastnosti obecných sekvenčních testů

Lemma 1.1.

Následující tři výroky jsou ekvivalentní:

1. Test skončí s pravděpodobností 1 pro $\forall\theta \in \Theta$
2. Platí
$$\lim_{n\to\infty} P_\theta(N > n) = 0 \quad \forall\theta \in \Theta$$
3. $P_S(\theta) + L_S(\theta) = 1 \quad \forall\theta \in \Theta$

Lemma 1.2.

Pro každý test S a libovolné $\theta', \theta'' \in \Theta$ splňují operační charakteristika $L_S(\theta)$ a silofunkce $P_S(\theta)$ následující vztahy

$$\mathbb{E}_{\theta'}(\exp\{Q_N(\theta', \theta'')\} \mid \text{přijímáme } H_0) = \frac{L_S(\theta'')}{L_S(\theta')},$$

$$\mathbb{E}_{\theta'}(\exp\{Q_N(\theta', \theta'')\} \mid \text{přijímáme } H_1) = \frac{P_S(\theta'')}{P_S(\theta')}.$$

Lemma 1.4.

Nechť $X_1, X2, \ldots$ jsou nezávislé stejně rozdělené náhodné veličiny s rozdělením závisejícím na parametru $\theta \in \Theta$.

1. Nechť pro statistiku $t(X_1)$ a rozsah výběru N platí $\mathbb{E}_\theta(|t(X_1)|) < \infty$ a $\mathbb{E}_\theta(N) < \infty$. Potom
$$\mathbb{E}_\theta(N)\, \mathbb{E}_\theta(t(X_1)) = \mathbb{E}_\theta\Big(\sum_{i=1}^{N} t(X_i)\Big).$$
2. Nechť navíc platí $\mathrm{var}_\theta(t(X_1)) < \infty$ a $\mathbb{E}_\theta(t(X_1)) = 0$. Potom
$$\mathbb{E}_\theta(N)\, \mathbb{E}_\theta(t^2(X_1)) = \mathbb{E}_\theta\Big(\big(\sum_{i=1}^{N} t(X_i)\big)^2\Big).$$

Věta 1.5.

Nechť $\mathbb{E}_{\theta'}(|Z_i(\theta', \theta'')|) < \infty$ a $\mathbb{E}_{\theta'}(N) < \infty$, kde $\theta', \ \theta'' \in \Theta$. Potom pro libovolný test S, který skončí s pravděpodobností 1, platí

$$\begin{aligned} -\mathbb{E}_{\theta'}(N)\,\mathbb{E}_{\theta'}(Z_1(\theta', \theta'')) \geq & L_S(\theta') \ \ln \frac{L_S(\theta')}{L_S(\theta'')} + \\ & + (1 - L_S(\theta')) \ \ln \frac{1 - L_S(\theta')}{1 - L_S(\theta'')}. \end{aligned}$$

Důsledek 1.6.

Nechť $\mathrm{E}_{\theta_i}(Z_1(\theta_0, \theta_1)) \neq 0$ a konečná pro $i = 0, 1$. Pak pro libovolný test **S** s $\mathrm{E}_{\theta_i}(N) < \infty$ pro $i = 0, 1$, pravděpodobnostmi chyb menší rovno α resp. β a $\beta < 1 - \alpha$ pro test $H_0 : \theta = \theta_0$ proti $H_1 : \theta = \theta_1$, pak platí

$$\mathrm{E}_{\theta_0}(N)\mathrm{E}_{\theta_0}(Z_1(\theta_0, \theta_1)) \geq (1 - \alpha) \ln(\tfrac{\beta}{1-\alpha}) + \alpha \ln(\tfrac{1-\beta}{\alpha}),$$

$$\mathrm{E}_{\theta_1}(N)\mathrm{E}_{\theta_1}(Z_1(\theta_0, \theta_1)) \geq \beta \ln(\tfrac{\beta}{1-\alpha}) + (1 - \beta) \ln(\tfrac{1-\beta}{\alpha}).$$

Kritérium optimality

Definice.

$\mathcal{C}(\alpha,\beta)$ třída všech testů pro $H_0 : \theta \in \Theta_0$ proti $H_1 : \theta \in \Theta_1$ splňujících:

1. test končí s pravděpodobností 1
2. $1 - L_S(\theta) \leq \alpha \quad \forall \theta \in \Theta_0$
3. $L_S(\theta) \leq \beta \quad \forall \theta \in \Theta_1$

Test $\mathbf{S}^* \in \mathcal{C}(\alpha,\beta)$ nazveme eficientnějším než $\mathbf{S} \in \mathcal{C}(\alpha,\beta)$ na množině $\Theta_{0,1} \subseteq \Theta$, jestliže

$$\mathrm{E}_\theta(N_{S^*}) < \mathrm{E}_\theta(N_S) \quad \forall \theta \in \Theta_{01}.$$

Test $\mathbf{S}^* \in \mathcal{C}(\alpha,\beta)$ nazveme stejnoměrně eficientním testem na množině $\Theta_{01} \subseteq \Theta$, jestliže

$$\inf_{S \in \mathcal{C}(\alpha,\beta)} \mathrm{E}_\theta(N_S) = \mathrm{E}_\theta(N_{S^*}) \quad \forall \theta \in \Theta_{01}.$$

Děkujeme za pozornost!