13. cvičení z NMSA202
(14. 5. 2015 & 18. 5. 2015)

Neymanova-Pearsonova věta, Test poměrem věrohodnosti, Regrese

1. Nechť $X_1, \ldots, X_n$ je náhodný výběr z alternativního rozdělení s parametrem $p \in (0, 1)$. Najděte pomocí Neymanovy-Pearsonovy věty kritický obor pro test hypotézy $H_0 : p = p_0$ proti alternativě $H_1 : p = p_1$, kde $p_0, p_1 \in (0, 1)$ jsou známé konstanty a $p_1 > p_0$.
2. Nechť $X_1, \ldots, X_n$ je náhodný výběr z exponenciálního rozdělení s parametrem $\lambda > 0$. Najděte pomocí Neymanovy-Pearsonovy věty kritický obor pro test hypotézy $H_0 : \lambda = \lambda_0$ proti alternativě $H_1 : \lambda = \lambda_1$, kde $\lambda_0, \lambda_1 > 0$ jsou známé konstanty a $\lambda_1 > \lambda_0$.
3. Nechť $X_1, \ldots, X_n$ je náhodný výběr z $\mathsf{N}(0, \sigma^2)$.
   (a) Najděte pomocí Neymanovy-Pearsonovy věty kritický obor pro test hypotézy $H_0 : \sigma = \sigma_0$ proti alternativě $H_1 : \sigma = \sigma_1$, kde $\sigma_0, \sigma_1 > 0$ jsou známé konstanty a $\sigma_1 < \sigma_0$.
   (b) Odvoďte kritický obor testu poměrem věrohodnosti pro test hypotézy $H_0 : \sigma = \sigma_0$ proti alternativě $H_1 : \sigma < \sigma_0$.
   (c) Odvoďte kritický obor testu poměrem věrohodnosti pro test hypotézy $H_0 : \sigma = \sigma_0$ proti alternativě $H_1 : \sigma \neq \sigma_0$.
4. Nechť naše pozorování $Y_1, \ldots, Y_n$ splňují model
$$Y_i = \beta_1 + \beta_2\, x_i + e_i, \qquad i = 1, \ldots, n,$$
kde $x_1, x_2, \ldots, x_n$ jsou známé konstanty, $e_1, \ldots, e_n$ jsou nezávislé stejně rozdělené náhodné veličiny s nulovou střední hodnotou a konečným rozptylem σ^2. Uvažujme následující odhady neznámých parametrů β_1 a β_2:
$$\widehat{\beta}_2 = \frac{\sum_{i=1}^n (x_i - \overline{x}_n) Y_i}{\sum_{i=1}^n (x_i - \overline{x}_n)^2}, \qquad \widehat{\beta}_1 = \overline{Y}_n - \widehat{\beta}_2\, \overline{x}_n,$$
kde $\overline{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$. Ukažte, že odhady $\widehat{\beta}_1$ a $\widehat{\beta}_2$ jsou nestranné a spočtěte rozptyl odhadu $\widehat{\beta}_2$.
5. Nechť $Y_1, Y_2, \ldots, Y_n$ jsou nezávislé náhodné veličiny a Y_i má normální rozdělení $\mathsf{N}(\beta_1 + \beta_2\, x_i, \sigma^2)$, kde $x_1, x_2, \ldots, x_n$ jsou známé konstanty a také σ^2 je známé. Odvoďte maximálně věrohodné odhady parametrů β_1 a β_2.

Opakování z přednášky

Neymanova-Pearsonova věta. Nechť $\mathbf{x} = (x_1, \ldots, x_n)$ a $p_0(\mathbf{x})$, $p_1(\mathbf{x})$ jsou nezáporné měřitelné funkce na $(\mathbb{R}^n, \mathcal{B}^n)$ takové, že

$$\int_{\mathbb{R}^n} p_0(\mathbf{x})\, d\nu_n(\mathbf{x}) = \int_{\mathbb{R}^n} p_1(\mathbf{x})\, d\nu_n(\mathbf{x}) = 1, \tag{14.1}$$

kde ν_n je σ-konečná míra na $(\mathbb{R}^n, \mathcal{B}^n)$. Nechť pro dané $\alpha \in (0, 1)$ existuje takové kladné číslo c, že pro množinu

$$W^* = \big\{\mathbf{x} \in \mathbb{R}^n : p_1(\mathbf{x}) \geq c\, p_0(\mathbf{x})\big\}$$

platí

$$\int_{W^*} p_0(\mathbf{x})\, d\nu_n(\mathbf{x}) = \alpha. \tag{14.2}$$

Pak pro libovolnou množinu $W \in \mathcal{B}^n$ splňující

$$\int_{W} p_0(\mathbf{x})\, d\nu_n(\mathbf{x}) \leq \alpha \tag{14.3}$$

platí

$$\int_{W^*} p_1(\mathbf{x})\, d\nu_n(\mathbf{x}) \geq \int_{W} p_1(\mathbf{x})\, d\nu_n(\mathbf{x}). \tag{14.4}$$

Aplikace v testování hypotéz. Nechť $X_1, \ldots, X_n$ je náhodný výběr z rozdělení s hustotou $f(x, \theta)$ vzhledem k nějaké σ-konečné míře ν. Chceme testovat hypotézu $H_0 : \theta = \theta_0$ proti alternativě $H_1 : \theta = \theta_1$, kde $\theta_1 \neq \theta_0$. Potom test s kritickým oborem

$$W^* = \left\{(x_1, \ldots, x_n) : \prod_{i=1}^{n} f(x_i, \theta_1) \geq c \prod_{i=1}^{n} f(x_i, \theta_0)\right\}, \tag{14.5}$$

kde

$$\int_{W^*} \prod_{i=1}^{n} f(x_i, \theta_0)\, d\nu(x_1) \ldots d\nu(x_n) = \alpha \tag{14.6}$$

má nejmenší pravděpodobnost chyby 2. druhu mezi všemi testy na hladině α.

Test poměrem věrohodnosti. Nechť náhodný vektor $\mathbf{X} = (X_1, \ldots, X_n)^\mathsf{T}$ pochází z rozdělení s hustotou $p(\mathbf{x}; \theta)$, kde $\theta \in \Theta$. Chceme testovat $H_0 : \theta \in \Theta_0$ proti alternativě $H_1 : \theta \in \Theta_1$, kde Θ_0 a Θ_1 jsou disjunktní podmnožiny Θ. Test poměrem čerpá inspiraci v Neymanově-Pearsonově větě a je daný kritický oborem ve tvaru

$$W^{**} = \left\{\mathbf{x} \in \mathbb{R}^n ; \frac{\sup_{\theta \in \Theta_1} p(\mathbf{x}; \theta)}{\sup_{\theta \in \Theta_0} p(\mathbf{x}; \theta)} \geq c\right\},$$

kde c je vhodná konstanta, aby test dodržoval hladinu.

Pokud $c \geq 1$ a $\Theta = \Theta_0 \cup \Theta_1$, pak se dá kritický obor W^{**} psát ve tvaru:

$$W^{**} = \left\{\mathbf{x} \in \mathbb{R}^n ; \frac{\sup_{\theta \in \Theta} p(\mathbf{x}; \theta)}{\sup_{\theta \in \Theta_0} p(\mathbf{x}; \theta)} \geq c\right\},$$

resp.

$$W^{**} = \left\{\mathbf{x} \in \mathbb{R}^n ; \frac{p(\mathbf{x}; \widehat{\theta})}{p(\mathbf{x}; \widehat{\theta}_0)} \geq c\right\},$$

kde $\widehat{\theta}$ je maximálně věrohodný odhad za předpokladu, že $\theta \in \Theta$ a $\widehat{\theta}_0$ je maximálně věrohodný odhad za platnosti H_0, tj. za předpokladu že $\theta \in \Theta_0$.

Výsledky

1. $\sum_{i=1}^n X_i \geq K$, , kde K je největší celé číslo, které splňuje

$$\mathsf{P}_{p_0}\big(\sum_{i=1}^n X_i \geq K\big) = \mathsf{P}\big(\mathrm{Bi}(n, p_0) \geq K\big) \leq \alpha.$$

2. $\sum_{i=1}^n X_i \leq \frac{\chi^2_{2n,\alpha}}{2\,\lambda_0}$
3. (a) $\sum_{i=1}^n X_i^2 \leq \sigma_0^2\,\chi^2_{n,\alpha}$, (b) $\sum_{i=1}^n X_i^2 \leq \sigma_0^2\,\chi^2_{n,\alpha}$, (c) $\left(\frac{n}{\sum_{i=1}^n X_i^2}\right)^{n/2} \exp\left\{\frac{\sum_{i=1}^n X_i^2}{2\,\sigma_0^2}\right\} \geq K$
4. $\mathsf{var}\left(\widehat{\beta}_2\right) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x}_n)^2}$.
5. $\widehat{\beta}_2 = \frac{\sum_{i=1}^n (x_i - \overline{x}_n) Y_i}{\sum_{i=1}^n (x_i - \overline{x}_n)^2}$, $\qquad \widehat{\beta}_1 = \overline{Y}_n - \widehat{\beta}_2\,\overline{x}_n$.