NMSA202 – PST & MAT STAT – KAPITOLY 7–9

Dk V Bodové vlastnosti ECDF,

(i) $E\left[\hat{F}_n(x)\right] = E\left[\frac{1}{n}\sum_{i=1}^{n} \underbrace{1\{X_i \le x\}}\right] = \frac{1}{n}\sum_{i=1}^{n} P(X_i \le x) = P(X_1 \le x) = F(x)$

(ii) $\text{var}\left[\hat{F}_n(x)\right] = \text{var}\left[\quad -\|- \quad\right] \overset{\text{IID} \Leftarrow X_i \text{ IID}}{=} \frac{1}{n^2}\sum_{i=1}^{n} \text{var}\underbrace{\left[1\{X_1 \le x\}\right]}_{\sim \text{Alt}(F(x))} = \frac{1}{n}F(x)\left[1-F(x)\right]$

(iii) $\text{bias}\left[\hat{F}_n(x)\right] = 0 \Rightarrow \text{MSE}\left(\hat{F}_n(x)\right) = \text{var}\left[\hat{F}_n(x)\right] \to 0, \; n \to \infty$

(iv) $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{1\{X_i \le x\}}_{\text{IID}} \xrightarrow[n\to\infty]{P} E\left[1\{X_1 \le x\}\right] = P(X_1 \le x) = F(x)$ ■

WLLN

Dk V Plug-in (empirický) odhad pro Lineární statistický funkcionál

Pro $\omega \in \Omega$ máme z definice emp. odhadu Lin. stat. fcionálu

$$T(\hat{P}_n)(\omega) = \int_{\mathbb{R}} r(x)\, d\hat{P}_n(\omega)(x) = \int_{\mathbb{R}} r(x)\, d\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}(x)\right) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} \int_{\mathbb{R}} r(x)\, d\delta_{X_i(\omega)}(x) = \frac{1}{n}\sum_{i=1}^{n} r\left(X_i(\omega)\right). \quad ■$$

Dk V MLE invariance,

Pro věrohodnostní fci parametru τ z definice plyne

$\tilde{L}_n(\tau) = \prod_{i=1}^{n} \tilde{f}(X_i;\tau)$, kde $\tilde{f}$ je hustota X_i parametrizovaná vzhledem k parametru τ.

Definujme

$L_n^g(\tau) := \sup_{\theta:\, g(\theta)=\tau} L_n(\theta)$, kde $L_n(\theta) = \prod_{i=1}^{n} f(X_i;\theta)$.

Potom pro MLE $\hat{\tau}_n$ platí

$$\tilde{L}_n(\hat{\tau}_n) = \max_{\tau} L_n^g(\tau) = \max_{\tau} \sup_{\theta:\, g(\theta)=\tau} L_n(\theta) =$$

$$= \sup_{\theta} L_n(\theta) = L_n(\hat{\theta}) \ge L_n(\theta) = \tilde{L}_n(\tau). \quad ■$$

Dk L E & Var skóre,

$$E_\theta[S(X;\theta)] = \int S(x;\theta) f(x;\theta)\,dx = \int_{M(X)} \frac{\partial \log f(x;\theta)}{\partial\theta} f(x;\theta)\,dx = \int_{M(X)} \frac{\frac{\partial f(x;\theta)}{\partial\theta}}{f(x;\theta)} f(x;\theta)\,dx =$$

$$= \int_{M(X)} \frac{\partial f(x;\theta)}{\partial\theta}\,dx = \frac{\partial}{\partial\theta}\int_{M(X)} f(x;\theta)\,dx = \frac{\partial}{\partial\theta} 1 = 0,$$ kde $M(X)$ je nosič X a $\int f = \int_{M(X)} f = 1$. ■

Dk L Alternativní definice informace

$$\frac{\partial^2 \log f(x;\theta)}{\partial\theta^T\partial\theta} = \frac{\partial}{\partial\theta^T}\left(\frac{\partial f(x;\theta)}{\partial\theta}\,\frac{1}{f(x;\theta)}\right) = \frac{1}{f(x;\theta)}\,\frac{\partial^2 f(x;\theta)}{\partial\theta^T\partial\theta} -$$

$$- \frac{1}{f^2(x;\theta)}\,\frac{\partial f(x;\theta)}{\partial\theta}\left[\frac{\partial f(x;\theta)}{\partial\theta}\right]^T$$

Potom $$-E_\theta\left[\frac{\partial^2 \log f(x;\theta)}{\partial\theta^T\partial\theta}\right] = E_\theta\Big[\underbrace{\frac{\partial \log f(x;\theta)}{\partial\theta}}_{S(X;\theta)}\Big]^{\otimes 2} - \underbrace{\int_{-\infty}^{+\infty}\frac{\partial^2 f(x;\theta)}{\partial\theta^T\partial\theta}\,dx}_{=\frac{\partial^2}{\partial\theta^T\partial\theta}\underbrace{\int_{M(x)} f(x;\theta)\,dx}_{=1} = 0} =$$

$$= \operatorname{var}_\theta S(X;\theta) = I(\theta).\ \blacksquare$$

Dk D Limitní informace

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underbrace{S(X_i;\theta)}_{\text{IID náh. vektory},\ E_\theta[\cdot]=\underset{\sim}{0},\ \operatorname{var}_\theta[\cdot]=I(\theta)} \xrightarrow[n\to\infty]{\mathcal{D}} N_d\left(\underset{\sim}{0}, I(\theta)\right) \text{ dle CLT pro IID.}$$

Dk V AN MLE

(Náznak) Rozvineme $\sum_{i=1}^{n} S(X_i;\hat\theta_n)$ Taylorovým rozvojem okolo θ:

$\underset{\sim}{0} = \sum_{i=1}^{n} S(X_i;\hat\theta_n) = \sum_{i=1}^{n} S(X_i;\theta) - \left[\frac{\partial}{\partial\theta^T}\sum_{i=1}^{n} S(X_i,\theta^*)\right](\hat\theta_n - \theta)$, kde θ^* leží mezi $\hat\theta_n$ a θ. Jelikož $\|\hat\theta_n - \theta\| \xrightarrow{P} 0 \Rightarrow \|\theta^* - \theta\| \xrightarrow{P} 0$.

Pak $$\sqrt{n}(\hat\theta_n - \theta) = \underbrace{\left[\frac{\partial}{\partial\theta^T}\sum_{i=1}^{n} S(X_i;\theta^*)\right]}_{\xrightarrow[n\to\infty]{P} I^{-1}(\theta)}\underbrace{\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S(X_i;\theta)\right\}}_{\xrightarrow[n\to\infty]{\mathcal{D}} \underset{\sim}{Z} \sim N_d(\underset{\sim}{0}, I(\theta))}$$

$\Uparrow$

Neboť $\theta^* \xrightarrow{P} \theta$, $\left[\frac{\partial}{\partial\theta^T}\sum_i S(X_i;\cdot)\right]$ je spojitá, $\left[\frac{\partial}{\partial\theta^T}\sum_i S(X_i;\theta_x)\right] \xrightarrow{P} I(\theta)$ a maticová inverze je spojitá fce.

Dle Slutského věty $\sqrt{n}(\hat\theta_n - \theta) \xrightarrow[n\to\infty]{\mathcal{D}} I^{-1}(\theta)\underset{\sim}{Z} \sim N_d\big(\underset{\sim}{0}, \underbrace{I^{-1}(\theta)I(\theta)I^{-1}(\theta)}_{=I^{-1}(\theta)}\big)$. $\blacksquare$

Dk V Asymptotická hladina Waldova testu

Za platnosti H_0 je $\theta = \theta_0$. Potom $(\hat\theta_n - \theta_0)/\hat{se}(\hat\theta_n) \xrightarrow[n\to\infty]{\mathcal{D}} N(0,1)$. Pak

$$\lim_{n\to\infty} P_{\theta_0}\left[|W| > u_{1-\alpha/2}\right] = \lim_{n\to\infty} P_{\theta_0}\left[\frac{|\hat\theta_n - \theta_0|}{\hat{se}(\hat\theta_n)} > u_{1-\frac{\alpha}{2}}\right] = \underset{Z\sim N(0,1)}{P}\left[|Z| > u_{1-\frac{\alpha}{2}}\right] = 1 - \left(1 - \frac{\alpha}{2} - \frac{\alpha}{2}\right) = \alpha.\ \blacksquare$$

Dk V Síla Waldova testu při alternativě

Za alternativy $H_1: \theta = \theta_1 \neq \theta_0$ máme $W \xrightarrow[n\to\infty]{\mathcal{D}} N\left(\frac{\theta_1-\theta_0}{\widehat{se}(\hat\theta_n)}, 1\right)$.

Přibližná síla testu je pak

$$\lim_{n\to\infty} P_{\theta_1}\left[|W| > u_{1-\frac{\alpha}{2}}\right] = \lim_{n\to\infty} P_{\theta_1}\Big[W < \underbrace{-u_{1-\frac{\alpha}{2}}}_{u_{\frac{\alpha}{2}}}\Big] + \lim_{n\to\infty} P_{\theta_1}\left[W > u_{1-\frac{\alpha}{2}}\right] =$$

$$= \lim_{n\to\infty} P_{\theta_1}\left[W - \frac{\theta_1-\theta_0}{\widehat{se}(\hat\theta_n)} < u_{\frac{\alpha}{2}} - \frac{\theta_1-\theta_0}{\widehat{se}(\hat\theta_n)}\right] + 1 - \lim_{n\to\infty} P_{\theta_1}\left[W - \frac{\theta_1-\theta_0}{\widehat{se}(\hat\theta_n)} \le u_{1-\frac{\alpha}{2}} - \frac{\theta_1-\theta_0}{\widehat{se}(\hat\theta_n)}\right] =$$

$$= \lim_{n\to\infty} P\left[Z < -u_{1-\frac{\alpha}{2}} - \frac{\theta_1-\theta_0}{\widehat{se}(\hat\theta_n)}\right] + 1 - \lim_{n\to\infty} P\left[Z < u_{1-\frac{\alpha}{2}} - \frac{\theta_1-\theta_0}{\widehat{se}(\hat\theta_n)}\right] =$$

$$= \lim_{n\to\infty} \left\{1 - \Phi\left(\frac{\theta_0-\theta_1}{\widehat{se}(\hat\theta_n)} + u_{1-\frac{\alpha}{2}}\right) + \Phi\left(\frac{\theta_0-\theta_1}{\widehat{se}(\hat\theta_n)} - u_{1-\frac{\alpha}{2}}\right)\right\}, \text{ kde } Z \sim N(0,1).$$ ■

Dk V Ekvivalence (dualita) testování hypotéz a intervalů spolehlivosti

Zamítáme H_0 na hladině $\alpha \Longleftrightarrow \alpha = \lim_{n\to\infty} P_{\theta_0}\left[|W| > u_{1-\frac{\alpha}{2}}\right] =$

$$= 1 - \lim_{n\to\infty} P_{\theta_0}\left[-u_{1-\frac{\alpha}{2}} < W < u_{1-\frac{\alpha}{2}}\right] = 1 - \lim_{n\to\infty} P_{\theta_0}\left[-u_{1-\frac{\alpha}{2}} < \frac{\hat\theta_n-\theta_0}{\widehat{se}(\hat\theta_n)} < u_{1-\frac{\alpha}{2}}\right] =$$

$$= 1 - \underbrace{\lim_{n\to\infty} P_{\theta_0}\left[\hat\theta_n - u_{1-\frac{\alpha}{2}}\,\widehat{se}(\hat\theta_n) < \theta_0 < \hat\theta_n + u_{1-\frac{\alpha}{2}}\,\widehat{se}(\hat\theta_n)\right]} \Longleftrightarrow$$ ■

$$\Longleftrightarrow \rightsquigarrow = 1-\alpha.$$

Dk V Výpočet p-hodnoty

Pro pevné $\omega \in \Omega$ je $\boldsymbol{X}(\omega) = \boldsymbol{x}$. Z definice p-hodnoty a z předpokladu věty je p-hodnota $= \inf\{\alpha : \boldsymbol{x} \in R_\alpha\} \overset{(*)}{=} \inf\{\alpha : T(\boldsymbol{x}) > c_\alpha\}$.

Nechť F_θ je distribuční funkce $T(\boldsymbol{X})$. Pro test s hladinou α platí $\alpha = \sup_{\theta\in\Theta_0} P_\theta[T(\boldsymbol{X}) > c_\alpha] = 1 - \inf_{\theta\in\Theta_0} F_\theta(c_\alpha)$. Pak $\inf_{\theta\in\Theta_0} F_\theta(c_\alpha) = 1-\alpha$. Potom upravíme $(*)$: p-hodnota $\underset{\substack{\uparrow \\ F_\theta \text{ je neklesající}}}{=} \inf\left\{\alpha : \inf_{\theta\in\Theta_0} F_\theta(T(\boldsymbol{x})) > \inf_{\theta\in\Theta_0} F_\theta(c_\alpha)\right\} =$

$$= \inf\left\{\alpha : \inf_{\theta\in\Theta_0} F_\theta(T(\boldsymbol{x})) > 1-\alpha\right\} = \inf\left\{\alpha : \alpha > \sup_{\theta\in\Theta_0}\left[1 - F_\theta(T(\boldsymbol{x}))\right]\right\} =$$

$$= \inf\left\{\alpha : \alpha > \sup_{\theta\in\Theta_0} P_\theta\left[T(\boldsymbol{X}) > T(\boldsymbol{x})\right]\right\}.$$ ■

Dk V p-hodnota pro Waldův test.

Aplikujeme předchozí větu na $T(\underset{\sim}{X}) = W$ a $T(\underset{\sim}{x}) = w$.
Pak využijeme, že $W \xrightarrow[n\to\infty]{\mathcal{D}} N(0,1)$ za platnosti H_0. $\blacksquare$

Dk V test podílem věrohodnosti.

Bez dk. Idea je rozvinout logaritmickou věrohodnost $\log L_n(\theta)$ Taylorovým rozvojem 2. řádu okolo $\hat{\theta}_n$.