1. NÁHODNÝ VÝBĚR

1.2 Statistiky

$\mathcal{F} = \mathcal{L}^2$, $EX_i \equiv \mu$, $\operatorname{var} X_i \equiv \sigma^2$

V1.1 (vlastnosti průměru)

Nechť $\operatorname{var} X_i < \infty$. Pak platí: (i) $E\bar{X}_n = \mu$, $\operatorname{var}\bar{X}_n = \frac{\sigma^2}{n}$; (ii) $\bar{X}_n \xrightarrow{P} \mu$; (iii) $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{D}} N(0,\sigma^2)$.

Dk (i) $E\bar{X}_n = E\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum EX_i = \frac{n\mu}{n} = \mu$

$$\operatorname{var}\bar{X}_n = \operatorname{var}\frac{1}{n}\sum X_i = \frac{1}{n^2}\operatorname{var}\sum X_i \underset{\text{nezáv.}}{=} \frac{1}{n^2}\sum \operatorname{var} X_i = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

(ii) použijeme ZVČ – Tvrz. P7.9

(iii) použijeme CLV – Tvrz. P7.10 pro $k=1$ ∎

Pozn. $X_i \sim N(\mu,\sigma^2)$ $i=1,\dots,n$; $\underset{\sim}{X} = (X_1,\dots,X_n)^T$, z nezávislosti $X_1,\dots,X_n \Rightarrow \underset{\sim}{X} \sim N_n(\underset{\sim}{\mu}, \sigma^2 I)$, kde $\underset{\sim}{\mu} = (\mu,\dots,\mu)$ $\rightsquigarrow$ $\bar{X}_n = \underset{\sim}{c}^T \underset{\sim}{X}$, kde $\underset{\sim}{c} = (\frac{1}{n},\dots,\frac{1}{n})^T \Rightarrow \underbrace{\underset{\sim}{c}^T\underset{\sim}{X}}_{\bar{X}_n} \sim N(\underbrace{\underset{\sim}{c}^T\underset{\sim}{\mu}}_{\mu}, \underbrace{\underset{\sim}{c}^T\sigma^2 I \underset{\sim}{c}}_{\sigma^2/n})$

Pozn. $S_n^2 = \frac{1}{n-1}\sum(X_i - \bar{X}_n)^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum X_i^2 - \bar{X}_n^2\right)$

Dk $S_n^2 = \frac{1}{n-1}\sum\left(X_i^2 - 2X_i\frac{1}{n}\sum X_i + \frac{1}{n^2}(\sum X_i)^2\right) = \frac{1}{n-1}\left[\sum X_i^2 - \frac{2}{n}(\sum X_i)^2 + \frac{1}{n}(\sum X_i)^2\right] =$

$$= \frac{n}{n-1}\left[\frac{1}{n}\sum X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2\right] ∎$$

V1.2 (vlastnosti výb. rozptylu)

(i) $S_n^2 \xrightarrow{P} \sigma^2$; (ii) $ES_n^2 = \sigma^2$; (iii) $\mathcal{F} = \mathcal{L}^4 \Rightarrow \sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{\mathcal{D}} N(0, \sigma^4(\mu_2 - 1))$, kde $\mu_2 = E(X_i - EX_i)^4/\sigma^4$ je špičatost; (iv) $\mathcal{F} = \mathcal{L}^4 \Rightarrow \sqrt{n}\left[\binom{\bar{X}_n}{S_n^2} - \binom{\mu}{\sigma^2}\right] \xrightarrow{\mathcal{D}} N_2\left(\binom{0}{0}, \begin{pmatrix}\sigma^2 & \sigma^3\mu_1 \\ \sigma^3\mu_1 & \sigma^4(\mu_2-1)\end{pmatrix}\right)$

Dk (i) $S_n^2 = \underbrace{\frac{n}{n-1}}_{\to 1}\Big[\underbrace{\frac{1}{n}\sum X_i^2}_{\xrightarrow{P} EX_i^2} - \underbrace{\bar{X}_n^2}_{\xrightarrow{P} (EX_i)^2}\Big]$ $\longleftarrow$ $\bar{X}_n \xrightarrow{P} \mu \underset{\text{věta o spoj. transf. Tvrz. P7.3}}{\Rightarrow} \bar{X}_n^2 \xrightarrow{P} \mu^2$

$$S_n^2 \xrightarrow{P} EX_i^2 - (EX_i)^2 = \operatorname{var} X_i = \sigma^2$$

(ii) $ES_n^2 = \frac{n}{n-1}\left[\frac{1}{n}\sum EX_i^2 - E\bar{X}_n^2\right]$, $\frac{1}{n}\sum EX_i^2 = EX_i^2 = \sigma^2 + \mu^2$ nezáv.

$$E\bar{X}_n^2 = \frac{1}{n^2}E\left(\sum X_i\right)^2 = \frac{1}{n^2}E\sum_{i=1}^{n}\sum_{j=1}^{n} X_iX_j = \frac{1}{n^2}\left[\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n}\sum_{j\neq i} EX_iX_j\right] =$$

$$= \frac{1}{n^2}\left[n(\sigma^2+\mu^2) + n(n-1)\mu^2\right] = \frac{\sigma^2}{n} + \mu^2$$

$$ES_n^2 = \frac{n}{n-1}\left[\sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2\right] = \frac{n}{n-1}\,\frac{n-1}{n}\,\sigma^2 = \sigma^2$$

(iii) BÚNO $\mu = 0$ (jinak vezmeme $X_i - \mu$, čímž nezměníme S_n^2)

Definujeme $\underset{\sim}{Q}_i = \binom{X_i}{X_i^2}$ iid vektory $\Rightarrow E\underset{\sim}{Q}_i = \binom{0}{\sigma^2}$, $\operatorname{var}\underset{\sim}{Q}_i = \begin{pmatrix}\sigma^2 & EX_i^3 \\ EX_i^3 & EX_i^4 - \sigma^4\end{pmatrix} \equiv \Sigma_Q$, $\bar{\underset{\sim}{Q}}_n = \binom{\bar{X}_n}{\frac{1}{n}\sum X_i^2}$

Z CLV (P7.10): $\sqrt{n}\left(\bar{\underset{\sim}{Q}}_n - \binom{0}{\sigma^2}\right) \xrightarrow{\mathcal{D}} N_2(\underset{\sim}{0}, \Sigma_Q)$

Z Δ-metody (P7.11): $\sqrt{n}\left(g(\bar{Q}_n) - g\binom{0}{\sigma^2}\right) \xrightarrow{\mathcal{D}} N(0, D_g \Sigma_Q D_g^T)$, kde $D_g = \frac{\partial g(q)}{\partial q}\Big|_{q=\binom{0}{\sigma^2}}$

Volíme $g\binom{q_1}{q_2} = q_2 - q_1^2$

Máme $g(\bar{Q}_n) = \frac{1}{n}\sum X_i^2 - \bar{X}_n^2 = \frac{n-1}{n} S_n^2$, $g\binom{0}{\sigma^2} = \sigma^2$, $D_g = (-2q_1, 1)\Big|_{\binom{0}{\sigma^2}} = (0,1)$, $D_g \Sigma_Q D_g^T = EX_i^4$

Tudíž $\sqrt{n}\left((1-\frac{1}{n})S_n^2 - \sigma^2\right) \xrightarrow{D} N(0, \underbrace{EX_i^4 - \sigma^4}_{=\sigma^4(\gamma_2-1)})$. Jelikož $\sqrt{n}\frac{1}{n}S_n^2 = \underset{\downarrow\, 0}{\frac{1}{\sqrt{n}}} \underset{\downarrow P\, \sigma^2}{S_n^2} \xrightarrow{P} 0$, máme

$$\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{D} N(0, \sigma^4(\gamma_2 - 1))$$

(iv) Jako předchozí část, ale volíme $g\binom{q_1}{q_2} = \binom{q_1}{q_2 - q_1^2}$ a máme $D_g = \begin{pmatrix}1&0\\-2q_1&1\end{pmatrix}\Big|_{\binom{0}{\sigma^2}} = I_2$, $D_g\Sigma_Q D_g = \Sigma_Q$. Čili

$$\sqrt{n}\left(\binom{\bar X_n}{S_n^2} - \binom{\mu}{\sigma^2}\right) \xrightarrow{D} N_2\left(0, \underbrace{\begin{pmatrix}\sigma^2 & EX_i^3\\ EX_i^3 & EX_i^4-\sigma^4\end{pmatrix}}_{=\begin{pmatrix}\sigma^2 & \sigma^3\gamma_1\\ \sigma^3\gamma_1 & \sigma^4(\gamma_2-1)\end{pmatrix}}\right) \blacksquare$$

$X_i \sim N(\mu, \sigma^2)$

V1.3 (i) $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$; (ii) $\bar X_n$ a S_n^2 jsou nezávislé.

Dk (i) Použijeme větu P6.3, bod 3: $X \sim N_n(0,\Sigma)$, $A_{n\times n}$, $A\Sigma$ idemp. $\Rightarrow X^TAX \sim \chi^2_{\mathrm{tr}\, A\Sigma}$

Vezmeme-li tedy $Y = \left(\frac{X_1-\mu}{\sigma}, \dots, \frac{X_n-\mu}{\sigma}\right)^T$ a $A = I_n - \frac1n 1_n 1_n^T = \begin{pmatrix}1-\frac1n & & -\frac1n\\ &\ddots& \\ -\frac1n & & 1-\frac1n\end{pmatrix}$, $1 = \begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}_{n\times1}$

dostaneme $Y\sim N_n(0, I_n)$

$Y^TAY = Y^TY - \frac1n(Y^T1_n)(1_n^TY) = \sum\left(\frac{X_i-\mu}{\sigma}\right)^2 - \frac1n\left(\sum\frac{X_i-\mu}{\sigma}\right)^2 = \frac{n}{\sigma^2}\left\{\frac1n\sum(X_i-\mu)^2 - \left[\frac1n\sum(X_i-\mu)\right]^2\right\}$

$= \frac{n}{\sigma^2}\frac1n\sum[(X_i-\mu)-(\bar X_n-\mu)]^2 = \frac{(n-1)S_n^2}{\sigma^2}$

$A\Sigma \equiv A$ je idempotentní: $AA = (I_n - \frac1n 1_n1_n^T)(I_n - \frac1n1_n1_n^T) = I_n - \frac2n 1_n1_n^T +$

$+ \frac1n 1_n \underbrace{1_n^T\frac1n 1_n}_{=n} 1_n^T = I_n - \frac1n 1_n1_n^T = A$

a $\mathrm{tr}\,A\Sigma = \mathrm{tr}\, I - \frac1n \mathrm{tr}\, 1_n1_n^T = n-1 \Rightarrow$ Tudíž $\frac{(n-1)}{\sigma^2}S_n^2 \sim \chi^2_{n-1}$

Pozn. $\frac{\sum(X_i-\mu)^2}{\sigma^2}\sim\chi^2_n$; $(X_i - \bar X_n)$, $i=1,\dots,n$ už nejsou nezávislé

(ii) Máme $S_n^2 = Y^TAY$, $Y\sim N_n(0,I_n)$ a $\bar Y_n = \underbrace{\frac1n 1_n^T}_{b^T} Y = \frac{\bar X_n-\mu}{\sigma}$.

Jsou Y^TAY a b^TY nezávislé?

Existuje (bez Dk) matice $D_{n\times(n-1)}$ s hodností $(n-1)$ taková, že $A = DD^T$. Jelikož A je idempotentní, $DD^TDD^T = DD^T$ musí být $D^TD = I_{n-1}$.

Máme $S_n^2 = (Y^TD)(D^TY)$. Ukážeme, že D^TY a b^TY jsou nezávislé.

Jelikož Y je normální, stačí ukázat, že $\mathrm{cov}(b^TY, D^TY) = b^T \mathrm{var}\, Y\, D = b^TD \overset{?}{=} 0$.

Máme $b^TA = b^TDD^T = \frac1n 1_n^T - \frac1n \underbrace{1_n^T \frac1n 1_n}_{=1} 1_n^T = 0$. Je tedy $b^TDD^T = 0 \Rightarrow b^TD\underbrace{D^TD}_{I_{n-1}} = 0 \Rightarrow b^TD = 0$ ∎

V1.4 $T = \sqrt n \frac{\bar X_n - \mu}{S_n} \xrightarrow{D} N(0,1)$

Dk CLV: $\sqrt n(\bar X_n - \mu) \xrightarrow{D} N(0,\sigma^2) \Rightarrow \sqrt n\frac{\bar X_n-\mu}{\sigma}\xrightarrow{D} N(0,1)$; $\sqrt n \frac{\bar X_n-\mu}{S_n} = \frac{\sigma}{S_n}\sqrt n\frac{\bar X_n-\mu}{\sigma} \xrightarrow{D} N(0,1)$

$\frac{1}{S_n}\xrightarrow{P}\frac1\sigma$ věta o spoj. transf. P7.3 Sluckého věta P7.6 $\left(\downarrow P\; 1 \quad \xrightarrow{D} N(0,1)\right)$ ∎

V1.5 $T=\sqrt{n}\,\frac{\bar X_n-\mu}{S_n}\sim t_{n-1}$ $\quad [X_i\sim N(\mu,\sigma^2)]$

Dk

$$T=\frac{\overbrace{\sqrt{n}\,\frac{\bar X_n-\mu}{\sigma}}^{\sim N(0,1)}}{\sqrt{\underbrace{\frac{(n-1)S_n^2}{\sigma^2}}_{\chi^2_{n-1}}/(n-1)}}\sim t_{n-1}$$

+ nezáv. $\bar X_n$ a S_n^2 dle V1.3 (ii) ∎

V1.6 $\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}\sim F_{n-1,m-1}$

Dk máme $\frac{(n-1)S_X^2}{\sigma_X^2}\sim\chi^2_{n-1}$, $\frac{(m-1)S_Y^2}{\sigma_Y^2}\sim\chi^2_{m-1}$. S_X^2 a S_Y^2 jsou nezávislé (nezáv. výběry).

$$\frac{\frac{(n-1)S_X^2}{\sigma_X^2}/(n-1)}{\frac{(m-1)S_Y^2}{\sigma_Y^2}/(m-1)}\sim F_{n-1,m-1}$$ ∎

1.3 Uspořádaný náhodný výběr

V1.7 $r=(r_1,\dots,r_n)\in P_n$: $p(y_1,\dots,y_n)=\begin{cases} n!\,f(y_1)\dots f(y_n), & y_1<\dots<y_n\\ 0, & \text{jinak}\end{cases}$

Dk Má-li mít $X_{(\cdot)}=(X_{(1)},\dots,X_{(n)})^T$ hustotu p, musí platit:

$\forall B\in\mathcal{B}: P[X_{(\cdot)}\in B]=\int\dots\int_B p(y_1,\dots,y_n)\,dy_1\dots dy_n$

Počítejme: $P[X_{(\cdot)}\in B]=\sum_{r\in P_n}P[X_{(\cdot)}\in B, R=r]=$ (kde $R=(R_1,\dots,R_n)^T$)

$=\sum_{r\in P_n}\int\dots\int f(x_1)\dots f(x_n)\,1_B(x_{(\cdot)})\,1(R=r)\,dx_1\dots dx_n=$ (kde $y=x_{(\cdot)}$, $y_1<\dots<y_n$)

$=\int\dots\int\underbrace{\sum_{r\in P_n}f(y_{r_1})\dots f(y_{r_n})}_{=p(y)}1_B(y)\,dy_1\dots dy_n=\int\dots\int n!\,f(y_1)\dots f(y_n)\,1_B(y)\,dy_1\dots dy_n$

neboť jsou stejně rozdělené

V1.8 – distr. fce k-té poř. statistiky

Dk Označme $Z_i=1_{(-\infty,x\rangle}(X_i)=\begin{cases}1, & X_i\le x\\ 0, & X_i>x\end{cases}\Rightarrow Z_i\sim Alt(F(x))$

Počet veličin, které jsou $\le x$ je $\sum_{i=1}^n Z_i$ a má binomické rozdělení $Bi(n,F(x))$

$\rightsquigarrow P[Z_i=j]=\binom{n}{j}F^j(x)(1-F(x))^{n-j}$

Jev $X_{(k)}\le x$ nastane právě když k nebo $k+1, k+2,\dots,n-1$, nebo n veličin je $\le x$.

Tudíž $P[X_{(k)}\le x]=\sum_{j=k}^n P[\sum Z_i\le x]=\sum_{j=k}^n\binom{n}{j}F^j(x)(1-F(x))^{n-j}$

Druhý zápis: zpětnou indukcí

nejprve $k=n$: $\frac{1}{B(n,1)}\int_0^{F(x)}t^{n-1}\,dt=\frac{1}{B(n,1)}\frac{1}{n}F^n(x)=\binom{n}{n}F^n(x)$ ✓ (neboť $\frac{1}{nB(n,1)}=\frac{\Gamma(n+1)}{n\Gamma(n)\Gamma(1)}=1=\binom{n}{n}$)

nyní krok $k\to k-1$. Předpokládáme, že $\sum_{i=k}^n\binom{n}{i}F^i(x)(1-F(x))^{n-i}=\frac{1}{B(k,n-k+1)}\int_0^{F(x)}t^{k-1}(1-t)^{n-k}\,dt$

a ukážeme $\sum_{i=k-1}^n$ —||— $\overset{(*)}{=}\frac{1}{B(k-1,n-k+2)}\int_0^{F(x)}t^{k-2}(1-t)^{n-k+1}\,dt$

Počítáme per partes $\int t^{k-2}(1-t)^{n-k+1}dt = \frac{1}{k-1} t^{k-1}(1-t)^{n-k+1} + \frac{n-k+1}{k-1}\int t^{k-1}(1-t)^{n-k}dt$

Pravá strana (*) je $\underbrace{\frac{1}{B(k-1,n-k+2)(k-1)}}_{\frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+2)} = \frac{n!}{(k-1)!(n-k+1)!} = \binom{n}{k-1}} F^{k-1}(x)(1-F(x))^{n-k+1} + \underbrace{\frac{n-k+1}{B(k-1,n-k+2)(k-1)}}_{\frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} = \frac{1}{B(k,n-k+1)}}\int_0^{F(x)} t^{k-1}(1-t)^{n-k}dt$ ■

Důsledek $X_i \sim R(0,1) \rightsquigarrow F(x) = x,\ x\in\langle 0,1\rangle$, jinak $\begin{cases} 0, & x<0 \\ 1, & x>1 \end{cases}$

- $P[X_{(k)} \le x] = \frac{1}{B(k,n-k+1)}\int_0^x t^{k-1}(1-t)^{n-k}dt$
- $X_{(k)} \sim B(k,n-k+1)$ $\quad EX_{(k)} = \frac{k}{k+n-k+1} = \frac{k}{n+1}$; $\operatorname{var} X_{(k)} = \frac{k(n-k+1)}{(n+1)^2(n+2)}$

V1.9 Hustota k-té poř. stat. $\quad$ {připomenutí: $\frac{d}{dx}\int_0^x G(u)du = G(x)$}

Dk) Máme $\frac{dF_{(k)}(x)}{dx} = \underbrace{\frac{n!}{(k-1)!(n-k)!}}_{=k\binom{n}{k}} F^{k-1}(x)(1-F(x))^{n-k}\cdot \underbrace{F'(x)}_{=f(x)}$ ■

V1.10 $P[\mathbf{R} = \mathbf{r}] = \frac{1}{n!}$; $\mathbf{r} \in \mathcal{P}_n$

Dk) $P[\mathbf{R}=\mathbf{r}] = \int\dots\int \mathbb{1}(\mathbf{R}=\mathbf{r}) f(x_1)\dots f(x_n)\,dx_1\dots dx_n = \int\dots\int_{y_1<\dots<y_n} f(y_{r_1})\dots f(y_{r_n})\,dy_1\dots dy_n =$

$= \int\dots\int_{y_1<\dots<y_n} f(y_1)\dots f(y_n)\,dy_1\dots dy_n = P[\mathbf{R}=(1,\dots,n)^T]$, $\forall \mathbf{r}\in\mathcal{P}_n$ ■

V1.11 momenty R_i (R_j)

Dk) (i) BÚNO $i=n$: $P[R_n=k] = \sum_{\mathbf{r}\in\mathcal{P}_{n-1}} P[R_1=r_1, R_2=r_2, \dots, R_n=k] = \sum_{\mathbf{r}\in\mathcal{P}_{n-1}} \frac{1}{n!} = \frac{(n-1)!}{n!} = \frac{1}{n}$

(ii) BÚNO $i=n-1$, $j=n$: $P[R_{n-1}=k, R_n=m] = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$

(iii) $ER_i = \sum_{k=1}^n k\frac{1}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$; $\operatorname{var} R_i = \sum_{k=1}^n k^2\frac{1}{n} - \left(\frac{n+1}{2}\right)^2 = \frac{n(n+1)(2n+1)}{6n} - \left(\frac{n+1}{2}\right)^2 =$

$= \frac{n+1}{2}\,\frac{4n+2-3n-3}{6} = \frac{(n+1)(n-1)}{12}$

(iv) $\operatorname{cov}(R_i,R_j) = \sum\sum_{k\neq m} km\frac{1}{n(n-1)} - \left(\frac{n+1}{2}\right)^2 = \frac{n(n+1)^2}{4(n-1)} - \frac{(n+1)(2n+1)}{6(n-1)} - \frac{(n+1)^2}{4} =$

$\sum\sum_{k\neq m} km = \left(\sum_1^n k\right)^2 - \left(\sum_1^n k^2\right) = \left(\frac{n(n+1)}{2}\right)^2 - \frac{n(n+1)(2n+1)}{6}$

$= \frac{n+1}{12(n-1)}\left[3n(n+1) - 2(2n+1) - 3(n+1)(n-1)\right] = -\frac{n+1}{12}$ ■

2. ZÁKLADY TEORIE ODHADU

2.1 Bodový odhad

Př) (str. 10, č. 3) $\mathcal{F} = \{Po(\lambda), \lambda>0\}$; $f_X(x) = \frac{\lambda^x}{x!}e^{-\lambda}$, $x = 0,1,2,\dots$

a) $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{0\}}(X_i)$; $\mathbb{1}_{\{0\}}(X_i) \sim Alt(p)$, $p = P[X_i=0] = e^{-\lambda}$

průměr je nestranný a konzistentní odhad; $E\mathbb{1}_{\{0\}}(X_i) = e^{-\lambda}$

b) $\tilde{\theta}_n = \left(\frac{n-1}{n}\right)^{\sum X_i}$; víme, že $\sum_1^n X_i \sim Po(n\lambda)$ ($\in$ V P5.6 – konvoluce nezáv. Poiss.)

$E\tilde{\theta}_n = \sum_{k=0}^{\infty} \left(\frac{n-1}{n}\right)^k \frac{(n\lambda)^k}{k!} e^{-n\lambda} = e^{-n\lambda} \sum_0^{\infty} \frac{[(n-1)\lambda]^k}{k!} = e^{-n\lambda} e^{(n-1)\lambda} = e^{-\lambda}$

konzistence: $\log \tilde{\theta}_n = \sum X_i \log\left(1-\frac{1}{n}\right) = \underbrace{\frac{1}{n}\sum X_i}_{\xrightarrow{P} EX_i = \lambda} \log \underbrace{\left(1-\frac{1}{n}\right)^n}_{\to e^{-1}} \xrightarrow{P} -\lambda \Rightarrow \tilde{\theta}_n \xrightarrow{P} e^{-\lambda}$

věta o spoj. transf. Tvrz P7.3

Př] $\theta_X = e^{-2\lambda X}$ (str. 10, cv. 4)

$X \sim Po(\lambda)$, $\lambda > 0$; hledáme nestranný odhad $e^{-2\lambda}$, $n=1$; nestranný odhad λ je X

nechť $T(X)$ je n. veličina taková, že $ET(X) = e^{-2\lambda} \Rightarrow$ musí platit: $\sum_{x=0}^{\infty} T(x) \frac{\lambda^x}{x!} e^{-\lambda} = e^{-2\lambda}$

$\Updownarrow$

$\sum_{x=0}^{\infty} T(x) \frac{\lambda^x}{x!} = e^{-\lambda} = \sum_{x=0}^{\infty} \frac{(-\lambda)^x}{x!}$

→ aby toto platilo $\forall \lambda > 0$, muselo by být $T(X) = (-1)^X$ ⇝ tj. odhad $e^{-2\lambda}$ je $\begin{cases} +1 & \text{pokud } X \text{ je sudé} \\ -1 & \text{pokud } X \text{ je liché} \end{cases}$

Ale $e^{-2\lambda}$ nemůže být záporné! A $e^{-2\lambda} = 1$ může nastat pouze pro $\lambda = 0$, což není přípustná hodnota parametru. ⇝ Existuje jediný nestranný odhad, ale nedává smysl!

V2.1 $E\hat{\theta}_n \to \theta_0$, $\mathrm{var}\, \hat{\theta}_n \to 0 \Rightarrow \hat{\theta}_n \xrightarrow{P} \theta_0$

Dk] viz Tvrz P7.5 ■

Př] (str. 10, dole) $MSE_{S_n^2} > MSE_{\hat{\sigma}_n^2}$

nechť $Y^* = \sum_1^n (Y_i - \bar{Y}_n)^2$; $S_n^2 = \frac{1}{n-1} Y^*$, $\hat{\sigma}_n^2 = \frac{1}{n} Y^*$; $\frac{Y^*}{\sigma_0^2} \sim \chi^2_{n-1}$ dle V1.3(i), $EY^* = (n-1)\sigma_0^2$, $\mathrm{var}\, Y^* = 2(n-1)\sigma_0^2$

$MSE_{S_n^2} = E\left(\frac{1}{n-1} Y^* - \sigma_0^2\right)^2 = \mathrm{var} \frac{1}{n-1} Y^* = \frac{2\sigma_0^4}{n-1}$

$MSE_{\hat{\sigma}_n^2} = E\left(\frac{1}{n} Y^* - \sigma_0^2\right)^2 = \frac{1}{n^2} EY^{*2} - \frac{2}{n}\sigma_0^2 EY^* + \sigma_0^4 = \frac{2(n-1)}{n^2}\sigma_0^4 + \frac{(n-1)^2}{n^2}\sigma_0^4 - \frac{2(n-1)}{n}\sigma_0^4 + \sigma_0^4 =$

$= \frac{\sigma_0^4}{n^2}\left(2n - 2 + \cancel{n^2} - \cancel{2n} + 1 - \cancel{2n^2} + \cancel{2n} + \cancel{n^2}\right) = \frac{2n-1}{n^2}\sigma_0^4 < \frac{2n}{n^2}\sigma_0^4 < \frac{2}{n-1}\sigma_0^4$

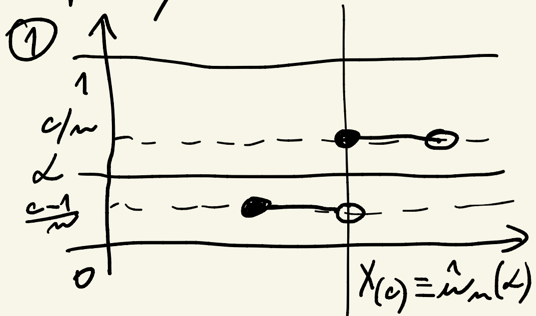
3. METODY PRO ODHADOVÁNÍ PARAMETRŮ

3.1 Empirické odhady a výběrové momenty

V3.1 Vlastnosti emp. distr. fce

Dk] (i)-(iv) triviální (v) bez dk

Empirický odhad kvantilu (str. 20) $\hat{F}_n^{-1}(\alpha) = \inf\{x : \hat{F}_n(x) \geq \alpha\} \equiv \hat{u}_n(\alpha)$



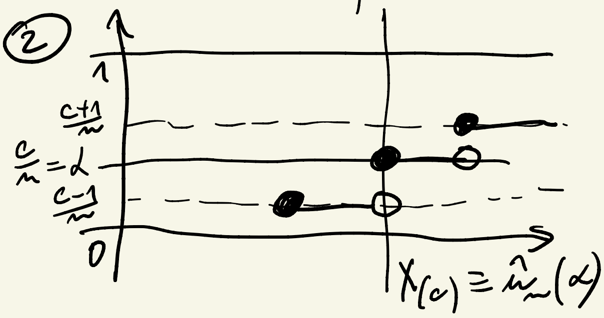
$(c-1)$ veličin je $< \hat{u}_n(\alpha)$

c-tá veličina je $> \hat{u}_n(\alpha)$

$n\alpha$ není celé

$\left.\begin{array}{l} \frac{c-1}{n} < \alpha \rightsquigarrow c-1 < n\alpha \\ \frac{c}{n} > \alpha \rightsquigarrow c > n\alpha \end{array}\right\} \Rightarrow c = [n\alpha]+1$

$\hat{u}_n(\alpha) = X_{([n\alpha]+1)}$



$(c-1)$ veličin $< \hat{u}_n(\alpha)$

c-tá veličina $= \hat{u}_n(\alpha)$

$n\alpha$ je celé $\Rightarrow c = n\alpha \rightsquigarrow \hat{u}_n(\alpha) = X_{([\alpha n])}$

V3.2 $X_1,\dots,X_n \sim F_X$, hustota f_X, F_X ostře rostoucí a spojitá

(i) $\hat{u}_n(\alpha)$ je konzistentní (ii) $\sqrt{n}\left(\hat{u}_n(\alpha) - u_X(\alpha)\right) \xrightarrow{\mathcal{D}} N(0, V(\alpha))$, $V(\alpha) = \dfrac{\alpha(1-\alpha)}{f_X^2(u_X(\alpha))}$

Dk: Nejprve pro $Y_i \sim R(0,1)$; $F_Y(x) = x$, $x \in (0,1)$; $u_Y(\alpha) = \alpha$, $\alpha \in (0,1)$

__Konzistence:__ $\hat{u}_n(\alpha) = Y_{(k_\alpha)}$, $k_\alpha = \begin{cases} [\alpha n]+1, & \alpha n \text{ není celé} \\ \alpha n, & \alpha n \text{ je celé} \end{cases}$ $\dfrac{k_\alpha}{n} \to \alpha$

Víme: $Y_{(k_\alpha)} \sim B(k_\alpha, n-k_\alpha+1)$ (dle V1.8) $\Rightarrow EY_{(k_\alpha)} = \dfrac{k_\alpha}{n+1} \to \alpha = u_Y(\alpha)$

& $\operatorname{var} Y_{(k_\alpha)} = \dfrac{k_\alpha(n-k_\alpha+1)}{(n+2)(n+1)^2} \to 0$ pro $n \to \infty$ $\Rightarrow$ dle V2.1: $Y_{(k_\alpha)} = \hat{u}_n(\alpha) \xrightarrow{P} \alpha = u_Y(\alpha)$

$\Rightarrow$ konzistence

__As. normalita:__ dle L3.3 (později): $\dfrac{\overbrace{\sum_{i=1}^{k} Z_i}^{U}}{\underbrace{\sum_{i=1}^{k} Z_i}_{U} + \underbrace{\sum_{i=k+1}^{n+1} Z_i}_{V}} \sim B(k, n-k+1)$, kde $Z_1,\dots,Z_{n+1} \overset{iid}{\sim} Exp(1)$

$\rightsquigarrow X_{(k_\alpha)}$ má toto rozdělení pro $k = k_\alpha \Rightarrow$ lze psát $Y_{(k_\alpha)} = \dfrac{U}{U+V}$

Z CLV: $\sqrt{k_\alpha}\left(\dfrac{1}{k_\alpha}U - \overset{EZ_i}{\overset{\parallel}{1}}\right) \xrightarrow{\mathcal{D}} N(0, \overset{\operatorname{var} Z_i}{\overset{\parallel}{1}}) \Rightarrow \sqrt{n}\sqrt{\dfrac{k_\alpha}{n}}\left(\dfrac{n}{k_\alpha}\dfrac{1}{n}U - 1\right) \xrightarrow{\mathcal{D}} N(0,1)$

a $\sqrt{n-k_\alpha+1}\left(\dfrac{1}{n-k_\alpha+1}V - 1\right) \xrightarrow{\mathcal{D}} N(0,1) \Rightarrow \sqrt{n}\underbrace{\sqrt{\dfrac{n-k_\alpha+1}{n}}}_{\to \sqrt{1-\alpha}}\left(\underbrace{\dfrac{n}{n-k_\alpha+1}}_{\to \frac{1}{1-\alpha}}\dfrac{1}{n}V - 1\right) \xrightarrow{\mathcal{D}} N(0,1)$ $\Bigg\} \Rightarrow$

$\sqrt{\alpha}$, $\dfrac{1}{\alpha}$; $n \to \infty$

U a V jsou nezávislé

$\Downarrow$

$\Rightarrow \sqrt{n}\left[\begin{pmatrix} \frac{1}{\alpha}\frac{U}{n} \\ \frac{1}{1-\alpha}\frac{V}{n} \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right] \xrightarrow{\mathcal{D}} N_2\left(\underset{\sim}{0}, \begin{pmatrix} \frac{1}{\alpha} & 0 \\ 0 & \frac{1}{1-\alpha} \end{pmatrix}\right) \rightsquigarrow$ Δ-metoda pro $g\binom{x}{y} = \dfrac{\alpha x}{\alpha x + (1-\alpha)y}$

Máme $g\begin{pmatrix} \frac{1}{\alpha}\frac{U}{n} \\ \frac{1}{1-\alpha}\frac{V}{n} \end{pmatrix} = \dfrac{U}{U+V} = Y_{(k_\alpha)}$, $g\binom{1}{1} = \alpha = u_Y(\alpha)$

$\dfrac{\partial g\binom{x}{y}}{\partial x} = \dfrac{\alpha}{\alpha x + (1-\alpha)y} - \dfrac{\alpha x \cdot \alpha}{(\alpha x + (1-\alpha)y)^2} \overset{x=1,\, y=1}{=} \alpha(1-\alpha)$, $\dfrac{\partial g\binom{x}{y}}{\partial y} = -\alpha x \dfrac{1-\alpha}{[\alpha x + (1-\alpha)y]^2} \overset{x=1,\, y=1}{=} -\alpha(1-\alpha)$

$\rightsquigarrow$ as. rozptyl po transformaci: $[\alpha(1-\alpha)]^2 (1,-1)\begin{pmatrix} \frac{1}{\alpha} & 0 \\ 0 & \frac{1}{1-\alpha} \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = [\alpha(1-\alpha)]^2\left(\dfrac{1}{\alpha} + \dfrac{1}{1-\alpha}\right) = \alpha(1-\alpha)$

Tedy $\sqrt{n}\left(Y_{(k_\alpha)} - \alpha\right) \xrightarrow{\mathcal{D}} N(0, \alpha(1-\alpha))$

Nyní X_i s rostoucí d. f. F_X a hustotou f_X: platí $X_i = F_X^{-1}(Y_i)$, kde $Y_i \sim R(0,1)$ a $X_{(k)} = F_X^{-1}(Y_{(k)})$

F_X^{-1} je spojitá funkce, $\dfrac{d}{du}F_X^{-1}(u) = \dfrac{1}{f_X(F_X^{-1}(u))}$

Víme: $Y_{(k_\alpha)} \xrightarrow{P} \alpha \Rightarrow \underbrace{F_X^{-1}(Y_{(k_\alpha)})}_{= X_{(k_\alpha)} = \hat{u}_n(\alpha)} \xrightarrow{P} \underbrace{F_X^{-1}(\alpha)}_{= u_X(\alpha)}$ z věty o spoj. transf.

a: $\sqrt{n}\left(Y_{(k_\alpha)} - \alpha\right) \xrightarrow{\mathcal{D}} N(0, \alpha(1-\alpha)) \underset{\Delta\text{-met.}}{\Longrightarrow} \sqrt{n}\left(\underbrace{F_X^{-1}(Y_{(k_\alpha)})}_{\hat{u}_n(\alpha)} - \underbrace{F_X^{-1}(\alpha)}_{u_X(\alpha)}\right) \xrightarrow{\mathcal{D}} N\left(0, \alpha(1-\alpha)\dfrac{1}{[f_X(\underbrace{F_X^{-1}(\alpha)}_{u_X(\alpha)})]^2}\right)$ $\blacksquare$

__Lemma 3.3__ $Z_i \overset{iid}{\sim} Exp(1)$, $U = \sum_{i=1}^{k} Z_i$, $V = \sum_{i=k+1}^{n+1} Z_i \Rightarrow \dfrac{U}{U+V} \sim B(k, n-k+1)$

Dk: $U \sim \Gamma(1,k)$ $\quad f_U(u) = \dfrac{1}{\Gamma(k)}u^{k-1}e^{-u}$, $u > 0$

$V \sim \Gamma(1, n-k+1)$ $\quad f_V(v) = \dfrac{1}{\Gamma(n-k+1)}v^{n-k}e^{-v}$, $v > 0$ $\Bigg\}$ nezáv.; transf. $h\binom{u}{v} = \begin{pmatrix} \frac{u}{u+v} \\ u+v \end{pmatrix} = \binom{y}{x} \Rightarrow$

$\Rightarrow u = yx$, $v = x(1-y)$, $x > 0$, $y \in (0,1)$

$\dfrac{\partial h^{-1}\binom{y}{x}}{\partial(y,x)} = \begin{pmatrix} x & y \\ -x & 1-y \end{pmatrix}$, $\det \cdot = x(1-y) + xy = x \rightsquigarrow f(y,x) = \dfrac{1}{\Gamma(k)\Gamma(n-k+1)}(yx)^{k-1}[(1-y)x]^{n-k}e^{-x}x \Rightarrow$

$\Rightarrow f(y) = \dfrac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)}y^{k-1}(1-y)^{n-k}$ $\quad \left[\int_0^\infty x^n e^{-x}\,dx = \Gamma(n+1)\right]$ $\blacksquare$

3.2 Momentová metoda

Př] $X_i \overset{iid}{\sim} Geo(p_X)$; $EX_i = \frac{1-p_X}{p_X} = \frac{1}{p_X} - 1$, $\operatorname{var} X_i = \frac{1-p_X}{p_X^2}$ $\rightsquigarrow p_X = \frac{1}{1+EX_i} \Rightarrow \hat{\theta}_n = \frac{1}{1+\bar X_n}$

As. rozdělení $\hat\theta_n$: $h(x) = \frac{1}{1+x}$, $h'(x) = -\frac{1}{(1+x)^2}$, $h'\left(\frac{1-p_X}{p_X}\right) = -p_X^2$

$\sqrt{n}\left(\bar X_n - \frac{1-p_X}{p_X}\right) \xrightarrow[CLV]{\mathcal{D}} N\left(0, \frac{1-p_X}{p_X^2}\right) \xRightarrow{\Delta\text{-met}} \sqrt{n}\left(h(\bar X_n) - h\left(\frac{1-p_X}{p_X}\right)\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{1-p_X}{p_X^2}\, p_X^4\right) \Rightarrow \sqrt{n}\left(\hat\theta_n - p_X\right) \xrightarrow{\mathcal{D}} N\left(0, p_X^2(1-p_X)\right)$

Př] $X_i \overset{iid}{\sim} R(0,\theta_X)$; $EX_i = \frac{\theta_X}{2}$, $\operatorname{var} X_i = \frac{\theta_X^2}{12}$ $\rightsquigarrow \theta_X = 2X_i \Rightarrow \hat\theta_n = 2\bar X_n$

As. rozdělení $\hat\theta_n$: $\sqrt{n}\left(\bar X_n - \frac{\theta_X}{2}\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{\theta_X^2}{12}\right) \Rightarrow \sqrt{n}\left(2\bar X_n - \theta_X\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{\theta_X^2}{3}\right)$

Př] $X_i \overset{iid}{\sim} \Gamma(a,p)$; $EX_i = \frac{p}{a}$, $\operatorname{var} X_i = \frac{p}{a^2}$ $\rightsquigarrow \operatorname{var} X_i = \frac{1}{a} EX_i \Rightarrow a = \frac{EX_i}{\operatorname{var} X_i}$, $p = \frac{(EX_i)^2}{\operatorname{var} X_i} \Rightarrow$

$\Rightarrow \hat a = \frac{\bar X_n}{S_n^2}$, $\hat p = \frac{(\bar X_n)^2}{S_n^2}$. Známe: $\gamma_1 = \frac{2}{\sqrt{p}}$, $\gamma_2 = \frac{6}{p} + 3$

As. rozdělení $\hat a$: $\sqrt{n}\left[\begin{pmatrix}\bar X_n \\ S_n^2\end{pmatrix} - \begin{pmatrix}p/a \\ p/a^2\end{pmatrix}\right] \xrightarrow{\mathcal{D}} N_2\left(0, \underbrace{\begin{pmatrix}\sigma^2 & \sigma^3\gamma_1 \\ \sigma^3\gamma_1 & \sigma^4(\gamma_2-1)\end{pmatrix}}_{V}\right)$ dle V1.2(iv); $V = \begin{pmatrix}\frac{p}{a^2} & \frac{p\sqrt{p}}{a^3}\frac{2}{\sqrt{p}} \\ \frac{p\sqrt{p}}{a^3}\frac{2}{\sqrt{p}} & \frac{p^2}{a^4}\left(\frac{6}{p}+2\right)\end{pmatrix}$

$h\binom{x}{y} = \frac{x}{y}$: $\frac{\partial h}{\partial x} = \frac{1}{y} \overset{y=p/a^2}{=} \frac{a^2}{p}$, $\frac{\partial h}{\partial y} = -\frac{x}{y^2} \overset{(a,p)}{=} -\frac{a^3}{p}$ $\Rightarrow D_h = \left(\frac{a^2}{p}, -\frac{a^3}{p}\right) = \frac{a^2}{p}(1,-a)$ ... atd.

Př] $X_i \overset{iid}{\sim} R(\theta_1,\theta_2)$; $\underset{\mu_X}{EX_i} = \frac{\theta_1+\theta_2}{2}$, $\underset{\sigma_X^2}{\operatorname{var} X_i} = \frac{(\theta_2-\theta_1)^2}{12}$ $\rightsquigarrow \theta_1 = 2\mu_X - \theta_2 \rightsquigarrow 12\sigma_X^2 = (2\theta_2 - 2\mu_X)^2 \rightsquigarrow (\theta_2 - \mu_X)^2 = 3\sigma_X^2 \rightsquigarrow$

$\rightsquigarrow \theta_2 - \mu_X = \pm\sqrt{3}\,\sigma_X$ (ale $\theta_2 > \mu_X$) $\rightsquigarrow \theta_2 = \mu_X + \sigma_X\sqrt{3} \rightsquigarrow \theta_1 = \mu_X - \sigma_X\sqrt{3}$

$\Rightarrow \hat\theta_1 = \bar X_n - \sqrt{3S_n^2}$, $\hat\theta_2 = \bar X_n + \sqrt{3S_n^2}$

Př] $X_i \overset{iid}{\sim} B(\alpha,\beta)$; $\underset{\mu_X}{EX_i} = \frac{\alpha}{\alpha+\beta}$, $\underset{\sigma_X^2}{\operatorname{var} X_i} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ $\rightsquigarrow \alpha+\beta = \frac{\alpha}{\mu_X}$, $\alpha\left(1-\frac{1}{\mu_X}\right) = -\beta$, $\alpha\left(\frac{1-\mu_X}{\mu_X}\right) = \beta$,

$\frac{\alpha\beta}{(\alpha+\beta)^2} = \mu_X(1-\mu_X)$, $(\alpha+\beta+1)\sigma_X^2 = \mu_X(1-\mu_X)$, $\left(\frac{\alpha}{\mu_X}+1\right) = \frac{\mu_X(1-\mu_X)}{\sigma_X^2}$ $\rightsquigarrow \alpha = \mu_X\left[\frac{\mu_X(1-\mu_X)}{\sigma_X^2} - 1\right]$

$\rightsquigarrow \beta = (1-\mu_X)\left[\frac{\mu_X(1-\mu_X)}{\sigma_X^2} - 1\right]$

$\Rightarrow \hat\alpha = \bar X_n\left[\frac{\bar X_n(1-\bar X_n)}{S_n^2} - 1\right]$, $\hat\beta = (1-\bar X_n)\left[\frac{\bar X_n(1-\bar X_n)}{S_n^2} - 1\right]$

3.3 Metoda maximální věrohodnosti

Tvrz 3.5 $\theta \neq \theta_X \Rightarrow P\left[\ell_n(\theta_X) > \ell_n(\theta)\right] \to 1$

Dk] $\frac{1}{n}\left[\ell_n(\theta_X) - \ell_n(\theta)\right] = \frac{1}{n}\sum_{i=1}^n \log\frac{f(X_i|\theta_X)}{f(X_i|\theta)} \xrightarrow[ZVČ]{P} E\log\frac{f(X_i|\theta_X)}{f(X_i|\theta)} \equiv K(\theta_X,\theta)$

Ukážeme, že $K(\theta_X,\theta) > 0$: $K(\theta_X,\theta) = E\left[-\log\frac{f(X_i|\theta)}{f(X_i|\theta_X)}\right] \underset{\text{Jensen}}{>} -\log E\frac{f(X_i|\theta)}{f(X_i|\theta_X)} =$

$= -\log\int\frac{f(x|\theta)}{f(x|\theta_X)} f(x|\theta_X)\,d\mu(x) = -\log\underbrace{\int f(x|\theta)\,\mathbb{1}_{S_{\theta_X}}(x)\,d\mu(x)}_{\leq 1} \geq -\log 1 = 0$

$\Rightarrow \frac{1}{n}\left[\ell_n(\theta_X) - \ell_n(\theta)\right] \xrightarrow{P} K(\theta_X,\theta) > 0 \rightsquigarrow P\left[K(\theta_X,\theta) - \varepsilon < \frac{1}{n}\left[\ell_n(\theta_X) - \ell_n(\theta)\right] < K(\theta_X,\theta) + \varepsilon\right] \equiv p_n$

ε tak, aby $K(\theta_X,\theta) - \varepsilon > 0$

$\Rightarrow P\left[0 < \frac{1}{n}\left[\ell_n(\theta_X) - \ell_n(\theta)\right]\right] \geq p_n \to 1$ ∎

Př] $X_1,\dots,X_n \overset{iid}{\sim} Exp(\lambda_X)$, $\mathcal{F} = \{Exp(\lambda), \lambda > 0\}$; $\Theta = (0,\infty)$; $f(x;\lambda) = \lambda e^{-\lambda x}\mathbb{1}_{(0,\infty)}(x)$; $EX_i = \frac{1}{\lambda_X}$

$L_n(\lambda) = \prod_{i=1}^n f(X_i;\lambda) = \lambda^n \exp\left\{-\lambda\sum_{i=1}^n X_i\right\}\prod_{i=1}^n \mathbb{1}_{(0,\infty)}(X_i) \Rightarrow \ell_n(\lambda) = n\log\lambda - \lambda\sum X_i \to \max$

$\left[U(\lambda,X_i) = \frac{\partial\log f(X_i;\lambda)}{\partial\lambda} = \frac{\partial}{\partial\lambda}\left(\log\lambda - \lambda X_i\right) = \frac{1}{\lambda} - X_i\right]$

$U_n(\lambda,X) = \frac{n}{\lambda} - \sum X_i \Rightarrow$ věroh. rovnice: $\frac{n}{\hat\lambda_n} - \sum X_i = 0 \Rightarrow \hat\lambda_n = \left(\frac{1}{n}\sum X_i\right)^{-1}$

$-\frac{\partial U_n(\lambda,X)}{\partial\lambda} = \frac{n}{\lambda^2} > 0$ $\forall\lambda\in(0,\infty)$, ℓ_n je konkávní $\rightsquigarrow$ maximum je $\hat\lambda_n$

Př] $X_1,\dots,X_n \overset{iid}{\sim} Alt(p_X)$, $\mathcal{F}=\{Alt(p), p\in(0,1)\}$, $\Theta=(0,1)$, $f(x;p)=p^x(1-p)^{1-x}$, $x\in\{0,1\}$

$L_n(p)=\prod_{i=1}^n p^{X_i}(1-p)^{1-X_i} = p^{\sum X_i}(1-p)^{n-\sum X_i} \rightsquigarrow \ell_n(p) = (\sum X_i)\log p + (n-\sum X_i)\log(1-p)$

$\left[U(p,X_i)=\frac{\partial}{\partial p}[X_i\log p + (1-X_i)\log(1-p)] = \frac{X_i}{p} - \frac{1-X_i}{p}\right]$

$U_n(p,X)=\frac1p\sum_{i=1}^n X_i - \frac{1}{1-p}(n-\sum_{i=1}^n X_i) \rightsquigarrow$ věroh. rovnice: $\frac{1}{\hat p}\sum X_i = \frac{1}{1-\hat p}(n-\sum X_i)$

$(1-\hat p)\sum X_i = n\hat p - \hat p\sum X_i \Rightarrow \hat p = \frac1n\sum X_i$

$\lim_{p\to0_+}\ell_n(p)=-\infty$, $\lim_{p\to1_-}\ell_n(p)=-\infty$, ℓ_n je spojitá, věroh. rovnice má právě 1 řešení $\Rightarrow$

$\Rightarrow$ je to globální maximum

Př] $X_1,\dots,X_n \overset{iid}{\sim} N(\mu_X,\sigma_X^2)$, $\mathcal F=\{N(\mu,\sigma^2),\mu\in\mathbb R,\sigma^2>0\}$, $\theta=\binom{\mu}{\sigma^2}$, $\Theta=\mathbb R\times(0,\infty)$, $f(x;\theta)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$

$L_n(\theta)=\frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}}\exp\{-\frac{1}{2\sigma^2}\sum(X_i-\mu)^2\} \rightsquigarrow \ell_n(\theta) = -\frac n2\log\sigma^2 - \frac{1}{2\sigma^2}\sum(X_i-\mu)^2 - \frac n2\log 2\pi$

$\frac{\partial\ell_n(\theta)}{\partial\mu}=\frac{1}{2\sigma^2}2\sum(X_i-\mu)$, $\frac{\partial \ell_n(\theta)}{\partial\sigma^2} = -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum(X_i-\mu)^2 \rightsquigarrow U_n(\theta,X)=\begin{pmatrix}\frac{1}{\sigma^2}\sum(X_i-\mu)\\ -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum(X_i-\mu)^2\end{pmatrix}$

Věroh. rovnice: $\frac{1}{\hat\sigma_n^2}\sum(X_i-\hat\mu_n)=0 \Rightarrow \sum X_i = n\hat\mu_n \Rightarrow \hat\mu_n=\bar X_n$

$-\frac{n}{2\hat\sigma_n^2}+\frac{1}{2\hat\sigma_n^4}\sum(X_i-\hat\mu_n)^2=0 \Rightarrow n\hat\sigma_n^2=\sum(X_i-\bar X_n)^2 \Rightarrow \hat\sigma_n^2=\frac1n\sum_{i=1}^n(X_i-\bar X_n)^2$

Př] $X_1,\dots,X_n \overset{iid}{\sim} \Gamma(a_X,p_X)$, $\mathcal F=\{\Gamma(a,p), a,p>0\}$, $\theta=\binom{a}{p}$, $\Theta=(0,\infty)\times(0,\infty)$, $f(x;\theta)=\frac{a^p}{\Gamma(p)}x^{p-1}e^{-ax}$, $x>0$

$L_n(\theta)=\frac{a^{np}}{\Gamma^n(p)}(\prod_i X_i)^{p-1}\exp\{-a\sum_i X_i\} \rightsquigarrow \ell_n(\theta)=np\log a - n\log\Gamma(p)+(p-1)\sum_i\log X_i - a\sum_i X_i$

$\frac{\partial\ell_n}{\partial a}=n\frac pa-\sum X_i$, $\frac{\partial\ell_n}{\partial p}=n\log a - n\frac{\Gamma'(p)}{\Gamma(p)}+\sum\log X_i \rightsquigarrow U_n(\theta,X)=\begin{pmatrix}\frac{np}{a}-\sum X_i\\ n\log a - n\frac{\Gamma'(p)}{\Gamma(p)}+\sum\log X_i\end{pmatrix}$

Věroh. rovnice: $\frac{n\hat p}{\hat a}=\sum X_i\Rightarrow \hat a=\frac{\hat p}{\bar X_n}$

$n\log\hat a - n\frac{\Gamma'(\hat p)}{\Gamma(\hat p)}+\sum\log X_i=0 \Rightarrow \frac1n\sum\log X_i - \log\bar X_n = \frac{\Gamma'(\hat p)}{\Gamma(\hat p)}-\log\hat p$ ... numericky

V3.6 – konzistence MLE

Dk] (náznak) Rozvineme $U_n(\hat\theta_n)$ okolo θ_X Taylorovým rozvojem:

$0=U_n(\hat\theta_n)=U_n(\theta_X)-nI_n(\theta^*)(\hat\theta_n-\theta_X)$, kde θ^* leží mezi $\hat\theta_n$ a θ_X

Tedy $\hat\theta_n-\theta_X=[I_n(\theta^*)]^{-1}\underbrace{\frac1n U(\theta_X)}_{\xrightarrow{P}0 \text{ dle V3.7}}$ & Ukáže se, že $[I_n(\theta^*)]^{-1}$ je omezená s pstí $\to 1$.

$\rightsquigarrow$ Musí tedy platit $\hat\theta_n-\theta_X\xrightarrow{P}0$. ∎

V3.7 – vlastnosti skóre

Dk] (i) $EU(\theta_X|X_i)=\int\frac{\partial\log f(x|\theta_X)}{\partial\theta_X}f(x|\theta_X)d\mu(x)=\int\frac{\partial f(x|\theta_X)}{\partial\theta_X}\frac{1}{f(x|\theta_X)}f(x|\theta_X)d\mu(x)=0$ dle $\overset{(R6)}{(3.1)}$

Počítejme: $\frac{\partial^2\log f(x|\theta_X)}{\partial\theta_X\partial\theta_X^T}=\frac{\partial}{\partial\theta_X^T}\left(\frac{\partial f(x|\theta_X)}{\partial\theta_X}\frac{1}{f(x|\theta_X)}\right)=\frac{1}{f(x|\theta_X)}\frac{\partial^2 f(x|\theta_X)}{\partial\theta_X\partial\theta_X^T}-\frac{1}{f^2(x|\theta_X)}\frac{\partial f(x|\theta_X)}{\partial\theta_X}\left(\frac{\partial f(x|\theta_X)}{\partial\theta_X}\right)^T$

Takže $I(\theta_X)=-E\frac{\partial^2\log f(X_i|\theta_X)}{\partial\theta_X\partial\theta_X^T}=E\Big[\underbrace{\frac{\partial\log f(X_i|\theta_X)}{\partial\theta_X}}_{U(\theta_X|X_i)}\Big]^{\otimes2}-\underbrace{\int_{-\infty}^{+\infty}\frac{\partial^2 f(x|\theta_X)}{\partial\theta_X\partial\theta_X^T}d\mu(x)}_{=0 \text{ dle (3.1) (R6)}}$

(ii) $\frac{1}{\sqrt n}U_n(\theta_X|X)=\frac1{\sqrt n}\underbrace{\sum_{i=1}^n U(\theta_X|X_i)}_{\text{iid vektory},\ E\cdot=0,\ var\cdot=I}\xrightarrow[CLV]{D}N_k(0,I(\theta_X))$ ∎

V3.8 – as. normalita MLE

Dk] (náznak) Rozvineme $U_n(\hat\theta_n)$ Taylorovým rozvojem okolo θ_X:

$0 = U_n(\hat{\theta}_n) = U_n(\theta_X) - n I_n(\theta^*)(\hat{\theta}_n - \theta_X)$, kde θ^* leží mezi $\hat{\theta}_n$ a θ_X; jelikož $\|\hat{\theta}_n - \theta_X\| \xrightarrow{P} 0 \Rightarrow \|\theta^* - \theta_X\| \xrightarrow{P} 0$

$$\sqrt{n}(\hat{\theta}_n - \theta_X) = \underbrace{[I_n(\theta^*)]^{-1}}_{\xrightarrow{P} I^{-1}(\theta_X)} \underbrace{\frac{1}{\sqrt{n}} U_n(\theta_X)}_{\xrightarrow{D} Z \sim N_p(0, I(\theta_X))}$$

↑ neboť $\theta^* \xrightarrow{P} \theta_X$, I_n je spojitá, $I_n(\theta_X) \xrightarrow{P} I(\theta_X)$ a maticová inverze je spojitá funkce. Dle Sluckého věty $\sqrt{n}(\hat{\theta}_n - \theta_X) \xrightarrow{D} I^{-1}(\theta_X) \cdot Z \sim N_p(0, \underbrace{I^{-1} I I^{-1}}_{I^{-1}(\theta_X)})$ ∎

V3.9 – as. rozd. věroh. poměru

Dk: Rozvineme $\ell_n(\theta_X)$ Taylorovým rozvojem okolo $\hat{\theta}_n$:

$$\ell_n(\theta_X) = \ell_n(\hat{\theta}_n) + \underbrace{U_n^T(\hat{\theta}_n)}_{=0}(\theta_X - \hat{\theta}_n) - \frac{1}{2}(\theta_X - \hat{\theta}_n)^T n I_n(\theta^*)(\theta_X - \hat{\theta}_n)$$

$$\leadsto 2[\ell_n(\theta_X) - \ell_n(\hat{\theta}_n)] = \underbrace{\sqrt{n}(\hat{\theta}_n - \theta_X)^T}_{\xrightarrow{D} Z} \underbrace{I_n(\theta^*)}_{\downarrow P \; I(\theta_X)} \underbrace{\sqrt{n}(\hat{\theta}_n - \theta_X)}_{\xrightarrow{D} Z \sim N_p(0, I^{-1}(\theta_X))}$$

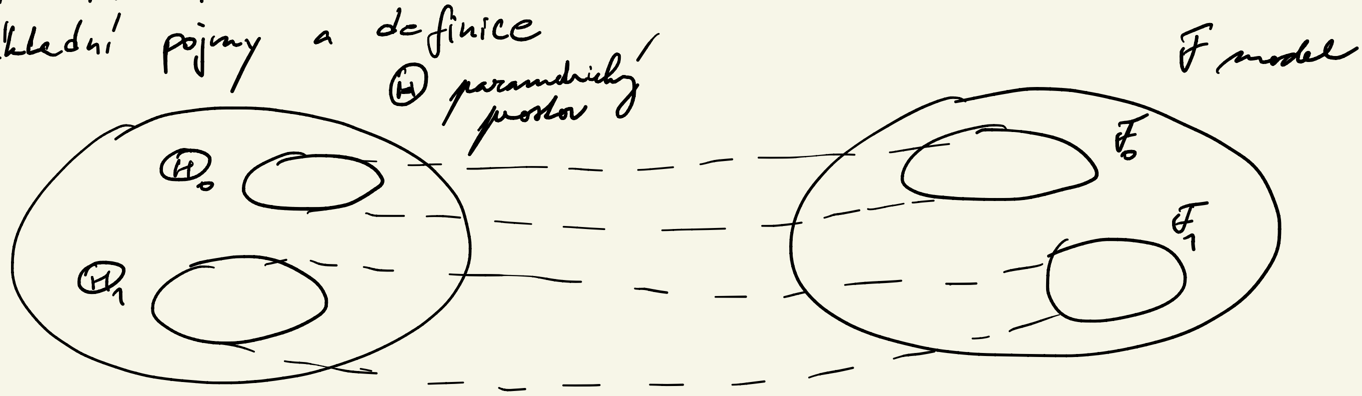
$\Rightarrow Z^T I(\theta_X) Z \sim \chi^2_p$ dle VP6.3 (ii) ∎

V3.10 – transformace parametrů

Dk: viz Δ metoda ∎

4. PRINCIPY TESTOVÁNÍ HYPOTÉZ

4.1 Základní pojmy a definice



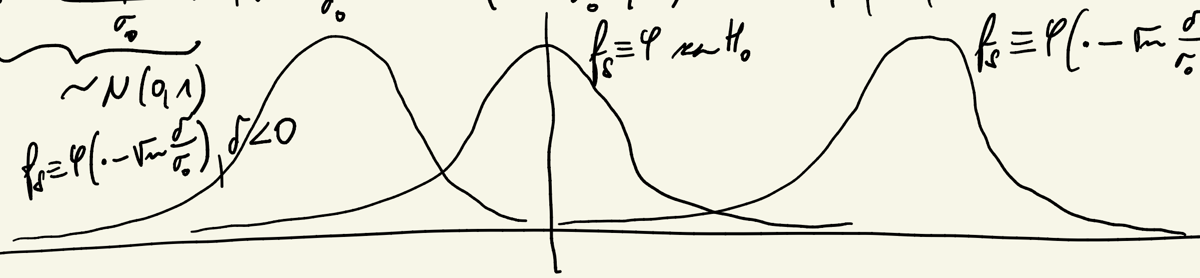
	H_0 nezamítneme $S(X) \notin C$	H_0 zamítneme $S(X) \in C$
H_0 platí	OK	chyba 1. druhu (pst. omezená hladinou)
H_1 platí	chyba 2. druhu (pst. není pod kontrolou)	OK

Př: $\mathcal{F} = \{N(\mu, \sigma_0^2), \mu \in \mathbb{R}, \sigma_0^2 > 0 \text{ známé}\}$, $X_1, \dots, X_n \overset{iid}{\sim} N(\mu_X, \sigma_0^2)$; $H_0: \mu_X = \mu_0$, $H_1: \mu_X \neq \mu_0$

Odvození testové statistiky: $\bar{X}_n$ je dobrý odhad μ_X, $\bar{X}_n \sim N(\mu_X, \frac{\sigma_0^2}{n})$ dle V1.1,

$S(X) = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma_0} \sim N(0,1)$ za platnosti H_0. Co když je $\mu_X = \mu_0 + \delta$? $\leadsto$

$$\leadsto S(X) = \underbrace{\sqrt{n} \frac{\bar{X}_n - \mu_0 - \delta}{\sigma_0}}_{\sim N(0,1)} + \sqrt{n} \frac{\delta}{\sigma_0} \sim N\left(\sqrt{n} \frac{\delta}{\sigma_0}, 1\right) \text{ za } H_1: \mu_X = \mu_0 + \delta.$$



Rozdělení $S(X)$ se mění podle toho, zda platí H_0 nebo ne.

Kritický obor: $P_{H_0}[S(\underset{\sim}{X})\in C]=\alpha$ ... ho lze volit mnoha způsoby

↳ volíme tak, aby f_S za H_0 byla v C malá, aby f_S za H_1 byla v C velká

↳ uvažujme různé alternativy $\Rightarrow$ dej $C=(-\infty,c_L)\cup(c_u,\infty)$

$$P_{H_0}[S(\underset{\sim}{X})\in C]=\underbrace{P[S(\underset{\sim}{X})<c_L]}_{\alpha/2}+\underbrace{P[S(\underset{\sim}{X})>c_u]}_{\alpha/2}=\alpha \quad\rightsquigarrow\quad c_L=\Phi^{-1}\left(\tfrac{\alpha}{2}\right)=u_{\alpha/2}=-u_{1-\frac{\alpha}{2}},\quad c_u=\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)=u_{1-\frac{\alpha}{2}}$$

(např.)

Zamítej $H_0 \iff S(\underset{\sim}{X})<-u_{1-\frac{\alpha}{2}}$ nebo $S(\underset{\sim}{X})>u_{1-\frac{\alpha}{2}} \iff |S(\underset{\sim}{X})|>u_{1-\frac{\alpha}{2}}$

$\frac{\sqrt{n}}{\sigma_0}|\bar{X}_n-\mu_0|>u_{1-\frac{\alpha}{2}}$ ... zamítáme pokud $\bar{X}_n$ je příliš daleko od μ_0

Síla proti alternativě $H_1:\mu_X=\mu_0+\delta,\ \delta>0 \rightsquigarrow S(\underset{\sim}{X})\sim N\left(\sqrt{n}\frac{\delta}{\sigma_0},1\right)\rightsquigarrow$

$$\rightsquigarrow P_{H_1}[S(\underset{\sim}{X})\in C]=\underbrace{P_{H_1}[S(\underset{\sim}{X})<-u_{1-\frac{\alpha}{2}}]}_{\approx 0}+\underbrace{P_{H_1}[S(\underset{\sim}{X})>u_{1-\frac{\alpha}{2}}]}_{=P_{H_1}\left[S(\underset{\sim}{X})-\sqrt{n}\frac{\delta}{\sigma_0}>u_{1-\frac{\alpha}{2}}-\sqrt{n}\frac{\delta}{\sigma_0}\right]}\doteq 1-\Phi\left(u_{1-\frac{\alpha}{2}}-\sqrt{n}\frac{\delta}{\sigma_0}\right)$$

Jak zvolit n, aby síla proti $\mu_X=\mu_0+\delta$ byla $\geq\beta$?

$$1-\Phi\left(u_{1-\frac{\alpha}{2}}-\sqrt{n}\frac{\delta}{\sigma_0}\right)\geq\beta \rightsquigarrow u_{1-\frac{\alpha}{2}}-\sqrt{n}\frac{\delta}{\sigma_0}\geq u_{1-\beta}\rightsquigarrow n\geq\left(u_{1-\frac{\alpha}{2}}+u_{\beta}\right)^2\frac{\sigma_0^2}{\delta^2}$$

Pomocí volby dostatečně velkého n můžeme zaručit, že síla proti danému rozdílu δ dosáhne požadované hodnoty (např. 0.95, 0.9, 0.8).

Tvrzení 4.1 zamítání na základě p-hodnoty

Dk (jednoduchá hypotéza, přesný test)

$C=(c_u,\infty)$

$c_u\equiv c_u(\alpha)=F_0^{-1}(1-\alpha)$

$p(\underset{\sim}{x})=1-F_0(s_{\underset{\sim}{x}})=P_0[S(\underset{\sim}{X})>s_{\underset{\sim}{x}}]\leq\alpha\iff s_{\underset{\sim}{x}}\geq c_u(\alpha)\iff s_{\underset{\sim}{x}}\in C$

$C=(-\infty,c_L)\cup(c_u,+\infty)$, $c_L=c_L(\alpha)=F_0^{-1}\left(\frac{\alpha}{2}\right)$, $c_u=c_u(\alpha)=F_0^{-1}\left(1-\frac{\alpha}{2}\right)$

Pokud $1-F_0(s_{\underset{\sim}{x}})<\frac{1}{2}$: $p(\underset{\sim}{x})=2\min\left(1-F_0(s_{\underset{\sim}{x}}),F_0(s_{\underset{\sim}{x}})\right)=2\left(1-F_0(s_{\underset{\sim}{x}})\right)\leq\alpha\iff 1-F_0(s_{\underset{\sim}{x}})\leq\frac{\alpha}{2}\iff s_{\underset{\sim}{x}}\geq c_u(\alpha)$

Podobně $1-F_0(s_{\underset{\sim}{x}})\geq\frac{1}{2}$: ... $p(\underset{\sim}{x})\leq\alpha\iff s_{\underset{\sim}{x}}\in C$

Tvrzení 4.2 ekvivalence int. spoleh. a testu

1. $P[c_L(\underset{\sim}{X})<\theta_X<c_u(\underset{\sim}{X})]=1-\alpha$

$$H_0:\theta_X=\theta_0\rightsquigarrow \underbrace{P[\theta_0<c_L(\underset{\sim}{X})]+P[\theta_0>c_u(\underset{\sim}{X})]}_{P[\text{zamítnutí platné } H_0]}=\alpha$$

2. $H_0:\theta_X=\theta^*$, $(S(\underset{\sim}{X},\theta^*),C)$

$$\underbrace{P_{\theta^*}[S(\underset{\sim}{X},\theta^*)\in C]=\alpha}$$

$$P[\theta_X\in\{\theta^*: S(\underset{\sim}{X},\theta^*)\notin C\}]=1-\alpha$$

4.5 As. testy založené na MLE

V4.3 (i) $2\log\lambda_n$ (ii) W_n (iii) R_n $\xrightarrow[H_0]{\mathcal{D}}\chi^2_d$

Dk (i) plyne z V3.9

(ii) V3.8: $\sqrt{n}(\hat{\theta}_n-\theta_X)\xrightarrow{\mathcal{D}}\underset{\sim}{Z}\sim N_d(\underset{\sim}{0},I^{-1}(\theta_X))\Rightarrow\sqrt{n}(\hat{\theta}_n-\theta_X)^T I(\theta_X)\sqrt{n}(\hat{\theta}_n-\theta_X)\xrightarrow{\mathcal{D}}\underset{\sim}{Z}^T I(\theta_X)\underset{\sim}{Z}\sim\chi^2_d$

věta o spoj. transformaci, V P6.3(ii)

$I_n(\hat{\theta}_n)\xrightarrow{P}I(\theta_X)$ (konzistence $\hat{\theta}_n$ + spoj. transf.)

Ukazuje $\theta_0 = \theta_X$: $n(\hat{\theta}_n - \theta_0)^T \hat{I}_n(\hat{\theta}_n)(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \chi^2_d$

(iii) V3.7: $\frac{1}{\sqrt{n}} U_n(\theta_X) \xrightarrow{\mathcal{D}} Z \sim N_d(0, I(\theta_X)) \Rightarrow \frac{1}{\sqrt{n}} U_n^T(\theta_X) I^{-1}(\theta_X) \frac{1}{\sqrt{n}} U_n(\theta_X) \xrightarrow{\mathcal{D}} Z^T I^{-1}(\theta_X) Z \sim \chi^2_d$

$\hat{I}_n(\theta_0) \xrightarrow{P} I(\theta_0) \rightsquigarrow \theta_X = \theta_0: \frac{1}{n} U_n^T(\theta_0) I^{-1}(\theta_0) U_n(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_d$

V4.5 (pouze (ii))

Dk (ii) $\sqrt{n}(\hat{\theta}_n - \theta_X) \xrightarrow{\mathcal{D}} Z \sim N_d(0, I^{-1}(\theta_X))$ dle V7.3

pouze m složek: $\sqrt{n}(\hat{\theta}_{An} - \theta_{AX}) \xrightarrow{\mathcal{D}} Z_A \sim N(0, I^{-1}_{AA\cdot B}) \Rightarrow n(\hat{\theta}_{An} - \theta_{AX})^T I_{AA\cdot B}(\hat{\theta}_{An} - \theta_{AX}) \xrightarrow{\mathcal{D}} \chi^2_m$

zatím je $\theta_{AX} = \theta_{A0}$

5.1 Kolmogorov – Smirnov

Tvrzení 5.1 $K_n^+ = \max_{1 \le i \le n} \left(\frac{i}{n} - F_0(X_{(i)})\right)$, $K_n^- = \max_{1 \le i \le n} \left(F_0(X_{(i)}) - \frac{i-1}{n}\right)$

Dk def. $X_{(0)} = -\infty$, $X_{(n+1)} \equiv +\infty \rightsquigarrow$ pro $x \in \langle X_{(i)}, X_{(i+1)})$ je $\hat{F}_n(x) = \frac{i}{n}$, $i = 0 \dots n$

máme $K_n^+ = \sup_{x \in \mathbb{R}} (\hat{F}_n(x) - F_0(x)) = \max_{0 \le i \le n} \sup_{X_{(i)} \le x \le X_{(i+1)}} \left[\frac{i}{n} - F_0(x)\right] = \max_{0 \le i \le n} \left[\frac{i}{n} - \inf_{X_{(i)} \le x \le X_{(i+1)}} F_0(x)\right] =$

$= \max_{0 \le i \le n} \left[\frac{i}{n} - F_0(X_{(i)})\right] = \max_{1 \le i \le n} \left[\frac{i}{n} - F_0(X_{(i)})\right]$; podobně pro K_n^- ∎

5.5 Jednovýběrový Wilcoxon

Tvrzení 5.4 $\mathcal{E} \perp 1$: $EW_s = \frac{n(n+1)}{4}$, $\operatorname{var} W_s = \frac{n(n+1)(2n+1)}{24}$

Dk nechť $\Delta_i = \begin{cases} 1 & \text{pokud } Z_i > 0 \\ -1 & \text{pokud } Z_i < 0 \end{cases} \Rightarrow E\Delta_i = 0$, $\operatorname{var} \Delta_i = E\Delta_i^2 = 1$, $P[\Delta_i = 1] = \frac{1}{2}$

A. ukažme, že $|Z_i|$ a Δ_i jsou nezávislé:

rozd. Z_i je symetrické kolem 0

$P[|Z_i| \le x, \Delta_i = 1] = P[0 \le Z_i \le x] = \frac{1}{2} P[-x \le Z_i \le x] = \frac{1}{2} P[|Z_i| \le x] = P[|Z_i| \le x] P[\Delta_i = 1]$

$\Rightarrow$ vektory $(|Z_1|, \dots, |Z_n|)$ a $(\Delta_1, \dots, \Delta_n)$ jsou nezávislé $\Rightarrow$ vektory $\underbrace{(R_1, \dots, R_n)}_{\text{funkce } |Z_1|, \dots, |Z_n|}$ a $(\Delta_1, \dots, \Delta_n)$ jsou nezávislé $\Rightarrow$

$\Rightarrow$ Jindy R_i a Δ_i jsou nezávislé.

B. máme $\sum_1^n R_i \mathbb{1}\{\Delta_i = 1\} + \sum_1^n R_i \mathbb{1}\{\Delta_i = -1\} = \sum_1^n R_i = \frac{n(n+1)}{2}$

$\underbrace{\sum_1^n R_i \mathbb{1}\{\Delta_i = 1\}}_{W_s} - \sum_1^n R_i \mathbb{1}\{\Delta_i = -1\} = \sum_1^n R_i \Delta_i$

$\Rightarrow 2W_s = \sum_1^n R_i \Delta_i + \frac{n(n+1)}{2} \Downarrow W_s = \frac{1}{2} \sum_1^n R_i \Delta_i + \frac{n(n+1)}{4}$

C. Počítejme: $EW_s = \frac{n(n+1)}{4} + \frac{1}{2} \sum_1^n ER_i\Delta_i$, $ER_i\Delta_i \overset{\text{nez. dle A.}}{=} ER_i E\Delta_i = 0$

$\operatorname{var} W_s = \frac{1}{4} \operatorname{var}(\sum R_i \Delta_i) = \frac{1}{4}\left[\sum_i \operatorname{var} R_i \Delta_i + \sum\sum_{i \ne j} \operatorname{cov}(R_i\Delta_i, R_j\Delta_j)\right] = (*)$

$\operatorname{var} R_i \Delta_i = E(R_i\Delta_i)^2 = ER_i^2 E\Delta_i^2 = \operatorname{var} R_i + (ER_i)^2 = \frac{n^2-1}{12} + \left(\frac{n+1}{2}\right)^2$ V1.11(iii)

$\operatorname{cov}(R_i\Delta_i, R_j\Delta_j) = E\Delta_i\Delta_j R_i R_j \overset{A}{=} \underbrace{E\Delta_i}_{=0} E\Delta_j ER_iR_j = 0$

$(*) = \frac{n}{4}\left[\frac{n^2-1}{12} + \frac{(n+1)^2}{4}\right] = \frac{n}{4} \frac{n^2 - 1 + 3n^2 + 6n + 3}{12} = \frac{n}{4} \frac{4n^2 + 6n + 2}{12} = \frac{n}{4} \frac{(n+1)(2n+1)}{6} = \frac{n(n+1)(2n+1)}{24}$

6. DVOUVÝBĚROVÉ PROBLÉMY NA SPOJITÁ DATA

6.2 Dvouvýběrový t-test

V6.2 $T = \sqrt{\frac{mn}{m+n}} \frac{\overline{X}_m - \overline{Y}_n - (\mu_X - \mu_Y)}{S_{m,n}} \sim t_{m+n-2}$

Dk
$$T = \frac{\dfrac{\overline{X}_m - \mu_X - (\overline{Y}_n - \mu_Y)}{\sqrt{\frac{1}{m} + \frac{1}{n}}\,\sigma}}{\sqrt{\dfrac{(m+n-2) S^2_{m,n}}{\sigma^2} \Big/ (m+n-2)}}$$

$\rightarrow$ čitatel: $\overline{X}_m - \overline{Y}_n \sim N\left(\mu_X - \mu_Y, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right)$ nezáv. $\Rightarrow \frac{\overline{X}_m - \mu_X - (\overline{Y}_n - \mu_Y)}{\sqrt{(\frac{1}{m} + \frac{1}{n})\sigma^2}} \sim N(0,1)$

$\rightarrow$ jmenovatel: $(m+n-2)\frac{S^2_{m,n}}{\sigma^2} = \underbrace{\frac{(m-1)S_X^2}{\sigma^2}}_{\sim \chi^2_{m-1}} + \underbrace{\frac{(n-1)S_Y^2}{\sigma^2}}_{\sim \chi^2_{n-1}} \sim \chi^2_{m+n-2}$

$(\bar X_n, S_X^2)$ nezáv. na $(\bar Y_m, S_Y^2)$; $\bar X_n$ a S_X^2 nezáv., $\bar Y_m$ a S_Y^2 nezáv. $\Rightarrow \bar X_n - \bar Y_m$ a $S^2_{n,m}$ jsou nezáv. $\blacksquare$

6.3 Dvouvýběrový z-test

V6.3 $$\frac{\bar X_n - \bar Y_m - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \xrightarrow{\mathcal{D}} N(0,1)$$

Dk Jelikož $S_X^2 \xrightarrow{P} \sigma_X^2$ a $S_Y^2 \xrightarrow{P} \sigma_Y^2$, stačí ukázat $\frac{\bar X_n - \bar Y_m - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \xrightarrow{\mathcal{D}} N(0,1)$

nechť $n = q_n m$, $q_n \to q \in (0,\infty)$, $n \to \infty$

$$\left.\begin{array}{l} \sqrt{n}(\bar X_n - \mu_X) \xrightarrow{\mathcal{D}} N(0, \sigma_X^2) \\ \sqrt{n}(\bar Y_m - \mu_Y) \xrightarrow{\mathcal{D}} N\left(0, \frac{\sigma_Y^2}{q}\right) \end{array}\right\} \text{nezávislé} \Rightarrow \sqrt{n}\begin{pmatrix} \bar X_n - \mu_X \\ \bar Y_m - \mu_Y \end{pmatrix} \xrightarrow{\mathcal{D}} N_2\left(\underset{\sim}{0}, \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \frac{\sigma_Y^2}{q} \end{pmatrix}\right)$$

$\sqrt{m}(\bar Y_m - \mu_Y) \xrightarrow{\mathcal{D}} N(0, \sigma_Y^2)$

$\sqrt{n q_n}(\bar Y_m - \mu_Y) \overset{\Downarrow}{\xrightarrow{\mathcal{D}}} N(0, \sigma_Y^2) \Rightarrow \sqrt{nq}(\bar Y_m - \mu_Y) \overset{\Uparrow}{\xrightarrow{\mathcal{D}}} N(0, \sigma_Y^2)$

$g(x_1, x_2) = x_1 - x_2$:

$\sqrt{n}\left(\bar X_n - \mu_X - (\bar Y_m - \mu_Y)\right) \xrightarrow{\mathcal{D}} N\left(0, \sigma_X^2 + \frac{\sigma_Y^2}{q}\right)$

$$\frac{\bar X_n - \bar Y_m - (\mu_X - \mu_Y)}{\sqrt{\frac{1}{n}\sigma_X^2 + \underbrace{\frac{1}{nq}}_{\approx \frac{1}{m}}\sigma_Y^2}} \xrightarrow{\mathcal{D}} N(0,1) \quad \blacksquare$$

6.4 Dvouvýběrový Wilcoxon

Tvrzení 6.4 $EW_{m,n} = \frac{n(m+n+1)}{2}$, $\operatorname{var} W_{m,n} = \frac{mn(m+n-1)}{12}$

Dk $EW_{m,n} = E\sum_{i=1}^{n} R_i = \sum_{i=1}^{n} ER_i = n\frac{m+n+1}{2}$ dle V1.11(iii) $\leadsto i = 1, \dots, N$: $ER_i = \frac{N+1}{2}$, $\operatorname{var} R_i = \frac{N^2-1}{12}$, $\operatorname{cov}(R_i, R_j) = -\frac{N+1}{12}$

$\operatorname{var} W_{m,n} = \operatorname{var}\sum_{i=1}^{n} R_i + \sum_{i \neq j}^{n}\sum^{n} \operatorname{cov}(R_i, R_j) = n\frac{(m+n)^2 - 1}{12} + (n^2 - n)\left(-\frac{m+n+1}{12}\right) = \frac{m+n+1}{12}\Big[\underbrace{n(m+n-1) - n^2 + n}_{mn}\Big]$ $\blacksquare$

Tvrzení 6.5 $W_{n,m} + W^*_{n,m} = nm + \frac{n(n+1)}{2}$; $EW^*_{n,m} = \frac{nm}{2}$, $\operatorname{var} W^*_{n,m} = \frac{mn(m+n-1)}{12}$; $\frac{1}{nm}W^*_{n,m} \xrightarrow{P} \frac{1}{2} = P[X_i < Y_j]$

Dk (a) $W_{n,m} + W^*_{n,m} = \sum_{i=1}^{n}\Big[\underbrace{\sum_{j=1}^{n} 1\{X_i \geq X_j\} + \sum_{j=1}^{m} 1\{X_i \geq Y_j\}}_{\equiv R_i}\Big] + \sum_{i=1}^{n}\sum_{j=1}^{m} 1\{X_i \leq Y_j\} =$

$>$ (dlky spojitosti) $\qquad <$

$= \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{n} 1\{X_i \geq X_j\}}_{= 1 + \dots + n = \frac{n(n+1)}{2}} + \underbrace{\sum_{i}^{n}\sum_{j}^{m} 1\{X_i > Y_j\} + \sum_{i}^{n}\sum_{j}^{m} 1\{X_i < Y_j\}}_{= nm \text{ ... všechny dvojice } (X_i, Y_j)}$

(b) $EW^*_{n,m} = nm + \frac{n(n+1)}{2} - \underbrace{EW_{n,m}}_{\frac{nm}{2} + \frac{n(n+1)}{2}} = \frac{nm}{2}$

(c) $$\left.\begin{array}{l} E\frac{1}{mn}W^*_{n,m} = \frac{1}{2} \\ \operatorname{var}\frac{1}{mn}W^*_{n,m} = \frac{1}{12}\left(\frac{1}{n} + \frac{1}{m} - \frac{1}{nm}\right) \to 0 \end{array}\right\} \Rightarrow \frac{1}{mn}W^*_{n,m} \xrightarrow{P} \frac{1}{2} \overset{H_0}{=} P[X_i < Y_j] \quad \blacksquare$$

7. JEDNOVÝBĚROVÉ PROBLÉMY PRO KATEGORIÁLNÍ DATA

7.1 Alt a Bi rozdělení

V7.2 $\hat\theta_n = \log\frac{\hat\mu_n}{1 - \hat\mu_n}$, $\theta_X = \log\frac{\mu_X}{1 - \mu_X}$

Dk (i) $\sqrt{n}(\hat\mu_n - \mu_X) \xrightarrow{\mathcal{D}} N(0, \mu_X(1 - \mu_X))$ ... V7.1(iv) $\leadsto g(x) = \log\frac{x}{1-x}$, $g'(x) = \frac{1}{x} + \frac{1}{1-x} = \frac{1}{x(1-x)}$

Δ-met: $\sqrt{n}(\hat\theta_n - \theta_X) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{\mu_X(1-\mu_X)}\right)$

(ii) $\sqrt{n}\sqrt{\mu_X(1-\mu_X)}(\hat\theta_n - \theta_X) \xrightarrow{\mathcal{D}} N(0,1) \Rightarrow$ Slucky: $\sqrt{n}\sqrt{\hat\mu_n(1 - \hat\mu_n)}(\hat\theta_n - \theta_X) \xrightarrow{\mathcal{D}} N(0,1) \Rightarrow \sqrt{\frac{X_n(n - X_n)}{n}}(\hat\theta_n - \theta_X) \xrightarrow{\mathcal{D}} N(0,1)$ $\blacksquare$

7.2 Multinomické rozdělení

V7.3 $\underset{\sim}{Y_i} \sim Mult_K(1, \underset{\sim}{\mu_X})$ nez. $\Rightarrow \underset{\sim}{X} = \sum_{i=1}^{n}\underset{\sim}{Y_i} \sim Mult_K(n, \underset{\sim}{\mu_X})$

Dk Hustota $\underset{\sim}{Y_i}$ jest $\mu_1^{x_1} \dots \mu_K^{x_K}$ pro $x_1, \dots, x_K \in \{0,1\}$, $\sum_{k=1}^{K}\mu_k = 1$. Indukce: $\underset{\sim}{Z} = \sum_{i=1}^{n-1}\underset{\sim}{Y_i} \sim Mult_K(n-1, \underset{\sim}{\mu_X}) \overset{②}{\Rightarrow} \underset{\sim}{Z} + \underset{\sim}{Y_n} \sim Mult_K(n, \underset{\sim}{\mu_X})$

$P[Z_1 + Y_{n1} = x_1, \dots, Z_K + Y_{nK} = x_K] = \sum_{\ell=1}^{K} P[\dots \mid Y_{n\ell} = 1]\underbrace{P[Y_{n\ell} = 1]}_{\mu_\ell} = \sum_{\ell=1}^{K}\frac{(n-1)!}{x_1! \dots (x_\ell - 1)! \dots x_K!}\mu_1^{x_1} \dots \mu_\ell^{x_\ell - 1} \dots \mu_K^{x_K} \cdot \mu_\ell =$

$$= \frac{(n-1)!}{(x_1 - 1)! \dots (x_K - 1)!}\mu_1^{x_1} \dots \mu_K^{x_K}\sum_{\ell=1}^{K}\frac{x_\ell}{\prod_{k=1}^{K} x_k} = \frac{n}{\prod x_k} \quad \blacksquare$$

V7.4 $\underset{\sim}{X} \sim Mult(n, \underset{\sim}{p_X})$: (i) $X_k \sim Bi(n, p_k)$ (ii) $EX_k = np_k$, $\operatorname{var} X_k = np_k(1-p_k)$ (iii) $\operatorname{cov}(X_j, X_k) = -np_j p_k$

(iv) $\operatorname{var} \underset{\sim}{X} = n \operatorname{diag} \sqrt{p_X} \left(I_K - \sqrt{p_X}^{\otimes 2}\right) \operatorname{diag} \sqrt{p_X}$

Dk (i) $\underset{\sim}{X} = \sum_{i=1}^{n} \underset{\sim}{Y_i}$ dle V7.3 ; $Z_{ik} = \mathbb{1}\{Y_{ik} = 1\} \underset{\text{nezáv.}}{\sim} Alt(p_k) \Rightarrow X_k = \sum_{i=1}^{n} Z_{ik} \sim Bi(n, p_k)$

(ii) plyne z (i)

(iii) $\operatorname{cov}(X_j, X_k) = \operatorname{cov}\left(\sum_{i=1}^{n} Z_{ij}, \sum_{l=1}^{n} Z_{lk}\right) = \sum_{i=1}^{n}\sum_{l=1}^{n} \operatorname{cov}(\underbrace{Z_{ij}, Z_{lk}}_{\text{nezáv. pokud } i \neq l}) = \sum_{i=1}^{n} \operatorname{cov}(Z_{ij}, Z_{ik}) = n \operatorname{cov}(Z_{ij}, Z_{ik}) =$

$= n\left[E Z_{ij} Z_{ik} - E Z_{ij} E Z_{ik}\right] = n\left(\underbrace{P[Z_{ij} = 1, Z_{ik} = 1]}_{=0} - P[Z_{ij} = 1] P[Z_{ik} = 1]\right) = -np_j p_k$

(iv) $n \operatorname{diag} \sqrt{p_X} \left(I_K - \sqrt{p_X}^{\otimes 2}\right) \operatorname{diag} \sqrt{p_X} = n\left[\begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_K \end{pmatrix} - \begin{pmatrix} p_1^2 & & p_j p_k \\ & \ddots & \\ p_j p_k & & p_K^2 \end{pmatrix}\right] = \operatorname{var} \underset{\sim}{X}$ ■

$\left(I_K - \sqrt{p_X}^{\otimes 2}\right)\left(I_K - \sqrt{p_X}^{\otimes 2}\right) = I_K - 2\sqrt{p_X}^{\otimes 2} + \sqrt{p_X} \underbrace{\sqrt{p_X}^T \sqrt{p_X}}_{= \sum p_k = 1} \sqrt{p_X}^T = I_K - \sqrt{p_X}^{\otimes 2}$

V7.5 $\underset{\sim}{X} \sim Mult_K(n, \underset{\sim}{p_X})$

(i) $\underset{\sim}{Z_n} = \frac{1}{\sqrt{n}} \operatorname{diag}\left(\sqrt{p_X}\right)^{-1}\left(\underset{\sim}{X} - n\underset{\sim}{p_X}\right) \xrightarrow{\mathcal{D}} N_K\left(\underset{\sim}{0}, I_K - \sqrt{p_X}^{\otimes 2}\right)$; (ii) $\underset{\sim}{Z_n}^T \underset{\sim}{Z_n} \xrightarrow{\mathcal{D}} \chi^2_{K-1}$

Dk $\underset{\sim}{X} = \sum_{1}^{n} \underset{\sim}{Y_i}$, $E\underset{\sim}{Y_i} = \underset{\sim}{p_X}$, $\operatorname{var} \underset{\sim}{Y_i} = \operatorname{diag} \sqrt{p_X} \left(I_K - \sqrt{p_X}^{\otimes 2}\right) \operatorname{diag} \sqrt{p_X}$

(i) $\frac{1}{\sqrt{n}} \sum_{1}^{n} \left(\underset{\sim}{Y_i} - E\underset{\sim}{Y_i}\right) \xrightarrow{\mathcal{D}} N_K\left(\underset{\sim}{0}, \operatorname{var} \underset{\sim}{Y_i}\right) \Rightarrow \frac{1}{\sqrt{n}}\left(\underset{\sim}{X} - n\underset{\sim}{p_X}\right) \xrightarrow{\mathcal{D}} N\left(\underset{\sim}{0}, \operatorname{var} \underset{\sim}{Y_i}\right) \Rightarrow$

$\Rightarrow \underset{\sim}{Z_n} = \frac{1}{\sqrt{n}}\left(\operatorname{diag}\sqrt{p_X}\right)^{-1}\left(\underset{\sim}{X} - n\underset{\sim}{p_X}\right) \xrightarrow{\mathcal{D}} \underset{\sim}{Z} \sim N_K\left(\underset{\sim}{0}, I_K - \sqrt{p_X}^{\otimes 2}\right)$

$Z_{nk} = \frac{X_k - np_k}{\sqrt{np_k}}$

(ii) $\underset{\sim}{Z_n} \xrightarrow{\mathcal{D}} \underset{\sim}{Z} \Rightarrow \underset{\sim}{Z_n}^T \underset{\sim}{Z_n} \xrightarrow{\mathcal{D}} \underset{\sim}{Z}^T \underset{\sim}{Z}$; $\underset{\sim}{Z} \sim N_K(\underset{\sim}{0}, \Sigma)$, Σ je idemp. $\Rightarrow \underset{\sim}{Z}^T \underset{\sim}{Z} \sim \chi^2_{\operatorname{tr} \Sigma = K-1}$ ■

9. ANALÝZA ROZPTYLU

V9.1 $SS_C = SS_A + SS_E$

Dk $SS_C = \sum_{ij}\left(Y_{ij} - \overline{Y}_{++}\right)^2 = \sum_{ij}\left(Y_{ij} - \overline{Y}_{i+} + \overline{Y}_{i+} - \overline{Y}_{++}\right)^2 = \underbrace{\sum_{ij}\left(Y_{ij} - \overline{Y}_{i+}\right)^2}_{SS_E} + \underbrace{\sum_i n_i\left(\overline{Y}_{i+} - \overline{Y}_{++}\right)^2}_{SS_A} + \underbrace{2\sum_{ij}\left(Y_{ij} - \overline{Y}_{i+}\right)\left(\overline{Y}_{i+} - \overline{Y}_{++}\right)}_{= \sum_i \left(\overline{Y}_{i+} - \overline{Y}_{++}\right) \underbrace{\sum_j \left(Y_{ij} - \overline{Y}_{i+}\right)}_{=0}}$ ■

V9.2

Dk 1. $SS_E/\sigma^2 \sim \chi^2_{n-k}$, $E \frac{SS_E}{n-k} = \sigma^2$

$Y_{ij} \sim N(\mu_i, \sigma^2)$: $\underbrace{\frac{\sum_j \left(Y_{ij} - \overline{Y}_{i+}\right)^2}{\sigma^2} \sim \chi^2_{n_i - 1}}_{\text{nezávislé pro } i = 1, \dots, k}$ dle V1.3 aplikované na $Y_{i1}, \dots, Y_{in_i}$ $\Rightarrow \sum_i \frac{\sum_j \left(Y_{ij} - \overline{Y}_{i+}\right)^2}{\sigma^2} = \frac{SS_E}{\sigma^2} \sim \chi^2_{\sum n_i - k} \equiv \chi^2_{n-k}$; $E \frac{SS_E}{\sigma^2} = n - k$

2. $\frac{SS_C}{\sigma^2} \sim \chi^2_{n-1}$ za H_0, $E\frac{SS_C}{n-1} = \sigma^2$ za H_0

H_0 platí $\Rightarrow \mu_1 = \dots = \mu_k \equiv \mu$: $Y_{ij} \sim N(\mu, \sigma^2)$ náh. výběr o rozsahu $n \Rightarrow \frac{\sum_{ij}\left(Y_{ij} - \overline{Y}_{++}\right)^2}{\sigma^2} \sim \chi^2_{n-1}$ dle V1.3 aplikované na $Y_{11}, \dots, Y_{kn_k}$

3. $\frac{SS_A}{\sigma^2} \sim \chi^2_{k-1}$ za H_0, $E\frac{SS_A}{k-1} = \sigma^2$ za H_0

$\mu_1 = \dots = \mu_k \equiv \mu$: vezmeme $\underset{\sim}{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_k \end{pmatrix}$, kde $Z_i = \overline{Y}_{i+} - \mu \overset{H_0}{\sim} N\left(0, \frac{\sigma^2}{n_i}\right)$ nezávislé

Napíšeme $\frac{SS_A}{\sigma^2}$ jako kvadratickou formu pro $\underset{\sim}{Z}$: $\frac{SS_A}{\sigma^2} = \frac{1}{\sigma^2}\sum_i n_i\left(\overline{Y}_{i+} - \overline{Y}_{++}\right)^2 = \frac{1}{\sigma^2}\sum_i \left(\underbrace{\overline{Y}_{i+} - \mu}_{Z_i} - \left(\overline{Y}_{++} - \mu\right)\right)^2$

$\overline{Y}_{++} - \mu = \frac{1}{n}\sum_i \sum_j \left(Y_{ij} - \mu\right) = \frac{1}{n}\sum_i n_i\left(\overline{Y}_{i+} - \mu\right) = \frac{1}{n}\sum_i n_i Z_i$ $\qquad \underset{\sim}{n} = \begin{pmatrix} n_1 \\ \vdots \\ n_k \end{pmatrix}$

$\frac{SS_A}{\sigma^2} = \frac{1}{\sigma^2}\sum_i n_i\left(Z_i - \frac{1}{n}\sum_j n_j Z_j\right)^2 = \frac{1}{\sigma^2}\left[\sum_i n_i Z_i^2 - \frac{2}{n}\left(\sum_i n_i Z_i\right)^2 + \frac{\sum n_i}{n}\left(\sum n_i Z_i\right)^2\right] =$

$= \frac{1}{\sigma^2}\left[\sum n_i Z_i^2 - \frac{1}{n}\sum\left(n_i Z_i\right)^2\right] = \frac{1}{\sigma^2}\left(\underset{\sim}{Z}^T \operatorname{diag}\{\underset{\sim}{n_i}\}\underset{\sim}{Z} - \frac{1}{n}\underset{\sim}{Z}^T \underset{\sim}{n}\,\underset{\sim}{n}^T \underset{\sim}{Z}\right) = \frac{\underset{\sim}{Z}^T}{\sigma}\underbrace{\left(\operatorname{diag}\{n_i\} - \frac{1}{n}\underset{\sim}{n}\,\underset{\sim}{n}^T\right)}_{A}\frac{\underset{\sim}{Z}}{\sigma}$

$N_k\left(\underset{\sim}{0}, \underbrace{\operatorname{diag}\left(\frac{1}{n_i}\right)}_{\Sigma}\right)$

Je $A\Sigma$ idempotentní? $A\Sigma = I_k - \frac{1}{n} n \mathbb{1}^T$

$A\Sigma A\Sigma = I_k - \frac{2}{n} n \mathbb{1}^T + \frac{1}{n^2} n \underbrace{\mathbb{1}^T \mathbb{1}}_{=n} n^T = I_k - \frac{1}{n} n \mathbb{1}^T = A\Sigma$, $\operatorname{tr} A\Sigma = k - \frac{1}{n}\sum n_i = k-1$ ∎

Tudíž $\frac{SS_A}{\sigma^2} \sim \chi^2_{k-1}$ za H_0, $E\frac{SS_A}{\sigma^2} = k-1$ za H_0

4. SS_A a SS_E jsou nezávislé ... bez Dk

V 9.3 $F_A = \frac{SS_A/(k-1)}{SS_E/(n-k)} \sim F_{k-1,n-k}$ za H_0

Dk $$F_A = \frac{\overbrace{\frac{SS_A}{\sigma^2}}^{\chi^2_{k-1}}/(k-1)}{\underbrace{\frac{SS_E}{\sigma^2}}_{\sim\chi^2_{n-k}}/(n-k)}$$ nezáv. $\overset{H_0}{\sim} F_{k-1,n-k}$ ∎

V 9.4 $k=2: F_A = T^2_{n_1,n_2}$

Dk $\frac{SS_A}{k-1} = \sum_{i=1}^{2} n_i(\bar Y_{i+} - \bar Y_{++})^2 = n_1(\bar Y_{1+}^2 - 2\bar Y_{1+}\bar Y_{++} + \bar Y_{++}^2) + n_2(\bar Y_{2+}^2 - 2\bar Y_{2+}\bar Y_{++} + \bar Y_{++}^2) = n_1\bar Y_{1+}^2 + n_2\bar Y_{2+}^2 + n\bar Y_{++}^2 -$

$-2\bar Y_{++}\underbrace{(n_1\bar Y_{1+} + n_2\bar Y_{2+})}_{=n\bar Y_{++}} = n_1\bar Y_{1+}^2 + n_2\bar Y_{2+}^2 - n\underbrace{\bar Y_{++}^2}_{=\frac{1}{n}\left(n_1^2\bar Y_{1+}^2 + n_2\bar Y_{2+}^2 + 2n_1n_2\bar Y_{1+}\bar Y_{2+}\right)} =$

$= \underbrace{\left(n_1 - \frac{n_1^2}{n_1+n_2}\right)}_{\frac{n_1n_2}{n_1+n_2}}\bar Y_{1+}^2 + \underbrace{\left(n_2 - \frac{n_2^2}{n_1+n_2}\right)}_{\frac{n_1n_2}{n_1+n_2}}\bar Y_{2+}^2 - 2\frac{n_1n_2}{n_1+n_2}\bar Y_{1+}\bar Y_{2+} = \frac{n_1n_2}{n_1+n_2}(\bar Y_{1+} - \bar Y_{2+})^2$

$\frac{SS_E}{n-k} = \frac{1}{n_1+n_2-1}\sum_i\sum_j (Y_{ij} - \bar Y_{i+})^2 = S^2_{n_1,n_2}$; $F_A = \Bigg(\underbrace{\sqrt{\frac{n_1n_2}{n_1+n_2}}\,\frac{\bar Y_{1+} - \bar Y_{2+}}{\sqrt{S^2_{n_1,n_2}}}}_{T_{n_1,n_2}}\Bigg)^2$ ∎

10. REGRESE

10.3 Odhady

V 10.3

Dk 1. $E\hat\beta = (X^TX)^{-1}X^T(X\beta) = \beta$; 2. $\operatorname{var}\hat\beta = (X^TX)^{-1}X^T\sigma^2 I X(X^TX)^{-1} = \sigma^2(X^TX)^{-1}$; 3. bez Dk

V 10.4 $\frac{SS_E}{\sigma^2} \sim \chi^2_{n-p}$

Dk $SS_E = u^Tu = Y^T\overset{\text{symetr., idemp.}}{(I-H)^T(I-H)}Y = Y^T(I-H)Y$; dále $(I-H)X = X - HX = X - X = 0$, díky čemuž

$SS_E = (Y - X\beta)^T(I-H)(Y - X\beta)$ a $\frac{SS_E}{\sigma^2} = \left(\frac{Y - X\beta}{\sigma}\right)^T\underbrace{(I-H)}_{\text{idemp.}}\underbrace{\left(\frac{Y - X\beta}{\sigma}\right)}_{\sim N(0,I)} \Rightarrow \frac{SS_E}{\sigma^2} \sim \chi^2_{h(I-H) = tr(I-H) = n-p}$ ∎

V 10.5 $\frac{c^T\hat\beta - c^T\beta}{\sqrt{\frac{SS_E}{n-p}c^T(X^TX)^{-1}c}} \sim t_{n-p}$

Dk $Y \sim N(X\beta, \sigma^2 I) \Rightarrow \hat\beta = (X^TX)^{-1}X^TY \sim N_p(\beta, \sigma^2(X^TX)^{-1}) \Rightarrow c^T(\hat\beta - \beta) \sim N_p(0, \sigma^2 c^T(X^TX)^{-1}c)$

navíc $\hat\beta = (X^TX)^{-1}X^TY$ a $SS_E = Y^T(I-H)Y$ jsou nezávislé neboť

$cov((X^TX)^{-1}X^TY, (I-H)Y) = \sigma^2(X^TX)^{-1}\underbrace{X^T(I-H)^T}_{=[(I-H)X]^T=0} = 0$; $\frac{SS_E}{\sigma^2} \sim \chi^2_{n-p}$

$$\frac{\dfrac{c^T\hat\beta - c^T\beta}{\sqrt{\sigma^2 c^T(X^TX)^{-1}c}} \sim N(0,1)}{\sqrt{\underbrace{\frac{SS_E}{\sigma^2}}_{\sim\chi^2_{n-p}}/(n-p)}} = \frac{c^T\hat\beta - c^T\beta}{\sqrt{\frac{SS_E}{n-p}c^T(X^TX)^{-1}c}} \sim t_{n-p}$$

nezáv. ∎

Dk) V10.1 ozn. $EY_i = \mu_Y$, $\operatorname{var} Y_i = \sigma_Y^2$, $EX_i = \mu_X$, $\operatorname{var} X_i = \sigma_X^2$

dle VP6.2: podmíněné rozdělení Y_i, je-li dáno $X_i = x_i$, je normální se stř. hodnotou $\mu_Y + \frac{\operatorname{cov}(X,Y)}{\sigma_X^2}(x_i - \mu_X)$, kde $\frac{\operatorname{cov}(X,Y)}{\sigma_X^2} = \rho \frac{\sigma_X}{\sigma_Y}$, a rozptylem, který nezávisí na x_i

Y_i splňuje lineární model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$; $i = 1, \dots, n$, $E\varepsilon_i = 0$, $\operatorname{var} \varepsilon_i = \sigma_\varepsilon^2$, kde

$\beta_1 = \rho \frac{\sigma_Y}{\sigma_X}$ a $\beta_0 = \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} \mu_X$ $\quad \dots \rho = 0 \iff \beta_1 = 0$

Odhad $\hat{\beta} = \begin{pmatrix} \hat\beta_0 \\ \hat\beta_1 \end{pmatrix}$ metodou nejm. čtverců je $\hat{\beta} = (X^TX)^{-1}(X^T\underset{\sim}{Y})$, kde $X = \begin{pmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}$, $X^TX = \begin{pmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix}$,

$X^T\underset{\sim}{Y} = \begin{pmatrix} \sum Y_i \\ \sum X_i Y_i \end{pmatrix} \rightsquigarrow \hat{\beta}$ řeší soustavu: $\left.\begin{matrix} n\hat\beta_0 + \sum X_i \hat\beta_1 = \sum Y_i \\ \sum X_i \hat\beta_0 + \sum X_i^2 \hat\beta_1 = \sum X_i Y_i \end{matrix}\right\} \Rightarrow \left.\begin{matrix} \hat\beta_0 + \overline{X}_n \hat\beta_1 = \overline{Y}_n \;/ \cdot \overline{X}_n \\ \overline{X}_n \hat\beta_0 + \frac{1}{n} \sum X_i^2 \hat\beta_1 = \frac{1}{n} \sum X_i Y_i \end{matrix}\right\} \Rightarrow$

$$\Rightarrow \left[\frac{1}{n} \sum X_i^2 - (\overline{X}_n)^2\right] \hat\beta_1 = \frac{1}{n} \sum X_i Y_i - \overline{X}_n \overline{Y}_n \Rightarrow \hat\beta_1 = \frac{\sum (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\sum (X_i - \overline{X}_n)^2} = \hat\rho \frac{S_X}{S_Y} \Rightarrow \hat\beta_0 = \overline{Y}_n - \overline{X}_n \hat\rho \frac{S_Y}{S_X}$$

$S_{XY} = \frac{1}{n-1} \sum (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)$

$$SS_E = \underset{\sim}{Y}^T (I - H) \underset{\sim}{Y} = \underset{\sim}{Y}^T \underset{\sim}{Y} - \underset{\sim}{Y}^T \widehat{H\underset{\sim}{Y}} = \sum Y_i^2 - \sum Y_i (\hat\beta_0 + \hat\beta_1 X_i) = \sum Y_i^2 - \sum Y_i \left(\overline{Y}_n - \overline{X}_n \hat\rho \frac{S_Y}{S_X}\right) - \left(\sum Y_i X_i\right) \hat\rho \frac{S_Y}{S_X} =$$

$$= (n-1) S_Y^2 - \hat\rho \underbrace{\frac{S_Y}{S_X} (n-1) S_{XY}}_{S_Y^2 (n-1) \hat\rho} = (n-1) S_Y^2 (1 - \hat\rho^2)$$

Test $H_0: \rho = 0$ je ekvivalentní testu $H_0: \beta_1 = 0$ $\rightsquigarrow$ zvolíme $\underset{\sim}{c} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ ve V10.5 $\rightsquigarrow$ máme testovou statistiku $\frac{\hat\beta_1}{\sqrt{\frac{SS_E}{n-2} v_2}}$, která má za H_0 rozdělení t_{n-2}, kde v_2 je pravý dolní prvek $(X^TX)^{-1}$; podle Lemma 4.4 je $v_2 = \left(\sum X_i^2 - \left(\sum X_i\right)^2 \frac{1}{n}\right)^{-1} = \left(n \cdot \frac{1}{n} \sum (X_i - \overline{X}_n)^2\right)^{-1} = \frac{1}{(n-1) S_X^2}$

Dosadíme za $\hat\beta_1$, SS_E, v_2 a dostaneme

$$\frac{\hat\rho \frac{S_Y}{S_X}}{\sqrt{\frac{n-1}{n-2} S_Y^2 (1 - \hat\rho^2) \frac{1}{(n-1) S_X^2}}} = \sqrt{n-2} \frac{\hat\rho}{\sqrt{1 - \hat\rho^2}} \overset{H_0}{\sim} t_{n-2} \quad \blacksquare$$