List formul za 4. kolokvij in izpite

Naj bo $X_1, X_2, \ldots, X_n \sim X$ slučajni vzorec. Z $\overline{X}$ označimo vzorčno povprečje. *Vzorčna disperzija* je enaka

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2 = \frac{1}{n-1}\left(\sum_{k=1}^{n} X_k^2 - n\overline{X}^2\right).$$

Naj bo F porazdelitvena funkcija slučajne spremenljivke X. Porazdelitvena funkcija $\min\{X_1, X_2, \ldots, X_n\}$ je enaka $1-(1-F)^n$. Porazdelitvena funkcija $\max\{X_1, X_2, \ldots, X_n\}$ je enaka F^n.

Intervali zaupanja za normalna vzorca

Naj bosta $X_1, X_2, \ldots, X_n$ in $Y_1, Y_2, \ldots, Y_m$ neodvisna slučajna vzorca, porazdeljena normalno - prvi z upanjem $\mu_1 = \mu$ in dispezijo $\sigma_1^2 = \sigma^2$ ter drugi z upanjem μ_2 in disperzijo σ_2^2. Naj bo $S_X^2 = S^2$ vzorčna disperzija prvega vzorca in S_Y^2 vzorčna disperzija drugega vzorca. Intervali zaupanja s stopnjo zaupanja β, kjer $(*)$ označuje približni interval zaupanja:

σ **znan** $\mu = \overline{X} \pm \Phi^{-1}\left(\frac{1+\beta}{2}\right)\frac{\sigma}{\sqrt{n}}$

σ **neznan** $\mu = \overline{X} \pm F_{t_{n-1}}^{-1}\left(\frac{1+\beta}{2}\right)\frac{S}{\sqrt{n}}$

μ **znan** $\frac{(n-1)\widehat{\sigma^2}}{x_2} \leq \sigma^2 \leq \frac{(n-1)\widehat{\sigma^2}}{x_1}$, kjer je $x_1 = F_{\chi_n^2}^{-1}\left(\frac{1-\beta}{2}\right)$, $x_2 = F_{\chi_n^2}^{-1}\left(\frac{1+\beta}{2}\right)$ in $\widehat{\sigma^2} = \frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2$

μ **neznan** $\frac{(n-1)S^2}{x_2} \leq \sigma^2 \leq \frac{(n-1)S^2}{x_1}$, kjer je $x_1 = F_{\chi_{n-1}^2}^{-1}\left(\frac{1-\beta}{2}\right)$ in $x_2 = F_{\chi_{n-1}^2}^{-1}\left(\frac{1+\beta}{2}\right)$

σ_1, σ_2 **znana** $\mu_1 - \mu_2 = \overline{X} - \overline{Y} \pm \Phi^{-1}\left(\frac{1+\beta}{2}\right)\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$

$(*)$ $\sigma_1 \neq \sigma_2$ **neznana** $\mu_1 - \mu_2 = \overline{X} - \overline{Y} \pm \Phi^{-1}\left(\frac{1+\beta}{2}\right)\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}$

$\sigma_1 = \sigma_2$ **neznana** $\mu_1 - \mu_2 = \overline{X} - \overline{Y} \pm F_{t_{n+m-2}}^{-1}\left(\frac{1+\beta}{2}\right) S_p\sqrt{\frac{1}{n} + \frac{1}{m}}$, kjer je $S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$

Intervali zaupanja za Bernoullijeva vzorca

Naj bosta slučajna vzorca $X_1, X_2, \ldots, X_n$ in $Y_1, Y_2, \ldots, Y_m$ porazdeljena Bernoullijevo, prvi s parametrom $p_1 = p$ in drugi s parametrom p_2. S $\hat{p_1} = \hat{p}$ in $\hat{p_2}$ označimo pripadajoči relativni frekvenci. Intervali zaupanja s stopnjo zaupanja β, kjer $(*)$ označuje približni interval zaupanja:

Clopper-Pearson $L(K) \leq p \leq U(K)$, kjer je $K = X_1 + X_2 + \ldots + X_n$,
$L(K) = F_{\mathrm{Beta(K,n-K+1)}}^{-1}\left(\frac{1-\beta}{2}\right)$ za $K > 0$ in $L(0) = 0$ ter
$U(K) = F_{\mathrm{Beta(K+1,n-K)}}^{-1}\left(\frac{1+\beta}{2}\right)$ za $K < n$ in $U(n) = 1$

$(*)$ **CLI in ZVŠ aproksimacija** $p = \hat{p} \pm \Phi^{-1}\left(\frac{1+\beta}{2}\right)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$(*)$ **neodvisna vzorca** $p_1 - p_2 = \hat{p_1} - \hat{p_2} \pm \Phi^{-1}\left(\frac{1+\beta}{2}\right)\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n} + \frac{\hat{p_2}(1-\hat{p_2})}{m}}$

$(*)$ $m = n$**, odvisna vzorca (na način z vaj)** $p_1 - p_2 = \hat{p_1} - \hat{p_2} \pm \Phi^{-1}\left(\frac{1+\beta}{2}\right)\sqrt{\frac{\hat{p_1}(1-\hat{p_1}) + \hat{p_2}(1-\hat{p_2}) + 2\hat{p_1}\hat{p_2}}{n}}$

Interval zaupanja za Poissonov vzorec

Naj bo slučajni vzorec $X_1, X_2, \ldots, X_n$ porazdeljen Poissonovo s parametrom λ in $K = X_1 + X_2 + \ldots + X_n$. Interval zaupanja za λ s stopnjo zaupanja β je enak $L(K) \leq \lambda \leq U(K)$, kjer je $L(K) = \frac{1}{2n}F_{\chi_{2K}^2}^{-1}\left(\frac{1-\beta}{2}\right)$ za $K > 0$ in $L(0) = 0$ ter $U(K) = \frac{1}{2n}F_{\chi_{2K+2}^2}^{-1}\left(\frac{1+\beta}{2}\right)$.

Asimptotične lastnosti cenilk

$\mathrm{SKN}(W) = \mathrm{D}(W) + \mathrm{b}(W)^2,\ \ \mathrm{b}(W) = \mathrm{E}(W) - \text{parameter}$
Metoda delta za CLI: če je h zvezno odvedljiva v μ, potem velja $(h(\overline{X}) - h(\mu))/\sqrt{\mathrm{D}(X)/n} \underset{n\to\infty}{\Longrightarrow} h'(\mu)\mathrm{N}(0,1)$.

Testiranje upanja in razlike upanj

Naj bosta vzorca neodvisna. Spodaj je $\nu = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{(S_1^2/n_1)^2/(n_1-1) + (S_2^2/n_2)^2/(n_2-1)}$.

predpostavke	testna statistika	porazdelitev ob H_0
Bernoullijev vzorec	$n\overline{X}$	$\mathrm{Bin}(n, p_0)$
Poissonov vzorec	$n\overline{X}$	$\mathrm{Poiss}(n\lambda_0)$
normalen vzorec, σ^2 znana	$\frac{\overline{X}-\mu_0}{\sqrt{\sigma^2/n}}$	$\mathrm{N}(0,1)$
normalen vzorec, σ^2 neznana	$\frac{\overline{X}-\mu_0}{\sqrt{S^2/n}}$	t_{n-1}
poljuben vzorec, velik n	$\frac{\overline{X}-\mu_0}{\sqrt{S^2/n}}$	$\approx \mathrm{N}(0,1)$
normalna vzorca, σ_1^2, σ_2^2 znani	$\frac{\overline{X}_1-\overline{X}_2-D_0}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$	$\mathrm{N}(0,1)$
normalna vzorca, σ_1^2, σ_2^2 neznani	$\frac{\overline{X}_1-\overline{X}_2-D_0}{\sqrt{S_1^2/n_1+S_2^2/n_2}}$	$\approx t_{\lfloor\nu\rfloor}$
poljubna vzorca, velika n_1 in n_2	$\frac{\overline{X}_1-\overline{X}_2-D_0}{\sqrt{S_1^2/n_1+S_2^2/n_2}}$	$\approx \mathrm{N}(0,1)$
Bernoullijeva vzorca, velika n_1 in n_2	$\frac{\overline{X}_1-\overline{X}_2-D_0}{\sqrt{\hat{p}_1(1-\hat{p}_1)/(n_1-1)+\hat{p}_2(1-\hat{p}_2)/(n_2-1)}}$	$\approx \mathrm{N}(0,1)$

Če sta vzorca odvisna in porazdeljena Bernoullijevo, potem za vzorec $D_i = X_i - Y_i$ velja
$S_D^2/n = (\hat{p}_1(1-\hat{p}_1) + \hat{p}_2(1-\hat{p}_2) + 2\hat{p}_1\hat{p}_2)/(n-1)$.

Razmerje verjetij

$\lambda = \frac{\sup\{L(\theta) \mid \theta \in H_0\}}{\sup\{L(\theta) \mid \theta \in \Theta\}}$; H_0 zavrnemo, če $\lambda < D$
Ob H_0 je $-2\ln\lambda \approx \chi^2_{\dim\Theta - \dim H_0}$; H_0 zavrnemo, če $-2\ln\lambda > C$.
Če je slučajni vzorec porazdeljen multinomsko s parametri $p_0, p_1, \ldots, p_m$, je $L(p_0, \ldots, p_m) = p_0^{T_0} p_1^{T_1} \ldots p_m^{T_m}$ (T_i pripadajoče frekvence) in cenilka največjega verjetja za $(p_0, p_1, \ldots, p_m)$ je $(\frac{T_0}{n}, \frac{T_1}{n}, \ldots, \frac{T_m}{n})$.

Hi-kvadrat prilagoditveni test

$T = \sum_k \frac{(O_k - P_k)^2}{P_k}$, kjer so O_k opazovane in P_k pričakovane frekvence

Enostavna linearna regresija

Model: $Y = \alpha + \beta X + \varepsilon$.
vzorčna kovarianca: $S_{x,y} = \frac{1}{n-1}\sum_k (x_k - \overline{x})(y_k - \overline{y}) = \frac{1}{n-1}\left(\sum_k x_k y_k - n\overline{x}\,\overline{y}\right)$
$\mathrm{VKR} = \sum_k (y_k - \hat{y}_k)^2,\ \ \mathrm{RVK} = \sum_k (\hat{y}_k - \overline{y})^2,\ \ \mathrm{SVK} = \sum_k (y_k - \overline{y})^2,\ \ \mathrm{SVK} = \mathrm{VKR} + \mathrm{RVK}$
$R^2 = \frac{\mathrm{RVK}}{\mathrm{SVK}} = \frac{(S_{x,y})^2}{S_x^2 S_y^2},\ \ \hat{\alpha} = \overline{y} - \hat{\beta}\,\overline{x},\ \ \hat{\beta} = \frac{S_{x,y}}{S_x^2}$
$(\hat{\alpha} - \alpha)/\sqrt{\frac{\mathrm{VKR}}{n-2}\frac{\sum_k x_k^2}{n(n-1)S_x^2}} \sim t_{n-2},\ \ (\hat{\beta} - \beta)/\sqrt{\frac{\mathrm{VKR}}{(n-2)(n-1)S_x^2}} \sim t_{n-2}$

Analiza variance (m skupin)

$\mathrm{VMS} = \sum_{j=1}^m n_j(\overline{X}_{\bullet j} - \overline{X}_{\bullet\bullet})^2,\ \ \mathrm{VZS} = \sum_{j=1}^m (n_j - 1)S_j^2,\ \ F = \frac{\mathrm{VMS}/(m-1)}{\mathrm{VZS}/(n-m)} \sim \mathrm{F}(m-1, n-m)$ ob H_0