

1 Statistická inference II - cvičení 16-02-23

Řešení příkladů z domácí úlohy - SI I

Příklad č.1 (dvojezměrné normální rozdělení)

Nechť náhodnou proměnnou X je největší výška mozkovny u mužů (skull.pH, v mm) a náhodnou proměnnou Y je morfologická výška tváře u mužů (face.H; v mm); data: one-sample-correlation-skull-mf.txt. Nechť $E[X] = \mu_1$ je střední hodnota největší výšky mozkovny a $Var[X] = \sigma_1^2$ je rozptyl největší výšky mozkovny, $E[Y] = \mu_2$ je střední hodnota morfologické výšky tváře a $Var[Y] = \sigma_2^2$ je rozptyl morfologické výšky tváře. Předpokládejme, že největší výška mozkovny X má normální rozdělení $N(\mu_1, \sigma_1^2)$ a morfologická výška tváře Y má normální rozdělení $N(\mu_2, \sigma_2^2)$. Potom $(X, Y)^T$ má dvojezměrné normální rozdělení $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ s parametry $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$, což je vektor středních hodnot a σ_1^2, σ_2^2 a ρ , což jsou parametry kovarianční matice $\boldsymbol{\Sigma}$, přičemž síla lineárního vztahu těchto dvou proměnných je daná velikostí a znaménkem ρ . Potom $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)^T$.

- Nakreslete hustotu dvojezměrného normálního rozdělení $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ pomocí funkce `image()` a superponujte ji konturovým grafem hustoty toho stejného rozdělení pomocí funkce `contour()`.
- Nakreslete dvojezměrný jádrový odhad hustoty pomocí funkcí `kde2d()` a `image()` a superponujte ho konturovým grafem hustoty dvojezměrného normálního rozdělení $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ pomocí funkce `contour()`. Namísto $\boldsymbol{\theta}$ použijte vektor $\hat{\boldsymbol{\theta}} = (\bar{x}_1, \bar{x}_2, s_1^2, s_2^2, r)^T$ odhadnutý z dat, kde r je Pearsonův korelační koeficient.

Příklad č.2 (směs dvou dvojezměrných normálních rozdělení)

Nechť $(X_1, Y_1)^T$ pochází z rozdělení $N_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, kde X_1 je průměrná délka dolní končetiny (lowex.L; v mm) a Y_1 je délka trupu (tru.L; v mm) u mužů. Nechť $(X_2, Y_2)^T$ pochází z rozdělení $N_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, kde X_2 je průměrná délka dolní končetiny (lowex.L; v mm) a Y_2 délka trupu (tru.L; v mm) u žen; data: two-samples-correlations-trunk.txt. Předpokládejme, že průměrná délka dolní končetiny X a délka trupu Y pochází (1) ze směsi $pN_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) + (1-p)N_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, kde $\boldsymbol{\theta} = (\mu_{11}, \mu_{12}, \sigma_{11}^2, \sigma_{12}^2, \rho_1, \mu_{21}, \mu_{22}, \sigma_{21}^2, \sigma_{22}^2, \rho_2)^T$ a (2) z dvojezměrného rozdělení $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, kde parametry představují společný vektor středních hodnot a společnou kovarianční matici, t.j. $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)^T$.

- Nakreslete teoretickou hustotu (2) pomocí funkce `image()` a superponujte ji konturovým grafem teoretické hustoty (2) pomocí funkce `contour()`.
- Nakreslete teoretickou hustotu (1) pomocí funkce `image()` a superponujte ji konturovým grafem teoretické hustoty (1) pomocí funkce `contour()`.
- Nakreslete dvojezměrný jádrový odhad hustoty realizací (1) pomocí funkce `image()` a superponujte ho konturovým grafem teoretické hustoty (1) pomocí funkce `contour()`.

Poznámka:

- $\hat{\boldsymbol{\theta}} = (\hat{\mu}_{11}, \hat{\mu}_{12}, \hat{\sigma}_{11}^2, \hat{\sigma}_{12}^2, \hat{\rho}_1, \hat{\mu}_{21}, \hat{\mu}_{22}, \hat{\sigma}_{21}^2, \hat{\sigma}_{22}^2, \hat{\rho}_2)^T$ a $p = n_1/(n_1 + n_2)$; parametry jsou odhadnuté z dat.
- $\hat{\boldsymbol{\theta}} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\rho})^T$; parametry jsou odhadnuté ze společného výběru.

Příklad č.3 (testovací statistika, simulační studie)

Na základě simulační studie proveďte, že pokud

- $X \sim N(\mu, \sigma^2)$, kde $\mu = 0, \sigma^2 = 1$;

(b) $X \sim [(1-p)N(\mu, \sigma^2) + pN(\mu, \sigma_1^2)]$, kde $p = 0.05$, $\mu = 0$ a $\sigma_1^2 = 2$,

potom testovací statistika

$$F = \frac{(n-1)S^2}{\sigma^2}$$

má asymptoticky χ^2 rozdělení s $n-1$ stupni volnosti. Použijte rozsahy náhodných výběrů $n = 15$ a $n = 100$. Pro každou simulaci X vypočítejte $F_{obs,m}$, kde $m = 1, 2, \dots, M$, přičemž $M = 1000$. Superponujte histogram vygenerovaných testovacích statistik v relativní škále s teoretickou křivkou hustoty F .

Příklad č.4 (kvadratická aproximace profilové funkce věrohodnosti)

- (a) Nakreslete škálovaný logaritmus profilové funkce věrohodnosti normálního rozdělení pro μ . Na ose x bude μ a na ose y $\ln \mathcal{L}_P(\mu|\mathbf{x}) = l_P(\mu|\mathbf{x}) - l_P(\hat{\mu}|\mathbf{x})$. Porovnejte $\ln \mathcal{L}_P(\mu|\mathbf{x})$ s kvadratickou aproximací vypočítanou pomocí Taylorova rozvoje $\ln \mathcal{L}_P(\mu|\mathbf{x}) = \ln\left(\frac{\mathcal{L}_P(\mu|\mathbf{x})}{\mathcal{L}_P(\hat{\mu}|\mathbf{x})}\right) \approx -\frac{1}{2}\mathcal{I}(\hat{\mu})(\mu - \hat{\mu})^2$.
- (b) Nechť skóre funkce $S(\mu) = \frac{\partial}{\partial \mu} \ln \mathcal{L}_P(\mu|\mathbf{x})$. Vezmeme-li derivaci kvadratické aproximace uvedené výše, dostaneme $S(\mu) \approx -\mathcal{I}(\hat{\mu})(\mu - \hat{\mu})$ nebo $-\mathcal{I}^{-1/2}(\hat{\mu})S(\mu) \approx \mathcal{I}^{1/2}(\hat{\mu})(\mu - \hat{\mu})$. Potom zobrazením pravé strany na ose x a levé strany na ose y dostaneme asymptoticky lineární funkci s jednotkovým sklonem. Asymptoticky také platí $\mathcal{I}^{1/2}(\bar{X})(\mu - \bar{X}) \stackrel{\mathcal{D}}{\sim} N(0, 1)$. Je postačující mít rozsah osy x rovný $\langle -2, 2 \rangle$, protože funkce je asymptoticky (lokálně) lineární na tomto intervalu. Rozumně škálujte osu y . Zobrazte pro (a) $n = 10$, (b) $n = 100$ a (c) $n = 1000$. Použijte (1) $X \sim N(0, 1)$ a (2) $X \sim (1-p)N(0, 1) + pN(0, 2)$, kde $p = 0.05$. Okomentujte rozdíly mezi (a), (b) a (c), stejně jako rozdíly mezi (1) a (2).

Příklad č.5 (maximálně věrohodný odhad μ a σ^2)

Vygenerujte pseudonáhodná čísla z $X \sim N(4, 1)$, $n = 1000$.

- (a) Napište logaritmus profilové funkce věrohodnosti pro μ a σ^2 a proveďte, zda jsou maximálně věrohodné odhady μ a σ^2 dostatečně blízko k jejich skutečným hodnotám. Nakreslete grafy $l(\mu|\mathbf{x})$ a $l(\sigma^2|\mathbf{x})$, kde zvýrazníte polohu maxim těchto funkcí.
- (b) Napište logaritmus funkce věrohodnosti pro $\theta = (\mu, \sigma^2)^T$ a proveďte, zda je maximálně věrohodný odhad $\hat{\theta}$ dostatečně blízko k jeho skutečné hodnotě.
- (c) Nakreslete graf $l(\theta|\mathbf{x})$ použitím funkce `image()` a superponujte ho s konturovým grafem použitím funkce `contour()`. Zvýrazněte polohu maxima.

Příklad č.6 (maximálně věrohodné odhady)

Za předpokladu normality rozdělení náhodné proměnné X vypočítejte maximálně věrohodné odhady střední hodnoty μ (ozn. $\hat{\mu}$) a rozptylu σ^2 (ozn. $\hat{\sigma}^2$) pomocí logaritmů funkcí věrohodnosti $l(\mu|\mathbf{x})$, resp. $l(\sigma^2|\mathbf{x})$. Porovnejte tyto odhady s aritmetický průměrem \bar{x} a rozptylem s^2 . Musí platit $\hat{\mu} = \bar{x}$ a $\hat{\sigma}^2 = \frac{n-1}{n}s^2$. Realizacemi náhodné proměnné X jsou hodnoty x_i , $i = 1, 2, \dots, n$, proměnných:

- (a) délka pravé klíční kosti (`length.R`; data: `paired-means-clavicle2.txt`);
- (b) morfologická výška tváře (`face.H`; data: `one-sample-correlation-skull-mf.txt`);
- (c) šířka lebky (`skull.B`; data: `one-sample-mean-skull-mf.txt`).

Příklad č.7 (maximálně věrohodné odhady multinomické rozdělení)

- (a) Mějme data `more-samples-probabilities-pubis.txt`. Nakreslete logaritmus standardizované $\mathcal{L}(\boldsymbol{\theta}|\mathbf{x})$, kde $\boldsymbol{\theta} = (p_1, p_2)^T$, Evropské populace ($n_1 = 30$, $n_2 = 20$ a $n_3 = 10$) pomocí funkce `contour()`. Dokreslete do obrázku její maximum v bodě $\hat{\boldsymbol{\theta}} = (\hat{p}_1, \hat{p}_2)^T$.