ML odhad $\boldsymbol{\mu}$ a $\boldsymbol{\Sigma}$

Ďalšie pomocné tvrdenia

Lema 1*. *Platí*

$$\frac{\partial \mathbf{m}'\mathbf{x}}{\partial \mathbf{x}} = \mathbf{m},$$

$$\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}.$$

Dôkaz. Urobte ako cvičenie.

Nech $\mathbf{B}$ je regulárna $n \times n$ matica, ktorej prvky sú diferencovateľnými funkciami premennej t, čiže $\{\mathbf{B}\}_{i,j} = b_{ij} = b_{ij}(t),\ i,j = 1,2,...,n,$

$$\frac{\partial \mathbf{B}}{\partial t} \text{ je } n \times n \text{ matica, ktorej prvky sú } \frac{\partial b_{ij}(t)}{\partial t},\ i,j = 1,2,...,n$$

$$\frac{\partial det \mathbf{B}}{\partial \mathbf{B}} \text{ je } n \times n \text{ matica, ktorej prvky sú } \frac{\partial det \mathbf{B}}{\partial b_{ij}},\ i,j = 1,2,...,n,$$

$$diag \mathbf{B} = \begin{pmatrix} \{\mathbf{B}\}_{1,1} & 0 & \dots & 0 & \\ 0 & \{\mathbf{B}\}_{2,2} & \dots & 0 & \\ \vdots & & & & \\ 0 & & & & \{\mathbf{B}\}_{n,n} \end{pmatrix}.$$

Značenie je rovnaké ako v kapitolke 8. "Pomocné tvrdenia" textu "Plánovanie regresného experimentu".

Lema 2*. *Pre symetrickú regulárnu $n \times n$ maticu $\mathbf{B}$ a symetrickú maticu $\mathbf{C}$ platí*

$$\frac{\partial tr \mathbf{B}^{-1}\mathbf{C}}{\partial \mathbf{B}} = -2\mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1} + diag(\mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1}).$$

Dôkaz. Platí

$$\frac{\partial tr \mathbf{B}^{-1}\mathbf{C}}{\partial b_{11}} = -tr \mathbf{C}\mathbf{B}^{-1}\frac{\partial \mathbf{B}}{\partial b_{11}}\mathbf{B}^{-1} =$$

$$= -tr \mathbf{C}\mathbf{B}^{-1} \begin{pmatrix} 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & \\ \vdots & & & \\ 0 & 0 & 0 & \end{pmatrix} \mathbf{B}^{-1} = -tr \begin{pmatrix} 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & \\ \vdots & & & \\ 0 & 0 & 0 & \end{pmatrix} \mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1} =$$

$$= -\{\mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1}\}_{11},$$

$$\frac{\partial tr \mathbf{B}^{-1}\mathbf{C}}{\partial b_{12}} = -tr \mathbf{C}\mathbf{B}^{-1}\frac{\partial \mathbf{B}}{\partial b_{12}}\mathbf{B}^{-1} =$$

$$= -tr\mathbf{C}\mathbf{B}^{-1}\begin{pmatrix} 0 & 1 & 0 & \dots \\ 1 & 0 & 0 & \\ 0 & 0 & 0 & \\ \vdots & & & \\ 0 & 0 & 0 & \end{pmatrix}\mathbf{B}^{-1} = -tr\begin{pmatrix} 0 & 1 & 0 & \dots \\ 1 & 0 & 0 & \\ 0 & 0 & 0 & \\ \vdots & & & \\ 0 & 0 & 0 & \end{pmatrix}\mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1} =$$

$$= -\{\mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1}\}_{21} - \{\mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1}\}_{21} = 2\{\mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1}\}_{12}.$$

Úplne analogicky dostávame

$$\frac{\partial tr\mathbf{B}^{-1}\mathbf{C}}{\partial b_{ii}} = -tr\mathbf{C}\mathbf{B}^{-1}\frac{\partial \mathbf{B}}{\partial b_{ii}}\mathbf{B}^{-1} =$$

$$= \{\mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1}\}_{ii}$$

a pre $i \neq j$

$$\frac{\partial tr\mathbf{B}^{-1}\mathbf{C}}{\partial b_{ij}} = -tr\mathbf{C}\mathbf{B}^{-1}\frac{\partial \mathbf{B}}{\partial b_{ij}}\mathbf{B}^{-1} = -2\{\mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1}\}_{ij},$$

teda

$$\frac{\partial tr\mathbf{B}^{-1}\mathbf{C}}{\partial \mathbf{C}} = -2\mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1} + diag(\mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1}). \quad \square$$

Lema 3^***.** *Pre symetrickú* $n \times n$ *maticu* $\mathbf{B}$ *platí*

$$2\mathbf{B} - diag\mathbf{B} = \mathbf{0} \Leftrightarrow \mathbf{B} = \mathbf{0}.$$

Dôkaz. Spravte ako cvičenie.

Združenú funkciu hustoty rozdelenia náhodného výberu $\mathbf{X}_{np,1} = (\mathbf{X}_1', ..., \mathbf{X}_n')'$ uvažovanú pri danom $\mathbf{x}$ (realizácia $\mathbf{X} \in \mathcal{R}^p$) ako funkciu vektorového parametra $\boldsymbol{\theta} \in \mathcal{R}^q$ nazývame funkciou vierohodnosti

$$L(\mathbf{x}; \boldsymbol{\theta}) = \prod_{i=1}^{n} f(\mathbf{x}_i; \boldsymbol{\theta}),$$

resp. jej logaritmus, teda

$$l(\mathbf{x}, \boldsymbol{\theta}) = l(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n, \boldsymbol{\theta}) = \ln L(\mathbf{x}; \boldsymbol{\theta}) = \sum_{i=1}^{n} \ln f(\mathbf{x}_i; \boldsymbol{\theta}).$$

Vierohodnostnými rovnicami rozumieme systém

$$\sum_{i=1}^{n} \frac{\partial \ln f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \theta_k} = 0, \quad k = 1, 2, ..., q.$$

Majme náhodný výber $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2', ..., \mathbf{X}_n')'$, kde $\mathbf{X}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}$ je regulárna. Potom

$$L(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = |2\pi\boldsymbol{\Sigma}|^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})},$$

čiže

$$l(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \ln L(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = -\frac{n}{2}\ln|2\pi\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu}).$$

Platí

$$(\mathbf{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu}) = (\mathbf{x}_i-\overline{\mathbf{x}}+\overline{\mathbf{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\overline{\mathbf{x}}+\overline{\mathbf{x}}-\boldsymbol{\mu}) =$$

$$= (\mathbf{x}_i-\overline{\mathbf{x}})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\overline{\mathbf{x}}) + (\overline{\mathbf{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}) +$$

$$+2(\mathbf{x}_i-\overline{\mathbf{x}})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}),$$

čiže

$$\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu}) = \sum_{i=1}^{n}(\mathbf{x}_i-\overline{\mathbf{x}})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\overline{\mathbf{x}}) + n(\overline{\mathbf{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}) +$$

$$+2\sum_{i=1}^{n}(\mathbf{x}_i-\overline{\mathbf{x}})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}) = tr\sum_{i=1}^{n}(\mathbf{x}_i-\overline{\mathbf{x}})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\overline{\mathbf{x}}) + n(\overline{\mathbf{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}) =$$

$$= tr\left[\boldsymbol{\Sigma}^{-1}\left\{\sum_{i=1}^{n}(\mathbf{x}_i-\overline{\mathbf{x}})(\mathbf{x}_i-\overline{\mathbf{x}})'\right\}\right] + n(\overline{\mathbf{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}) +$$

$$\text{(A)} \qquad = ntr\left\{\boldsymbol{\Sigma}^{-1}\mathbf{S}^{(real)}\right\} + n(\overline{\mathbf{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}),$$

lebo

$$\mathbf{S}^{(real)} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\overline{\mathbf{x}})(\mathbf{x}_i-\overline{\mathbf{x}})' \quad \text{a} \quad 2\sum_{i=1}^{n}(\mathbf{x}_i-\overline{\mathbf{x}})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}) =$$

$$= 2\left\{n\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}) - n\overline{\mathbf{x}}'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu})\right\} = 0.$$

Dostávame

$$\text{(B)} \quad l(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = -\frac{n}{2}\ln|2\pi\boldsymbol{\Sigma}| - \frac{n}{2}tr\left\{\boldsymbol{\Sigma}^{-1}\mathbf{S}^{(real)}\right\} - \frac{n}{2}tr\left\{\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu})(\overline{\mathbf{x}}-\boldsymbol{\mu})'\right\} =$$

$$= -\frac{np}{2}\ln 2\pi - \frac{n}{2}\ln(det(\boldsymbol{\Sigma})) - \frac{n}{2}tr\left\{\boldsymbol{\Sigma}^{-1}\left[\mathbf{S}^{(real)} - (\overline{\mathbf{x}}-\boldsymbol{\mu})(\overline{\mathbf{x}}-\boldsymbol{\mu})'\right]\right\}.$$

Teda vierohodnostné rovnice sú

$$\left.\frac{\partial l}{\partial \boldsymbol{\mu}}\right|_{\mu=\hat{\mu}^{(real)},\boldsymbol{\Sigma}=\hat{\boldsymbol{\Sigma}}^{(real)}} = \mathbf{0},$$

$$\left.\frac{\partial l}{\partial \boldsymbol{\Sigma}}\right|_{\mu=\hat{\mu}^{(real)},\boldsymbol{\Sigma}=\hat{\boldsymbol{\Sigma}}^{(real)}} = \mathbf{0}.$$

Ak $n \geqq p+1$, pomocou Lemy 1* dostávame z prvého systému vierohodnostných rovníc

$$-2(\hat{\mathbf{\Sigma}}^{(real)})^{-1}\overline{\mathbf{x}} + 2(\hat{\mathbf{\Sigma}}^{(real)})^{-1}\boldsymbol{\mu}^{(real)} = \mathbf{0},$$

čiže

$$\hat{\boldsymbol{\mu}}^{(real)} = \overline{\mathbf{x}},$$

teda

$$\hat{\boldsymbol{\mu}} = \overline{\mathbf{X}}.$$

Ďalej budeme pokračovať bez komplikovaného značenia a využijeme ostatné lemy. Dostávame z druhého systému vierohodnostných rovníc

$$-\frac{n}{2}\frac{\partial}{\partial \mathbf{\Sigma}}\left\{\ln(det(\mathbf{\Sigma})) + tr\left[\mathbf{\Sigma}^{-1}(\mathbf{S} + (\overline{\mathbf{x}} - \boldsymbol{\mu})(\overline{\mathbf{x}} - \boldsymbol{\mu})')\right]\right\} = \mathbf{0},$$

$$2\left\{\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}(\mathbf{S} + (\overline{\mathbf{x}} - \boldsymbol{\mu})(\overline{\mathbf{x}} - \boldsymbol{\mu})')\mathbf{\Sigma}^{-1}\right\} -$$

$$-diag\left\{\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}(\mathbf{S} + (\overline{\mathbf{x}} - \boldsymbol{\mu})(\overline{\mathbf{x}} - \boldsymbol{\mu})')\mathbf{\Sigma}^{-1}\right\} = \mathbf{0},$$

čiže

$$\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}(\mathbf{S} + (\overline{\mathbf{x}} - \boldsymbol{\mu})(\overline{\mathbf{x}} - \boldsymbol{\mu})')\mathbf{\Sigma}^{-1} = \mathbf{0}.$$

Výsledne

$$\hat{\mathbf{\Sigma}} = \mathbf{S}.$$

Chceme ukázať, že pre každú realizáciu $\mathbf{x}_1, ..., \mathbf{x}_n$ je

$$l(\mathbf{x}_1, ..., \mathbf{x}_n; \overline{\mathbf{x}}, \mathbf{S}^{(real)}) - l(\mathbf{x}_1, ..., \mathbf{x}_n; \boldsymbol{\mu}, \mathbf{\Sigma}) \geqq 0.$$

Budeme ešte potrebovať niekoľko pomocných tvrdení.

Lema 4*. *Čísla $\lambda_1, ..., \lambda_p$ sú korene rovnice $|\boldsymbol{S}^{(real)} - \lambda\mathbf{\Sigma}| = 0$ práve vtady, ak sú koreňmi rovnice $|\mathbf{\Sigma}^{-\frac{1}{2}}\boldsymbol{S}^{(real)}\mathbf{\Sigma}^{-\frac{1}{2}} - \lambda\boldsymbol{I}| = 0$.*

Dôkaz. vyplýva z implikácií

$$|\boldsymbol{S}^{(real)} - \lambda\mathbf{\Sigma}| = 0 \Longleftrightarrow |\mathbf{\Sigma}^{\frac{1}{2}}(\mathbf{\Sigma}^{-\frac{1}{2}}\boldsymbol{S}^{(real)}\mathbf{\Sigma}^{-\frac{1}{2}} - \lambda\boldsymbol{I})\mathbf{\Sigma}^{\frac{1}{2}}| = 0$$

$$\Longleftrightarrow |\mathbf{\Sigma}^{\frac{1}{2}}||\mathbf{\Sigma}^{-\frac{1}{2}}\boldsymbol{S}^{(real)}\mathbf{\Sigma}^{-\frac{1}{2}} - \lambda\boldsymbol{I}||\mathbf{\Sigma}^{\frac{1}{2}}| = 0 \Longleftrightarrow |\mathbf{\Sigma}^{-\frac{1}{2}}\boldsymbol{S}^{(real)}\mathbf{\Sigma}^{-\frac{1}{2}} - \lambda\boldsymbol{I}| = 0,$$

prčom navyše platí (podľa Lemy 8.9 v texte "Plánovanie regresného experimentu"), že

$$tr(\mathbf{\Sigma}^{-\frac{1}{2}}\boldsymbol{S}^{(real)}\mathbf{\Sigma}^{-\frac{1}{2}}) = \sum_{=1}^{p}\lambda_i, \quad |\mathbf{\Sigma}^{-\frac{1}{2}}\boldsymbol{S}^{(real)}\mathbf{\Sigma}^{-\frac{1}{2}}| = \lambda_1 ... \lambda_p.$$

Lema 5^***.** *Pre* $x > 0$ *je* $\ln x + 1 - x \leqq 0$.

Dôkaz. : Pre funkciu $f(x) = e^{x-1} - x$ platí, že $f(0) = \frac{1}{e}$, $\quad f(1) = 0$. Pre $x > 0$ je minimum tejto funkcie v tom čísle x, pre ktoré $f'(x) = 0$. Teda $f'(x) = e^{x-1} - 1 = 0$, čiže $x = 1$. Pretože $f''(1) = e^{x-1}|_{x=1} = 1 > 0$, v bode $x = 1$ funkcia $f(x)$ nadobúda minimum. A minimálna hodnota funkcie je $f(1) = 0$. Pre $x > 0$ preto $e^{x-1} - x \geqq 0$, $\quad e^{x-1} \geqq x$, $\quad x - 1 \geqq \ln x$ a konečne $\ln x + 1 - x \leqq 0$.

Pomocou (A), (B), Lemy 4^* a Lemy 5^* dostávame:

$$l(\mathbf{x}_1, ..., \mathbf{x}_n; \overline{\mathbf{x}}, \mathbf{S}^{(real)}) - l(\mathbf{x}_1, ..., \mathbf{x}_n; \boldsymbol{\mu}, \boldsymbol{\Sigma}) =$$

$$= -\frac{n}{2} \ln(det\mathbf{S}^{(real)}) - \frac{n}{2} tr{\mathbf{S}^{(real)}}^{-1}\mathbf{S}^{(real)} - \frac{n}{2}(\overline{\mathbf{x}} - \overline{\mathbf{x}})'{\mathbf{S}^{(real)}}^{-1}(\overline{\mathbf{x}} - \overline{\mathbf{x}}) +$$

$$+ \frac{n}{2} \ln(det\boldsymbol{\Sigma}) + \frac{n}{2} tr\boldsymbol{\Sigma}^{-1}\mathbf{S}^{(real)} + \frac{n}{2}(\overline{\mathbf{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}) \geqq$$

$$\geqq -\frac{n}{2} \ln \frac{det\mathbf{S}^{(real)}}{det\boldsymbol{\Sigma}} - \frac{np}{2} + \frac{n}{2} tr\boldsymbol{\Sigma}^{-1}\mathbf{S}^{(real)} =$$

$$= -\frac{n}{2} \ln \frac{det\mathbf{S}^{(real)}}{det\boldsymbol{\Sigma}^{\frac{1}{2}} det\boldsymbol{\Sigma}^{\frac{1}{2}}} - \frac{np}{2} + \frac{n}{2} tr\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{S}^{(real)}\boldsymbol{\Sigma}^{-\frac{1}{2}} =$$

$$= -\frac{n}{2} \ln det(\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{S}^{(real)}\boldsymbol{\Sigma}^{-\frac{1}{2}}) - \frac{np}{2} + \frac{n}{2} tr\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{S}^{(real)}\boldsymbol{\Sigma}^{-\frac{1}{2}} =$$

$$= -\frac{n}{2} \ln(\lambda_1 ... \lambda_p) - \frac{np}{2} + \frac{n}{2}(\lambda_1 + ... + \lambda_p) =$$

$$= -\frac{n}{2}\{\ln \lambda_1 + ... + \ln \lambda_p + 1 + ... + 1 - \lambda_1 - ... - \lambda_p\} =$$

$$= -\frac{n}{2}\{(\ln \lambda_1 + 1 - \lambda_1) + ... + (\ln \lambda_p + 1 - \lambda_p)\} \geqq 0.$$