MNOHOROZMĚRNÉ NORMÁLNÍ ROZDĚLENÍ

1. CVIČENÍ S ŘEŠENÍM

DEFINICE: Nechť $\mathbf{Z} = (Z_1, \ldots, Z_r)^T$, kde $Z_i \sim \mathsf{N}(0,1)$ jsou nezávislé náhodné veličiny. Nechť $\boldsymbol{\mu} \in \mathbb{R}^n$ a $\mathbf{A}_{n\times r} \in \mathbb{R}^{n\times r}$ je matice. Pak řekneme, že $\mathbf{X} = \boldsymbol{\mu} + \mathbf{AZ}$ má n-**rozměrné normální rozdělení** $\mathsf{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, kde $\boldsymbol{\Sigma} = \mathbf{AA}^T$.

DISKUZE K DEFINICI:
- Je-li $\boldsymbol{\Sigma}_{n\times n}$ pozitivně semidefinitní matice s hodností $h(\boldsymbol{\Sigma}) = r \leq n$, pak existuje matice $\mathbf{A}_{n\times r}$ taková, že $h(\mathbf{A}) = r$ a $\mathbf{AA}^T = \boldsymbol{\Sigma}$.
- Rozdělení $\boldsymbol{\mu} + \mathbf{AZ}$ z definice nezávisí na konkrétní volbě $\mathbf{A}$, závisí pouze na $\boldsymbol{\Sigma} = \mathbf{AA}^T$.

Ukažte první bod (např. pomocí spektrálního rozkladu nebo Choleského rozkladu). Druhý bod zdůvodněte pomocí vhodné vlastnosti níže pro případ, kdy je $\mathbf{A}$ regulární (tj. pozitivně definitivní).

VLASTNOSTI: Ukažte, že platí:

1. Má-li $\mathbf{X} \sim \mathsf{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, pak $\mathsf{E}\,\mathbf{X} = \boldsymbol{\mu}$, $\mathsf{Var}\,\mathbf{X} = \boldsymbol{\Sigma}$.
2. Je-li $\mathbf{B} \in \mathbb{R}^{k\times n}$, $\boldsymbol{\mu}_2 \in \mathbb{R}^k$, pak $\mathbf{BX} + \boldsymbol{\mu}_2$ má rozdělení $\mathsf{N}_k(\mathbf{B}\boldsymbol{\mu} + \boldsymbol{\mu}_2, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T)$.

   Poznámka: Dokonce platí, že $\mathbf{X}$ má mnohorozměrné normální rozdělení $\Leftrightarrow$ $\mathbf{c}^T\mathbf{X}$ má normální rozdělení pro každé $\mathbf{c} \in \mathbb{R}^n$.
3. Existence hustoty:
   - Je-li $\boldsymbol{\Sigma}$ regulární, pak existuje hustota vzhledem k Lebesgueově míře a je tvaru
   $$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{\Sigma}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}, \quad \mathbf{x} \in \mathbb{R}^n.$$
   - Je-li $\boldsymbol{\Sigma}$ singulární, pak hustota vzhledem k Lebesgueově míře neexistuje.
4. Nechť $\mathbf{X} \sim \mathsf{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ a označme $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)^T$, kde $\mathbf{X}_1$ tvoří prvních k složek $\mathbf{X}$.

   (i) Marginální rozdělení $\mathbf{X}_1$ i $\mathbf{X}_2$ je normální, speciálně $\mathbf{X}_1 \sim \mathsf{N}_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$, kde $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)^T$ a $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{pmatrix}$.

   (ii) Je-li $\boldsymbol{\Sigma}_{12} = \mathbf{0}$, pak jsou $\mathbf{X}_1$ a $\mathbf{X}_2$ nezávislé.
5. Nechť $X \sim \mathsf{N}(0,1)$ a nechť $1/2 = \mathsf{P}(Z = 1) = \mathsf{P}(Z = -1)$ jsou nezávislé náhodné veličiny. Definujme $Y = Z \cdot X$. Ukažte, že

   (a) X i Y mají rozdělení $\mathsf{N}(0,1)$,

Riešenie: Nájdime distribučnú funkciu Y. Pre $y \in \mathbb{R}$ môžeme písať

$$\begin{aligned} \mathsf{P}\left(Y \leq y\right) &= \mathsf{P}\left(X \cdot Z \leq y\right) \\ &= \mathsf{P}\left(X \cdot Z \leq y \mid Z = 1\right) \mathsf{P}\left(Z = 1\right) + \mathsf{P}\left(X \cdot Z \leq y \mid Z = -1\right) \mathsf{P}\left(Z = -1\right) \\ &= \frac{1}{2}\left(\mathsf{P}\left(X \leq y\right) + \mathsf{P}\left(-X \leq y\right)\right) = \frac{1}{2}\left(\mathsf{P}\left(X \leq y\right) + \left(1 - \mathsf{P}\left(X < -y\right)\right)\right) \\ &= \mathsf{P}\left(X \leq y\right), \end{aligned}$$

kde sme využili symetriu a absolútnu spojitosť rozdelenia X.

(b) veličiny X a Y jsou nekorelované, ale závislé,

Riešenie: Máme

$$\mathsf{E}\left(X \cdot Y\right) = \mathsf{E}\left(X^2 \cdot Z\right) = \mathsf{E}\, X^2 \mathsf{E}\, Z = 0,$$

a tiež $\mathsf{E}\, X = 0$ a $\mathsf{E}\, Z = 0$, teda X a Y sú nekorelované. Sú však závislé, pretože napríklad pre obdĺžnik $[1,2] \times [-1,1] \subset \mathbb{R}^2$ platí

$$0 = \mathsf{P}\left(X \in [1,2], Y \in [-1,1]\right) \neq \mathsf{P}\left(X \in [1,2]\right) \mathsf{P}\left(Y \in [-1,1]\right) > 0.$$

(c) sdružené rozdělení X a Y není normální.

Riešenie: Ak by združené rozdelenie X a Y bolo normálne, z príkladu 5, časť (ii) by už musela z nekorelovanosti plynúť nezávislosť. Tým by sme boli v spore s časťou (b).

6. Nechť $\mathbf{X} \sim \mathsf{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, kde $\boldsymbol{\Sigma}$ je regulární. Pak

$$(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2_n.$$

Riešenie: Opäť použijeme reprezentáciu $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$ pre $\mathbf{Z}$ nezávislé $\mathsf{N}(0,1)$ veličiny a $\mathbf{A}$ čtvorcovú a regulárnú. Dostaneme

$$(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) = (\mathbf{A}\mathbf{Z})^T \left(\mathbf{A}\mathbf{A}^T\right)^{-1} (\mathbf{A}\mathbf{Z}) = \mathbf{Z}^T \mathbf{A}^T \left(\mathbf{A}^T\right)^{-1} \mathbf{A}^{-1} \mathbf{A}\mathbf{Z} = \mathbf{Z}^T \mathbf{Z},$$

kde vpravo je súčet štvorcov n nezávislých $\mathsf{N}(0,1)$ náhodných veličín. To je definícia χ^2_n rozdelenia.

Doplňující příklady k procvičení

D1. Nechť $\mathbf{X} \sim \mathsf{N}_n(\mathbf{0}, \boldsymbol{\Sigma})$. Pak náhodná veličina $||\mathbf{X}||^2 = \mathbf{X}^T\mathbf{X} = \sum_{i=1}^n X_i^2$ má stejné rozdělení jako $\sum_{i=1}^n \lambda_i Z_i^2$, kde Z_i jsou iid veličiny s $\mathsf{N}(0,1)$ rozdělením a $\lambda_1, \ldots, \lambda_n$ jsou vlastní čísla $\boldsymbol{\Sigma}$. Speciálně pak platí, že

- $\mathsf{E}\, ||\mathbf{X}||^2 = \mathrm{tr}(\boldsymbol{\Sigma})$,

– pokud má $\mathbf{\Sigma}$ vlastní čísla pouze 0 nebo 1, pak má veličina $\|\mathbf{X}\|^2$ rozdělení χ^2 se stupni volnosti rovnými $\mathrm{h}(\mathbf{\Sigma}) = \mathrm{tr}(\mathbf{\Sigma})$.

Riešenie: Použime reprezentáciu $\mathbf{X} = \mathbf{AZ}$ pre $\mathbf{Z}$ nezávislé $\mathsf{N}(0,1)$ veličiny, a spektrálny rozklad matice $\mathbf{\Sigma} = \mathbf{Q}$, kde $\mathbf{Q}$ je ortonormálna a $\mathbf{\Lambda}$ diagonálna. Pre $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}^{1/2}$

$$\mathbf{X}^T\mathbf{X} = (\mathbf{AZ})^T(\mathbf{AZ}) = \mathbf{Z}^T\left(\mathbf{A}^T\mathbf{A}\right)\mathbf{Z} = \mathbf{Z}^T\mathbf{\Lambda}^{1/2}\mathbf{Q}^T\mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Z} = \sum_{i=1}^{n}\lambda_i Z_i^2,$$

kde λ_i sú diagonálne členy matice $\mathbf{\Lambda}$, teda vlastné čísla matice $\mathbf{\Sigma}$.

D2. Nalezněte asymptotickou aproximaci rozdělení χ^2_n pro velké n pomocí normálního rozdělení.

Riešenie: Vieme, že rozdelenie χ^2_n môžeme reprezentovať ako súčet n nezávislých náhodných veličín s rozdelením Z_i^2, kde $Z_i \sim \mathsf{N}(0,1)$. Označme $Y_i = Z_i^2$. Potom Y_i sú nezávislé, rovnako rozdelené náhodné veličiny so strednou hodnotou $\mathsf{E}\,Y_i = \mathsf{E}\,Z_i^2 = 1$ a rozptylom $\mathsf{Var}\,Y_i = \mathsf{E}\,Z_i^4 - 1 = 2$. Centrálna limitná veta dáva

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i - 1\right) \xrightarrow[n\to\infty]{\mathsf{d}} \mathsf{N}(0,2).$$

To môžeme prepísať ako

$$\frac{\sum_{i=1}^{n} Z_i^2 - n}{\sqrt{2n}} \xrightarrow[n\to\infty]{\mathsf{d}} \mathsf{N}(0,1),$$

čo znamená že pre n veľké, χ^2_n bude rozdelenie blízke $\mathsf{N}(n, 2n)$.

D3. Pro $n = 2$ ukažte, jak lze generovat náhodný vektor z rozdělení $\mathsf{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ pro nějaké zadané $\boldsymbol{\mu} = (\mu_1, \mu_2)^T \in \mathbb{R}^2$ a $\mathbf{\Sigma}_{2\times 2} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ pozitivně semidefinitní, umíme-li generovat náhodné veličiny z $\mathsf{N}(0,1)$.

Riešenie: Stačí nájsť maticu $\mathbf{A}$ takú, že $\mathbf{A}\mathbf{A}^T = \mathbf{\Sigma}$ a použiť reprezentáciu $\mathbf{X} = \boldsymbol{\mu} + \mathbf{AZ}$. Spektrálny rozklad $\mathbf{\Sigma}$ vedie k symetrickej odmocninovej matici $\mathbf{A} = \mathbf{\Sigma}^{1/2}$, ktorá má ale všeobecne pomerne zložitý tvar. Jednoduchšia (nesymetrická) matica s danou vlastnosťou je napr.

$$\mathbf{A} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1-\rho^2}\sigma_2 \end{pmatrix}.$$

Je jednoduché overiť, že platí $\mathbf{A}\mathbf{A}^T = \mathbf{\Sigma}$. Môžeme teda voliť

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \mathbf{A}\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix},$$

čiže

$$X_1 = \mu_1 + \sigma_1 Z_1,$$
$$X_2 = \mu_2 + \sigma_2 \left(\rho Z_1 + \sqrt{1-\rho^2} Z_2\right).$$

Opakování

Spektrální rozklad matice. Nechť $\mathbf{\Sigma}_{n\times n}$ je symetrická reálná matice. Pak existuje $\mathbf{Q}$ orthonormální a $\mathbf{\Lambda}$ diagonální matice takové, že

$$\mathbf{\Sigma} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T.$$

χ^2 rozdělení. Nechť $Z_1, \ldots, Z_n$ jsou nezávislé náhodné veličiny s rozdělením $\mathsf{N}(0,1)$. Pak $Y = \sum_{i=1}^n Z_i^2$ má χ^2 rozdělení s n stupni volnosti.

Význam normálního rozdělení: Mnohorozměrná centrální limitní věta.

Mějme náhodný výběr $\mathbf{X}_1, \ldots, \mathbf{X}_n$ z p-rozměrného rozdělení se střední hodnotou $\boldsymbol{\mu}$ a konečnou rozptylovou maticí $\mathbf{\Sigma}$. Pak platí

$$\sqrt{n}\,(\overline{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow[n\to\infty]{\mathsf{d}} \mathsf{N}_p(\mathbf{0}, \mathbf{\Sigma}).$$