Alkalmazott statisztika

1. feladatsor: lineáris algebrai ismétlés, többdimenziós normális

1. Egy szimmetrikus $\Sigma \in \mathbb{R}^{k\times k}$ mátrix pozitív szemidefinit (vagy nemnegatív definit), ha tetszőleges $\mathbf{x} \in \mathbb{R}^k$ esetén $\mathbf{x}^\top \Sigma \mathbf{x} \geq 0$. Akkor pozitív definit, ha egyenlőség csak $\mathbf{x} = 0$ esetén áll fenn. Igazoljuk, hogy minden kovarianciamátrix pozitív szemidefinit!

2. Legyen Σ pozitív definit szimmetrikus mátrix. Igazoljuk, hogy ha $\Sigma\mathbf{x} = \lambda\mathbf{x}$, akkor $\Sigma^{-1}\mathbf{x} = \lambda^{-1}\mathbf{x}$, és Σ^{-1} pozitív definit!

3. Egy négyzetes mátrix *nyoma* a főátlóban lévő elemeinek összege, azaz $\mathrm{tr}(A) = \sum_{i=1}^k a_{ii}$. Igazoljuk, hogy

1. $\mathrm{tr}(BC) = \mathrm{tr}(CB)$, ahol $B \in \mathbb{R}^{k\times \ell}$, $C \in \mathbb{R}^{\ell\times k}$;
2. $\mathbf{x}^\top A\mathbf{x} = \mathrm{tr}(\mathbf{x}^\top A\mathbf{x}) = \mathrm{tr}(A\mathbf{x}\mathbf{x}^\top)$, ahol $A \in \mathbb{R}^{k\times k}$, $\mathbf{x} \in \mathbb{R}^k$;
3. $\mathrm{tr}A = \sum_{i=1}^k \lambda_i$, ahol λ_i-k az A sajátértékei.

4. Legyen $\mathbf{X}$ véletlen vektor *szimmetrikus normális eloszlású*, azaz minden komponens azonos eloszlású, és bármely két (különböző) kovarianciája ugyanakkora. Adjuk meg a korrelációs mátrix spektrálfelbontását!

5. Határozzuk meg a $N_2(\mathbf{m}, \Sigma)$ kétdimenziós normális eloszlás sűrűségfüggvényének szintvonalait, ahol

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{pmatrix}.$$

R **6.** Szimuláljunk kétdimenziós normális eloszlást

$$\Sigma_1 = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad \Sigma_3 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},$$

kovarianciamátrixokkal. Rajzoltassuk ki az eredményt, és nézegessük!

7. Legyen $\mathbf{X} \sim N_2(\mathbf{m}, \Sigma)$. Adjuk meg a komponensek tetszőleges $aX_1 + bX_2$ lineáris kombinációinak eloszlását!

8. Legyen $\mathbf{X} \sim N_3(\mathbf{m}, \Sigma)$. Adjuk meg $(X_1 - X_2, X_2 - X_3)^\top$ eloszlását!

9. Legyen $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)^\top \sim N_k(\mathbf{m}, \Sigma)$, ahol $\mathbf{m} = (\mathbf{m}_1, \mathbf{m}_2)^\top$,

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Ekkor $\mathbf{X}_1$ és $\mathbf{X}_2$ pontosan akkor függetlenek, ha $\Sigma_{12} = \mathbf{0}$.

10. Igazoljuk, hogy a standard normális eloszlás forgatásinvariáns; azaz ha $\mathbf{Z} \sim N_k(0, I_k)$, és $U \in \mathbb{R}^{k\times k}$ ortogonális, akkor $U\mathbf{Z} \sim N_k(0, I_k)$.

11. Legyen $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)^\top \sim N_k(\mathbf{m}, \Sigma)$, ahol $\mathbf{m} = (\mathbf{m}_1, \mathbf{m}_2)^\top$,

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}, \end{pmatrix}$$

ahol $|\Sigma_{22}| = \det(\Sigma_{22}) > 0$. Igazoljuk, hogy $\mathbf{X}_1$ vektor $\mathbf{X}_2 = \mathbf{x}_2$ feltételre vett feltételes eloszlása

$$N_{k_1}\left(\mathbf{m}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mathbf{m}_2)), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right).$$

12. Legyenek $\mathbf{X}_1, \ldots, \mathbf{X}_n$ független, azonos eloszlású véletlen k-dimenziós vektorváltozók $\mathbf{m}$ várható értékkel és Σ kovarianciamátrixszal. Határozzuk meg az $\mathbf{X}_1 + \ldots + \mathbf{X}_n$ véletlen vektor várható érték és kovarianciamátrixát!

13. *Többdimenziós Steiner-formula.* Igazoljuk, hogy tetszőleges $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^k$, $\mathbf{v} \in \mathbb{R}^k$ esetén

$$\sum_{i=1}^{n}(\mathbf{x}_k - \mathbf{v})(\mathbf{x}_k - \mathbf{v})^\top = \sum_{i=1}^{n}(\mathbf{x}_k - \overline{\mathbf{x}})(\mathbf{x}_k - \mathbf{x})^\top + n(\overline{\mathbf{x}} - \mathbf{v})(\overline{\mathbf{x}} - \mathbf{v})^\top,$$

ahol $\overline{\mathbf{x}} = n^{-1}\sum_{i=1}^{n}\mathbf{x}_i$.

14. Az $F(x) = \mathbf{P}(X \le x)$ eloszlásfüggvényhez tartozó kvantilisfüggvény

$$Q(s) = \inf\{x : F(x) \ge s\}, \ s \in (0,1).$$

Ha F szigorúan monoton növő folytonos függvény, akkor Q éppen F inverze.

Mutassuk meg, hogy $Q(s) \le x$ pontosan akkor, ha $s \le F(x)$.

15. *Kvantilis transzformáció.* Legyen U egyenletes eloszlású $[0,1]$-en. Ekkor $Q(U)$ véletlen változó eloszlásfüggvénye F, ahol Q az F kvantilisfüggvénye.

16. Határozzuk meg az exponenciális eloszlás kvantilisfüggvényét!

17. *Pareto-eloszlás.* Az $\alpha > 0$ paraméterű Pareto-eloszlás eloszlásfüggvénye $F_\alpha(x) = 1 - x^{-\alpha}$ ha $x \ge 1$, és 0, különben. Határozzuk meg a kvantilisfüggvényt.

18. *Q-Q plot.* Legyen Q_F és Q_G az F és G eloszlásfüggvényekhez tartozó kvantilisfüggvény. Ekkor az F és G egymásra vonatkozó Q-Q plot grafikonja a $\{(Q_F(u), Q_G(u)) : u \in (0,1)\}$ görbe.

Legyenek F és G folytonos eloszlásfüggvények. Igazoljuk, hogy ha az eloszlásokhoz tartozó elméleti Q-Q plot lineáris, akkor valamilyen a, b számokra $F(ax + b) = G(x)$, minden x-re; azaz a két véletlen változó egymás lineáris transzformáltja.

19. Gondoljuk végig, hogy mit jelent az $X \sim F$ és $Y \sim G$ véletlen változók viszonyára, ha az $F - G$ Q-Q plot a kis értékekre / nagy értékekre az $x = y$ egyenes alatt van?

R **20.** Vegyünk mintát különböző eloszlásokból (normális, exponenciális, Cauchy), és vegyük a minta normális szerinti Q-Q plotját (qqnorm). Nézegessük az eredményt!

R **21.** Szimuláljunk Pareto-eloszlást a (korábban meghatározott) kvantilisfüggvénye segítségével! Rajzoljuk ki más eloszlásokkal vett Q-Q plotját!