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. Adjuk meg az

$$\Sigma = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$$

mátrix sajátértékeit, sajátvektorait, inverzét, és spektrálfelbontását!

5. Legyen $\mathbf{X}$ kovarianciamátrixa

$$\Sigma = \begin{pmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{pmatrix}.$$

i) Adjuk meg az R korrelációmátrixot és a megfelelő D_0 diagonális mátrixot (azaz $\Sigma = D_0RD_0$)!

ii) Adjuk meg X_1 és $(X_2 + X_3)/2$ kovarianciáját és korrelációját!

6. Az X_1, X_2, X_3 változók várható értékeivel és kovarianciáival fejezzük ki a következő változók szórásnégyzetét: a) $X_1 - 2X_2$; b) $-X_1 + X_3$; c) $2X_1 + X_2 - 3X_3$.

7. Legyenek X_1, X_2 független véletlen változók, 0 várható értékkel és 1 szórással. Tekintsük az $Y_1 = X_1 + X_2$, $Y_2 = 2X_1$, $Y_3 = X_1 - 3X_2$ változókat! Számítsuk ki az $\mathbf{Y} = (Y_1, Y_2, Y_3)^\top$ vektorváltozó kovarianciamátrixát!

8. Legyen $\mathbf{X}$ kovarianciamátrixa

$$\Sigma = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 5 \end{pmatrix}.$$

Igazoljuk, hogy X_1, X_2, X_3 lineárisan függőek! Legyen $\mathbf{EX} = (0, 1, 0)^\top$. Milyen altérre koncentrált $\mathbf{X}$?

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

10. Legyen (X_1, X_2) kétváltozós normális eloszlású. Adjuk meg explicit alakban az X_1 véletlen változó X_2-re vett feltételes sűrűségfüggvényét!

11. 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 **12.** 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!

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

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

15. 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 $\Sigma_{12} = \mathbf{0}$.

16. 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

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

17. 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!

18. 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 $U\mathbf{Z} \sim N_k(0, I_k)$.

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