Lektion 8

TAMS24 – Statistisk teori

Skriven av Oliver Wettergren
oliwe188@student.liu.se
https://www.instagram.com/olwettergren/

Lektionsgenomgång

n-dimensionell fördelning, $X_1, X_2, \ldots, X_n$ s.v

$$C = \left[\operatorname{cov}(X_i, X_j) \right]_{i,j = 1, \ldots, n}$$ kovariansmatris

$$\mu = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{bmatrix}$$ väntevärdesvektor

$$\mathbb{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$ n-dimensionell normalfördelad

Om

$$f_{\mathbb{X}}(x_1, \ldots, x_n) = f_{\mathbb{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{2} \sqrt{\det C}} \cdot \left\{ -\frac{1}{2} (\mathbf{x}-\mu)^T C_{\mathbb{X}}^{-1} (\mathbf{x}-\mu) \right\},$$

$\mathbf{x} \in \mathbb{R}.$

Räkneregler

(1) $\operatorname{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$ (förkortningsformel)

där

$$E(X \cdot Y) = \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y \, f(x,y) \, dx dy, & X, Y \text{ kontinuerliga} \\ \sum_{\text{alla } k} \sum_{\text{alla } j} j \, k \, p(j,k), & X, Y \text{ diskreta} \end{cases}$$

(2) $V(aX + bY + c) = C(aX + bY, aX + bY) =$

$= a^2 C(X, X) + 2ab \, C(X, Y) + b^2 C(Y, Y) =$

$= a^2 V(\mathbf{X}) + 2ab\, C(\mathbf{X},\mathbf{Y}) + b^2\, V(\mathbf{Y}).$

(3) $\begin{bmatrix} Y_1 \\ \vdots \\ Y_k \end{bmatrix} = A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$ // $\mathbf{Y} = A\mathbf{X} + b$

där

$$\mathbf{X} \sim N(C_{\mathbf{X}}, \mu_{\mathbf{X}})$$

medför

$$\mathbf{Y} \sim N(C_{\mathbf{Y}}, \mu_{\mathbf{Y}})$$

där

$$C_{\mathbf{Y}} = A C_{\mathbf{X}} A^T, \quad \mu_{\mathbf{Y}} = A\mu_{\mathbf{X}} + b$$

(4) $\mathbf{X}, \mathbf{Y}$ gemensamt normalfördelade.
Det gäller att

$$\mathbf{X}, \mathbf{Y} \text{ oberoende} \iff C(\mathbf{X},\mathbf{Y}) = 0$$

OBS: "$\Leftarrow$" är fel om $\mathbf{X}, \mathbf{Y}$ inte gemensamt normalfördelad.

G.1.1.

X_1, X_2 oberoende $N(0,1)$ medför

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{\mu_X}, \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{C_X} \right)$$

$$Y = AX + b$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\mu_Y = A\mu_X = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0$$

$$C_Y = A C_{\underline{X}} A^T = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix}$$

$$\det C_Y = \frac{1}{9} \begin{bmatrix} 5 & -1 \\ -1 & 2 \end{bmatrix}$$

$$f_Y = \frac{1}{\sqrt{(2\pi)^2}\sqrt{9}} \cdot \exp\left\{ -\frac{1}{2} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}^T \frac{1}{9} \begin{bmatrix} 5 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right\} =$$

$$= \frac{1}{6\pi} \exp\left\{ -\frac{1}{2} (y_1 \;\; y_2) \cdot \frac{1}{9} \begin{pmatrix} 5y_1 - y_2 \\ -y_1 + 2y_2 \end{pmatrix} \right\}$$

$$= \frac{1}{6\pi} \exp\left\{-\frac{1}{18}\left(5y_1^2 - y_1y_2 - y_1y_2 + 2y_2^2\right) =\right.$$

$$= \frac{1}{6\pi} \exp\left\{-\frac{1}{18}\left(5y_1^2 - 2y_1y_2 + 2y_2^2\right)\right\}$$

G1.2

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 60 \\ 60 \\ 60 \end{pmatrix}, \begin{pmatrix} 100 & 80 & 20 \\ 80 & 100 & 10 \\ 20 & 10 & 80 \end{pmatrix}\right)$$

$$Y = (X_1 + X_2 + 2X_3)/4$$

a) Find the distribution for Y

$$X_1 \sim N(60, 10)$$
$$X_2 \sim N(60, 10)$$
$$X_3 \sim N(60, \sqrt{80})$$

$$Y = (X_1 + X_2 + 2X_3)/4 = \frac{1}{4}\begin{pmatrix} 1 & 1 & 2 \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix} \mathbf{X} \sim N.$$

$$\mu_Y = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix} \mu_{\mathbf{X}} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} 60 \\ 60 \\ 60 \end{pmatrix} = 60$$

$$\sigma_Y^2 = V(Y) = C\left(\frac{1}{4}X_1 + \frac{1}{4}X_2 + \frac{1}{2}X_3, \frac{1}{4}X_1 + \frac{1}{4}X_2 + \frac{1}{2}X_3\right) =$$

$$= \frac{1}{16} C(X_1, X_1) + \frac{1}{16} C(X_1, X_2) + \frac{1}{8} C(X_1, X_3) + \dots + \dots =$$

$$= 50$$

$$Y \sim N(60, \sqrt{50})$$

$$0.9 = P(Y > a) = P\left(N(0,1) > \frac{a-60}{\sqrt{50}}\right) = 1 - \Phi\left(\frac{a-60}{\sqrt{50}}\right) \Rightarrow$$

$$-\Phi\left(\frac{a-60}{\sqrt{50}}\right) = 0.9 \Rightarrow -\frac{(a-60)}{\sqrt{50}} = 1.28$$

$$\Leftrightarrow a = -\sqrt{50} \cdot 1.28 + 60$$

b) X_1, X_2, X_3 oberoende

$$X \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) = N(\mu_x, C_x)$$

$$U = X_1 - 2X_2 + X_3$$

$$V = c_1 X_1 + c_2 X_2 + c_3 X_3$$

$$W = \begin{bmatrix} U \\ V \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & -2 & 1 \\ c_1 & c_2 & c_3 \end{bmatrix}}_{A} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

Sökt är villkoret sådana att

$$\mathrm{Cov}(U, V) = 0$$

$$\mu_U = A\mu_X + b = A\begin{bmatrix}0\\0\\0\end{bmatrix} + \begin{bmatrix}0\\0\\0\end{bmatrix} = \bar{0}$$

$$C_U = A C_X A^T = \underbrace{\begin{bmatrix}1 & -2 & 1\\ C_1 & C_2 & C_3\end{bmatrix}}\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & C_1\\-2 & C_2\\1 & C_3\end{bmatrix} =$$

$$= \begin{bmatrix}6 & C_1 - 2C_2 + C_3\\ C_1 - 2C_2 + C_3 & C_1^2 + C_2^2 + C_3^2\end{bmatrix}$$

Det sökta villkoret

$$C_1 - 2C_2 + C_3 = 0.$$

G.1.6.

Random variables

$X_1, X_2, \dots, X_5$, independent

$X \sim N(10, 4)$

$$Y_1 = \frac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5)$$

$$Y_2 = 4X_1 + X_2 - X_3 - X_4 - X_5$$

a) $$\begin{bmatrix}Y_1\\Y_2\end{bmatrix} = \underbrace{\begin{bmatrix}1/5 & 1/5 & 1/5 & 1/5 & 1/5\\ 4 & 1 & -1 & -1 & -1\end{bmatrix}}_{A}\begin{bmatrix}X_1\\ \vdots\\ X_5\end{bmatrix} \qquad b = \begin{bmatrix}0\\0\end{bmatrix}$$

$$\mu_X = \begin{bmatrix}10\\ \vdots\\ 10\end{bmatrix}, \quad C_X = \begin{bmatrix}4 & & 0\\ & \ddots & \\ 0 & & 4\end{bmatrix}$$

$$\mu_Y = A\mu_X + b = \begin{bmatrix} 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 4 & 1 & -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 10 \\ \vdots \\ 10 \end{bmatrix} = \underline{\underline{\begin{bmatrix} 10 \\ 20 \end{bmatrix}}}$$

$$C_Y = A C_X A^T = \dots = \underline{\underline{\begin{bmatrix} 0.8 & 1.6 \\ 1.6 & 80 \end{bmatrix}}}$$

$$Y \sim N\left(\begin{bmatrix} 10 \\ 20 \end{bmatrix}, \begin{bmatrix} 0.8 & 1.6 \\ 1.6 & 80 \end{bmatrix} \right)$$

b) $P(Y_1 > Y_2) = P(Y_1 - Y_2 > 0)$

Vi behöver fördelningen $Y_1 - Y_2$

Som linjärkombination av de gemensamt normalfördelade Y_1, Y_2 är $Y_1 - Y_2$ också normalfördelade

$$N(\mu_{Y_1 - Y_2}, \sigma_{Y_1 - Y_2})$$

$$\mu_{Y_1 - Y_2} = E(Y_1 - Y_2) = \underbrace{E(Y_1)}_{10} - \underbrace{E(Y_2)}_{20} = -10$$

$$\sigma^2_{Y_1 - Y_2} = V(Y_1 - Y_2) = V(Y_1) - 2C(Y_1, Y_2) + V(Y_2) =$$

$$= 0.8 - 2 \cdot 1.6 + 80 = 77.6$$

$$P(Y_1 > Y_2) = P(Y_1 - Y_2 > 0) = P\left(\frac{Y_1 - Y_2 - (-10)}{\sqrt{77.6}} > \frac{-(-10)}{\sqrt{77.6}}\right) =$$

$$= P\left(N(0,1) > \frac{10}{\sqrt{77.6}}\right) = 1 - \Phi\left(\underbrace{\frac{10}{\sqrt{77.6}}}_{1.13}\right) = 1 - 0.8686 =$$

$$= \underline{\underline{0.1314}}$$

c) $\rho = \dfrac{C(Y_1, Y_2)}{\sqrt{V(Y_1)\cdot V(Y_2)}} = \dfrac{1.6}{\sqrt{0.8\cdot 80}} = \dfrac{1.6}{8} = 0.2$

6.1.711

Three dimension normal distribution with mean vector

$$\mu = \begin{pmatrix} 71 \\ 53 \\ 18 \end{pmatrix} \qquad C = \begin{pmatrix} 100 & 64 & 38 \\ 64 & 64 & 28.8 \\ 38 & 28.8 & 16 \end{pmatrix}$$

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{bmatrix} 71 \\ 53 \\ 18 \end{bmatrix}, \begin{bmatrix} 100 & 64 & 38 \\ 64 & 64 & 28.8 \\ 38 & 28.8 & 16 \end{bmatrix} \right)$$

$\hat{X}_3 = a + bX_1 + cX_2$

(i) $E(\hat{X}_3) = E(X_1)$

(ii) $V(X_3 - \hat{X}_3)$ is minimized.