# Stokastiska vektorer och multivariat normalfördelning

Johan Thim (johan.thim@liu.se)

13 november 2018

## 1 Repetition

**Definition.** Låt $X$ och $Y$ vara stokastiska variabler med $E(X) = \mu_X$, $V(X) = \sigma_X^2$, $E(Y) = \mu_Y$ samt $V(Y) = \sigma_Y^2$. Kovariansen $C(X,Y)$ definieras enligt

$$C(X,Y) = E\big((X-\mu_X)(Y-\mu_Y)\big)$$

och korrelationen mellan $X$ och $Y$ enligt

$$\rho(X,Y) = \frac{C(X,Y)}{\sigma_X\,\sigma_Y}.$$

Både kovarians och korrelation är ett mått på linjärt beroende mellan $X$ och $Y$ där korrelationen är normerad så det går att jämföra olika fall. Vi listar lite kända egenskaper.

(i) Om $C(X,Y) = 0$ kallas $X$ och $Y$ för okorrelerade.

(ii) $C(X,Y) = E(XY) - E(X)E(Y)$.

(iii) Om $X$ och $Y$ är oberoende så är $C(X,Y) = 0$.

(iv) $C(X,X) = V(X)$.

(v) $C\left(a_0 + \sum_{i=1}^{m} a_i X_i,\, b_0 + \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j C(X_i, Y_j)$.

(vi) $|\rho(X,Y)| \leq 1$ med likhet om och endast om det finns ett linjärt samband mellan $X$ och $Y$.

Observera att $C(X,Y) = 0$ *inte* nödvändigtvis innebär oberoende. Låt till exempel $X$ vara rektangelfördelad enligt $X \sim \mathrm{Re}(-1,1)$ och definiera $Y = X^2$. Uppenbarligen beroende variabler, men

$$C(X,Y) = E(XY) - E(X)E(Y) = E(X^3) - 0\cdot E(Y) = E(X^3) = \int_{-1}^{1} x^3 \cdot \frac{1}{2}\,dx = 0,$$

så $X$ och $Y$ är okorrelerade.

## 2 Vektorer av stokastiska variabler

Låt $\boldsymbol{X} = (X_1 \; X_2 \; \cdots \; X_n)^T$ vara en vektor vars komponenter är stokastiska variabler. Vi strävar efter att skriva vektorer som kolonnvektorer. Det faller sig naturligt att definiera väntevärdet av $\boldsymbol{X}$ genom **väntevärdesvektorn**

$$E(\boldsymbol{X}) = (E(X_1) \; E(X_2) \; \cdots \; E(X_n))^T.$$

På samma sätt definierar vi väntevärdet av en matris av stokastiska variabler. Variansen blir lite konstigare så vi introducerar **kovariansmatrisen** mellan två vektorer (av samma dimension). Låt $\boldsymbol{Y} = (Y_1, Y_2, \dots, Y_n)^T$ och definiera $C(\boldsymbol{X}, \boldsymbol{Y})$ enligt

$$C(\boldsymbol{X}, \boldsymbol{Y}) = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix} = \begin{pmatrix} C(X_1, Y_1) & C(X_1, Y_2) & \cdots & C(X_1, Y_n) \\ C(X_2, Y_1) & C(X_2, Y_2) & \cdots & C(X_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(X_n, Y_1) & C(X_n, Y_2) & \cdots & C(X_n, Y_n) \end{pmatrix}$$

där $c_{ij}$ är kovariansen mellan $X_i$ och $Y_j$.
En stor anledning att blanda in vektorer och matriser är givetvis att få tillgång till maskineriet från linjär algebra. Kovariansen mellan två vektorer $\boldsymbol{X}$ och $\boldsymbol{Y}$ kan då lite mer kompakt skrivas

$$C(\boldsymbol{X}, \boldsymbol{Y}) = E((\boldsymbol{X} - E(\boldsymbol{X}))(\boldsymbol{Y} - E(\boldsymbol{Y}))^T) = E(\boldsymbol{X}\boldsymbol{Y}^T) - E(\boldsymbol{X})E(\boldsymbol{Y})^T,$$

där $(\,\cdot\,)^T$ innebär transponering. En produkt $A = \boldsymbol{x}\boldsymbol{y}^T$ brukar kallas för den yttre produkten och består av element $(a)_{ij} = x_i y_j$, $i, j = 1, 2, \dots, n$. Detta är alltså *inte* skalärprodukten $(\boldsymbol{X}^T\boldsymbol{Y})$. Låt $A, B \in \mathbf{R}^{n \times n}$ vara matriser. Då är $A\boldsymbol{X}$ en linjärkombination av $X_1, X_2, \dots, X_n$ och $B\boldsymbol{Y}$ en linjärkombination av $Y_1, Y_2, \dots, Y_n$. Dessutom kan *alla* linjärkombinationer skrivas på detta sätt. Vidare gäller nu tack varje linjäriteten att $E(A\boldsymbol{X}) = AE(\boldsymbol{X})$ och

$$\begin{aligned} C(A\boldsymbol{X}, B\boldsymbol{Y}) &= E(A\boldsymbol{X}(B\boldsymbol{Y})^T) - E(A\boldsymbol{X})E(B\boldsymbol{Y})^T = AE(\boldsymbol{X}\boldsymbol{Y}^T)B^T - AE(\boldsymbol{X})E(\boldsymbol{Y})^T B^T \\ &= AC(\boldsymbol{X}, \boldsymbol{Y})B^T. \end{aligned}$$

Notationen $\mathrm{cov}(\boldsymbol{X}, \boldsymbol{Y})$ är också vanligt förekommande, och i fallet då $\boldsymbol{Y} = \boldsymbol{X}$ skriver vi ofta

$$C_{\boldsymbol{X}} = \mathrm{cov}(\boldsymbol{X}) = C(\boldsymbol{X}, \boldsymbol{X}).$$

**Exempel**

Låt $\boldsymbol{X} = (X_1 \; X_2)^T$ vara en stokastisk variabel med $E(\boldsymbol{X}) = (1 \; 2)^T$ och $C_{\boldsymbol{X}} = \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}$.
Hitta en prediktor $\widehat{X}_2 = aX_1 + b$ så att $E(\widehat{X}_2) = E(X_2)$ och $V(X_2 - \widehat{X}_2)$ är minimal.

**Lösning**. Vi ser direkt att

$$E(aX_1 + b) = aE(X_1) + b = a + b \quad \text{och} \quad E(X_2) = 2,$$

så $a + b = 2$. Vidare gäller att

$$X_2 - (aX_1 + b) = \begin{pmatrix} -a & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} - b,$$

så

$$\begin{aligned} V(X_2 - aX_1 - b) &= V\left( \begin{pmatrix} -a & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \right) = \begin{pmatrix} -a & 1 \end{pmatrix} C_{\boldsymbol{X}} \begin{pmatrix} -a \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} -a & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} -a \\ 1 \end{pmatrix} = a^2 + 4a + 4 = (a + 2)^2. \end{aligned}$$

Minimum sker uppenbarligen när $a = -2$, vilket ger att $b = 4$.

# 3 Skattningar för kovarians och korrelation

Om vi har ett stickprov $(x_k, y_k)$, $k = 1, 2, \ldots, n$, där $(X_k, Y_k)$ är stokastiska variabler med samma fördelning, så skattar vi kovariansen $C$ med

$$\widehat{c} = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y})$$

och korrelationen med

$$\widehat{\rho} = \frac{\widehat{c}}{s_x\, s_y} = \frac{\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y})}{\left(\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})^2\right)^{1/2}\left(\frac{1}{n-1}\sum_{k=1}^{n}(y_k - \overline{y})^2\right)^{1/2}}.$$

Av tradition betecknar man ofta $\widehat{\rho} = r$. En naturlig fråga i detta skede är om vi kan säga något om fördelningen för den skatta korrelationen under något lämpligt antagande om det slumpmässiga stickprovet. Vi återkommer i fallet med normalfördelning i nästa avsnitt.

# 4 Multivariat normalfördelning

*"Pain has a face. Allow me to show it to you."*
–Pinhead

Vi har stött på den flerdimensionella normalfördelningen tidigare, men vi kan formulera det hela lite mer kompakt på följande sätt.

**Multivariat normalfördelning**

**Definition.** Vi säger att $\boldsymbol{Y}$ har en **multivariat normalfördelning** om det finns en konstant vektor $\boldsymbol{\mu} \in \mathbf{R}^n$ och en konstant matris $\boldsymbol{A} \in \mathbf{R}^{n\times m}$ så att $\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{AX}$, där $\boldsymbol{X}$ är en vektor med stokastiska variabler, $\boldsymbol{X} = (X_1\ X_2\ \cdots\ X_m)^T$, och $X_i \sim N(0,1)$ är oberoende.

Är definitionen vettig? Ja, den reducerar åtminstone till det förväntade resultatet om $n = 1$: $Y = \mu + \sigma^2 X$ där $X \sim N(0,1)$. Vidare gäller så klart att

$$E(\boldsymbol{Y}) = \boldsymbol{\mu} + E(\boldsymbol{AX}) = \boldsymbol{\mu}$$

och

$$C_{\boldsymbol{Y}} = \boldsymbol{A}C_{\boldsymbol{X}}\boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{A}^T$$

eftersom $C_{\boldsymbol{X}}$ är identitetsmatrisen (variablerna är oberoende om har varians 1).

**Exempel**

Låt $\boldsymbol{X} \sim \mathrm{N}\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\right)$. Bestäm väntevärde och varians för $Y = X_1 + X_2$.

**Lösning**. Vi skriver $Y = (1\ 1)(X_1\ X_2)^T = \boldsymbol{A}\boldsymbol{X}$. Då blir

$$E(Y) = \boldsymbol{A}E(X) = (1\ 1)(1\ \ -1)^T = 0$$

och

$$C_Y = \boldsymbol{A}C_{\boldsymbol{X}}\boldsymbol{A}^T = (1\ 1)\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}(1\ 1)^T = 5.$$

Är det klart att $Y \sim \mathrm{N}(0, \sqrt{5})$?

**Sats.** Om $\boldsymbol{Y}$ har väntevärdesvektorn $\boldsymbol{\mu}$ och en kovariansmatris $\boldsymbol{\Sigma}$ som uppfyller att $\det\boldsymbol{\Sigma} \neq 0$ så gäller att $\boldsymbol{Y}$ har multivariat normalfördelning om och endast om $\boldsymbol{Y}$ har den simultana täthetsfunktionen

$$f_{\boldsymbol{Y}}(y_1, y_2, \ldots, y_n) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right), \quad \boldsymbol{y} \in \mathbf{R}^n.$$

**Bevis**. Eftersom kovariansmatrisen $\boldsymbol{\Sigma}$ alltid är positivt semidefinit (varför?) och vi antar att determinanten $|\boldsymbol{\Sigma}| := \det\boldsymbol{\Sigma} \neq 0$, så är $\boldsymbol{\Sigma}$ positivt semidefinit och då finns alltid en inverterbar matris $\boldsymbol{A} \in \mathbf{R}^{n\times n}$ sådan att $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^T$. Definiera $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{\mu}$, där $\boldsymbol{X} = (X_1\ X_2\ \cdots\ X_n)^T$ och $X_k \sim N(0,1)$ är oberoende. Täthetsfunktionen för $\boldsymbol{X}$ ges då av

$$f_{\boldsymbol{X}}(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x}\right), \quad \boldsymbol{x} \in \mathbf{R}^n.$$

Enligt transformationssatsen för flerdimensionella stokastiska variabler så kommer

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}\left((\boldsymbol{A}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right)\left|\frac{d(x_1, x_2, \ldots, x_n)}{d(y_1, y_2, \ldots, y_n)}\right|.$$

eftersom $\boldsymbol{X} = \boldsymbol{A}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})$. Vi ser att jacobianen ges av

$$\frac{d(x_1, x_2, \ldots, x_n)}{d(y_1, y_2, \ldots, y_n)} = |\boldsymbol{A}^{-1}| = |\boldsymbol{A}|^{-1},$$

så

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) &= \frac{1}{(2\pi)^{n/2}}|\boldsymbol{A}|^{-1} \exp\left(-\frac{1}{2}\left(\boldsymbol{A}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right)^T\left(\boldsymbol{A}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right)\right) \\ &= \frac{1}{(2\pi)^{n/2}}|\boldsymbol{A}\boldsymbol{A}^T|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T(\boldsymbol{A}^{-1})^T\boldsymbol{A}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right) \\ &= \frac{1}{(2\pi)^{n/2}}|\boldsymbol{A}\boldsymbol{A}^T|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right), \end{aligned}$$

där vi utnyttjat att $|\boldsymbol{A}|^{-1} = |\boldsymbol{A}|^{-1/2}|\boldsymbol{A}^T|^{-1/2} = |\boldsymbol{A}\boldsymbol{A}^T|^{-1/2}$.
Omvänt, om $\boldsymbol{Y}$ är normalfördelad så säger definitionen att det finns en matris $\boldsymbol{A} \in \mathbf{R}^{n\times m}$ och en vektor $\boldsymbol{\mu} \in \mathbf{R}^n$ så att $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{\mu}$ för $\boldsymbol{X} = (X_1\ X_2\ \cdots\ X_m)^T$ där $X_k \sim N(0,1)$ är oberoende. Faktum är att $m = n$ är nödvändigt då $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^T$ antas vara inverterbar, eftersom $n = \mathrm{rank}(\boldsymbol{A}\boldsymbol{A}^T) \leq \min\{\mathrm{rank}(\boldsymbol{A}), \mathrm{rank}(\boldsymbol{A}^T)\}$ (ty vid produkter av matriser vinner alltid den med lägst rank) och $\mathrm{rank}(\boldsymbol{A}^T) = \mathrm{rank}(\boldsymbol{A})$, så $\mathrm{rank}(\boldsymbol{A}) = n$ eftersom vi har $n$ kolonner. Argumentet ovan visar nu att täthetsfunktionen ges av uttrycket i satsen. $\square$

## Exempel

Låt $X_1,\ X_2 \sim \mathrm{N}(0,1)$ vara oberoende och definiera $\boldsymbol{Y} = (X_1 - X_2,\ 2X_1 + X_2)$. Bestäm täthetsfunktionen för $\boldsymbol{Y}$.

**Lösning**. Vi skriver

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \boldsymbol{A} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \text{där } \boldsymbol{A} = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}.$$

Då blir

$$\boldsymbol{\mu}_{\boldsymbol{Y}} = E\left(\boldsymbol{A}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\right) = \boldsymbol{A} \cdot \boldsymbol{0} = \boldsymbol{0}$$

och

$$C_{\boldsymbol{Y}} = \boldsymbol{A} C_{\boldsymbol{X}} \boldsymbol{A}^T = \boldsymbol{A} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \boldsymbol{A}^T = \boldsymbol{A}\boldsymbol{A}^T = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}.$$

Således blir det $C_{\boldsymbol{Y}} = 9$ och

$$C_{\boldsymbol{Y}}^{-1} = \frac{1}{9} \begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix}.$$

Alltså blir

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

ty

$$(y_1\ y_2)\, \frac{1}{9} \begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{1}{9}\left(5y_1^2 - 2y_1y_2 + 2y_2^2\right).$$

**Sats.** Låt $\boldsymbol{Z} = \boldsymbol{d} + \boldsymbol{B}\boldsymbol{Y}$, där $\boldsymbol{Y}$ är multivariat normalfördelad. Då är även $\boldsymbol{Z}$ multivariat normalfördelad.

**Bevis**. Följer direkt från definitionen.

**Sats.** För $\boldsymbol{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ gäller att komponenterna i $\boldsymbol{Y}$ är oberoende om och endast om $\boldsymbol{\Sigma}$ är en diagonalmatris (under förutsättning att $\boldsymbol{A}$ är inverterbar).

**Bevis**. Kravet på att $\boldsymbol{A}$ ska vara inverterbar följer av att om så icke är fallet så är fördelningen degenererad eftersom $\boldsymbol{A}\boldsymbol{x} = 0$ har oändligt många lösningar. Det är alltså självklart i detta läge att komponenterna i $\boldsymbol{Y}$ inte kan vara oberoende. Så antag nu att $\det \boldsymbol{A} \neq 0$.
Den ena riktningen är mer eller mindre självklar eftersom om komponenterna i $\boldsymbol{Y}$ är oberoende kommer $C(Y_i, Y_j) = 0$ för $i \neq j$ och $C(Y_i, Y_i) = \sigma_i^2$, så $C_{\boldsymbol{Y}}$ blir en diagonalmatris.
Antag nu att $C_{\boldsymbol{Y}}$ är en diagonalmatris, säg

$$\begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}.$$

Eftersom $C_{\boldsymbol{Y}} = \boldsymbol{A}\boldsymbol{A}^T$ kommer $C_{\boldsymbol{Y}}$ att vara inverterbar, vilket innebär att samtliga $\sigma_i^2 \neq 0$. Inversen $C_{\boldsymbol{Y}}^{-1}$ är även den en diagonalmatrisen med diagonalelementen $\sigma_i^{-2}$. Således blir den simultana täthetsfunktionen

$$\begin{aligned}
f_{\boldsymbol{Y}}(\boldsymbol{y}) &= \frac{1}{(2\pi)^{n/2}\sqrt{\det C_{\boldsymbol{Y}}}} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right) \\
&= \frac{1}{(\sqrt{2\pi})^n \sigma_1\sigma_2\cdots\sigma_n} \exp\left(-\frac{1}{2}\sum_{j=1}^{n}(y_j-\mu_j)\sigma_j^{-2}(y_j-\mu_j)\right) \\
&= \prod_{j=1}^{n} \frac{1}{\sigma_j\sqrt{2\pi}} \exp\left(-\frac{(y_j-\mu_j)^2}{2\sigma_j^2}\right) = \prod_{j=1}^{n} f_{Y_j}(y_j).
\end{aligned}$$

Eftersom den simultana täthetsfunktionen ges av produkten av täthetsfunktionerna för $Y_j$ följer det att variablerna är oberoende. $\square$