

Sandsynlighedsregning

12. forelæsning

Bo Friis Nielsen

Matematik og Computer Science

Danmarks Tekniske Universitet

2800 Kgs. Lyngby – Danmark

Email: bfni@imm.dtu.dk

Dagens nye emner afsnit 6.5



- Den bivariate normalfordeling

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}}$$

$$Y = \rho X + \sqrt{1-\rho^2} Z, \quad X, Z \text{ uafhængige standard NF}$$

- $V = \mu_V + \sigma_V X$, $W = \mu_W + \sigma_W Y$ er bivariat normalfordelt med $E(V) = \mu_V$, $E(W) = \mu_W$, $\text{Var}(V) = \sigma_V^2$, $\text{Var}(W) = \sigma_W^2$, samt $\text{Cov}(V, W) = \rho\sigma_V\sigma_W$.

$$V = \sum_i a_i Z_i, \quad W = \sum_i b_i Z_i \text{ med } Z_i \sim \text{normal}(\mu_i, \sigma_i^2) \text{ uafhængige}$$

$$\Rightarrow V \sim \text{normal} \left(\sum a_i \mu_i, \sum a_i^2 \sigma_i^2 \right), W \sim \text{normal} \left(\sum b_i \mu_i, \sum b_i^2 \sigma_i^2 \right),$$

$$\text{Cov}(V, W) = \sum_i a_i b_i \sigma_i^2 \quad (V, W) \text{ bivariat normalfordelte}$$

Bivariat NF - en lille forhistorie



- givet X med $f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$
- Vi indfører en stokastisk variabel Y der for givet $X = x$ har
 - ◇ $E(Y|X = x) = \rho x$, hvor $-1 < \rho < 1$
 - ◇ $\text{Var}(Y|X = x) = 1 - \rho^2$
 - ★ Y givet $X = x$ normalfordelt.

$$f_Y(y|X = x) = \frac{1}{\sqrt{2\pi} \sqrt{1 - \rho^2}} e^{-\frac{1}{2} \frac{(y - \rho x)^2}{1 - \rho^2}}$$

forhistorie fortsat



- Den simultane fordeling $f(x, y)$

$$\begin{aligned} f(x, y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \frac{1}{\sqrt{2\pi} \sqrt{1-\rho^2}} e^{-\frac{1}{2} \frac{(y-\rho x)^2}{1-\rho^2}} \\ &= \frac{1}{2\pi \sqrt{1-\rho^2}} e^{-\frac{1}{2} \frac{x^2 - 2\rho xy + y^2}{1-\rho^2}} \end{aligned}$$

- Udtrykket er symmetrisk i x og y så Y er også standard normal fordelt.

Et kapitel til i forhistorien

- Vi betragter nu $Z = \frac{Y - \rho X}{\sqrt{1 - \rho^2}}$, og får
- $E(Z) = 0$
- $E(Z^2) = E\left(\frac{Y^2 + \rho^2 X^2 - 2\rho XY}{1 - \rho^2}\right)$.
- $E(X^2) = E(Y^2) = 1$
- $E(XY) = E_X(E_Y(XY|X)) = E_X(\rho X^2) = \rho$
- Alt i alt: $E(Z^2) = 1$.

Slut på forhistorie



Fordelingen af (X, Z) ?

$$\begin{aligned} P(X \leq x, Z \leq z) &= P\left(X \leq x, \frac{Y - \rho X}{\sqrt{1 - \rho^2}} \leq z\right) = \\ P\left(X \leq x, Y \leq \rho X + \sqrt{1 - \rho^2} z\right) &= \\ \int_{-\infty}^x \int_{-\infty}^{\rho u + \sqrt{1 - \rho^2} z} f(u, y) dy du &= \\ \int_{-\infty}^x \int_{-\infty}^{\rho u + \sqrt{1 - \rho^2} z} \frac{1}{2\pi\sqrt{1 - \rho^2}} e^{-\frac{u^2 - 2\rho u y + y^2}{2(1 - \rho^2)}} dy du &= \\ \int_{-\infty}^x \int_{-\infty}^{\rho u + \sqrt{1 - \rho^2} z} \frac{1}{2\pi\sqrt{1 - \rho^2}} e^{-\frac{(y - \rho u)^2 + (1 - \rho^2)u^2}{2(1 - \rho^2)}} dy du &= \\ \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \left[\int_{-\infty}^{\rho u + \sqrt{1 - \rho^2} z} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1 - \rho^2}} e^{-\frac{(y - \rho u)^2}{2(1 - \rho^2)}} dy \right] du &= \\ = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \Phi\left(\frac{\rho u + \sqrt{1 - \rho^2} z - \rho u}{\sqrt{1 - \rho^2}}\right) du &= \Phi(x)\Phi(z) \end{aligned}$$

- så X og Z er uafhængige standard normalfordelte

Slut på forhistorie

Fordelingen af (X, Z) ? $P(X \in dx, Z \in dz) = g(x, z)dx dz$



Vi finder - noget heuristisk -

$$\begin{aligned} f(x, y)dx dy &= f(x, \rho x + \sqrt{1 - \rho^2} z)dx dy \\ dx dy &= dx \left(\rho(x + dx) + \sqrt{1 - \rho^2}(z + dz) - (\rho x + \sqrt{1 - \rho^2} z) \right) \\ &\approx \sqrt{1 - \rho^2} dx dz \end{aligned}$$

$$\begin{aligned} g(x, z)dx dz &= f(x, y)dx dy = f\left(x, \rho x + \sqrt{1 - \rho^2} z\right) dx \sqrt{1 - \rho^2} dz \\ &= \frac{1}{2\pi \sqrt{1 - \rho^2}} e^{-\frac{1}{2} \frac{x^2 - 2\rho x(\rho x + \sqrt{1 - \rho^2} z) + (\rho x + \sqrt{1 - \rho^2} z)^2}{1 - \rho^2}} dx \sqrt{1 - \rho^2} dz \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2} dx dz \end{aligned}$$

- så X og Z er uafhængige standard normalfordelte
- og vi kan skrive $Y = \rho X + \sqrt{1 - \rho^2} Z$

På en passende normeret skala beskriver X og Y vægten af henholdsvis mørbraden og svinekammen hos slagtesvin. Man kan antage, at vægtene kan beskrives ved en standardiseret bivariat normalfordeling med korrelationskoefficient $\rho = \frac{3}{5}$.

Spørgsmål 1

Bestem andelen af slagtesvin, hvor summen af de normerede vægte overstiger 1.

1 $1 - \Phi\left(\sqrt{\frac{5}{16}}\right)$

2 $1 - \Phi\left(\frac{1}{2}\right)^2$

3 $1 - \Phi\left(\frac{1}{\sqrt{2}}\right)$

4 $1 - \Phi(1)$

5 $1 - \Phi\left(\sqrt{\frac{7}{8}}\right)$

6 Ved ikke

Standardiseret bivariat normal fordeling



- Samtidig fordeling af to ikke-uafhængige (korrelerede størrelser)
 - ◊ Højde og vægt af mennesker
 - ◊ En lang række biologiske og tekniske målinger
- For to standardiserede normalfordelte variable finder vi

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}}$$

- Hvis X og Y er standardiseret bivariat normalfordelt med korellation ρ , kan vi skrive Y som

$$Y = \rho X + \sqrt{1 - \rho^2} Z$$

- hvor X og Z er uafhængige standardiserede normalfordelte variable

Betinget tæthed



$$f_X(x|Y = y) = \frac{f(x, y)}{f_Y(y)} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1 - \rho^2}} e^{-\frac{1}{2} \frac{(x - \rho y)^2}{1 - \rho^2}}$$

- De betingede tætheder i normalfordelingen er igen normale

Generel bivariat normalfordeling



- U og V er bivariat normalfordelt hvis (X, Y) med

$$X = \frac{U - \mu_u}{\sigma_U}, \quad Y = \frac{V - \mu_V}{\sigma_V}$$

- er standardiseret bivariat normalfordelt

Fædre og sønners højder



- Baseret på 1078 par af engelske mænd og deres (voksne) sønner har man fundet, at den simultane fordeling af højderne kan beskrives ved en bivariat normalfordling.
- Lad U være en faders højde, og V være en søns højde. Eksperimentelt vides da, at
- $E(U) = 172,5\text{cm}$, $E(V) = 175\text{cm}$, $SD(U) = SD(V) = 5\text{cm}$ samt $\rho = 0,5$.
- Man ønsker at bestemme den forventede højde af en søn, hvis fader er 185cm.

Generel bivariat normalfordeling



- Vi danner (X, Y) ud fra U og V

$$X = \frac{U - 172,5cm}{5cm}, \quad Y = \frac{V - 175cm}{5cm}$$

- $U = 185cm \Leftrightarrow X = 2,5$
- $E(Y|X = 2,5) = 0,5 \cdot 2,5 = 1,25$
- $E(V|U = 185cm) = 175cm + 1,25 \cdot 5cm = 181,25cm$

Regression towards the mean

Følgende tabel angiver amerikanske studerendes resultater ved kvalifikationsprøver til college



PSAT score	gennemsnit:1200	SD:100
SAT score	gennemsnit:1300	SD:90
korrelation: 0,6		

Spørgsmål 2

Hvor stor en andel af de studerende, der fik 1000 i PSAT score scorede samtidigt over gennemsnittet i SAT score?

- 1 0,9332
- 2 0,8413
- 3 0,5
- 4 0,1587
- 5 0,0668
- 6 Ved ikke

Linearkombinationer af uafhængige normalfordelte variable

- $Z_i \in normal(\mu_i, \sigma_i^2)$ uafhængige

$$V = \sum_i a_i Z_i, \quad W = \sum_i b_i Z_i$$

- Parret V, W er bivariat normalfordelt med

$$\mu_V = \sum_i a_i \mu_i \quad \mu_W = \sum_i b_i \mu_i$$

$$\sigma_V^2 = \sum_i a_i^2 \sigma_i^2 \quad \sigma_W^2 = \sum_i b_i^2 \sigma_i^2$$

$$\text{Cov}(V, W) = \sum_i a_i b_i \sigma_i^2$$

- Vi har $Z_1 \in normal(3, 4)$ og $Z_2 \in normal(-1, 9)$.
- Vi danner $V = Z_1 + 2Z_2$ og $W = 2Z_1 - Z_2$.

$$E(V) = 3 + 2 \cdot (-1) = 1 \quad E(W) = 2 \cdot 3 - (-1) = 7$$

$$\text{Var}(V) = 4 + 4 \cdot 9 = 40 \quad \text{Var}(W) = 4 \cdot 4 + 9 = 25$$

$$\text{Cov}(V, W) = 1 \cdot 2 \cdot 4 + 2 \cdot (-1) \cdot 9 = -10$$

$$\text{Corr}(V, W) = \frac{-10}{\sqrt{40}\sqrt{25}} = \frac{-1}{\sqrt{10}}$$

Ukorrelerede variable er uafhængige




- To linear kombinationer af $V = \sum_i a_i Z_i$ og $W = \sum_i b_i Z_i$ af uafhængige $normal(\mu_i, \sigma_i^2)$ fordelte variable er uafhængige hvis og kun hvis de er ukorrelerede. Det vil sige hvis $\sum_i a_i b_i \sigma_i^2 = 0$
- Med $Z_i \in normal(0, 1)$ $i = 1, 2$ og $V = \frac{Z_1 + Z_2}{\sqrt{2}}$, $W = \frac{Z_1 - Z_2}{\sqrt{2}}$ får vi

$$\text{Cov}(V, W) = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} \cdot 1 + \frac{(-1)}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} \cdot 1 = 0$$

- dvs. V og W er uafhængige
- Summen og differensen af to standardiserede normalfordelte variable er uafhængige

Eksempel 2 side 457...

- Parret (X, Y) er bivariat standardiseret normalfordelt med  korrelation ρ .
- Hvad er sandsynligheden for at punktet ligger i første kvadrant?
- $P(X > 0, Y > 0) = P(X > 0, \rho X + \sqrt{1 - \rho^2} Z > 0) =$
 $P\left(X > 0, Z > \frac{-\rho}{\sqrt{1 - \rho^2}} X\right)$

Herefter bruger vi rotationsinvariansen af den bivariante normal fordeling for uafhængige variable

$$\begin{aligned} P(X > 0, Y > 0) &= P\left(X > 0, Z > \frac{-\rho}{\sqrt{1 - \rho^2}} X\right) \\ &= \frac{\frac{\pi}{2} + \text{Arctan}\left(\frac{\rho}{\sqrt{1 - \rho^2}}\right)}{2\pi} \end{aligned}$$

Den simultane fordeling af mødres og døtres højder kan beskrives ved en bivariat normalfordeling med korrelation $\frac{1}{2}$. På passende standardiseret skala gælder for begge enkeltvariable, at middelværdien er 0 og standardafvigelsen er 1.



Spørgsmål 3

Hvad er andelen af døtre, der er over gennemsnitshøjde og samtidigt mindre end deres mødre?

- 1 $\frac{1}{10}$
- 2 $\frac{1}{8}$
- 3 $\frac{1}{6}$
- 4 $\frac{1}{4}$
- 5 $\frac{1}{3}$
- 6 Ved ikke

Afsnit 6.5



- Den standardiserede bivariate normalfordeling

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}}$$

$$Y = \rho X + \sqrt{1-\rho^2}Z, \quad X, Z \text{ uafhængige standard NF}$$

- $V = \mu_V + \sigma_V X$, $W = \mu_W + \sigma_W Y$ er bivariat normalfordelt med $E(V) = \mu_V$, $E(W) = \mu_W$, $\text{Var}(V) = \sigma_V^2$, $\text{Var}(W) = \sigma_W^2$, samt $\text{Cov}(V, W) = \rho\sigma_V\sigma_W$.

$$V = \sum_i a_i Z_i, \quad W = \sum_i b_i Z_i \text{ med } Z_i \sim \text{normal}(\mu_i, \sigma_i^2) \text{ uafhængige}$$

$$\Rightarrow V \sim \text{normal}\left(\sum a_i \mu_i, \sum a_i^2 \sigma_i^2\right), W \sim \text{normal}\left(\sum b_i \mu_i, \sum b_i^2 \sigma_i^2\right),$$

$$\text{Cov}(V, W) = \sum_i a_i b_i \sigma_i^2 \quad (V, W) \text{ bivariat normalfordelte}$$