

Sandsynlighedsregning

9. forelæsning

Bo Friis Nielsen

Matematik og Computer Science
Danmarks Tekniske Universitet
2800 Kgs. Lyngby – Danmark
Email: bfn@dtu.dk

Simultan fordeling for kontinuerte variable

$$P((X, Y) \in B) = \int_B f(x, y) dx dy$$


- Marginale fordelinger

$$f_X(x) = \int_y f(x, y) dy \quad f_Y(y) = \int_x f(x, y) dx$$

- Hvis X og Y er uafhængige

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

Dagens emner afsnit 5.3 og 5.4

- Simultane kontinuerte fordelinger 
 $P(X \in dx, Y \in dy) = f(x, y) dx dy$
- Uafhængige normalfordelte variable ($\sim normal(\mu, \sigma^2) = N(\mu, \sigma^2)$)
 - ◇ $f(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$
 - ◇ $F_R(r) = P(\sqrt{X^2 + Y^2} \leq r) = 1 - e^{-\frac{1}{2}r^2}$ (Rayleigh)
 - ◇ $X \sim N(\lambda, \sigma^2), Y \sim N(\mu, \tau^2) \Rightarrow X + Y \sim N(\lambda + \mu, \sigma^2 + \tau^2)$
 - ◇ For $X_i \sim N(0, 1)$ er $Y = \sum_{i=1}^n X_i^2$ χ^2 -fordelt (gamma).
- Operationer med stokastiske variable
 - ◇ $f_Z(z) dz = P(Z = X + Y \in dz) = \int_{-\infty}^{\infty} f(x, z - x) dx \quad dz$
 - ◇ $f_Z(z) dz = P(Z = \frac{Y}{X} \in dz) = \int_{-\infty}^{\infty} |x| f(x, zx) dx \quad dz$

Uafhængige normalfordelte variable

- Tætheden for en standardiseret normalfordelt variabel er

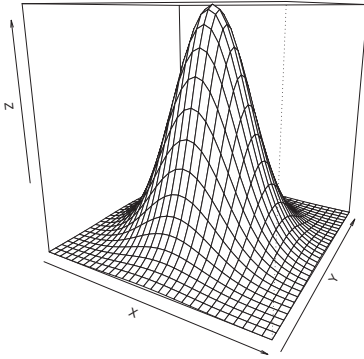
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

- for to uafhængige $normal(0, 1)$ variable X og Y får vi

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

- Bemærk rotationssymmetrien.

Den simultane tæthed for (X, Y)



Fordelingen er rotationsinvariant

Linearkombination



- Summer af normalfordelte variable side 363

Hvis X og Y er uafhængige og normalfordelt med henholdsvis $N(\lambda, \sigma^2)$ og $N(\mu, \tau^2)$ så er $Z = X + Y$ $normal(\lambda + \mu, \sigma^2 + \tau^2)$ fordelt.

- For X_i $normal(\mu_i, \sigma_i^2)$ fordelte og uafhængige er $Z = b + \sum_{i=1}^n a_i X_i$ $normal(b + \sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2)$ fordelt.

(Denne står ikke direkte i Pitman, men kan dog let udledes af resultatet side 364.)

Linearkombinationer af to normalfordelte variable

Lad X og Y være uafhængige $normal(0, 1)$ -fordelte variable.

Betragt $Z = aX + bY$.

1. Hvis $a^2 + b^2 = 1$, da er Z også $normal(0, 1)$ -fordelt.
Det skyldes rotationssymmetrien, Z er den nye førstekoordinat i et drejet koordinatsystem.
2. Generelt:

$$Z = \sqrt{a^2 + b^2} \left(\frac{a}{\sqrt{a^2 + b^2}} X + \frac{b}{\sqrt{a^2 + b^2}} Y \right)$$

... altså er Z **normalfordelt** med middelt 0 og varians $a^2 + b^2$.
Resultatet kan også vises analytisk ved brug af afsnit 5.4

- Lad X og Y være uafhængige stokastiske variable, med

$$E(X) = 1, E(Y) = 2, Var(X) = 3, Var(Y) = 4$$



- Bestem $E(10X^2 + 8Y^2 - XY + 8X + 5Y - 1)$

$$E(10X^2 + 8Y^2 - XY + 8X + 5Y - 1)$$

$$= E(10X^2) + E(8Y^2) - E(XY) + E(8X) + E(5Y) - E(1)$$

(Middelværdien af en sum er summen af komponenternes middelværdier)

$$= 10E(X^2) + 8E(Y^2) - E(XY) + 8E(X) + 5E(Y) - E(1)$$

($E(aX) = aE(X)$; vi ser, at middelværdien af en linearkombination er linearkombinationen af middelværdierne)

... vi fortsætter med hvert led i dette udtryk



$$E(X^2) = \text{Var}(X) + (E(X))^2 = 4$$

$$E(Y^2) = \text{Var}(Y) + (E(Y))^2 = 8$$

(Beregningsformlen for varians, $\text{Var}(X) = E(X^2) - (E(X))^2$)

$$E(XY) = E(X)E(Y) = 2$$

($E(XY) = E(X)E(Y)$ for uafhængige variable)

... så i alt

$$\mu = 10 \cdot 4 + 8 \cdot 8 - 2 + 8 \cdot 1 + 5 \cdot 2 - 1 = 119$$

Antag at R_1 og R_2 er to uafhængige stokastiske variable med den samme tæthedsfunktion $f(x) = xe^{-\frac{1}{2}x^2}$ for $x \geq 0$.



Spørgsmål 1

Tætheden af $Y = \min(R_1, R_2)$ findes til

- 1 $f_Y(y) = 2ye^{-y^2}$
- 2 $f_Y(y) = 1 - (1 - ye^{-\frac{1}{2}y^2})(1 - ye^{-\frac{1}{2}y^2})$
- 3 $f_Y(y) = e^{-y}$
- 4 $f_Y(y) = 2y^2e^{-y^2}$
- 5 $f_Y(y) = 2y^2e^{-\frac{y^2}{2}}$
- 6 Ved ikke

- Bestem $P(2X > 3Y - 5)$ under den antagelse, at X og Y er normalfordelte.

Definer $Z = 3Y - 2X$, vi søger da $P(Z < 5)$

Z er normalfordelt med middel

$$E(Z) = 3 \cdot 2 - 2 \cdot 1 = 4$$

og varians

$$V(Z) = 3^2 \cdot 4 + (-2)^2 \cdot 3 = 48$$

$$P(Z < 5) = P\left(\frac{Z - 4}{\sqrt{48}} < \frac{5 - 4}{\sqrt{48}}\right) = \Phi\left(\frac{1}{\sqrt{48}}\right) \approx 0.557$$

Konstanten $1/\sqrt{2\pi}$

(X, Y) har simultan tæthed $f_{XY}(x, y) = c^2 e^{-\frac{1}{2}(x^2 + y^2)}$.

(Bemærk! Vi har hidtil taget $c = 1/\sqrt{2\pi}$ for givet, men nu udleder vi det!)

$$P(r \leq R \leq r + dr) = P(r \leq \sqrt{X^2 + Y^2} \leq r + dr)$$

$$= c^2 e^{-\frac{1}{2}r^2} (\pi(r + dr)^2 - \pi r^2) = 2\pi c^2 r e^{-\frac{1}{2}r^2} dr + o(dr)$$

... tegn en skitse af X , Y og R for at checke dette. Vi får tætheden af R direkte:

$$f_R(r) = 2\pi c^2 r e^{-\frac{1}{2}r^2}$$

Eftersom $\int_0^\infty f_R(r) dr = 1$, må vi have $c = 1/\sqrt{2\pi}$.

$$f_R(r) = r e^{-\frac{1}{2}r^2} \quad P(R \leq r) = F_R(r) = 1 - e^{-\frac{1}{2}r^2}$$

Størrelse af normalfordelt vektor side 358-361

- Hvis talparret (X, Y) er uafhængige standardiserede normalfordelte ($normal(0, 1)$) variable, så er $R = \sqrt{X^2 + Y^2}$ Rayleigh fordelt.

$$f_R(r) = re^{-\frac{1}{2}r^2}, \quad F_R(r) = 1 - e^{-\frac{1}{2}r^2}$$

- Middelværdi og varians er

$$E(R) = \sqrt{\frac{\pi}{2}}, \quad Var(R) = E(R^2) - E(R)^2 = 2 - \frac{\pi}{2}$$

Fordelingen af $S = R^2 = X^2 + Y^2$



Fra variabelskift:

$$F_S(s) = F_R(\sqrt{s}) = 1 - e^{-\frac{1}{2}s}$$

... altså er S eksponentialfordelt med middelværdi 2.

Deraf fås i øvrigt

$$E(X^2) = 1 \quad (!)$$

En skytte skyder mod en skydeskive. Skuddenes placering på skiven kan beskrives ved 2 uafhængige standardiserede normalfordelte variable, hvor skivens centrum er koordinatsystemets nulpunkt.



Spørgsmål 2

Sandsynligheden for, at et skud rammer indenfor en cirkel med radius $\frac{1}{2}$, er

- $\Phi\left(\frac{1}{2}\right) - \Phi\left(-\frac{1}{2}\right)$
- $\frac{1}{4}$
- $\frac{1}{2}$
- $1 - e^{-\frac{1}{8}}$
- $\left(\Phi\left(\frac{1}{2}\right) - \Phi\left(-\frac{1}{2}\right)\right)^2$
- Ved ikke

Hvor $\Phi(x)$ er fordelingsfunktionen for en standardiseret normalfordelt variabel.

En skytte skyder mod en skydeskive. Skuddenes placering på skiven kan beskrives ved 2 uafhængige standardiserede normalfordelte variable, hvor skivens centrum er koordinatsystemets nulpunkt.



Spørgsmål 3

Sandsynligheden for, at et skud rammer indenfor en afstand 1 fra koordinatsystemets anden-akse, er

- $\Phi(1)$
- $\frac{1}{4}$
- $2\Phi(1) - 1$
- $1 - e^{-\frac{1}{2}}$
- $1 - e^{-\frac{1}{4}}$
- Ved ikke

Hvor $\Phi(x)$ er fordelingsfunktionen for en standardiseret normalfordelt variabel.

Operationer med stokastiske variable 5.4

- Vi har set max og min af stokastiske variable, samt specielle former for summer.
- For uafhængige diskrete variable har vi tidligere set

$$P(Z = X + Y = z) = \sum_{x=-\infty}^{\infty} f_X(x)f_Y(z-x)$$

- En foldning. For ikke negative uafhængige X og Y

$$P(Z = X + Y = z) = \sum_{x=0}^z f_X(x)f_Y(z-x)$$

- For kontinuerte tætheder får vi

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x) dx \stackrel{uaf}{=} \int_{-\infty}^{\infty} f_X(x)f_Y(z-x) dx$$

- Den simultane tæthed af (X_1, X_2) er

$$f(x_1, x_2) = \begin{cases} \frac{1}{2} & 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 2 \\ 0 & \text{ellers} \end{cases}$$

$$P(X_1 + X_2 \leq 2) = \int_{x_1+x_2 \leq 2} \frac{1}{2} dx_1 dx_2$$

- Eller vi kan bruge arealbetrægtninger og få $P(X_1 + X_2 \leq 2) = \frac{3}{4}$

Lad X_1 være ligeligt fordelt på enhedintervallet og X_2 være ligeligt fordelt på intervallet $(0;2)$ og uafhængig af X_1 .

Spørgsmål 4

Sandsynligheden $P(X_1 + X_2 \leq 2)$ findes til

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

Sum af max og min af to eksponentialfordelte variable

Vi har tidligere set at for $W_i \in \exp(\lambda)$, $X = \min(W_1, W_2)$ og $Y = \max(W_1, W_2)$ er den simultane tæthed:

$$f(x, y) = 2\lambda^2 e^{-\lambda(x+y)}, 0 < x \leq y.$$

Fordelingen af $Z = X + Y$ findes til

$$f_Z(z) = \int_0^{\frac{z}{2}} 2\lambda^2 e^{-\lambda(x+(z-x))} dx = \lambda(\lambda z) e^{-\lambda z}$$

Overraskende?

Summer af uafhængige variable

- Med $X \in \text{bin}(n, p)$ og $Y \in \text{bin}(m, p)$ er $Z = X + Y \in \text{bin}(n + m, p)$
- Med $X \in \text{NB}(n, p)$ og $Y \in \text{NB}(m, p)$ er $Z = X + Y \in \text{NB}(n + m, p)$
- Med $X \in P(\lambda)$ og $Y \in P(\mu)$ er $Z = X + Y \in P(\lambda + \mu)$
- Med $X \in \text{Gamma}(r, \lambda)$ og $Y \in \text{Gamma}(s, \lambda)$ er $Z = X + Y \in \text{Gamma}(r + s, \lambda)$
- Summer af mange har normalfordelingen som grænse
- Summer af normalfordelte variable side 363
- Med $X \in N(\mu_1, \sigma_1^2)$ og $Y \in N(\mu_2, \sigma_2^2)$ er $Z = X + Y \in N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
 - ◊ Med $X \in N(\mu_1, \sigma_1^2)$ og $Y \in N(\mu_2, \sigma_2^2)$ er $W = aX + bY \in N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$



Fordeling af forhold side 383

- Hvis $Z = \frac{Y}{X}$, hvor X og Y er uafhængige

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(xz) dx$$

- Se eksempel 5 side 383



χ^2 fordelingen side 364-366, opgave 5.3.15



Summer af kvadrerede uafhængige standard normalfordelte variable: $X_i \sim \text{normal}(0, 1)$, uafhængige. Da har $Y = \sum_{i=1}^n X_i^2$ tæthed

$$f_Y(y) = \frac{1}{2^{n/2} \Gamma(n/2)} y^{(n/2)-1} e^{-y/2}$$

... og siges at være χ^2 fordelt med n frihedsgrader.

(... det er også en Gamma-fordeling $\Gamma(n/2, 1/2)$)

Reproduktion: Hvis $Y_1 \sim \chi^2(n)$ og $Y_2 \sim \chi^2(m)$, så er $Y_1 + Y_2 \sim \chi^2(n + m)$.

Opgave 5.4.10



- Find tætheden af $Y = \frac{V}{U}$, hvor U og V er uafhængige ligefordelte $(0, 1)$ variable.
- Vi benytter $f_Y(y) = \int_{-\infty}^{\infty} |u| f_U(u) f_V(uy) du$.
- Integrationsgrænserne for u er fra 0 til $\min\left(1, \frac{1}{y}\right)$. Så

$$f_Y(y) = \int_0^1 u du = \frac{1}{2}, \quad \text{for } y < 1$$

$$f_Y(y) = \int_0^{\frac{1}{y}} u du = \frac{1}{2y^2}, \quad \text{for } 1 < y$$

$$F_Y(y) = \begin{cases} \frac{y}{2}, & y \leq 1 \\ 1 - \frac{1}{2y}, & 1 < y \end{cases}$$

F-fordelingen: Forholdet mellem to uafhængige χ^2 fordelte variable



Lad $Y \in \chi^2(n)$ og $X \in \chi^2(m)$ og betragt $Z = \frac{Y}{X}$

$$f_Z(z) = \int_0^\infty \frac{n}{m} x f_X(x) f_Y\left(\frac{n}{m}xz\right) dx$$

... efter en længere række udregninger finder vi

$$f_Z(z) = \frac{1}{B\left(\frac{n}{2}, \frac{m}{2}\right)} \left(\frac{n}{m}\right)^{\frac{n}{2}} \frac{z^{\frac{n}{2}-1}}{\left(1 + \frac{n}{m}z\right)^{\frac{n+m}{2}}} \quad \text{hvor } B(r, s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$$

Dette kaldes en $F(n, m)$ fordeling - den optræder hyppigt som en fordeling for teststørrelser i statistikken.

Udledningen af F-fordelingen



- Med $Y \in \chi^2(n)$ og $X \in \chi^2(m)$, vil vi undersøge fordelingen af $Z = \frac{Y}{X} = \frac{Y_1}{X_1}$

$$f_Z(z) = \int_0^\infty x f_{X_1}(x) f_{Y_1}(xz) dx$$

- Af tætheden for Y $f_Y(y)$, findes tætheden $f_{Y_1}(y)$ for $Y_1 = \frac{Y}{n}$ til $n f_Y(n \cdot y)$
- Ved indsættelse af $f_Z(z) =$

$$\int_0^\infty x \cdot m \cdot 2^{-\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)^{-1} (mx)^{\frac{m}{2}-1} e^{-\frac{mx}{2}} n 2^{-\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)^{-1} (nxz)^{\frac{n}{2}-1} e^{-\frac{nxz}{2}} dx$$

$$\begin{aligned} & \int_0^\infty x 2^{-\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)^{-1} (mx)^{\frac{m}{2}-1} m e^{-\frac{mx}{2}} 2^{-\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)^{-1} (nxz)^{\frac{n}{2}-1} n e^{-\frac{nxz}{2}} dx \\ &= 2^{-\frac{n+m}{2}} \left(\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)\right)^{-1} z^{\frac{n}{2}-1} m^{\frac{m}{2}} n^{\frac{n}{2}} \int_0^\infty x^{\frac{m+n}{2}-1} e^{-\frac{m+nz}{2}x} dx \end{aligned}$$

- Funktionen under integralet er næsten en Γ tæthed

$$\begin{aligned} &= \int_0^\infty \frac{m+nz}{2} \frac{\left(\frac{m+nz}{2}x\right)^{\frac{m+n}{2}-1}}{\Gamma\left(\frac{m+n}{2}\right)} e^{-\frac{m+nz}{2}x} dx \\ &= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\left(\frac{m+nz}{2}\right)^{\frac{m+n}{2}}} 2^{-\frac{n+m}{2}} \left(\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)\right)^{-1} z^{\frac{n}{2}-1} m^{\frac{m}{2}} n^{\frac{n}{2}} \end{aligned}$$

$$\begin{aligned} &= \int_0^\infty \frac{m+nz}{2} \frac{\left(\frac{m+nz}{2}x\right)^{\frac{m+n}{2}-1}}{\Gamma\left(\frac{m+n}{2}\right)} e^{-\frac{m+nz}{2}x} dx \\ &= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\left(\frac{m+nz}{2}\right)^{\frac{m+n}{2}}} 2^{-\frac{n+m}{2}} \left(\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)\right)^{-1} z^{\frac{n}{2}-1} m^{\frac{m}{2}} n^{\frac{n}{2}} \end{aligned}$$



- Integralet er 1. Efter de sidste modifikationer får vi

$$f_Z(z) = \frac{1}{B\left(\frac{n}{2}, \frac{m}{2}\right)} \left(\frac{n}{m}\right)^{\frac{n}{2}} \frac{z^{\frac{n}{2}-1}}{\left(1 + \frac{n}{m}z\right)^{\frac{n+m}{2}}}$$

- Idet

$$B(r, s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$$

- En $F(n, m)$ fordeling.

Lad Y_1, Y_2 og Y_3 være tre punkter, der vælges tilfældigt og uafhængigt i intervallet $(0, 1)$. Lad X være punktet tættest på 1.



Spørgsmål 5

Fordelingsfunktionen for X er

- 1 $F(x) = 1 - (1 - x)^3$
- 2 $F(x) = x$
- 3 $F(X) = \Phi\left(\frac{x - \frac{2}{3}}{\sqrt{12}}\right)$
- 4 $F(x) = x^3$
- 5 $F(x) = x^{\frac{1}{3}}$
- 6 Ved ikke

Hvor $\Phi(x)$ er fordelingsfunktionen for en standardiseret normalfordelt variabel.



Farvede kugler trækkes tilfældigt fra en æske. Indholdet af æsken er givet ved

| Farve | Rød | blå | grøn | gul |
|-------|-----|-----|------|-----|
| Andel | 0.1 | 0.2 | 0.3 | 0.4 |

Spørgsmål 6



Hvad er sandsynligheden for at få præcis 5 gule kugler i 20 trækninger

- 1 Ingen af de nedenstående
- 2 $\binom{20}{5} (0.4)^5 (0.6)^{15}$
- 3 $\Phi\left(\frac{5 - 0.1 \cdot 20}{\sqrt{20 \cdot 0.1 \cdot 0.9}}\right)$
- 4 $5 \cdot 0.1$
- 5 $1 - \Phi\left(\frac{5 - 0.1 \cdot 20}{\sqrt{20 \cdot 0.1 \cdot 0.9}}\right)$
- 6 Ved ikke

Spørgsmål 7



Netop 2 røde, 4 blå, 6 grønne og 8 gule i 20 trækninger.

- 1 $\binom{20}{4} (0.1)^2 (0.2)^4 (0.3)^6 (0.4)^8$
- 2 $8 \cdot 6 \cdot 4 \cdot 2 (0.1)^2 (0.2)^4 (0.3)^6 (0.4)^8$
- 3 0.5
- 4 $\frac{20!}{(5!)^4} (0.1)^2 (0.2)^4 (0.3)^6 (0.4)^8$
- 5 $\frac{20!}{2!4!6!8!} (0.1)^2 (0.2)^4 (0.3)^6 (0.4)^8$
- 6 Ved ikke

Spørgsmål 8

Netop 25 trækninger for at få 3 røde.



- 1 $\binom{25}{3} (0.1)^3 (0.9)^{22}$
- 2 $((0.1)(0.9)^{24})^3$
- 3 $(0.1)^3 (0.9)^{22}$
- 4 $\binom{24}{2} (0.1)^3 (0.9)^{22}$
- 5 $((0.1)^{24} (0.9))^3$
- 6 Ved ikke

Afsnit 5.3 og 5.4



- Simultane kontinuerte fordelinger
 $P(X \in dx, Y \in dy) = f(x, y) dx dy$
- Uafhængige normalfordelte variable ($\sim normal(\mu, \sigma^2) = N(\mu, \sigma^2)$)
 - ◇ $f(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$
 - ◇ $F_R(r) = P(\sqrt{X^2 + Y^2} \leq r) = 1 - e^{-\frac{1}{2}r^2}$ (Rayleigh)
 - ◇ $X \sim N(\lambda, \sigma^2), Y \sim N(\mu, \tau^2) \Rightarrow X + Y \sim N(\lambda + \mu, \sigma^2 + \tau^2)$
 - ◇ For $X_i \sim N(0, 1)$ er $Y = \sum_{i=1}^n X_i^2$ χ^2 -fordelt (gamma).
- Operationer med stokastiske variable
 - ◇ $f_Z(z) dz = P(Z = X + Y \in dz) = \int_{-\infty}^{\infty} f(x, z - x) dx dz$
 - ◇ $f_Z(z) dz = P(Z = \frac{Y}{X} \in dz) = \int_{-\infty}^{\infty} |x| f(x, zx) dx dz$