

# Sandsynlighedsregning

## 8. forelæsning

Bo Friis Nielsen

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

Lad  $X$  være en stokastisk variabel med fordelingsfunktion

$$F(x) = x^3 \text{ for } 0 < x < 1.$$

### Spørgsmål 1

Tætheden  $f(x)$  for  $X$  er

- 1   $3x^2$  for  $0 < x < 1$ .
- 2   $\frac{1}{4}x^4$  for  $0 < x < 1$ .
- 3  1 for  $0 < x < 1$ .
- 4   $1 - x^3$  for  $0 < x < 1$ .
- 5   $\frac{3x^2}{x^3}$
- 6  Ved ikke



## Dagens emner 5.1 og 5.2

- Ligefordeling med to variable (5.1):  $P((x, y) \in C) = \frac{A(C)}{A(D)}$  DTU
- Simultane tætheder (5.2)  
 $f(x, y)dxdy = P(x \leq X \leq x + dx, y \leq Y \leq y + dy), f(x, y) \geq 0$
- Simultan fordelingsfunktion (5.2)

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^y \int_{-\infty}^x f(x, y)dxdy$$

- Repetition af maximum/minimum - 4.5  
 $F_{max}(x) = \prod_{i=1}^n F_i(x)$   $G_{min}(x) = \prod_{i=1}^n G_i(x)$
- Simultan fordeling af minimum og maximum
- Repetition af "order statistics" - 4.6  
 $f_{(k)}(x) = nf(x) \binom{n-1}{k-1} (F(x))^{k-1} (1 - F(x))^{n-k}$
- Simultan fordeling af "order statistics" (Example 3. p.352-355)

$$f(x, y) = 5!x(y-x)(1-y) \text{ , for } 0 < x < y < 1$$

## Flerdimensionale uniformt fordelte variable

- Sandsynligheder som andele af areal DTU
- Antag, at et punkt med sikkerhed falder i området  $D$  med areal  $A(D)$ .
- Sandsynligheden for, at et punkt  $(X, Y)$  falder i området  $C$ , hvor  $C \subset D$ , er lig den andel som  $A(C)$  udgør af det totale areal:  $P((X, Y) \in C) = \frac{A(C)}{A(D)}$ .

Hvis  $X$  og  $Y$  begge er uniformt fordelt på et interval og uafhængige, da er  $(X, Y)$  uniformt fordelt på et rektangel (side 340).

Lad  $(X, Y)$  være ligefordelt på mængden  $\{(x, y) \in \mathbf{R} \times \mathbf{R} : 0 < x < 2 \cap 0 < y < 4 \cap x < y\}$ .



### Spørgsmål 2

Hvad er  $P(X < 1)$

- 1   $\frac{1}{2}$
- 2   $\frac{3}{5}$
- 3   $\frac{5}{8}$
- 4   $\frac{6}{10}$
- 5   $\frac{7}{12}$
- 6  Ved ikke

### Opgave 5.1.3



Antag at  $X$  og  $Y$  er uafhængige ligefordelte på enhedsintervallet  $(0,1)$ .

Spm: Find

$$P\left(Y \geq \frac{1}{2} \mid Y \geq 1 - 2X\right)$$

Svar: Først benytter vi  $P(A|B) = \frac{P(A \cap B)}{P(B)}$

$$P\left(Y \geq \frac{1}{2} \mid Y \geq 1 - 2X\right) = \frac{P\left(Y \geq \frac{1}{2}, Y \geq 1 - 2X\right)}{P(Y \geq 1 - 2X)}$$

- Først nævneren



$$P(Y \geq 1 - 2X) = 1 - P(Y < 1 - 2X) = 1 - \frac{1}{4} = \frac{3}{4}$$

- så tælleren

$$P\left(Y \geq \frac{1}{2}, Y \geq 1 - 2X\right) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{3}{4} = \frac{7}{16}$$

- således at

$$P\left(Y \geq \frac{1}{2} \mid Y \geq 1 - 2X\right) = \frac{\frac{7}{16}}{\frac{3}{4}} = \frac{7}{12}$$

### Simultan fordeling for kontinuerte variable



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

- Marginale fordelinger

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

Antag at  $(X, Y)$  er ligefordelt over området  $\{(x, y) \in \mathbf{R} \times \mathbf{R} : 0 < |y| < x < 1\}$ .



### Spørgsmål 3

Den simultane (joint) tæthed  $f(x, y)$  af  $(X, Y)$  findes til

- 1   $f(x, y) = \frac{1}{2}, \quad 0 < |y| < x < 1$
- 2   $f(x, y) = 1, \quad 0 < |y| < x < 1$
- 3   $f(x, y) = x|y|, \quad 0 < |y| < x < 1$
- 4   $f(x, y) = x, \quad 0 < |y| < x < 1$
- 5   $f(x, y) = |y|, \quad 0 < |y| < x < 1$
- 6  Ved ikke

Antag at  $(X, Y)$  er ligefordelt over området  $\{(x, y) \in \mathbf{R} \times \mathbf{R} : 0 < |y| < x < 1\}$ .



### Spørgsmål 4

Den marginale tæthed  $f_X(x)$  af  $X$  findes til

- 1   $f_X(x) = 1, \quad 0 < x < 1$
- 2   $f_X(x) = 2, \quad 0 < x < 1$
- 3   $f_X(x) = 2x, \quad 0 < x < 1$
- 4   $f_X(x) = 3x^2, \quad 0 < x < 1$
- 5   $f_X(x) = 2x|y|, \quad 0 < x < 1$
- 6  Ved ikke

## Opgave 5.2.3



Et punkt vælges i enhedskvadratet i henhold til den simultane tæthed  $f(x, y) = c(x^2 + 4xy)$  for  $0 < x < 1$  og  $0 < y < 1$ , for en konstant  $c$ .

*Spm:* Bestem  $c$

*Svar:*

$$\begin{aligned} 1 &= \int \int f(x, y) dx dy = \int_0^1 \int_0^1 c(x^2 + 4xy) dy dx \\ &= \int_0^1 c(x^2 + 2x) dx = \frac{4c}{3} \end{aligned}$$

Så  $c = \frac{3}{4}$

*Spm:* Find  $P(X \leq a), \quad 0 < a < 1$



*Svar:*

$$\int_0^a \int_0^1 \frac{3}{4}(x^2 + 4xy) dy dx = \int_0^a \frac{3}{4}(x^2 + 2x) dx = \frac{3}{4} \left( \frac{a^3}{3} + a^2 \right)$$

dette er *fordelingsfunktionen* for  $X$  evalueret i  $a$

*Spm:* Find  $P(Y \leq b), \quad 0 < Y < 1$

*Svar:*

$$\int_0^b \int_0^1 \frac{3}{4}(x^2 + 4xy) dx dy = \int_0^b \frac{3}{4} \left( \frac{1}{3} + 2y \right) dy = \frac{3}{4} \left( \frac{b}{3} + b^2 \right)$$

*fordelingsfunktionen* for  $Y$  evalueret i  $b$

## Uafhængige kontinuerte variable



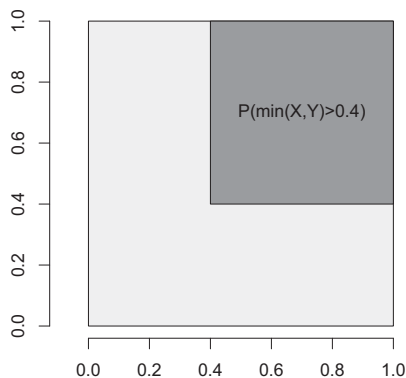
- De stokastiske variable  $X$  og  $Y$  med simultan tæthed  $f(x, y)$  og marginale tætheder henholdsvis  $f_X(x)$  og  $f_Y(y)$  er uafhængige hvis og kun hvis

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

for alle  $(x, y)$ .

## Opgave 5.1.7

- Lad  $U$  og  $V$  være to uafhængige ligefordelte  $(0,1)$  variable. Lad  $X$  være den mindste af  $U$  og  $V$ .
  - ◊ Repræsenterer hændelsen  $(X \geq x)$  ( $0 < x < 1$ ) som et område i planen og find  $P(X \geq x)$  som arealet af dette område.



$$P(X \geq x) = (1 - x)^2.$$

$$\begin{aligned} \text{Eller: } F(x) &= P(X \leq x) \\ &= 1 - (1 - x)^2 \\ (F_{min}(x) &= P(X_{(1)} \leq x) \\ &= 1 - (1 - F(x))^n \end{aligned}$$

## Fordeling af maximum og minimum



- $n$  identisk fordelte stokastiske variable  $X_i$  med fordelingsfunktion  $F(x)$
- Den sorterede sekvens indekseres med  $X_{(i)}$  således at  $X_{(min)} = X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)} = X_{(max)}$

$$P(X_{(n)} \leq x) = P(\max_i(X_i) \leq x) = P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x)$$

- Hvis vi har uafhængighed får vi

$$P(X_{(n)} \leq x) = P(X_1 \leq x)P(X_2 \leq x) \dots P(X_n \leq x) = F(x)^n$$

$$\begin{aligned} P(X_{(1)} \leq x) &= 1 - P(X_1 > x)P(X_2 > x) \dots P(X_n > x) \\ &= 1 - (1 - F(x))^n \end{aligned}$$

- Find fordelingsfunktionen og tætheden af  $X$  og skitser grafen af disse funktioner.



$$P(X \leq x) = F_X(x) = 1 - P(X \geq x) = 1 - (1 - x)^2$$

- Vi finder tætheden ved differentiation

$$f_X(x) = \frac{dF_X(x)}{dx} = 2(1 - x)$$

Lad  $X = \min(U, V)$  og  $Y = \max(U, V)$ , hvor  $U$  og  $V$  er uafhængige uniformt(0,1) fordelte.



### Spørgsmål 5

Sandsynligheden  $P(x < X, Y \leq y)$  hvor  $x \leq y$  findes til

- 1   $y^2 - x^2$
- 2   $(y - x)^2$
- 3   $\frac{1}{y-x}$
- 4   $y - x$
- 5   $2xy - x^2$
- 6  Ved ikke

$$P(x < X, Y \leq y) = P(x < U \leq y \wedge x < V \leq y) \\ = P(x < U \leq y)P(x < V \leq y) = (y - x)^2$$



Idet hændelsen  $(Y \leq y)$  kan skrives som foreningen af de to disjunkte hændelser  $(x < X \wedge Y \leq y)$  og  $(X \leq x \wedge Y \leq y)$  får vi

$$P(X \leq x, Y \leq y) = F(x, y) = P(Y \leq y) - P(x < X, Y \leq y) \\ = y^2 - (y - x)^2 = 2xy - x^2$$

Den simultane fordeling af maksimum og minimum af to ligefordelte variable.

Alternativt til den første udledning kunne vi benytte:

$$P(x < X, Y \leq y) = \int_x^y \int_u^y 2dvdu = \int_x^y 2[y - u] du \\ [2yu - u^2]_x^y = 2y^2 - y^2 - 2xy + x^2 = (y - x)^2$$



Tilsvarende kunne vi udlede den simultane fordeling

$$P(X \leq x, Y \leq y) = \int_0^x \int_u^y 2dvdu = \int_0^x 2[y - u] du = \\ [2yu - u^2]_0^x = 2xy - x^2$$

### Fordeling af ordnede variable side 326



Lad  $X_i$ , være i.i.d. ( $i = 1, \dots, n$ ) med fordelingsfunktion  $F(x)$  og tæthed  $f(x)$ .

Lad  $X_{(k)}$  igen benævne den  $k$ 'te mindste, så

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n-1)} \leq X_{(n)}.$$

$X_{(k)}$  har tæthed  $f_k(x)$  givet ved

$$f_k(x) = n \binom{n-1}{k-1} f(x)F(x)^{k-1}(1-F(x))^{n-k}$$

## Ordning af uniforme variable



Hvis  $X_i$  er  $U(0,1)$  (d.v.s. uniformt fordelt på  $[0,1]$ ) er  $F(x) = x$ .

Dermed bliver tæthedsfunktionen for  $X_{(k)}$ :

$$f_k(x) = n \binom{n-1}{k-1} x^{k-1} (1-x)^{n-k}$$

(Tætheder side 328)

## Example 3. p.352-355



Vi har  $U_i \in U(0,1), i = 1, \dots, 5$ . Definer  $X = U_{(2)}$  og  $Y = U_{(4)}$ .

Spm: bestem den simultane tæthed  $f(x, y)$  af  $(X, Y)$ .

Svar:

$$f(x, y) dx dy = P(x \leq U_{(2)} \leq x + dx, y \leq U_{(4)} \leq y + dy)$$

$$= P(U_{(1)} \leq x \leq U_{(2)} \leq x + dx \leq U_{(3)} \leq y \leq U_{(4)} \leq y + dy \leq U_{(5)} \leq 1)$$

$$= 5 \cdot x \cdot 4 \cdot dx \cdot 3 \cdot (y - x) \cdot 2 \cdot dy \cdot (1 - y)$$

$$= 5! x(y - x)(1 - y) dx dy$$

Så:

$$f(x, y) = 5! x(y - x)(1 - y)$$

## Betafordelingen



Lidt mere generelt siges en stokastisk variabel med tæthed

$$f(x) = \frac{1}{B(r, s)} x^{r-1} (1-x)^{s-1}, \quad 0 < x < 1$$

for vilkårlige  $r > 0, s > 0$  at være  $\text{beta}(r, s)$  fordelt.

Her er

$$B(r, s) = \int_0^1 x^{r-1} (1-x)^{s-1} dx = \frac{\Gamma(r) \Gamma(s)}{\Gamma(r+s)}$$

Idet  $\Gamma(r) = (r-1)!$  for  $r$  heltallig.

Dermed er  $X_{(k)}$  fra før  $\text{beta}(k, n - k + 1)$  -fordelt.

## Marginale fordelinger:



$$f_X(x) = \int_x^1 5! x(y-x)(1-y) dy$$

$$= 20x(1-x)^3$$

$$f_Y(y) = \int_0^y 5! x(y-x)(1-y) dx$$

$$= 20y^3(1-y)$$

Begge er *Beta* fordelinger som de skulle være!

## Opgave 5.2.9

Lad  $X = \min(S, T)$  og  $Y = \max(S, T)$ , hvor  $S$  og  $T$  er



uafhængige exponentialfordelte med intensitet  $\lambda$ , i.e.

$F(t) = 1 - e^{-\lambda t}$ .  
Spørgsmål: Find den simultane fordeling af  $X$  og  $Y$ . Er  $X$  og  $Y$  uafhængige? Svar:

$$P(x < X, Y \leq y) = P(x < S \leq y \wedge x < T \leq y) = (F(y) - F(x))^2 \\ = (e^{-\lambda x} - e^{-\lambda y})^2$$

$$F_{X,Y}(x, y) = F_Y(y) - P(x < X, Y \leq y) = (1 - e^{-\lambda y})^2 - (e^{-\lambda x} - e^{-\lambda y})^2 \\ 1 - e^{-2\lambda x} - 2e^{-\lambda y} + 2e^{-\lambda(x+y)} \quad f_{X,Y}(x, y) = 2\lambda^2 e^{-\lambda(x+y)}$$

Vi har  $F_X(x) = 1 - e^{-2\lambda x}$  så  $F_{X,Y}(x, y) \neq F_X(x)F_Y(y)$  - dvs.  $X$  og  $Y$  ikke uafhængige. Overraskende?

## Opgave 5.2.9 alternativt



Svar:

Vi kan alternativt benytte et hazardrate argument til at udlede den simultane tæthed for  $X$  og  $Z$ :

$$f(x, z) = 2\lambda e^{-2\lambda x} \lambda e^{-\lambda z}$$

- Bestem den marginale fordeling af  $X$  og  $Z$ .

$$f(x) = 2\lambda e^{-2\lambda x} \quad f(z) = \lambda e^{-\lambda z}$$

- $X$  og  $Z$  er uafhængige

Definer yderligere  $Z = Y - X$ . Spørgsmål: Find den simultane fordeling af  $X$  og  $Z$ . Er  $X$  og  $Z$  uafhængige? Svar: Vi har

$$F_{X,Z}(x, z) = P(X \leq x, Z \leq z) = P(X \leq x, Y - X \leq z) \\ = P(X \leq x, Y \leq X + z)$$

Ved brug af definitionen på tætheden får vi

$$F_{X,Z}(x, z) = \int_0^x \int_t^{t+z} 2\lambda^2 e^{-\lambda(t+y)} dy dt \\ = (1 - e^{-\lambda z}) (1 - e^{-2\lambda x}) \quad x \geq 0, z \geq 0$$

Vi finder tætheden ved differentiation  $f(x, z) = 2\lambda e^{-2\lambda x} \lambda e^{-\lambda z}$  og den marginale fordeling af  $X$  og  $Z$

$$f(x) = 2\lambda e^{-2\lambda x} \quad f(z) = \lambda e^{-\lambda z} \quad X \text{ og } Z \text{ er uafhængige}$$

## 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
- Hvad med summer af normalfordelte variable?
- 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  $W = aX + bY \in N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$



Hændelserne  $A, B$  og  $C$  er defineret i et udfaldsrum. Find udtryk for de følgende sandsynligheder i form af

$P(A), P(B), P(C), P(A \cap B), P(A \cap C), P(B \cap C)$  og  $P(A \cap B \cap C)$ .



### Spørgsmål 6

Sandsynligheden for at netop to ud af  $A, B$  og  $C$  indtræffer.

- 1   $P(A \cap B) + P(A \cap C) + P(B \cap C)$
- 2   $P(A)P(B)(1 - P(C)) + P(A)(1 - P(B))P(C) + (1 - P(A^c))P(B)P(C)$
- 3   $P(A \cap B) + P(A \cap C) + P(B \cap C) - 3P(A \cap B \cap C)$
- 4   $1 - P(A)P(B)P(C) - P(A^c)P(B^c)P(C^c)$
- 5  Ingen af de ovenstående
- 6  Ved ikke

### Spørgsmål 8



Tjener denne udskiftningsstrategi et fornuftigt formål, og med hvilken begrundelse?

- 1  Ja der bliver færre nedbrud
- 2  Nej, der bruges flere komponenter end nødvendigt uden at få færre nedbrud
- 3  Ja, middellevetiden af komponenterne øges
- 4  Nej, der bruges det samme antal komponenter og man får hverken færre eller flere nedbrud
- 5  Det gør ingen forskel
- 6  Ved ikke

En type af elektriske komponenter har exponentialfordelte levetider med middelværdi 48 timer. I en type af udstyr udskiftes komponenten straks hvis den fejler og altid efter 48 timers funktionstid.



### Spørgsmål 7

Middelfunktionstiden af en komponent findes til (afrund eventuelt til nærmeste heltal)

- 1  24 timer
- 2  30 timer
- 3  36 timer
- 4  42 timer
- 5  48 timer
- 6  Ved ikke

## Dagens emner 5.1 og 5.2

- Ligeformdeling med to variable (5.1):  $P((x, y) \in C) = \frac{A(C)}{A(D)}$
- Simultane tætheder (5.2)  
 $f(x, y)dx dy = P(x \leq X \leq x + dx, y \leq Y \leq y + dy), f(x, y) \geq 0$
- Simultan fordelingsfunktion (5.2)

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^y \int_{-\infty}^x f(x, y) dx dy$$

- Repetition af maximum/minimum - 4.5  
 $F_{max}(x) = \prod_{i=1}^n F_i(x) \quad G_{min}(x) = \prod_{i=1}^n G_i(x)$
- Simultan fordeling af minimum og maximum
- Repetition af "order statistics" - 4.6  
 $f_{(k)}(x) = n f(x) \binom{n-1}{k-1} (F(x))^{k-1} (1 - F(x))^{n-k}$
- Simultan fordeling af "order statistics" (Example 3. p.352-355)

$$f(x, y) = 5!x(y-x)(1-y) \quad , \quad \text{for } 0 < x < y < 1$$