Sandsynlighedsregning 8. forelæsning

Bo Friis Nielsen

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

Dagens emner 5.1 og 5.2

- Ligefordeling med to variable (5.1): $\mathsf{P}((x,y)\in C) = \frac{A(C)}{A(D)}$
- Simultane tætheder (5.2)

  $f(x,y)\mathrm{d}x\mathrm{d}y \stackrel{\sim}{=} \mathsf{P}(x \le X \le x+\mathrm{d}x, y \le Y \le y+\mathrm{d}y), f(x,y) \ge 0$
- Simultan fordelingsfunktion (5.2)

$$F(x,y) = \mathsf{P}(X \le x, Y \le y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)\mathrm{d}x\mathrm{d}y$$

- 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

  $f_{(k)}(x) = nf(x)\begin{pmatrix} n-1 \\ k-1 \end{pmatrix}(F(x))^{k-1}(1-F(x))^{n-k}$
- Simultan fordeling af "order statistics" (Example 3. p.352-355)

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

Lad X være en stokastisk variabel med fordelingsfunktion $F(x)=x^3$ 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

Flerdimensionale uniformt fordelte variable

- Sandsynligheder som andele af areal
- 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: $\mathsf{P}((X,Y)\in C) = \frac{A(C)}{A(D)}$.

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 $\{(x, y) \in \mathbf{R} \times \mathbf{R} : 0 < x < 2 \wedge 0 < y < 4 \wedge x < y\}$.

Spørgsmål 2

Hvad er $\mathsf{P}(X < 1)$

1 ☐ $\frac{1}{2}$

2 ☐ $\frac{3}{5}$

3 ☐ $\frac{5}{8}$

4 ☐ $\frac{7}{10}$

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

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

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

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

- Først nævneren

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

- 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

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

Simultan fordeling for kontinuerte variable

$$f(x, y) \geq 0, \quad \int_{-\infty}^{\infty} f(x, y) \mathrm{d}x \mathrm{d}y = 1$$

$$\mathsf{P}((X, Y) \in B) = \int_B f(x, y) \mathrm{d}x \mathrm{d}y$$

- Marginale fordelinger

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) \mathrm{d}y \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \mathrm{d}x$$

$$\mathsf{P}(x \leq X \leq x + \mathrm{d}x, y \leq Y \leq y + \mathrm{d}y) \simeq f(x, y) \mathrm{d}x \mathrm{d}y$$

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 ☐ $f(x,y) = \frac{1}{2}, \qquad 0 < |y| < x < 1$

2 ☐ $f(x,y) = 1, \qquad 0 < |y| < x < 1$

3 ☐ $f(x,y) = x|y|, \qquad 0 < |y| < x < 1$

4 ☐ $f(x,y) = x, \qquad 0 < |y| < x < 1$

5 ☐ $f(x,y) = |y|, \qquad 0 < |y| < x < 1$

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 ☐ $f_X(x) = 1, \qquad 0 < x < 1$

2 ☐ $f_X(x) = 2, \qquad 0 < x < 1$

3 ☐ $f_X(x) = 2x, \qquad 0 < x < 1$

4 ☐ $f_X(x) = 3x^2, \qquad 0 < x < 1$

5 ☐ $f_X(x) = 2x|y|, \qquad 0 < x < 1$

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:

$$1 = \int\int f(x,y)\mathrm{d}x\mathrm{d}y = \int_0^1\int_0^1 c(x^2+4xy)\mathrm{d}y\mathrm{d}x$$

$$= \int_0^1 c(x^2+2x)\mathrm{d}x = \frac{4c}{3}$$

Så $c = \frac{3}{4}$

Spm: Find $\mathsf{P}(X \le a), \qquad 0 < a < 1$

Svar:

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

dette er *fordelingsfunktionen* for X evalueret i a

Spm: Find $\mathsf{P}(Y \le b), \qquad 0 < Y < 1$

Svar:

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

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).

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 $X_{(min)} = X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)} = X_{max}$

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

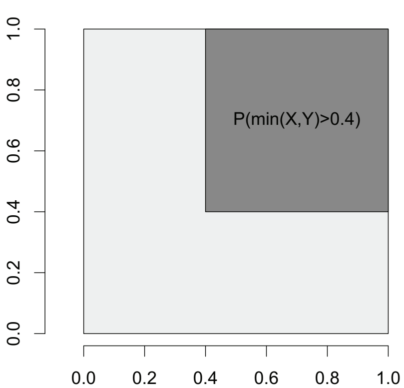
- Hvis vi har uafhængighed får vi

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

$$\mathsf{P}(X_{(1)} \leq x) = 1 - \mathsf{P}(X_1 > x)\mathsf{P}(X_2 > x)\cdots\mathsf{P}(X_n > x)$$
$$= 1 - (1 - F(x))^n$$

Opgave 5.1.7

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



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

Eller: $F(x) = \mathsf{P}(X \leq x)$
$= 1 - (1-x)^2$
$(F_{min}(x) = \mathsf{P}(X_{(1)} \leq x)$
$= 1 - (1 - F(x))^n)$

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

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

- Vi finder tætheden ved differentation

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x} = 2(1-x)$$

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

Spørgsmål 5

Sandsynligheden $\mathsf{P}(x < X, Y \leq y)$ 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

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

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

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

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

Alternativt til den første udledning kunne vi benytte:

$$\mathsf{P}(x < X, Y \leq y) = \int_x^y \int_u^y 2 \mathrm{d}v \mathrm{d}u = \int_x^y 2\,[y - u]\, \mathrm{d}u =$$
$$\left[2yu - u^2\right]_x^y = 2y^2 - y^2 - 2xy + x^2 = (y - x)^2$$

Tilsvarende kunne vi udlede den simultane fordeling

$$\mathsf{P}(X \leq x, Y \leq y) = \int_0^x \int_u^y 2 \mathrm{d}v \mathrm{d}u = \int_0^x 2\,[y - u]\, \mathrm{d}u =$$
$$\left[2yu - u^2\right]_0^x = 2xy - x^2$$

Fordeling af ordnede variable side 326

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

Lad $X_{(k)}$ igen benævne den k'te mindste, så $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n-1)} \leq X_{(n)}$.

$X_{(k)}$ 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)

Betafordelingen

Lidt mere generelt siges en stokastisk variabel med tæthed

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

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

Her er

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

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

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

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:

$$f(x,y)\mathrm{d}x\mathrm{d}y = \mathrm{P}(x \le U_{(2)} \le x + \mathrm{d}x, y \le U_{(4)} \le y + \mathrm{d}y)$$

$$= \mathrm{P}(U_{(1)} \le x \le U_{(2)} \le x + \mathrm{d}x \le U_{(3)} \le y \le U_{(4)} \le y + \mathrm{d}y \le U_{(5)} \le 1)$$

$$= \frac{5!}{1!1!1!1!1!} x^1 \mathrm{d}x^1 (y-x)^1 \mathrm{d}y^1 (1-y)^1$$

$$= 5 \cdot x \cdot 4 \cdot \mathrm{d}x \cdot 3 \cdot (y-x) \cdot 2 \cdot \mathrm{d}y \cdot (1-y)$$

$$= 5!x(y-x)(1-y)\mathrm{d}x\mathrm{d}y$$

Så: $$f(x,y) = 5!x(y-x)(1-y)$$

Marginale fordelinger:

$$f_X(x) = \int_x^1 5!x(y-x)(1-y)\mathrm{d}y$$

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

$$f_Y(y) = \int_0^y 5!x(y-x)(1-y)\mathrm{d}x$$

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

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 λ, i.e. $F(t) = 1 - e^{-\lambda t}$.

Spørgsmål: Find den simultane fordeling af X og Y. Er X og Y uafhængige? *Svar:*

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

$$= \left(e^{-\lambda x} - e^{-\lambda y}\right)^2$$

$$F_{X,Y}(x,y) = F_Y(y) - \mathsf{P}(x < X, Y \le y) = \left(1 - e^{-\lambda y}\right)^2 - \left(e^{-\lambda x} - e^{-\lambda y}\right)^2$$

$$1 - e^{-2\lambda x} - 2e^{-\lambda y} + 2e^{-\lambda(x+y)} \qquad f_{X,Y}(x,y) = 2\lambda^2 e^{-\lambda(x+y)}$$

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?

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

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

$$= \mathsf{P}(X \le x, Y \le X + z) \quad .$$

Ved brug af definitionen på tætheden får vi

$$F_{X,Z}(x,z) = \int_0^x \int_t^{t+z} 2\lambda^2 e^{-\lambda(t+y)} \mathrm{d}y \mathrm{d}t$$

$$= \left(1 - e^{-\lambda z}\right)\left(1 - e^{-2\lambda x}\right) \qquad x \ge 0,\ z \ge 0$$

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

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

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.

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

- X og Z er uafhængige

Summer af uafhængige variable

- Med $X \in bin(n,p)$ og $Y \in bin(m,p)$ er $Z = X + Y \in bin(n+m,p)$
- Med $X \in NB(n,p)$ og $Y \in NB(m,p)$ er $Z = X + Y \in NB(n+m,p)$
- Med $X \in \mathsf{P}(\lambda)$ og $Y \in \mathsf{P}(\mu)$ er $Z = X + Y \in \mathsf{P}(\lambda + \mu)$
- Med $X \in Gamma(r,\lambda)$ og $Y \in Gamma(s,\lambda)$ er $Z = X + Y \in Gamma(r+s,\lambda)$
- 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 $\mathsf{P}(A), \mathsf{P}(B), \mathsf{P}(C), \mathsf{P}(A \cap B), \mathsf{P}(A \cap C), \mathsf{P}(B \cap C)$ og $\mathsf{P}(A \cap B \cap C)$.

Spørgsmål 6

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

1 ☐ $\mathsf{P}(A \cap B) + \mathsf{P}(A \cap C) + \mathsf{P}(B \cap C)$

2 ☐ $\mathsf{P}(A)\mathsf{P}(B)(1 - \mathsf{P}(C)) + \mathsf{P}(A)(1 - \mathsf{P}(B))\mathsf{P}(C) + (1 - \mathsf{P}(A^c))\mathsf{P}(B)\mathsf{P}(C)$

3 ☐ $\mathsf{P}(A \cap B) + \mathsf{P}(A \cap C) + \mathsf{P}(B \cap C) - 3\mathsf{P}(A \cap B \cap C)$

4 ☐ $1 - \mathsf{P}(A)\mathsf{P}(B)\mathsf{P}(C) - \mathsf{P}(A^c)\mathsf{P}(B^c)\mathsf{P}(C^c)$

5 ☐ Ingen af de ovenstående

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

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

Dagens emner 5.1 og 5.2

- Ligefordeling med to variable (5.1): $\mathsf{P}\left((x,y) \in C\right) = \frac{A(C)}{A(D)}$
- Simultane tætheder (5.2)
  $f(x,y)\mathrm{d}x\mathrm{d}y \tilde{=} \mathsf{P}(x \le X \le x + \mathrm{d}x, y \le Y \le y + \mathrm{d}y), f(x,y) \ge 0$
- Simultan fordelingsfunktion (5.2)
  $$F(x,y) = \mathsf{P}(X \le x, Y \le y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u,v)\mathrm{d}u\mathrm{d}v$$
- 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)
  $$f(x,y) = 5!x(y-x)(1-y) \quad , \qquad \text{for} \quad 0 < x < y < 1$$