

# Sandsynlighedsregning

## 11. forelæsning

Bo Friis Nielsen

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

### Betingede fordelinger i det kontinuerte tilfælde

Første skud kunne være at se på de betingede sandsynligheder:

$$P(Y \leq y | X = x)$$

... men de er ikke veldefinerede fordi den hændelse vi betinger med,  $X = x$ , har sandsynlighed 0.

I stedet betinger vi på  $X \in dx$ , og lader intervallet  $dx$  gå mod punktet  $x$ .

Grænsen er veldefineret når  $(X, Y)$  har simultan tæthed  $f_{XY}(x, y)$  og den marginale tæthed af  $X$  er forskellig for 0 i  $x$ :

$$P(Y \in dy | X \in dx) = \frac{f_{XY}(x, y) dx dy}{f_X(x) dx} + o(dy) = \frac{f_{XY}(x, y) dy}{f_X(x)} + o(dy)$$

Det geometriske billede på s. 412 er meget nyttigt!

### Dagens nye emner afsnit 6.3 (og 6.4)

- Betingede kontinuerte fordelinger  $f(y|X = x) = \frac{f(x,y)}{f_X(x)}$
- Integralformel for gennemsnit af betingede sandsynligheder  $P(A) = \int P(A|X = x) f_X(x) dx$
- Betinget forventning  $E(Y) = E(E(Y|X))$
- Kovarians  $Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$ ,
- Korellation  $Corr(X, Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)}$
- Varians af sum af variable  $Var\left(\sum_{k=1}^N X_k\right) = \sum_{k=1}^n Var(X_k) + 2 \sum_{1 \leq j < k \leq n} Cov(X_j, X_k)$
- Bilinearitet af kovariansen  $Cov\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j Cov(X_i, Y_j)$



### Kontinuerte betingede fordelinger



Maskineriet kører som i det diskrete tilfælde; erstat summer med integraler.

F.eks. multiplikationsreglen:

$$f_{XY}(x, y) = f_X(x) f_Y(y | X = x)$$

Betinget forventningsværdi af en funktion:

$$E(g(Y) | X = x) = \int g(y) f_Y(y | X = x) dy$$

Ubetingede forventningsværdier:

$$E(g(Y)) = \int E(g(Y) | X = x) f_X(x) dx$$

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



### Spørgsmål 1

Den betingede tæthed af  $X$  givet  $Y$   $f(x|Y = y)$  findes til

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

### Integralformel for gennemsnit af betingede sandsynligheder



$$P(A) = \int P(A|X = x)f_X(x)dx$$

- Betragt  $(\Pi, Y)$  hvor  $\Pi$  er ligefordelt  $(0,1)$  og  $P(Y = y|\Pi = p) = \binom{n}{y}p^y(1-p)^{n-y}$
- Hvad er  $P(Y = y)$ ?

$$\begin{aligned} P(Y = y) &= \int_0^1 \binom{n}{y} p^y (1-p)^{n-y} \cdot 1 \cdot dp \\ &= \frac{1}{n+1} \int_0^1 \frac{(n+1)!}{y!(n-y)!} p^y (1-p)^{n-y} dp \end{aligned}$$

### Eksempel på betinget fordeling

- For givet  $X = x$  er  $Y \in \text{exponential}(x)$ , hvor  $X \in \text{exponential}(1)$
- Vi finder den marginale tæthed af  $Y$  ved



$$\begin{aligned} f_Y(y) &= \int_0^\infty f(x, y) dx = \int_0^\infty f_Y(y|X = x)f_X(x) dx \\ &= \int_0^\infty x e^{-xy} 1 \cdot e^{-x} dx = \frac{1}{y+1} \int_0^\infty x(y+1)e^{-(y+1)x} dx \end{aligned}$$

- Integralet er middelværdien af en eksponentialfordelt variabel med middelværdi  $\frac{1}{y+1}$ . Vi finder


$$f_Y(y) = \frac{1}{(1+y)^2}$$

Man har den stokastiske variabel  $X \geq 0$  med tæthed  $f_X(x) = \lambda^2 x e^{-\lambda x}$  samt den betingede fordeling af  $Y$  givet  $X = x$ , der er Poisson( $x$ ).

*Spm:* Hvad er fordelingen af  $Y$ ?

*Svar:*

$$\begin{aligned} P(Y = y) &= \int_0^\infty P(Y = y|X = x)f_X(x) dx \\ &= \int_0^\infty \frac{x^y}{y!} e^{-x} \lambda^2 x e^{-\lambda x} dx = \frac{\lambda^2}{y!} \int_0^\infty x^{y+1} e^{-(\lambda+1)x} dx \\ &= \frac{\lambda^2}{y!} \int_0^\infty (\lambda+1) \frac{((\lambda+1)x)^{y+1}}{(y+1)!} e^{-(\lambda+1)x} dx \frac{(y+1)!}{(\lambda+1)^{y+2}} \\ &= \frac{\lambda^2 (y+1)!}{(\lambda+1)^{y+2} y!} = y \left( \frac{\lambda}{\lambda+1} \right)^2 \left( \frac{1}{\lambda+1} \right)^y \end{aligned}$$

Den kontinuerte stokastiske variabel  $0 \leq Y \leq 1$  har tætheden   
 $f_Y(y) = 6y(1 - y)$ , yderligere har man at den betingede fordeling af  $X$  givet  $Y = y$  er *binomial*(3,  $y$ ).

### Spørgsmål 2

I  $X$  værdimængde finder man  $P(X = x)$  til

- 1   $\frac{1}{4}$
- 2   $\frac{1}{20}(x + 1)(4 - x)$
- 3   $\frac{x}{10}$
- 4   $\binom{3}{x} \frac{1}{8}$
- 5   $\binom{3}{x} y^x (1 - y)^{3-x}$
- 6  Ved ikke

## Betinget forventning



$$E(Y|X = x) = \int_y y f(y|X = x) dy$$

Regneregler for betinget forventning er uændret fra det diskrete tilfælde

Man har stadig fortolkningen af  $E(Y|X)$  som en stokastisk variabel, og

$$E(Y) = E(E(Y|X))$$

## Betinget forventning



Man har den ligefordelte ( $[0;1]$ ) kontinuerte stokastiske variabel  $X$ , og den stokastiske variabel  $Y$ , der givet  $X = x$ , følger en *normal* ( $x^3 - x, x$ ) fordeling.

*Spm:* Hvad er middelværdien  $E(Y)$  af  $Y$ ?

*Svar:* Vi har  $E(Y|X = x) = x^3 - x$  så  $E(Y|X) = X^3 - X$

$$E(Y) = E(E(Y|X)) = E(X^3 - X) = \frac{1}{4} - \frac{1}{2} = -\frac{1}{4}$$

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



### Spørgsmål 3

Den betingede forventning af  $X$  givet  $Y = y$  findes til

- 1   $\frac{1}{2}$
- 2   $\frac{1+2|y|}{1+|y|}$
- 3   $\frac{1}{2(1-|y|)}$
- 4   $\frac{1+|y|}{2}$
- 5   $\frac{1+|y|}{1-|y|}$
- 6  Ved ikke

## Kontinuerte og diskrete variable

Man har den stokastiske variabel  $X \geq 0$  med tæthed

$f_X(x) = \lambda^2 x e^{-\lambda x}$  samt den betingede fordeling af  $Y$  givet  $X = x$ , der er  $\text{Poisson}(x)$ .

Spm: Hvad er  $E(Y)$ ?

Svar: Vi kan benytte resultatet fra før, at  $Y$  følger en negativ binomialfordeling. Eller:

$$E(Y|X) = X, \text{ så } E(Y) = E(E(Y|X)) = E(X) = \frac{2}{\lambda}$$

Ved det sidste lighedstegn har vi udnyttet, at  $X$  følger en  $\text{gamma}(2, \lambda)$  fordeling. Som kontrol kan vi beregne middelværdien i den negative binomialfordeling

$$E(Y) = \frac{2q}{p} = \frac{2 \frac{1}{\lambda+1}}{\frac{\lambda}{\lambda+1}} = \frac{2}{\lambda}$$

## Kovarians/korellation

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

- hvor (for kontinuerte variable)

$$E(XY) = \int_x \int_y x \cdot y f(x, y) dy dx$$

- Generel regel for varians af sum

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\text{SD}(X)\text{SD}(Y)}$$

- En slags mål for afhængigheden mellem  $X$  og  $Y$

## Nøgletal for fordelinger



- Vi har middelværdi og varians som nøgletal for marginale fordelinger  $f_X(x)$  og  $f_Y(y)$ .
  - ◊ Mål for fordelings placering og variation, med varierende informationsindhold.
- Kovarians, korrelation er et tilsvarende første mål for en eventuel sammenhæng/afhængighed af to variable.

## Relative andele



Antal beboere	0	1	2	3	4	5	6
1 værelse	0.011	0.085	0.007	0.001	0.000	0.000	0.000
2 værelser	0.017	0.264	0.096	0.014	0.003	0.001	0.001
3 værelser	0.009	0.112	0.104	0.032	0.013	0.003	0.002
4 værelser	0.005	0.044	0.052	0.023	0.016	0.004	0.002
5 værelser	0.002	0.010	0.016	0.009	0.007	0.002	0.001
6 værelser	0.002	0.006	0.010	0.006	0.006	0.002	0.002

$$E(XY) = 1 \cdot 1 \cdot 0.085 + 1 \cdot 2 \cdot 0.007 + \dots + 6 \cdot 6 \cdot 0.002 = 4.9711$$

- For korrelationskoefficienten finder vi

$$\text{Corr}(X, Y) = \frac{E(XY) - E(X)E(Y)}{SD(X)SD(Y)} = 0.4257$$

Man har de to hændelser  $A$  og  $B$ , således at  $P(A) = 0.3$   
 $P(B) = 0.4$  og  $P(A \cup B) = 0.5$ .

#### Spørgsmål 4

Hændelserne  $A$  og  $B$  er

- 1  Positivt afhængige
- 2  Uafhængige
- 3  Negativt afhængige
- 4   $A$  afhænger af  $B$ , medens  $B$  er uafhængig af  $A$
- 5   $A$  og  $B$  er gensidigt udelukkende.
- 6  Ved ikke

## Positiv/negativ afhængighed

- For to hændelser  $A$  og  $B$  har vi defineret uafhængighed, hvis  $P(A|B) = P(A|B^c) = P(A)$ , der er ensbetydende med  $P(A \cap B) = P(A)P(B)$ .
- Når dette ikke er tilfældet, kan vi have  $P(A \cap B) > P(A)P(B)$  positiv afhængighed, eller  $P(A \cap B) < P(A)P(B)$  negativ afhængighed.
- Sammenhæng med correlationskoefficienten for indikatorvariablene  $I_A$  og  $I_B$ .

## Bilinearitet af kovariansen

$$\text{Var} \left( \sum_{k=1}^N X_k \right) = \sum_{k=1}^n \text{Var}(X_k) + 2 \sum_{1 \leq j < k \leq n} \text{Cov}(X_j, X_k)$$

$$\text{Cov} \left( \sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j \right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(X_i, Y_j)$$

- Bemærk iøvrigt:  $\text{Cov}(X, X) = \text{Var}(X)$
- Kovariansen kan ses som en generalisering af variansen

Man har de to hændelser  $A$  og  $B$ , således at  $P(A) = 0.3$   
 $P(B) = 0.4$  og  $P(A \cup B) = 0.5$ . Lad nu  $X = I_A$  og  $Y = I_B$ ,  
hvor  $I_C$  er indikatorfunktionen for hændelsen  $C$ .



### Spørgsmål 5

Kovariansen  $Cov(X, Y)$  mellem  $X$  og  $Y$  findes til

- 1  -0.16
- 2  0
- 3  0.04
- 4  0.08
- 5  0.16
- 6  Ved ikke

## Afsnit 6.3 og 6.4

- Betingede kontinuerte fordelinger  $f(y|X = x) = \frac{f(x,y)}{f_X(x)}$
- Integralformel for gennemsnit af betingede sandsynligheder  
 $P(A) = \int P(A|X = x)f_X(x)dx$
- Betinget forventning  $E(Y) = E(E(Y|X))$
- Kovarians/korellation



$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

$$Corr(X, Y) = \frac{Cov(X, Y)}{SD(X)SD(Y)}$$

- $(X, Y)$  uafhængige  $\Rightarrow Corr(X, Y) = 0$
- Varians af sum af variable  
 $Var\left(\sum_{k=1}^N X_k\right) = \sum_{k=1}^N Var(X_k) + 2 \sum_{1 \leq j < k \leq n} Cov(X_j, X_k)$
- Bilinearitet af kovariansen  
 $Cov\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j Cov(X_i, Y_j)$