

Sandsynlighedsregning

10. forelæsning

Bo Friis Nielsen

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

En virksomhed, der producerer elektriske pærer har afdelinger i to byer. Fabrikken i by A står for $\frac{2}{3}$ af produktionen. Resten produceres i by B , udaf disse er 1% defekte.



Spørgsmål 1

Hvor stor en andel af det samlede antal pærer der produceres af virksomheden fremstilles som intakte i by B ?

- 1 $\frac{1}{300}$
- 2 $\frac{198}{300}$
- 3 $\frac{1}{100}$
- 4 $\frac{99}{300}$
- 5 Spørgsmålet kan ikke besvares pga. manglende oplysninger
- 6 Ved ikke

Dagens emner afsnit 6.1 og 6.2



- Betingede diskrete fordelinger $P(Y = y|X = x) = \frac{P(X=x, Y=y)}{P(X=x)}$
 - ◇ $P(Y = y) = \sum_x P(X = x)P(Y = y|X = x)$
- Betinget middelværdi (det diskrete tilfælde)
 - ◇ $E(Y|X = x) = \sum_y yP(Y = y|X = x)$
 - ◇ $E(Y|X)$ en stokastisk variabel
 - ◇ $E(Y) = E_X(E_Y(Y|X)) = \sum_x P(X = x)E(Y|X = x)$
 - ◇ $E(g(X, Y)|X = x) = E(g(x, Y)|X = x)$
 - ◇ $E(XY|X) = XE(Y|X)$

Spørgsmål 2

Man får nu den yderligere oplysning at defektprocenten for pærer produceret i by A er 0.005 (altså er 1 ud af 200 i gennemsnit defekt).



Givet en pære er defekt, hvad er da sandsynligheden for, at den er produceret i by B .

- 1 1
- 2 $\frac{2}{3}$
- 3 $\frac{0.01 \cdot \frac{1}{3}}{0.01 \cdot \frac{1}{3} + 0.005 \cdot \frac{2}{3}}$
- 4 $\frac{1}{3}$
- 5 $\frac{\frac{102}{300} \cdot \frac{1}{3}}{\frac{102}{300} \cdot \frac{1}{3} + 0.005 \cdot \frac{2}{3}}$
- 6 Ved ikke

Københavnske boliger 1989 (antal)



Antal beboere:	0	1	2	3	4	5	6
1 værelse	3166	23554	2038	260	69	19	8
2 værelser	4874	73508	26670	3826	955	248	145
3 værelser	2408	31193	29023	8781	3735	925	565
4 værelser	1343	12301	14374	6487	4380	1064	585
5 værelser	436	2877	4376	2487	1977	538	264
6 værelser	442	1553	2748	1775	1543	615	521

- En bolig er således her karakteriseret ved to egenskaber størrelse - X og antal beboere - Y

Antallet af beboere i en 3 værelses bolig

- Hvis „nogen“ har valgt en tilfældig bolig, og fortæller os at den har tre værelser, hvilken sandsynlighedsfordeling vil vi så tillægge antallet af beboere?

Fra rækken med 3 værelses boliger i tabellerne har vi:

Beboere:	0	1	2	3	4	5	6	Alle
Antal:	2408	31193	29023	8781	3735	925	565	76630
Andel:	0,009	0,112	0,104	0,032	0,013	0,003	0,002	0,275

Andel af 3

værelses	0,031	0,407	0,379	0,115	0,049	0,012	0,007	1,000
----------	-------	-------	-------	-------	-------	-------	-------	-------

- Når vi ved at boligen har 3 værelser, normerer vi sandsynlighederne så de summerer til 1 *henover 3-værelses boliger*.

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

- Disse relative andele kan fortolkes som sandsynligheder, f.eks. $P(X = 1, Y = 0) = 0,011$ - for hvilket stokastisk eksperiment?

Betinget sandsynlighed/fordeling



Husk formelen for den betingede sandsynlighed af hændelsen $Y = y$, givet hændelsen $X = x$ (kapitel 1):

$$P(Y = y|X = x) = \frac{P(Y = y \cap X = x)}{P(X = x)} = \frac{P(Y = y, X = x)}{P(X = x)}$$

Betragter vi x som en fast parameter i dette udtryk og y som en variabel, er $P(Y = y|X = x)$ en sandsynlighedsfordeling for en stokastisk variabel (check det!).

Vi kalder den *Den betingede fordeling af Y givet $X = x$* .

Konkret får vi

$$P(Y = y|X = 3) = \frac{P(Y = y, X = 3)}{P(X = 3)} = \frac{P(X = 3, Y = y)}{0,275}$$

som i tabellen

Betinget fordeling afsnit 6.1



- For fastholdt værdi af $X = x$ kan Y variere over hele eller dele af sit udfaldsrum.
- Vi kalder fordelingen af Y for fastholdt $X = x$ for den betingede fordeling af Y givet $X = x$

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

Vi har den meget vigtige:

$$P(Y = y) = \sum_x P(Y = y \cap X = x) = \sum_x P(Y = y|X = x)P(X = x)$$

Reglen om gennemsnit af betingede fordelinger

Vi fandt for en familie med $C = c$ børn

$$P(G = i|C = c) = \binom{c}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{c-i}, \quad i = 0, \dots, c$$

Vi skal nu finde den *ubetingede* fordeling af antallet af piger G . Vi benytter reglen om gennemsnittet af betingede sandsynligheder (boksen side 41 eller Chapter 1 summary side 73).

$$\begin{aligned} P(G = 0) &= P(G = 0|C = 0)P(C = 0) + P(G = 0|C = 1)P(C = 1) \\ &\quad + \dots + P(G = 0|C = 4)P(C = 4) \\ &= \sum_{c=0}^4 P(G = 0|C = c)P(C = c) \\ &= \left(\frac{1}{2}\right)^0 \frac{1}{10} + \left(\frac{1}{2}\right)^1 \frac{2}{10} + \left(\frac{1}{2}\right)^2 \frac{4}{10} + \left(\frac{1}{2}\right)^3 \frac{2}{10} + \left(\frac{1}{2}\right)^4 \frac{1}{10} = \frac{53}{160} \end{aligned}$$

Opgave 6.1.2



I en bestemt by har 10% af familierne ingen børn, 20 % har et barn, 40% har to børn, 20 % har tre børn og 10 % har fire børn. Antag, at der er 50% chance for, at det enkelte barn er en pige uafhængigt af antallet af børn i familien. Bestem sandsynlighedsfordelingen for antallet af piger i en familie.

- Vi definerer først relevante stokastiske variable: C antallet af børn i en familie, og G antallet af piger i en familie.
- I en familie med $C = c$ børn finder vi

$$P(G = i|C = c) = \binom{c}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{c-i}, \quad i = 0, \dots, c$$

Det var for 0 piger. For de andre muligheder får vi:

$$P(G = i) = \sum_{j=0}^4 P(G = i|C = j)P(C = j)$$



Det vil sige:

$$\begin{aligned} P(G = 4) &= \frac{1}{16} \cdot \frac{1}{10} = \frac{1}{160} & P(G = 3) &= \frac{1}{8} \cdot \frac{1}{5} + \frac{4}{16} \cdot \frac{1}{10} = \frac{1}{20} \\ P(G = 1) &= \frac{2}{5}, & P(G = 2) &= \frac{17}{80} \end{aligned}$$

Den generelle formel er:

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

Man kaster tre mønter. De mønter, der lander "plat" kastes igen. Lad X betegne antallet af "krone" efter første kast og Y betegne antallet af "krone" efter andet kast ($0 \leq X \leq Y \leq 3$).

Spørgsmål 3

Angiv den betingede fordeling $f_Y(y|X = x) = P(Y = y|X = x)$ af Y for $X = x$

- 1 $f_Y(y|X = x) = \frac{3!}{y!(3-y)!} \frac{1}{8}, \quad y = x, \dots, 3$
- 2 $f_Y(y|X = x) = \frac{(3-x)!}{(y-x)!(3-y)!} \frac{1}{2^{3-x}}, \quad y = x, \dots, 3$
- 3 $f_Y(y|X = x) = \frac{1}{4-x}, \quad y = x, \dots, 3$
- 4 $f_Y(y|X = x) = \frac{(3+x)!}{(y+x)!(3-y)!} \frac{1}{2^{3-x}}, \quad y = x, \dots, 3$
- 5 $f_Y(y|X = x) = \frac{(3+x)!}{(y+x)!(3-y)!} \frac{1}{8}, \quad y = x, \dots, 3$
- 6 Ved ikke

Middelantallet af beboere i en treværelses bolig

Beboere:	0	1	2	3	4	5	6	Alle
Andele	0,031	0,407	0,379	0,115	0,049	0,012	0,007	1,000

Hvis vi ved at boligen er treværelses, beregner vi middelværdien i den betingede fordeling:

Betinget middelværdi afsnit 6.2

$$E(Y|X = x) = \sum_y yP(Y = y|X = x)$$

$$E(Y|X = 3) = 0,031 \cdot 0 + \dots + 6 \cdot 0,007 = 1,808$$

Uafhængighed og betingede fordelinger



Hvis X og Y er uafhængige har vi

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

og dermed

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

Omvendt, hvis dette holder $\forall x, y$, må X og Y være uafhængige.

Altså: X og Y er uafhængige, hvis og kun hvis den betingede fordeling af Y givet X er identisk med den ubetingede fordeling af Y .

Med andre ord: Hvis X ikke indeholder information af Y .

Betinget middelværdi som stokastisk variabel

Hvad er en stokastisk variabel?



- Vi kan tænke på den som en variabel hvis værdi fastlægges gennem et (stokastisk) eksperiment, vi ikke har udført endnu

Her er sådan en konstruktion:

1. Vælg en tilfældig bolig.
2. Lad X være antallet af værelser i boligen.
3. Lad Z være det gennemsnitlige antal beboere blandt alle de boliger, der har X værelser.

Vi ser at $Z = E(Y|X)$ er en stokastisk variabel.

- Vi bruger sandsynlighedsregning til at udtrykke, hvad vi ved om dens opførsel. I et eksperiment, hvor vi måler både X og Y , udtrykker den simultane fordeling, hvad vi ved om X 's og Y 's opførsel før eksperimentet.

Betingede forventningsværdier



som stokastiske variable

$Z = E(Y|X)$ er en stokastisk variabel.

Den kan fastlægges uden at kende andet til realisationen (altså, hvilken bolig vi valgte) end X .

Vi kan altså skrive

$$Z = g(X)$$

for en passende funktion g . I dette tilfælde er $g(x)$ givet ved algoritmen: Tag det gennemsnitlige antal beboere blandt de boliger der har x værelser.

Information og betingede fordelinger



Lad X og Y være to stokastiske variable knyttet til samme eksperiment, og forestil jer en observatør der kun får adgang til X .

For hende vil den realiserede værdi af Y være ukendt, både før (*a priori*) og efter (*a posteriori*) eksperimentet udføres.

For hende vil Y *a priori* være fordelt efter den *ubetingede* fordeling $P(Y = y)$. Den vigtigste størrelse i denne fordeling er forventningsværdien $E(Y)$.

Efter eksperimentet vil hun have adgang til informationen $X = x$. Derfor reviderer hun sin opfattelse af Y 's fordeling til den *betingede* fordeling $P(Y = y|X = x)$. Den vigtigste størrelse i denne fordeling er den betingede forventningsværdi $E(Y|X = x)$.

Hvad er fordelingen af $Z = E(Y|X)$?



For boligeksemplet finder vi:

x	1	2	3	4	5	6
$P(X = x)$	0,104	0,396	0,275	0,145	0,046	0,033
$E(Y X = x)$	0,9902	1,3087	1,8079	2,1429	2,4139	2,6908

Dermed har vi direkte fordelingen af $Z = E(Y|X)$:

z	0,9902	1,3087	1,8079	2,1429	2,4139	2,6908
$P(Z = z)$	0,104	0,396	0,275	0,145	0,046	0,033

Hvad er forventningsværdien af $Z = E(Y|X)$?

Man har

$$E(Y) = E(Z) = E(E(Y|X)) = \sum_x E(Y|X = x)P(X = x)$$

Intuitivt: Informationen $X = x$ vil ændre observatørens forventningsværdi af Y , men ikke *i middel*.

- I behøver ikke forstå dette i dybden, men I skal kunne anvende ovenstående formel (opgave 6.2.4 træner dette).
- Den *betingede middelværdi* er et vigtigt begreb/redskab i anvendelserne (statistik, signalbehandling, billedbehandling)

For boligeeksemplet:



Vi havde fordelingen af $Z = E(Y|X)$:

z	0,9902	1,3087	1,8079	2,1429	2,4139	2,6908
$P(Z = z)$	0,104	0,396	0,275	0,145	0,046	0,033

hvoraf

$$\begin{aligned} E(Y) &= E(Z) \\ &= 0,9902 \cdot 0,104 + 1,3087 \cdot 0,396 + \dots + 2,6908 \cdot 0,033 \\ &= 1,6250 \end{aligned}$$

hvilket stemmer med hvad vi får direkte fra den marginale fordeling af Y .

Lad X_1 og X_2 være antal øjne i to kast med en almindelig terning. Lad nu X betegne minimum og Y betegne maximum af X_1 og X_2 .

Spørgsmål 5

Bestem $E(Y|X)$

- 1 $\frac{7}{2}$ med $P = \frac{1}{6}$, 4 med $P = \frac{1}{6}$, $\frac{9}{2}$ med $P = \frac{1}{6}$,
5 med $P = \frac{1}{6}$, $\frac{11}{2}$ med $P = \frac{1}{6}$, 6 med $P = \frac{1}{6}$
- 2 3.5
- 3 $\frac{41}{11}$ med $P = \frac{11}{36}$, $\frac{38}{9}$ med $P = \frac{9}{36}$, $\frac{33}{7}$ med $P = \frac{7}{36}$,
 $\frac{26}{5}$ med $P = \frac{5}{36}$, $\frac{17}{3}$ med $P = \frac{3}{36}$, 6 med $P = \frac{1}{36}$
- 4 $\sum_{i=X}^6 i \cdot \frac{1}{7-X}$
- 5 $\frac{X+6}{2}$
- 6 Ved ikke



Lad X_1 og X_2 være antal øjne i to kast med en almindelig terning. Lad nu X betegne minimum og Y betegne maximum af X_1 og X_2 .

Spørgsmål 4

Bestem $E(Y|X = 4)$

- 1 $4 \cdot \frac{1}{4} + 5 \cdot \frac{1}{2} + 6 \cdot \frac{1}{4}$
- 2 3.5
- 3 4.5
- 4 $4 \cdot \frac{1}{3} + 5 \cdot \frac{1}{3} + 6 \cdot \frac{1}{3}$
- 5 $4 \cdot \frac{1}{5} + 5 \cdot \frac{2}{5} + 6 \cdot \frac{2}{5}$
- 6 Ved ikke

Afsnit 6.1 og 6.2



- Betingede diskrete fordelinger $P(Y = y|X = x) = \frac{P(X=x, Y=y)}{P(X=x)}$
 - ◇ $P(Y = y) = \sum_x P(X = x)P(Y = y|X = x)$
- Betinget middelværdi (det diskrete tilfælde)
 - ◇ $E(Y|X = x) = \sum_y yP(Y = y|X = x)$
 - ◇ $E(Y|X)$ en stokastisk variabel
 - ◇ $E(Y) = E_X(E_Y(Y|X)) = \sum_x P(X = x)E(Y|X = x)$
 - ◇ $E(g(X, Y)|X = x) = E(g(x, Y)|X = x)$
 - ◇ $E(XY|X) = XE(Y|X)$