Contents

1 Disclaimer

Disse noter blev udarbejdet i forbindelse med jeg underviste i kurset **Sandsynlighedsteori og statistik** udbudt af Økonomisk Institut, Københavns Universitet.

Dette er ikke blevet gennemlæst, rettet eller på anden måde redigeret af en tredje person, som ville kunne fange evt. fejl og mangler. Derfor **forvent** at der er fejl i dette dokument. Forhold dig kritisk til resultaterne, og hvis du er sikker på der er en fejl, så tag udgangspunkt i det.

Dokumentet indeholder rettevejledninger til øvelsesseddlerne forbundet med faget. Der er et tilhørende github-repository:

https://github.com/JakartaLaw/statistik2018.

Lecture Notes

Jeppe Johansen

December 4, 2018

2 Topics

readings: Sørensen 1.2-1.3

2.1 Det endelige sandsynlighedsfelt

Et endelig sandsynlighedsfelt har følgende egenskaber:

- En endelig mængde $E = \{e_1, e_2, \ldots, e_j\}$
- En funktion p fra E ind i intervallet $[0, 1]$
- Summen af samtlige sandsynligheder skal være 1:

$$\sum_{j=1}^{k} p(e_j) = 1 \tag{1}$$

2.1.1 Hændelser

A er en hændelse:

$$A \subseteq E \tag{2}$$

Sandsynligheden for A:

$$P(A) = \sum_{x \in A} p(x) \tag{3}$$

Sandsynligheden for to disjunkte mængder A, B:

$$P(A \cup B) = P(A) + P(B), \qquad P(A) \cap P(B) = \emptyset \tag{4}$$

2.2 Ved konstant ssh funktion

$$\text{ssh. for hændelse} = \frac{\#\text{ gunstige udfald}}{\#\text{ mulige udfald}} \tag{5}$$

2.2.1 Termer

- **Udfald** de enkelte elementer i E
- **Hændelse** en delmængde
- **Sandsynlighedsfunktionen** er $p(\cdot)$
- **punktsandsynligheden** for e_j er $p(e_j)$
- **disjunkte** er to mængde som har den tomme mængde som fællesmængde
- **Sandsynlighedsmål** er funktionen P fra klassen af delmængder af E. (har ekstra krav, se p. 14 i **Sørensen**)
- # antal elementer i et sæt

2.3 Det generelle sandsynlighedsfelt

Hvis mængden er uendelig stor, (både tællelig og utællelig) kigger man på delintervaller af E.

Under antagelse af ligefordeling:

$$P(I) = c|I| \tag{6}$$

$|I|$ betegner længden af linjestykket på den reelle akse.

2.3.1 Definition af sandsynlighedsfelt

- Et udfaldsrum E
- En klasse $\mathcal{E}$ af delmængder fra E

- En funktion P fra $\mathcal{E}$ ind i $[0,1]$
- $\mathcal{E}$ skal indeholde både E og $\emptyset$
- P skal opfylde

$$P(E) = 1 \tag{7}$$
$$P(A \cup B) = P(A) + P(B), \quad A \cap B = \emptyset \tag{8}$$

$\mathcal{E}$ er kun en klasse af pæne delmængder. Dette er ikke et problem på dette kursus (eller andre på økonomisk institut).

2.3.2 Indikator funktion

En indikator funktion kan tage en af de to værdier $\{0,1\}$. $\mathbb{1}_A(x) = 1$ hvis $x \in A$, ellers 0

2.3.3 Regneregler for sandsynlighedsmål P (sætning 1.3.4)

1. regler hvis $B \subseteq A$:

$$P(A \setminus B) = P(A) - P(B) \tag{9}$$
$$P(B) \leq P(A) \tag{10}$$

2. regler for den komplementære hændelse til B. i.e. $E \setminus B$

$$P(E \setminus B) = 1 - P(B) \tag{11}$$

3.

$$P(\emptyset) = 0 \tag{12}$$

4.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{13}$$

5.

$$P(A) + P(B) \leq P(A) + P(B) \tag{14}$$

ligning 14 kan udvides til vilkårligt mange mængder. Det bliver en lighed hvis samtlige vilkårlige mængder er disjunkte.

---

readings: Sørensen 1.4-1.5

2.4 Betingede sandsynligheder og uafhængighed

2.4.1 Den betingede sandsynlighed af B giver A skrevet $P(B \mid A)$, er defineret ved (definition 1.4.1

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \tag{15}$$

2.4.2 Regneregler for betingede sandsynligheder

$A_1, A_2, \cdots A_n$ være n hændelser, hvor $P(A_1 \cap A_2 \cap \cdots A_{n-1} > 0$. Da:

$$P(A_1 \cap A_2 \cap \cdots A_n) = \tag{16}$$

$$P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \cdots P(A_n \mid A_1 \cap \cdots \cap A_{n-1}) \tag{17}$$

Endnu en regneregel

Hvis $A_1, A_2, \cdots A_n$ er n disjunkte hændelser, hvor at $E = \bigcup_{i=1}^{n} A_i$ samt $P(A_i) > 0$, da gælder for en vilkærlig hændelse B:

$$P(B) = \sum_{j=1}^{n} P(B \mid A_j)P(A_j) \tag{18}$$

2.4.3 Omvendingsformel - simpel bayes'

$$P(A \mid B) = P(B \mid A)\frac{P(A)}{P(B)} \tag{19}$$

2.4.4 Bayes' formel

$A_1, A_2, \cdots A_n$ er n disjunkte hændelser, hvor at $E = \bigcup_{i=1}^{n} A_i$ samt $P(A_i) > 0$. For en hændelse B med $P(B) > 0$, da gælder for en enhver hændelse k:

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{\sum_{j=1}^{n} P(B \mid A_j)P(A_j)} \tag{20}$$

2.5 Stokastisk Uafhængighed

Uafhængighed tænkes oftest som:

$$P(A \mid B) = P(A) \tag{21}$$

Altså at sandsynligheden for A ikke er påvirket af udfaldet af B.

2.5.1 Definition af uafhængighed

hændelse A og B er uafhængige siges at være stokastisk uafhængige når (definition 1.5.1) :

$$P(A \cap B) = P(A) \cdot P(B) \tag{22}$$

Dette udsagn kan let udvides til n hændelser (se p.34 **definition 1.5.4**)

2.5.2 Regler for indbyrdes uafhængighed

Tegn for uafhængighed $\perp\!\!\!\perp$.

A, B og C er indbyrdes uafhængige hændelser. Følgende gælder:

1. $A \setminus B \perp\!\!\!\perp C$
2. $A \cap B \perp\!\!\!\perp C$
3. $A \cup B \perp\!\!\!\perp C$
4. $E \setminus A, B \perp\!\!\!\perp C$

2.5.3 forenings mængdens uafhængighed

A, B, C er hændelser. A og B er betinget afhængige givet C hvis:

$$P(A \cap B \mid C) = P(A \mid C) \cdot P(B \mid C) \tag{23}$$

denne kan generaliseres (se p. 37 **definition 1.5.7**)

2.6 Fordelinger på endelige mængder

2.6.1 Centrale begreber

Adskiller sig fra diskrete fordelinger, som ikke er på endelige mængder, men fx. alle de positive tal

Binomialfordelingen sandsynlighedsfunktion er givet ved

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \tag{24}$$

$$\binom{n}{x} = \frac{n!}{x!((n-x)!} \tag{25}$$

Transformation af fordelinger

Middelværdi

$$\mathrm{E}(X) = \sum_{i=1}^{k} a_i p(a_i) \tag{26}$$

Varians

$$\mathrm{Var}(X) = \mathrm{E}([X - E(X)]^2) \tag{27}$$

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 \tag{28}$$

Kovarians

$$\mathrm{Cov} = \mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))) \tag{29}$$

$$\mathrm{Cov} = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) \tag{30}$$

Korrelation

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \tag{31}$$

2.7 Flerdimensionale stokastiske variable, uafhængighed

readings: Sørensen 3.5, 4.1, 4.4

2.7.1 Polynomial fordelingen

2.7.2 Poisson fordelingen

$$p(x) = \frac{\lambda^x}{x!}e^{-x}, x \in N_0 \tag{32}$$

2.7.3 Generelle diskrete fordelinger

- Fordelinger på uendelige tællelige mængder
- $\sum_{i=1}{}^{\infty} p(x_i) = 1$
- Kontinuert fordeling af 1 dimension
- middelværdi
- varians
- normalfordelingen
- transformationer
- χ^2 fordeling

2.8 Middelværdi

X har en middelværdi hvis:

$$\int_{-\infty}^{\infty} |x|p(x)dx < \infty \tag{33}$$

Middelværdien for X er:

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} xp(x)dx < \infty \tag{34}$$

2.9 Varians

Variansen ekstisterer hvis:

$$\int_{-\infty}^{\infty} x^2 p(x)dx < \infty \tag{35}$$

Variansen er:

$$\mathrm{Var}(X) = \mathrm{E}([X - \mathrm{E}(X)]^2) \tag{36}$$

2.9.1 Normalfordelingen

Standard normalfordeling:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad x \in \mathbb{R} \tag{37}$$

Fordelingsfunktionen (CDF)

$$\Phi(x) = \int_{-\infty}^{x} \phi(y)dy \tag{38}$$

Den generelle statndard normalfordeling:

$Y = \mu + \sigma X$

$$p(y) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \tag{39}$$

med middelværdi μ og varians σ^2

2.10 Transformationer

$$q(y) = \begin{cases} p(t^{-1}(y))|\frac{d}{dy}t^{-1}(y)|, & y \in (v,h) \\ 0, & y \notin (v,h) \end{cases} \tag{40}$$

hvor $v = \inf t(I), h = \sup t(I)$ og I er intervallet (a,b)

2.11 Fler dimensionelle kontinuerte fordelinger

Man har $p(x,y)$ og $A \in \mathbb{R}$

$$P(X \in A) = P((X,Y) \in A \times \mathbb{R}) \tag{41}$$

Hvilket kan skrives som:

$$q(x) = \int_{\mathbb{R}} p(x,y)dy \tag{42}$$

Altså man kan integrere irrelevante variable ud!

2.11.1 Uafhængighed

$X_1, X_2, \cdots X_n$ er uafhængige. Dette betyder:

$$p(x_1, x_2, \ldots, x_n) = p(x_1) \cdot p(x_2) \cdot \ldots \cdot p(x_n) \tag{43}$$

Sætning:

> *Hvis vi ikke kan finde en produktmængde $T_1 \times T_2$, sæledes at (X_1, X_2) er koncentreret på $T_1 \times T_2$, og således at $p(x_1, X_2) > 0$ for alle $(x_1, x_2) \in T_1 \times T_2$ så kan X_1, X_2 ikke være uafhængige*

2.11.2 Transformation af kontinuerte variable

Kig **6.3.2**, **6.3.5**, **6.3.6**, **6.3.7** for eksempler på to dimensionelle transformationer. (X + Y), (X / Y) og lignende.

Sætning 6.3.10 viser sandsynlighedstætheden for $q(y_1, y_2) = q(t_1(x_1, x_2), t_2(x_1, x_2))$. Hvor transformationen er på formen $Y_1 = aX_1 + bX_2$ og $Y_2 = cX_1 + dX_2$

Generelt:

sætning 6.3.11

$$Y = AX \tag{44}$$

hvor at $\det(A) \neq 0$ og A er en $n \times n$ matrice og Y er n-dimensionel. da er Y's tæthed:

$$q(y) = \frac{p(A^{-1}y)}{|\det(A)|} \tag{45}$$

2.11.3 Middelværdi, Varians og Kovarians

Resultater vist for diskrete stokastiske fordelinger er de samme som for kontinuerete (integraler i stedet for summer)

$$E(X_1 + X_2 + \ldots X_n) = E(X_1) + E(X_2) \ldots + E(X_n) \tag{46}$$

Hvis de er uafhængige, da:

$$E(X_1 \cdot X_2 \cdot \ldots \cdot X_n) = E(X_1) \cdot E(X_2) \cdot \ldots \cdot E(X_n) \tag{47}$$

2.11.4 Kontinuerte betingede fordelinger

Situation hvor man vil betinge på at $X = x$:

$$q(y) = p(x, y)/p_1(x) \tag{48}$$

hvor $p_1(x)$ bare er hvor y er integreret ud af $p(x, y)$.

2.12 Normalfordelingsteori

χ^2 fordelingens tæthed med k-frihedsgrader

$$p(x) = \frac{x^{\frac{k}{2}} e^{-x/2}}{2^{k/2} c_k} \tag{49}$$

Kig nærmere i bogen for den 2-dimensionelle normalfordeling!

2.13 Grænseresultater for stokastiske variable

readings: Sørensen 7

2.13.1 Store tals lov

For n ukorrelerede stokastiske variable med middelværdi μ og varians σ^2 vil gennemsnittet (det empiriske) være tæt på middelværdien μ.

Altså gennemsnittet vil konvergere med middelværdien. Gennemsnittet er stadig en stokastisk variabel

2.13.2 Den centrale grænseværdi sætning

Her er det værd at notere:

$$\mathrm{E}(\bar{X}_n) = \mu \tag{50}$$

Og måske mindre klart:

$$\mathrm{Var}(\bar{X}_n) = \frac{1}{n^2}(\mathrm{Var}(X_1)+\mathrm{Var}(X_2)+\cdots+\mathrm{Var}(X_n)) = \frac{1}{n^2}(n\mathrm{Var}(X_i)) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n} \tag{51}$$

Man altså se at vi kan standardisere den stokastiske variabel som er middelværdien:

$$U_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \tag{52}$$

Nu er reskaberne til sætningen klar:

En række af stokastiske uafhængige varible, som er identiske. Da vil fordelingen af U_n konvergere mod en standard normalfordeling når $n \to \infty$

Dette kaldes konvergens i fordeling

2.14 Deskriptiv statistik

Vi har indsamlet noget data:

$$y_1, y_2, \cdots, y_n \tag{53}$$

Man kan forestille sig en DGP (data generende proces) have forskellige karakteristika, hvormed den mapper til en respons variabel, som kan være:

- binær
- tælle
- diskret
- kontinuær

frekvens for j'te element i $\mathbb{Y}$ udregnes ved:

$$f_{y=j} \sum_{i=0}^{N} \mathbb{1}(y_i = j) \tag{54}$$

Den empiriske cumulative distribution

$$F(y) = \sum_{i=1}^{N} \frac{\mathbb{1}(y_i \leq y)}{n} \tag{55}$$

Empriske momenter:

- mean (gennemsnit) $= \frac{1}{N} \sum_{i=1}^{N} y_i$
- Varians $= \frac{1}{N} \sum (y_i - \bar{y})^2$
- Skewness $= \frac{1}{N} \sum (y_i - \bar{y})^3$
- Kurtosis $= \frac{1}{N} \sum (y_i - \bar{y})^4$

standard afvigelse:

$$std(Y) = \sqrt{Var(Y)} \tag{56}$$

Excess Kurtosis vil sige Kurtosis - 3. Da en standard normal fordeling har kurtosis på 3.

Man vil ofte standardisere data når man kigger på skewness og kurtosis.

Standardisering af data er:

$$y_{i,standardised} = \frac{y_i - \bar{y}}{std(y)} \tag{57}$$

hvor *std* betyder standard deviation (standard afvigelse).

2.15 Kapitel 3: Likelihood funktionen

Under i.i.d antagelser kan man udlede

$$\prod_{i=1}^{n} f_{Y_i}(y_i \mid \theta) \tag{58}$$

Likelihoid contribution

$$l(\theta \mid y_i) = f_{Y_i}(y_i \mid \theta) \tag{59}$$

$$L(\theta \mid y_1, y_2, \cdots, y_n) = \prod_{i=1}^{n} l(\theta \mid y) \tag{60}$$

Tag logaritmen:

$$\sum_{i=1}^{n} \log l(\theta \mid y_i) \tag{61}$$

Maximér via differentiation.

2.15.1 Egenskaber ved Maximum Likelihood estimatoren

Nielsen kap. 4

Kig i discussion afsnit for opremsning af egenskaber!

2.16 Kapitel 6: Konfidens Intervaller og hypotese test

2.16.1 Konfindens Intervaller

Vi har fra Theorem 4.1

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N(0, \Omega_0) \tag{62}$$

Side (98) viser vi kan herfra komme til udtrykket:

$$\frac{\hat{\theta} - \theta_0}{\mathbf{se}(\hat{\theta})} \sim N(0, 1) \tag{63}$$

Altså vi har nu konstrueret en stokastisk variabel som er standard normalt fordelt. Vi ved at 95% af sandsynlighedsmassen ligger i intervallet:

$$-1.96 < Z < 1.96 \tag{64}$$

Hvor Z er en standard normalt fordelt stokastisk variabel. Alternativt kunne man sige:

$$P(-1.96 < Z < 1.96) = 0.95 \tag{65}$$

Derfra kan man altså let udlede at:

$$P(\hat{\theta} - 1.96 \cdot \mathbf{se}(\hat{\theta}) < \theta_0 < \hat{\theta} + 1.96 \cdot \mathbf{se}(\hat{\theta})) = 0.95 \tag{66}$$

2.16.2 Hypotese test

$$H_0 : \theta_0 = a \tag{67}$$

$$H_A : \theta_0 \neq a \tag{68}$$

Altså vi har en H_0 (det vi tester). Og en H_A, alternativet. Den urestriktere model kaldes H_U. Den urestrikterede model er den vi har arbejdet med i kurset op til dette punkt. Vi husker at Θ (vores parameter rum) er defineret af H_U Vi kan nu sige at:

Under H_0 er parameterrummet:

$$\theta \in \Theta_0 = \{a\} \tag{69}$$

Undwe H_A er parameterrummet

$$\theta \in \Theta_A = \{\theta \in \Theta : \theta \neq a\} \tag{70}$$

Man ser altså at: $\Theta_0 \cap \Theta_A = \emptyset$ og at: $\Theta_0 \cup \Theta_A = \Theta$

	H_0 er sand	H_0 er falsk
Test afvises ikke	Korrekt	Type 2 fejl
test afvises	Type 1 fejl	Korrekt

2.16.3 Wald-Test

Kig bog for eksempel.

$H_0 : \theta_0 = a$ og $H_A : \theta_0 \neq a$

$$p(\theta_0 = a) = \frac{\hat{\theta} - a}{\mathbf{se}(\hat{\theta})} = k \tag{71}$$

Hvor vi afviser H_0 hvis k er større end 1.96 eller mindre end -1.96.

Læs nærmere i kapitel for p-værdi.

Man kan også lave en *squared wald-test* Hvor man kvadrere Z. Her skal man teste i en χ^2-fordeling.

2.16.4 LR-test

Vi definere $\tilde{\theta}$ som:

$$\tilde{\theta} = \arg\max_{\theta \in \Theta_0} \sum_{i=1}^{n} \log l(\theta \mid y_i) \tag{72}$$

Faldet i likelihood under restriktionen af H_0 can blive måldt med

$$\frac{L_n \tilde{\theta}}{L_n \hat{\theta}} \tag{73}$$

Log likelihood ratio (LR)

$$LR_n(H_0) = -2 \log \left(\frac{L_n \tilde{\theta}}{L_n \hat{\theta}} \right) \tag{74}$$

$$= 2 \left[\log L_n(\hat{\theta}) - \log L_n(\tilde{\theta}) \right] \tag{75}$$

Det kan vises at $LR_n(H_0)$ konvergerer mod en $\chi^2(v)$-fordeling, under en sand 0 hypotese, hvor v er *antal frihedsgrader*. Som i dette tilfælde er antal restriktioner under H_0.

Kig bog for eksempler.

3 Lectures

3.1 Øvelse 1

10/09/2018, opgaver: 1.1, 1.4, 1.17, 1.24 (og 1.13 hvis der er tid)

3.1.1 Opgave 1.1

- en fair mønt
- 3 kast

Udfaldsrummet E har 2^3 udfald:

$E = \{(0,0,0),(0,0,1),(0,1,0),(0,1,1),(1,0,0),(1,0,1),((1,1,0),(1,1,1)\}$

SSH for 1 mønt=krone:

$$p((1,1,1)) = \frac{\text{Gunstige udfald}}{\text{mulige udfald}} = \frac{1}{8} \tag{76}$$

SSH for mindst 1 mønt=krone.

brug den komplementere sandsynlighed: $p((0,0,0)) = \frac{1}{8}$. Kald denne hændelse B.

$$P(E \setminus B) = 1 - \frac{1}{8} = \frac{7}{8} \tag{77}$$

SSH for præcis et kast viser mønt=krone

Vi definere 3 hændelser

- $A = \{\text{Det første kast bliver krone}\}$,
- $B = \{\text{Det andet kast bliver krone}\}$,
- $C = \{\text{Det tredje kast blive krone}\}$

Undersøg om dette er korrekt:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) = 3 \times \frac{1}{8} = \frac{3}{8} \tag{78}$$

spørgsmål: hvorfor kan vi ignorere fællesmængden: denoted $A \cup B \cup C$? =¿ Den er disjunkt. Kig i opgave 1.13 for at se hvordan man skulle have inkluderet fællesmængderne

3.1.2 Opgave 1.4

- 3 slag med terninger $\{1, 2, 3, 4, 5, 6\}$
- Ssh summen er 10

1) Vi ser at summen kan antage alle hele tal mellem 3 og 18.

2) Vi kan se det ikke er en ligefordeling af summer: dvs. summen 3 er ikke så hyppig som summen 10.

3) Det samlede antal udfald er 6^3

4) Via computer fandt jeg det gunstige antal udfald:

$$\frac{\text{antal gunstige udfald}}{\text{antal mulige udfald}} = \frac{27}{6^3} = \frac{27}{216} \tag{79}$$

3.1.3 Opgave 1.17

- 1 sort terning
- 1 hvid terning

del 1) Hvad er den betingede ssh. for at summen er 12 givet summen er mindst 11

Brug reglen for betingede sandsynligheder

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \tag{80}$$

Lad A være sandsynligheden for summen er 12.

Lad B være sandsynligheden for summen er mindst 11.

$$P(A) = p((6, 6)) = \frac{1}{36} \tag{81}$$

$$P(B) = P(\{(6, 6), (5, 6), (6, 5)\}) = \frac{3}{36} \tag{82}$$

vi ser at $A \subset B \Rightarrow P(A \cap B) = P(A)$

$$P(A \mid B) = \frac{P(A)}{P(B)} = \frac{\frac{1}{36}}{\frac{3}{36}} = \frac{1}{3} \tag{83}$$

Del 2) Hvad er den betingede SSH for at de to terninger viser det samme, givet summen er 7:

A er hændelsen for begge er terninger viser det samme.

B er hændelsen summer af terningerne er 7.

$A = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)$

$B = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$

Vi ser at $P(A \cap B) = \emptyset$

Man husker $P(\emptyset) = 0$ Givet fra definitioner af sandsynlighedsmål.

$$P(A \mid B) = 0 \tag{84}$$

Del 3) Ssh for den hvide terning viser 3, givet den sorte viser 5

A: er hændelsen at den hvide terning er 3.

B: er hændelsen den sorte terning er 5.

Hændelserne er uafhængige!

$$A = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\} \tag{85}$$

$$B = \{(1,5),(2,5),\cdots,(6,5)\} \tag{86}$$

Vi ser: $P(A \cup B) = P\{(3,5)\} = \frac{1}{6^2}$.

Vi ser: $P(B) = \frac{1}{6}$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6} \tag{87}$$

Del 4) Ssh. for den mindste terning viser 2, givet den terning med det højeste andel højest viser 5

A: hændelsen at den mindste terning viser 2.

B: hændelsen at den terning med det højeste antal øjne viser 5.

$A = \{(2,2),(2,3),\cdots,(2,6),(3,2),(4,2),\cdots,(6,2)$

$B = \bigcup_{i,j\in\{1,2,3,4,5\}}(i,j) = E\backslash\{(1,6),(2,6),\cdots,(5,6),(1,6),(2,6)\cdots(5,6),(6,6)\}$

Vi kan finde $A \cap B$:

$$A \cap B = (2,2),(2,3),(2,4),(2,5),(3,2),(4,2),(5,2) \tag{88}$$

$$P(A \cap B) = \frac{7}{6^2} = \frac{7}{36} \tag{89}$$

$$P(B) = P(E) - P(\{(1,6),(2,6),\cdots,(5,6),(1,6),(2,6)\cdots(5,6),(6,6)\}) \tag{90}$$

$$= 1 - \frac{11}{36} = \frac{25}{36} \tag{91}$$

$$P(A \mid B) = \frac{\frac{7}{36}}{\frac{25}{36}} = \frac{7}{25} \tag{92}$$

3.1.4 Opgave 1.24

- 1 hvid terning
- 1 sort terning
  - A = {den hvide terning viser 4}
  - B = {den sorte terning viser 1}
  - C = {terningen med det højeste antal øjne viser 4 }
  - D = {summen af øjene er 5 }
  - F = {summen af øjnene er 7 }

hvilke par er indbyrdes uafhængige:

Husk uafhængighed er: $P(A \cap B) = P(A)P(B)$

$$P(A) = \frac{1}{6} = \frac{6}{36} \tag{93}$$

$$P(B) = \frac{1}{6} = \frac{6}{36} \tag{94}$$

$$P(C) = P(\{(1,4),(2,4),(3,4),(4,4),(4,1),(4,2),(4,3)\}) = \frac{7}{36} \tag{95}$$

$$P(D) = P(\{(1,4),(2,3),(3,2),(4,1) = \frac{4}{36} \tag{96}$$

$$P(F) = \{(1,6),(6,1)\} = \frac{1}{12} = \frac{3}{36} \tag{97}$$

Elementer i hver fællesmængde:

A	B	C	D	F	set
6	1	4	1	1	A
1	6	1	1	1	B
4	1	7	2	2	C
1	1	2	4	0	D
1	1	2	0	6	F

Figure 1: opg. 1.24 - elementer i hver fællesmængde

A	B	C	D	F	set
0.167	0.028	0.111	0.028	0.028	A
0.028	0.167	0.028	0.028	0.028	B
0.111	0.028	0.194	0.056	0.056	C
0.028	0.028	0.056	0.111	0.000	D
0.028	0.028	0.056	0.000	0.167	F

Figure 2: opg 1.24 - Sandsynlighed for fællesmængde

A	B	C	D	F	set
0.028	0.028	0.032	0.019	0.028	A
0.028	0.028	0.032	0.019	0.028	B
0.032	0.032	0.038	0.022	0.032	C
0.019	0.019	0.022	0.012	0.019	D
0.028	0.028	0.032	0.019	0.028	F

Figure 3: opg. 1.24 - Sandsynligheden for $P(A) \cdot P(B)$

De uafhængige par er: $(A, B), (A, F), (B, F)$

3.1.5 Opgave 1.13

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B \cup C) - P(A \cap (B \cup C) && (98) \\ &= P(A) + P(B) + P(C) - P(B \cup C) - P(A \cap (B \cup C) && (99) \end{aligned}$$

Hvis vi ser nærmere på den sidste del **Lav tegning af mængder! A, B, C har en intersektion**.

$$P(A \cap (B \cup C) = P(A \cup B) + P(A \cup C) - P(A \cup B \cup C) \tag{100}$$

Man husker at der er minus foran denne mængde, sådan at:

$$\begin{aligned} &P(A \cup B \cup C) = && (101) \\ &\quad P(A) + P(B) + P(C) - P(B \cup C) && (102) \\ &\quad - (P(A \cup B) + P(A \cup C) - P(A \cup B \cup C)) && (103) \end{aligned}$$

3.2 Øvelse 2

15/09/2018, opgaver: 1.6, 1.7, 1.9, 1.15, 1.18, 1.28 og 1.30 (og 1.12 hvis der er tid)

3.2.1 1.6

- 1 terning
- 2 slag

Ssh for mindst 1 sekser

$$P(\{\text{mindst en sekser}\}) = \tag{104}$$
$$P(\{(1,6),(2,6),\cdots,(6,6),(6,1),\cdots,(6,5) = \tag{105}$$
$$\frac{5+6}{36} = \frac{11}{36} \tag{106}$$

Ssh. for mindst 1 sekser eller mindst 1 toer

$$P(\{\text{mindst en sekser}\}) = \tag{107}$$
$$P(\{(1,6),(2,6),\cdots,(6,6),(6,1),\cdots,(5,6), \tag{108}$$
$$(1,2),\cdots(5,2),(2,1)\cdots(2,5)\}) = \tag{109}$$
$$\frac{6+5+5+4}{36} = \frac{20}{36} \tag{110}$$

3.2.2 Opgave 1.7

- 1 mønt
- 10 kast

Hvad er ssh. for mindst 2 plat

Find sandsynligheden for komplimenter hændelsen:

A : Er hændelsen for at få mindst 2 plat.

A^C : Er Komplementær hændelsen - altså maks 1 plat:

$$A^C = \{\text{slå 0 plat}\} \cup \{\text{slå 1 plat}\} \tag{111}$$

$$P(\{\text{slå 0 plat}\}) = \frac{1}{2^{10}} \tag{112}$$

$$P(\{\text{slå 1 plat}\}) = \frac{10}{2^{10}} \tag{113}$$

Noter at $\{\text{slå 0 plat}\} \cap \{\text{slå 1 plat}\} = \emptyset$

$$P(A^C) = \frac{1}{2^{10}} + \frac{10}{2^{10}} = \frac{11}{2^{10}} \tag{114}$$

$$P(A) = 1 - P(A^C) = 1 - \frac{11}{2^{10}} = \frac{1013}{2^{10}} \tag{115}$$

3.2.3 Opgave 1.9

- 1 spil kort (52 kort)
- 13 kort trækkes

Hvad er Ssh. for 0 billedkort eller esser

Antal billedkort og esser (kaldet billedkort fra nu): $4 * 4 = 16$

Kig på komplementær hændelsen:

$$P(\{\text{kort 1 ikke billedkort}\}) = \frac{52-16}{52} \tag{116}$$

Vi har trukket 1 kort nu $\implies$ 51 kort tilbage, men stadig 12 billedkort

$$P(\{\text{kort 2 er billedkort}\}) = \frac{51-16}{51} \tag{117}$$

$$P(\{\text{man trækker 0 billedkort}\}) = \prod_{i=0}^{12} \frac{52-i-16}{52-i} = 0.0036 \tag{118}$$

Alternativt

$$\#E = 52 \cdot 51 \cdots 40 = \frac{52!}{39!} \tag{119}$$

$$\#A = 36 \cdot 35 \cdots 24 = \frac{36!}{23!} \tag{120}$$

$$P(\{\text{man trækker 0 billedkort}\}) = \frac{\#A}{\#E} = 0.0036 \tag{121}$$

3.2.4 Opgave 1.15

- 4 slag med terning
- mindst 1 sekser
- demere mente $4 \times \frac{1}{6}$

Hvorfor tog han fejl?

Klasse diskussion:

Kig på komplementærhændelsen: *Ingen seksere*

$$P(\{\textbf{Ingen seksere}\}) = (\frac{5}{6})^4 = \frac{5^4}{6^4} = 0.49 \tag{122}$$

Da dette er komplementær hændelsen kan vi i stedet sige:

$$P(\{\textbf{mindst 1 sekser}\} = 1 - 0.49 = 0.51 \tag{123}$$

Ssh for en dobbelt sekser i 24 kast

- 24 kast
- mindst 1 dobbelt sekser

Sandsynligheden for 1 dobbelt sekser i et slag.

$$P(\{\textbf{En dobbelt sekser}\}) = \frac{1}{6}\frac{1}{6} = \frac{1}{36} \tag{124}$$

Brug komplementær hændelsen: Dvs. ssh for ikke at få en dobbelt sekser i 24 slag:

$$P(\{\textbf{Ingen dobbelt sekser i 24 slag}\}) = (\frac{35}{36})^{24} = 0.509 \tag{125}$$

$$P(\{\textbf{mindst en dobbelt sekser i 24 slag}\}) = 1 - 0.509 = 0.491 \tag{126}$$

Så ikke langt fra!

3.2.5 Opgave 1.18

- 1 mønt
- 10 kast

Hvad er ssh. for at få krone den 10'ende gang givet 9 plat

Lad os definerer hændelserne:

A : Man har fået 9 plat på de første 9 slag af de 10 slag

B : Man får krone på det sidste slag ud af de 10 slag

Brug definition for betingede ssh (1.4.1):

$$P(A \mid B) = \frac{P(A \cap B)}{P(A)} \tag{127}$$

$$P(A \cap B) = (\frac{1}{2})^{10} = \frac{1}{2^{10}} \tag{128}$$

$$P(A) = (\frac{1}{2})^{9} = \frac{1}{2^{9}} \tag{129}$$

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{2^{10}}}{\frac{1}{2^{9}}} = \frac{1}{2} \tag{130}$$

SSh for den 10 bliver krone, givet 9 af de 10 kast blive plat

Lad os definerer hændelserne:

A : Man har fået 9 plat ud af de 10 slag

B : Man får krone på det sidste slag ud af de 10 slag

$$P(A \cap B) = (\frac{1}{2})^{10} = \frac{1}{2^{10}} \tag{131}$$

$$P(A) = 10 \times (\frac{1}{2})^{10} = \frac{10}{2^{10}} \tag{132}$$

$$P(B \mid A) = \frac{1}{10} \tag{133}$$

3.2.6 Opgave 1.28

- 1 terning
- 1 kast
- Hændelse A : kast er 1,2,3
- Hændelse B : kast er 1 eller 4

Vis at A og B er uafhængige

Brug Definition 1.5.1:

$$P(A \cap B) = P(A)\dot{P}(B) \tag{134}$$

$$P(A) = \frac{1}{2} \tag{135}$$

$$P(B) = \frac{1}{3} \tag{136}$$

Hvad er fælles mængden af de to hændelser: *at terningen bliver 1*

$$P(A \cap B) = P(\{\textbf{Terningen bliver 1}\}) = \frac{1}{6} = P(A)\dot{P}(B) \tag{137}$$

Og vi har herved vist, at hændelserne er uafhængige!

3.2.7 Opgave 1.30

Lad eleverne prøve!

- 3 hændelser: A, B, C
- $A \perp\!\!\!\perp B$
- $A \perp\!\!\!\perp C$

Kan man fra ovenstående slutte at: $A \perp\!\!\!\perp B \cup C$

$$A \perp\!\!\!\perp B \implies P(A) \cdot P(B) = P(A \cap B) \tag{138}$$

$$A \perp\!\!\!\perp C \implies P(A) \cdot P(C) = P(A \cap C) \tag{139}$$

Bevis via. modeksempel

$A = \{$**Spar eller hjerter**$\}$

$B = \{$**Spar eller ruder**$\}$

$C = \{$**hjerter eller ruder**$\}$

$P(A \cap B) = \frac{1}{4} = P(A)P(B)$

$P(A \cap C) = \frac{1}{4} = P(A)P(C)$

$P(A \cap (B \cup C)) = \frac{1}{2} \neq P(A)P(B \cup C) = \frac{1}{2}\frac{3}{4}$

3.2.8 Opgave 1.12

- 1 slag
- 5 terninger

Sandsynligheden for at få mindst 1 sekser

Udregn ssh for komplementærhændelsen at få 0 seksere!

Definér hændelsen A : At få mindst 1 sekser

$$P(A^C) = P(\{\textbf{0 seksere}\} = \left(\frac{5}{6}\right)^5 = 0.402 \tag{140}$$

$$1 - A^C = 0.598 \tag{141}$$

3.3 Øvelse 3

17/9/2017, Opgaver: 2.1 og 2.3 fra Sørensen (2015) samt opgaverne B.1, B.2 og B.3

3.3.1 Opgave 2.1

- 1 rød terning
- 1 sort terning
- $Y := \min(r, s)$
- $Z := \max(r, s)$

Fordelingen for Y

TEGN TERNINGEMATRICEN

$$P(Y = 1) = P(\{(1,1),(1,2),\cdots,(1,6),(2,1),\cdots,(6,1)\}) = \frac{11}{36} \tag{142}$$

$$P(Y = 2) = P(\{(2,2),(2,3),\cdots(2,6),(3,2)\cdots(6,2)\}) = \frac{9}{36} \tag{143}$$

$$P(Y = 3) = \cdots = \frac{7}{36} \tag{144}$$

Den resterende fordeling for Y er: $P(Y = 4) = \frac{5}{36}, P(Y = 5) = \frac{3}{36}, P(Y = 6) = \frac{1}{36}$.

Fordelingen for Z

TEGN TERNINGEMATRICEN

$$P(Z = 1) = P(\{(1,1)\}) = \frac{1}{36} \tag{145}$$

$$P(Z = 2) = P(\{(2,1),(2,2),(2,1) = \frac{3}{36} \tag{146}$$

$$P(Z = 3) = \cdots = \frac{5}{36} \tag{147}$$

Den resterende fordeling for Z er $P(Z=4)=\frac{7}{36}, P(Z=5)=\frac{9}{36}, P(Z=6)=\frac{11}{36}$.

Den simultane fordeling er 3.3.1:

Y er vandret, Z lodret: Vi ved at det må være en øvre trekantsmatrice.

Til diagonalen: Vi ved at der er kun måde at min og maks kan være ens $min(T_1,T_2)=max(T_1,T_2) \implies T_1=T_2$.

Til den øvre trekant: $Y=1, Z_2 \implies T_1=1, T_2=2 \vee T_1=2, T_2=1$. Dette kan gøres for alle elementer af den øvre trekant

Table 1: Simultan fordeling

	$Y=1$	$Y=2$	$Y=3$	$Y=4$	$Y=5$	$Y=6$
$Z=1$	1/36	2/36	2/36	2/36	2/36	2/36
$Z=2$	0	1/36	2/36	2/36	2/36	2/36
$Z=3$	0	0	1/36	2/36	2/36	2/36
$Z=4$	0	0	0	1/36	2/36	2/36
$Z=5$	0	0	0	0	1/36	2/36
$Z=6$	0	0	0	0	0	1/36

3.3.2 Opgave 2.3

- Stokastisk variabel er beskrevet i bogen

$Y=t(X)$

$$P(Y=1)=P(X\in\{1,2,3\})=0.12+0.8+0.20=0.4 \tag{148}$$
$$P(Y=2)=P(X\in\{4,5\})=0.11+0.19=0.30 \tag{149}$$
$$P(Y=3)=P(X\in\{6,7\})=0.14+0.06=0.20 \tag{150}$$
$$P(Y=4)=P(X\in\{8\})=0.10 \tag{151}$$

Fordelingsfunktion (CDF):

$$P(Y \leq 0) = 0 \tag{152}$$
$$P(Y \leq 1) = 0.4 \tag{153}$$
$$P(Y \leq 2) = 0.7 \tag{154}$$
$$P(Y \leq 3) = 0.9 \tag{155}$$
$$P(Y \leq 4) = 1.0 \tag{156}$$

3.3.3 Opgave B.1

- stokastiske variable X_1, X_2
- $X_1 = 1$ hvis der var en stor nyhed (ellers 0)
- $X_2 = 1$ hvis aktiemarkedet steg/faldt (0 hvis ikke)
- $P(X_1 = 1) = \frac{6}{10}$
- $P(X_2 = 1) = \frac{3}{10}$

Simultane fordeling under antagelse af uafhængighed!

Brug definition 2.4.1 (sørensen)

$$P(X_1 = 0, X_2 = 0) = P(X_1 = 0)P(X_2 = 0) = \frac{4}{10}\frac{7}{10} = \frac{28}{100} \tag{157}$$
$$P(X_1 = 0, X_2 = 1) = P(X_1 = 0)P(X_2 = 1) = \frac{4}{10}\frac{3}{10} = \frac{12}{100} \tag{158}$$
$$P(X_1 = 1, X_2 = 0) = P(X_1 = 1)P(X_2 = 0) = \frac{6}{10}\frac{7}{10} = \frac{42}{10} \tag{159}$$
$$P(X_1 = 1, X_2 = 1) = P(X_1 = 1)P(X_2 = 1) = \frac{6}{10}\frac{3}{10} = \frac{18}{10} \tag{160}$$

TEGN BI-MATRICE

DEL 2: Antag IKKE uafhængighed - Hvad er den simultane fordeling (X_1, X_2)

tegn bimatrice og fyld værdier i løbende!

$$P(X_2 = 1 \mid X_1 = 1) = \frac{4}{10} \tag{161}$$

Udvid **Definition 1.4.3**

$$P(B) = \sum_{j=1}^{n} P(B \mid A_j)P(A_j) = \sum_{j=1}^{n} P(B, A_j) \tag{162}$$

$$P(X_2 = 1) = P(X_2 = 1 \mid X_1 = 0)P(X_1 = 0) + P(X_2 = 1 \mid X_1 = 1)P(X_1 = 1) \tag{163}$$

$$P(X_2 = 1) = P(X_2 = 1, X_1 = 0) + P(X_2 = 1, X_1 = 0) \tag{164}$$

Vi husker at $P(X_2) = \frac{3}{10}$

$$\frac{3}{10} = \underbrace{\frac{4}{10}\frac{6}{10}}_{P(X_1=1,X_2=1)} + P(X_2 = 1 \mid X_1 = 0)P(X_1 = 0) \tag{165}$$

$$\implies P(X_1 = 0, X_2 = 1) = \frac{6}{100} \tag{166}$$

Vi har allerede set at:

$$P(X_1 = 1, X_2 = 1) = \frac{24}{100} \tag{167}$$

Vi går videre:

$$P(X_1 = 1) = P(X_1 = 1, X_2 = 0) + P(X_1 = 1, X_2 = 1) \tag{168}$$

Vi indsætter de værdier vi kender:

$$\frac{6}{10} = P(X_1 = 1, X_2 = 0) + \frac{24}{100} \implies P(X_1 = 1, X_2 = 0) = \frac{36}{100} \tag{169}$$

Vi mangler kun sidste værdi nu:

$$P(X_2 = 0) = P(X_2 = 0, X_1 = 0) + P(X_2 = 0, X_1 = 1) \tag{170}$$

Husker værdier: $P(X_2 = 0) = \frac{7}{10}$ og $P(X_1 = 1, X_2 = 0) = \frac{36}{100}$

$$\frac{7}{10} = \frac{36}{100} + P(X_1 = 0, X_2 = 0) \implies P(X_1 = 0, X_2 = 0) = \frac{34}{100} \tag{171}$$

Ændrer fordelingen sig for X_1, X_2, X

Spørg klassen

De marginale distributioner er ens, De betingede og den simultane er forskellig

3.3.4 Opgave B.2

- Test for cancer
- Den gætter rigtig med 95 % ssh.
- 1 ud af 100.000 mennesker har denne kræft form

Lad X for cancer testen $X = 1$ implicerer positiv test . Lad Y være en stokastisk variabel som angiver om man har kræft $Y = 1$ betyder man har kræft.

Vi kan skitserer nogle sandsynligheder:

$$P(X = 1 \mid Y = 1) = 0.95, \quad P(X = 0 \mid Y = 1) = 0.05 \tag{172}$$

$$P(X = 0 \mid Y = 0) = 0.95, \quad P(X = 1 \mid Y = 0) = 0.05 \tag{173}$$

$$P(Y = 1) = \frac{1}{100000} = 0.00001 \tag{174}$$

Brug bayes formel (sætning 1.4.7):

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{\sum_{j=1}^{n} P(B \mid A_j)P(A_j)} \tag{175}$$

$$
\begin{aligned}
&P(Y = 1 \mid X = 1) && (176)\\
&= \frac{P(X = 1 \mid Y = 1)P(Y = 1)}{P(X = 1 \mid Y = 1)P(Y = 1) + P(X = 1 \mid Y = 0)P(Y = 0)} && (177)
\end{aligned}
$$

$$
P(Y = 1 \mid X = 1) = \frac{0.95 \cdot 0.00001}{0.95 \cdot 0.00001 + 0.05 \cdot 0.99999} = 0.0001899 \tag{178}
$$

3.3.5 Opgave B.3

Kig github!

3.4 Øvelse 4

21/9/2018, Øvelser: B.4 og 2.4, 2.5, og 2.9 fra Sørensen (2015)

3.4.1 Opgave B.4

Lav i klassen

- 1 mønt
- 1 terning
- X er stokastisk variabel med summen af antal øjne på terning + (0/1) (1 hvis krone).

$$
T := \text{Ternings øjne}, \qquad M := \text{Mønt} \tag{179}
$$

$$
X := T + M \tag{180}
$$

Del 4 - Find $P(X > 3)$

Definer hændelser:

$A = \{X > 3\}$

$A^C = \{X \leq 3\}$

udfaldsrummet for den simultane fordeling af T of M $\{0,1\}\times\{1,2,3,4,5,6\}$

$$P(A^C) = \tag{181}$$
$$P(\{(M=0,T=1),(M=0,T=2),(M=0,T=3), \tag{182}$$
$$(M=1,T=1),(M=1,T=2)\}) \tag{183}$$

Dette var kun komplementær hændelsen

$$P(A^C) = \frac{5}{12} \tag{184}$$

$$P(A) = \frac{7}{12} \tag{185}$$

Del 5 - SSh for ulige nummer

Definér hændelsen.

$A = \{X \in \textbf{Ulige numre}\}$

Disse er alle indbyrdes disjunkte hændelser $A = \{X = 1\} \cup \{X = 3\} \cup \{X = 5\} \cup \{X = 7\}$

$$P(A) = P(\{X=1\}) + P(\{X=3\}) + P(\{X=5\}) + P(\{X=7\}) \tag{186}$$

$$P(A) = \frac{1}{12} + \frac{2}{12} + \frac{2}{12} + \frac{1}{12} = \frac{1}{2} \tag{187}$$

3.4.2 Opgave 2.4

L

- X er en stokastisk variabel som kan antage værdierne $\{1,2,3\}$
- $P(X=1) = P(X=2) = P(X=3) = \frac{1}{3}$
- En stokastisk variabel $Y = 1/X$

Tegn fordelingsfunktionen for X og Y

Kig Github!

3.4.3 Opgave 2.5

Lav første del i klassen

- X_1, X_2 er stokastiske variable.
- begge har udfaldsrummet $\{0, 1\}$
- X_1 marginale fordeling:
  - $P(X_1 = 0) = 0.4$
  - $P(X_1 = 1) = 0.6$
- X_2 marginale fordeling
  - $P(X_2 = 0) = 0.3$
  - $P(X_2 = 1) = 0.7$
- Vi har en stokastisk vektor $X = (X_1, X_2)$

Del 1) Undersøg uafhængighed når den simultane fordeling af X er:

Table 2: Simultan fordeling af X

	$X_1 = 0$	$X_1 = 1$
$X_2 = 0$	0.12	0.18
$X_2 = 1$	0.28	0.42

Se **definition 2.4.1**: Skriv den op på tavlen!

Vi tester for uafhængighed:

$$P(X_1 = 0)P(X_2 = 0) = 0.4 \cdot 0.3 = 0.12 \quad (188)$$
$$P(X_1 = 0)P(X_2 = 1) = 0.4 \cdot 0.7 = 0.28 \quad (189)$$
$$P(X_1 = 1)P(X_2 = 0) = 0.6 \cdot 0.3 = 0.18 \quad (190)$$
$$P(X_1 = 1)P(X_2 = 1) = 0.6 \cdot 0.7 = 0.42 \quad (191)$$

Vi ser at X_1 er uafhængig af X_2.

Del 2) Undersøg uafhængighed når den simultane fordeling af X er:

Til klassen: Er *dette overhovedet muligt - givet ovenstående resultat?*

Table 3: Simultan fordeling af X

	$X_1 = 0$	$X_1 = 1$
$X_2 = 0$	0.15	0.15
$X_2 = 1$	0.25	0.45

Del 3) gør rede for at begge simulatane fordelinger er i overensstemmelse med de angivne marginale fordelinger

$$P(X_1 = 0) = P((0,0)) + P((0,1)) = 0.4 \tag{192}$$
$$P(X_1 = 1) = P((1,0)) + P((1,1)) = 0.6 \tag{193}$$
$$P(X_2 = 0) = P((0,0)) + P((1,0)) = 0.3 \tag{194}$$
$$P(X_2 = 1) = P((0,1)) + P((1,1)) = 0.7 \tag{195}$$

3.4.4 Opgave 2.9

Note brug min() og maks() som funktioner istedet for bogens notation.

- 2 terninger, T_1, T_2
- T_1, T_2 er ligefordelt på $\{1, 2, 3, 4, 5, 6\}$
- $Y = min(T_1, T_2)$
- $Z = max(T_1, T_2)$

Hvad er den simultane fordeling?

Y er vandret, Z lodret: Vi ved at det må være en øvre trekantsmatrice.

Til diagonalen: Vi ved at der er kun måde at min og maks kan være ens $min(T_1, T_2) = max(T_1, T_2) \implies T_1 = T_2$.

Til den øvre trekant: $Y = 1, Z_2 \implies T_1 = 1, T_2 = 2 \vee T_1 = 2, T_2 = 1$. Dette kan gøres for alle elementer af den øvre trekant

Er Y, Z uafhængige

Husk:

$$P(Y = A, Z = B) = P(Y = A)P(Z = B) \quad \forall A, B \in \{1, 2, 3, 4, 5, 6\} \tag{196}$$

Table 4: Simultan fordeling

	$Y=1$	$Y=2$	$Y=3$	$Y=4$	$Y=5$	$Y=6$
$Z=1$	1/36	2/36	2/36	2/36	2/36	2/36
$Z=2$	0	1/36	2/36	2/36	2/36	2/36
$Z=3$	0	0	1/36	2/36	2/36	2/36
$Z=4$	0	0	0	1/36	2/36	2/36
$Z=5$	0	0	0	0	1/36	2/36
$Z=6$	0	0	0	0	0	1/36

Vi skal bare have et modeksempel. Eftersom: $P(Y=1, Z=2) = 0$ kan vi konkluderer ikke uafhægighed. *Overvej dette !*

3.5 Øvelse 5

24/09/2018 - C.1, C.2, C.3 & 3.20, 3.24, 3.27 (optional 3.2) sørensen

3.5.1 Opgave C.1

- Basketball player
- 10 skud
- ssh for at ramme 0.5

Binomial fordeling

Hvad er SSh for at ramme 8 skud med ssh 0.5

$$p(x) = \binom{10}{8} 0.5^8 (1-0.5)^{10-8} = 0.04394 \tag{197}$$

Hvad er SSh for at ramme med ssh 0.6

$$p(x) = \binom{10}{8} 0.6^8 (1-0.6)^{10-8} = 0.1209 \tag{198}$$

Ssh på 0.5 - hvad er varians of middelværdi

$$\mathrm{E}(X) = n \cdot p = 0.5 \cdot 10 = 5 \tag{199}$$

fra wikipedia

$$\mathrm{Var}(X) = n \cdot p \cdot (1-p) = 2.5 \tag{200}$$

3.5.2 Opgave C.2

- X er stokastisk variabel
- diskret pdf $f(x) = \frac{x}{8}$
- $x \in \{1, 2, 5\}$

Hvad er E(X)

$$\mathrm{E}(X) = \sum_{i=1}^{n} p_i \cdot x_i = 1 \cdot \frac{1}{8} + 2 \cdot \frac{2}{8} + 5 \cdot \frac{5}{8} = \frac{1+4+25}{8} = 3.75 \tag{201}$$

Hvad er Var(X)

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 \tag{202}$$

$$E(X^2) = 1^2 \cdot \frac{1}{8} + 2^2 \cdot \frac{2}{8} + 5^2 \cdot \frac{5}{8} = \frac{1+8+125}{8} = 16.75 \tag{203}$$

$$\mathrm{Var}(X) = 16.75 - 3.75^2 = 16.75 - 14.0625 = 2.6875 \tag{204}$$

Hvad er $E(2X+3)$

Vi bruger:

$$\mathrm{E}(a + bX) = a + b\mathrm{E}(X) \tag{205}$$

Husk $\mathrm{E}(X) = 3.75$

$$2 \cdot 3.75 + 3 = 7.5 + 3 = 10.5 \tag{206}$$

3.5.3 Opgave C.3

- Efterspørgsel for software er X
- købspris 10
- salgspris 35
- Ved årets ende er softwaren intet værd
- køber 4 kopier af software

Find $\mathbf{E}(X)$

$$\mathrm{E}(X) = 0.1{\cdot}0+0.3{\cdot}1+0.3{\cdot}2+0.2{\cdot}3+0.1{\cdot}4 = 0.3+0.6+0.6+0.4 = 1.9 \tag{207}$$

Find $\mathbf{Var}(X)$

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 \tag{208}$$

$$\mathrm{E}(X^2) = 0.1{\cdot}0+0.3{\cdot}1+0.3{\cdot}4+0.2{\cdot}9+0.1{\cdot}16 = 0.3+1.2+1.8+1.6 = 4.9 \tag{209}$$

$$\mathrm{Var}(X) = 4.9 - 1.9^2 = 4.9 - 3.61 = 1.29 \tag{210}$$

Efterspørgselsfunktion Y, samt $\mathbf{E}(Y)$ og $\mathbf{Var}(Y)$

man køber 4 stykker software 4×10. og sælger x af dem som er en realisation af X.

$$Y := 35X - 40 \tag{211}$$

husk

$$\mathrm{E}(a + bX) = a + b\mathrm{E}(X) \tag{212}$$

$$\mathrm{E}(Y) = \mathrm{E}(35X - 40) = 35 \cdot \mathrm{E}(X) - 40 = 3.5 \cdot 1.9 - 40 = 26.5 \tag{213}$$

Normalt ville vi sige:

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 \tag{214}$$

Vi gør noget smartere her (kig bog s. 93):

$$\mathrm{Var}(aX + b) = b^2 \mathrm{Var}(X) \tag{215}$$

$$\mathrm{Var}(Y) = \mathrm{Var}(35X - 40) = 35^2 \cdot \mathrm{Var}(X) = 35^2 \cdot 1.29 = 1580.25 \tag{216}$$

3.5.4 Opgave 3.20

- en stokastisk variabel X er ligefordelt på $\{1, 2, 3, 4, 5, 6\}$ (en terning)
- stokastisk variabel $Y := R + H$, hvor er og R, H er terninger
- Z er stokastisk variabel som er for uniform på $\{1, 2, 3 \cdots, n$.

Find middelværdi og varians for X

Man siger at $X := unif\{a, b\} = unif\{1, 6\}$

Middelværdi

$$\mathrm{E}(X) = \sum_{i=1}^{6} \frac{1}{6} i = 3.5 \tag{217}$$

Fra wikipedia om diskrete uniform fordeling

`https://en.wikipedia.org/wiki/Discrete_uniform_distribution`
`Varians`

Generelt er der gode informationer om distributioner på wiki!

$$\mathrm{Var}(X) = \frac{(b - a + 1)^2 - 1}{12} \tag{218}$$

$$\mathrm{Var}(X) = \frac{(6 - 1 + 1)^2 - 1}{12} = \frac{35}{12} = 2.92 \tag{219}$$

For Y

$R, H := unif\{1, 6\}$. $Y = R + H$

Vi ved at $R \perp\!\!\!\perp H$

brug Sætning 3.7.7 (s. 91) - (uafhængighed er ikke nødvendig)

$$\mathrm{E}(Y) = \mathrm{E}(R + H) = \mathrm{E}(R) + \mathrm{E}(H) = 3.5 + 3.5 = 7 \tag{220}$$

Grundet uafhængighed kan vi nu bruge sætning 3.8.8 (s. 101)

$$\mathrm{Var}(X_1 + X_2 + \cdots X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) \cdots \mathrm{Var}(X_n) \tag{221}$$

$$\mathrm{Var}(Y) = \mathrm{Var}(R) + \mathrm{Var}(H) = 2.92 + 2.92 = 5.84 \tag{222}$$

Middelværdi og varians for Z

Vi kan definere den stokastiske variabel $Z := unif(1, n)$

$$\mathrm{E}(Z) = \sum_{i=1}^{n} \frac{1}{n} i = \frac{1}{n} \sum_{i=1}^{n} i \tag{223}$$

summen er $\frac{n(n+1)}{2}$. Vis gaus beviset: vi har $n/2$ gange (1 + n). 1 + 50 = 51, 2 + 49 = 51 osv det kan vi gøre 25 gange.

$$\mathrm{E}(Z) = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2} \tag{224}$$

Nu skal variansen udregnes!

$$\mathrm{Var}(Z) = \mathrm{E}(Z^2) - (E(Z))^2 \tag{225}$$

I bogen har vi opgivet at:

$$\sum_{i=1}^{n} i^2 = \frac{1}{6} n(2n+1)(n+1) \tag{226}$$

Vi ved derfor at:

$$E(Z^2) = \sum_{i=1}^{n} \frac{1}{n} i^2 = \frac{1}{n} \sum_{i=1}^{n} i^2 = \frac{1}{n} \frac{1}{6} n(2n+1)(n+1) = \frac{1}{6}(2n+1)(n+1) \tag{227}$$

(Andel af udtrykket er $E(Z)^2$)

$$\mathrm{Var}(Z) = \frac{1}{6}(2n+1)(n+1) - \frac{n+1}{2}\frac{n+1}{2} \tag{228}$$

Vi ser udtrykket kan forkortes:

$$\mathrm{Var}(Z) = \left(\frac{1}{6}(2n+1) - \frac{n+1}{2^2}\right)(n+1) \tag{229}$$

3.5.5 Opgave 3.24

- en stokastisk variabel X
- $\mathrm{E}(X) = 5$
- $\mathrm{Var}(X) = 2$

Find E$(7 + 8X + X^2)$

$$\mathrm{E}(7 + 8X + X^2) = \mathrm{E}(7) + \mathrm{E}(8X) + \mathrm{E}(X^2) \tag{230}$$

Først ved vi at $\mathrm{E}(7) = 7$.

Dernæst

$$\mathrm{E}(8X) = 8 \cdot \mathrm{E}(X) = 8 \cdot 5 = 40 \tag{231}$$

Til sidst

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 \tag{232}$$

Vi kender variansen og $\mathrm{E}(X)$:

$$2 = \mathrm{E}(X^2) - 5^2 \implies \mathrm{E}(X^2) = 2 + 5^2 = 27 \tag{233}$$

$$\mathrm{E}(7 + 8X + X^2) = 7 + 40 + 27 = 74 \tag{234}$$

3.5.6 Opgave 3.27

- 3 stokastiske variable
- X_1, X_2, X_3
- identiske og uafhængige

Vis at

$$\text{Corr}(X_1 + X_2, X_2 + X_3) = \frac{1}{2} \tag{235}$$

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \tag{236}$$

$$\text{Cov}(X, Y) = (X - \text{E}(X))(Y - \text{E}(Y)) \tag{237}$$

Indsæt vores stokastiske variable $X_1 + X_2$ og $X_2 + X_3$.

$$\text{Cov}(X_1 + X_2, X_2 + X_3) = \tag{238}$$
$$(X_1 + X_2 - \text{E}(X_1) + \text{E}(X_2))(X_2 + X_3 - \text{E}(X_2) + \text{E}(X_3)) = \tag{239}$$
$$([X_1 - \text{E}(X_1)] + [X_2 - \text{E}(X_2)])([X_2 - \text{E}(X_2)] + [X_3 - \text{E}(X_3)]) = \tag{240}$$
$$\text{Cov}(X_1, X_2) + \text{Cov}(X_1, X_3) + \text{Cov}(X_2, X_3) + \text{Var}(X_2) \tag{241}$$

Vi ved at uafhængighed implicerer ar covariancen er lig 0. Det betyder:

$$\text{Cov}(X_1 + X_2, X_2 + X_3) = \text{Var}(X_2) = \sigma^2 \tag{242}$$

Brug **sætning 3.8.8** (s. 101). Man kan splitte variansen op af ukorrelerede stokastiske variabler til en sum

$$\text{Var}(X_1 + X_2)\text{Var}(X_2 + X_3) = \tag{243}$$
$$(\text{Var}(X_1) + \text{Var}(X_2))(\text{Var}(X_2) + \text{Var}(X_3)) = \tag{244}$$
$$\tag{245}$$

Vi ved variansen er ens for alle stokastiske variable sådan at: $Var(X_1) = \text{Var}(X_2) = \text{Var}(X_3) = \sigma^2$

$$2\sigma^2 \cdot 2\sigma^2 \tag{246}$$

$$\sqrt{2\sigma^2 \cdot 2\sigma^2} = 2\sigma^2 \tag{247}$$

Vi har herved fundet det ønskede resultat!

$$\text{Corr}(X_1 + X_2, X_2 + X_3) = \frac{\sigma^2}{2\sigma^2} = \frac{1}{2} \tag{248}$$

3.5.7 Opgave 3.2

- 5 Cola-smagere
- 2 Cola-mærker $\{C, P\}$
- med sandsynlighed p gætter de rigtigt
- 4 ud af 5 gætte på cola P. 1 gættede C

Hvad er den betingede ssh for at det var cola C der blev serveret

Definér to stokastiske variable:

$S :=$ {**Hvilke cola der blev serveret**}

$C :=$ {**hvilken cola der blev serveret**}

$P(D) = \frac{1}{2}$

$P(S \mid D) \sim Bin(5, p)$

$$P(S = 4 \mid D = C) \binom{5}{4} p(1-p)^4 \tag{249}$$

3.6 Øvelse 6

28/09/2018 - C.4 & Opgave 1

3.6.1 Opgave C.4

- Poisson distribution
- Antal opkald kan modelleres med en stokastisk variabel kaldet $X := Poisson(\lambda)$.

Om Poisson fordelingen: En ventetidsfordeling! Citat wikipedia:

> "[Poisson fordelingen] is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant rate and independently of the time since the last event." - Wikipedia

Den har den egenskab at: $\mathrm{E}(X) = \mathrm{Var}(X) = \lambda$

Om denne fordeling kan vi sige at sandsynligheden for et givent udfald er (pdf):

$$p(x) = \frac{\lambda^x}{x!}e^{-\lambda} \tag{250}$$

Lad dem regne selv

Ssh for præcis 7

$$p(7) = \frac{7^{10}}{7!}e^{-10} = 0.090079 \tag{251}$$

Ssh for max 7 opkald

$$P(X \leq 7) = \sum_{i=0}^{7} \frac{i^{10}}{i!}e^{-10} = 0.22022 \tag{252}$$

$$P(3 \leq X \leq 7) = \sum_{i=3}^{7} \frac{i^{10}}{i!}e^{-10} = \sum_{i=0}^{7} \frac{i^{10}}{i!}e^{-10} - \sum_{i=0}^{2} \frac{i^{10}}{i!}e^{-10} \tag{253}$$

Indsæt værdier udregnet i python

$$0.22022 - 0.002769 = 0.217451 \tag{254}$$

3.6.2 Opgave 1

- Værdi af cykel 4000 kr
- ssh for den bliver stjålet 5 %
- man kan tegne en cykel så den bliver erstattet for hele dens værdi

Del 1) Hvor meget er man villig til at betale for en sådan forsikring?

Spørg klassen - Intet rigtigt svar?

del 2) Udregn værdi af cykel (på et år)

Vi definerer X stokastiske variable:

$X :=$ **Cykel værdi**

$P(X = 0) = 0.05$ og $P(X = 4000) = 0.95$

Så ganger vi værdien på X bagefter.

$$\mathrm{E}(X) = 0.95 \cdot 4000 = 3800 \tag{255}$$

del 3) Cykel forsikring!

$$Y := \textbf{Værdi af cykel minus forsikring 1} \tag{256}$$

$$(Y \mid X = 0) = 0 - 400 + 4000 = 3600 \tag{257}$$
$$(Y \mid X = 4000) = 4000 - 400 = 3600 \tag{258}$$

$$\mathrm{E}(Y) = 0.95 \cdot 3600 + 0.05 \cdot 3600 = 3600 \tag{259}$$

Del 4) Forsikring med selvrisiko på 1000 kr!

pris = 150 årligt, selvrisiko = 1000.

$Z :=$ **Værdi af cykel minus forsikring 2**

$$(Z \mid X = 0) = 0 - 150 - 1000 + 4000 = 2850 \tag{260}$$
$$(Z \mid X = 4000) = 4000 - 150 = 3850 \tag{261}$$

$$\mathrm{E}(Z) = 0.05 \cdot 2850 + \cdot 0.95 \cdot 3850 = 3800 \tag{262}$$

Del 5) Sammenlign middel værdier

Klassediskussion

Del 6) Nytte af af X, Y, Z

nyttefunktion:

$$u(v) = 10v - 0.001v^2, \qquad v \in \{0, 1, \cdots 4000\} \tag{263}$$

Transformér de enkelte stokastiske variable først! X:

$$u(X \mid X = 0) = 0 \tag{264}$$
$$u(X \mid X = 1) = 10 \cdot 4000 - 0.001 \cdot 4000^2 = 24000 \tag{265}$$

transformation af Y:

$$u(Y \mid Y = 3600) = 10 \cdot 3600 - 0.001 \cdot 3600^2 = 23040 \tag{266}$$
$$\tag{267}$$

Transformation af Z:

$$u(Z \mid Z = 3850) = 10 \cdot 3850 - 0.001 \cdot 3850^2 = 23677.5 \tag{268}$$
$$u(Z \mid Z = 2850) = 10 \cdot 2850 - 0.001 \cdot 2850^2 = 20377.5 \tag{269}$$

$$\mathrm{E}(u(X)) = 0.95 \cdot 24000 + 0.05 \cdot 0 = 22800 \tag{270}$$

$$\mathrm{E}(u(Y)) = 0.95 \cdot 23040 + 0.05 \cdot 23040 = 23040 \tag{271}$$

$$\mathrm{E}(u(Z)) = 0.95 \cdot 23677.5 + 0.05 \cdot 20377.5 = 23512.5 \tag{272}$$

Del 7) Vis generelt udtryk for den forventede værdi af $u(W)$

$$u(v) = 10v - 0.001v^2, \qquad v \in \{0, 1, \cdots 4000\} \tag{273}$$

lad W være koncentreret på mængden T:

$$\mathrm{E}(u(W)) = \sum_{w \in T} (10 \cdot w - 0.001w^2)p(w) \tag{274}$$

$$\mathrm{E}(u(W)) = \sum_{w \in T} (10 \cdot w)p(w) - \sum_{w \in T} (0.001w^2)p(w) \tag{275}$$

$$\mathrm{E}(u(W)) = 10 \cdot \sum_{w \in T} (w)p(w) - 0.001 \cdot \sum_{w \in T} (w^2)p(w) \tag{276}$$

$$\mathrm{E}(u(W)) = 10\mathrm{E}(W) - 0.001 \cdot \mathrm{E}(W^2) \tag{277}$$

Vi ved at:

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 \implies \mathrm{Var}(X) + (\mathrm{E}(X))^2 = \mathrm{E}(X^2) \tag{278}$$

Vi bruger dette:

$$\mathrm{E}(u(W)) = 10 \cdot \mathrm{E}(W) - 0.001 \cdot (\mathrm{E}(W))^2 + \mathrm{Var}(W) \tag{279}$$

Som var det ønskede udtryk

Del 8) Udregn variansen af X, Y, Z

Vi bruger formlen for den varians:

$$\sum_{x \in T} (x - \mathrm{E}(X))^2 p(x) \tag{280}$$

Varians af X

$$0.95 \cdot (3800 - 4000)^2 + 0.05 \cdot (0 - 4000)^2 = 760000 \tag{281}$$

Varians af Y: Den er $\mathrm{Var}(Y) = 0$. Vi får altid udbetalt det samme! **Definition 3.7.13**

Varians af Z

$$0.95 \cdot (3850 - 3800)^2 + 0.05 \cdot (2850 - 3800)^2 = 47500 \tag{282}$$

3.7 Øvelse 7

Opgaver: 3.4, 3.13, 3.14, 4.5, 4.6, (4.14)

3.7.1 Opgave 3.4

- 5 terninger kastes

SSH for 3 seksere

Man kan bruge både binomial fordelingen og Polynomialfordelingen.

Vi bruger binomialfordelingen $X := Binom(n = 5, p = 1/6)$

$$p = \frac{1}{6} \tag{283}$$

VI har antalsparameter $n = 5$, og antal succeser $x = 3$

$$P(X = 3) = \binom{5}{3}\left(\frac{1}{6}\right)^3\left(1 - \frac{1}{6}\right)^{5-3} = 0.0321 \tag{284}$$

SSH for mindst 3 seksere

$$P(X \geq 3) = \sum_{i=3}^{n}\binom{5}{i}\left(\frac{1}{6}\right)^i\left(1 - \frac{1}{6}\right)^{5-i} = 0.03549 \tag{285}$$

SSh for præcis 3 ens

Brug hvad vi har udregnet tidligere. SSH for præcis 3 seksere, kan vi gange med 6 for at finde det for alle!

$$P(Z = 3) = 6 \cdot 0.0321 = 0.1929 \tag{286}$$

SSH for mindst 3 ens

Brug hvad vi regnede ud tidligere for mindst 3 seksere

$$P(Z = 3) = 6 \cdot 0.03549 = 0.2129 \tag{287}$$

3.7.2 Opgave 3.13

- $X, Y \sim Uni(0, N)$
- $X \perp\!\!\!\perp Y$

Find $P(X > Y)$

Find middelværdien for X, Y.

Vi ser: $P(X > Y \mid Y = 0) = P(X > 0)$, $P(X > Y \mid Y = 1) = P(X > 1)$. Vi ved at Y, X er ligefordelt sådan at alle ting er lige sandsynlige. Dette implicerer $P(Y = y) = \frac{1}{N+1}, \forall y \in Y$.

Vi kender CDF af den diskrete uniforme fordeling:

$$P(Y \geq k) = \frac{k - a + 1}{n} \tag{288}$$

Sæt det hele sammen:

$$P(X \geq Y) = \frac{1}{N+1} \sum_{i=0}^{N} \frac{i - 0 + 1}{N+1} = \frac{1}{N+1} \frac{1}{N+1} \sum_{i=0}^{N} i + 1 \tag{289}$$

Husk at summen fra 1 til N kan skrives som $= (n+1)n/2$. I vores tilfælde $(n+1+1)(n+1)/2$,grundet vi har $i+1$ i vores sum.

$$P(X \geq Y) = \frac{1}{(N+1)^2} \frac{(N+1)(N+1+1)}{2} = \frac{(N+2)}{2(N+1)} \tag{290}$$

Find $P(X = Y)$

Der er $N + 1$ udfald.

$$P(X = Y, Y = y) = \frac{1}{(1+N)^2} \tag{291}$$

Dette er klart tænk på terninger ssh for 1 dobbelt sekser $1/6^2$.

Vi har $1 + N$ måder at dette kan ske på:

$$P(X = Y)(N+1) \frac{1}{(N+1)^2} = \frac{1}{N+1} \tag{292}$$

Find $P(Z)$ **hvor** $Z \sim max(X, Y)$

Vi ser at: $P(Z = 0) = P(X = 0, Y = 0)$

Og at: $P(Z = 1) = P(X = 1, Y = 1) + P(Y = 1, X = 0) + P(X = 0, Y = 1)$.

Vi prøver at generaliserer observationen:

Find $P(V)$ **hvor** $V \sim min(X, Y)$

DROP AT LAVE

Find $P(W)$ **hvor** $W \sim |X - Y|$

DROP AT LAVE

3.7.3 Opgave 3.14

LAV I KLASSEN

- (X_1, X_2) er en stokastisk vektor
- SE OPLÆG for den simultane fordeling

SSH X_1 **er et lige tal**

Vi husker relationen mellem marginale, betingede og simultane fordelinger!

$$P(X_1 = k) = \sum_{i=1}^{n} P(X_1 = k, X_2 = x_i) \tag{293}$$

Vi ser at X_1 skal være et lige tal:

$$P(X_1 \in \textbf{Lige tal}) = P(X_1 = 0) + P(X_1 = 2) + P(X_1 = 6) = 1 - P(X_1 = -1) \tag{294}$$

$$\begin{aligned} P(X_1 = -1) =& P(X_1 = -1, X_2 = 3) && (295)\\ &+ P(X_1 = -1, X_2 = 1) && (296)\\ &+ P(X_1 = -1, X_2 = -2) && (297) \end{aligned}$$

$$P(X_1 = -1) = 0 + \frac{2}{9} + \frac{1}{9} = \frac{3}{9} \tag{298}$$

Vi finder den sandsynlighed vi ønskede fra start:

$$P(X_1 \in \textbf{Lige tal}) = 1 - P(X_1 = -1) = 1 - \frac{3}{9} = \frac{6}{9} \tag{299}$$

SSH, X_1X_2 er et ulige tal

Kravet er at produktet af de to stokastiske variable skal være et ulige tal. Dette vil implicere at $X_1 \in \{\textbf{ulige tal}\}, X_2 \in \{\textbf{ulige tal}\}$.

$$\begin{align} P(X_1X_2 \in \{\textbf{Ulige tal}\}) &= P(X_1 = -1, X_2 = 3) \tag{300} \\ &+ P(X_1 = -1, X_2 = 1) \tag{301} \\ &= \frac{2}{9} \tag{302} \end{align}$$

SSH for $X_2 > 0$ og $X_1 \geq 0$

$$\begin{align} P(X_2 > 0, X_1 \geq 0) &= P(X_2 = 3, X_1 = 2) \tag{303} \\ &+ P(X_2 = 3, X_1 = 6) \tag{304} \\ &+ P(X_2 = 1, X_1 = 2) \tag{305} \\ &+ P(X_2 = 1, X_1 = 6) \tag{306} \\ &= \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{4}{27} = \frac{13}{27} \tag{307} \end{align}$$

3.7.4 Opgave 4.5

Lav i klassen

- shh for sikring defekt 0.03
- køber pakke med 100 sikringer

SSH for at i en pakke med 100 sikringer maks 2 er er defekte

Brug sætning 4.1.2

VI lader altså vores antal parameter gå mod uendelig. Vi bruger nu en poisson fordeling!

Vi ser at $n \cdot p = \lambda = 100 \cdot 0.03 = 3$

Vi definerer vores stokastiske variabel $X \sim Poisson(\lambda = 3)$

$$P(X \leq 2) = \sum_{i=0}^{2} \frac{\lambda^i}{i!} e^{-\lambda} = \sum_{i=0}^{2} \frac{3^i}{i!} e^{-3} \approx 0.42 \tag{308}$$

3.7.5 Opgave 4.6

- En terning kastes indtil den første sekser opnås

Hvad er ssh for at en sekser opnås inden 6 kast.

$$P(X < 6) = 1 - P(X \geq 5) = 1 - (1 - 1/6)^{5+1} = 0.665 \tag{309}$$

5 + 1 fordi 0 skal tælles med

Hvad er den største værdi af $i \in \mathbb{N}$ hvor $P(X > i) \geq \frac{1}{2}$

$$P(X > 0) = (1 - 1/6)^1 = 0.8333 \tag{310}$$

$$P(X > 1) = (1 - 1/6)^2 = 0.6944 \tag{311}$$

$$P(X > 2) = (1 - 1/6)^3 = 0.5787 \tag{312}$$

$$P(X > 3) = (1 - 1/6)^3 = 0.4822 \tag{313}$$

VI ser at $i = 2$ er det største!!

3.7.6 Optional (4.14)

- En stokastisk variabel X
- $X \sim Poisson(\lambda)$

Hvad er $E\left(2^X\right)$

$Z = 2^X$. VI har så at

$$p(z) = \frac{\lambda^{2^x}}{2^x!}e^{-\lambda} \tag{314}$$

$$\mathrm{E}(Z) = \sum_{i=0}^{\infty} 2^i \frac{\lambda^{2^i}}{2^i!}e^{-\lambda} \tag{315}$$

Vi kan trække en fra i nævneren da den bliver ganget på!

$$= \sum_{i=0}^{\infty} \frac{\lambda^{2^i}}{(2^i-1)!}e^{-\lambda} \tag{316}$$

Man trækker et lambda fra tælleren ud foran sumtegnet!

$$= \lambda \sum_{i=0}^{\infty} \frac{\lambda^{2^i-1}}{(2^i-1)!}e^{-\lambda} \tag{317}$$

Hvad er $E\left((1+X)^{-1}\right)$

3.8 Øvelse 8

Opgaver: 4.4, opgave A, (Opgave H)

3.8.1 Opgave 4.4

- A står ved en lidet trafikkeret vej
- Antal taxaer pr. minut, er poisson fordelt med $\lambda = \frac{1}{30}$

Del 1) Hvad er ssh for A må vente mere end en halv time

Altså poisson fordelingen måler "antal observationer" som vores x. og vores λ som vores parameter. Vi bliver nødt til at gange lambda (det er på minut basis, og vi skal have det på halv time basis) t.

$Y \sim Poisson(\lambda = 1/30t)$

$\lambda = 1/3 * 30$

$$P(Y=0) = \frac{\lambda^x}{x!}e^{-1} = \frac{1^0}{0!}e^{-1} = 0.36787 \tag{318}$$

Del 2) Hvad er ssh for at vente 1 1/2 time.

$\lambda = 1/30 * 90 = 3$

$$P(Y=0) = \frac{3^x}{x!}e^{-3} = \frac{3^0}{0!}e^{-3} = 0.04978 \tag{319}$$

Del 3) SSh for $Y > 0$ Taxa er der før 10 minutter

$\lambda = 1/30 * 10 = 1/3$

$$P(Y>0) = 1 - P(Y=0) = 1 - \frac{(1/3)^0}{0!}e^{-(1/3)} = 0.28346 \tag{320}$$

Del 4) Vis at ventetiden, afrundet nedad til helt minuttal, er geometrisk fordelt med $p = 1 - e^{1/30}$

Den geometriske fordeling:

Antal forsøg inden succes

$$pdf = (1-p)^k p \tag{321}$$

Først ser vi at:

$$P(Y=y) = P(X_y > 0, X_{y-1} = 0) \tag{322}$$

Altså ventetiden må være sådan at man ikke har fået taxa i sidste minut, men har i dette minut.

Brug nu at en simultan fordeling kan skrives som en betinget fordeling

$$P(X_y > 0, X_{y-1} = 0) \tag{323}$$
$$= P(X_y > 0 \mid X_{y-1} = 0) P(X_{y-1} = 0) \tag{324}$$
$$\overset{(*)}{=} (1 - P(X_{y=1} = 0)) P(X_{y-1} = 0) \tag{325}$$

Vi har i (*) brugt at $P(X_y > 0 \mid X_{y-1} = 0)$ Svarer til $P(X_1 > 0)$ som svarer til $1 - P(X_1 = 0)$

Indsæt nødvendige tal:

$$\left(1 - \frac{(1/30)^0}{0!} e^{-1/30}\right) \left(\frac{((t-1)/30)^0}{0!} e^{-(t-1)/30}\right) \tag{326}$$

Vi ser at: $\frac{(t-1/30)^0}{0!} = \frac{1}{1} = 1$

Hvilket betyder:

$$\left(1 - \frac{(1/30)^0}{0!} e^{-1/30}\right) \left(\frac{((t-1)/30)^0}{0!} e^{-(t-1)/30}\right) \tag{327}$$

$$= \left(1 - e^{-1/30}\right) \left(e^{-(t-1)/30}\right) \tag{328}$$

$$\approx \left(1 - e^{-1/30}\right) \left(e^{-1/30 \cdot t}\right) \tag{329}$$

Vi skulle have i den geometriske fordeling: $p = 1 - e^{-1/30}$

$$(1-p)^k p = (1 - (1 - e^{-1/30}))^t (1 - e^{-1/30}) \tag{330}$$

Vi forkorter

$$(1-p)^k p = e^{-1/30 \cdot t}(1 - e^{-1/30}) \tag{331}$$

Vi har vist udtrykket!

3.8.2 Opgave A

Lav i klassen!!!

Cykelforsikring fortsat!

- udbetaling ved mistet cykel 4000
- ssh for cykel stjålet pr. år: 5%
- Maks en cykel stjålet om året
- forsikring pris 400

Del 1)

10 cyklister tegner forsikring:

$$Y \sim Binomial(n = 10, p = 0.05) \tag{332}$$

Del 2) Udregn Forventet antal stjålne cykler, samt forventet udgift

$$\mathrm{E}(Y) = n \cdot p = 10 \cdot 0.05 = 0.5 \tag{333}$$

Forventet udgift:

$$\mathrm{E}(Y) \cdot 4000 = 2000 \tag{334}$$

Del 3) SSh for mere end en cykel bliver stjålet

Få folk til at opskrive binomial koefficienter osv.

$$P(Y > 1) = 1 - P(Y = 1) - P(Y = 0) = 0.08613 \tag{335}$$

Del 4) Antag nu 100 cyklister

$$Z \sim Binomial(n = 100, p = 0.05) \tag{336}$$

$$\mathrm{E}(Z) = 100 \cdot 0.05 = 5 \tag{337}$$

Forventede indtægter:

$$400 \cdot 100 = 40000 \tag{338}$$

Forventede udgifter:

$$4000 \cdot \mathrm{E}(Z) = 4000 \cdot 5 = 20000 \tag{339}$$

Del 5)Ssh for man udgifter overstiger indtægter

Udgifer overstiger indtægter når der er 11, som får stjålet sin cykel:

Med binomial (udregnet på com):

$$P(Z > 10) = 1 - \sum_{i=0}^{10} P(Z = i) = 0.01147 \tag{340}$$

Med poisson:

$lambda = 100 \cdot 0.05 = 5$

$$P(Z > 10) = 1 - P(Z \leq 10) = 0.013695 \tag{341}$$

Del 6) Antag nu nu n=200, Ssh udgifter over indtægter

Dette sker når der er 21 som får stjålet cyklen

$lambda = 200 \cdot 0.05 = 10$

$W \sim Poisson(10)$

$$P(W > 20) = 0.0015882 \tag{342}$$

Del 7)

Klasse diskussion!!!

3.8.3 (Optional) Opgave H

3.9 Øvelse 9

Opgaver: 5.2, 5.3, 5.7, U41.1, U41.2

3.9.1 Opgave 5.2

- X er en kontinuær stok var
- $p(x) = \alpha x^{-(\alpha+1)}$ for $x > 1, \alpha > 0$

Find fordelingsfunktionen for X

Vi ved at $p(x) = F'(x)$ Hvis vi skulle finde sandsynligheden for et udfald ville vi bruge tætheden $p(x)$ lad os sige vi ville finde ssh for at X er i intevallet a til b: da

$$\int_a^b p(x)dx \tag{343}$$

Fordelingsfunktionen er kendetegnen ved for intervallet $(-\infty, \infty)$:

$$\int_{-\infty}^x p(x)dx \tag{344}$$

Vi har dog intervallet $(1, \infty)$

Vi opskriver integralet:

$$\int_1^x \alpha x^{-(\alpha+1)} \tag{345}$$

$$\left[\frac{\alpha}{-\alpha+1-1}x^{-\alpha+1-1}\right]_1^x = [-x^{-\alpha}]_1^x = -x^{-\alpha} + 1 \tag{346}$$

3.9.2 Opgave 5.3

LAV I KLASSEN

fordelingsfunktionen fotr X er givet ved:

$$F(x) = \begin{cases} 0 \quad \text{for } x \leq 0 \\ x/3 \quad \text{for } 0 < x \leq 1 \\ (2x-1)/3 \quad \text{for } 1 < x \leq 2 \\ 1 \quad \text{for } x > 2 \end{cases}$$

Find de følgende sandsynligheder

$$P(0.5 < X < 1) = F(1) - F(0.5) = \frac{1-0.5}{3} = \frac{1}{6} \tag{347}$$

Vi kan ignorere punktsandsynligheden da denne er 0 (i forhold til $\leq$ udtryk i oplæg).

$$P(1 \leq X < 1.5) = F(1.5) - F(1) = \frac{3-1}{3} - \frac{1}{3} = \frac{1}{3} \tag{348}$$

$$P(2/3 < X < 4/3) = F(4/3) - F(2/3) = \frac{2(4/3)-1}{3} - \frac{2/3}{3} \tag{349}$$

$$= \frac{8/3 - \frac{3}{3} - 2/3}{3} = \frac{3/3}{3} = \frac{1}{3} \tag{350}$$

Redegør for kontinuitet

Vi viser kontinuæritet via et lille $\delta > 0$

Først se om: $F(0+\delta) \to 0$ og $F(0-\delta) \to 0$ for $\delta \to 0$. Man ser at for $x/3$ går mod 0, hvis x er tæt på 0. (Trivielt at se 0 går mod 0 for lille x).

Undersøg i en omegn af punktet $x = 1$: $F(x \pm \delta) \to \frac{1}{3}$ for $\delta \to 0$. Det er klart da: $x/3 \to \frac{1}{3}$, for $x = 1 - \delta$ og $(2x-1)/3 \to \frac{1}{3}$, for $x = 1 + \delta$

Undersøg i en omegn af punktet $x = 2$: $F(x \pm \delta) \to 1$ for $\delta \to 0$. man ser at $(2(2-\delta)-3)/3 \to 1$ for $\delta \to 0$. (trivilt at 1 går mod 1)

Kontinuitet er vist. Vi noterer at fordelingsfunktionen overholder at $F : \mathbb{R} \mapsto [0,1]$ og at $F(x) \leq F(x+h), \quad h > 0$. Altså den er defineret på hele den reelle akse, samt at den er monotont voksende!

Find tæthedsfunktionen for X

Vi differentiere de enkelte udtryk og får:

$$p(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{3} & 0 < x \leq 1 \\ \frac{2}{3} & 1 < x \leq 2 \\ 0 & x > 2 \end{cases} \tag{351}$$

3.9.3 Opgave 5.7

LAV I KLASSEN!

- 5.1.5 i bogen:

$$p(x) = \beta x^{\beta-1} \tag{352}$$

- $x \in [0, 1]$

Vis at 5.1.5 (i bogen) har middelværdi $\beta/(\beta+1)$

Vi behøver ikke at teste om middelværdien eksisterer!

Definition på middelværdi!

$$E(X) = \int_{-\infty}^{\infty} xp(x)dx < \infty \tag{353}$$

$$E(X) = \int_{-\infty}^{\infty} x\beta x^{\beta-1} = \int_{-\infty}^{\infty} \beta x^{\beta} \tag{354}$$

Vi ved at x er koncentrerer på intervallet 0 til 1: $x \in (0, 1)$

$$E(X) = \int_{-\infty}^{\infty} x\beta x^{\beta-1} = \int_{-\infty}^{\infty} \beta x^{\beta} \tag{355}$$

Vi har her et uendeligt integrale, men x er koncentreret på en mindre mængde. Vi bruger at $P(\emptyset) = 0$ og at vi må splitte integralerne op (indskudssætningen):

$$\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx, \qquad \text{(Indskudssætningen)} \tag{356}$$

Vi ser at integralerne i intervallet $(-\infty, 0[$ og $]1, \infty)$ er lig 0.

$$E(X) \int_0^1 \beta x^{\beta} = \left[\frac{\beta}{\beta+1} x^{\beta+1}\right]_0^1 = \frac{\beta}{\beta+1} \tag{357}$$

Vi finder variansen

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 \tag{358}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \beta x^{\beta-1} = \int_{-\infty}^{\infty} \beta x^{\beta+1} \tag{359}$$

Analogt med før

$$\mathrm{E}(X^2) = \left[\frac{\beta}{\beta+2} x^{\beta+2}\right]_0^1 = \frac{\beta}{\beta+2} \tag{360}$$

Variansen findex:

$$\mathrm{Var}(X) = \frac{\beta}{\beta+2} - \left(\frac{\beta}{\beta+1}\right)^2 \tag{361}$$

kan evt. forkortes

3.9.4 Opgave U41.1

- $X, Y \sim Uni(0,1)$
- den uniforme fordeling er kontinuær

$$\mathrm{E}(X) = \mathrm{E}(Y) = \frac{1}{2}(a+b) = \frac{1}{2}(1+0) = \frac{1}{2} \tag{362}$$

Brug sætning 6.4.2 - man kan splitte forventinger op.

find E$(6X + 32Y)$

$$\mathrm{E}(6X + 32Y) = \frac{6+32}{2} = 19 \tag{363}$$

Find E(X^3) **og E**$(X^3 + Y^3)$

$$\mathrm{E}(X^3) = \int_0^1 x^3 p(x) = \left[\frac{1}{4}x^4\right]_0^1 = \frac{1}{4} \tag{364}$$

Vi har derfor selvfølgelig $\mathrm{E}(X^3 + Y^3) = 2 \cdot \frac{1}{4} = \frac{1}{2}$

Find Var$(X) = \mathbf{E}(X^2) - [\mathbf{E}(X)]^2$

Vi ved at $\mathrm{E}(X)^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$

$$\mathrm{E}(X^2) = \int_0^1 x^2 p(X) = \left[\frac{1}{3}x^3\right]_0^1 = \frac{1}{3} \tag{365}$$

Varians:

$$\mathrm{Var}(X) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \tag{366}$$

Find tæthed for $Z = X - \frac{1}{2}$

$$p(z) = 1, \quad z \in [-0.5, 0.5] \tag{367}$$

Find E(Z)

Brug sætning 5.2.5. lineær transformation.

$$\mathrm{E}(Z) = \mathrm{E}\left(X - \frac{1}{2}\right) = \mathrm{E}(X) - \frac{1}{2} = 0 \tag{368}$$

Find $F(Z)$

$$F(Z) = z - \frac{1}{2}, \qquad z \in [-0.5, 0.5] \tag{369}$$

3.9.5 Opgave U41.2

- stokastisk variabel X
- $p(x) = \lambda \exp(-\lambda x)$

Del 1 - A) Opskriv fordelingsfunktionen for X **og vis at** $Y = F(X)$ **er ligefordelt på** $[0, 1]$

$$F(x) = 1 - \exp(-\lambda x) \tag{370}$$

Vis at $Y = F(X)$ er ligefordelt på $[0, 1]$

$$P(Y \leq y) = P(F(X) \leq y) = P(X \leq F^{-1}(y)) = P(X \leq x) = F(X) = y \tag{371}$$

$$x = F^{-1}(y) = \ln\left(\frac{1}{1-\lambda}\right)/\lambda \tag{372}$$

$t(X) = F(X) = y$ bruges i sidste led af ligningen!

Vi ser at $P(Y \leq y) = y$ Hvor vi ved at y er fordelingsfunktionen for en uniform fordeling!

Del 2)

3.10 Øvelse 10

12/10/2018, Opgaver: 5.1, 5.5, 5.13, 5.15, U41.3 og U41.4

3.10.1 Opgave 5.1

- $X \sim exponential(\lambda)$
- pdf: $\lambda e^{-\lambda x}$

Find $P(X > x)$**, for alle** $x > 0$

Vi ved at fordelingsfunktion $F(x)$ svarer til $P(X < x)$ hvilket betyder at $P(X > x) = 1 - F(x)$.

Kommentar: vi bruger lille x i fordelingsfunktionen. hvorfor? fordi det er en funktion der tager et tal (en realisation) af X

Vi kan se på wikipedia at exponential fordelingens fordelingsfunktionen CDF er:

$$F(x) = 1 - e^{\lambda x} \tag{373}$$

Så vi har at:

$$P(X > x) = 1 - (1 - e^{\lambda x}) = e^{\lambda x} \tag{374}$$

SSH $P(1 < X < 2)$ **,hvor** $\lambda = 1$

brug (hvor lambda er 1):

$$F(x) = 1 - e^{1x} \tag{375}$$

$$P(1 < X < 2) = F(2) - F(1) = (1 - e^2) - (1 - e^1) = 0.2325 \tag{376}$$

3.10.2 Opgave 5.5

Lav i klassen!

- Laplace-fordelingen
- defineret på hele $\mathbb{R}$
- funktionsforskrift:

$$f(x) = \frac{1}{2}e^{-|x|}, \qquad x \in \mathbb{R} \tag{377}$$

Find fordelingsfunktionen F

Fordelingsfunktionen er: $F(k) = \int_{-\infty}^{k} f(x)dx$

Vi ser, vi må skære integralet op i to dele på grund af normerings operatoren på x.

Først $x < 0$

$$F(a) = \int_{-\infty}^{a} \frac{1}{2}e^{x} = \left[\frac{1}{2}e^{x} + k\right]_{-\infty}^{a} = \left(\frac{1}{2}e^{a} + k\right) - \left(\frac{1}{2}e^{-\infty} + k\right) = \frac{1}{2}e^{a} \tag{378}$$

Nu $x \geq 0$

$$F(a) = \int_{-\infty}^{0} \frac{1}{2}e^{x} + \int_{0}^{a} \frac{1}{2}e^{-x} = \frac{1}{2} + \left[\frac{1}{-1}\frac{1}{2}e^{-x}\right]_{0}^{a} = \frac{1}{2} + \left[-\frac{1}{2}e^{-x}\right]_{0}^{a} \tag{379}$$

$$F(a) = \frac{1}{2} + \left(-\frac{1}{2}e^{-a}\right) - \left(-\frac{1}{2}e^{0}\right) = \frac{1}{2} + \frac{1}{2} - \frac{1}{2}e^{-a} = 1 - \frac{1}{2}e^{-a} \tag{380}$$

Vi kan opskrive fordelingsfunktionen!

$$F(x) = \begin{cases} \frac{1}{2}e^{x}, & x < 0 \\ 1 - \frac{1}{2}e^{-x}, & x \geq 0 \end{cases} \tag{381}$$

Del 2) Find middelværdi

Vi behøver ikke at vise middelværdi og varians eksisterer!

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} xp(x)dx \tag{382}$$

Vi splitter intergralet op i intervallerne $(-\infty, 0)$ og $[0, \infty)$:

(man har her brugt reglen for partiel integration - kig Thomas note/formelsamling) $f(x) = exp(x), g(x) = x$:

for integralet i intervallet $(-\infty, 0)$:

$$\int \frac{1}{2}xe^x dx = \frac{1}{2}(x-1)e^x \tag{383}$$

for integralet i intervallet $[0, \infty)$

$$\int \frac{1}{2}xe^{-x} dx = -\frac{1}{2}(x+1)e^{-x} \tag{384}$$

vi ved at:

$$\int_{-\infty}^{\infty} x\frac{1}{2}e^{-|x|}dx = \int_{-\infty}^{0} \frac{1}{2}xe^x dx + \int_{0}^{\infty} \frac{1}{2}xe^{-x}dx \tag{385}$$

Vi sætter integralernes grænser ind i stamfunktioner udledt ovenfor:

$$\int_{-\infty}^{0} \frac{1}{2}xe^x dx = \left(\frac{1}{2}(0-1)e^0\right) - \left(\frac{1}{2}(-\infty-1)e^{-\infty}\right) = -\frac{1}{2} - 0 = -\frac{1}{2} \tag{386}$$

$$\int_{0}^{\infty} \frac{1}{2}xe^{-x} dx = \left(-\frac{1}{2}(\infty+1)e^{-\infty}\right) - \left(-\frac{1}{2}(0+1)e^0\right) = 0 + \frac{1}{2} = \frac{1}{2} \tag{387}$$

Så vi har at:

$$\mathrm{E}(X) = -\frac{1}{2} + \frac{1}{2} = 0 \tag{388}$$

Find variansen Var(X)

VI ved at $\mathrm{E}(X) = 0$ det betyder at $Var(X) = \mathrm{E}(X^2)$. *Husk på formlen for varians.*

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 p(x) dx \tag{389}$$

vi deler igen integralet op. og bruger reglerne for partiel integration. *Vi ender med at få integralet fra før som et del element.*

I intervallet $(-\infty, 0)$:

$$\frac{1}{2}\int_{-\infty}^{0} x^2 e^x dx = \left[\left(\frac{1}{2}x^2 - x + 1\right) e^x\right]_{-\infty}^{0} = 1 - 0 = 1 \tag{390}$$

I intervallet $[0, \infty)$:

$$\frac{1}{2}\int_{0}^{\infty} x^2 e^{-x} dx = \left[-\left(\frac{1}{2}x^2 + x + 1\right) e^{-x}\right]_{0}^{\infty} = 0 - (-1) = 1 \tag{391}$$

Vi har at:

$$\mathrm{Var}(X) = \frac{1}{2}\int_{-\infty}^{0} x^2 e^x dx + \frac{1}{2}\int_{0}^{\infty} x^2 e^{-x} dx = 1 + 1 = 2 \tag{392}$$

3.10.3 Opgave 5.13

LAV I KLASSEN

- X er en kontinuær stokastisk variabel i intervallet (a, b)
- X har en kontinuer sandsynlighedstæthed p på (a, b)

Vi bruger sætning 5.4.1

$$q(y) = \begin{cases} p(t^{-1}(y))\left|\frac{d}{dy}t^{-1}(y)\right|, & y \in (v, h) \\ 0, & y \notin (v, h) \end{cases} \tag{393}$$

hvor $v = \inf t(I), h = \sup t(I)$ og I er intervallet (a, b)

Til de kommende opgaver kan der siges generalt at: $x = t^{-1}(y)$

Og der skippes ofte (y) fra notation, således at: $\frac{d}{dy}t^{-1}(y)$ bliver til $\frac{d}{dy}t^{-1}$

Del 1) Find tætheden for $exp(X)$

vi har vores transformation givet som $t = exp(\cdot)$ som implicerer at $t^{-1} = ln(\cdot)$.

Vi finder den afledte af vores inverse transformation

$$\frac{d}{dy}t^{-1}(y) = \frac{d}{dy}ln(y) = \frac{1}{y} \tag{394}$$

Vi opskriver:

$$q(y) = \begin{cases} p(ln(y)) \cdot |\frac{1}{y}|, & y \in (e^a, e^b) \\ 0, & \text{ellers} \end{cases} \tag{395}$$

Man ser at faktisk $y \in \mathbb{R}_+ \forall y \in Y$, hvilket betyder, man ikke ville behøve at lave normeringstegnet

Antag resten af opgaven at $a > 0$

Del 2) Find tætheden for $\sqrt{X}$

Vi finder transformationens inverse $t^{-1} = y^2$. og herfra den afledte: $\frac{d}{dy}t^{-1} = 2y$.

$$q(y) = \begin{cases} p(y^2) \cdot 2y, & y \in (\sqrt{a}, \sqrt{b}) \\ 0, & \text{ellers} \end{cases} \tag{396}$$

Vi bemærker at y ikke kan antage værdier under 0, grundet $a > 0$.

Del 3) Find tætheden for $\frac{1}{X}$

Vi finder transformationens inverse $t^{-1} = \frac{1}{y}$

den inverse transformations afledte: $\frac{d}{dy}t^{-1} = -\frac{1}{y^2}$. *Det huskes at man tager den absolutte værdi $\implies$ man fjerner minuset*

$$q(y) = \begin{cases} p\left(\frac{1}{y}\right)\frac{1}{y^2} & y \in \left(\frac{1}{a}, \frac{1}{b}\right) \\ 0, & \text{ellers} \end{cases} \tag{397}$$

Del 4) Find tætheden for X^2

Vi finder den inverse transformation: $t^{-1} = \sqrt{y}$

Den afledte af den inverse transformation: $\frac{d}{dy}t^{-1} = \frac{1}{2}y^{-1/2}$

$$q(y) = \begin{cases} p(\sqrt{y})\frac{1}{2}y^{-1/2}, & (a^2, b^2) \\ 0, & \text{ellers} \end{cases} \tag{398}$$

3.10.4 Opgave 5.15

- $X \sim N(\mu, \sigma^2)$
- $Y = \exp(X)$

Del 1) Find sandsynlighedstætheden for Y

tæthedsfunktionen for normal fordlingen:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{399}$$

Vi finder den inverse transformation: $x = t^{-1}(y) = ln(y)$

Den inverse transformations afledte mht y: $\frac{d}{dy}t^{-1}(y) = \frac{1}{y}$

læg mærke til $\ln(y)$ ind i udtrykket

$$q(y) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(\ln(y)-\mu)^2}{2\sigma^2}\right) \cdot \frac{1}{y}, & y \in (0, \infty) \\ 0, & \text{ellers} \end{cases} \tag{400}$$

Del 2) Vis at $Y = \beta X$ er scala invariant

Vi finder den inverse transformation $x = t^{-1}(\beta y) = ln(\beta y)$. Vi husker at: $ln(\beta y) = ln(\beta) + \ln(y)$

Den inverse transformations afledte mht y:

$$\frac{d}{dy}t^{-1}(\beta y) = \frac{d}{dy}ln(y) + ln(\beta) = \frac{1}{y} \tag{401}$$

Vi indsætter de fundne værdier

$$q(y) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(\ln(\beta)+\ln(y)-\mu)^2}{2\sigma^2}\right) \cdot \frac{1}{y}, & y \in (0,\infty) \\ 0, & \text{ellers} \end{cases} \tag{402}$$

Vi ser den transformerede fordeling stadig er logaritmisk normalfordelt!

Del 3

Vi husker en detalje: $\int_{-\infty}^{\infty} p(x)dx = 1$. Dette betyder, at hvis vi kan skabe det ovenstående integrale, og få det resterende ud foran integralet, så har vi fundet resultatet!

Husk $q(y)$ er 0 når ikke $y \in (0,\infty)$

$$\int_{-\infty}^{\infty} q(y)dy = \int_{0}^{\infty} q(y)dy \tag{403}$$

$$\mathrm{E}(Y) = \int_{0}^{\infty} yq(y)dy \tag{404}$$

$$\mathrm{E}(Y) = \int_{0}^{\infty} y\frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(\ln(y)-\mu)^2}{2\sigma^2}\right) \cdot \frac{1}{y} \tag{405}$$

Vi ser at y går ud med $\frac{1}{y}$ Vi indsætter $\mu = 0, \sigma = 1$ som angivet i opgaveteksten.

$$\mathrm{E}(Y) = \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{(\ln(y))^2}{2}\right) \tag{406}$$

Det bagerste udtryk manipuleres:

$$exp\left(-\frac{ln(y)^2}{2}\right) = exp\left(-\frac{\ln(y)\ln(y)}{2}\right) = exp\left(-\frac{1}{2}\right) exp\left(\ln(y)\ln(y)\right) \tag{407}$$

Går i stå her!

3.10.5 Opgave U41.3

- X er ligefordelt på $(0, 1)$.

Del 1) $S = \mathbb{1}_{(0,0.25)}$ **Find** $P(S = 1)$

$$P(X \in (0, 0.25)) = F(0.25) = \frac{1}{4} \tag{408}$$

Del 2) $S = \mathbb{1}_{(0,p)}$. **Find** $P(S = 1)$

$$P(X \in (0, p)) = F(p) = p \tag{409}$$

Del 3) Beskriv hvordan du kan simulere en trækning fra en stokastisk variabel Y

$P(Y = 1) = \frac{1}{9}$ og $P(Y = 2) = \frac{8}{9}$

Vi ved at fordelingsfunktionen $F : \mathbb{R} \mapsto [0, 1]$. Det betyder at den inverse $F^{-1} : [0, 1] \mapsto \mathbb{R}$. Overvej dette.

Vi kan altså sample fra intervallet $[0, 1]$ og mappe det til en real værdi gennem den inverse fordelingsfunktion:

Vi har implicit givet fordelingsfunktionen ovenfor:

$$F(y) = \begin{cases} 0, & y < 1 \\ \frac{1}{9}, & 1 \leq y < 2 \\ 1, & 2 \leq y \end{cases} \tag{410}$$

Tegn fordelingsfunktionen og den inverse fordelingsfunktion

Det betyder at vi kunne sample således:

$Y = 1$ når $x \in \left(0, \frac{1}{9}\right)$.

$Y = 2$ når $x \in \left(\frac{1}{9}, 1\right)$

3.10.6 Opgave U41.4

- $X \sim N(\mu, \sigma^2)$

Del 1) Hvad er fordelingen af $Y = (X - \mu)/\sigma$

Denne er let, da dette bare er en tilbage skalering af normalfordelingen! Dvs. en standard normalfordeling:

$$Y \sim N(0,1) \tag{411}$$

Del 2) Hvad er fordelingen af $Z = (X - \mu)^2/\sigma^2$

Vi ser dette er:

$$Z = \frac{(X-\mu)^2}{\sigma^2} = \left(\frac{X-\mu}{\sigma}\right)^2 \tag{412}$$

Dette svarer altså til den kvadrerede standard normalfordeling: χ^2-fordelingen.

3.11 Øvelse 11

22/10/2018, opgaver: U43.1.1, U43.1.2, U43.1.3 U43.1.4

3.11.1 Opgave U43.1.1

- X, Y er ligefordelt på A
- $A = [0,1] \times [0,1]$
- $p(x,y) = \mathbb{1}_A(x,y)$

Tegn 2-D sketch af definitionsmængden

Del 1) Udregn $P(X < 0.1, Y < 0.6)$

$$\begin{aligned}
P(X < 0.1, Y < 0.6) &= \int_0^{0.6}\int_0^{0.1} \mathbb{1}_A(x,y)dxdy && (413)\\
&= \int_0^{0.6} [x]_0^{0.1}\mathbb{1}_A(y)dy && (414)\\
&= [x]_0^{0.1}[y]_0^{0.6} && (415)\\
&= (0.1-0)\cdot(0.6-0) && (416)\\
&= 0.1\cdot 0.6 = 0.06 && (417)
\end{aligned}$$

Del 2) Udregn $P(0.25 < X < 0.75, 0.4 < Y < 0.6$

Analogt med før - opskrivningen er ikke nødvendig:

$$P(0.25 < X < 0.75, 0.4 < Y < 0.6) = 0.5 \cdot 0.2 = 0.1 \tag{418}$$

Del 3) Udregn $P(X < 0.1)$

Her bruges at man kan integrere irrelevante variable ud: **sætning 6.1.3**

$$q(x) = \int_{\mathbb{R}} p(x,y)dy \tag{419}$$

dvs:

$$q(x) = \mathbb{1}_{[1,0]}(x) \tag{420}$$

Vi finder nu det ønskede udtryk

$$P(X < 0.1) = \int_0^{0.1} \mathbb{1}_{[0,1]}(x) = [x]_0^{0.1} = 0.1 \tag{421}$$

Del 4) Find den marginale fordeling for X

Igen bruges sætning 6.1.3

$$q(x) = \int_{\mathbb{R}} p(x,y)dy \tag{422}$$

dvs:

$$q(x) = \mathbb{1}_{[0,1]}(x) \tag{423}$$

Altså vi svarede indirekte på det problem før!

$p_x(x) = \mathbb{1}_{[0,1]}(x)$ og lige så $p_y(y) = \mathbb{1}_{[0,1]}(y)$ Vi ser altså nu at $p(x,y) = p_x(x) \cdot p_y(y)$

3.11.2 Opgave U43.1.2

- X, Y er uafhængige
- X, Y er ligfordelte på intervallet $[0, 1]$
- $Y* = 2Y$

Find $E(Y*), V(Y*)$

Brug sætning **6.3.2** som viser at hvis $X \perp\!\!\!\perp Y \implies X \perp\!\!\!\perp \phi(Y)$

Vi har uafhængighed hvilket implicerer:

$$p(x, y*) = p(x)p(y*) \tag{424}$$

Nu integreres X ud:

$$p(y*) = p(y*) \int_{\mathbb{R}} p(x)dx = p(y*) \tag{425}$$

Vi finder den forventede værdi:

2 er den øvre grænse, 0 er den nedre grænse for Y.

$$\mathrm{E}(Y*) = 2 \cdot \mathrm{E}(Y) = 2 \cdot 0.5 = 1 \tag{426}$$

Variansen findes ved: $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$.

$$\mathrm{Var}(Y) = \frac{1}{12}(0 - 1)^2 = \frac{1}{12} \tag{427}$$

$$\mathrm{Var}(Y*) = 2^2\mathrm{Var}(Y) = 4 \cdot \frac{1}{12} = \frac{1}{3} \tag{428}$$

Del 2) Tætheden for $Y*$

Tætheden er:

tætheden for en uniform (kontinuær) distribution er: $p(x) = \frac{1}{b-a}\mathbb{1}_{x\in[a,b]}(x)$

Vi bruger dette:

$$p(y*) = \frac{1}{2 - 0}\mathbb{1}_{x\in[0,2]}(y*) \tag{429}$$

Del 3) $Z = X + Y*$ **Find tætheden for** Z**,** $q(z)$

Vi bruger **korollar 6.3.2** (få en studerende til at læse op).

$$q(z) = \int_{-\infty}^{\infty} p_1(x)p_2(z-x)dx \tag{430}$$

$$p_x(x)p_{y*}(z-x) = \mathbb{1}_{[0,1]\times[0,2]}\frac{1}{2}(x)(z-x) = \frac{1}{2}(xz - x^2) \tag{431}$$

Nu integreres denne:

$$\begin{aligned} q(z) &= \int_{\mathbb{R}} \frac{1}{2}(xz - x^2)dx && (432)\\ &= \frac{1}{2}\int_{\mathbb{R}} (xz - x^2)dx && (433)\\ &= \left[\frac{1}{2}\frac{1}{2}x^2 z - \frac{1}{2}\frac{1}{3}x^3\right]_0^1 = \frac{1}{4}z - \frac{1}{6} && (434) \end{aligned}$$

NOGET ER GALT

3.11.3 Opgave U43.1.3

- $X, Y \in [5, 10] \times [3, 7]$
- $p(x, y) = \frac{1}{20}\mathbb{1}_{[5,10]\times[3,7]}(x, y)$

Skitser definition mængden.

Del 1) Forklar hvorfor $p(x, y)$ **er en tæthedsfunktion**

notér at $(10 - 5) \times (7 - 3) = 20$, således at den samlede areal under kurven er 1.

Find $P(6 \leq X \leq 10, 4 \leq Y \leq 6)$

$$P(6 \leq X \leq 10, 4 \leq Y \leq 6) = \int_6^{10} \int_4^6 \frac{1}{20} \mathbb{1}_{[5,10]\times[3,7]}(x,y)dydx \tag{435}$$

$$= \frac{1}{20} \int_6^{10} \int_4^6 \mathbb{1}_{[5,10]\times[3,7]}(x,y)dydx \tag{436}$$

$$= \frac{1}{20} \int_6^{10} \mathbb{1}_{[5,10]}(x) \left[y\right]_4^6 dx \tag{437}$$

$$= \frac{1}{20} \int_6^{10} \mathbb{1}_{[5,10]}(x)(6-4)dx \tag{438}$$

$$= \frac{2}{20} \int_6^{10} \mathbb{1}_{[5,10]}(x)dx \tag{439}$$

$$= \frac{2}{20} \left[x\right]_6^{10} \tag{440}$$

$$= \frac{2}{20}(10-6) = \frac{8}{20} \tag{441}$$

Del 3) Find de marginale fordelinger

$$p(x) = \frac{1}{20} \int_3^7 \mathbb{1}_{[5,10]\times[3,7]}(x,y)dy = \frac{4}{20} \mathbb{1}_{[5,10]}(x,y) \tag{442}$$

Omvendt for Y:

$$p(y) = \frac{5}{20} \mathbb{1}_{[3,7]}(x,y) \tag{443}$$

Del 4) Find $\mathbf{E}(X)$

For en ligefordeling har man middelværdi ved (a og b er enderne):

$$\mathrm{E}(X) = \frac{a+b}{2} \tag{444}$$

Vi bruger dette

$$\mathrm{E}(X) = \frac{5+10}{2} = 7.5 \tag{445}$$

3.11.4 Opgave U43.1.4

- $X, Y \in [0, \infty)$
- $p(x,y) = 6\exp(-2x-3y)$

Praktisk at vide:

$$\int \exp(-bx)dx = -\frac{\exp(-bx)}{b} \tag{446}$$

Del 1 - a) find $P(X \leq 2, Y \leq 4)$

$$\begin{aligned}
P(X \leq 2, Y \leq 4) &= \int_0^2 \int_0^4 6\exp(-2x-3y)dydx && (447)\\
&= \int_0^2 \int_0^4 6\exp(-2x)\exp(-3y)dydx && (448)\\
&= 6\int_0^2 \exp(-2x)\left(\int_0^4 \exp(-3y)dy\right)dx && (449)\\
&= 6\int_0^2 \exp(-2x)\left(\left[-\frac{\exp(-3y)}{3}\right]_0^4\right)dx && (450)
\end{aligned}$$

Vi løser det indre problem:

$$\begin{aligned}
\left[-\frac{\exp(-3y)}{3}\right]_0^4 &= \left(-\frac{\exp(-12)}{3}\right) - \left(-\frac{1}{3}\right) && (451)\\
&= \frac{1}{3} + \frac{\exp(-12)}{3} && (452)\\
&= \frac{1-\exp(-12)}{3} && (453)
\end{aligned}$$

Vi indsætter dette!

$$6\int_0^2 \exp(-2x)\left(\frac{1-\exp(-12)}{3}\right)dx = 6\left(\frac{1-\exp(-12)}{3}\right)\int_0^2 \exp(-2x)dx \tag{454}$$

$$= 6\left(\frac{1-\exp(-12)}{3}\right)\left[-\frac{\exp(-2x)}{2}\right]_0^2 \tag{455}$$

Vi udregner det inderste:

$$\left[-\frac{\exp(-2x)}{2}\right]_0^2 = \left(-\frac{\exp(-2\cdot 2)}{2}\right) - \left(-\frac{1}{2}\right) \tag{456}$$

$$= \frac{1}{2} - \frac{\exp(-4)}{2} \tag{457}$$

$$= \frac{1-\exp(-4)}{2} \tag{458}$$

Dette indsættes:

$$6\left(\frac{1-\exp(-12)}{3}\right)\left(\frac{1-\exp(-4)}{2}\right) = (1-\exp(-12))\,(1-\exp(-4)) \tag{459}$$

Del 1 - b) find $P(X > 1, Y \leq 3)$

Lav i klassen! Efter samme opskrift som ovenfor:

Resultat:

$$P(X > 1, Y \leq 3) = \exp(-2)\,(1-\exp(9)) \tag{460}$$

Find de marginale fordelinger $p_y(y), p_x(x)$

Man integrere den ene variabel ud: dvs, integrer y ud, hvis man ønsker at finde $p_x(x)$, og vice versa.

$$p_y(y) = \int_{\mathbb{R}} p(x,y)dx = \int_{\mathbb{R}} 6\exp(-2x-3y)dx \tag{461}$$
$$= 6\exp(-3y)\int_{\mathbb{R}} \exp(-2x)dx \tag{462}$$
$$= 6\exp(-3y)\cdot\frac{1}{2} \tag{463}$$
$$= 3\exp(-3y) \tag{464}$$

Hvor man har udnyttet at $\int_{\mathbb{R}} \exp(-2x)dx = \frac{1}{2}$

Lad klassen lave anden halvdel!

Resultatet er analogt for Y, bare hvor

$$p_x(x) = \int_{\mathbb{R}} p(x,y)dy = \frac{1}{3}\cdot 6\exp(-2x) = 2\exp(-2x) \tag{465}$$

Del 3) Find fordelingsfunktionen for X

Jeg udskifter x med a for ikke at gøre notationen forvirrende!

$$F(a)\int_0^a p(x)dx = \int_0^a 2\exp(-2x)dx \tag{466}$$
$$= 2\left[-\frac{-\exp(-2x)}{2}\right]_0^a \tag{467}$$
$$= 2\,(1) - 2\left(-\frac{\exp(-2a)}{2}\right) \tag{468}$$
$$= 1 - \exp(-2a) \tag{469}$$

Dette indsættes:

$$F(x) = 1 - \exp(-2x) \tag{470}$$

Medianen findes

$$
\begin{aligned}
0.5 = 1 - \exp(-2x) &\Leftrightarrow 0.5 = \exp(-2x) && (471)\\
&\Leftrightarrow ln(0.5) = -2x && (472)\\
&\Leftrightarrow -\frac{ln(0.5)}{2} = x && (473)
\end{aligned}
$$

Del 4) Vis uafhængighed

Vi ser at $p(x)p(y) = p(x,y)$ - Dette er sætning **6.2.1**

$$
(2\exp(-2x))\,(3\exp(-3y)) = 6\exp(-2x-3y) \tag{474}
$$

3.12 Øvelse 12

26/10/2018, opgaver: U43.2.1, U43.2.2, U43.2.3, U43.2.4, U43.2.5 fra bogen: 6.4, 6.21

3.12.1 Opgave U43.2.1

- $A = \{x, y \mid x^2 + y^2 \leq 1$
- $p(x,y) = \frac{1}{\pi}\mathbb{1}_A(x,y)$

Del 1) Tegn $p(x,y)$

- Tegn på tavlen en tredimensionel enhedscirkel. Højden: $\frac{1}{\pi} \approx \frac{1}{3}$

Del 2) Find den marginale tæthed for X

For at finde den marginale tæthed skal man integrere Y ud af udtrykket $p(x,y)$

Først noteres at:

$$
X^2 + y^2 \leq 1 \Leftrightarrow y \leq \sqrt{1-x^2} \tag{475}
$$

$$p_X(x) = \int_{\mathbb{R}} \frac{1}{\pi} \mathbb{1}_A(x,y)dy \tag{476}$$
$$= 2\frac{1}{\pi} \int_0^{\sqrt{1-x^2}} \mathbb{1}_A(x,y)dy \tag{477}$$
$$= 2\frac{1}{\pi} \left(\left(\sqrt{1-x^2} \right) - (0) \right) \tag{478}$$
$$= \frac{2\sqrt{1-x^2}}{\pi} \tag{479}$$

2-tallet kommer fra at y både kunne have været positivt og negativt!!!

Del 2) Find den marginale tæthed

Kig på github!

3.12.2 Opgave U43.2.2

- $A = \{x, y \mid x \in [1,2], y \in [1,2]\}$
- $p(x,y) = \mathbb{1}_A p(x,y)$

Del 1) Tegn $p(x,y)$

Gør på tavlen. 3-dimensionel tegning.

Del 2) Find de marginale tætheder p_Y, p_X

$$p_X(x) = \int_{\mathbb{R}} \mathbb{1}_A(x,y)dy = \int_1^2 \mathbb{1}_A(x,y)dy = \mathbb{1}_{[1,2]}(x) \int_1^2 \mathbb{1}_A(y)dy = \mathbb{1}_{[1,2]}(x) \tag{480}$$

Analogt for $p_Y(y) =$

$$p_Y(y) = \mathbb{1}_{[1,2]}(y) \tag{481}$$

Del 3) Definér $Z = X + Y$. Find $\mathbf{E}(Z), \mathbf{Var}(Z)$

Vi ser at $X \perp\!\!\!\perp Y$

Det implicerer at:

$$\mathrm{E}(Z) = \mathrm{E}(X) + \mathrm{E}(Y) = \frac{3}{2} + \frac{3}{2} = 3 \tag{482}$$

Hvor man har udnyttet at $\mathrm{E}(X) = \mathrm{E}(Y) = \frac{a+b}{2} = \frac{1+2}{2} = \frac{3}{2}$. Man husker at $\frac{a+b}{2}$ er middelværdien for den uniforme fordeling!

Variansen findes:

VI husker de er uafhængige hvilket gør vi kan sige - Fundet på wikipedia - generelt er wikipedia bedre til egenskaber end bogen - bogen er meget rodet opbygget:

$$\mathrm{Var}(Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) = \frac{1}{12} + \frac{1}{12} = \frac{1}{6} \tag{483}$$

Vi finder variansen af X:

$$\mathrm{Var}(X) = \frac{1}{12}(a-b)^2 = \frac{1}{12} \tag{484}$$

Find tætheden $q(z)$ for Z

Vi bruger Korollar 6.3.2

$$\begin{aligned} p(x, x-z) &= \mathbb{1}(1 \leq x \leq 2)\mathbb{1}(1 \leq z - x \leq 2) && (485)\\ &= \mathbb{1}(2 \leq z \leq 3)\mathbb{1}(1 \leq x \leq z-1) && (486)\\ &+ \mathbb{1}(3 \leq z \leq 4)\mathbb{1}(z-2 \leq x \leq 2) && (487) \end{aligned}$$

Kig github for illustration!

Vi bruger dette:

$$\begin{aligned} q(z) &= \int_{\mathbb{R}} \mathbb{1}(2 \leq z \leq 3)\mathbb{1}(1 \leq x \leq z-1) + \mathbb{1}(3 \leq z \leq 4)\mathbb{1}(z-2 \leq x \leq 2)dx && (488)\\ &= \int_{\mathbb{R}} \mathbb{1}(2 \leq z \leq 3)\mathbb{1}(1 \leq x \leq z-1)dx + \int_{\mathbb{R}} \mathbb{1}(3 \leq z \leq 4)\mathbb{1}(z-2 \leq x \leq 2)dx && (489)\\ &= \mathbb{1}(2 \leq z \leq 3)\int_{1}^{z-1} \mathbb{1}dx + \mathbb{1}(3 \leq z \leq 4)\int_{z-2}^{2} \mathbb{1}dx && (490) \end{aligned}$$

Indsætter i stamfunktionen giver:

$$q(z) = \mathbb{1}(2 \leq z \leq 3)(z-2) + \mathbb{1}(3 \leq z \leq 4)(4-z) \tag{491}$$

del 4) Benyt $q(z)$ til at udregne $\mathbf{E}(z)$

$$\begin{aligned}
\int_{\mathbb{R}} zq(z)dz &= \int_{\mathbb{R}} \mathbb{1}(2 \leq z \leq 3)(z^2-2z) + \mathbb{1}(3 \leq z \leq 4)(4z-z^2)dz && (492)\\
&= \int_2^3 (z^2-2z)dz + \int_3^4 (4z-z^2)dz && (493)\\
&= \left[\frac{1}{3}z^3 - z^2\right]_2^3 + \left[\frac{1}{2}z^2 - \frac{1}{3}z^3\right]_3^4 && (494)\\
&= 3 && (495)
\end{aligned}$$

Del 5) Udregn $\mathbf{Cov}(X, Z)$

$$\begin{aligned}
\mathrm{Cov}(X, Z) &= \mathrm{Cov}(X, X+Y) && (496)\\
&= \mathrm{Cov}(X, X) + \mathrm{Cov}(X, Y) && (497)\\
&= \mathrm{Var}(X) = \frac{1}{12} && (498)
\end{aligned}$$

Vi husker at X, Y er uafhængige

Vi husker at variansen af X er fundet tidligere

3.12.3 Opgave U43.2.3

- $p_X(x) = \exp(-x)$
- $p_Y(y) = \exp(-y)$
- $X \perp\!\!\!\perp Y$
- X, Y er defineret på $\mathbb{R}_+$

Del 1) Find tætheden $p(x, y)$

Grundet uafhængighed mellem X, Y ved vi at: $p(x, y) = p_X(x)p_Y(y)$

Vi bruger dette:

$$p(x, y) = \exp(-x)\exp(-y) \tag{499}$$

Del 2) find tætheden for $Z = X + Y$

Vi gør som tidligere:

$$\begin{aligned} p(x, z - x) &= \mathbb{1}(0 < x < z - x)\exp(-x)\exp(-(z - x)) && (500)\\ &= \mathbb{1}(0 < x < z - x)\exp(-x)\exp(x)\exp(-z) && (501)\\ &= \mathbb{1}(0 < x < z - x)\exp(-z) && (502) \end{aligned}$$

udtrykket $\cdot 1$ er kun for at understrege der altid står 1.

$$q(z) = \int_0^z \exp(-z)dx = \exp(-z)\int_0^z \mathbb{1}dx = z\exp(-z) \tag{503}$$

Del 3) Find tætheden for $Z = X - Y$

Vi bruger korollar 6.3.2 og indser at: $Z = X - Y \implies Y = X - Z$

$$\begin{aligned} p(x, x - z) &= \mathbb{1}(0 < z < x)\exp(-x)\exp(-(x - z)) && (504)\\ &= \mathbb{1}(0 < z < x)\exp(-2x)\exp(-z) && (505) \end{aligned}$$

Vi finder tætheden $q(z)$

$$\begin{aligned} q(z) &= \int_{\mathbb{R}} \mathbb{1}(0 < z < x)\exp(-2x)\exp(-z)dx && (506)\\ &= \exp(-z)\int_0^\infty \exp(-2x)dx && (507)\\ &= \exp(-z)\left[-\frac{\exp(-2x)}{2}\right]_0^\infty && (508)\\ &= \frac{1}{2}\exp(-z) && (509) \end{aligned}$$

VI husker at i anden nederste ligning skal kun x indsættes i square brackets.

3.12.4 Opgave U43.2.4

- X_1, X_2, X_3, X_4 er identiske og uafhængige
- $\mathrm{E}(X_i) = 5, \mathrm{Var}(X_i) = 9$
- $Y = X_1 + 2X_2 - X_4$

$$\mathrm{E}(Y) = 5 + 2 \cdot 5 - 5 = 10 \tag{510}$$

Man husker de stokastiske variable er uafhængige

$$\mathrm{Var}(Y) = \mathrm{Var}(X_1 + 2X_2 - X_4) = \mathrm{Var}(X_1) + 2^2\mathrm{Var}(X_2) + \mathrm{Var}(X_4) = 6 \cdot 9 = 54 \tag{511}$$

3.12.5 Opgave U43.2.5

Drop denne opgave! Tidspres gør det umuligt at nå!

3.12.6 Opgave 6.4

- Man har $p(x, y)$ givet ved:

$$p(x,y) = \begin{cases} 3xy^{-2}, & x \in (0,1), y \in (1,3) \\ 0 & \text{ellers} \end{cases} \tag{512}$$

Find de marginale fordelinger for X, Y og vis uafhængighed!

Lad $A = \{x, y \mid x \in (0,1), y \in (1,3)\}$

$$
\begin{aligned}
p_X(x) &= \int_{\mathbb{R}} \mathbb{1}_A(x,y) 3xy^{-2} dy && (513)\\
&= \mathbb{1}_{[0,1]}(x) x \int_1^3 3y^{-2} dy && (514)\\
&= \mathbb{1}_{[0,1]}(x) x \left[3 \cdot \frac{1}{-1} y^{-1}\right]_1^3 && (515)\\
&= \mathbb{1}_{[0,1]}(x) x \left[\frac{-3}{y}\right]_1^3 && (516)\\
&= \mathbb{1}_{[0,1]}(x) x \left(-1+3\right) && (517)\\
&= \mathbb{1}_{[0,1]}(x) x \cdot 2 && (518)\\
&= 2x, \quad x \in (0,1) && (519)
\end{aligned}
$$

Det samme gøres for y

$$
\begin{aligned}
p_Y(y) &= \int_{\mathbb{R}} \mathbb{1}_A(x,y) 3xy^{-2} dy && (520)\\
&= \mathbb{1}_{[1,3]}(y) y^{-2} \cdot 3 \int_0^1 x dx && (521)\\
&= \mathbb{1}_{[1,3]}(y) y^{-2} \cdot 3 \left(\frac{1}{2}\right) && (522)\\
&= \frac{3}{2} y^{-2}, \qquad y \in (1,3) && (523)
\end{aligned}
$$

Vi tester for uafhængighed:

$$
\begin{aligned}
p_X(x) \cdot p_Y(y) &= \frac{3}{2} y^{-2} 2x && (524)\\
&= 3y^{-2} x && (525)\\
&= p(x,y) && (526)
\end{aligned}
$$

Hvilket viser uafhængighed.

3.12.7 Opgave 6.21

- $X \sim Uni(-1,1)$

- $Y = X^2$

Vis at $\text{Corr}(X, Y) = 0$

$$\text{Corr} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \tag{527}$$

Vi finder Covariansen:

$$\text{Cov}(X, Y) = \text{E}(X - \text{E}(X))\text{E}(Y - \text{E}(Y)) = \text{E}(XY) - \text{E}(X)\text{E}(Y) \tag{528}$$

Vi ser hurtigt at $\text{E}(X) = \text{E}(Y) = 0$. (evt - tegn for at overbevise klasse).

$$\text{E}(X \cdot Y) = \int_{-1}^{-1} x \cdot x^2 dx = \left(\frac{1}{4}\right) - \left(\frac{1}{4}\right) = 0 \tag{529}$$

Herfra ser vi let at:

$$\text{Cov}(X, Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y) = 0 - 0 = 0 \tag{530}$$

De er ikke uafhængige! Kan vises formelt, men bedre med intuition ved at tegne!

Vis github!

3.13 Øvelse 13

28/10/2018, opgaver: 44.1.1, 44.1.2, 44.1.3 44.1.4

3.13.1 Opgave 44.1.1

- to terninger (stokastiske variable) X_1, X_2
- $(X_1, X_2) \in \{1, 2, 3, 4, 5, 6\}^2 = \{x_{1,i}\} \times \{x_{2,j}\}, \quad i, j \in \{1, 2, 3, 4, 5, 6\}$
- $Z = X_1 + X_2$

Den vigtige regel:

$$p_x(x) = \int_y p_{x,y}(x, y)dy = \int_y p_{x|y}(x \mid y)p_y(y)dy \tag{531}$$

Find $P(X_1 = i \mid Z \geq 4)$

Vi noterer først vi ikke har kontinuerte stokastiske variable!

Man får en god idé

$$P(X_1 = i \mid Z \geq 4) = \frac{P(X_1 = i, Z \geq 4)}{P(Z \geq 4)} \tag{532}$$

(skits summen af to terninger på tavlen)

Vi indser hurtigt at $P(Z \geq 4) = \frac{33}{36}$

Vi indser også at:

$$P(X_1 = i \mid Z \geq 4) = \frac{P(X_1 = i, X_1 + X_2 \geq 4)}{33/36} = \frac{P(X_1 = i, X_2 \geq 4 - i)}{33/36} \tag{533}$$

Vi indser at i er en konstant og vi nu har uafhængighed i den simultane sandsynlighed således at:

$$P(X_1 = i)P(X_2 \geq 4 - i) \tag{534}$$

Husk at $P(X_1 = i) = \frac{1}{6}$

Vi kan opskrive det hele i et samlet udtryk:

$$P(X_1 = i)P(X_2 \geq 4 - i) = \frac{1}{6}P(X_2 \geq 4 - i) \tag{535}$$

$$P(X_1 = i)P(X_2 \geq 4 - i) = \frac{1}{6} \cdot \frac{4 - 1 + i}{6} \quad i < 4 \tag{536}$$

over i siger man bare $1/6 \times 1/6$

$$P(X_1 = i)P(X_2 \geq 4 - i) = \frac{3 + i}{36} \quad i < 4 \tag{537}$$

Vi husker at dele med 33/36

$$P(X_1 = i \mid Z \geq 4) = \begin{cases} 4/33 & i = 1 \\ 5/33 & i = 2 \\ 6/33 & i \geq 3 \end{cases} \tag{538}$$

Find E$(X_1 = i \mid Z \geq 4)$

vi ved at $P(X_1 = i) = \frac{1}{36}$. VI kan derfor sige:

$$\sum_{i=1}^{6} i \cdot \min\left(\frac{3+i}{33}, \frac{6}{33}\right) = \frac{4 \cdot 1 + 5 \cdot 2 + (3 + 4 + 5 + 6) \cdot 6}{33} \tag{539}$$

3.13.2 Opgave 44.1.2

- $Z \in \{1, 2\}$ angiver kommune
- $V \in \{0, 1\}$ angiver om man er velhavende
- $P(V = 1 \mid Z = 1) = 0.8 = 1 - P(V = 0 \mid Z = 0)$
- $P(V = 1 \mid Z = 2) = 0.1$

Udregn E$(V \mid Z = 1)$ **og E**$(V \mid Z = 2$

Udtrykkene er udtryk for sandsynligheden for at være velhavende betinget på hvilken kommune man kommer fra.

$$\mathrm{E}(V \mid Z = 1) = 0 \cdot P(V = 0 \mid Z = 0) + 1 \cdot (V = 1 \mid Z = 1) = 0.2 \cdot 0 + 0.8 \cdot 1 = 0.8 \tag{540}$$

For kommune 2:

$$\mathrm{E}(V \mid Z = 2) = 1 \cdot P(V = 1 \mid Z = 2) + 0 \cdot P(V = 0 \mid Z = 2) = 0.1 \tag{541}$$

Vis udtrykket:

$$\mathrm{E}(V \mid Z = z) = f(z) = 0.8 \cdot \mathbb{1}(z = 1) + 0.2 \cdot \mathbb{1}(Z = 2) \tag{542}$$

Man ser at hvis $z = 1 \implies E(V \mid Z = 1) = 0.8$

og omvendt: $z = 2 \implies E(V \mid Z = 2) = 0.1$

Hvad udtrykker E$(V \mid Z = z)$

Det betyder at vores forventning er afhængig af realization af z.

Del 4)

Man definerer nu den stokastiske variabel *Den betingede middelværdi af V givet Z.*

$$\mathrm{E}(V \mid Z) = f(z) \tag{543}$$

Vis at:

$$\mathrm{E}(f(z)) = \mathrm{E}(\mathrm{E}(V \mid Z)) = 0.8P(Z = 1) + 0.1P(Z = 2) \tag{544}$$

Det følger næsten naturligt:

$$\mathrm{E}(f(z)) = \mathrm{E}(0.8 \cdot \mathbb{1}(z = 1) + 0.1 \cdot \mathbb{1}(z = 2)) \tag{545}$$

Herfra følger det da $V \in \{0, 1\}$

$$\mathrm{E}(f(z)) = 0.8 \cdot P(Z = 1) + 0.1 \cdot P(Z = 2) \tag{546}$$

3.13.3 Opgave 44.1.3

- X er ligefordelt på $A = [0, 10]$

Del 1) Opskriv tætheden $p(x)$ **for** X **og vis** $P(X) > 5 = \frac{1}{2}$

Tegn tæthedsfunktionen.

Man ved at $F(x) \to 1$ for $x \to \infty$. nærmere bestemt ved man at $F(10) = 1$. Man ved at $\int \mathbb{1}_A(x)$ vil være x, så man skal gange en konstant på for at få $F(10) = 1$. Hel konkret $10 \cdot c = 1 \implies c = 1/10$

$$p(x) = \frac{1}{10}\mathbb{1}_A(x) \tag{547}$$

$$P(X > 5) = \int_0^5 \frac{1}{10}\mathbb{1}_A(x) = \frac{1}{10}\int \mathbb{1}_A(x) = \frac{1}{10}[x]_0^5 = \frac{1}{10}(5-0) = 0.5 \tag{548}$$

Del 2) Find $\mathbf{E}(X)$

$$\mathrm{E}(x) = \int_{-\infty}^{\infty} p(x)x \tag{549}$$

Vi ved at indikator funktionen kun er defineret i intervallet $[0, 10]$. så vi kan skrive:

$$\mathrm{E}(X) = \int_0^{10} \frac{1}{10}x \cdot \mathbb{1}_A(x) \tag{550}$$
$$= \frac{1}{0}\int_0^{10} x \tag{551}$$
$$= \frac{1}{10}\left[\frac{1}{2}x^2\right]_0^{10} \tag{552}$$
$$= \frac{1}{10} \cdot \frac{1}{2} \cdot 10^2 = 5 \tag{553}$$

Vis at tætheden for $X \mid X > 5$ kan skrive som:

Skitser det givne på en tegning!

$$q(x) = \frac{2}{10}\mathbb{1}(5 < x < 10) \tag{554}$$

Man indser hurtigt at: $X \in [0, 5] \cap X \in (5, 10] = \emptyset$. Vi kan altså herfra konkludere at $X \mid X > 5$ kun er defineret på intervallet $(5, 10]$.

$X \mid X > 5$ er stadig uniformt fordelt, og vi kan derfor sige at: $q(x) = c \cdot \mathbb{1}(5 < x \leq 10)$. Igen ved vi også at $Q(10) = 1$. Vi kan hurtige udlede at $c = \frac{1}{5}$. hvormed det ønskede resultat er vist.

Del 4) Er $\mathbf{E}(X \mid X > 5) = 7.5$?

Først se på tegningen. Herfra burde det fremgår tydeligt. Mere formelt:

$$
\begin{aligned}
\mathrm{E}(X \mid X > 5) &= \int_{-\infty}^{\infty} x \cdot \frac{1}{5}\mathbb{1}(5 < x < 10) && (555)\\
&= \frac{1}{5}\int_{5}^{10} x \cdot \mathbb{1}(5 < x < 10) && (556)\\
&= \frac{1}{5}\left[\frac{1}{2}x^2\right]_5^{10} && (557)\\
&= \frac{1}{5}\frac{1}{2}(10^2 - 5^2) && (558)\\
&= \frac{1}{5}\frac{1}{2} \cdot 75 && (559)\\
&= 7.5 && (560)
\end{aligned}
$$

3.13.4 Opgave 44.1.4

- X angiver ratingen fra 0 til 1
- Y angiver værdipapirets værdi i 1000 \$
- X, Y er ligefordelt på mængden B
- $B = \{(x, y) \in \mathbb{R}^2 \mid 0 < x < 1, 0.5 + 2x \leq y \leq 2.5 + 2x\}$

Lad os starte med at tegne B. Kig github!

Find tæthedsfunktionen $f_{X,Y}(x, y)$ for den simultane fordeling for (X, Y)

Vi hurtigt indser at de marginale fordelinge må blive 1. Den hurtigste måde at konstanten c på (tænk simultan fordeling $f(x, y) = c\mathbb{1}_B(x, y)$). er at finde arealet af B.

$$
\frac{1}{c} = h \cdot l = 1 \cdot 2 = 2 \implies c = \frac{1}{2} \qquad (561)
$$

tæthedsfunktionen er:

$$
f_{x,y}(x, y) = \frac{1}{2}\mathbb{1}_B(x, y) \qquad (562)
$$

Del 2) Find $P(Y > 2)$

Tegn på tegningen hvad det egentlig medfører. Altså på mængden B.

Først og fremmest ved vi at vi må integrere X ud af tætheden.

$$p_y(y) = \int_{\mathbb{R}} \frac{1}{2} \mathbb{1}_B(x,y)dx \tag{563}$$

Vi lavet et trick og skærer mængden B ud i to mængder M_1, M_2.

$$M_1 = \{x, y \mid 0 < x < 1, 0.5 + 2x < y < 2.5\} \tag{564}$$

$$M_2 = \{x, y \mid 0 < x < 1, 2.5 < y < 2.5 + 2x\} \tag{565}$$

$$p_Y(y) = \frac{1}{2} \int_{\mathbb{R}} \mathbb{1}_{M_1}(x,y)dx + \frac{1}{2} \int_{\mathbb{R}} \mathbb{1}_{M_2}(x,y)dx \tag{566}$$

Vi håndterer først M_1:

Vi ser at vi skal differentiere x ud. Mængden er er altså defineret i y-intervallet [0.5, 2.5]. Vi isolerer x som en funktion af y:

NOTE: Tegn diagrammet på tavlen og forklar intuitionen!

$$y = 0.5 + 2x \implies \frac{1}{2}(y - 0.5) = x \tag{567}$$

Hvor vi husker at: $y \in [0.5, 2.5]$

Vi kan nu finde at arealet for M_1 :

$$\int_0^{\frac{1}{2}(y-0.5)} 1dx = [x]_0^{\frac{1}{2}(y-0.5)} = \frac{1}{2}y - 0.25 \tag{568}$$

Analogt for M_2:

(KIG PÅ TAVLESKITSE)

$$y = 2.5 + 2x \implies \frac{1}{2}(y - 2.5) \tag{569}$$

$$\int_{\frac{1}{2}(y-2.5)}^1 1dx = [x]_{\frac{1}{2}(y-2.5)}^1 = 1 - \left(\frac{1}{2}y - 1.25\right) = 2.25 - \frac{1}{2}y \tag{570}$$

hvor vi husker at $y \in (2.5, 4.5]$

Vi opskriver $p_Y(y)$. Man husker at gange konstanten $\frac{1}{2}$ på.

$$p_Y(y) = \begin{cases} \frac{1}{2}\left(2.25 - \frac{1}{2}y\right) & , y \in (2.5, 4.5] \\ \frac{1}{2}\left(\frac{1}{2}y - 0.25\right) & , y \in [0.5, 2.5] \end{cases} \tag{571}$$

Vi kan opskrive $P(Y > 2) = 1 - P(Y \leq 2) = 1 - \int_{0.5}^{2} \frac{1}{2}\left(\frac{1}{2}y - 0.25\right) dy$

$$1 - \int_{0.5}^{2} \frac{1}{2}\left(\frac{1}{2}y - 0.25\right) dy = 1 - \frac{1}{4}\int_{0.5}^{2} y - 0.5dy \tag{572}$$

$$= 1 - \frac{1}{4}\left[\frac{1}{2}y^2 - 0.5y\right] \tag{573}$$

$$= 1 - \frac{1}{4}\left(\left(\frac{1}{2}2^2 - \frac{1}{2}\cdot 2\right) - \left(\frac{1}{2}0.5^2 - 0.5\cdot 0.5\right)\right) \tag{574}$$

$$= 1 - \frac{1}{4}(2-1) + \frac{1}{4}\left(\frac{1}{8} - \frac{1}{4}\right) \tag{575}$$

$$= 1 - \frac{1}{4} - \frac{1}{4}\frac{1}{8} \tag{576}$$

$$= 0.71875 \tag{577}$$

Del 4) Angiv den betingede fordeling af X givet $Y = 1$

Vi skal finde $p_{X|Y=1}(x)$

Vi kan altså bruge vores regel:

$$p_{X|Y}(x)p_Y(y) = p(x,y) \implies p_{X|Y} = \frac{p(x,y)}{p_Y(y)} \tag{578}$$

Vi ved at $Y = 1$. Vi bruger dette:

$$p_Y(1) = \frac{1}{2}\left(\frac{1}{2}(1) - 0.25\right) = \frac{1}{4} - \frac{1}{8} = \frac{1}{8} \tag{579}$$

Vi indsætter $Y = 1$ i den øverste del af brøken. Vi ved vi er i den nederste mængde M_1. Dette implicerer:

$$0.5 + 2x < y \wedge y = 1 \implies 0.5 + 2x < 1 \implies x < \frac{1}{4} \tag{580}$$

Vi kan herfra konkludere at:

$$p_{X|Y=1}(x) = \frac{1}{2}\frac{\mathbb{1}_{[0,0.25]}(x)}{1/8} = 4 \cdot \mathbb{1}_{[0,0.25]}(x) \tag{581}$$

Del 5) udregn forventede rating når $Y = 1$ og når $Y = 2$

Vi kender formlen for forventningen af en ligefordeling: $\mathrm{E}(x) = \frac{a+b}{2}$

Vi har svaret for $E(X \mid Y = 1) = \frac{0+0.25}{2} = \frac{1}{8}$

Vi skal nu analogt finde den betingede tæthed når $Y = 2$

$$p_Y(2) = \frac{1}{2}\left(\frac{1}{2}(2) - 0.25\right) = \frac{1}{2} - \frac{1}{8} = \frac{3}{8} \tag{582}$$

Vi ser igen på mængden M_2 :

$$0.5 + 2x < y \wedge y = 2 \implies 0.5 + 2x < 2 \implies x < \frac{3}{4} \tag{583}$$

Vi kan herfor konkluderer at den betingede fordeling for $X \mid Y = 2$ må være:

$$p_{X|Y=2}(x) = \frac{1}{2}\frac{\mathbb{1}_{[0,\frac{3}{4}]}(x)}{\frac{3}{8}} = \frac{4}{3} \cdot \mathbb{1}_{[0,\frac{3}{4}]} \tag{584}$$

Vi finder forventningen som må være:

$$\mathrm{E}(X \mid Y = 2) = \frac{1}{2}\frac{3}{4} = \frac{3}{8} \tag{585}$$

Del 6) Find variansen Var$(X \mid Y = 1$ og Var$(X \mid Y = 2)$

I stedet for at bruge hintet kigger vi på distributionen og bruger regnereglen for ligefordelinger:

$$Var(X) = \frac{1}{12}(a-b)^2 \tag{586}$$

$$\mathrm{Var}(X \mid Y=1) = \frac{1}{12}\left(0 - \frac{1}{4}\right)^2 = \frac{1}{16}\frac{1}{12} = \frac{1}{192} \tag{587}$$

$$\mathrm{Var}(X \mid Y=2) = \frac{1}{12}\left(0 - \frac{3}{4}\right)^2 = \frac{1}{12}\frac{9}{16} = \frac{9}{192} \tag{588}$$

Hvornår er den betingede varians størst - dvs. variansen af ratingen betinget på prisen

Kig på tegningen: Det rigtige svar må være $Y = 2500$.

Man overvejer følgende:

$$Var(X) = \frac{1}{12}(a-b)^2 \tag{589}$$

I intervallet $y \in [0.5, 2.5]$ ved vi at:

$$Var(X \mid Y = y) = \frac{1}{12}(0-a)^2 \tag{590}$$

Hvor at er øvre grænse:

$$0.5 + 2x < y \implies x = \frac{1}{2}(y - 0.5) \tag{591}$$

er monotont stigende med højere i y i intervallet $[0.5, 2.5)$.

Vi kan derfor sige at:

I intervallet $[0.5, 2.5)$ finder vi den højeste varians ved $Y = 2.5$.

Analogt kan man den højeste varians i i intervallet $[2.5, 4.5]$ til at være $Y = 2.5$

Illustrer på tavle!

3.14 Øvelse 14

01/11/2018, opgaver: 44.2.1, 44.2.2, 44.2.3

3.14.1 Opgave 44.2.1

- $U, V \sim N(0,1)$
- $U \perp\!\!\!\perp V$
- derudover er følgende variable defineret:

$$X = \alpha_1 + \beta_1 U \tag{592}$$
$$Y = \alpha_2 + \beta_2 U + \delta_2 V \tag{593}$$

Del 1) Udregn $P(0.1 < U < 0.5)$

Kig github! Gjort i python, så det er let tilgængeligt for alle online.

$$P(0.1 < U < 0.5) = 1 - F_U(0.1) - (1 - F_U(0.5)) = -F_U(0.1) + F_U(0.5) = 0.15163 \tag{594}$$

Del 2) Udregn $E(X), E(Y)$ **samt** $V(X), V(Y)$

$$\mathrm{E}(X) = \mathrm{E}(\alpha_1 + \beta_1 U) = \mathrm{E}(\alpha_1) + \beta_1 \mathrm{E}(U) = \alpha_1 \tag{595}$$

$$\mathrm{E}(Y) = \mathrm{E}(\alpha_2 + \beta_2 U + \delta_2 V) = \alpha_2 \tag{596}$$

Igen fordi at $\mathrm{E}(U) = \mathrm{E}(V) = 0$ eftersom vi har standard normalt fordelte stokastiske variable U, V

Variansen kan nu findes:

Man husker at standard normalt fordelte stokastiske variable har egenskaben: $\sigma^2 = \sigma = 1$.

$$\mathrm{Var}(X) = \mathrm{Var}(\alpha_1 + \beta_1 U) = \beta_1^2 \mathrm{Var}(U) = \beta_1^2 \tag{597}$$

$$\mathrm{Var}(Y) = \mathrm{Var}(\alpha_2 + \beta_2 U + \delta_2 V) = \mathrm{Var}(\beta_2 U + \delta_2 V) \tag{598}$$

Vi husker at når $V \perp\!\!\!\perp U \implies \mathrm{Var}(V + U) = \mathrm{Var}(V) + \mathrm{Var}(U)$

$$\begin{aligned}
\mathrm{Var}(\beta_2 U + \delta_2 V) &= \mathrm{Var}(\beta_2 U) + \mathrm{Var}(\delta_2 V) &(599)\\
&= \beta_2^2 \mathrm{Var}(U) + \delta^2 \mathrm{Var}(V) &(600)\\
&= \beta_2^2 + \delta_2^2 &(601)
\end{aligned}$$

Del 3) Udregn E$(X \cdot Y)$ **og Cov**(X, Y)

$$\begin{aligned}
\mathrm{E}(X \cdot Y) &= \mathrm{E}\left[(\alpha_1 + \beta_1 U)(\alpha_2 + \beta_2 U + \delta_2 V)\right] &(602)\\
&= \alpha_1\alpha_2 + \beta_1\beta_2 \mathrm{E}(U^2) &(603)\\
&= \alpha_1\alpha_2 + \beta_1\beta_2 &(604)
\end{aligned}$$

Dette er klart da $\mathrm{E}(V) = \mathrm{E}(U) = 0 \wedge V \perp\!\!\!\perp U \implies \mathrm{E}(V \cdot U) = 0$

Covariansen findes:

$$\mathrm{Cov}(X, Y) = \mathrm{E}(X \cdot Y)\mathrm{E}(X)\mathrm{E}(Y) = (\alpha_1\alpha_2 + \beta_1\beta_2) - (\alpha_1\alpha_2) = \beta_1\beta_2 \quad (605)$$

Del 4) Hvad skal α_1, β_1 **sættes til for at E**$(X) = 10,$ **Var**$(X) = 4$

Vi har tidligere fundet: $\mathrm{E}(X) = \alpha_1$, $\mathrm{Var}(X) = \beta_1^2$.

Vi kan derfor hurtigt konkluderer at: $\alpha_1 = 10 \implies \mathrm{E}(X) = 10$ og at $\beta_1^2 = 2 \implies \mathrm{Var}(X) = 4$

Del 5)

Find $\alpha_1, \alpha_2 + \beta_1\beta_2, \delta_2$ for at $\mathrm{Cov}(X, Y) = 4$

Vi kan hurtigt konkludere at: $\mathrm{Cov}(X, Y) = \beta_1\beta_2$, betyder at vi kan sige $\beta_1 = \beta_2 = 2 \implies \mathrm{Cov}(X, Y) = 4$. De resterende parametre kan sættes tilfældigt.

3.14.2 Opgave 44.2.2

- Y_1 Er timeløn i periode 1
- Y_2 Er timeløn i periode 2
- $Y_1 \sim N(\mu, \sigma^2)$

- $Y_2 \sim \alpha + \beta Y_1 + U$
- $U \sim N(0, v^2)$
- $Y \perp\!\!\!\perp U$

Del 1) lad $\mu = 350$ og $\sigma^2 = 12365$

- **Del A)** Udregn Ssh for timelønnen i periode 1 er højst 275
- **Del B)** Udregn Ssh for timelønnen i periode 1 er mindst 425

Kig i Github!

$$P(Y_1 < 275) = F_{Y_1}(275) = 0.25 \tag{606}$$

$$P(Y_1 > 425) = 1 - F_{Y_1}(425) \tag{607}$$

Del 2) Udregn middelværdi og varians af Y_2

$$\mathrm{E}(Y_2) = \mathrm{E}(\alpha + \beta Y_1 + U) = \mathrm{E}(\alpha) + \beta \mathrm{E}(Y_1) + \mathrm{E}(U) \tag{608}$$

Vi husker $\mathrm{E}(U) = 0$ (U var fordelt omkring 0). Derudover husker vi at $\mathrm{E}(Y_1) = \mu$

$$\mathrm{E}(Y_2) = \alpha + \beta\mu \tag{609}$$

Vi finder variansen.

$$\begin{aligned} \mathrm{Var}(Y_2) &= \mathrm{Var}(\alpha + \beta Y_1 + U) && (610) \\ &= \mathrm{Var}(\beta Y_1 + U) && (611) \\ &= \beta^2 \mathrm{Var}(Y_1) + \mathrm{Var}(U) && (612) \\ &= \beta^2\sigma^2 + v^2 && (613) \end{aligned}$$

Hvor vi har udnyttet at man kan splitte variansen op til en sum af to uafhængige stokastisk variable.

Del 3) er $Y_1 \perp\!\!\!\perp Y_2$

Nej! Dette kan ses uden yderligere udregninger, da Y_2 er en transformation af Y_1.

Note: Dette afhænger af den Strukturalle kausale model er *faithful*, hvilket vi kan se i dette tilfælde den er. (Læs mere i Jonas Peters bog om causality hvis man er interesseret)

Del 4) Angiv den betingede middelværdi og varians af Y_2 betinget på $Y_1 = y_1$

$$p_{Y_2|Y_1=y_1}(y_2) = \alpha + \beta y_1 + U \tag{614}$$

Vi kan hermed sige at:

$$\mathrm{E}(Y_2 \mid Y_1 = y_1) = \mathrm{E}(\alpha + \beta y_1 + U) = \alpha + \beta y_1 \tag{615}$$

Inden variansen findes, da overvej: $\mathrm{Var}(\beta y_1) = 0$. Da y_1 er en realisation of derfor ikke har noget stokastisk element.

Variansen er:

$$\mathrm{Var}(Y_2 \mid Y_1 = y_1) = \mathrm{Var}(\alpha + \beta y_1 + U) = \mathrm{Var}(U) = v^2 \tag{616}$$

Del 5)

Antagelser

- $\mu = 350$
- $\sigma^2 = 12365$
- $\alpha = 350(1 - \beta)$
- $v^2 = 12365(1 - \beta^2)$
- (ikke i opgaven - men nødvendigt) $|\beta| < 1$

Angiv de marginale fordelinger af Y_1 og Y_2

Man ser hurtigt at:

$$Y_1 := N(350, 12365) \tag{617}$$

Den marginale fordeling for Y_2:

$$
\begin{aligned}
Y_2 &:= \alpha + \beta Y_1 + U && (618)\\
&= 350(1-\beta) + \beta Y_1 + U && (619)\\
&= 350(1-\beta) + \beta N(350, 12365) + N(0, 12365(1-\beta^2)) && (620)\\
& && (621)
\end{aligned}
$$

Vi kan se at $\mathrm{E}(350(1-\beta) + \beta N(350, 12365) + N(0, 12365(1-\beta^2))) = 350$. Fra $\beta 350 + (1-\beta)350 = 350$

For variansen ser vi følgende: Husk $k^2\mathrm{Var}(X) = \mathrm{Var}(kX)$. Det vil sige: $\beta N(350, 12365) = N(350, \beta^2 12365)$. Variansen for 2 ukorrelerede stokastiske variable: $\mathrm{Var}(X) + \mathrm{Var}(Y) = \mathrm{Var}(X+Y)$. Hvilker betyder at variansen i dette tilfælde er:

$$
\mathrm{Var}(N(350, \beta^2)) + \mathrm{Var}(N(0, 12365(1-\beta^2)) = 12365(\beta^2 + (1-\beta^2) = 12365 \tag{622}
$$

Antag nu $\beta > 0$**. Vis at**$\mathbf{E}(Y_2 \mid Y_1 = 275) < \mathbf{E}(Y_2)$

Vi kan se at $\mathrm{E}(Y_2) = 350$. læs ovenfra:

Vi kan se at:

$$
\mathrm{E}(Y_2 \mid Y_1 = 275) = \mathrm{E}(350(1-\beta) + \beta 275 + N(0, 12365(1-\beta^2)) \tag{623}
$$

Hvilket betyder:

$$
\mathrm{E}(Y_2 \mid Y_1 = 275) = 350(1-\beta) + \beta 275 < 350 \tag{624}
$$

Del C)

Lad nu $\beta = 0.90$

NOGET GÅR GALT

- $P(Y_1 \leq 275, Y_2 \leq 275) \approx 0.193$
- $P(Y_1 \leq 275, Y_2 \leq 425) \approx 0.250$

Vi ved at $P(A,B) = P(A \mid B)P(B)$

Så vi kan sige:

$$P(Y_2 \leq 275 \mid Y_1 \leq 275) = \frac{P(Y_1 \leq 275, Y_2 \leq 275)}{P(Y_1 \leq 275)} = \frac{0.193}{0.25} = 0.772 \quad (625)$$

3.14.3 Opgave 44.2.3

- (X, Y) er en todimensionel stokastiske vektor
- $f_{X,Y} = 6\exp(-2x - 3y) \quad x, y \in [0, \infty)$
- $A = \{(x,y) : 0 \leq x + y < 1, x > 0, y > 0\}$

Del 1) Hvorfor er $X \perp\!\!\!\perp Y$

Lavet i en tidligere ugeseddel. Kan splittes op i to tæthedsfunktioner:

$$f_X(x) = 2\exp(-2x), \quad f_Y(y) = 3\exp(-3y) \quad (626)$$

Disse kan ganges sammen til $f_{X,Y}$

Del 2) Udregn $P((X,Y) \in A)$

Først tegn A: Kig github!

Note: $\int e^{bx}dx = \frac{e^{bx}}{b}$

Vi ser$x + y = 1 \implies y - 1 = x$. Og $y = x - 1$

$$\begin{aligned}
P((X,Y) \in A) &= \int_0^1 \int_0^{1-y} 6\exp(-2x - 3y)dxdy && (627)\\
&= \int_0^1 \int_0^{1-y} 6\exp(-2x)\exp(-3y)dxdy && (628)\\
&= 6\int_0^1 \exp(-3y) \int_0^{1-y} \exp(-2x)dxdy && (629)\\
&\dots && (630)\\
&= [-\exp(-2x)]_0^1 - 2\exp(-3)[\exp(x)]_0^1 && (631)\\
&= [1 - \exp(-2)] - 2\exp(-3)[\exp(1) - 1] && (632)\\
&\approx 0.6935 && (633)
\end{aligned}$$

Find tætheden for (X,Y) givet $(X,Y) \in A$

$$\begin{aligned} p(x, y \mid (X,Y) \in A) &= \frac{p(x,y,(X,Y) \in A)}{P((X,Y) \in A)} && (634) \\ &= \frac{p(x,y)\mathbb{1}_A(x,y)}{0.6935} && (635) \\ &= 6\exp(-2x-3y)/0.6935 \qquad 0 < x+y < 1 && (636) \end{aligned}$$

Kig Thomas rettevejledning for integrale show!

Del 4) Er $X \perp\!\!\!\perp Y \mid A$

Nej! A er ikke en produkt mængde!

Del 5) Find $\mathbf{E}(X \mid (X,Y) \in A$ **og ligeledes for** Y

3.15 Øvelse 15

5/11/2018, opgaver: U45.1, U45.2, U45.3, U45.4

3.15.1 Opgave U45.1

- $X \sim N(\mu, \sigma^2)$

Vis at $Y = \frac{1}{\sqrt{\sigma^2}}(X-\mu)$ **er standard normal fordelt** $N(0,1)$

Vi kan starte med at opskrive den parametriske form op for tætheden af en normalfordeling med mu og $sigma^2$ som henholdsvis middelværdi og varians

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad (637)$$

Vi husker et par regneregler:

Vi ved at $\mathrm{E}(X) = \mu$.

Det vil sige at:

$$\mathrm{E}[Y] = \mathrm{E}\left[\frac{1}{\sqrt{\sigma^2}}(X-\mu)\right] = \frac{1}{\sqrt{\sigma^2}}\mathrm{E}\left[(X-\mu)\right] \tag{638}$$

Som sagt: $\mathrm{E}(X) = \mu$.

$$\frac{1}{\sqrt{\sigma^2}}\mathrm{E}\left[(X-\mu)\right] = \frac{1}{\sqrt{\sigma^2}}(\mathrm{E}[X] - \mathrm{E}[\mu]) = \frac{1}{\sqrt{\sigma^2}}\left(\mu - \mu\right) = 0 = \mathrm{E}[Y] \tag{639}$$

Vi finder variansen. Man husker at $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$

Vi ved at $\mathrm{Var}(X) = \sigma^2$

$$\mathrm{Var}(Y) = \mathrm{Var}\left(\frac{1}{\sqrt{\sigma^2}}X\right) \tag{640}$$

Vi ignorerer μ da dette er en konstant

$$\mathrm{Var}\left(\frac{1}{\sqrt{\sigma^2}}X\right) = \frac{1}{\sigma^2}\mathrm{Var}(X) = \frac{1}{\sigma^2}\sigma^2 = 1 = \mathrm{Var}(Y) \tag{641}$$

Og vi har nu vidst at $Y \sim N(0,1)$

Del 2) (X,Y) **er en 2-dimensionel stokastisk vektor med middelværdi** mu **og covarians matrice** $= \Omega$

Vis at $Z = \frac{1}{\sqrt{\sigma_X^2}}(X - \mu_X)$ **er standard normalfordelt**

$$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \qquad \Omega = \begin{pmatrix} \sigma_Y^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_X^2 \end{pmatrix} \tag{642}$$

Samme argument som før!

Lad (X,Y) **være som i spørgsmål 2, men med** $\mu = (0,0)^T$**. Vis at** $Z = Y - \beta X$ **er normaltfordelt med** $N(0,\sigma^2)$

Hvor σ^2 er

$$\sigma^2 = \sigma_Y^2 - \beta\sigma_{XY} \tag{643}$$

og β er:

$$\beta = \sigma_{X,Y}/\sigma_X^2 \tag{644}$$

Vi husker at summen af to normalfordelte stokastiske variable er normalfordelt. Dvs: $Y - \beta X$ nødvendigvis må være normalfordelt!

Vi finder middelværdien først:

$$\mathrm{E}[Z] = \mathrm{E}[Y - \beta X] = \mathrm{E}[Y] - \beta \mathrm{E}[X] = 0 - \beta \cdot 0 = 0 \tag{645}$$

Nu finder vi variansen:

$$\mathrm{Var}(Z) = \mathrm{Var}(Y - \beta Z) \tag{646}$$

Vi husker der er covarians mellem X og Y.

Det vil sige:

$$\begin{align}
\mathrm{Var}(Z) &= \mathrm{Var}(Y - \beta x) \tag{647}\\
&= \mathrm{Var}(Y) + \beta^2 \mathrm{Var}(X) + 2\mathrm{Cov}(Y, -\beta X) \tag{648}\\
&= \mathrm{Var}(Y) + \beta^2 \mathrm{Var}(X) - 2\beta \mathrm{Cov}(Y, X) \tag{649}\\
&= \sigma_Y^2 + \beta^2 \sigma_X^2 - 2\beta \sigma_{X,Y} \tag{650}\\
&= \sigma_Y^2 + (\sigma_{X,Y}/\sigma_X^2)^2 \sigma_X^2 - 2(\sigma_{X,Y}/\sigma_X^2)\sigma_{X,Y} \tag{651}\\
&= \sigma_Y^2 + \frac{\sigma_{X,Y}^2}{\sigma_X^2} - 2\frac{\sigma_{X,Y}^2}{\sigma_X^2} \tag{652}\\
&= \sigma_Y^2 - \frac{\sigma_{X,Y}^2}{\sigma_X^2} \tag{653}
\end{align}$$

Hvilket var det vi ønskede at vise:

$$Z \sim N(0, \sigma^2) = N\left(0, \sigma_Y^2 - \frac{\sigma_{X,Y}^2}{\sigma_X^2}\right) \tag{654}$$

Del 4)

Vis at:

$$\mathrm{E}((Y - \beta X)X) = 0 \tag{655}$$

Hvilket betyder at Z og X er uafhængige!

$$\begin{aligned} \mathrm{E}((Y-\beta X)X) &= \mathrm{E}(YX-\beta X^2) && (656) \\ &= \mathrm{E}(YX)-\beta\mathrm{E}(X^2) && (657) \\ & && (658) \end{aligned}$$

Her ser man at $\mathrm{E}(X)=\mathrm{E}(Y)=0 \implies \mathrm{E}(XY)=\sigma_{XY}$ og at $\mathrm{E}(X)=0 \implies \mathrm{E}(X^2)=\sigma_X^2$

$$\begin{aligned} \mathrm{E}(YX)-\beta\mathrm{E}(X^2) &= \sigma_X Y-\beta\sigma_X^2 && (659) \\ &= \sigma_{XY}-(\sigma_{X,Y}/\sigma_X^2)\sigma_X^2 && (660) \\ &= 0 && (661) \end{aligned}$$

For at afgøre om de er uafhængige definerer vi først fejlleddet ϵ:

$$\epsilon = Y-\beta X \tag{662}$$

$$\mathrm{Cov}(\epsilon,X)=\mathrm{E}(\epsilon X)-\mathrm{E}(\epsilon)\mathrm{E}(X)=\mathrm{E}(\epsilon X)=0 \tag{663}$$

Det betyder at ϵ er ukorreleret med X. Og da både X og ϵ er normalfordelte kan vi konkludere de er uafhængige!

3.15.2 Opgave U45.2

Laves i klassen!

- stokastisk vektor (X,Y)
- distribueret med $N(m,\Omega)$

$$m=\begin{pmatrix}1\\0\end{pmatrix} \qquad \Omega=\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix} \tag{664}$$

Del 1) Hvad er E(X) og Var(X)

Vi kan direkte aflæse svarene $\mathrm{E}(X) = 1$ og $\mathrm{Var}(X) = 1$.

Del 2) Hvordan er Y fordelt:

Aflæses i m og Ω

$$Y \sim N(0,1) \tag{665}$$

Del 3) Hvad er Cov(X, Y)

Dette kan aflæses i kovarians-matricen off-diagonal elementer: $\sigma_{XY} = \rho$

Del 4) Hvad er E$(Y \mid X = x)$

Fra Rahbeks note (property G.3) finder vi formlen:

$$\mathrm{E}(Y \mid X = x) = \mu_{Y|X} = \mu_Y + \omega(x - \mu_X) \tag{666}$$

Hvor $\omega = \sigma_{YX}/\sigma_X^2$

Hvilket implicerer at:

$$\mathrm{E}(Y \mid X = x) = \mu_{Y|X} = \mu_Y + (\sigma_{YX}/\sigma_X^2)(x - \mu_X) \tag{667}$$

Vi finder de passende værdier i kovarians-matricen: $\mu_X = 1$, $\mu_Y = 0$, $\sigma_X^2 = 1$ og $\sigma_{XY} = \rho$

$$\mathrm{E}(Y \mid X = x) = 0 + \left(\frac{\rho}{1}\right)(x - 1) = \rho(x - 1) \tag{668}$$

Del 5) Hvad er E$(X \mid Y = y)$

Vi bruger samme formel som før, bare hvor: $\omega = \sigma_{YX}/\sigma_Y^2$

$$\mathrm{E}(X \mid Y = y) = \mu_x + \omega(y - \mu_Y) = 1 + \rho y \tag{669}$$

Del 6) Hvad er Var$(Y \mid X = x)$

Vi kigger igen i Rahbeks note (property G.3).

Finder formlen:

$$\mathrm{Var}(Y \mid X = x) = \sigma_Y^2 - \omega\sigma_{XY} = \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} \tag{670}$$

Og vi har stadig $\omega = \sigma_{YX}/\sigma_Y^2$. Altså samme ω som i del 5.

Vi har de passende værdier: $\sigma_X^2 = 1$, $\sigma_Y^2 = 1$ og $\sigma_{XY} = \rho$

$$\mathrm{Var}(Y \mid X = x) = \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} = 1 - \frac{\rho^2}{1} = 1 - \rho^2 \tag{671}$$

Del 7) Hvad gælder for (X, Y) hvis $\rho = 0.9$ og $\rho = 0$

Hvis $\rho = 0.9$ er X, Y stærkt positivt korrelerede. Det vil sige en høj realisering af Y implicerer en høj realisering X.

Omvendt $\rho = 0$ implicerer X og Y er ukorrelerede! Da de begge er normalt fordelte er de nødvendigvis også uafhængige!

3.15.3 Opgave U45.3

- (X, Y) er stokastisk vektor med som er normalfordelt med $N(\mu, \Omega)$

$$\mu = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \qquad \Omega = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \tag{672}$$

Opskriv tætheden $p(x, y)$

Kig formlen på s.237 Sørensen (ligning 8.3.6)

$$p(x,y) = \frac{1}{2\pi}\frac{1}{\sqrt{\det(\Omega)}} \exp\left(-\frac{1}{2}(x-\mu_X, y-\mu_Y)\Omega^{-1}\begin{pmatrix} x-\mu_X \\ y-\mu_Y \end{pmatrix}\right) \tag{673}$$

Hvis det ser uklart ud, da kan vi hurtigt lige definerer vektoren $K = (x - \mu_X, y - \mu_Y)$

Hvilket forsimpler utrykket til (Gøres for klargøre at det er en vektor K man "opløfter i 2"):

$$p(x,y) = \frac{1}{2\pi}\frac{1}{\sqrt{\det(\Omega)}} \exp\left(-\frac{1}{2}K\Omega^{-1}K^T\right) \tag{674}$$

Vi finder de to centrale ting:

$$\det(\Omega) = 1^2 - 0.5^2 = 0.75 \tag{675}$$

Kig i Sørensen s.237

$$\Omega^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix} = \begin{pmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{pmatrix} \tag{676}$$

Og vi har fundet tætheden!

3.15.4 Opgave U45.4

- Y er afkast på amerikansk aktie (Microsoft)
- X er et aktie-indeks (SP500)
- $Y := \beta X + \epsilon$
- $\epsilon \sim N(0, \sigma^2)$

Del 1) Fortolkning af β

β er relateret til kovariansen mellem X og Y og angiver altså samvariansen mellem en given aktie og hvordan hele markedet bevæger sig. En aktie med negativ β vil altså kunne reducerer volatiliteten i en portfølje da den er modsat korreleret med de andre aktier.

$$\beta = \frac{\sigma_{XY}}{\sigma_X^2} \tag{677}$$

Står beskrevet i Rahbeks Note s. 10

Del 2) En anden model blev repræsenteret

$$Y := \epsilon_Y, \qquad X := \epsilon_X \tag{678}$$

hvor $\epsilon_Y \sim N(0, \sigma_Y^2)$ og $\epsilon_X \sim N(0, \sigma_X^2)$

Forklar hvordan denne kan stemme overens med modellen præsenteret ovenfor

$$\mathrm{E}(Y) = 0 \tag{679}$$

Modellen ovenfor var implicit en betinget model for Y:

$$\begin{aligned} \mathrm{E}(Y \mid X) &= \mathrm{E}(\beta X + \epsilon \mid X) && (680) \\ &= \beta X && (681) \end{aligned}$$

Da $\mathrm{E}(\epsilon \mid X) = 0$

Altså så før så vi på en betinget model, men nu er det to marginale modeller, som er opstillet!

Man bruger altså i den betingede model information om hvordan en aktie samvarierer med markedet

---

3.16 Øvelse 16

9/11/2018, opgaver: 45.5, 45.6, 45.7

3.16.1 Opgave U45.5

Lav i klassen!

- (Y, X) er en 2-dimensionel stokastisk vektor
- (X, Y) er fordelt med $N(\mu, \Omega)$
- Vi ved at:
$$\mathrm{E}(Y \mid X) = X \qquad \mathrm{Var}(Y \mid X) = 1 \tag{682}$$
- derudover ved vi:
$$\mathrm{E}(X) = 0 \qquad \mathrm{Var}(X) = 1 \tag{683}$$

Vi kan først konkludere at:

husk at $\omega = \sigma_{YX}/\sigma_X^2$

$$
\begin{aligned}
\mathrm{E}(Y \mid X) &= \mu_Y + \omega(X - \mu_X) && (684)\\
&= \mathrm{E}(\mu_Y + \omega X) && (685)\\
&= \mathrm{E}(\mu_Y + (\sigma_{XY}/\sigma_X^2)X) && (686)\\
&= \mathrm{E}(\mu_Y + \sigma_{XY}X) && (687)\\
&= \mathrm{E}(\mu_Y) + \sigma_{XY}\mathrm{E}(X) && (688)
\end{aligned}
$$

Fra det kan vi konkludere at: $\mathrm{E}(\mu_Y) + \sigma_{XY}\mathrm{E}(X) = X \implies \mu_Y = 0$ og $\sigma_{XY} = 1$.

Vi udnytter dette:

$$
\begin{aligned}
\mathrm{Var}(Y \mid X) &= \sigma_Y^2 - \omega\sigma_{XY} && (689)\\
&= \sigma_Y^2 - \frac{\sigma_{YX}}{\sigma_X^2}\sigma_{XY} && (690)\\
&= \sigma_Y^2 - \frac{1^2}{1^2} && (691)\\
&= \sigma_Y^2 - 1 = 1 && (692)
\end{aligned}
$$

Hvorfra vi kan konkludere at $\sigma_Y^2 = 2$

Vi kan herfra opskrive den funktionelle form:

$$
\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \Omega = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \tag{693}
$$

3.16.2 Opgave U45.6

- $Z_1 \perp\!\!\!\perp Z_2$
- $Z_1, Z_2 \sim N(0,1)$
- VI har følgende to stokastiske variable:

$$
\begin{aligned}
Y &:= 2Z_1 + Z_2 && (694)\\
X &:= 3Z_1 && (695)
\end{aligned}
$$

Del 1) Hvordan er (Y, X) fordelt

Denne opgave følger eksemplet i Sørensen 8.3.3 meget tæt

Husk $\mathrm{E}(Z_1) = \mathrm{E}(Z_2) = 0$ og $\mathrm{Var}(Z_1) = \mathrm{Var}(Z_2) = 1$

Vi ser hurtigt:

$$\mathrm{E}(Y) = \mathrm{E}(2Z_1 + Z_2) = 2\mathrm{E}(Z_1) + \mathrm{E}(Z_2) = 0 + 2 \cdot 0 = 0 \tag{696}$$

$$\mathrm{E}(X) = \mathrm{E}(3Z_1) = 3\mathrm{E}(Z_1) = 0 \tag{697}$$

Variansen af de to er:

$$\mathrm{Var}(Y) = \mathrm{Var}(2Z_1 + Z_2) = 2^2\mathrm{Var}(Z_1) + \mathrm{Var}(Z_2) = 4 + 1 = 5 \tag{698}$$

$$\mathrm{Var}(X) = \mathrm{Var}(3Z_1) = 3^2\mathrm{Var}(Z_1) = 9 \tag{699}$$

Nu skal kovariansen findes mellem X og Y:

Kig formlen på s. 236 Sørensen. Her ser vi at:

$$\mathrm{Cov}(X, Y) = ab \tag{700}$$

hvor a og b er de konstanter der ganget på den stokastiske variabel som er går igen i udtrykket for henholdsvis X og Y. Altså i vores tildælde 2 for Y, og 3 for X.

$$\mathrm{Cov}(X, Y) = 2 \cdot 3 = 6 \tag{701}$$

Vi har med andre ord (Y, X) er distribueret med $N(\mu, \Omega)$, hvor:

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \Omega = \begin{pmatrix} 5 & 6 \\ 6 & 9 \end{pmatrix} \tag{702}$$

Del 2) Find $\mathbf{E}(Y \mid Z_1)$

$$\begin{aligned} \mathrm{E}(Y \mid Z_1) &= \mathrm{E}(2Z_1 + Z_2 \mid Z_1) && (703)\\ &= 2\mathrm{E}(Z_1 \mid Z_1) + \mathrm{E}(Z_2 \mid Z_1) && (704)\\ &= 2Z_1 && (705) \end{aligned}$$

Del 3) Find E$(X \mid Z_2)$

$$\begin{aligned} \mathrm{E}(X \mid Z_2) &= \mathrm{E}(3Z_1 \mid Z_2) && (706)\\ &= 0 && (707) \end{aligned}$$

Del 4) Find E$(Y \mid X)$

Husk $\omega = \sigma_{XY}/\sigma_X^2$

$$\begin{aligned} \mathrm{E}(Y \mid X) &= \mu_Y + \omega(X - \mu_X) && (708)\\ &= 0 + \frac{\sigma_{XY}}{\sigma_X^2}(X - 0) && (709)\\ &= \frac{6}{9}X = \frac{2}{3}X && (710) \end{aligned}$$

3.16.3 Opgave U45.7

- X er diskret ligefordelt på $\{-1, 0, 1\}$
- Y er kontinuært ligefordelt på intervallet $(-1, 1)$.

Del 1) Find E(X)**, Var**(X)**,** $P(X > 0)$

Find relevante formler på wiki

$$\mathrm{E}(X) = (a + b)/2 = (-1 + 1)/2 = 0 \quad (711)$$

$$\mathrm{Var}(X) = \frac{(b - a + 1)^2 - 1}{12} = \frac{(1 - (-1) + 1)^2 - 1}{12} = \frac{8}{12} = \frac{3}{4} \quad (712)$$

$$P(X > 0) = \frac{\#\ \text{Gunstige}}{\#\ \text{Mulige}} = \frac{1}{3} \quad (713)$$

Del 2) Find $\mathbf{E}(X \mid X > 0)$

$$\mathrm{E}(X \mid X > 0) = 1 \tag{714}$$

klart da $X \mid X > 0$ kun kan antage værdien 1.

Del 3) Find $\mathbf{E}(Y)$ og $\mathbf{Var}(Y)$ samt $P(Y > 0)$

$$\mathrm{E}(Y) = (a+b)/2 = (-1+1)/2 = 0 \tag{715}$$

$$\mathrm{Var}(Y) = \frac{(b-a)^2}{12} = \frac{((1-(-1))^2}{12} = \frac{4}{12} = \frac{1}{3} \tag{716}$$

$$P(Y > 0) = \int_0^1 \frac{1}{2}\mathbb{1}_{(-1,1)}(x)dx = \frac{1}{2}\left((1)-(0)\right) = \frac{1}{2} \tag{717}$$

Del 4) Find $\mathbf{E}(Y \mid Y > 0)$

husk $P(X,Y) = P(X \mid Y)P(Y)$ - hvor X, Y er arbitrære navne for at illustrere matematikken.

$$\mathrm{E}(Y \mid Y > 0) = \int_{-\infty}^{\infty} y\frac{P(Y, Y > 0)}{P(Y > 0)}dy \tag{718}$$

$$= \int_{-\infty}^{\infty} y\frac{\frac{1}{2}\mathbb{1}_{(-1,1)}(y)\mathbb{1}_{(0,1)}(y)}{\frac{1}{2}}dy \tag{719}$$

$$= \int_0^1 y\mathbb{1}_{(0,1)}(y) \tag{720}$$

$$= \left[\frac{1}{2}y^2\right]_0^1 = \frac{1}{2} \tag{721}$$

3.17 Øvelse 17

22/10/2018, opgaver: 1, 2, 3

3.17.1 Opgave 1

Kig do file do_1_17

3.17.2 Opgave 2

Del 1

Figur A: Histogram (og kernel density estimation). Viser tæthedsfunktionen (eller en approksimation).

Figur B: Viser den empiriske CDF.

Figur C: Q-Q plot er et plot der viser der modholder den empiriske distribution med en parametrisk - i dette tilfælde den gaussiske distribution. Dette er gjort ved at sammenligner quantiler.

Figur D: Boxplot - giver indblik i antal outliers samt hvordan de kvartiler, og median er fordelt.

Del 2

A) Medianen er den observation som er svarer til det punkt hvor $F(x) = 0.5$. Altså med andre ord $F^{-1}(0.5) = \text{median}$

B) Kig boxplot. Ja det synes der at være. Det er dog altid svært at vurdere outliers.

C) Ja, vi ser at fordelingen er centreret omkring en middelværdi og er stort set symmetrisk og unimodal.

D) Kig CDF. Omkring halvdelen af alle firmaerne.

C) 10% fraktilen (hvilket kaldes 10% percentilen). Angiver det punkt hvor $F(x) = 0.1$. I vores konkrete tilfælde cirka -500

E) Vi kan se på Q-Q plottet at distribution er lidt lang i halerne, men det er ikke klart om den er venstre eller højre skæv. Derudover er det ikke klart om disse afvigelser i halerne er nok, til at antage den skulle være venstre skæv eller højre skæv.

3.17.3 Opgave 3

Del 1

gns_gym: Man finder at gennemsnittet fra gymnasiet er kontinuær.

studietimer: Man finder at antal timer brugt på studie er tælle data.

3.18 Øvelse 19

19/11/2018, opgaver: 1, 2

3.18.1 Opgave 1

Lav i klassen!

Del 1) Opskriv udfaldsrummet $\mathbb{Y}$

Vi ser hurtigt at en Y_i må være defineret på de positive naturlige tal: $\mathbb{N}_+$ inkusiv 0.

Vi ved parameter rummet Θ må have følgende restriktioner: i poisson fordelingen er middelværdi og varians begge λ. Da vi ved at variansen er positive må λ være defineret på den reello postive akse: dvs $\Theta = \mathbb{R}_+$

Del 2)

Kig side 62. eksempel

sandsynlighedsfunktion

$$l(\lambda \mid y_i) = f_{Y_i}(y_i \mid \lambda) = \frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}, \qquad y_i \in \mathbb{N} \tag{722}$$

Hvis man har mere information om branchen, ville man skulle betinge på det.

Del 3)

Man husker at uafhængighed har implikationen:

$$f(x,y) = f(x)f(y) \tag{723}$$

Vi opskriver den samlede sandsynlighedsfunktion!

$$f_{Y_1,Y_2\cdots Y_n}(y_1, y_2, \cdots, y_n \mid \lambda) = \prod_{i=1}^{n} \frac{\exp(-\lambda)\lambda^{y_i}}{y_i!} \tag{724}$$

Del 4)

Igen kig side 62

$$L(\lambda \mid y_1, y_2 \cdots y_n) = \prod_{i=1}^{n} \frac{\exp(-\lambda)\lambda^{y_i}}{y_i!} \tag{725}$$

Nu opskrives log-likelihood funktionen:

$$\log L(\lambda \mid y_1, y_2 \cdots y_n) = \sum_{i=1}^{n} y_i \log(\lambda) - \lambda - \log(y_i!) \tag{726}$$

Dette kan igen om skrives til

$$= \log(\lambda) \sum_{i=1}^{n} y_i - \lambda n - \sum_{i=1}^{n} log(y_i!) \tag{727}$$

Forskellen mellem y_i og Y_i er om vi overvejer problem som værende realisationer eller stokastiske variable. DVS. vores endelig estimat $\hat{\theta}$ er enten et tal (realisation) eller en stokastisk variabel (stokastiske variable).

Del 5) Find første ordens betingelsen

$$\frac{\partial}{\partial \lambda} L(\lambda \mid y_1, y_2, \cdots, y_n) = \frac{1}{\lambda} \sum_{i=1}^{n} y_i - n \tag{728}$$

Del 6) Løs of find $\hat{\lambda}$

Kig side 69.

Vi ved (eller antager) at vi har et har at gøre med et konvekst optimerings problem. Vi finder maksimum ved at sætte:

$$\frac{1}{\hat{\lambda}}\sum_{i=1}^{n} y_i - n = 0 \tag{729}$$

$$\implies \quad \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{730}$$

Del 7) Find 2. ordens betingelsen og argumenter for det er et maksimum

$$H(\lambda) = \frac{-\sum_{i=1}^{n} y_i}{\lambda^2} \tag{731}$$

som er negativt, hvilket betyder vi har vist at et hvert ekstremum må være et unikt ekstremum.

Del 8)

Vi har i opgave teksten oplyst at $\sum_{i=1}^{n} y_i = 63$

Vi bruger 731 og indsætter summen af y_i:

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{21}63 = 3 \tag{732}$$

Vi har altså fundet estimatet!

Del 9) Hvad hvis summen var 0?

Så ville $\lambda = 0$, hvilket ikke ville være i overensstemelse med vores parameter rum Θ, som kræver at $\lambda > 0$.

3.18.2 Opgave 2

- $Y_i \in \mathbb{Y} = \{y \in \mathbb{R} \mid y > 0\}$
- $Y_i \sim Exponential(\theta)$
- $f_{Y_i}(y \mid \theta) = \theta \exp(-\theta y)$

Del 1) Hvad er Y

Y_i er en stokastisk variabel hvor y_i er en realisation. Table 3.1 på side 55, beskriver forskellen. Det betyder også om vi ender med et estimat (et tal). Eller en Estimator som er en stokastisk variabel.

y indgår i tæthedsfunktionen og relaterer til hvordan sandsynlighedsmassen er ved punktet y. Man kan ikke sige at det er sandsynligheden da punktsandsynligheden er 0 - dvs: $P(Y = y) = 0$.

Del 2)

Man kan se at θ ikke er på individ niveau. Altså vi antager at $\theta_i = \theta$. Og altså at alle stokastiske variable er identisk distribueret.

Dette er vigtigt, da ellers ville man ikke kunne lave maksimerering. Vi antager nemlig at vi kan finde et parameter θ. Hvis de ikke var identisk fordelt skulle vi finde et parameter for hvert enkelt stokastisk variabel.

Del 3) Opskrive sample likelihood funktionen

Vi opskriver likelihood contribution:

$$l(\theta \mid y_i) = f_{Y_i} = (y_i \mid \theta) = \theta \exp(-\theta y_i) \tag{733}$$

Vi kan nu opskrive sample likelihood funktionen:

$$L(\theta \mid y_1, \cdots, y_n) = \prod_{i=1}^{n} l(\theta \mid y_i) = \prod_{i=1}^{n} \theta \exp(-\theta y_i) \tag{734}$$

Vi opskriver nu log-likelihood funktionen:

$$\log L(\theta \mid y_1, \cdots, y_n) = \sum_{i=1}^{n} \log(\theta) - \theta y_i = n \log(\theta) - \theta \sum_{i=1}^{n} y_i \tag{735}$$

OPSKRIV OGSÅ MED Y_i altså så vi modellerer med stokastiske variable i stedet for realisationer.

Dette vil have implikationen at vi fik en estimator til sidst i stedet for et estimat.

Del 4) HVorfor er uafhængighed en vigtig antagelse.

Uden uafhængighed kunne man ikke skrive den simultane tæthed op som produktet af marginale tætheder. Dette er kun muligt under antagelse af uafhængighed. Derfor meget vigtigt!

Del 5) Find maximum likelihood estimatoren

Spørgsmål til klassen: " Skal vi nu bruge store Y eller lille y"

Svar: store Y

$$S(\theta) = \frac{\partial}{\partial\theta} = n\frac{1}{\theta} - \sum_{i=1}^{n} Y_i \tag{736}$$

Vi finder nu estimatoren:

$$n\frac{1}{\hat{\theta}} - \sum_{i=1}^{n} Y_i = 0 \implies \frac{n}{\sum_{i=1}^{n} Y_i} = \hat{\theta} \tag{737}$$

Del 6) Find maximum likelihood estimatet

Vi ved at $\sum_{i=1}^{n} y_i = 252$. $n = 120$

Nu finder vi et estimat $\implies$ vi går fra lille y til store Y.

$$\hat{\theta} = \frac{120}{252} = 0,476190 \tag{738}$$

Del 7) Forskellen på estimat og estimator

Estimatet er et tal! Estimator er en stokastiske variabel

Hvis man læser om et estimat på $\theta = 0.2$ så snakkes der om et estimat.

En estimator kan have lille varians.

3.19 Øvelse 20

19/11/2018, opgaver: 3, 4, 5

3.19.1 Opgave 3

- Pærerne fra opgave 2 er ikke identiske fordelt! Deres distribution er rigtigt:

$$Y_i = \begin{cases} exponential(\theta_1) & i \in \{1, 2, 3, \dots, 75\} \\ exponential(\theta_2) & \in \{76, 77, \dots, 120\} \end{cases} \tag{739}$$

- $Y_i \perp\!\!\!\perp Y_j$, for $i \neq j$

Parameterrummet:

$$\theta = (\theta_1, \theta_2) \in \Theta \tag{740}$$

Hvor at $\Theta = (0, \infty)^2$

Vi har likelood sample funktionen:

$$L_n(\theta \mid y_i) = \prod_{i=1}^{75} \theta_1 \exp(-\theta_1 y_i) \prod_{i=76}^{120} \theta_2 \exp(-\theta_2 y_i) \tag{741}$$

Opskriv log-likelihood funktionen for datasættet $\{y_i\}_{i=1}^n$

Vi tager logaritmen til udtrykket ovenfor:

$$\log L_n(\theta \mid y_i) = 75 \log(\theta_1) - \theta_1 \sum_{i=1}^{75} y_i + (120 - 75) \log(\theta_2) - \theta_2 \sum_{i=76}^{120} y_i \tag{742}$$

Del 3) Find maximum likelihood estimatoren $\hat{\theta}(Y_1, \dots, 120)$

Vi tager udgangspunkt i store Y nu, da vi skal finde estimatoren. Vi finder første ordens betingelserne:. θ_1, θ_2

$$S(\hat{\theta_1}) = \frac{\partial}{\partial \theta_1} \log L_n(\theta \mid Y_i) = \frac{75}{\hat{\theta_1}} - \sum_{i=1}^{75} Y_i = 0 \tag{743}$$

Hvilket løst for θ_1 bliver:

$$\hat{\theta_1} = \frac{75}{\sum_{i=1}^{75} Y_i} \tag{744}$$

Ligeledes ville det for θ_2 være:

$$\hat{\theta}_2 = \frac{(120 - 75)}{\sum_{i=76}^{120} Y_i} \tag{745}$$

Del 4) Vi får nu fortalt at: $\sum_{i=1}^{75} y_i = 160$ og $\sum_{i=76}^{120} y_i = 92$

Vi indsætter i estimatoren og finder estimatet (estimaterne):

$$\theta = (\theta_1, \theta_2) = (0.468, 0.489) \tag{746}$$

Spørg i klassen: Er de nye pærer bedre end de gamle?

Del 5) Forklar hvordan man fortolke dette som en betinget model

Man kan nu betinge levetiden på om det er en ny eller gammel pære! Hvis man har information om dette, kan man sige mere præcist noget om forventet levetid af pæreren.

3.19.2 Opgave 4

- Fan chart for inflation fra bank of England.

Del 1) Er fordelingen skewed. Hvad kan skyldes dette?

Man kan forestille sig at historisk har man udregnet momenterne og fundet fordelingen skewed opad. Det vil sige at engang imellem har man historisk oplevet meget høj inflation, men sjældent meget lav inflation.

Uddyb mere i klassen? Spørg klassen.

Del 2) Næste periode (dvs kvartal 4. 2004) er fordelt $N(\mu = 2.6, \sigma^2 = 0.56)$

Kig github.

3.19.3 Opgave 5

- $\mathbb{Y} = \{0, 1, 2, 3\}$

- $$P(Y_i = y_i) = \begin{cases} \frac{2\theta}{3} & y_i = 0 \\ \frac{\theta}{3} & y_i = 1 \\ \frac{2(1-\theta)}{3} & y_i = 2 \\ \frac{1-\theta}{3} & y_i = 3 \end{cases} \tag{747}$$

- θ er mellem 0 og 1

Del 1) Vis sandsynlighedsfunktionen kan skrives som nedenfor

$$f_{Y_i}(y \mid \theta) = \left(\frac{2}{\theta}\right)^{\mathbb{1}(y_i=0)} \left(\frac{\theta}{3}\right)^{\mathbb{1}(y_i=1)} \left(\frac{2(1-\theta)}{3}\right)^{\mathbb{1}(y_i=2)} \left(\frac{1-\theta}{3}\right)^{\mathbb{1}(y_i=3)} \tag{748}$$

Dette er klar hvis man har et udtryk:

$$(\cdot)^{\mathbb{1}(\text{condition})} \tag{749}$$

Hvis betingelsen er sand får man udtrykket i parantesen ellers får man 1, da ethvert udtryk opløftet i 0 er 1.

Uddyb eksempel på tavlen.

Opskriv sample likelihood funktionen og loglikelihood funktionen

VI ser at likelikehood bidraget er:

$$l(\theta \mid y_i) = f_{Y_i}(y_i \mid \theta) \tag{750}$$

Og vi har at sample likelihood funktionen er:

$$L(\theta \mid y_i) = \prod_{i=1}^{n} l(\theta \mid y_i) \tag{751}$$

$$L(\theta \mid y_i) = \prod_{i=1}^{n} \left(\frac{2}{\theta}\right)^{\mathbb{1}(y_i=0)} \left(\frac{\theta}{3}\right)^{\mathbb{1}(y_i=1)} \left(\frac{2(1-\theta)}{3}\right)^{\mathbb{1}(y_i=2)} \left(\frac{1-\theta}{3}\right)^{\mathbb{1}(y_i=3)} \tag{752}$$

vi tager logaritmen:

$$\log L(\theta \mid y_i) = \sum_{i=1}^{n} \mathbb{1}(y_i = 0) \log\left(\frac{2}{\theta}\right) + \mathbb{1}(y_i = 1) \log\left(\frac{\theta}{3}\right) \tag{753}$$

$$+ \mathbb{1}(y_i = 2) \log\left(\frac{2(1-\theta)}{3}\right) + \mathbb{1}(y_i = 3) \log\left(\frac{1-\theta}{3}\right) \tag{754}$$

Vi kan transformere summen, da vi indser vi har:

- 2 observationer = 0
- 3 observationer = 1
- 3 observationer = 2
- 2 observationer = 3

$$\log L(\theta \mid y_i) = 2 \log\left(\frac{2}{\theta}\right) + 3 \log\left(\frac{\theta}{3}\right) \tag{755}$$

$$+ 3 \log\left(\frac{2(1-\theta)}{3}\right) + 2 \log\left(\frac{1-\theta}{3}\right) \tag{756}$$

Dette kan omskrives til:

$$\log L(\theta \mid y_1, y_2 \cdots y_n) = 5 \left[\log\left(\frac{2}{9}\right) + \log(\theta - \theta^2)\right] \tag{757}$$

(Folk kan lave omskrivningen i klassen) - Hint dem til Frederiks rettevejledning.

Del 3) Tegn $\log L_n$ som funktion af θ

Kig github.

Find maximum likelihood estimatet

$$S(\hat{\theta}) = \frac{\partial}{\partial \hat{\theta}} \log L_n(\hat{\theta} \mid y_i) = 5 \frac{1}{\theta - \theta^2}(1 - 2\theta) = 0 \implies \hat{\theta} = 0.5 \tag{758}$$

Vi har med givne $\hat{\theta} = 5$:

$$P(Y_i = y_i) = \begin{cases} \frac{2 \cdot 0.5}{3} = 1/3 & y_i = 0 \\ \frac{0.5}{3} = 1/6 & y_i = 1 \\ \frac{2(1-0.5)}{3} = 1/3 & y_i = 2 \\ \frac{1-0.5}{3} = 1/6 & y_i = 3 \end{cases} \tag{759}$$

Observeret:

2/10, 3/10, 3/10, 2/10

Men vi har et småt sample. Det er meget svært at lave konklusioner!

3.20 Øvelse 21

26/11/2018, Opgaver: 1, 2, 3

3.20.1 Opgave 1

Lav i klassen!

Kig do-files

3.20.2 Opgave 2

- $n = 573$
- 121 har $y = 1$
- 251 har $y = 2$
- 201 har $y = 3$
- Vi betegner de absolutte frekvenser som s_j:

$$s_j = \sum_{i=1}^{n} \mathbb{1}(y_i = j) \tag{760}$$

- Vi ser at $s_1 + s_2 + s_3 = n$

- Opstiller en statistik model:

$$P(Y_i = y) = \begin{cases} p_1 & \text{hvis } y = 1 \\ p_2 & \text{hvis } y = 2 \\ p_3 & \text{hvis } y = 3 \end{cases} \tag{761}$$

Del 1) Hvis at sandsynlighedsunktionen kan skrives som:

$$f_{Y_i}(y \mid p_1, p_2, p_3) = p_1^{\mathbb{1}(y=1)} p_2^{\mathbb{1}(y=2)} p_3^{\mathbb{1}(y=3)} \tag{762}$$

Minder om opgaven til sidste uge:

Man indser først at vi har:

$$S = (\text{expression})^{\mathbb{1}(\text{condition})} \tag{763}$$

Hvis betingelsen (condition) er sand da må $S = (\text{expression})$. Hvis condition er falsk: $S = (\text{expression})^0 = 1$.

Hvormed det er klart: hvis eksempelvis $(y = 1$

$$f_{Y_i}(1 \mid p_1, p_2, p_3) = p_1 \cdot 1 \cdot 1 \tag{764}$$

Hvilket er det som modellen skulle overholde!

Del 2) Opskriv likelihood bidraget samt sample likelihood funktionen

$$l(p_1, p_2, p_3 \mid y_i) = p_1^{\mathbb{1}(y_i=1)} p_2^{\mathbb{1}(y_i=2)} p_3^{\mathbb{1}(y_i=3)} \tag{765}$$

Vi ser nu sample likelihood funktionen kan skrives op:

$$\begin{align} L(p_1, p_2, p_3 \mid y_1, y_2, \cdots, y_n) &= \prod_{i=1}^{n} l(p_1, p_2, p_3 \mid y_i) \tag{766} \\ &= p_1^{s_1} p_2^{s_2} p_3^{s_3} \tag{767} \end{align}$$

Overvej dette:

for $y_i = 1$ vi har at $l(p_1, p_2, p_3 \mid y_i) = p_1$. Vi har s_1 at sådanne tilfælde og vi kan konkludere at vi må have: $p_1 \cdot p_1 \cdots p_1$ i alt s_1 gange, som vil svare til $p_1^{s_1}$

Del 3) Opskriv θ, og log likelihood funktionen

Først hvor mange frie parametre har vi?

Vi ved at $p_1 + p_2 + p_3 = 1$. Dette må implicere at $p_3 = 1 - p_1 - p_2$. Vi har altså 2 frie parametre:

$$\theta = (\theta_1, \theta_2) \in \Theta \tag{768}$$

hvor at $\Theta = [0, 1] \times [0, 1]$

Vi opskriver likelihood funktionen som funktion af de frie parametre θ_1, θ_2:

$$L(\theta_1, \theta_2 \mid y_1, y_2, \cdots, y_n) = \theta_1^{s_1} \theta_2^{s_2} (1 - \theta_1 - \theta_2)^{s_3} \tag{769}$$

Vi kan opskrive Log likelihood funktionen

$$\log L(\theta_1, \theta_2 \mid y_1, y_2, \cdots, y_n) = s_1 \log(\theta_1) + s_2 \log(\theta_2) + s_3 \log(1 - \theta_1 - \theta_2) \tag{770}$$

Del 4) Angiv antagelser

Vi har antaget at vores stokastiske variable er *i.i.d.* Altså de er identiske, uafhængigt distribueret. Uafhængigheden gør at man kan skrive det op som et produkt, og identiske betyder at $\theta_i = \theta$ for alle Y_i.

Antagelserne er nok rimelige, da vi ikke har betinget på størrelsen af familen og ikke har informationen. Man altså man ville nok ikke forestille sig den betingede ssh for en familie med en person har 2 biler er den samme som en familie med 2 personer har 2 biler.

Del 5) Find estimator og estimat

Vi finder estimatoren.

FOC:

$$\frac{\partial}{\partial \theta_1} \log L(\theta_1, \theta_2 \mid Y_1, Y_2, \cdots, Y_n) = 0 \tag{771}$$

$$\frac{\partial}{\partial \theta_2} \log L(\theta_1, \theta_2 \mid Y_1, Y_2, \cdots, Y_n) = 0 \tag{772}$$

Notér: Vi bruger store Y da vi skal finde estimatoren!

Vi ser at for θ_1 får vi:

noter: $\frac{\partial}{\partial \theta_1} log(1 - \theta_1 - \theta_2) = -\frac{1}{1-\theta_1-\theta_2}$

$$\frac{s_1}{\hat{\theta}_1} - \frac{s_3}{1 - \hat{\theta}_1 - \hat{\theta}_2} = 0 \tag{773}$$

Ligeledes finder vi:

$$\frac{s_2}{\hat{\theta}_2} - \frac{s_3}{1 - \hat{\theta}_1 - \hat{\theta}_2} = 0 \tag{774}$$

Hvilket betyder at:

$$\frac{s_2}{\hat{\theta}_2} = \frac{s_1}{\hat{\theta}_1} \tag{775}$$

Vi isolerer $\hat{\theta}_2 = \frac{\hat{\theta}_1 s_2}{s_1}$

Vi indsætter så vi har:

$$\frac{s_1}{\hat{\theta}_1} - \frac{s_3}{1 - \hat{\theta}_1 - \frac{\hat{\theta}_1 s_2}{s_1}} = 0 \tag{776}$$

$$\frac{s_1}{\hat{\theta}_1} = \frac{s_3}{1 - \hat{\theta}_1(1 + s_2/s_1)} \tag{777}$$

Herfra kan man yderligere isolere således at man finder:

$$\hat{\theta}_1 = \frac{s_1}{s_1 + s_2 + s_3} = \frac{s_1}{n} \tag{778}$$

Ligeledes ville man kunne finde for θ_2.

Vi kan nu finde estimatet:

$$\hat{\theta}_1 = \frac{s_1}{n} = \frac{121}{573} = 0.211 \tag{779}$$

$$\hat{\theta}_2 = \frac{s_2}{n} = \frac{251}{573} = 0.438 \tag{780}$$

Og vi kan finde $p_3 = 1 - \theta_1 - \theta_2 = 1 - 0.211 - 0.438 = 0.351$

3.20.3 Opgave 3

- Vi har en rebproducent
- producerer 2 m i minuttet
- Standard afvigelse på 10 cm
- lad rebproduktionen være beskrevet ved Y_i, hvor $i \in \{1, 2, \cdots, n\}$
- Den gennemsnitlige reblængde på n minutter er derfor:

$$X_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{781}$$

- Den samlede reblængde er:

$$nX_n = \sum_{i=1}^{n} Y_i \tag{782}$$

Del 1) Brug den centrale grænseværdisætning til at karakterisere den approksimative fordeling af den gennemsnitlige reb-længde, når n bliver stor

Vi husker fra sørensen at:

$$U_n = \frac{X_n - \mu}{\sigma/\sqrt{n}} \tag{783}$$

Og vi ved at U_n vil være standard normalt fordelt for $n \to \infty$

Det betyder også at vi kunne sige at:

$$X_n \sim N(200, 100/n) \tag{784}$$

hvor vi har at $\sigma = 10$ og at $\sigma^2 = 10^2 = 100$.

Transformation af U_n er klar:

$$U_n = \frac{X_n - \mu}{\sigma} \tag{785}$$
$$= \frac{X_n - \mu}{\sqrt{\sigma^2/n}} \tag{786}$$
$$\implies \quad \sqrt{\sigma^2/n} \cdot U_n = X_n - \mu \tag{787}$$
$$\implies \quad \sqrt{\sigma^2/n} \cdot U_n + \mu = X_n \tag{788}$$

Vi indser at:

$$\sqrt{\sigma^2/n} \cdot U_n \sim N(0, \sigma^2/n) \tag{789}$$

og herfra trækker man bare middelværdien fra!

Del 2) Brug den asymptotiske distribution til at udregne ssh for at der på en time produceres 125 meter.

vi kan altså herfra sige:

$$125 \cdot 100/60 = 208,33 \tag{790}$$

Vi kan transformere dette!

$$\frac{208.33 - 200}{10/\sqrt{60}} = 8.33/1.2909 = 6.4528 \tag{791}$$

Vi kan nu spørge:

$$P(U_n > 6.45) = 1 - \Phi(6.45) = 1 - 0.99999999 = 0 \tag{792}$$

hvor $\Phi(\cdot)$ er fordelingsfunktionen CDF'en for en standard normalfordeling.

Del 3) Hvad hvis det var cauchy fordelingen?

Så nej! Denne fordeling har ingen momenter!

3.21 Øvelse 22

30/11/2018, Opgaver: 4, 5

3.21.1 Opgave 4

side 65 i Heino's bog

- $\{Y_i\}_{i=1}^{n}$ er i.i.d. normalt fordelte stokastiske variable
- altså:
$$Y_i \sim N(\mu, \sigma^2) \tag{793}$$
- tæthedsfunktionen:
$$f_{Y_i}(y \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right) \tag{794}$$

Del 1.a) Opskriv likelihood funktionen $L(\mu, \sigma^2 \mid Y_1, Y_2 \cdots Y_n)$

likelihood contribution:

$$l(\mu, \sigma^2 \mid Y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(Y_i-\mu)^2}{2\sigma^2}\right) \tag{795}$$

Likelihood funktionen:

$$L(\mu, \sigma^2 \mid Y_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(Y_i-\mu)^2}{2\sigma^2}\right) \tag{796}$$

Del 1.b) Opskriv log likelihood funktionen

Vi tager logaritmen:

$$\log L(\mu, \sigma^2 \mid Y_1, Y_2, \cdots Y_n) = \sum_{i=1}^{n} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) + \frac{-(Y_i-\mu)^2}{2\pi^2} \tag{797}$$

Del 1.c) Find estimatorerne

Kig side 70.

$$\frac{\partial}{\partial \mu} \log L(\mu, \sigma) = \sum_{i=1}^{n} \frac{Y_i - \mu}{\sigma^2} \tag{798}$$

$$-\frac{1}{2\sigma^2} + \frac{(Y_i - \mu)^2}{2\sigma^4} \tag{799}$$

Sæt disse lig med 0 og vi kan se at:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} Y_i \tag{800}$$

$$\hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\mu})^2 \tag{801}$$

Dette er de samme formler som blev præsenteret i starten af bogen for det empiriske gennemsnit og varians.

Del 2) Find estimaterne

Vi har givet at $\sum_{i=1}^{n} y_i = 627.6$ og $\sum_{i=1}^{n} y_i^2 = 3807.2$. v har at $n = 125$

Vi bruger hintet givet i opgaven:

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{\mu})^2 = \sum_{i=1}^{n} y_i^2 + n\hat{\mu}^2 - 2\hat{\mu} \sum_{i=1}^{n} y_i \tag{802}$$

Vi finder først estimatet for μ

$$\hat{\mu} = \frac{1}{125} \cdot 627.6 = 5.0208 \tag{803}$$

Herefter finder vi det for σ^2

note: Skriv måske formlen op igen for variansen!

$$\hat{\sigma^2} = 3807.2 + 125 \cdot 5.0208^2 - 2 \cdot 5.0208 \cdot 627.6 \tag{804}$$

Likelihood og Log Likelihood (under equi-dispersion $\mu = \sigma^2 = \phi$)

Dette tilfælde kaldes *equi dispersion.*

tætheden er givet ved:

$$f_{Y_i}(y \mid \phi) = \frac{1}{\sqrt{2\pi\phi}} \exp\left(\frac{-(y-\phi)^2}{2\phi}\right) \tag{805}$$

Vi opskriver Likelihood funktionen:

$$L(\phi \mid Y_1, Y_2 \cdots Y_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\phi}} \exp\left(\frac{-(Y_i-\phi)^2}{2\phi}\right) \tag{806}$$

Vi opskriver også log likelihood funktionen

$$\log L(\phi) = \sum_{i=1}^{n} \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\phi) + \left(-\frac{(Y_i-\phi)^2}{2\phi}\right) \tag{807}$$

Som kan omskrives til:

$$\log L(\phi) = \frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\phi) - \frac{\sum_{i=1}^{n}(Y_i-\phi)^2}{2\phi} \tag{808}$$

Del 4) Udled scoren

$$S(\phi) = \frac{\partial}{\partial\phi}\log L(\phi \mid Y_1, \cdots, Y_n) = \frac{\sum_{i=1}^{n} Y_i - n\phi^2 - n\phi}{2\phi^2} \tag{809}$$

Man ser man kan brække udtrykket (fra før) over i to dele når man differentiere:

Den første del $\frac{n}{2}\log(\phi)$

$$\frac{\partial}{\partial\phi}\left(\frac{n}{2}\log(\phi)\right) = \frac{n}{2\phi} \tag{810}$$

Den anden del:

$$\frac{\partial}{\partial\phi}\left(-\frac{\sum_{i=1}^{n}(Y_i-\phi)^2}{2\phi}\right) = -\sum_{i=1}^{n}\frac{\partial}{\partial\phi}\left(\frac{(Y_i-\phi)^2}{2\phi}\right) \tag{811}$$

Dette skrives som (husk hint i første skridt - differentiation af brøker) :

$$-\sum_{i=1}^{n} \frac{\partial}{\partial \phi} \left(\frac{(Y_i - \phi)^2}{2\phi} \right) = -\sum_{i=1}^{n} \frac{(-2(Y_i - \phi)2\phi) - (2(Y_i - \phi)^2)}{4\phi^2} \tag{812}$$

$$= -\sum_{i=1}^{n} \frac{-4Y_i\phi + 4\phi^2 - 2Y_i^2 - 2\phi^2 + 4Y_i\phi}{4\phi^2} \tag{813}$$

$$= -\sum_{i=1}^{n} \frac{2\phi^2 - 2Y_i^2}{4\phi^2} \tag{814}$$

$$= -\sum_{i=1}^{n} \frac{\phi^2 - Y_i^2}{2\phi^2} \tag{815}$$

$$= \frac{\sum_{i=1}^{n}(Y_i) - n\phi^2}{2\phi^2} \tag{816}$$

$$\tag{817}$$

Hvis vi sætter resultaterne sammen får vi:

(husk minusset fra det orindelige udtryk)

$$\frac{\sum_{i=1}^{n}(Y_i) - n\phi^2}{2\phi^2} - \frac{n}{2\phi} = \tag{818}$$

VI ser at:

$$\frac{n}{2\phi} = \frac{n\phi}{2\phi^2} \tag{819}$$

sådan at:

$$\frac{\sum_{i=1}^{n}(Y_i) - n\phi^2}{2\phi^2} - \frac{n\phi}{2\phi^2} = \frac{\sum_{i=1}^{n}(Y_i) - n\phi^2 - n\phi}{2\phi^2} \tag{820}$$

Som var det vi skulle vise!

Del 5) Find estimatoren

$$\frac{\sum_{i=1}^{n}(Y_i) - n\hat{\phi}^2 - n\hat{\phi}}{2\hat{\phi}^2} = 0 \tag{821}$$

$$\implies \qquad \sum_{i=1}^{n}(Y_i) = n\hat{\phi}^2 + n\hat{\phi} \tag{822}$$

Kan løses med solver:

$$\hat{\phi} = \frac{\sqrt{4\frac{1}{n}\sum_{i=1}^{n}(Y_i) + 1} - 1}{2} \tag{823}$$

Del 6) Find estimatet

indsætter $n = 125$ og $\sum_{i=1}^{n} Y_i = 627.6$

$$\hat{\phi} = \frac{\sqrt{4 \cdot \frac{627.6}{125} + 1} - 1}{2} = (\sqrt{21,08} - 1)/2 = (4.591 - 1)/2 = 1,795 \tag{824}$$

3.22 Øvelse 23

30/11/2018, Opgaver: 1, 2, 2

3.22.1 Opgave 1

- Sample likelihood funktionen angivet
- $L(\theta \mid y_1, \cdots y_n)$
- $\theta \in \Theta = \mathbb{R}$
- $\hat{\theta} = 12.41$
- $\mathbf{se}(\hat{\theta}) = 1.07$

Vi har fået angivet en sample likelihood fi

Del 1) Angiv konfidensinterval for den sande parameterværdi θ_0 - giv fortolkning

$$P(\hat{\theta} - 1.96 \cdot \mathbf{se}(\hat{\theta}) < \theta_0 < \hat{\theta} + 1.96 \cdot \mathbf{se}(\hat{\theta})) = 0.95 \tag{825}$$

$$P(12.41 - 1.96 \cdot 1.07) < \theta_0 < 12.41 + 1.96 \cdot 1.07) = 0.95 \tag{826}$$

$$P(10.3128 < \theta_0 < 14.5072) = 0.95 \tag{827}$$

Under antagelse af at vi ikke har misspecificeret vore model, ligger den sande parameter θ_0 i intervallet mellem 10.3 til 14.5.

Del 2) teori implicerer $\theta_0 = 10$, formuler som statistik hypotese

$$H_0 : \theta_0 = 10 \qquad H_a : \theta_0 \neq a \tag{828}$$

Vi kan derfor skrive:

$$\Theta_0 = \{10\} \qquad \Theta_A = \Theta \setminus \Theta_0 = \mathbb{R} \setminus \{10\} \tag{829}$$

Hvor parameterrummet for den uretstrikterede model er Θ.

Del 3) Test hypotesen med en z-test

Test størrelse:

$$\frac{\hat{\theta}_n - 10}{\mathbf{se}(\hat{\theta}_n)} = \frac{12.41 - 10}{1.07} = 2.252 \tag{830}$$

Den kritiske værdi: Den kritiske værdi blev fundet tidligere: 1.96

Konklusion: Vi kan afvise med over 95% sandsynlighed at θ_0 er 10.

p-værdi:

$$2(1 - \Phi(2.252)) = 2(1 - 0.9878) = 0.0244 \tag{831}$$

Her indser vi det følgende: $\Phi(z)$ er CDF'en for en standard normal fordeling. Vi bruger konstanten 2 fordi det er en to sidet test.

Del 4) Brug en kvadreret Wald-test

Nu sætter man teststørrelsen i anden:

$$\left(\frac{\hat{\theta}_n - 10}{\mathbf{se}(\hat{\theta}_n)}\right)^2 = 2.252^2 = 5.071 \tag{832}$$

Denne skal evalueres i en $\chi^2(v)$-fordeling. Først skal vi finde ud af hvad v er:

Vi indser at vi har 1 frit parameter under vores restriktion - altså $v = 1$.

Vores kritiske værdi med 1 frihedsgrad er: 3.8414. Fundet ved kommandoen i stata:

```
disp invchi2(1, 0.95)
```

Vi kan altså afvise vores H_0, da vores test-størrelse er over vores kritiske værdi.

Vores p-værdi er:

$$\chi^2_{df=1}(5.071) = 1 - 0.9757 \tag{833}$$

også fundet i stata med kommandoen:

```
disp chi2(1, 5.071)
```

Vi ganger ikke med to denne gang da alt massen i χ^2 fordelingen er på den positive reelle akse.

Noter om fordelingsantagelsen:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \sigma^2) \tag{834}$$

$$\sqrt{n}(\hat{\theta}_n - a - (\theta_0 - a)) \xrightarrow{d} N(0, \sigma^2) \qquad \Longrightarrow \tag{835}$$

$$\sqrt{n}\left(\frac{\hat{\theta}_n - a}{\sigma}\right) - \sqrt{n}\left(\frac{\theta_0 - a}{\sigma}\right) \xrightarrow{d} N(0, 1) \tag{836}$$

Man ser at hvis H_0 ikke gælder, vil det bagerste led ikke forsvinde. Husk $H_0 : \theta_0 = a$. Derfor er vores teststørrelse kun normaltfordelt hvis H_0 er sand!

3.22.2 Opgave 2

- lys-pærer eksempel fra tidligere igen
- $f_{Y_i}(y \mid \theta) = \theta \exp(-\theta y)$
- $y \in \mathbb{Y} = \mathbb{R}_+$

- $\theta \in \Theta = \mathbb{R}_+$
- Vi har at vores stokatiske variable Y_i er fordelt

$$Y_i \sim \text{Exponential}(\theta_0) \tag{837}$$

Del 1) tætheden skrive til tider: $\mu = \mathbf{E}(Y) = \theta^{-1}$ find maximum likelihood estimatoren

sample likelihood funktionen:

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\mu} \exp\left(-\frac{y_i}{\mu}\right) \tag{838}$$

log likelihood funktionen

$$\log L(\mu) = -n \log(\mu) - \frac{1}{\mu} \sum_{i=1}^{n} y_i \tag{839}$$

Scoren:

$$S(\mu) = \frac{\partial}{\partial \mu} \log L(\mu) = \frac{-n}{\mu} + \frac{1}{\mu^2} \sum_{i=1}^{n} y_i \tag{840}$$

Vi finder estimatoren:

$$\frac{-n}{\hat{\mu}} + \frac{1}{\hat{\mu}^2} \sum_{i=1}^{n} y_i = 0 \tag{841}$$

$$\implies \qquad n\hat{\mu} = \sum_{i=1}^{n} y_i \tag{842}$$

$$\implies \qquad \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{843}$$

Del 2) Vis estimatoren er middelret

$$\begin{aligned}
\mathrm{E}(\mu_0) &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] &(844)\\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(Y_i) &(845)\\
&= \frac{1}{n}\sum_{i=1}^{n} \mu_0 = \mu_0 &(846)
\end{aligned}$$

Del 3) Find variansen af estimatoren $\hat{\mu}_n$

$$\mathrm{Var}(\hat{\mu}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(Y_i) \tag{847}$$

Husk nu at normalt er tætheden af en exponential fordeling:

$$f_{Y_i}(y \mid \lambda) = \lambda \exp(-\lambda y) \tag{848}$$

I dette tilfælde ville $\mathrm{E}(Y_i) = \lambda^{-1}$ og $\mathrm{Var}(Y_i) = \lambda^{-2}$.

Udnyt det vi ved om variansen:

$$\frac{1}{n^2}\sum_{i=1}^{n} Var(Y_i) = \frac{1}{n^2}\sum_{i=1}^{n} \mu_0^2 = \frac{1}{n^2} \cdot n\mu_0^2 = \frac{\mu_0^2}{n} \tag{849}$$

Del 4) Angiv den asympotiske fordeling af estimatoren

Vi er under antagelse af en urestrikteret model. Dvs sige vores parameterrum er $\mu \in \Theta = \mathbb{R}$

Vi ved at den centrale grænseværdi sætning vil gælde:

$$\sqrt{n}(\mu - \hat{\mu}_n) \xrightarrow{d} N(0, \sigma^2) = N(0, \mu^2) \tag{850}$$

HVor man kan gange igennem men $\sqrt{n}$ og få:

$$(\mu - \hat{\mu}_n) \xrightarrow{d} N(0, \mu^2/n) \tag{851}$$

3.22.3 Opgave 3

Fortsær med udgangspunkt i opgaven ovenfor:

Del 1) Opskriv log likelihood bidraget for den generelle model

$$\log l(\mu \mid y_i) = \log\left(\frac{1}{\mu}\exp\left(-\frac{y_i}{\mu}\right)\right) = -\mu - \frac{y_i}{\mu} \tag{852}$$

Del 2) Hesse bidraget betinget på man har maximeret likelihood-scoren

$$H_i(\mu_0) = \left.\frac{\partial^2 \log l(\mu \mid y_i)}{\partial \mu^2}\right|_{\mu=\mu_0} \tag{853}$$

Først finder vi den første afledte:

$$\frac{\partial}{\partial \mu}\log l(\mu) = -\mu^{-1} + y_i\mu^{-2} \tag{854}$$

Hernest findes den anden afledte:

$$\frac{\partial^2 \mu}{\partial \mu^2}\log l(\mu) = \mu^{-2} - 2y_i\mu^{-3} \tag{855}$$

beting nu på at $\mu = \mu_0$.

$$\left.\frac{\partial^2 \mu}{\partial \mu^2}\log l(\mu)\right|_{\mu=\mu_0} = \mu_0^{-2} - 2y_i\mu_0^{-3} \tag{856}$$

Del 3) Find informationsmatricen

$$\begin{aligned}
I(\mu_0) = \mathrm{E}(-H_i(\mu_0)) = -\mathrm{E}(H_i(\mu_0)) &= -\mathrm{E}(\mu_0^{-2} - 2y_i\mu_0^{-3}) && (857)\\
&= \mathrm{E}(-\mu_0^{-2} + 2y_i\mu_0^{-3}) && (858)\\
&= \mathrm{E}(-\mu_0^{-2}) + \mathrm{E}(2y_i\mu_0^{-3}) && (859)\\
&= -\mu_0^{-2} + 2\mu_0^{-3}\mathrm{E}(y_i) && (860)\\
&= -\mu_0^{-2} + 2\mu_0^{-3}\mu_0 && (861)\\
&= -\mu_0^{-2} + 2\mu_0^{-2} && (862)\\
&= \mu_0^{-2} && (863)
\end{aligned}$$

Evalueret i estiamtoren bliver det:

$$I(\hat{\mu}_n) = \hat{\mu}_n^{-2} \tag{864}$$

Del 4) Find variansen på estimatoren

$$V(\hat{\mu}_n^{-2}) = \frac{I(\hat{\mu}_n^{-1})}{n} = \frac{\mu_0^2}{n} \tag{865}$$

Hvilket er det samme som tidligere fundet!

3.23 Øvelse 24

30/11/2018, Opgaver: 4, 5, 6

3.23.1 Opgave 4

Kig opgave tekst for beskrivelse. Der er følgende variable: Survived, female, child, class.

Del 1) Opstil en statistisk model for variablen *survived*

Vi lader y_i være variablen survived. Vi betinger altså ikke på noget af den information, som ellers er tilgængelig i datasættet. Det vil sige vi sandsynligheden for at overleve ens, om man er kvinde, barn eller hvilken klasse man rejste på.

Vi ser at Y_i er en binær variable. Vi vælger altså at modellere med en bernoulli fordeling.

Uddybning: Flere hat spurgt hvordan man vælger fordeling når man modellerer sin data. Her er et godt eksempel. Vi kigger på dataens struktur (den er binær) og herfra beslutter vi os for hvad den korrekte fordeling er. Men bare fordi vi kun observerer to udfald, er det ikke nødvendigvis korrekt at bruge bernoulli fordelingen.

Antag man havde adspurgt 3. årgang på en folkeskole om hvor mange kæledyr de har. Alle svarene er enten $\{0, 1\}$. De besvarer ikke et sandt falsk spørgsmål, og derfor ville jeg ikke mene bernoulli-fordelingen er korrekt at bruge. Dels fordi man kan have flere kæledyr end et, men også fordi at fortolkningen af bernoulli ville være underlig i denne sammenhæng.

Tilbage til eksemplet.

Vi indser at vores parameterrum $\Theta =]0, 1[$. Således at $\theta \in \Theta$.

$$P(Y_i) = \begin{cases} \theta & , Y_i = 1 \\ 1 - \theta & , Y_1 = 0 \end{cases} \tag{866}$$

Vores tæthed er:

$$f_{Y_i}(y \mid \theta) = \theta^y (1 - \theta)^{(1-y)} \tag{867}$$

Hvorfra vi finder vores likelihood contribution:

$$l(\theta \mid y_i) = \theta^{y_i} (1 - \theta)^{(1-y_i)} \tag{868}$$

Vi har selvfølgelig antaget $i.i.d$ observationer (selvom det er i strid, med vores observationer i datasættet).

Vi opskriver sample likelihood funktionen:

(Kig i heinos bog s. 58)

$$L(\theta) = \sum_{i=1}^{n} l(\theta \mid y_i) = \sum_{i=1}^{n} \theta^{y_i} (1 - \theta)^{(1-y_i)} \tag{869}$$

Vi finder log likelihood funktionen

$$\log L(\theta) = \sum_{i=1}^{n} y_i \log(\theta) + (1 - y_i) \log(1 - \theta) \tag{870}$$

definér $R_n = \sum_{i=1}^{n} y_i$

$$\log L(\theta) = R_n \log(\theta) + (n - R_n) \log(1 - \theta) \tag{871}$$

Herfra udledes scoren:

Fra side 68 i bogen finder vi estimatoren $\hat{\theta}(Y_1, Y_2, \cdots Y_n)$: (bare differentier og sæt scoren lig 0, og løs for $\hat{\theta}_n$)

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{872}$$

Vi bruger dette til at finde estimatet $\hat{\theta}(y_1, y_2, \cdots y_n)$

Vi kan finde i datasættet at: $n = 2201$ og at $\sum_{i=1}^{n} y_i = 711$

$$\hat{\theta}_n = \frac{1}{2201} \cdot 711 = 0.323 \tag{873}$$

Del 2) Find variansen på estimatoren og angiv den asymptotiske fordeling på estimatoren

Vi finder variansen på vores estimator på følgende måde (husk man kan også gøre det ved at bruge informations matricen):

$$\begin{aligned}
\text{Var}(\hat{\theta}_n) &= \text{Var}(\frac{1}{n}\sum_{i=1}^{n} y_i) && (874)\\
&= \frac{1}{n^2}\text{Var}(\sum_{i=1}^{n} Y_i) \qquad \text{husk i.i.d } Y_i && (875)\\
&= \sum_{i=1}^{n} \text{Var}(Y_i) && (876)\\
&= \frac{n}{n^2}\text{Var}(Y_i) && (877)\\
&= \frac{1}{n}\text{Var}(Y_i) && (878)
\end{aligned}$$

Vi finder den asympotiske fordeling:

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_n - \theta_0) &\xrightarrow{d} N(0, \Omega_0) = N\left(0, \text{Var}(Y_i)\right) \implies && (879)\\
\sqrt{n}\frac{(\hat{\theta}_n - \theta_0)}{\mathbf{se}(\hat{\theta}_n)} &\xrightarrow{d} N(0, 1) && (880)
\end{aligned}$$

Del 3

Kig github for do-file. Får samme koefficient.

Del 4)

Kig do-file

Del 5) Undersøg om sandsynligheden for at overleve variere mellem mænd og kvinder

$$P(Y_i = 1) = \theta_i = \theta + \delta \cdot female_i \tag{881}$$

$$L_n(\theta \mid \delta) = \prod_{i=1}^{n} \theta_i^{y_i}(1-\theta_i)^{(1-y_i)} \tag{882}$$

$$\log L(\theta, \delta) = \sum_{i=1}^{n} y_i \log(\theta_i) + (1-y_i)\log(1-\theta_i) \tag{883}$$

Dette kan igen transformeres:

lad $f_i = female_i$ of husk at$\theta_i = \theta + \delta \cdot f_i$

$$\log L(\theta, \delta) = \sum_{i=1}^{n} y_i \log(\theta + \delta \cdot f_i) + (1-y_i)\log(1-(\theta + \delta \cdot f_i)) \tag{884}$$

Vi ser altså, at vi ikke er i stand til at sætte noget udenfor sum-tegnet.

Kig do-file for estimation:

Coefficienter:

- $\theta = 0.212$
- $\delta = 0.519$

Vi ser altså sandsynligheden for at overleve som mand nu falder til 0.212. Sandsynligheden for at overleve som Kvinde er noget højere. Vi ser det handler om at $P(Y_i = 1) = \theta_i = \theta + \delta \cdot f_i$. Altså Sandsynlighden for at overleve hvis man er kvinde er $\theta + \delta = 0.212 + 0.519 = 0.731$

Del 6) Forklar, hvorfor dette er et eksempel på en betinget fordeling

Vores første model tog ikke højde for om en person var kvinde eller ej når vi udregnede sandsynligheden for om hvorvidt en person overlevede.

Nu tilføjer vi information og siger: GIVET man er kvinde, hvor stor var sandsynligheden så for, at man overlevede, eller omvendt: GIVET man IKKE var kvinde, hvor stor var ssh. for man overlevede.

Vi udleder de betingede estimatorer nu:

Altså sandsynligheden for en kvinde overlever GIVET det var en kvinde:

$$P(Y_i = 1 \mid F_i = 1) = \frac{1}{\sum_{i=1}^{n} F_i} \sum_{i=1}^{n} Y_i F_i \tag{885}$$

Her ses at $\sum_{i=1}^{n} F_i$ må være antallet af kvinder, og at: $Y_i F_i$ er en interaktionsvariabel som kun er 1, hvis Y_i er 1 og F_1. Altså hvis man var kvinder OG man overlevede!

Omvendt ville man kunne lave for mænd:

$$P(Y_i = 1 \mid F_i = 0) = \frac{1}{\sum_{i=1}^{n}(1 - F_i)} \sum_{i=1}^{n} Y_i(1 - F_i) \tag{886}$$

Hvor man ser at $Y_i(1 - F_i)$ kun kan være 1, hvis $F_i = 0 \wedge Y_i = 1$.

3.23.2 Opgave 5

Fortsæt med ovenstående analyse af Titanics forlis.

Del 1) Opstil ideen om at mænd og kvinder blev behandlet ens under titanics forlis.

$$H_0 : \delta = 0 \qquad H_A : \delta \neq 0 \tag{887}$$

Det er sværere her at restriktere vores parameter rum Θ. Vi kan sige i den urestrikterede model, må $\theta \in]0, 1[$, da dette er sandsynligheden for mænd overlever. Vi ser nu, at det bliver mere besværligt for δ. Delta kan altså derfor være $\delta \in] - 1, 1[$.

Men faktisk kan vi lave en strammere restriktion. Dette skyldes, at $0 < \delta + \theta < 1$

$$(\theta, \delta)^T \in \Theta = \{(\theta, \delta)^T \in \mathbb{R}^2 : 0 < \theta < 1, 0 < \theta + \delta < 1\} \tag{888}$$

Vi kan skrive:

$$\Theta_0 = \{(\theta, \delta)^T \in \mathbb{R}^2 : 0 < \theta < 1, \delta = 0\} \tag{889}$$

$$\Theta_A = \{(\theta, \delta)^T \in \mathbb{R}^2 : 0 < \theta < 1, \delta \neq 0\} \tag{890}$$

Del 2) Lav en Z-test for vores hypotese.

Hvis nul hypotesen er forkert divergere man enten mod $-\infty$ eller ∞. Hvis H_0 er sand, da vil testen være normalt fordelt.

Teststørrelse findes i vores do-file:

$$z = 22.93 \tag{891}$$

Alternativt kunne det regnes selv:

Dette gøres ved:

$$\frac{\hat{\theta}_n - \theta_0}{\mathbf{se}(\delta)} = \frac{0.519 - 0}{\mathbf{se}(0.0226)} = 22,9 \tag{892}$$

Hvor der er antaget at vores H_0 er sand, således at $\theta_0 = 0$.

Vi afviser vores test med 95 % ¿ sandsynligheden, at dødeligheden skulle være det samme for mænd og kvinder.

Dette er klart da den kritiske værdi for en dobbelt sidet test er -1.96 og 1.96, på et 95% konfidens niveau.

Del 3) Test hypotesen med en kvadreret Wald-test

Vi kvadrerer vores test-størrelse og evaluere i en χ^2-fordeling

$$z^2 = 22.93^2 = 525.78 \tag{893}$$

Hvilket skal evalueres i χ^2-fordelingen. Vi indser vi har en frihedsgrad: altså $\chi^2_{df=1}(525.78) =$

Hvilket betyder vi afviser da den kritiske værdi for en $\chi^2_{df=1}$, med 95 % ssh er 3.84.

Del 4) test hypotesen med en Likelihood ration test (LR-test)

Vi husker at $\tilde{\theta}_n$ er vores restrikterede model.

$$LR(H_0) = -2\log\left(\frac{L_n(\tilde{\theta}_n)}{L_n(\hat{\theta}_n)}\right) = 2\left[\log L_n(\hat{\theta}_n) - \log(\tilde{\theta}_n)\right] \tag{894}$$

Vi kigger på vores værdier i vores output.

- Den restrikterede model: -1384.7284 Hvor det forstås at, den restrikterede model antog $\delta = 0$. Hvilket svarer til den første model.
- Den urestrikterede model: -1167.4939. Altså den model, hvor vi tillader δ at antage hvilken som helst værdi.

$$LR = 2((-1167.4939) - (-1384.7284)) \approx 434 \tag{895}$$

Vi evaluere testen i en $\chi^2_{df=1}$-fordeling. VI ved den kritiske værdi er som før 3.84.

Og vi kan afvise testen.

$$p\textbf{-value} = 1 - \chi^2_{df=1}(434) = 0.000 \tag{896}$$

3.23.3 Opgave 6

Vi fortsætter ovenstående analyse af Titanics forlis.

Del 1) Antag at mænd og kvinder blev behandlet ens med $\theta_i = 0$. Opskriv hypotesen

Antag at mænd og kvinder blev behandlet ens og værdien var $\theta_i = 0.35$. (husk $\theta_i = \theta + \delta \cdot f_i$). Vi kan altså se at $\delta = 0$, da kvinder og mænd antages at behandles ens. Det betyder:

$$\theta_i = \theta + \delta \cdot f_i \qquad (\delta = 0) \tag{897}$$

$$\implies \quad \theta_i = \theta + 0 \cdot f_i \tag{898}$$

$$\implies \quad \theta_i = \theta = 0.35 \tag{899}$$

$$H_0 : \theta = 0.35 \wedge \delta = 0, \qquad H_A : \theta \neq 0.35 \vee \delta \neq 0 \tag{900}$$

Vi opskriver parameterrummet. Θ er samme som tidligere. Tegn skitse af parameterrummet!

$$(\theta,\delta)^T \in \Theta_0 = \{(\theta,\delta)^T \in \mathbb{R}^2 : 0 < \theta < 1, 0 < \delta + \theta < 1\} \tag{901}$$

$$(\theta,\delta)^T \in \Theta_0 = \{(\theta,\delta)^T = (0.35, 0)^T\} \tag{902}$$

Del 2) test hypotesen med likelihood-ratio test.

Vi har den urestrikterede model fra tidligere.

Nu skal vi estimere den restrikterede model.

Løses i stata (kig do-file):

For den restrikterede model får vi værdien:

$$L_n = -1388.2901 \tag{903}$$

Indsættes nu i vores likelihood ratio test:

$$LR(H_0) = 2((-1167.4939) - (-1388.2901)) = 442 \tag{904}$$

hvor (-1167.4939) er værdien for likelikehood under antagelse af at θ, δ har de værdier som maximere scoren.

Vi skal evaluere test-værdien i en $\chi^2_{df=2}$ da vi har 2 parametre som estimeres. Vi har en kritisk værdi på 5.99 for 2 parametre, afviser vi vores nul-hypotese.

Del 3)

Nej z-test kan kun bruges, når der er er en enkel restriktion på data!