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

November 27, 2018

2 Topics

2.1 Øvelse 21

26/11/2018, Opgaver: 1, 2, 3

2.1.1 Opgave 1

Lav i klassen!

Kig do-files

2.1.2 Opgave 2

- $n = 573$
- 121 har $y = 1$
- 251 har $y = 2$
- 201 har $y = 3$
- Vi betegner de absolutte frekvenser som s_j:

$$s_j = \sum_{i=1}^{n} \mathbb{1}(y_i = j) \tag{1}$$

- Vi ser at $s_1 + s_2 + s_3 = n$
- Opstiller en statistik model:

$$P(Y_i = y) = \begin{cases} p_1 & \text{hvis } y = 1 \\ p_2 & \text{hvis } y = 2 \\ p_3 & \text{hvis } y = 3 \end{cases} \tag{2}$$

Del 1) Hvis at sandsynlighedsunktionen kan skrives som:

$$f_{Y_i}(y \mid p_1, p_2, p_3) = p_1^{\mathbb{1}(y=1)} p_2^{\mathbb{1}(y=2)} p_3^{\mathbb{1}(y=3)} \tag{3}$$

Minder om opgaven til sidste uge:

Man indser først at vi har:

$$S = (\text{expression})^{\mathbb{1}(\text{condition})} \tag{4}$$

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 $(y = 1$

$$f_{Y_i}(1 \mid p_1, p_2, p_3) = p_1 \cdot 1 \cdot 1 \tag{5}$$

Hvilket er det som modellen skulle overholde!

Del 2) Opskriv likelihood bidraget samt sample likelihood funktionen

$$l(p_1, p_2, p_3 \mid y_i) = p_1^{\mathbb{1}(y_i=1)} p_2^{\mathbb{1}(y_i=2)} p_3^{\mathbb{1}(y_i=3)} \tag{6}$$

Vi ser nu sample likelihood funktionen kan skrives op:

$$\begin{aligned} L(p_1, p_2, p_3 \mid y_1, y_2, \cdots, y_n) &= \prod_{i=1}^{n} l(p_1, p_2, p_3 \mid y_i) && (7) \\ &= p_1^{s_1} p_2^{s_2} p_3^{s_3} && (8) \end{aligned}$$

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:

$$\theta = (\theta_1, \theta_2) \in \Theta \tag{9}$$

hvor at $\Theta = [0, 1] \times [0, 1]$

Vi opskriver likelihood funktionen som funktion af de frie parametre θ_1, θ_2:

$$L(\theta_1, \theta_2 \mid y_1, y_2, \cdots, y_n) = \theta_1^{s_1} \theta_2^{s_2} (1 - \theta_1 - \theta_2)^{s_3} \tag{10}$$

Vi kan opskrive Log likelihood funktionen

$$\log L(\theta_1, \theta_2 \mid y_1, y_2, \cdots, y_n) = s_1 \log(\theta_1) + s_2 \log(\theta_2) + s_3 \log(1 - \theta_1 - \theta_2) \tag{11}$$

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:

$$\frac{\partial}{\partial \theta_1} \log L(\theta_1, \theta_2 \mid Y_1, Y_2, \cdots, Y_n) = 0 \tag{12}$$

$$\frac{\partial}{\partial\theta_2}\log L(\theta_1,\theta_2 \mid Y_1,Y_2,\cdots,Y_n) = 0 \tag{13}$$

Notér: Vi bruger store Y da vi skal finde estimatoren!

Vi ser at for θ_1 får vi:

noter: $\frac{\partial}{\partial\theta_1}log(1-\theta_1-\theta_2) = -\frac{1}{1-\theta_1-\theta_2}$

$$\frac{s_1}{\hat{\theta}_1} - \frac{s_3}{1-\hat{\theta}_1-\hat{\theta}_2} = 0 \tag{14}$$

Ligeledes finder vi:

$$\frac{s_2}{\hat{\theta}_2} - \frac{s_3}{1-\hat{\theta}_1-\hat{\theta}_2} = 0 \tag{15}$$

Hvilket betyder at:

$$\frac{s_2}{\hat{\theta}_2} = \frac{s_1}{\hat{\theta}_1} \tag{16}$$

Vi isolerer $\hat{\theta}_2 = \frac{\hat{\theta}_1 s_2}{s_1}$

Vi indsætter så vi har:

$$\frac{s_1}{\hat{\theta}_1} - \frac{s_3}{1-\hat{\theta}_1-\frac{\hat{\theta}_1 s_2}{s_1}} = 0 \tag{17}$$

$$\frac{s_1}{\hat{\theta}_1} = \frac{s_3}{1-\hat{\theta}_1(1+s_2/s_1)} \tag{18}$$

Herfra kan man yderligere isolere således at man finder:

$$\hat{\theta}_1 = \frac{s_1}{s_1+s_2+s_3} = \frac{s_1}{n} \tag{19}$$

Ligeledes ville man kunne finde for θ_2.

Vi kan nu finde estimatet:

$$\hat{\theta}_1 = \frac{s_1}{n} = \frac{121}{573} = 0.211 \tag{20}$$

$$\hat{\theta}_2 = \frac{s_2}{n} = \frac{251}{573} = 0.438 \tag{21}$$

Og vi kan finde $p_3 = 1 - \theta_1 - \theta_2 = 1 - 0.211 - 0.438 = 0.351$

2.1.3 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:

$$X_n = \frac{1}{n} \sum_{i=1}^{n} Y_i \tag{22}$$

- Den samlede reblængde er:

$$nX_n = \sum_{i=1}^{n} Y_i \tag{23}$$

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:

$$U_n = \frac{X_n - \mu}{\sigma/\sqrt{n}} \tag{24}$$

Og vi ved at U_n vil være standard normalt fordelt for $n \to \infty$

Det betyder også at vi kunne sige at:

$$X_n \sim N(200, 100/n) \tag{25}$$

hvor vi har at $\sigma = 10$ og at $\sigma^2 = 10^2 = 100$.

Transformation af U_n er klar:

$$
\begin{aligned}
U_n &= \frac{X_n - \mu}{\sigma} && (26)\\
&= \frac{X_n - \mu}{\sqrt{\sigma^2/n}} && (27)\\
\implies \qquad \sqrt{\sigma^2/n} \cdot U_n &= X_n - \mu && (28)\\
\implies \qquad \sqrt{\sigma^2/n} \cdot U_n + \mu &= X_n && (29)
\end{aligned}
$$

Vi indser at:

$$\sqrt{\sigma^2/n} \cdot U_n \sim N(0, \sigma^2/n) \tag{30}$$

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:

$$125 \cdot 100/60 = 208,33 \tag{31}$$

Vi kan transformere dette!

$$\frac{208.33 - 200}{10/\sqrt{60}} = 8.33/1.2909 = 6.4528 \tag{32}$$

Vi kan nu spørge:

$$P(U_n > 6.45) = 1 - \Phi(6.45) = 1 - 0.99999999 = 0 \tag{33}$$

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!

2.2 Øvelse 22

30/11/2018, Opgaver: 4, 5

2.2.1 Opgave 4

side 65 i Heino's bog

- $\{Y_i\}_{i=1}^n$ er i.i.d. normalt fordelte stokastiske variable
- altså:

$$Y_i \sim N(\mu, \sigma^2) \tag{34}$$

- tæthedsfunktionen:

$$f_{Y_i}(y \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right) \tag{35}$$

Del 1.a) Opskriv likelihood funktionen $L(\mu, \sigma^2 \mid Y_1, Y_2 \cdots Y_n)$

likelihood contribution:

$$l(\mu, \sigma^2 \mid Y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(Y_i-\mu)^2}{2\sigma^2}\right) \tag{36}$$

Likelihood funktionen:

$$L(\mu, \sigma^2 \mid Y_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(Y_i-\mu)^2}{2\sigma^2}\right) \tag{37}$$

Del 1.b) Opskriv log likelihood funktionen

Vi tager logaritmen:

$$\log L(\mu, \sigma^2 \mid Y_1, Y_2, \cdots Y_n) = \sum_{i=1}^{n} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) + \frac{-(Y_i-\mu)^2}{2\pi^2} \tag{38}$$

Del 1.c) Find estimatorerne

Kig side 70.

$$\frac{\partial}{\partial \mu} \log L(\mu, \sigma) = \sum_{i=1}^{n} \frac{Y_i - \mu}{\sigma^2} \tag{39}$$

$$-\frac{1}{2\sigma^2} + \frac{(Y_i - \mu)^2}{2\sigma^4} \tag{40}$$

Sæt disse lig med 0 og vi kan se at:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} Y_i \tag{41}$$

$$\hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\mu})^2 \tag{42}$$

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:

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{\mu})^2 = \sum_{i=1}^{n} y_i^2 + n\hat{\mu}^2 - 2\hat{\mu} \sum_{i=1}^{n} y_i \tag{43}$$

Vi finder først estimatet for μ

$$\hat{\mu} = \frac{1}{125} \cdot 627.6 = 5.0208 \tag{44}$$

Herefter finder vi det for σ^2

note: Skriv måske formlen op igen for variansen!

$$\hat{\sigma^2} = 3807.2 + 125 \cdot 5.0208^2 - 2 \cdot 5.0208 \cdot 627.6 \tag{45}$$

Likelihood og Log Likelihood (under equi-dispersion $\mu = \sigma^2 = \phi$)

Dette tilfælde kaldes *equi dispersion.*

tætheden er givet ved:

$$f_{Y_i}(y \mid \phi) = \frac{1}{\sqrt{2\pi\phi}} \exp\left(\frac{-(y-\phi)^2}{2\phi}\right) \tag{46}$$

Vi opskriver Likelihood funktionen:

$$L(\phi \mid Y_1, Y_2 \cdots Y_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\phi}} \exp\left(\frac{-(Y_i-\phi)^2}{2\phi}\right) \tag{47}$$

Vi opskriver også log likelihood funktionen

$$\log L(\phi) = \sum_{i=1}^{n} \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\phi) + \left(-\frac{(Y_i-\phi)^2}{2\phi}\right) \tag{48}$$

Som kan omskrives til:

$$\log L(\phi) = \frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\phi) - \frac{\sum_{i=1}^{n}(Y_i-\phi)^2}{2\phi} \tag{49}$$

Del 4) Udled scoren

$$S(\phi) = \frac{\partial}{\partial\phi}\log L(\phi \mid Y_1, \cdots, Y_n) = \frac{\sum_{i=1}^{n} Y_i - n\phi^2 - n\phi}{2\phi^2} \tag{50}$$

Man ser man kan brække udtrykket (fra før) over i to dele når man differentiere:

Den første del $\frac{n}{2}\log(\phi)$

$$\frac{\partial}{\partial\phi}\left(\frac{n}{2}\log(\phi)\right) = \frac{n}{2\phi} \tag{51}$$

Den anden del:

$$\frac{\partial}{\partial\phi}\left(-\frac{\sum_{i=1}^{n}(Y_i-\phi)^2}{2\phi}\right) = -\sum_{i=1}^{n}\frac{\partial}{\partial\phi}\left(\frac{(Y_i-\phi)^2}{2\phi}\right) \tag{52}$$

Dette skrives som (husk hint i første skridt - differentiation af brøker) :

$$-\sum_{i=1}^{n} \frac{\partial}{\partial \phi} \left(\frac{(Y_i - \phi)^2}{2\phi} \right) = -\sum_{i=1}^{n} \frac{(-2(Y_i - \phi)2\phi) - (2(Y_i - \phi)^2)}{4\phi^2} \tag{53}$$

$$= -\sum_{i=1}^{n} \frac{-4Y_i\phi + 4\phi^2 - 2Y_i^2 - 2\phi^2 + 4Y_i\phi}{4\phi^2} \tag{54}$$

$$= -\sum_{i=1}^{n} \frac{2\phi^2 - 2Y_i^2}{4\phi^2} \tag{55}$$

$$= -\sum_{i=1}^{n} \frac{\phi^2 - Y_i^2}{2\phi^2} \tag{56}$$

$$= \frac{\sum_{i=1}^{n}(Y_i) - n\phi^2}{2\phi^2} \tag{57}$$

$$\tag{58}$$

Hvis vi sætter resultaterne sammen får vi:

(husk minusset fra det orindelige udtryk)

$$\frac{\sum_{i=1}^{n}(Y_i) - n\phi^2}{2\phi^2} - \frac{n}{2\phi} = \tag{59}$$

VI ser at:

$$\frac{n}{2\phi} = \frac{n\phi}{2\phi^2} \tag{60}$$

sådan at:

$$\frac{\sum_{i=1}^{n}(Y_i) - n\phi^2}{2\phi^2} - \frac{n\phi}{2\phi^2} = \frac{\sum_{i=1}^{n}(Y_i) - n\phi^2 - n\phi}{2\phi^2} \tag{61}$$

Som var det vi skulle vise!

Del 5) Find estimatoren

$$\frac{\sum_{i=1}^{n}(Y_i) - n\hat{\phi}^2 - n\hat{\phi}}{2\hat{\phi}^2} = 0 \tag{62}$$

$$\implies \qquad \sum_{i=1}^{n}(Y_i) = n\hat{\phi}^2 + n\hat{\phi} \tag{63}$$

Kan løses med solver:

$$\hat{\phi} = \frac{\sqrt{4\frac{1}{n}\sum_{i=1}^{n}(Y_i) + 1} - 1}{2} \tag{64}$$

Del 6) Find estimatet

indsætter $n = 125$ og $\sum_{i=1}^{n} Y_i = 627.6$

$$\hat{\phi} = \frac{\sqrt{4 \cdot \frac{627.6}{125} + 1} - 1}{2} = (\sqrt{21,08} - 1)/2 = (4.591 - 1)/2 = 1,795 \tag{65}$$