Eksamen på Økonomistudiet vinteren 2018-19

Sandsynlighedsteori og Statistik

2. årsprøve

5. januar, 2019

(3-timers prøve med hjælpemidler)

Rettevejledning

Opgave 1

1. $P(X_1 = 0, X_2 = 0) = P(X_1 = 0)P(X_2 = 0) = (1 - 0.05)(1 - 0.07) = 0.8835$

2. Lad Y være indkomsten. Vi har at der er 4 kombinationer af tilbud:

$$\begin{aligned}
\mathbb{E}(Y) = &\ \mathbb{E}(Y|\text{job i type 1})P(\text{tilbud fra type 1})P(\text{ikke tilbud fra type 2}) \\
&+ \mathbb{E}(Y|\text{job i type 1})P(\text{tilbud fra type 1})P(\text{tilbud fra type 2}) \\
&+ \mathbb{E}(Y|\text{job i type 2})P(\text{ikke tilbud fra type 1})P(\text{tilbud fra type 2}) \\
&+ \mathbb{E}(Y|\text{intet job})P(\text{ikke tilbud fra type 1})P(\text{ikke tilbud fra type 2}) \\
&= 500 \cdot 0.05 \cdot (1 - 0.07) + 500 \cdot 0.05 \cdot 0.07 + 300 \cdot (1 - 0.05) \cdot 0.07 + 0 \cdot (1 - 0.05) \cdot (1 - 0.07) \\
&= 44.95
\end{aligned}$$

   dvs. 44950kr.

3. Lad Z_2 være antallet af jobs i virksomhedstype 2 tilbudt til den arbejdsløse. Vi har at $Z_2 \sim Bin(7, 0.07)$ og at $P(Z_2 > 0) = 1 - P(Z_2 = 0) = 1 - \begin{pmatrix} 7 \\ 0 \end{pmatrix} 0.07^0(1 - 0.07)^7 \approx 1 - 0.6017 = 0.3983$

4. Vi approksimerer nu Binomialfordelingen med Poissonfordelingen. Vi er interesserede i den stokastiske variabel $Z = Z_1 + Z_2 = \sum_{j=1}^{12} X_j$ som måler antal job-tilbud den arbejdsløse har fået. Da Z_1 og Z_2 er (approksimativt) Poissonfordelt med parametre $\lambda_1 = n_1p_1 = 5 \cdot 0.05 = 0.25$ og $\lambda_2 = n_2p_2 = 7 \cdot 0.07 = 0.49$ er summen, Z, Poissonfordelt med parameteren $\lambda = \lambda_1 + \lambda_2 = 0.74$. Vi har således, at $P(Z > 0) = 1 - P(Z = 0) = 1 - \frac{\lambda^0}{0!}\exp(-\lambda) = 1 - \exp(-\lambda) = 1 - \exp(-0.74) \approx 0.5229$

Opgave 2

1. Tætheden er

$$p_X(x) = 2\exp(-2x)$$

2. Vi benytter regneregler for middelværdi til at beregne

$$\mathbb{E}(Y) = \mathbb{E}(2 \cdot X - 1) = 2 \cdot \mathbb{E}(X) - 1 = 2 \cdot \frac{1}{2} - 1 = 0$$

Vi kan ligeledes bruge regneregler for varians til at finde

$$Var(Y) = Var(2 \cdot X - 1) = 2^2 Var(X) = 4 \cdot \frac{1}{4} = 1$$

3. Vi definerer $Y = t(X) = 2X - 1$ og har

$$\begin{aligned} t^{-1}(y) &= \frac{1}{2}(y+1) \\ \frac{dt^{-1}(y)}{dy} &= \frac{1}{2} \end{aligned}$$

sådan at for $y \in (-1, \infty)$ er tæthedsfunktionen for Y

$$\begin{aligned} q(y) &= p_X(t^{-1}(y)) \left| \frac{dt^{-1}(y)}{dy} \right| \\ &= 2\exp(-2\frac{1}{2}(y+1))\frac{1}{2} \\ &= \exp(-(y+1)) \end{aligned}$$

og vi har

$$q(y) = \begin{cases} 0 & \text{hvis } y \leq -1 \\ \exp(-(y+1)) & \text{hvis } y > -1 \end{cases}$$

Opgave 3

1. Parameterrummet er $\Theta = \{\theta : 0 < \theta < 1\}$.

2. Vi har at likelihood bidragende for hvert par er $\ell(\theta|y_i) = p(y_i) = (1-\theta)^{y_i - 1}\theta$ og log-likelihood bridraget er $\log(\ell(\theta|y_i)) = (y_i - 1)\log(1-\theta) + \log(\theta)$. Log-likelihood funktionen bliver således

$$\begin{aligned} \log L_n(\theta) = \log L(\theta|y_1, \ldots, y_n) &= \sum_{i=1}^{n} [(y_i - 1)\log(1-\theta) + \log(\theta)] \\ &= n \cdot (\log(\theta) - \log(1-\theta)) + \log(1-\theta)\sum_{i=1}^{n} y_i \end{aligned}$$

3. Første ordens betingelsen (FOC) for den givne model er, at scoren skal være nul,

$$S(\hat{\theta}_n) = \left.\frac{\partial \log L_n(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_n} = 0$$

hvor vi her har at

$$\begin{aligned}
\frac{\partial \log L_n(\theta)}{\partial \theta} &= \sum_{i=1}^{n} \frac{\partial \log(\ell(\theta|Y_i))}{\partial \theta} \\
&= \sum_{i=1}^{n} s_i(\theta) \\
&= \sum_{i=1}^{n} \left[\frac{1}{\theta} - \frac{Y_i - 1}{1-\theta}\right] \\
&= \frac{n}{\theta} - \frac{(\sum_{i=1}^{n} Y_i) - n}{1-\theta}
\end{aligned}$$

således at maximum likelihood **estimatoren** kan findes som løsningen til ligningen

$$\begin{gathered}
\frac{n}{\hat{\theta}} = \frac{(\sum_{i=1}^{n} Y_i) - n}{1-\hat{\theta}} \\
\Updownarrow \\
n(1-\hat{\theta}) = \hat{\theta}\left(\left(\sum_{i=1}^{n} Y_i\right) - n\right) \\
\Updownarrow \\
n = \hat{\theta} \sum_{i=1}^{n} Y_i \\
\Updownarrow \\
\hat{\theta} = \frac{n}{\sum_{i=1}^{n} Y_i}
\end{gathered}$$

Ved at bruge $n = 67$ og $\sum_{i=1}^{67} y_i = 107$ kan vi udlede **estimatet** for vores data som

$$\begin{aligned}
\hat{\theta}_n = \hat{\theta}(y_1, \ldots, y_{67}) &= \frac{67}{\sum_{i=1}^{67} y_i} \\
&= \frac{67}{107} \\
&\approx 0.6262
\end{aligned}$$

4. Hesse-matricen er en skalar i dette tilfælde og givet ved den anden-afledte. Vi får at

bidraget for par i er

$$H_i(\theta) = \frac{\partial^2 \log(\ell(\theta|y_i))}{\partial^2 \theta} = \frac{\partial s_i(\theta)}{\partial \theta} = -\frac{1}{\theta^2} - \frac{Y_i - 1}{(1-\theta)^2}$$

og dermed er informationen

$$\begin{aligned}
I(\theta_0) &= \mathbb{E}(-H_i(\theta_0)) \\
&= \mathbb{E}\left(\frac{1}{\theta_0^2} + \frac{Y_i - 1}{(1-\theta_0)^2}\right) \\
&= \frac{1}{\theta_0^2} + \frac{\mathbb{E}(Y_i) - 1}{(1-\theta_0)^2} \\
&= \frac{1}{\theta_0^2} + \frac{\frac{1}{\theta_0} - 1}{(1-\theta_0)^2} \\
&= \frac{1}{\theta_0^2(1-\theta_0)}
\end{aligned}$$

Ved at indsætte vores estimat fås

$$\begin{aligned}
I(\hat{\theta}_n) &= \frac{1}{\hat{\theta}_n^2(1-\hat{\theta}_n)} \\
&= \frac{1}{0.6262^2(1-0.6262)} \\
&\approx 6.8224
\end{aligned}$$

således at variansen

$$\begin{aligned}
Var(\hat{\theta}) &= \frac{1}{n} I(\theta_0)^{-1} \\
&= \frac{\theta_0^2(1-\theta_0)}{n}
\end{aligned}$$

kan approksimeres som

$$\begin{aligned}
Var(\hat{\theta}) &\approx \frac{1}{n} I(\hat{\theta}_n)^{-1} \\
&= \frac{1}{67 \cdot 6.8224} \\
&= 0.0022
\end{aligned}$$

Det ses at standardafvigelsen bliver $se(\hat{\theta}) = \sqrt{Var(\hat{\theta})} = \sqrt{0.0022} \approx 0.0469$.
[opgaveformuleringen var uklar, idet der blev bedt om at udlede variansen på *estima-*

toren mens denne blev referret til som $Var(\hat{\theta}_n)$, hvor $\hat{\theta}_n$ er *estimatet*.]

5. Vi kan bruge den estimerede model til at beregne sandsynligheden for, at et par skal føde 3 drengebørn før de får en pige:

$$P(Y=4) = p(4) = (1-\hat{\theta}_n)^3\hat{\theta}_n = .3738^3 \cdot 0.6262 \approx 0.0327$$

6. Vi skal teste om $\mathbb{E}(Y_i) = \frac{1}{\theta} = 2$ ved hjælp af et Wald test. Det svarer til restriktionen $\theta_0 = 0.5$ og vi kan opstille vores nul-hypotese

$$\mathcal{H}_0 : \theta_0 = 0.5$$

og alternativ hypotese

$$\mathcal{H}_A : \theta_0 \neq 0.5.$$

Vi beregner vores z-statistik som

$$z_n(\theta_0 = 0.5) = \frac{\hat{\theta}_n - 0.5}{se(\hat{\theta})} = \frac{0.6262 - 0.5}{0.0469} \approx 2.6908.$$

Vi ved at $z_n(\theta_0 = 0.5) \overset{a}{\sim} \mathcal{N}(0,1)$ under $\mathcal{H}_0$. Så vi kan beregne den kristiske værdi på et 5% signifikans-niveau, $\alpha = 0.05$, som $c = \Phi^{-1}(0.975) \approx 1.96$ (to-sidet test). Da $z_n > |c|$ kan vi afvise på et 5% signifikansniveau, at det forventede antal børn er 2. (p-værdien er $2 \cdot (1 - \Phi(2.6908)) \approx 0.0071$, hvilket er noget lavere end de 5%)

7. Log-likelihood funktionen for den betingede model er

$$\begin{aligned}\log L_n(\theta,\delta) &= \log L(\theta,\delta|y_1,\ldots,y_{67},x_1,\ldots,x_{67}) \\ &= \log\left(\prod_{i=1}^{67}(1-[\theta+\delta x_i])^{y_i-1}[\theta+\delta x_i]\right) \\ &= \sum_{i=1}^{67}\{(y_i-1)\log(1-[\theta+\delta x_i]) + \log(\theta+\delta x_i)\}\end{aligned}$$

og δ måler nu, hvor meget mere det er sandsynligt for en kvinde på 40 år eller mere at få et pigebarn.

8. Vi kunne estimere den betingede model i STATA ved at skrive

```
mlexp ((y-1)*log(1-{theta}-{delta}*x) + log({theta}+{delta}*x))
```