Eksamen på Økonomistudiet vinteren 2018-19

Sandsynlighedsteori og Statistik

2. årsprøve

13. februar, 2019

(3-timers prøve med hjælpemidler)

Rettevejledning

Opgave 1

1. Den betingede sandsynlighed for at få en pige hvis man i forvejen har en dreng er

$$P(X_1 = 1|X_2 = 1) = \frac{P(X_1 = 1, X_2 = 1)}{P(X_2 = 1)} = \frac{0.2}{0.55} \approx 0.3636$$

2. X_1 og X_2 er IKKE uafhængige. Dette kan eksempelvis ses ved at den simultane sandsynlighed $P(X_1 = 1, X_2 = 1) = 0.2$ ikke er lig produktet af de marginale sandsynligheder

$$\begin{aligned} P(X_1 = 1) \cdot P(X_2 = 1) &= 0.45 \cdot 0.55 \approx 0.2475 \\ &> P(X_1 = 1, X_2 = 1) = 0.2 \end{aligned}$$

   og vi kan se, at der er en tendens til en lavere sandsynlighed for at få et barn, hvis man allerede har ét barn.

3. Vi lader $Y = X_1 + X_2$ angive antallet af børn i en husholdning. Vi har så, at det forventede antal børn er

$$\mathbb{E}(Y) = \mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2) = P(X_1 = 1) + P(X_2 = 1) = 0.45 + 0.55 = 1$$

4. Vi lader omkostningerne være givet ved $Z = aX_1 + bX_2 + cX_1X_2$. Den forventede omkostning er så

$$\begin{aligned} \mathbb{E}(Z) &= \mathbb{E}(aX_1 + bX_2 + cX_1X_2) \\ &= a\mathbb{E}(X_1) + b\mathbb{E}(X_2) + c\mathbb{E}(X_1X_2) \\ &= aP(X_1 = 1) + bP(X_2 = 1) + c\mathbb{E}(X_1X_2) \\ &= 0.45a + 0.55b + 0.2c \end{aligned}$$

   da

$$\mathbb{E}(X_1X_2) = \sum_{i=0}^{1} \sum_{j=0}^{1} i \cdot j \cdot p(i, j) = p(1, 1) = 0.2$$

Opgave 2

1. Tæthedsfunktionen for hhv. X og Y er

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(- \frac{(x-\mu_X)^2}{2\sigma_X^2} \right) = \frac{1}{\sqrt{0.2\pi}} \exp\left(- \frac{(x-1)^2}{0.2} \right)$$
$$p_Y(y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(- \frac{(y-\mu_Y)^2}{2\sigma_Y^2} \right) = \frac{1}{\sqrt{2\pi}} \exp\left(- \frac{(y+0.5)^2}{2} \right)$$

2. Vi lader nu $Z = 0.2 \cdot X + 0.1 \cdot Y$. Vi har middelværdi

$$\mu_Z = \mathbb{E}(Z) = \mathbb{E}(0.2X + 0.1Y) = 0.2\mathbb{E}(X) + 0.1\mathbb{E}(Y) = 0.2\mu_X + 0.1\mu_Y = 0.2 - 0.05 = 0.15$$

og varians

$$\begin{aligned} \sigma_Z^2 = Var(Z) &= Var(0.2X + 0.1Y) \\ &= 0.2^2 Var(X) + 0.1^2 Var(Y) + 0.2 \cdot 0.1 Cov(X, Y) \\ &= 0.04\sigma_X^2 + 0.01\sigma_Y^2 \\ &= 0.04 \cdot 0.1 + 0.01 \\ &= 0.014 \end{aligned}$$

Da Z er summen af to normalfordelinger er Z også normalfordelt, $Z \sim N(\mu_Z, \sigma_Z^2) = N(0.15, 0.014)$. (Tæthedsfunktionen er givet ved)

$$p_Z(z) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp\left(- \frac{(z-\mu_Z)^2}{2\sigma_Z^2} \right) = \frac{1}{\sqrt{0.028\pi}} \exp\left(- \frac{(z-0.15)^2}{0.028} \right)$$

3. Vi kan beregne kovarianserne og korrelationerne som

$$\begin{aligned}
Cov(Z,X) &= Cov(0.2 \cdot X + 0.1 \cdot Y, X) \\
&= Cov(0.2 \cdot X, X) + Cov(0.1 \cdot Y, X) \\
&= 0.2 \cdot Cov(X,X) \\
&= 0.2 \cdot Var(X) \\
&= 0.2 \cdot 0.1 \\
&= 0.02
\end{aligned}$$

$$corr(Z,X) = \frac{Cov(Z,X)}{\sqrt{Var(Z)Var(X)}} = \frac{0.02}{\sqrt{0.014 \cdot 0.1}} = 0.5345$$

$$\begin{aligned}
Cov(Z,Y) &= Cov(0.2 \cdot X + 0.1 \cdot Y, Y) \\
&= Cov(0.2 \cdot X, Y) + Cov(0.1 \cdot Y, Y) \\
&= 0.1 \cdot Cov(Y,Y) \\
&= 0.1 \cdot Var(Y) \\
&= 0.1
\end{aligned}$$

$$corr(Z,Y) = \frac{Cov(Z,Y)}{\sqrt{Var(Z)Var(Y)}} = \frac{0.1}{\sqrt{0.014 \cdot 1}} = 0.8452$$

4. Den betingede middelværdi er

$$\begin{aligned}
\mathbb{E}(Z|X=1) &= \mathbb{E}(0.2 \cdot X + 0.1 \cdot Y|X=1) \\
&= \mathbb{E}(0.2 \cdot X|X=1) + \mathbb{E}(0.1 \cdot Y|X=1) \\
&= 0.2 \cdot 1 + 0.1 \cdot \mathbb{E}(Y) \\
&= 0.2 - 0.1 \cdot 0.5 \\
&= 0.15
\end{aligned}$$

Opgave 3

1. Fordelingsfunktionen er for $y \in [0,1]$,

$$P(Y \leq y) = \int_{-\infty}^{y} p(y)\mathbf{1}(y \in [0,1])dy = \int_{0}^{y} \beta x^{\beta-1}dx = [x^{\beta}]_0^y = y^{\beta}$$

dvs. vi har at fordelingsfunktionen er

$$F(y) = \begin{cases} 0 & \text{hvis } y < 0 \\ y^{\beta} & \text{hvis } y \in [0,1] \\ 1 & \text{hvis } y > 1 \end{cases}$$

2. Vi har, at likelihood bidragende for hver container er $\ell(\beta|y_i) = p_Y(y_i) = \beta y_i^{\beta-1}$ og log-likelihood bridraget er $\log(\ell(\beta|y_i)) = \log\beta + (\beta-1)\log(y_i)$. Log-likelihood funktionen bliver, grundet uafhænighed mellem containerne, således

$$\begin{aligned} \log L_n(\beta) = \log L(\beta|y_1,\ldots,y_{82}) &= \sum_{i=1}^{n} \log \ell(\beta|y_i) \\ &= \sum_{i=1}^{n} [\log(\beta) + (\beta-1)\log(y_i)] \\ &= n \cdot \log(\beta) + (\beta-1)\sum_{i=1}^{n}\log(y_i) \end{aligned}$$

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

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

hvor vi her har at

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

således at maximum likelihood **estimatoren**, $\hat{\beta}$, kan findes som løsningen til ligningen

$$\begin{gathered}
\frac{n}{\hat{\beta}} + \sum_{i=1}^{n} \log(Y_i) = 0 \\
\Updownarrow \\
n + \hat{\beta} \sum_{i=1}^{n} \log(Y_i) = 0 \\
\Updownarrow \\
\hat{\beta} = -\frac{n}{\sum_{i=1}^{n} \log(Y_i)}
\end{gathered}$$

Ved at bruge $n = 82$ og $\frac{1}{82}\sum_{i=1}^{82} \log(y_i) = -2.809$ kan vi udlede **estimatet** for vores data som

$$\begin{aligned}
\hat{\beta}_n = \hat{\beta}(y_1, \ldots, y_{82}) &= -\frac{82}{\sum_{i=1}^{82} \log(y_i)} \\
&= -\frac{1}{\frac{1}{82}\sum_{i=1}^{82} \log(y_i)} \\
&= -\frac{1}{-2.809} \\
&\approx 0.356 \approx 0.360
\end{aligned}$$

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

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

og dermed er informationen

$$
\begin{aligned}
I(\beta_0) &= \mathbb{E}(-H_i(\beta_0)) \\
&= \mathbb{E}\left(\frac{1}{\beta_0^2}\right) \\
&= \frac{1}{\beta_0^2}
\end{aligned}
$$

Ved at indsætte vores estimator og information fra teksten fås

$$
\begin{aligned}
I(\hat{\beta}_n) &= \frac{1}{\hat{\beta}_n^2} \\
&= \frac{1}{0.360^2} \\
&\approx 7.716
\end{aligned}
$$

således at variansen

$$
\begin{aligned}
Var(\hat{\beta}) &= \frac{1}{n} I(\beta_0)^{-1} \\
&= \frac{\beta_0^2}{n}
\end{aligned}
$$

kan approksimeres som

$$
\begin{aligned}
Var(\hat{\beta}) &\approx \frac{1}{n} I(\hat{\beta}_n)^{-1} \\
&= \frac{1}{82 \cdot 7.716} \\
&\approx 0.0016
\end{aligned}
$$

Det ses at standardafvigelsen bliver $se(\hat{\beta}) = \sqrt{Var(\hat{\beta})} = \sqrt{0.0016} \approx 0.040$.

5. Vi kan bruge den estimerede model til at beregne sandsynligheden for at højest 20 procent er defekte:

$$
P(Y \leq 0.2) = \int_{-\infty}^{0.2} p(y)\mathbf{1}(y \in [0,1])dy = \int_0^{0.2} \beta x^{\beta-1} dx = [x^\beta]_0^{0.2} = 0.2^\beta
$$

og hvis vi indsætter vores estimat får vi $P(Y \leq 0.2) = 0.2^{\hat{\beta}_n} = 0.2^{0.360} \approx 0.560$. Dette findes også ved direkte indsættelse i svaret på spørgsmål 3.1.

6. Vi skal teste om $\mathbb{E}(Y_i) = \frac{\beta}{1+\beta} = 0.2$ ved hjælp af et Wald test. Det svarer til restriktionen

$\beta_0 = 0.25$ og vi kan opstille vores nul-hypotese

$$\mathcal{H}_0 : \beta_0 = 0.25$$

og alternativ hypotese

$$\mathcal{H}_A : \beta_0 \neq 0.25.$$

Vi beregner vores z-statistik som

$$z_n(\beta_0 = 0.25) = \frac{\hat{\beta} - 0.25}{se(\hat{\beta})} = \frac{0.360 - 0.25}{0.040} \approx 2.750.$$

Vi ved at $z_n \overset{a}{\sim} \mathcal{N}(0,1)$ hvis $\mathcal{H}_0$ er sand, 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 den forventede andel defekte iPhones er 20%. (p-værdien er $2 \cdot (1 - \Phi(2.750)) \approx 0.0060$, hvilket er noget lavere end de 5%)

7. Log-likelihood funktionen for den betingede model er

$$\begin{aligned}\log L_n(\beta, \delta) &= \log L(\beta, \delta | y_1, \ldots, y_{82}, x_1, \ldots, x_{82}) \\ &= \log \left(\prod_{i=1}^{82} \beta \cdot (1 + \delta x_i) y_i^{\beta(1+\delta x_i)-1} \right) \\ &= \sum_{i=1}^{82} \{ \log(\beta \cdot (1 + \delta x_i)) + [\beta \cdot (1 + \delta x_i) - 1] \log(y_i) \}\end{aligned}$$

og $100 \cdot \delta$ måler nu hvor mange procent højere parameteren i beta-fordelingen er for virksomheden AlwaysOnTime.

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

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