Eksamen på Økonomistudiet sommeren 2019

Sandsynlighedsteori og Statistik

2. årsprøve

12. Juni, 2019

(3-timers prøve med hjælpemidler)

RETTEVEJLEDNING

Opgaven består af tre delopgaver, som alle skal besvares. De tre opgaver kan regnes uafhængigt af hinanden. Opgave 1 og 2 indgår tilsammen med samme vægt som opgave 3.

Opgave 1

1. $P(X = 2|Y = 1) = \frac{P(X=2,Y=1)}{P(Y=1)} = \frac{0.1}{0.4+0.1} = 0.2.$
2. $\mathbb{E}[X|Y = 1] = 1 \cdot P(X = 1|Y = 1) + 2 \cdot P(X = 2|Y = 1) = \frac{P(X=1,Y=1)}{P(Y=1)} + 2 \cdot 0.2 = \frac{0.4}{0.5} + 0.4 = 1.2.$
3. Den ubetingede middelværdi af produktet er:

$$\begin{aligned} \mathbb{E}[X \cdot Y] &= 1 \cdot 1 \cdot P(X = 1, Y = 1) \\ &+ 2 \cdot 1 \cdot P(X = 2, Y = 1) \\ &+ 1 \cdot 2 \cdot P(X = 1, Y = 2) \\ &+ 2 \cdot 2 \cdot P(X = 2, Y = 2) \\ &= 0.4 + 0.2 + 0.4 + 1.2 \\ &= 2.2 \end{aligned}$$

.

4. De to stokastiske variable er IKKE uafhængige. Det kan bl.a. ses ved at

$$P(X = 1, Y = 1) = 0.4 \neq 0.3 = P(X = 1) \cdot P(Y = 1)$$

sådan at den simultane sandsynlighed ikke er lig produktet af de marginale.

Opgave 2

1. For $x \leq 20$ har vi

$$\begin{aligned} P(X \leq x) &= \int_{-\infty}^{x} p(y)dy \\ &= \int_{-10}^{x} \frac{1}{30} dy \\ &= \frac{1}{30}[y]_{-10}^{x} \\ &= \frac{x + 10}{30} \end{aligned}$$

sådan at vi har Fordelingsfunktionen:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{hvis } x < -10 \\ \frac{x+10}{30} & \text{hvis } x \in [-10, 20] \\ 1 & \text{hvis } x > 20 \end{cases}$$

2. Vi kan beregne sandsynligheden ved at bruge Fordelingsfunktionen:

$$\begin{aligned} P(X \in [0, 10]) &= P(X \leq 10) - P(X < 0) \\ &= P(X \leq 10) - P(X \leq 0) + \underbrace{P(X = 0)}_{=0} \\ &= \frac{10 + 10}{30} - \frac{0 + 10}{30} \\ &= \frac{1}{3} \end{aligned}$$

hvilket vi også kan se må være tilfældet, da intervallet $[0, 10]$ dækker én tredjedel af det samlede interval for uniform-fordelingen.

3. Vi beregner den betingede middelværdi

$$\begin{aligned} \mathbb{E}[X | X \in [0, 10]] &= \int_{-\infty}^{\infty} x \cdot p(x | x \in [0, 10]) dx \\ &= \int_{-\infty}^{\infty} x \cdot p(x) \cdot \mathbf{1}\{x \in [0, 10]\} / P(X \in [0, 10]) dx \\ &= P(X \in [0, 10])^{-1} \int_{0}^{10} x \cdot \frac{1}{30} \cdot 3 dx \\ &= \frac{1}{10} [\frac{1}{2} x^2]_0^{10} \\ &= \frac{100}{20} \\ &= 5 \end{aligned}$$

hvilket igen også er intuitivt, da det er midtpunktet i den betingede uniform fordeling.

4. Den betingede middelværdi er $\mathbb{E}[Y | X \in [0, 10]] = \mathbb{E}[2 \cdot X - 10 | X \in [0, 10]] = 2 \cdot \mathbb{E}[X | X \in [0, 10]] - 10 = 2 \cdot 5 - 10 = 0$.

Opgave 3

1. Vi har at likelihood bidragene for hver kasse er $\ell(\theta|y_i) = p(y_i) = \frac{6!}{(6-y_i)!\cdot y_i!}\theta^{y_i}(1-\theta)^{6-y_i}$ og log-likelihood bridraget er $\log(\ell(\theta|y_i)) = \log(\frac{6!}{(6-y_i)!\cdot y_i!}) + y_i \log(\theta) + (6-y_i)\log(1-\theta)$. Log-likelihood funktionen bliver således

$$\begin{aligned}\log L_n(\theta) = \log L(\theta|y_1,\ldots,y_n) &= \sum_{i=1}^{n}[\log(\frac{6!}{(6-y_i)!\cdot y_i!}) + y_i\log(\theta) + (6-y_i)\log(1-\theta)] \\ &= n\log(\frac{6!}{(6-y_i)!\cdot y_i!}) + \log(\theta)\sum_{i=1}^{n} y_i + \log(1-\theta)(6n - \sum_{i=1}^{n} y_i)\end{aligned}$$

2. 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

$$\begin{aligned}\frac{\partial \log L_n(\theta)}{\partial\theta} &= \sum_{i=1}^{n}\frac{\partial \log(\ell(\theta|Y_i))}{\partial\theta} \\ &= \sum_{i=1}^{n} s_i(\theta) \\ &= \sum_{i=1}^{n}\left[\frac{Y_i}{\theta} - \frac{6-Y_i}{1-\theta}\right] \\ &= \frac{(\sum_{i=1}^{n} Y_i)}{\theta} - \frac{6n - (\sum_{i=1}^{n} Y_i)}{1-\theta}\end{aligned}$$

sådan at **estimatoren** $\hat{\theta}(Y_1, \ldots, Y_n)$ kan udledes til at være

$$S(\hat{\theta}) = 0$$
$$\Updownarrow$$
$$\frac{(\sum_{i=1}^n Y_i)}{\hat{\theta}} = \frac{6n - (\sum_{i=1}^n Y_i)}{1-\hat{\theta}}$$
$$\Updownarrow$$
$$(1-\hat{\theta})\sum_{i=1}^n Y_i = \hat{\theta}(6n - \sum_{i=1}^n Y_i)$$
$$\Updownarrow$$
$$\sum_{i=1}^n Y_i = \hat{\theta}6n$$
$$\Updownarrow$$
$$\hat{\theta} = \frac{1}{6}\frac{1}{n}\sum_{i=1}^n Y_i.$$

Ved at indsætte informationen givet i opgaven får vi **estimatet**

$$\begin{aligned}\hat{\theta}_n = \hat{\theta}(y_1, \ldots, y_{101}) =& \frac{1}{6}\frac{1}{101}\sum_{i=1}^{101} y_i \\ =& \frac{167}{606} \\ \approx& 0.2760.\end{aligned}$$

Der er altså ca. 27.6% sandsynlighed for, at en flaske er udrikkelig.

3. Hesse-matricen er en skalar i dette tilfælde og givet ved den anden-afledte.

$$H_i(\theta) = \frac{\partial^2 \log(\ell(\theta|y_i))}{\partial^2\theta} = \frac{\partial s_i(\theta)}{\partial\theta} = -\frac{Y_i}{\theta^2} - \frac{6-Y_i}{(1-\theta)^2}$$

og dermed er informationen

$$
\begin{aligned}
I(\theta_0) &= \mathbb{E}(-H_i(\theta_0)) \\
&= \mathbb{E}\left(\frac{Y_i}{\theta_0^2} + \frac{6 - Y_i}{(1-\theta_0)^2}\right) \\
&= \frac{\mathbb{E}(Y_i)}{\theta_0^2} + \frac{6 - \mathbb{E}(Y_i)}{(1-\theta_0)^2} \\
&= \frac{6\theta_0}{\theta_0^2} + \frac{6 - 6\theta_0}{(1-\theta_0)^2} \\
&= \frac{6}{\theta_0} + \frac{6}{(1-\theta_0)} \\
&= \frac{6}{\theta_0(1-\theta_0)}
\end{aligned}
$$

Ved at indsætte vores estimat fås

$$
\begin{aligned}
I(\hat{\theta}_n) &= \frac{6}{\hat{\theta}_n(1-\hat{\theta}_n)} \\
&= \frac{6}{0.276(1-0.276)} \\
&\approx 30.055
\end{aligned}
$$

således at variansen

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

kan approksimeres som

$$
\begin{aligned}
Var(\hat{\theta}) &\approx \frac{1}{n} I(\hat{\theta}_n)^{-1} \\
&= \frac{1}{101 \cdot 30.055} \\
&= 0.00033
\end{aligned}
$$

Det ses at standardafvigelsen bliver $se(\hat{\theta}) = \sqrt{Var(\hat{\theta})} = \sqrt{0.00033} \approx 0.0182$.

4. Vi bliver bedt om at beregne

$$\begin{aligned}
P(Y_i \leq 1) &= p(0) + p(1) \\
&= \frac{6!}{(6-0)! \cdot 0!} \hat{\theta}_n^0 (1-\hat{\theta}_n)^{6-0} + \frac{6!}{(6-1)! \cdot 1!} \hat{\theta}_n^1 (1-\hat{\theta}_n)^{6-1} \\
&= (1-\hat{\theta}_n)^6 + 6\hat{\theta}_n (1-\hat{\theta}_n)^5 \\
&= (1-0.276)^6 + 6 \cdot 0.276 \cdot (1-0.276)^5 \\
&= 0.1445 + 0.3298789 \\
&= 0.4744
\end{aligned}$$

dvs der er ca. 47.5% sandsynlighed for, at der højest er 1 udrikkelig flaske i en tilfældig kasse med 6 flasker.

5. Vi skal teste om $\mathbb{E}(Y_i) = 1.8$. Vi ved at $\mathbb{E}(Y_i) = 6\theta$ så det svarer til restriktionen $\theta_0 = 1.8/6 = 0.3$ og vi kan opstille vores nul-hypotese

$$\mathcal{H}_0 : \theta_0 = 0.3$$

og alternativ hypotese

$$\mathcal{H}_A : \theta_0 \neq 0.3.$$

Vi beregner vores z-statistik som

$$z_n(\theta_0 = 0.3) = \frac{\hat{\theta}_n - 0.3}{se(\hat{\theta})} = \frac{0.276 - 0.3}{0.0182} \approx -1.319.$$

Vi ved at $z_n(\theta_0 = 0.3) \overset{a}{\sim} \mathcal{N}(0,1)$ under $\mathcal{H}_0$. Vi kan derfor 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 IKKE afvise på et 5% signifikansniveau, at det forventede antal udrikkelige flasker i en tilfældig kasse er 1.8. (p-værdien er $2 \cdot (1 - \Phi(1.319)) \approx 0.1873$, hvilket er noget højere end de 5%)

6. Vi har nu, at log-likelihood funktionen er

$$\begin{aligned}\log L_n(\theta,\delta) &= \log L(\theta,\delta|y_1,\ldots,y_{101},z_1,\ldots,z_{101})\\ &= \log\left(\prod_{i=1}^{101}\frac{6!}{(6-y_i)!\cdot y_i!}(\theta+\delta z_i)^{y_i}(1-\theta-\delta z_i)^{6-y_i}\right)\\ &= 101\cdot\log(\frac{6!}{(6-y_i)!\cdot y_i!})+\sum_{i=1}^{101}\{y_i\log(\theta+\delta z_i)+(6-y_i)\log(1-\theta-\delta z_i)\}\end{aligned}$$

7. Vi får nu givet, at $L_u = -351.321$ og $L_r = -356.769$. Vi opstiller vores nul-hypotese

$$\mathcal{H}_0 : \delta_0 = 0$$

og alternativ hypotese

$$\mathcal{H}_A : \delta_0 \neq 0.$$

Vi kan beregne vores Likelihood Ratio (LR) test-størrelse som

$$LR(\delta_0 = 0) = 2\cdot(L_u - L_r) = 2\cdot(-351.321 + 356.769) = 10.896.$$

Vi ved, at under $\mathcal{H}_0$ er teststørrelsen asymptotisk χ^2 fordelt med 1 frihedsgrad, $LR(\delta_0 = 0) \overset{a}{\sim} \chi^2_1$. Så vi kan beregne den kristiske værdi på et 5% signifikans-niveau, $\alpha = 0.05$, som $c = (\chi^2_1)^{-1}(0.95) \approx 3.84$. Da $LR > 3.84$ kan vi afvise på et 5% signifikansniveau, at $\delta_0 = 0$.

Vi finder altså, at vi på et 5% signifikans niveau kan *afvise*, at der *ikke er forskel* på sandsynligheden for en udrikkelig flaske på tværs af leverandører. Resultatet tyder på, at den mistænkte leverandør har en større sandsynlighed for at levere udrikkelige flasker, sammenlignet med de andre leverandører.