Eksamen på Økonomistudiet vinteren 2016-17

Sandsynlighedsteori og Statistik

2. årsprøve

23. januar, 2017

(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

I denne opgave undersøges, hvordan et forsikringsselskab skal fastlægge sine præmier på forskellige forsikringsprodukter. Antag, at vi har et forsikringsselskab, som udbyder cykelforsikringer. Vi antager, at alle deres forsikringstagere har en cykel med værdien 5.000 kr. Sandsynligheden per år for at få stjålet en cykel er $p = 0.06$. Vi antager, at hver forsikringstager højst kan få stjålet én cykel om året. Forsikringsselskabet tilbyder en forsikring uden selvrisiko. Præmien for denne forsikring er 350 kr. om året. Vi antager, at forsikringsselskabet har 10 kunder, som har tegnet en forsikring uden selvrisiko. Lad X angive antallet af kunder som får stjålet en cykel i løbet af et år.

1. Sandsyndsynligheden for at 2 af de 10 kunder får deres cykel stjålet, $P(X = 2)$ udregnes ved brug af formlen for punktsandsynligheden for for binomialfordelingen. De studerende skal indse at X er binomialfordelt med antalsparameter $n = 10$ og sandsynlighedsparameter $p = 0.06$:

   $$P(X = 2) = \binom{10}{2}(0.06)^2(1 - 0.06)^{10-2} = 0.099.$$

   De studerende skal angive, at man her antager, at alle forsikringstagerne har samme sandsynlighed for at få stjålet en cykel og at der er uafhængighed mellem forsikringstagerne.

2. Forsikringsselskabets indtjening (i kr.), Z er givet ved

   $$Z = 350 \cdot 10 - 5000 \cdot X.$$

   Den forventede indtjening kan beregnes ved brug af regneregler for middelværdi og formlen for middelværdien i en binomialfordeling:

   $$\begin{aligned} E(Z) &= 350 \cdot 10 - 5000 \cdot E(X) \\ &= 3500 - 5000 \cdot n \cdot p \\ &= 3500 - 5000 \cdot 10 \cdot 0.06 \\ &= 500 \end{aligned}$$

Variansen på indtjeningen er

$$\begin{aligned}
Var(Z) &= (-5000)^2 Var(X) \\
&= (-5000)^2 \cdot n \cdot p \cdot (1-p) \\
&= (-5000)^2 \cdot 10 \cdot 0.06 \cdot (1-0.06) \\
&= 1.41 \cdot 10^7.
\end{aligned}$$

Standardafvigelsen på Z er

$$\sqrt{Var(Z)} = \sqrt{1.41 \cdot 10^7} = 3755$$

3. Forsikringsselskabet har også en anden type forsikring med en selvrisiko på 1.000 kr. Prisen for denne forsikring er 290 kr. Antag, at 10 nye forsikringstagere vælger denne forsikring. Det antages, at de har samme sandsynlighed for at få stjålet en cykel: $p = 0.06$. Lad Y angive antallet af stjålne cykler for de nye forsikringstagere (med en selvrisiko). Sandsynlighed for at der i alt bliver stjålet 4 cykler blandt de 20 forsikringstagere, $P(X+Y=4)$ kan beregnes ved at anvende formlen for summen af to binomialfordelinger: $X \sim Bin(10, 0.06)$ og $Y \sim Bin(10, 0.06)$ det følger så at $X + Y \sim Bin(20, 0.06)$,

$$P(X+Y=4) = \binom{20}{4}(0.06)^4(1-0.06)^{20-4} = 0.023.$$

4. Den forventede indtjening på de 10 nye kunder kan bestemmes som $W = 290 \cdot 10 - 4000 \cdot Y$

$$E(W) = 2900 - 4000 \cdot 10 \cdot 0.06 = 500.$$

Variansen og standardafvigelsen for fortjeneste på denne forsikring er

$$\begin{aligned}
Var(W) &= (4000)^2 \cdot 10 \cdot 0.06 \cdot (1-0.06) \\
&= 9.024 \cdot 10^6. \\
\sqrt{Var(W)} &= 3004.
\end{aligned}$$

Det kan være attraktivt for forsikringsselskabet at udbyde forsikring med selvrisiko, selvom den forventede indtjening er den samme for begge typer forsikring, da det giver en mindre varians på indtjeningen.

Opgave 2

1. Lad X være normalfordelt $N\left(0, \sigma_x^2\right)$. Find fordelingen af $Z_1 = X + 1$.

   Facit: Z_1 er $N\left(1, \sigma_x^2\right)$.

2. Lad Y være normalfordelt $N\left(0, \sigma_y^2\right)$. Find fordelingen af $Z_2 = 2Y$.

   Facit: Z_2 er $N\left(0, 4\sigma_y^2\right)$ fordelt.

3. Antag at X og Y er uafhængige. Vis at (forklar hvilke regneregler du benytter):

$$\begin{aligned} E\left(Z_1 | Z_2 = z_2\right) &= 1. \\ Var\left(Z_2 | Z_1 = z_1\right) &= 4\sigma_y^2. \end{aligned}$$

   Facit: 1) $E\left(Z_1 | Z_2 = z_2\right) = E\left(Z_1\right) = 1$ pga uafhængighed. 2) Tilsvarende: $Var\left(Z_2 | Z_1 = z_1\right) = Var\left(Z_2\right) = 4\sigma_y^2$.

4. Under antagelsen at X og Y er uafhængige, er (X, Y) $N(\mu, \Omega)$ hvor

$$\mu = (0,0)' \text{ og } \Omega = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}.$$

   Definer nu vektoren $W = \left(W_1, W_2\right)'$ ved

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} X \\ X + 2Y \end{pmatrix}.$$

   Find $E\left(W_1 W_2\right)$ og $E\left(W_2 | W_1\right)$.

   Facit: $E\left(W_1 W_2\right) = E\left(X^2 + 2XY\right) = E(X^2) + 2E(X)E(Y) = E(X^2) = Var(X) = \sigma_x^2$. Og

$$E\left(W_2 | W_1\right) = E\left(X + 2Y | X\right) = X + 2E(Y) = X.$$

5. Vis at $\left(W_1, W_2\right)$ er normalfordelt,

$$N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_x^2 \\ \sigma_x^2 & 4\sigma_y^2 \end{pmatrix}\right).$$

Facit: I denne opgave er der en fejl.Variansen af W_2 er $\sigma_x^2 + 4\sigma_y^2$. Med $A = \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}$ er det $(A) = 2$ så fra Sætning 6.3.1 eller Regnergel G.5 at

$$\begin{aligned}\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} &= A \begin{pmatrix} X \\ Y \end{pmatrix} = N\left(0, A\Omega A'\right) \\ &= N\left(0, \begin{pmatrix} \sigma_x^2 & 0 \\ \sigma_x^2 & 2\sigma_y^2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}\right) \\ &= N\left(0, \begin{pmatrix} \sigma_x^2 & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2 + 4\sigma_y^2 \end{pmatrix}\right).\end{aligned}$$

Opgave 3

1. Parameterrummet for en sandsynlighed er $\theta \in \Theta =]0,1[$.

   Likelihood bidraget er

   $$\ell\left(\theta \mid y_i\right) = f_{Y_i}\left(y_i \mid \theta\right) = \left(y_i - 1\right)\theta^2 (1-\theta)^{(y_i - 2)},$$

   så

   $$\log \ell\left(\theta \mid y_i\right) = \log\left(y_i - 1\right) + 2\log(\theta) + \left(y_i - 2\right)\log(1-\theta).$$

   Den samlede log-likelihood funktion er så

   $$\log L_n(\theta) = \sum_{i=1}^n \log \ell\left(\theta \mid y_i\right) = \sum_{i=1}^n \left\{\log\left(y_i - 1\right) + 2\log(\theta) + \left(y_i - 2\right)\log(1-\theta)\right\}.$$

2. Scorebidraget er givet ved

   $$s_i(\theta) = \frac{\partial \log \ell\left(\theta \mid y_i\right)}{\partial \theta} = \frac{2}{\theta} - \frac{y_i - 2}{1-\theta},$$

   så scoren er

   $$S(\theta) = \sum_{i=1}^n \left\{\frac{2}{\theta} - \frac{y_i}{1-\theta}\right\} = \frac{2n}{\theta} - \frac{\sum_{i=1}^n y_i - 2n}{1-\theta}.$$

   Så er førsteordensbetingelsen givet ved

   $$S(\hat{\theta}_n) = 0,$$

sådan at $\hat{\theta}_n$ er bestemt som

$$\begin{aligned}\frac{2n}{\hat{\theta}_n} &= \frac{\sum_{i=1}^n y_i - 2n}{1-\hat{\theta}_n} \\ 2n - 2n\hat{\theta}_n &= \hat{\theta}_n \sum_{i=1}^n y_i - 2n\hat{\theta}_n \\ \hat{\theta}_n &= \frac{2n}{\sum_{i=1}^n y_i}.\end{aligned}$$

Vi får oplyst at $\sum_{i=1}^n y_i = 748$, så estimatet bliver

$$\hat{\theta}_n = \frac{120}{748} = 0.160.$$

3. Bidraget til Hesse-matricen er defineret som

$$\begin{aligned}H_i(\theta) &= \frac{\partial^2 \log \ell(\theta \mid Y_i)}{\partial\theta\partial\theta} \\ &= \frac{\partial}{\partial\theta} s_i(\theta) \\ &= \frac{\partial}{\partial\theta}\left(\frac{2}{\theta} - \frac{Y_i - 2}{1-\theta}\right) \\ &= -\frac{2}{\theta^2} - \frac{Y_i - 2}{(1-\theta)^2}.\end{aligned}$$

Vi bruger nu, at

$$E(Y_i) = 2 + \frac{2(1-\theta_0)}{\theta_0},$$

og finder for informationen,

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

Det følger, at variansen på estimatoren er

$$V(\hat{\theta}_n) = \frac{I(\theta_0)^{-1}}{n} = \frac{\theta_0^2 (1-\theta_0)}{2n},$$

som kan approksimeres med

$$V(\hat{\theta}_n) \approx \frac{\hat{\theta}_n^2 \left(1-\hat{\theta}_n\right)}{kn} = \frac{0.160^2 (1-0.160)}{120},$$

og $\mathsf{se}(\hat{\theta}_n) = \sqrt{V(\hat{\theta}_n)} = \sqrt{\frac{0.160^2(1-0.160)}{120}} = 0.0134.$

4. Det forventede antal kontroller med estimatet indsat er

$$E(Y_i) = 2 + \frac{2\left(1-\hat{\theta}_n\right)}{\hat{\theta}_n} = 2 + \frac{2(1-0.160)}{0.160} = 12.5.$$

Hvis der er budgetteret med 10 kontroller, svarer det til

$$E(Y_i) = 2 + \frac{2(1-\theta)}{\theta} = 10$$

eller

$$\theta = 0.20.$$

5. For at teste hypotesen at $\theta_0 = 0.20$ anvendes et $z-$test. Der anvendes hypoteser

$$H_0 : \theta_0 = 0.20 \quad \text{mod} \quad H_A : \theta_0 \neq 0.20.$$

Test-statistikken er

$$z_n(\theta_0 = 0.20) = \frac{\hat{\theta}_n - 0.20}{\mathsf{se}(\hat{\theta}_n)} = \frac{0.160 - 0.20}{0.0134} = -2.99.$$

Hvis nul-hypotesen er korrekt, er $z_n(\theta_0 = 0.20)$ fordelt som en standard normal-fordeling, $N(0,1)$, så på $\alpha = 0.05$ signifikans-niveau er den kritiske værdi for $|z_n(\theta_0 = 0.20)|$ bestemt som 1.96. Hypotesen afvises ret klart.

6. Likelihood bidraget for den betingede model kan skrives som

$$\begin{aligned}
\ell(\alpha, \beta \mid y_i, x_i) &= f_{Y_i|X_i}(y_i \mid x_i; \alpha, \beta) \\
&= (y_i - 1)\,\theta_i^2 (1-\theta_i)^{(y_i-2)} \\
&= (y_i - 1)\,(\alpha + \beta x_i)^2\,(1 - (\alpha + \beta x_i))^{(y_i-2)}.
\end{aligned}$$

OxMetrics koden til estimation af den betingede model kan fx være

```
actual = Y;
theta  = &1 + &2*X;
fitted = theta;
loglik = log(Y-1)+2*log(theta)+(Y-2)*log(1-theta);
&1     = 0.5;
&2     = 0.1;
```