Eksamen på Økonomistudiet sommeren 2016
Sandsynlighedsteori og Statistik
2. årsprøve
14. juni, 2016
(3-timers prøve med hjælpemidler)

RETTEVEJLEDNING

Dette eksamenssæt består af 5 sider (forsiden inklusiv).

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.

Den simultane fordeling af X og Y er givet ved

	$X = 0$	$X = 1$	$X = 2$
$Y = 0$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
$Y = 1$	$\frac{1}{4}$	0	$\frac{1}{8}$
$Y = 2$	0	$\frac{1}{8}$	0

Opgave 1

1. Beregn middelværdi og varians af den marginale fordeling for Y. Dvs. beregn $E(Y)$ og $V(Y)$.

   $P(Y = 0) = \frac{4}{8}$; $P(Y = 1) = \frac{3}{8}$; $P(Y = 2) = \frac{1}{8}$;

   $E(Y) = 0 * \frac{4}{8} + 1 * \frac{3}{8} + 2 * \frac{1}{8} = \frac{5}{8} = 0,625$

   $E(Y^2) = 0^2 * \frac{4}{8} + 1^2 * \frac{3}{8} + 2^2 * \frac{1}{8} = \frac{7}{8}$

   $V(Y) = E(Y^2) - [E(Y)]^2 = \frac{7}{8} - (\frac{5}{8})^2 = \frac{56-25}{64} = \frac{31}{64} = 0,48$

2. Angiv fordelingen af Z hvor $Z = X * Y$. Udregn også $E(Z)$.

   $P(Z = 0) = \frac{6}{8}; P(Z = 1) = 0; P(Z = 2) = \frac{2}{8}; P(Z = 4) = 0;$

   $E(Z) = 0 * (\frac{6}{8}) + 2 * (\frac{2}{8}) = \frac{4}{8} = 0,5.$

3. Udregn den betingede middelværdi og den betingede varians af Y givet $X = 1$. dvs. udregn $E(Y|X = 1)$ og $V(Y|X = 1)$.

   $P(Y = 0|X = 1) = \frac{1}{2}; P(Y = 1|X = 1) = 0; P(Y = 2|X = 1) = \frac{1}{2};$

   $E(Y|X = 1) = 0 * (\frac{1}{2}) + 2 * (\frac{1}{2}) = 1$

   $E(Y^2|X = 1) = 0^2 * (\frac{1}{2}) + 2^2 * (\frac{1}{2}) = 2$

   $V(Y|X = 1) = 2 - 1^2 = 1$

4. Er X og Y uafhængige? Begrund svaret.

   Hvis der skulle være uafhængighed så skal der gælde i alle celler at

   $P(Y = 1, X = 1) = \quad P(Y = 1) * P(X = 1)$

   0=$P(Y = 1, X = 1)$ =? $\quad P(Y = 1) * P(X = 1) = (\frac{3}{8}) * (\frac{2}{8}) = \frac{6}{64}$

   altså ikke uafhængighed.

Opgave 2

I den sidste PISA undersøgelse foretaget i 2012 deltog 7.500 elever fra Danmark og 4.700 elever fra Sverige. De danske elevers score i matematik kan beskrives med en normalfordeling, der har middelværdi 500 og en spredning (standard afvigelse) på 80. Dvs. $N(500, 80^2)$. De svenske elevers score i matematik kan beskrives med en normalfordeling, der har middelværdi 480 og en spredning på 80. Dvs. $N(480, 80^2)$.Det antages, at alle elevers score er uafhængige af hinanden.

1. Beregn sandsynligheden for at en tilfældig svensk elev får en score, der er mindst 500.

   $X \sim N(480, 80^2).P(X => 500) = 1 - P(X <= 500) = 1 - P(\frac{X-480}{80} <= \frac{500-480}{80}) = 1 - \Phi(\frac{1}{4}) = \Phi(-\frac{1}{4}) = 0,40$

   I alt er der i de to lande 12.200 elever. Der er nu udvalgt en elev, hvis score er mindst 500.

2. Hvad er sandsynligheden for, at eleven kommer fra Sverige?

   P(tilfældig elev får mindst 500)=0,5* $\frac{7500}{12200} + 0,4 * \frac{4700}{12200} = 0,46$

   P(elev fra Sverige|mindst 500)=$\frac{0{,}4*\frac{4700}{12200}}{0{,}5*\frac{7500}{12200}+0{,}4*\frac{4700}{12200}} = \frac{0{,}15}{0{,}46} = 0,33$

   I Danmark udvælges de elever, der har en score på mindst 600. Antallet af elever der har en score på mindst 600 kaldes Z.

3. Hvilken fordeling kan beskrive Z? Beregn $E(Z)$.

   $Y \sim N(500, 80^2).P(Y => 600) = 1 - P(Y <= 600) = 1 - P(\frac{Y-500}{80} <= \frac{600-500}{80}) = 1 - \Phi(\frac{5}{4}) = \Phi(-\frac{5}{4}) = 0,11$

   $Z \sim Bin(7500; 0,11) \quad E(Z) = 7500 * 0,11 = 792$

   Eleverne er testet både i matematik og læsning. Læsescoren i Danmark kan beskrives med en normalfordeling med middelværdi 490 og en spredning på 80. Dvs. $N(490, 80^2)$.

Korrelationskoefficienten mellem læse- og matematikscoren er på 0,9. dvs

$$\rho = \frac{cov(X,Y)}{\sqrt{V(X) * V(Y)}} = 0,9.$$

Her angiver X læsescoren og Y matematikscoren.

Lidt uheldigt at der skiftes notation! X var tidligere matematik.

Undervisningsministeriet ønsker at anvende en samlet score som kaldes S. Den samlede score beregnes udfra formlen:

$S = \frac{1}{3}X + \frac{2}{3}Y$. Man ønsker altså at lægge mere vægt på matematik end læsning i dette samlede mål

1. Angiv fordelingen for S.

   $E(S) = E(\frac{1}{3}X + \frac{2}{3}Y) = E(\frac{1}{3}X) + E(\frac{2}{3}Y) = \frac{1}{3}E(X) + \frac{2}{3}E(Y) = \frac{1}{3} * 500 + \frac{2}{3} * 490 = 493,3$

   $V(S) = V(\frac{1}{3}X + \frac{2}{3}Y) = V(\frac{1}{3}X) + V(\frac{2}{3}Y) + 2COV(\frac{1}{3}X, \frac{2}{3}Y) = (\frac{1}{3})^2 V(X) + (\frac{2}{3})^2 V(Y) + 2(\frac{1}{3})(\frac{2}{3})(0,9) * (80)^2 = (\frac{1}{3})^2 80 + (\frac{2}{3})^2 80 + 2(\frac{1}{3})(\frac{2}{3})(0,9) * (80)^2 =$

   80*$((\frac{1}{3})^2 + (\frac{2}{3})^2 + 2(\frac{1}{3})(\frac{2}{3})(0,9) * (80)) = (51,0)^2$

   $S \sim N(496,7; (51,0)^2)$

Opgave 3

På to store indfaldsveje (A og B) til København har Vejdirektoratet i henholdsvis 35 og 33 uger registreret antallet af trafikuheld på en hverdag.

Registreringerne er angivet i nedenstående tabel.

antal trafikuheld	A	B
0	10	13
1	13	10
2	8	7
3	3	3
4	1	0
5+	0	0
i alt	35	33

Der opstilles følgende model:

$X_1, ...X_{35} \sim Poisson(\lambda)$ og $Y_1, ...Y_{33} \sim Poisson(\mu)$ alle stokastiske variable er uafhængige af hinanden.

Ud fra tabellen fås

$\sum_{i=1}^{35} X_i = 42$ og $\overline{X} = \frac{1}{35}\sum_{i=1}^{35} X_i = 1,2$ samt

$s^2 = \frac{1}{35-1}\sum_{i=1}^{35}(X_i - 1,2)^2 = 1,11$

$\sum_{i=1}^{33} Y_i = 33$ og $\overline{Y} = \frac{1}{33}\sum_{i=1}^{33} Y_i = 1,0$ samt

$\mathrm{s}^2 = \frac{1}{33-1}\sum_{i=1}^{33}(Y_i - 1,2)^2 = 1,00$

1. Argumenter for at det er en rimelig model, der opstilles. Angiv fordelingerne for $\sum_{i=1}^{35} X_i$ og $\sum_{i=1}^{33} Y_i$

   Poissonfordelingen er er fordeling der tæller ankomster og

   I Poissonfordelingen er middelværdien = variansen [E(X)=V(X)=λ].

   Summen af uafhængige poissonfordelinger er igen en poissonfordeling

   $\sum_{i=1}^{35} X_i \sim Poisson(35\lambda)$ og $\sum_{i=1}^{33} Y_i \sim Poisson(33\mu)$

2. Opskriv likelihood funktionen $L(\lambda,\mu)$ samt log-likelihood funktionen $log[L(\lambda,\mu)]$ for det samlede datamateriale $X_1, ...X_{35}$ og $Y_1, ...Y_{33}$

   dvs. for alle 68 observationer. Angiv scorefunktionerne og Hesse-matricen. Udregn MLE (maksimumlikelihood-estimaterne) for λ og μ.

   $L(\lambda,\mu) = \prod_{i=1}^{35} \frac{\lambda^{x_i}\exp(-\lambda)}{x_i!} \prod_{i=1}^{33} \frac{\mu^{y_i}\exp(-\mu)}{y_i!}$

   $log[L(\lambda,\mu)] = \sum_{i=1}^{35}[x_i \log(\lambda) - \lambda - \log(x_i!)] + \sum_{i=1}^{33}[y_i \log(\mu) - \mu - \log(y_i!)]$

   $$S(\lambda,\mu) = \begin{array}{|c|}\hline [\sum_{i=1}^{35} x_i] * \lambda^{-1} - 35 \\ \hline [\sum_{i=1}^{33} y_i] * \mu^{-1} - 33 \\ \hline\end{array}$$

   $$H(\lambda,\mu) = \begin{array}{|l|l|}\hline -[\sum_{i=1}^{35} x_i] * \lambda^{-2} & 0 \\ \hline 0 & -[\sum_{i=1}^{33} y_i] * \mu^{-2} \\ \hline\end{array}$$

   $$S(\lambda,\mu) = \begin{array}{|c|}\hline 0 \\ \hline 0 \\ \hline\end{array} \text{ giver } \begin{array}{|c|}\hline \widehat{\lambda} \\ \hline \widehat{\mu} \\ \hline\end{array} = \begin{array}{|c|}\hline \overline{x} \\ \hline \overline{y} \\ \hline\end{array} = \begin{array}{|c|}\hline 1,2 \\ \hline 1,0 \\ \hline\end{array}$$

3. Angiv et 95% (approximativt) konfidensinterval for λ.

   $V(\widehat{\lambda}) = \frac{\lambda}{n} \qquad \widehat{V(\widehat{\lambda})} = \frac{\widehat{\lambda}}{n} = \frac{1,2}{35} = (0,19)^2$

1,2+-1,96*0,19 [0,84 - 1,57]

Der ønskes nu undersøgt om $\lambda = \mu$. Kald den fælles intensitet for γ.

4. Vis at MLE for γ bliver $\frac{42+33}{35+33} = 1,13$. Test $H_0 : \lambda = \mu$ mod $H_A : \lambda \neq \mu$ ved brug af et likelihood ratio test.

$log[L(\gamma)] = \sum_{i=1}^{35}[x_i \log(\gamma) - \gamma - \log(x_i!)] + \sum_{i=1}^{33}[y_i \log(\gamma) - \gamma - \log(y_i!)] =$

$S(\gamma) = [\sum_{i=1}^{35} x_i] * \gamma^{-1} - 35 + [\sum_{i=1}^{33} y_i] * \gamma^{-1} - 33 =$

$[\sum_{i=1}^{35} x_i + \sum_{i=1}^{33} y_i] * \gamma^{-1} - (35 + 33)$

$S(\gamma) = 0$ giver $\widehat{\gamma} = \frac{\sum_{i=1}^{35} x_i + \sum_{i=1}^{33} y_i}{35+33} = \frac{42+33}{35+33} = 1,103$

$log[L(\widehat{\lambda}, \widehat{\mu})] = \sum_{i=1}^{35}[x_i \log(\widehat{\lambda}) - \widehat{\lambda} - \log(x_i!)] + \sum_{i=1}^{33}[y_i \log(\widehat{\mu}) - \widehat{\mu} - \log(y_i!)]$

$log[L(\widehat{\gamma})] = \sum_{i=1}^{35}[x_i \log(\widehat{\gamma}) - \widehat{\gamma} - \log(x_i!)] + \sum_{i=1}^{33}[y_i \log(\widehat{\gamma}) - \widehat{\gamma} - \log(y_i!)]$

$log[L(\widehat{\lambda}, \widehat{\mu})] - log[L(\widehat{\gamma})] =$

$\sum_{i=1}^{35} x_i * [\log(\widehat{\lambda}) - \log(\widehat{\gamma})] + \sum_{i=1}^{35} y_i * [\log(\widehat{\mu}) - \log(\widehat{\gamma})] +$

$-35\widehat{\lambda} - 33\widehat{\mu} + 35\widehat{\gamma} + 33\widehat{\gamma} = -35\widehat{\lambda} - 33\widehat{\mu} + (35 + 33) * \widehat{\gamma} =$

$42 * \log(\frac{1,2}{1,13}) + 33 * \log(\frac{1,0}{1,13}) - 42 - 33 + 42 + 33 =$

$42 * \log(\frac{1,2}{1,13}) + 33 * \log(\frac{1,0}{1,13}) = 0,30$

Dermed teststørerrelse på 2*0,30 =0,60 som er χ^2 med 1 frihedsgrad

p-værdi=62%

Antag nu at alle 68 observationer er Poissonfordelt med parameteren γ.

5. Test $H_0 : \gamma = 1$ mod $H_A : \gamma \neq 1$.

$\widehat{\gamma} = \frac{\sum_{i=1}^{35} x_i + \sum_{i=1}^{33} y_i}{35+33} = \frac{42+33}{35+33} = 1,103$

$V(\widehat{\gamma}) = \frac{\gamma}{35+33}$ $\widehat{V(\widehat{\gamma})} = \frac{\widehat{\gamma}}{35+33} = \frac{1,103}{68} = (0,13)^2$

wald test $= \frac{1,103-1}{0,13} = 0,81$ giver p-værdi =42%