Eksamen på Økonomistudiet sommeren 2017
Sandsynlighedsteori og Statistik

2. årsprøve

12. juni, 2017

udkast til rettevejledning

Der kan komme en mere forbedret rettevejledning senere

(3-timers prøve med hjælpemidler)

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.

Opgave 1

Den simultane fordeling af X og Y er givet ved

	$Y = 0$	$Y = 1$
$X = 0$	$0,1$	$0,2$
$X = 1$	$0,3$	$0,1$
$X = 2$	$0,1$	$0,2$

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

   $E(X) = 0 * 0,3 + 1 * 0,4 + 2 * 0,3 = 1$

   $E(X^2) = 0^2 * 0,3 + 1^2 * 0,4 + 2^2 * 0,3 = 1,6$

   $V(X) = E(X^2) - [E(X)]^2 = 1,6 - 1 = 0,6$

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

   Z's udfaldsrum {0,1,2,4,5}.

   $P(Z = 0) = P(X = 0, Y = 0) = 0,1.$

   $P(Z = 1) = P(X = 1, Y = 0) + P(X = 0, Y = 1) = 0,3 + 0,2 = 0,5.$

   $P(Z = 2) = P(X = 1, Y = 1) = 0,1.$

   $P(Z = 4) = P(X = 2, Y = 0) = 0,1.$

   $P(Z = 5) = P(X = 2, Y = 1) = 0,2.$

   $E(Z) = 0 * 0,1 + 1 * 0,5 + 2 * 0,1 + 4 * 0,1 + 5 * 0,2 = 2,1$

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

$P(Y=0|Z=1) = \frac{P(Y=0,X=1)}{P(Z=1)} = \frac{0,3}{0,5} = 0,6.$

$P(Y=1|Z=1) = \frac{P(Y=1,X=0)}{P(Z=1)} = \frac{0,2}{0,5} = 0,4.$

$E = 0 * 0,6 + 1 * 0,4 = 0,4.$

$E^2 = 0 * 0,6 + 1 * 0,4 = 0,4.$

$V = 0,4 - 0,4 * 0,4 = 0,24$

4. Er X og Y uafhængige? Begrund svaret.

   $P(X=0,Y=0) = P(X=0) * P(Y=0) \quad 0,1 =? = 0,5 * 0,3 = 0,15$

Opgave 2

På to indfaldsveje til København optælles antallet af BMW'er, der ankommer i tidsrummet 9.00 til 9.30 en hverdag.

På indfaldsvej A kan dette antal (X) beskrives med en poissonfordeling med $\lambda = 10$.

På indfaldsvej B kan dette antal (Y) beskrives med en poissonfordeling med $\lambda = 20$.

De tilsvarende stokastiske variable antages at være uafhængige.

1. Beregn sandsynligheden for at der ankommer mere end 12 BMW'er på indfaldsvej A.

   P(X>12)=1-P(X<=12)=1-0,79=0,21

2. Angiv fordelingen for Z=X+Y.

   Z bliver poisson($\lambda = 30$)

3. Givet at der en tilfældig dag samlet er registreret 35 ankomster af mærket BMW, hvad er sandsynligheden for, at 10 af disse ankomster er sket ved indfaldsvej A?

   dvs. udregn P(X=10|Z=35).

   P(X=10|Z=35)=$\frac{P(X=10,Y=25)}{P(Z=35)} = \frac{P(X=10)*P(Y=25)}{P(Z=35)} = \frac{\frac{10^{10}\exp(-10)}{10!}\frac{20^{25}\exp(-25)}{25!}}{\frac{30^{35}\exp(-30)}{35!}} =$
   $\frac{35!(10^{10})(20^{25})\exp(-35)}{10!25!(30^{10})(30^{25})\exp(-35)} =$
   $\frac{35!}{10!25!}\frac{10^{10}}{30^{10}}\frac{20^{25}}{30^{25}} = \binom{35}{10}(\frac{1}{3})^{10}(\frac{2}{3})^{20} = P(W = 10) = 0,12$

   hvor W er $BIN(35, \frac{1}{3})$.

   $[\frac{P(X=10)*P(Y=25)}{P(Z=35)} = \frac{0,12511*0,04459}{0,04531} = 0,12]$

4. Udregn sandsynligheden for at W=10, hvor W er $BIN(35, \frac{1}{3})$ (dvs. binomialfordelt med antalsparameter 35 og sandsynlighedsparameter $\frac{1}{3}$).

   Sammenlign dette resultat med sp. 3.

   Eksempel på at hvis man betinger med summen af to uafhængige poissonfordelinger så fås en binomialfordeling

Opgave 3

I en større europæisk undersøgelse blandt borgerne i de enkelte lande, er der blevet stillet spørgsmålet "Hvor tilfreds er De med Europarlamentet".

I nedenstående tabel er vist fordelingen for landene Belgien, Danmark og Sverige

Land	meget tilfreds (X_i)	alt andet	i alt (n_i)	$\widehat{p}_i$
Belgien (i=1)	222	1.526	1.748	0,127
Danmark (i=2)	182	1.241	1.423	0,128
Sverige (i=3)	164	1.479	1.643	0,100
i alt	568		4.814	0,118

kilde: European Social Survey 2014

Beregninger til LR test

Land	i alt (n_i)	$\widehat{p}_i$	$x_i * \ln(\widehat{p}_i)$	$(n_i - x_i) * \ln(1 - \widehat{p}_i)$	sum
Belgien (i=1)	1.748	0,127	-458,1	-207,3	-665,4
Danmark (i=2)	1.423	0,128	-374,3	-169,8	-544,1
Sverige (i=3)	1.643	0,100	-377,9	-155,5	-533,5
sum					-1742,9
i alt	4.814	0,118	-1213,9	-533,1	-1747,0

Der opstilles følgende model:

X_i = antallet af personer der svarer "meget tilfreds". i=1,2,3

X_i er $BIN(n_{i,}p_i)$. Dvs at $P(X_i = x_i) = \binom{n_i}{x_i} p_i^{x_i}(1-p_i)^{n_i - x_i}$

De tre stokastiske variable er uafhængige af hinanden.

1. Argumentér for at det er en rimeligt at bruge en binomialfordeling i de enkelte lande og giv en fortolkning af p_i.

2. Opskriv likelihood funktionen $L(p_{1,}p_{2,}p_3)$ og vis at log-likelihood funktionen $log[L(p_{1,}p_{2,}p_3)]$ bliver

   $\sum_{i=1}^{3} \ln[\binom{n_i}{x_i}] + \sum_{i=1}^{3} x_i \ln(p_i) + \sum_{i=1}^{3}(n_i - x_i)\ln(1-p_i)$

   Angiv de tre scorefunktioner og Hesse-matricen.

   $p_1 : \quad \frac{x_1}{p_1} - \frac{n_1 - x_1}{1 - p_1}.$

$p_2 : \ \frac{x_2}{p_2} - \frac{n_2 - x_2}{1 - p_2}$.

$p_3 : \ \frac{x_3}{p_3} - \frac{n_3 - x_3}{1 - p_3}$.

.

$\frac{x_i}{p_i} - \frac{n_i - x_i}{1 - p_i} = 0$ giver $\widehat{p}_i = \frac{x_i}{n_i}$

Hesse bliver en diagonal matrice

$$\begin{bmatrix} H_1 & 0 & 0 \\ 0 & H_2 & 0 \\ 0 & 0 & H_3 \end{bmatrix}$$

$H_i = -x_i p_i^{-2} - (n_i - x_i)(1 - p_i)^{-2}$ H_i angiver det i'te diagonalelement

$E(-H_i) = -E[-x_i p_i^{-2} - (n_i - x_i)(1 - p_i)^{-2}] =$

$E(x_i) p_i^{-2} + (n_i - E(x_i))(1 - p_i)^{-2} =$

$n_i p_i p_i^{-2} + (n_i - n_i p_i)(1 - p_i)^{-2} =$

$n_i [p_i^{-1} + (1 - p_i)^{-1}] = \frac{n_i}{p_i(1 - p_i)}$

.

$E(-H_i)^{-1} = \frac{p_i(1 - p_i)}{n_i}$

3. Udregn konfidensintervallet for p_2, dvs. for Danmarks andel.

   $\widehat{p}_2 = \frac{x_2}{n_2} = \frac{182}{1423} = 0,13$. $V(\widehat{p}_2) = \frac{p_2 * (1 - p_2)}{n_2} = (0,0089)^2$

   0,13+-1,96*0,0089 [0,11-0,15]

4. Det antages nu at $p_1 = p_2 = p_3$. Den fælles parameter kaldes p.

   Opskriv likelihood funktionen $L(p)$ samt log-likelihood funktionen $log[L(p)]$.

   Udregn MLE for p som kaldes $\widehat{p}$.

   Bemærk at $Y = X_1 + X_2 + X_3$ er $BIN(n_1 + n_2 + n_3, p)$

   under antagelsen af at $p_1 = p_2 = p_3 = p$

   $\widehat{p} = \frac{x_1 + x_2 + x_3}{n_1 + n_2 + n_3} = \frac{222 + 182 + 164}{1748 + 1423 + 1643} = \frac{568}{4814} = 0,118$

5. Angiv den approksimative fordeling for $\widehat{p}$.

   Er normalfordelt med middelværdi p og varians $\frac{p * (1 - p)}{n_1 + n_2 + n_3}$

6. Test $H_0 : p_1 = p_2 = p_3 = p$ mod $H_A : H_0^C$.

   Ved brug af et likelihood ratio test.

   $log[L(\widehat{p}_1, \widehat{p}_2, \widehat{p}_3)] = \sum_{i=1}^{3} x_i \ln(\widehat{p}_i) + \sum_{i=1}^{3} (n_i - x_i) \ln(1 - \widehat{p}_i) = -1742,94$

   $log[L(\widehat{p})] = (x_1 + x_2 + x_3) \ln(\widehat{p}) + (n_1 + n_2 + n_3 - x_1 - x_2 - x_3) \ln(1 - \widehat{p}) = -1747$

   teststørrelse=2*(-1742,94–1747)=8,11

   som chi-i-anden fordelt med DF=3-1=2.

   Sss=1,7% forkast

   se regneark for beregninger, som er lagt på Absalon.

7. Giv en samlet konklusion.