Eksamen på Økonomistudiet sommeren 2016
Sandsynlighedsteori og Statistik

2. årsprøve

22. august, 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.

Opgave 1

Firmaet straks.com bringer dagligvarer ud til forbrugerne i bokse. Fra tidligere undersøgelser er det vurderet, at vægten (i kg) på en boks kan beskrives

med en normalfordeling $N(30, 6^2)$, endvidere kan det antages, at der er uafhængighed mellem boksene.

1. Hvad er sandsynligheden for, at en tilfældig udvalgt boks vejer mere end 40 kg?

   Z$\sim N(30, 6^2)$

   P(Z=>40)=1-P($Z <= 40) = 1 - P(\frac{Z-30}{6} <= \frac{40-30}{6}) = 1 - \Phi(\frac{5}{3}) = \Phi(-\frac{5}{3}) = 0,048$

   Arbejdstilsynet udvælger i alt 200 tilfældige bokse og registrerer antallet af bokse, der vejer mere end 40 kg. Lad X være antallet af bokse, der vejer mere end 40 kg blandt 200 udtrukne bokse.

2. Angiv fordelingen for X samt middelværdi og varians. dvs. $E(X)$ og $V(X)$.

   $X \sim Bin(200, p)$ hvor p=$0,048$

   E(X)=200*0,048= 9,6   V(X)=200*0,048*(1-0,048)=9,1

   Undersøgelsen resulterer i, at 2 ud af de 200 (dvs. 1%) vejer mere end 40 kg. Antag at udgangspunktet er, at 5% af samtlige bokse vejer mere end 40 kg.

3. Hvad er sandsynligheden for, at 2 kasser blandt de 200 udtrukne vejer mere end 40 kg?

   P(X=2|$X \sim Bin(200, 0.05)$)=0,0019

Arbejdstilsynet beslutter nu at udtage bokse, indtil man første gang observerer en kasse der vejer mere end 40 kg. Lad Y angive antallet af udtagne bokse, indtil man har obsereveret en boks der vejer mere end 40 kg. Det forudsættes, at sandsynligheden for at en boks vejer mere end 40 kg fortsat er 5%.

4. Hvordan er Y fordelt? Angiv endvidere $E(Y)$ og $V(Y)$.

   P(Y=y)=(1-θ)$^y\theta$ y=0,1,2,.... $\theta = 0,05$ se side 119 i Michael Sørensens bog

   E(Y)=$\frac{1-\theta}{\theta} = 19,0$ side 132 V(Y)=$\frac{1-\theta}{\theta^2} = 380,0$ side 138

Opgave 2

Den simultane fordeling af X og Y er givet ved

	$X = 0$	$X = 1$	$X = 2$	$X = 3$
$Y = 0$	0,1	0,1	0,1	0,2
$Y = 1$	0,05	0	0,1	0,1
$Y = 2$	0,05	0,1	0	0,1

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

   $E(X)$=0*0,2+1*0,2+2*0,2+3*0,4=1,8

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

   V(X)=$E(X^2) - [E(X)]^2 = 4,6 - (1,8)^2 = 1,36$

   Lad $Z = X + Y$.

2. Angiv fordelingen for X givet Z=3.

   P(X=0|Z=3)=0 P(X=1|Z=3)=$\frac{1}{4}$ P(X=2|Z=3)=$\frac{1}{4}$ P(X=3|Z=3)=$\frac{2}{4}$

   Lad $W = X * Y$.

3. Angiv fordelingen for W og udregn $E(W)$.

   P(W=0)=0,6 P(W=1)=0 P(W=2)=0,2 P(W=3)=0,1 P(W=4)=0 P(W=6)=0,1

   E(W)=0*0,6+2*0,2+3*0,1+6*0,1=1,3

4. Er X og Y uafhængige? Begrund svaret.

   nogle celler har sandsynlighed 0

Opgave 3

I en større opinionsmåling er der udspurgt 1.500 personer. De skal svare på spørgsmålet: Hvilken blok vil De stemme på, hvis der var valg i morgen?. Svarkategorierne og fordelingen af svar er vist i nedenstående tabel.

kategori	Z	antal svar
Blå blok	$Z = 1$	650
Rød blok	$Z = 2$	600
Ved ikke	$Z = 3$	250
I alt		1.500

Der opstilles følgende model:

Z$_i \overset{d}{=} categorical(p_1, p_2, p_3)$.
Dvs. $P(Z_i = 1) = p_1$; $P(Z_i = 2) = p_2$; $P(Z_i = 3) = p_3$
Her gælder at $p_1 + p_2 + p_3 = 1$
lad S_1=$\sum_{i=1}^{1500} \mathbb{I}(Z_i = 1)$, S_1=$\sum_{i=1}^{1500} \mathbb{I}(Z_i = 2)$ og S_3=$\sum_{i=1}^{1500} \mathbb{I}(Z_i = 3)$.
Bemærk at $S_1 + S_2 + S_3 = 1.500$

1. Kommenter kort den opstillede model. Angiv hvor mange frie parametre modellen indeholder.

   Man kunne vælge at bruge at $(S_1, S_2, S_3) \sim Multinomial(1500, p_1, p_2, p_3)$.

   der er to frie parametre. Multinomialfordelingen (=Polynomialfordeling) er omtalt side 78 i Sørensens bog.

2. Opskriv likelihood funktionen $L(p_1, p_2)$ samt log-likelihood funktionen $log[L(p_1, p_2)]$. Angiv scorefunktionerne og vis at løsningen til scorefunktionerne bliver $\widehat{p}_1 = \frac{S_1}{1500} = 0,43$ og $\widehat{p}_2 = \frac{S_2}{1500} = 0,40$

   $L(p_1, p_2) = \prod_{i=1}^{1500} (p_1)^{\mathbb{I}(Z_i=1)} (p_2)^{\mathbb{I}(Z_i=2)} (1 - p_1 - p_2)^{\mathbb{I}(Z_i=3)} =$
   $(p_1)^{s_1} (p_2)^{s_2} (1 - p_1 - p_2)^{s_3}$

   $log[L(p_1, p_2)] = s_1 \ln(p_1) + s_2 \ln(p_2) + s_3 \ln(1 - p_1 - p_2)$

   $$S(p_1, p_2) = \begin{array}{|c|} \hline \frac{s_1}{p_1} - \frac{s_3}{1-p_1-p_2} \\ \hline \frac{s_2}{p_2} - \frac{s_3}{1-p_1-p_2} \\ \hline \end{array}$$

$S(p_1, p_2) = \boxed{\begin{array}{c}0\\0\end{array}}$ giver $\frac{s_1}{p_1} = \frac{s_2}{p_2}$giver $s_1 * \widehat{p}_2 = s_2 * \widehat{p}_1$

$s_1 * (1 - p_1 - p_2) = (1500 - s_1 - s_2) * p_1$

Man ønsker nu at undersøge om blå blok og rød blok er lige store. Der opstilles derfor følgende model: $p_1 = p_2 (= p)$ og dermed $p_3 = 1 - 2p$

3. Opskriv likelihood funktionen samt log-likelihoodfunktionen.

   $L(p) = (p)^{s_1}(p)^{s_2}(1-2p)^{s_3} = (p)^{s_1+s_2}(1-2p)^{s_3}$

   $log[L(p)] = (s_1 + s_2)\ln(p) + (1500 - s_1 - s_2)\ln(1 - 2p)$

   S(p)=$\frac{s_1+s_2}{p} - \frac{2(1500-s_1-s_2)}{1-2p} = 0$ giver $(s_1 + s_2) * (1 - 2p) = 2(1500p) - (s_1 + s_2)2p$

   giver $s_1 + s_2 = 2(1500p)$

   Angiv scorefunktionen og vis at løsningen bliver $\widehat{p} = \frac{S_1+S_2}{2*1500} = 0,42$

4. Vis at Hesse-matricen kan skrives som $H(p) = \frac{-(S_1+S_2)}{p^2} - \frac{(1500-S_1-S_2)*4}{(1-2p)^2}$.

   Begrund at løsningen i 3. er et maksimum.

   diffenrentier S(p)=$\frac{s_1+s_2}{p} - \frac{2(1500-s_1-s_2)}{1-2p}$

   er negativ dermed maksimum

5. Test $H_0 : p_1 = p_2$ mod $H_A : p_1 \neq p_2$. Brug et Likelihood Ratio Test.

   $log[L(\widehat{p}_1, \widehat{p}_2)] = s_1 \ln(\widehat{p}_1) + s_2 \ln(\widehat{p}_2) + s_3 \ln(1 - \widehat{p}_1 - \widehat{p}_2) =$

   $650 * \ln(\frac{650}{1500}) + 600 \ln(\frac{600}{1500}) + 250 \ln(\frac{250}{1500}) = -1541,28$

   $log[L(\widehat{p})] = (s_1 + s_2)\ln(\widehat{p}) + (1500 - s_1 - s_2)\ln(1 - 2\widehat{p}) =$

   $(650 + 600)\ln(\frac{650+600}{2*1500}) + (250)\ln(1 - 2 * \frac{650+600}{2*1500}) = -1542,28$

   $-2 * \{\frac{650+600}{2*1500} - log[L(\widehat{p}_1, \widehat{p}_2)]\} = 2,0$ som χ^2 med 1 DF

   p-værdi =15,7%

6. Brug oplysningerne i 4. til at udregne variansen for $\widehat{p}$.

   $E[-H(p)] = E[\frac{(S_1+S_2)}{p^2} + \frac{(1500-S_1-S_2)*4}{(1-2p)^2}] =$

   $\frac{E(S_1+S_2)}{p^2} + \frac{E[(1500-S_1-S_2)*4]}{(1-2p)^2} = \frac{2*1500*p}{p^2} + \frac{4*[1500-1500*p-1500*p]}{(1-2p)^2} =$

   $\frac{2*1500}{p} + \frac{4*1500[1-2p]}{(1-2p)^2} = \frac{2*1500}{p} + \frac{4*1500}{(1-2p)} = \frac{2*1500*(1-2p)+4*1500*p}{p(1-2p)} =$

$\frac{2*1500(1-2p+2p)}{p(1-2p)} = \frac{2*1500}{p(1-2p)}$

$V(\widehat{P}) = \frac{p(1-2p)}{2*1500} = \frac{0{,}42*(1-0{,}84)}{3000} = (0,0048)^2$

alternativt $V(\widehat{p}) = V(\frac{S_1+S_2}{2*1500}) = \frac{1500*2p(1-2p)}{4*1500*1500} = \frac{p(1-2p)}{2*1500}$

7. Angiv et 95% (approximativt) konfidensinterval for p.

   0,42+-1,96*0,0048=0,42+-0,009