Eksamen på Økonomistudiet vinteren 2019-20

Sandsynlighedsteori og Statistik

2. årsprøve

4. januar, 2020

(3-timers prøve med hjælpemidler)

Dette eksamenssæt består af 4 sider inkl. denne forside.
Til dette eksamenssæt hører 0 bilag.

Syg under eksamen:

Bliver du syg under selve eksamen på Peter Bangs Vej, skal du

- kontakte en eksamensvagt for at få hjælp til registreringen i systemet som syg og til at aflevere blankt.
- forlade eksamen
- kontakte din læge og indsende lægeerklæring til Det Samfundsvidenskabelige Fakultet senest 5 dage efter eksamensdagen.

Pas på, du ikke begår eksamenssnyd!

Det er eksamenssnyd, hvis du under prøven

- Bruger hjælpemidler, der ikke er tilladt
- Kommunikerer med andre eller på anden måde modtager hjælp fra andre
- Kopierer andres tekster uden at sætte citationstegn eller kildehenvise, så det ser ud som om det er din egen tekst
- Bruger andres idéer eller tanker uden at kildehenvise, så det ser ud som om det er din egen idé eller dine egne tanker
- Eller hvis du på anden måde overtræder de regler, der gælder for prøven

Du kan læse mere om reglerne for eksamenssnyd på Din Uddannelsesside og i Rammestudieordningens afs. 4.12.

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

Bent er på ski-ferie med sin familie i Norge. På en af T-lifterne mangler lift nummer 43. Der er i alt 90 lifte på T-liften. Ud af i alt 25 ture med liften i løbet af dagen ankommer Bent 3 gange til liften netop i samme øjeblik som den manglene lift nummer 43. Dette betyder, at Bent således har måttet vente på den næste lift mens den manglende lift kørte forbi. I frustration herover beder han ved middagsbordet sin familie om at hjælpe ham med at forstå sandsynligheden for denne oplevelse. For at hjælpe Bent vil vi antage, at det er tilfældigt hvilken af de 90 lifter Bent ankommer til på hver af de 25 ture. Således er der uafhængighed mellem hver ankomst til liften.

1. Lad $X_i \in \{0, 1\}$ være en Bernoulli-fordelt stokastisk variabel. Vi lader $X_i = 1$ hvis Bent møder den manglene lift nummer 43 på tur nummer i og $X_i = 0$ ellers. Hvad er sandsynligheden for, at Bent ankommer til den manglende lift 43 på tur i, $P(X_i = 1)$?

2. Lad $Y = \sum_{i=1}^{25} X_i$ angive antallet af gange Bent har mødt den manglende lift nummer 43 ud af sine 25 uafhængige ture. Hvad er sandsynligheden for, at Bent møder denne manglende lift 3 gange ud af sine 25 ture, $P(Y = 3)$?

3. Hvad er sandsynligheden for, at Bent møder den manglende lift mindst 1 gang på sine 25 ture?

4. Nu kommer et andet familie-medlem i tanke om, at der for de sidste 10 ture i virkeligheden var 2 lifter, der manglede. Lift nummer 6 manglede altså også for de sidste 10 ture. Hvad er det forventede antal gange, Bent møder én af de to manglende lifter?

Opgave 2

Lad X være en uniform kontinuert stokastisk variabel på intervallet $[1, 5]$ med tæthedsfunktion

$$p(x) = \frac{1}{4}\mathbf{1}(x \in [1, 5]),\ x \in [1, 5]$$

og lad $Y = 6X + 4$.

1. Hvad er middelværdien af Y, $\mathbb{E}[Y]$?

2. Hvad er variansen af Y, $Var(Y)$?

3. Lad $Z = \log(Y)$. Hvad er tæthedsfunktionen for Z, $q(z)$?

Opgave 3

Vi er blevet kontaktet af Styrelsen for Arbejdsmarked og Rekruttering (STAR), da de ønsker en analyse af job-søgningsadfæren for arbejdsløse. Til det formål har de indsamlet data for arbejdsløse borgere og registreret antallet af ansøgninger, hver borger har sendt ud i løbet af en måned. Vi lader en stokastiske variabel $X_i \in \{0, 1, 2, 3, \dots\}$ angive antal afsendte ansøgninger for borger i. Data fra STAR indeholder antal ansøgninger for $n = 251$ borgere, hvor vi får oplyst at det gennemsnitlige antal ansøgninger sendt er $\frac{1}{251}\sum_{i=1}^{251} x_i = 2.291$.

Vi vil opstille en statistisk model til at belyse ansøgningsadfærden. Vi vil antage, at beslutningen om at afsende en ansøgning er uafhængig af alle andre ansøgninger sendt af borgeren selv og andre medborgere. Til sidst antager vi, at antallet af ansøgninger er Poisson-fordelt med sandsynlighedsfunktionen

$$p(x) = \frac{\theta^{2x}}{x!}\exp(-\theta^2),\ x \in \mathbb{N}_0$$

hvor $\theta > 0$. Vi vil betegne denne parametrisering som den modificerede Poisson-fordeling så

$$X_i \sim \text{ModPois}(\theta) \text{ for } i = 1, \dots, n.$$

Det forventede antal ansøgninger er givet ved

$$\mathbb{E}[X_i] = \theta^2.$$

1. Opskriv likelihood bidragene for hver borger, $\ell(\theta|x_i)$, log-likelihood bidragene for hver borger og log-likelihood funktionen.

2. Angiv første ordens betingelsen (FOC) for maksimering af log-likelihood funktionen og udled maximum likelihood *estimatoren*, $\hat{\theta}_X = \hat{\theta}(X_1, \dots, X_{251})$ for den givne model. Brug derefter information givet i opgaveteksten til at beregne *estimatet*, $\hat{\theta}_x = \hat{\theta}(x_1, \dots, x_{251})$, for de $n = 251$ observationer, vi har fået givet.

3. Angiv bidraget fra hver borger til Hesse-matricen og beregn variansen på estimatoren fundet i forrige spørgsmål, $Var(\hat{\theta}_X)$.

4. Følgende STATA-kode er blevet kørt og har genereret outputtet herunder
```
mlexp ( 2*x*log({theta}) - {theta}^2)
```
Vi ser, at estimatet er $\hat{\theta}_x = 1.514$ og standard afvigelsen er $se(\hat{\theta}_X) = 0.0316$.
Vi bliver bedt om at bruge den estimerede model til at beregne hvad sandsynligheden

er for at en borger sender mindst én ansøgning.

```
Maximum likelihood estimation

Log likelihood = -98.372667                         Number of obs     =        251
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /theta |   1.513551   0.0315597    47.96   0.000     1.451695    1.575407
------------------------------------------------------------------------------
```

5. Vi får nu oplyst, at STAR ønsker at undersøge, om der er en effekt af at tvinge arbejdsløse til at bygge miniature huse af pasta og skumfiduser. Ideen er, at dette vil være så meningsløst for de arbejdsløse, at de hellere vil sende flere ansøgninger i håbet om af finde et job hurtigere. Til det formål har STAR udvalgt tilfældige arbejdsløse borgere til at skulle bygge huse af pasta. Vi vil lade $D_i = 1$ angive, at den arbejdsløse har skullet bygge huse af pasta og $D_i = 0$ ellers. Vi har nu oplysninger om både antal ansøgninger og deltagelse i forsøget, $\{x_i, d_i\}_{i=1}^{251}$. Vi er nu interesseret i den betingede model

$$X_i|D_i \sim \text{ModPois}(\theta + \delta D_i).$$

   Opskriv log-likelihood funktionen for den betingede model.

6. Hvad siger det om forsøget, hvis $\delta \neq 0$?

7. Angiv hvordan STATA-koden fra tidligere kunne ændres for at estimere den betingede model.

8. Vi får nu følgende estimations-output fra STATA ved estimation af den betingede model ovenover. Test om der er en signifikant effekt på antallet af ansøgninger fra at tvinge arbejdsløse til at bygge huse af pasta.
   Vær præcis med hypoteser, teststørrelse og kritisk værdi.

```
Maximum likelihood estimation

Log likelihood = -95.916305                         Number of obs     =        251
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /theta |   1.433197   0.0474579    30.20   0.000    1.340181     1.526213
      /delta |   0.1411508   0.063545     2.22   0.026    0.0166049    0.2656967
------------------------------------------------------------------------------
```