Eksamen på Økonomistudiet vinteren 2018-19
Sandsynlighedsteori og Statistik
2. årsprøve
5. januar, 2019

(3-timers prøve med hjælpemidler)

Dette eksamenssæt består af 4 sider (forsiden inklusiv).
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. Derefter forlader du eksamen. Når du kommer hjem, skal du kontakte din læge og indsende lægeerklæring til Det Samfundsvidenskabelige Fakultet senest en uge 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

Vi vil her undersøge overgangen fra arbejdsløshed til beskæftigelse. Forestil dig, at en arbejdsløs kan få job i to typer af virksomheder. Virksomhedstype 1 giver en højere indkomst end virksomhedstype 2, idet indkomsten i de to virksomheder er hhv. $y_1 = 500$ og $y_2 = 300$ (målt i 1000kr). Til gengæld tilbyder en virksomhed af type 1 et job med 5% sandsynlighed mens en virksomhed af type 2 tilbyder et job med 7% sandsynlighed. Jobtilbudene fra de to virksomheder er uafhængige og den arbejdsløse *skal* tage imod et job, hvis der tilbydes et job. Hvis et job fra hver af de to typer virksomheder modtages, så skal den ledige tage det bedst-betalte job. Der er ingen understøttelse ved arbejdsløshed, dvs. en indkomst på 0 hvis intet job tilbydes. Antag indledningsvist, at der kun er én arbejdsløs og 2 virksomheder; én af type 1 og én af type 2.

1. Lad $X_1 \in \{0, 1\}$ og $X_2 \in \{0, 1\}$ indikere et jobtilbud i virksomheds-type 1 og type 2, henholdsvis. Hvad er sandsynligheden for, at den arbejdsløse ikke bliver tilbudt et job i nogen af de to virksomheder, $P(X_1 = 0, X_2 = 0)$?

2. Hvad er den forventede indkomst for en arbejdsløs?

3. Vi antager nu, at der er 5 type 1 virksomheder ($n_1 = 5$) og 7 type 2 virksomheder ($n_2 = 7$) og at tilbud om ansættelse i hver af disse virksomheder er uafhængige Bernoulli-fordelte stokastiske variable, $X_i \in \{0, 1\}$ for $i = 1, 2, \ldots, 12$. Lad Z_j være antal tilbudte ansættelser i virksomheds-type $j = 1, 2$. Hvad er sandsynligheden for, at den arbejdsløse modtager tilbud om ansættelse i mindst én af de 7 type 2 virksomhed, $P(Z_2 > 0)$?

4. Vi approksimerer nu fordelingen af antal ansættelser i virksomhed af type 1 og 2 med Poisson fordelinger med parametre λ_1 og λ_2 henholdsvis. Hvad er sandsynligheden for, at den arbejdsløse bliver ansat i én af de i alt 12 virksomheder?
   Hint: Poisson parameteren kan approksimeres som $\lambda = np$, hvor n og p er hhv. antalsparameteren og sandsynlighedsparameteren i Binomialfordelingen. Vi bruger Poisson-approksimationen, selvom vi kun har henholdsvis $n_1 = 5$ og $n_2 = 7$ type 1 og type 2 virksomheder.

Opgave 2

Lad X være en Eksponentialfordelt stokastisk variabel på $(0, \infty)$ med parameter $\lambda = 2$. Vi har at $\mathbb{E}(X) = \frac{1}{\lambda} = \frac{1}{2}$ og $Var(X) = \frac{1}{\lambda^2} = \frac{1}{4}$.

1. Angiv tæthedsfunktionen for X, $p(x)$.

2. Lad $Y = 2 \cdot X - 1$. Find middelværdi, $\mathbb{E}(Y)$, og varians, $Var(Y)$, af Y.

3. Find tæthedsfunktionen for Y, $q(y)$.

Opgave 3

En fertilitetsklinik er interesseret i at modellere danske pars fertilitet. Vi arbejder hos et konsulentbureau og vi bliver nu bedt om at undersøge fertilitetsadfærden hos $n = 67$ par. Vi får oplyst, at fertilitetsklinikken har observeret, at mange af parrene ønsker at få én pige og at de fortsætter med at få børn indtil de får en pige, hvorefter de ikke får flere børn. Vi lader Y_i angive antal børn født af par $i = 1, \dots, 67$ og vi får oplyst, at antallet af børn i vores sample på $n = 67$ par er $\sum_{i=1}^{67} y_i = 107$. Udfaldsrummet for antal fødsler er $\mathbb{N} = \{1, 2, \dots\}$ og vi vil lade θ angive sandsynligheden for at føde en pige og $1 - \theta$ for sandsynligheden for at føde en dreng. Antal børn har dermed sandsynlighedsfunktionen

$$p(y) = (1-\theta)^{y-1}\theta, \ y \in \mathbb{N}$$

som vi vil betegne den ¿¿modificerede geometriske fordeling¡¡ så,

$$Y_i \sim \text{ModGeo}(\theta).$$

Det forventede antal børn er givet ved

$$\mathbb{E}(Y_i) = \frac{1}{\theta}.$$

1. Definér parameter-rummet Θ så $\theta \in \Theta$.

2. Opskriv likelihood bidragene for hvert par, $\ell(\theta|y_i)$, log-likelihood bidragene for hvert par og log-likelihood funktionen.

3. Angiv første ordens betingelsen (FOC) for maksimering af log-likelihood funktionen og udled maximum likelihood *estimatoren*, $\hat{\theta}(Y_1, \dots, Y_{67})$ for den givne model. Brug derefter information givet i opgaveteksten til at beregne *estimatet*, $\hat{\theta}_n = \hat{\theta}(y_1, \dots, y_{67})$, for de $n = 67$ observationer, vi har fået givet.

4. Angiv bidraget fra hvert par til Hesse-matricen og beregn variansen på estimatoren fundet i forrige spørgsmål, $Var(\hat{\theta}_n)$.

5. Vi får nu oplyst, at estimatet er $\hat{\theta}_n = 0.6262$ og standard afvigelsen er $se(\hat{\theta}_n) = 0.0469$. Vi bliver bedt om at bruge den estimerede model til at beregne hvad sandsynligheden er for et par vil skulle have 3 drenge før de får en pige, dvs. 4 børn i alt.

6. Vi bliver nu bedt om, at teste om det forventede antal børn er 2 ved hjælp af et Wald-test. Vær præcis med hypoteser, teststørrelse og kritisk værdi.

7. Vi får nu oplyst, at fertilitetsklinikken ønsker, at undersøge om kvindens alder påvirker barnets køn. Vi vil lade x_i være alderen på kvinden i par i, hvor $x_i = 1$ hvis kvinden er 40 år eller ældre og nul ellers. Fertilitetsklinikken har indsamlet alderen på deres klienter og vi har nu oplysninger om både antal børn og moderens alder, $\{y_i, x_i\}_{i=1}^{67}$. Vi er nu interesseret i den betingede model

$$Y_i|X_i \sim \text{ModGeo}(\theta + \delta X_i).$$

   Opskriv log-likelihood funktionen for den betingede model. Hvad er fortolkningen af δ?

8. STATA-koden, der estimerede den oprindelige model, er
```
mlexp ((y-1)*log(1-{theta}) + log({theta}))
```
   Angiv hvordan STATA-koden kunne ændres for at estimere den betingede model.