Eksamen på Økonomistudiet sommeren 2019

Sandsynlighedsteori og Statistik

2. årsprøve

28. August, 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

Du har følgende information vedr. to binære stokastisk variable, $X \in \{1, 2\}$ og $Y \in \{1, 2\}$:

$$P(X = 1, Y = 1) = 0.2$$
$$P(X = 1, Y = 2) = 0.1$$

1. Angiv den marginale sandsynlighedsfunktion for X, $p(x) = P(X = x)$.
2. Beregn middelværdien af X kvadreret, $\mathbb{E}(X^2)$.
3. Beregn variansen af X, $Var(X)$.
4. Lad $Z = -2 \cdot X + 3$. Beregn variansen af Z, $Var(Z)$, og korrelationen mellem X og Z, $corr(X, Z)$.

Opgave 2

Lad $X \geq 0$ være en Eksponential-fordelt stokastisk variabel med tæthedsfunktionen

$$p(x) = \frac{1}{2}\exp(-\frac{1}{2}x), \ x \geq 0$$

1. Angiv Fordelingsfunktionen, $P(X \leq x)$.
2. Beregn sandsynligheden for at X mindst er 0.5, $P(X \geq 0.5)$.
3. Lad $Y = \log(X)$. Hvilket interval er Y fordelt på?
4. Angiv tæthedsfunktionen for den transformerede stokastiske variabel, $q(y)$.

Opgave 3

Vi er blevet kontaktet af Danmarks Lærerforening (DLF), som beder os analyserer det årlige skolefravær for landets 9. klasse-elever. Til det formål har vi fået data fra $n = 75$ tilfældige skoleklasser med information om det årlige antal fraværstimer for hver klasse. I vores analyse vil vi definere en diskret stokastisk variabel $Y_i \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$ som antallet af fraværstimer for klasse i. Vi har også fået at vide, at det samlede antal fraværstimer i data er $\sum_{i=1}^{75} y_i = 1599$. Vi antager, at klasserne er uafhængige, og vil benytte Poisson-fordelingen til at modellere antallet af fraværstimer. Konkret vil vi benytte sandsynlighedsfunktionen

$$p(y_i) = \exp(-\exp(\theta))\frac{\exp(\theta \cdot y_i)}{y_i!},$$

hvilket er en Poisson-fordeling med $\lambda = \exp(\theta)$. Vi kalder den derfor for den modificerede Poisson fordeling,

$$Y_i \sim \text{ModPois}(\theta)$$

og vi ved, at middelværdien er lig

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

1. Definér parameter-rummet Θ så $\theta \in \Theta$.
2. Opskriv likelihood bidragene for hver klasse, $\ell(\theta|y_i)$, log-likelihood bidragene for hver klasse og log-likelihood funktionen.
3. Angiv første ordens betingelsen (FOC) for maksimering af log-likelihood funktionen og udled maximum likelihood *estimatoren*, $\hat{\theta}(Y_1, \dots, Y_n)$ for den givne model. Brug derefter information givet i opgaveteksten til at beregne *estimatet*, $\hat{\theta}_n = \hat{\theta}(y_1, \dots, y_{75})$.
4. Angiv bidraget fra hver klasse til Hesse-matricen og beregn variansen på estimatoren fundet i forrige spørgsmål, $Var(\hat{\theta})$.
5. Vi får nu oplyst, at estimatet er $\hat{\theta}_n = 3.0596$ og standardafvigelsen er $se(\hat{\theta}) = 0.0251$. Med dette estimat bliver vi bedt om at beregne sandsynligheden for, at en tilfældig klasse har 22 timers fravær.
6. Vi bliver nu bedt om, at teste om middelværdien af antal fraværstimer er 22 ved hjælp af et Wald-test. Vær præcis med hypoteser, teststørrelse og kritisk værdi.

7. DLF vil nu gerne have os til at undersøge, om der er forskel på tværs af Danmark. De vil konkret have os til at undersøge, om der er forskel på fraværet i Jylland i forhold til resten af landet. Til det formål definerer vi en ny stokastisk variabel $X_i \in \{0, 1\}$, som er lig 1 hvis klasse i er fra en skole i Jylland og 0 ellers. Vi har nu data for både antal fraværstimer og placeringen af klassen, $\{y_i, x_i\}_{i=1}^{75}$.

   Vi opstiller en betinget model for fravær givet den geografiske placering af klassen,

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

   Opskriv log-likelihood funktionen for den betingede model.

8. Vi får nu oplyst, at log-likelihood funktionen fra estimation af den betingede (urestrikterede) model er $L_u = L(\hat{\theta}_n, \hat{\delta}_n) = -213.10$ hvor $\hat{\delta}_n < 0$ mens log-likelihood funktionen ved estimation af den restrikterede model i starten af opgave 3 med $\delta = 0$ var $L_r = L(\hat{\theta}_n, 0) = -216.06$.

   Test om δ er signifikant forskellig fra nul ved hjælp af et Likelihood Ratio (LR) test. Vær præcis med hypoteser, teststørrelse og kritisk værdi. Fortolk dit test-resultat.