Eksamen på Økonomistudiet vinteren 2018-19
Sandsynlighedsteori og Statistik
2. årsprøve
13. februar, 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 børne-relaterede omkostninger. Til det formål definerer vi $X_1 \in \{0, 1\}$ som antal piger og $X_2 \in \{0, 1\}$ som antal drenge i en husholdning. Vi antager altså, at en husholdning maksimalt kan få to børn; ét af hvert køn. Vi får angivet, at den marginale sandsynlighed for at have fået en pige er $P(X_1 = 1) = 0.45$ og den marginale sandsynlighed for at have fået en dreng er $P(X_2 = 1) = 0.55$. Vi får også oplyst, at sandsynligheden for både at have fået en dreng og en pige er $P(X_1 = 1, X_2 = 1) = 0.2$.

1. Hvad er den betingede sandsynlighed for at få en pige, hvis man i forvejen har en dreng, $P(X_1 = 1|X_2 = 1)$?

2. Er antal piger og drenge, X_1 og X_2, uafhængige? Fortolk din konklusion.

3. Vi kan lade $Y = X_1 + X_2$ angive antallet af børn i en husholdning. Hvad er det forventede antal børn, $\mathbb{E}(Y)$?

4. Vi får nu oplyst, at omkostningerne relateret til at opfostre én pige er a mens det koster b kroner at opfostre én dreng og omkostningerne vokser yderligere med c hvis man har 2 børn. Vi kan altså lade $Z = aX_1 + bX_2 + cX_1X_2$ angive børne-relaterede omkostninger. Hvad er de forventede omkostninger, $\mathbb{E}(Z)$?

Opgave 2

Lad X og Y være to uafhængige Normalfordelte stokastisk variable på $\mathbb{R}^2$. Vi har, at middelværdierne er $\mu_X = \mathbb{E}(X) = 1$ og $\mu_Y = \mathbb{E}(Y) = -0.5$ mens varianserne er $\sigma_X^2 = Var(X) = 0.1$ og $\sigma_Y^2 = Var(Y) = 1$.

1. Angiv de marginale tæthedsfunktioner for henholdsvis X, $p_X(x)$, og Y, $p_Y(y)$.

2. Lad $Z = 0.2 \cdot X + 0.1 \cdot Y$. Find middelværdi, $\mathbb{E}(Z)$, og varians, $Var(Z)$ af Z. Angiv fordelingen af Z.

3. Beregn kovariansen og korrelationen mellem Z og X og Y. Det vil sige, beregn $Cov(Z, X)$, $\text{corr}(Z, X)$, $Cov(Z, Y)$ og $\text{corr}(Z, Y)$.

4. Beregn den betingede middelværdi, $\mathbb{E}(Z|X = 1)$.

Opgave 3

Vi arbejder nu hos en stor dansk shipping-virksomhed, og er blandt andet ansvarlige for import af iPhones. Apple har kontaktet os, da de ønsker, at vi analyserer tilfælde af defekte iPhones. Helt specifikt vil de gerne have os til at estimere en model for andelen af defekte iPhones i en container. Vi vil definere Y_i på $[0,1]$ som andelen af defekte iPhones i container i. Vi har data på $n = 82$ container-leverancer af iPhones og får oplyst fra skades-afdelingen, at den gennemsnitlige andel af defekte iPhones er $\frac{1}{82}\sum_{i=1}^{82} y_i = 0.078$ og at den gennemsnitlige log-andel er $\frac{1}{82}\sum_{i=1}^{82}\log(y_i) = -2.809$. Vi vil approksimere andelen af defekte iPhones som uafhængige Beta-fordelte stokastiske variable, således at

$$Y_i \sim \text{Beta}(\beta).$$

Beta-fordeling med parameter $\beta > 1$ har tæthedsfunktionen

$$p_Y(y) = \beta y^{\beta-1},\ y \in [0,1]$$

og den forventede andel defekte iPhones er

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

1. Hvis Y_i har tæthedsfunktionen som angivet ovenover, hvad er Fordelingsfunktionen så?
2. Opskriv likelihood bidragene for hver container, $\ell(\beta|y_i)$, log-likelihood bidragene for hver container og log-likelihood funktionen.
3. Angiv første ordens betingelsen (FOC) for maksimering af log-likelihood funktionen og udled maximum likelihood *estimatoren*, $\hat{\beta} = \hat{\beta}(Y_1, \ldots, Y_{82})$ for den givne model. Brug derefter information givet i opgaveteksten til at beregne *estimatet*, $\hat{\beta}_n = \hat{\beta}(y_1, \ldots, y_{82})$, for de $n = 82$ observationer, vi har fået givet.
4. Angiv bidraget for hver container til Hesse-matricen og beregn variansen på estimatoren fundet i forrige spørgsmål, $Var(\hat{\beta})$.
5. Vi får nu oplyst, at estimatet er $\hat{\beta}_n = 0.360$ og standard afvigelsen er $se(\hat{\beta}) = 0.040$. Vi bliver bedt om at bruge den estimerede model til at beregne, hvad sandsynligheden er for at højest 20 procent af en sending er defekte.
6. Vi bliver nu bedt om at teste om den forventede andel defekte iPhones er 20% ved brug af et Wald test. Vær præcis med hypoteser, teststørrelse og kritisk værdi.

7. Vi får nu oplyst, at shipping-virksomheden ønsker at undersøge om der er en systematisk forskel på andelen af defekte iPhones ud fra hvilket fragtselskab, der har håndteret containerne. Der er to fragtselskaber: AlwaysOnTime og NeverLate. Vi lader x_i indikere hvilket selskab, der har fragtet varerne og angiver med $x_i = 1$ hvis det var selskabet AlwaysOnTime og $x_i = 0$ hvis det var NeverLate. Vi har nu oplysninger både om andelen af defekte iPhones og hvilket fragtselskab, der har transporteret dem, $\{y_i, x_i\}_{i=1}^{82}$. Vi er nu interesseret i den betingede model

$$Y_i | X_i \sim \text{Beta}(\beta \cdot (1 + \delta X_i))$$

hvor $\delta \in [-1, 1]$ mens $\beta \in \Theta$. Opskriv log-likelihood funktionen for den betingede model. Hvad er fortolkningen af δ?

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