Eksamen på Økonomistudiet sommeren 2019

Sandsynlighedsteori og Statistik

2. årsprøve

12. Juni, 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

Forestil dig to stokastiske variable med hver to mulige udfald, $X \in \{1, 2\}$ og $Y \in \{1, 2\}$. Den simultane sandsynlighedsfunktion, $p(x, y)$, er angivet herunder:

		Y = 1	Y = 2
X	1	0.4	0.2
	2	0.1	0.3

1. Beregn den betingede sandsynlighed $P(X = 2|Y = 1)$.
2. Beregn den betingede middelværdi $\mathbb{E}[X|Y = 1]$.
3. Beregn den ubetingede middelværdi af produktet mellem X og Y, $\mathbb{E}[X \cdot Y]$.
4. Er de to stokastiske variable uafhængige?

Opgave 2

Vi er blevet hyret til at undersøge en nystartet virksomheds profit fra salg af lysestager. Selvom efterspørgslen efter lysestager i realiteten er diskret, vil vi approksimere profitten (målt i 100,000 kr.) som en kontinuert stokastisk variabel, X. Da virksomheden er nystartet og ikke har indgående kenskab til markedet, vælger de, at modellere profitten som en kontinuert uniform fordeling på intervallet -10 til 20. Dvs. $X \in [-10, 20]$ med tæthedsfunktion givet ved

$$p(x) = \frac{1}{30}\mathbf{1}\{X \in [-10, 20]\}$$

hvor $\mathbf{1}\{\bullet\}$ er indikatorfunktionen, som er lig med 1 hvis udtrykket i $\{\}$ er sandt og nul ellers.

1. Givet den givne tæthedsfunktion, hvad er så Fordelingsfunktionen, $P(X \leq x)$?
2. Ejeren af virksomheden er bekymret for den potentielle negative profit. Han beder dig beregne hvad sandsynligheden for en profit mellem 0 og 10 er. Det vil sige beregn $P(X \in [0, 10])$.
3. Ejeren vil også gerne vide, hvad den forventede profit er, givet profitten er mellem 0 og 10. Det vil sige beregn $\mathbb{E}[X|X \in [0, 10]]$.

4. Ejeren vil nu af uransagelige årsager gerne have os til at beregne den betingede middelværdi af transformationen $Y = 2 \cdot X - 10$. Det vil sige beregn $\mathbb{E}[Y|X \in [0, 10]]$.

Opgave 3

Vi vil nu analysere en virksomhed, "VinVin", der sælger vintage vin online. VinVin sælger vin i kasser af 6 flasker, og de er blevet gjort opmærksomme på, at en relativt stor andel af vinen er udrikkelig. Derfor er vi blevet bedt om at analysere et tilfældigt sample af deres solgte vin. Til det formål definerer vi en stokastisk variabel, $X_j \in \{0, 1\}$, der er lig 0 hvis flaske j er drikkelig eller 1 hvis flaske j er udrikkelig. Lad $Y_i = \sum_{j=1}^{6} X_j$ være antallet af udrikkelige flasker i kasse i. Vi har fået data for $n = 101$ solgte kasser samt fået informationen at det samlede antal udrikkelige flasker i data er $\sum_{i=1}^{101} y_i = 167$.

Vi antager, at samtlige flasker er uafhængige af hinanden, så sandsynligheden for at antallet af udrikkelige flasker i en kasse kan beskrives med en Binomial-fordeling med antalsparameter 6 og sandsyndlighedsparameter θ. Y_i har altså sandsynlighedsfunktionen

$$p(y_i) = \frac{6!}{(6 - y_i)! \cdot y_i!} \theta^{y_i} (1 - \theta)^{6 - y_i}$$

og er fordelt

$$Y_i \sim Bin(6, \theta).$$

Det forventede antal udrikkelige flasker er

$$\mathbb{E}(Y_i) = 6 \cdot \theta.$$

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

2. Angiv første ordens betingelsen (FOC) for maksimering af log-likelihood funktionen og udled maximum likelihood *estimatoren*, $\hat{\theta}(Y_1, \ldots, Y_n)$ for den givne model. Brug derefter information givet i opgaveteksten til at beregne *estimatet*, $\hat{\theta}_n = \hat{\theta}(y_1, \ldots, y_{101})$.

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

4. Vi får nu oplyst, at estimatet er $\hat{\theta}_n = 0.276$ og standard afvigelsen er $se(\hat{\theta}) = 0.0182$. Vi bliver nu bedt om at beregne sandsynligheden for, at der højest er 1 udrikkelig flaske i

en tilfældig kasse, $P(Y_i \leq 1)$. Brug den givne sandsynlighdsfunktion og den estimerede parameter til at beregne dette.

5. VinVin vil nu gerne have os til formelt at teste om det forventede antal udrikkelige flasker vin i en tilfældig kasse er 1.8 ved hjælp af et Wald-test. Vær præcis med hypoteser, teststørrelse og kritisk værdi.

6. VinVin mistænker, at én af deres leverandører er skyld i det høje antal udrikkelige vine ved ikke at opbevare vinen korrekt under transport. Vi definerer nu en stokastisk variabel $Z_i \in \{0, 1\}$, hvor 1 angiver at kassen med vin er blevet leveret af den mistænkte leverandør og $Z_i = 0$ ellers. Vi har nu fået data stillet til rådighed med information omkring leverandørene, så vi ved hvilken leverandør, der leverede hvilke kasser, $\{y_i, z_i\}_{i=1}^{101}$. Vi antager, at fordelingen af kasser til leverandører er tilfældig, så Z_i er eksogen. Vi opstiller derfor den betingede model

$$Y_i | X_i \sim Bin(6, \theta + \delta Z_i).$$

   Opskriv log-likelihood funktionen for den betingede model.

7. Vi får nu oplyst, at log-likelihood funktionen fra estimation af den betingede (urestrikterede) model er $L_u = L(\hat{\theta}_n, \hat{\delta}_n) = -351.321$ hvor $\hat{\delta}_n > 0$ mens log-likelihood funktionen ved estimation af den restrikterede model i starten af opgave 3 med $\delta = 0$ var $L_r = L(\hat{\theta}_n, 0) = -356.769$.
   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.