Eksamen på Økonomistudiet vinteren 2019-20

Sandsynlighedsteori og Statistik

2. årsprøve

14. februar, 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

Vi får angivet den simultane sandsynlighedsfunktion for to stokastiske variable, $X \in \{0, 1\}$ og $Y \in \{1, 2\}$ i nedenstående tabel.

Tabel 1: Simultan Sandsynlighedsfunktion, $p(x, y)$.

	$X = 0$	$X = 1$	
$Y = 1$	0.47	0.18	0.65
$Y = 2$	0.09	0.26	0.35
	0.56	0.44	1

1. Beregn $\mathbb{E}(Y|X = 0)$.

2. Beregn $\mathbb{E}(Y)$ og $\mathbb{E}(X)$.

3. Beregn $Var(Y)$.

4. Beregn $Cov(X, Y)$.

Opgave 2

Betragt to stokastiske variable $X \in \mathbb{R}$ og $Y \in \mathbb{R}$, som vi antager er bivariat Normalfordelt med middelværdier $\mathbb{E}[X] = \mu_X$ og $\mathbb{E}[Y] = \mu_Y$, varianser $Var(X) = \sigma_X^2$ og $Var(Y) = \sigma_Y^2$, samt covarians $Cov(X, Y) = \sigma_{X,Y}$.

1. Lad $M = 1 + 2X - 3Y$. Hvad er middelværdien, $\mathbb{E}[M]$, og variansen af M, $Var(M)$, og hvilken fordeling følger M?

2. Hvis $\mu_X = 0.1$, $\mu_Y = 0.5$, $\sigma_X^2 = 0.2$, $\sigma_Y^2 = 0.5$ og $\sigma_{X,Y} = 0.1$, hvad er så den betingede middelværdi af Y givet X er 0, $\mathbb{E}[Y|X = 0]$

3. Forklar hvorfor $\mathbb{E}[Y|X = 0] < \mathbb{E}[Y]$ i spørgsmålet ovenover.

4. Lad $Z = \exp(X)$. Hvad er tæthedsfunktionen for Z?

Opgave 3

Vi er blevet kontaktet af et stor dansk el-selskab, der ønsker at analysere deres kunders daglige el-forbrug. De har til det formål indsamlet data for $n = 178$ kunder's daglige el-forbrug som en andel af kundens samlede forbrug i alt den dag. Vi lader $Y_i \in (0,1)$ angive el-forbrugs andelen for kunde i. El-selskabet oplyser, at den gennemsnitlige log-forbrugsandel i deres data er $\frac{1}{178}\sum_{i=1}^{178}\log(y_i) = -0.3634$. Til at udføre vores analyse vil vi antage at forbrugsandelen følger tæthedsfunktionen

$$p(y) = [\exp(\theta) + 1]y^{\exp(\theta)}, \; y \in (0,1)$$

som vi kalder den modificerede beta-fordeling

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

for alle i. Vi antager at alle forbrugernes el-forbrugs andele er uafhængige.

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

2. Angiv første ordens betingelsen (FOC) for maksimering af log-likelihood funktionen og udled maximum likelihood *estimatoren*, $\hat{\theta}_Y = \hat{\theta}(Y_1, \ldots, Y_{178})$ for den givne model. Brug derefter information givet i opgaveteksten til at beregne *estimatet*, $\hat{\theta}_y = \hat{\theta}(y_1, \ldots, y_{178})$, for de $n = 178$ observationer, vi har fået givet.

3. Angiv bidraget fra hver kunde til Hesse-matricen og beregn variansen på estimatoren fundet i forrige spørgsmål, $Var(\hat{\theta}_Y)$.
   Brug at vi får at vide, at informationen er $I(\theta) \approx 0.4055$.

4. Følgende STATA-kode er blevet kørt og har genereret outputtet herunder
   `mlexp( log(exp({theta}) + 1) + exp({theta})*log(y))`
   Vi ser, at estimatet er $\hat{\theta}_y = 0.560$ og standard afvigelsen er $se(\hat{\theta}_Y) = 0.1177$.
   Brug den estimerede model til at beregne, hvad sandsynligheden er for, at en tilfældig kunde's elforbrug højest udgør halvdelen af kunden's samlede forbrug, $P(Y_i \leq 0.5)$.

```
Maximum likelihood estimation
Log likelihood =  66.865291                        Number of obs     =        178
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /theta |   0.5606289   0.1177402    4.76   0.000     0.3298624  0.7913954
------------------------------------------------------------------------------
```

5. Vi får nu oplyst, at virksomheden ønsker at undersøge, om forbruget er forskelligt i weekenden i forhold til hverdagene. Til det formål har el-selskabet registreret hvilken dag på ugen, forbrugsandelen er registreret. Vi vil lade $D_i = 1$ angive at forbruget er foretaget en weekend og $D_i = 0$ ellers. Vi har nu oplysninger om både forbrugsandelen og ugedagen, $\{y_i, d_i\}_{i=1}^{178}$. Vi er nu interesseret i den betingede model

$$Y_i|D_i \sim \text{ModBeta}(\theta + \delta D_i).$$

   Opskriv log-likelihood funktionen for den betingede model.

6. Hvad siger det om el-forbruget, 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 forskel på el-forbrugsandele på tværs af weekend og hverdage.
   Vær præcis med hypoteser, teststørrelse og kritisk værdi.

```
Maximum likelihood estimation
Log likelihood =  68.763549                        Number of obs     =        178
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /theta |   0.31883   0.1907152     1.67   0.095    -0.0549649    0.6926248
      /delta |   0.46155   0.2419145     1.91   0.056    -0.0125891    0.9356984
------------------------------------------------------------------------------
```