Eksamen på Økonomistudiet vinteren 2015-16
Sandsynlighedsteori og Statistik
2. årsprøve
18. februar, 2016
(3-timers prøve med hjælpemidler)

Dette eksamenssæt består af 6 sider (forsiden inklusiv).

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

I denne opgave betragter vi en virksomhed, som producerer træskåle. Vi antager, at produktionen består af to arbejdsprocesser, som udføres efter hinanden. Den tid som bruges på den første proces kan beskrives som en stokastisk variabel X_1, og tiden som bruges på den anden proces kan beskrives som X_2. Den simultane fordeling af tidsforbruget fremgår af tabel 1.

Tabel 1: *Den simultane fordeling af tidsforbrug*

Arbejdstid i timer		Proces 1 X_1		
		1	2	3
	1	0.10	0.10	0.05
Proces 2 X_2	2	0.15	0.10	0.15
	3	0.20	0.10	0.05

1. Angiv den marginale fordeling for tidsforbruget ved proces 1 X_1.

2. Angiv om tidsforbruget ved de to processer er stokastisk uafhængigt. Begrund dit svar.

3. Udregn kovariansen mellem de to tidsforbrug: $Cov(X_1, X_2)$.

   Beskriv i ord hvad den angiver.

4. Udregn sandsynligheden for at den samlede arbejdstid er 3 timer.

Opgave 2

For virksomheden, som producerer træskåle, varierer kvaliteten af det færdige produkt. Kvaliteten af skålen beskrives med et indeks, Q, der antager værdier på den reelle akse mellem -1 og 5, dvs.

$$Q \in \{q \in \mathbb{R} : -1 \leq q \leq 5\} = [-1, 5].$$

Vi antager at indekset for kvalitet, Q, kan beskrives som en ligefordeling på $[-1, 5]$. Varer, som har en kvalitet på 0 eller derunder, kasseres.

1. Opskriv sandsynlighedstætheden for Q og udregn sandsynligheden for at en vare må kasseres $P(Q \leq 0)$.

2. Vi er interesseret i den forventede kvalitet.

   (a) Udregn den forventede kvalitet af alle varer som produceres: $E(Q)$.

   (b) Vis, at den forventede kvalitet af de varer som ikke kasseres er

   $$E(Q|Q > 0) = 2.5.$$

3. Virksomheden kan sælge skålen til en pris (målt i kr.), som afhænger af den forventede kvalitet af produkterne som sælges:

   $$p = 1000 \cdot E(Q|Q > 0).$$

   Varer, som kasseres, har ingen værdi.

   Udregn den forventede indtjening for hver vare som produceres.

4. Virksomheden overvejer nu at introducere en ny teknologi, som giver mindre variation i kvaliteten. Med den nye teknologi vil kvaliteten kunne beskrives som en ligefordeling på intervallet $[-0.5, 4]$.

   Vil den nye teknologi øge indtjeningen per produceret vare (begrund dit svar)?

Opgave 3

Denne opgave handler om levetiden af en central komponent i et fladskærms-TV. Vi har observeret en række levetider i måneder, kaldet $\{y_i\}_{i=1}^n$, hvor n er antallet af observationer. Vigtig for levetiden er tykkelsen af de anvendte ledninger i komponenten og jo tykkere ledningerne er des længere er levetiden. Ledningstykkelsen er målt i micro-meter og kaldes $\{x_i\}_{i=1}^n$. Vi betragter observationerne som realisationer af stokastiske variable, $\{Y_i\}_{i=1}^n$ henholdsvis $\{X_i\}_{i=1}^n$, med udfaldsrum givet ved

$$Y_i \in \mathbb{Y} = \{y \in \mathbb{R} : y > 0\} \quad \text{og} \quad X_i \in \mathbb{X} = \{x \in \mathbb{R} : x > 0\}.$$

For at modellere levetiden som funktion af tykkelsen af de anvendte ledninger antager vi, baseret på tidligere undersøgelser, at levetiden for komponenten, Y_i, givet ledningstykkelsen, $X_i = x$, er beskrevet ved en betinget fordeling med tæthedsfunktion givet ved

$$f_{Y_i|X_i}(y \mid x; \theta) = (\theta \cdot x)^{-1} \cdot \exp\left(\frac{-y}{x \cdot \theta}\right), \quad i = 1, 2, ..., n,$$

hvor $\theta \in \Theta$ er en ukendt parameter med parameterrum givet ved $\Theta = \{\theta \in \mathbb{R} : \theta > 0\}$. For denne fordeling gælder, at

$$E(Y_i \mid X_i = x) = x \cdot \theta.$$

Antag endelig, at de n par af stokastiske variable, $\{Y_i, X_i\}_{i=1}^n$, er uafhængige.

1. Vis at fordelingsfunktionen for $Y_i \mid X_i = x$ er givet ved

$$F_{Y_i|X_i}(y \mid x; \theta) = \begin{cases} 0 & \text{hvis } y \leq 0 \\ 1 - \exp\left(\frac{-y}{x \cdot \theta}\right) & \text{hvis } y > 0. \end{cases}$$

2. Opskriv sample likelihood-bidraget,

$$\ell(\theta \mid y_i, x_i),$$

sample likelihood funktionen,

$$L(\theta \mid y_1, ..., y_n, x_1, ..., x_n),$$

og den tilsvarende log-likelihood funktion. Angiv de antagelser du anvender undervejs.

3. Find et udtryk for maksimum likelihood estimatet, $\hat{\theta}_n$, som funktion af observationerne, $\{y_i, x_i\}_{i=1}^n$.

4. Betragt nedenstående tabel med beskrivende statistik for $n = 200$ observationer for variablene, $\{y_i\}_{i=1}^n$, $\{x_i\}_{i=1}^n$, $\{1/x_i\}_{i=1}^n$, $\{y_i x_i\}_{i=1}^n$, og $\{y_i/x_i\}_{i=1}^n$:

	y	x	$1/x$	$y \cdot x$	y/x
Mean	25.97	4.98	0.206	131.00	5.275
Variance	620.94	0.61	0.0012	17780	24.96
Standard dev.	24.92	0.78	0.0349	133.34	5.00
Skewness	1.56	0.04	1.078	1.89	1.57
Kurtosis	5.22	3.14	4.845	7.05	5.76

   (a) Er der nogle af variablene i tabellen, som ud fra den beskrivende statistik kan tænkes at være normalfordelt.

   (b) Brug informationen i tabellen til at finde værdien for estimatet i den konkrete model for de observerede levetider.

5. Vi vil nu udlede variansen på estimatoren:

   (a) Vis først, at bidraget til Hessematricen kan skrives som

$$H_i(\theta) = \frac{\partial^2 \log \ell (\theta \mid Y_i, X_i)}{\partial\theta\partial\theta} = \frac{1}{\theta^2} - \frac{2}{\theta^3}\frac{Y_i}{X_i}.$$

   (b) Find derefter informationen,

$$I(\theta) = -E\left[H_i(\theta) \mid X_i = x_i\right],$$

   hvor $E[\cdot \mid X_i = x_i]$ angiver den betingede forventning.

   (c) Brug informationen, $I(\theta)$, til at finde et udtryk for variansen på estimatoren.

   (d) Indsæt information for det konkrete eksempel og udregn $V(\hat{\theta}_n)$ og $\mathsf{se}(\hat{\theta}_n)$.

6. Der findes følgende resultater af estimationen

$$\hat{\theta}_n = 5.275 \quad \text{og} \quad V(\hat{\theta}_n) = 0.139.$$

For en enkelt enhed er ledningstykkelsen givet ved $x = 5$.

Udregn på baggrund af den estimerede model sandsynlighed for at levetiden er større end $y = 50$ måneder for denne enhed.

7. Vi vil til slut undersøge, om den sande værdi af θ kan tænkes at være givet ved $\theta_0 = 5$.

Formuler spørgsmålet som en statistisk hypotese og test hypotesen for det konkrete tilfælde. Vær præcis om nul-hypotese, alternativ hypotese, test-størrelse og den relevante kritiske værdi.