Eksamen på Økonomistudiet vinteren 2016-17
Sandsynlighedsteori og Statistik
2. årsprøve
21. februar, 2017
(3-timers prøve med hjælpemidler)

Dette eksamenssæt består af 5 sider (forsiden inklusiv).

OBS: Bliver du syg under selve eksamen på Peter Bangsvej, skal du kontakte et tilsyn, blive registreret som syg hos denne. Derefter afleverer du en blank besvarelse i systemet og forlader eksamen. Når du kommer hjem, skal du kontakte din læge og indsende lægeerklæring til Det Samfundsvidenskabelige Fakultet senest en uge efter eksamensdagen.

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 ses på et marked for brugte cykler. Vi antager, at værdien af brugte cykler kan beskrives ved en normalfordeling. Lad X være den stokastiske variabel, som angiver værdien af den brugte cykel (i kr). Der gælder, at $X \sim N(1000, 8100)$.

1. Udregn sandsynligheden for at en brugt cykler på markedet har en værdi der er større end 1100 kr., $P(X > 1100)$.

Vi antager nu, at prisen på en brugt cykel afhænger af den sande værdi X, men at det ikke er muligt præcist at vurdere værdien af cyklen. Derfor afhænger prisen også af en "målefejl", Z. Z er en stokastisk variabel, som er normalfordelt, $Z \sim N(0, 100)$ og uafhængig af X. Prisen, Y, er givet ved

$$Y = 50 + X + Z.$$

2. Angiv fordelingen af Y og udregn den forventede værdi af prisen, $E(Y)$ og variansen af $Y, Var(Y)$.

3. Udregn kovariansen mellem værdien, X, og prisen, $Y : Cov(X, Y)$ og angiv om X og Y er uafhængige.

4. Udregn den forventede pris, når kvaliteten er lig 1100, $E(Y|X = 1100)$.

Opgave 2

Lad X være ligefordelt på $(1, 2)$ dvs. X har tæthed $p(x) = \mathbf{1}(1 < x < 2)$.

1. Find $E(X)$ og $Var(X)$.

2. Vi sætter nu $Y = X - 1$. Find $E(Y)$ og $Var(Y)$.

3. Find tætheden $q(y)$ for Y.

4. Find $E\left(X|X > \frac{3}{2}\right)$.

Opgave 3

Vi betragter et forsikringsselskab som forsikrer en række meget værdifulde kunstgenstande. Selskabets ejere er nervøse for udbetalingen af store erstatninger, dvs. erstatninger større end en given værdi, $c = 1000\$$, og selskabet har observeret $n = 400$ erstatningsudbetalinger større end 1000\$, kaldet $\{y_i\}_{i=1}^n$. Observationerne betragtes som realisationer af stokastiske variable, $\{Y_i\}_{i=1}^n$, med udfaldsrum givet ved $Y_i \in \mathbb{Y} = \{y \in \mathbb{R} : y > 1000\}$.

Det antages at Y_i og Y_j er uafhængigt og identisk fordelt for alle $i \neq j$, og som statistisk model for de store udbetalinger anvendes en Pareto-fordeling, der har tæthedsfunktion givet ved

$$f_{Y_i}(y_i \mid \theta) = \theta c^\theta y_i^{-(1+\theta)}, \quad y_i > c, \tag{1}$$

hvor $c = 1000$ er kendt og parameteren er givet ved $\theta \in \Theta = \{\theta \in \mathbb{R} : \theta > 0\}$. Figuren nedenfor viser tætheds-funktionerne for to Pareto-fordelinger med $c = 1000$ og parametre henholdsvis $\theta = 1.5$ og $\theta = 2.5$ i intervallet $y \in [1000 : 4000]$. Tabellen nedenfor viser beskrivende statistik for de $n = 400$ observerede udbetalinger som skal modelleres.

1. Kommenter på den beskrivende statistik og sammenlign med egenskaberne for en normalfordeling.

2. Vis, at fordelingsfunktionen for en Pareto-fordeling kan skrives som

$$F_{Y_i}(y_i \mid \theta) = 1 - c^\theta y_i^{-\theta}. \tag{2}$$

3. Opskriv likelihood-bidraget, $\ell(\theta \mid y_i)$, for den valgte model.

   Opskriv også likelihood funktionen, $L(\theta \mid y_1, ..., y_n)$, og den tilsvarende log-likelihood funktion, $\log L(\theta \mid y_1, ..., y_n)$. Angiv de antagelser du anvender undervejs.

4. Vis at maksimum-likelihood estimatet kan skrives som

$$\hat{\theta}(y_1, ..., y_n) = \frac{n}{\sum_{i=1}^{n} \log(y_i/c)},$$

som i det konkrete tilfælde kan udregnes til $\hat{\theta}(y_1, ..., y_n) = 2.64$.

5. Brug den estimerede parameter, $\hat{\theta}$, til at udregne sandsynligheden for at erstatningsudbetalingen en givet dag er større en 2500\$ hhv. 3000\$.

   Forklar hvordan resultaterne passer med den beskrivende statistik i tabellen og hvordan denne sammenligning kan bruges til model-kontrol.

6. Find bidraget til Hessematricen fra observation i, dvs.

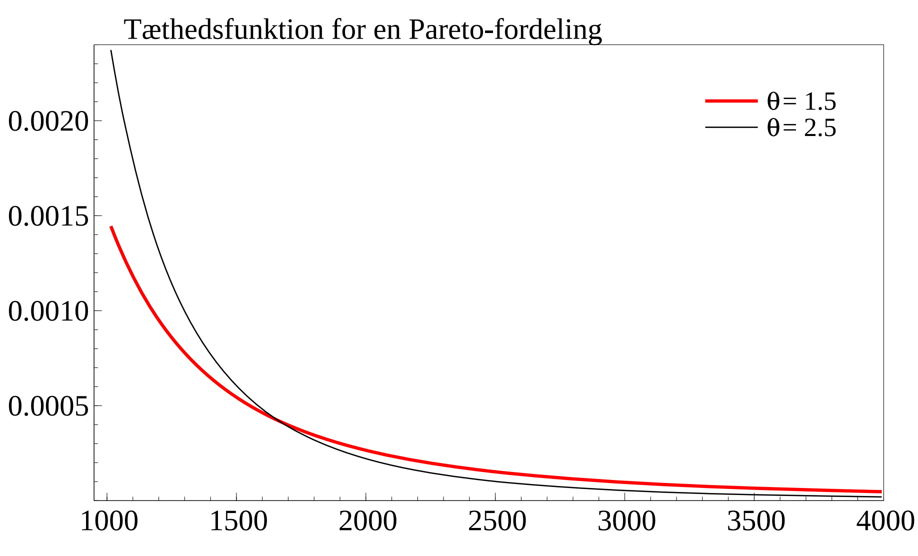
$$H_i(\theta_0) = \left. \frac{\partial^2 \log \ell(\theta \mid Y_i)}{\partial\theta\partial\theta} \right|_{\theta=\theta_0},$$

   hvor θ_0 angiver den sande værdi af parameteren.

   Brug dette til at finde variansen og standardfejlen på estimatoren, dvs. $V(\hat{\theta}_n)$ og $\mathsf{se}(\hat{\theta}_n)$.

7. Opskriv den asymptotiske fordeling at estimatoren $\hat{\theta}(Y_1, ..., Y_n)$, dvs. den approximative fordeling når n bliver stor.

   Angiv et 95% konfidens interval for θ_0 baseret på den asymptotiske fordeling og forklar hvordan det skal fortolkes.



Beskrivende statistik for udbetalinger	
Minimum	1001.3
Maximum	13367.0
Mean	1604.1
Standard deviation	1011.1
Skewness	5.7
Kurtosis	53.3
5% quantile	1021.8
10% quantile	1041.9
25% quantile	1133.3
Median	1293.5
75% quantile	1658.8
90% quantile	2428.9
95% quantile	3137.2
n	400