Eksamen på Økonomistudiet vinteren 2015-16
Sandsynlighedsteori og Statistik
2. årsprøve
15. januar, 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.

I de to første opgaver ser vi på et forsikringsselskab. Forsikringsselskabet forsikrer to typer af personer.

For personer af type I antager vi, at antallet af skader i løbet af 5 år kan beskrives ved en Poissonfordeling med parameter $\lambda_I = 0.5$. Lad X_I angive antallet af skader for en person af type I.

For personer af type II antager vi, at antallet af skader kan beskrives ved en Poissonfordeling med parameter $\lambda_{II} = 1$. Lad X_{II} angive antallet af skader for en person af type II.

Opgave 1

1. Udregn sandsynligheden for at en person af hhv. type I og II ingen skader har.

2. Udregn det forventede antal skader for henholdsvis type I og II.

3. Antag at forsikringsselskabet har 300 forsikringstagere, og at de fordeler sig således: 200 personer af type I og 100 af type II. Lad Z være antallet af skader for alle de forsikrede. Vi antager, at de enkelte forsikringstagere er uafhængige.

   Angiv det forventede antal skader $E(Z)$.

4. Angiv fordelingen af det samlede antal skader Z.

Opgave 2

I denne opgave undersøges, hvordan forsikringsselskabet skal fastlægge sine præmier, hvis forsikringsselskabet ikke ved, om personen er type I eller type II.

Vi antager, at forsikringsselskabet har kunder i to forskellige regioner, og at man kender fordelingen af type I og type II kunder i de to regioner A og B. Vi antager også, at Y angiver, om kunden er type I eller type II og R angiver regionen. Lad Y og R være stokastiske variable, hvor deres simultane fordeling er angivet i tabel 1. $Y = 1$ betyder, at personen er type I.

Tabel 1: *Den simultane fordeling af type Y og region R*

		Y	
		$Y = 1$ (type I)	$Y = 0$ (type II)
	Region A	0.10	0.20
R	Region B	0.35	0.35

1. Angiv den marginale fordeling for region (R).

2. Beskriv i ord hvad sandsynligheden $P(Y = 1|R = A)$ angiver. Udregn derefter sandsynligheden.

3. Angiv om Y og R er uafhængige. Begrund dit svar.

4. Antallet af skader for en tilfældig person kan beskrives som

$$Y \cdot X_I + (1 - Y) \cdot X_{II},$$

   hvor X_I og X_{II} er Poisson-fordelt med parametrene $\lambda_I = 0.5$ og $\lambda_{II} = 1$. Desuden antages at X_I og X_{II} er uafhængige af Y og R.

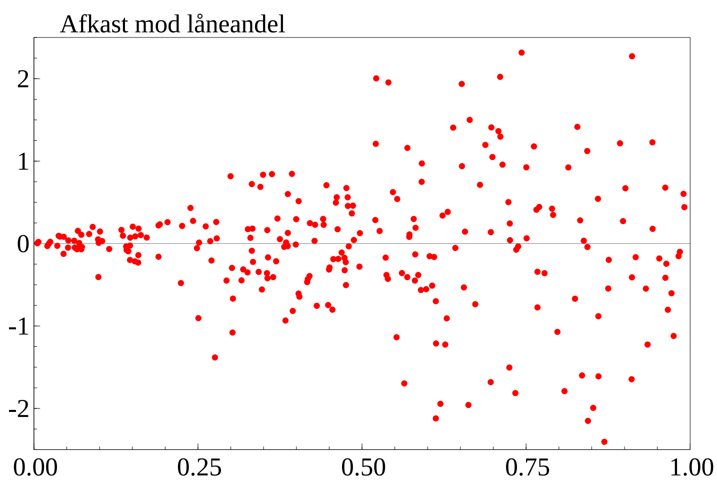
   Beregn det forventede antal skader for en person som bor i region A.

   Forsikringsselskabet ønsker at fastlægge prisen på præmien, således at præmien dækker den forventede udgift til skader. Forsikringsselskabet kan ikke observere, om personen er type I eller type II, kun hvilken region personen bor i.

   Diskuter hvorfor det er optimalt for forsikringsselskabet at have en prispolitik, som diskriminerer mellem regionerne.

Opgave 3

Denne opgave handler om usikkerhed på aktie-afkast fra finansielle virksomheder. Vi har observeret afkastene fra n forskellige virksomheder, kaldet $\{y_i\}_{i=1}^n$, som vi betragter som realisationer af stokastiske variable, $\{Y_i\}_{i=1}^n$, med udfaldsrum givet ved $Y_i \in \mathbb{Y} = \mathbb{R}$. En vigtig bestemmende faktor for usikkerheden på afkastene er andelen af virksomhedernes samlede udlån, der går til særligt udsatte brancher. Vi har observeret denne andel, kaldet $\{x_i\}_{i=1}^n$. Igen betragtes observationerne som realisationer af stokastiske variable, $\{X_i\}_{i=1}^n$, og udfaldsrummet er givet ved $X_i \in \mathbb{X} = \{x \in \mathbb{R} : 0 < x < 1\}$. Et krydsplot af afkastene for $n = 250$ virksomheder imod deres låneandel til udsatte brancher er vist i følgende figur.



For at modellere usikkerheden som funktion af udlånsandelen, antager vi at afkastet, Y_i, givet låneandelen til udsatte brancher, $X_i = x_i$, er beskrevet ved en betinget normalfordeling,

$$(Y_i \mid X_i = x_i) \overset{d}{=} N(\mu_i, \sigma_i^2), \quad i = 1, 2, ..., n, \tag{1}$$

med

$$\mu_i = 0,$$

så det forventede afkast er nul for alle virksomheder, mens variansen er en funktion af udlånsandelen,

$$\sigma_i^2 = \beta \cdot x_i^2.$$

Her er $\beta \in \Theta$ parameteren vi er interesseret i at estimere, med parameterrum givet ved $\Theta = \{\beta \in \mathbb{R} \mid \beta > 0\}$.
Vi antager, at de n par af stokastiske variable, $\{Y_i, X_i\}_{i=1}^n$, er uafhængige.

1. Forklar hvorfor modellen er en relevant beskrivelse af data i figuren.

2. Vis at modellen også kan skrives som

$$Y_i = \sigma_i \cdot Z,$$

   hvor $\sigma_i = \sqrt{\beta x_i^2}$ og $(Z \mid X_i = x_i) \overset{d}{=} N(0,1)$.

3. Skriv tæthedsfunktionen for fordelingen i (1).

   Opskriv likelihood-bidraget, $\ell(\beta \mid y_i, x_i)$, likelihood funktionen, $L(\beta \mid y_1, ..., y_n, x_1, ..., x_n)$, og den tilsvarende log-likelihood funktion. Angiv de antagelser du anvender undervejs.

4. Find et udtryk for maksimum likelihood estimatoren, $\hat{\theta}_n$, som funktion af $\{Y_i, X_i\}_{i=1}^n$.

5. Udled variansen på estimatoren, ved først at finde bidraget til Hessematricen fra observation i,

$$H_i(\beta_0) = \left. \frac{\partial^2 \log \ell \left(\beta \mid Y_i, X_i\right)}{\partial\beta\partial\beta} \right|_{\beta=\beta_0}.$$

   Dernæst informationen,

$$I(\beta_0) = -E\left[H_i(\beta_0) \mid X_i = x_i\right],$$

   hvor $E\left[\cdot \mid X_i = x_i\right]$ angiver den betingede forventning. Find endelig variansen på estimatoren, $V(\hat{\beta}_n)$, som funktion af informationen.

6. Forklar med dine egne ord, hvad det betyder at estimatoren, $\hat{\beta}_n$, er konsistent.

   Angiv den asymptotiske fordeling af estimatoren, $\hat{\beta}_n$.

   Forklar hvorfor den sande værdi af parameteren, β_0, indgår i udtrykket for variansen af estimatoren.

   Den sande værdi, β_0, er ukendt. Forklar hvordan man så bruger resultatet i praksis.

7. I det konkrete eksempel med $n = 250$ virksomheder finder man følgende resultater

$$\hat{\beta}_n = 2.037 \quad \text{og} \quad V(\hat{\beta}_n) = 0.0333.$$

Angiv et 95% konfidens interval for β_0 baseret på den asymptotiske fordeling og forklar hvad det betyder.

Test med et hypotese-test om den sande værdi af parameteren kan tænkes at være givet ved $\beta_0 = 2.5$. Vær præcis om nul-hypotese, alternativ hypotese, test-størrelse og den relevante kritiske værdi.