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

Dette eksamenssæt består af 6 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 undersøges, hvordan et forsikringsselskab skal fastlægge sine præmier på forskellige forsikringsprodukter. Antag at vi har et forsikringsselskab, som udbyder cykelforsikringer. Vi antager, at alle deres forsikringstagere har en cykel med værdien 5.000 kr. Sandsynligheden per år for at få stjålet en cykel er $p = 0.06$. Vi antager, at hver forsikringstager højst kan få stjålet én cykel om året. Forsikringsselskabet tilbyder en forsikring uden selvrisiko. Præmien for denne forsikring er 350 kr. om året. Vi antager, at forsikringsselskabet har 10 kunder, som har tegnet en forsikring uden selvrisiko. Lad $X$ angive antallet af kunder som får stjålet en cykel i løbet af et år.

1. Udregn sandsynligheden for at 2 af de 10 kunder får deres cykel stjålet, $P(X = 2)$. Angiv de antagelser man er nødt til at lave for at kunne lave beregningen.

Vi definerer nu en stokastisk variabel, som angiver forsikringsselskabets indtjening (i kr.), $Z$. Indtjeningen er givet ved

$$Z = 350 \cdot 10 - 5000 \cdot X.$$

2. Udregn den forventede indtjening $E(Z)$ og variansen på indtjeningen $Var(Z)$ og standardafvigelsen på $Z$.

Forsikringsselskabet har også en anden type forsikring med en selvrisiko på 1.000 kr. Prisen for denne forsikring er 290 kr. Antag at 10 nye forsikringstagere vælger denne forsikring. Det antages, at de har samme sandsynlighed for at få stjålet en cykel: $p = 0.06$. Lad $Y$ angive antallet af stjålne cykler for de nye forsikringstagere (med en selvrisiko).

3. Udregn sandsynlighed for at der i alt bliver stjålet 4 cykler blandt de 20 forsikringstagere, $P(X + Y = 4)$.

4. Udregn den forventede indtjening på de 10 nye kunder. Udregn derefter variansen og standardafvigelsen for fortjeneste på denne forsikring. Forklar hvorfor det kan være attraktivt for forsikringsselskabet at udbyde denne type forsikring i stedet for forsikringen uden selvrisiko.

## Opgave 2

1. Lad $X$ være normalfordelt $N\left(0, \sigma_x^2\right)$. Find fordelingen af $Z_1 = X + 1$.

2. Lad $Y$ være normalfordelt $N\left(0, \sigma_y^2\right)$. Find fordelingen af $Z_2 = 2Y$.

3. Antag at $X$ og $Y$ er uafhængige. Vis at (forklar hvilke regneregler du benytter):

$$\begin{aligned} E\left(Z_1 | Z_2 = z_2\right) &= 1 \\ V\left(Z_2 | Z_1 = z_1\right) &= 4\sigma_y^2. \end{aligned}$$

4. Under antagelsen at $X$ og $Y$ er uafhængige, er $(X, Y)$ $N(\mu, \Omega)$ hvor

$$\mu = (0,0)' \text{ og } \Omega = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}.$$

   Definer nu vektoren $W = (W_1, W_2)'$ ved

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} X \\ X + 2Y \end{pmatrix}.$$

   Find $E\left(W_1 W_2\right)$ og $E\left(W_2 | W_1\right)$.

5. Vis at $(W_1, W_2)$ er normalfordelt,

$$N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_x^2 \\ \sigma_x^2 & 4\sigma_y^2 \end{pmatrix}\right).$$

## Opgave 3

Denne opgave handler om en politi-enhed som gennemfører paragraf 77 eftersyn på indfaldsveje til København. I $n = 60$ tilfælde har man haft politi patruljer ude og gennemføre kontroller og de har hver kontrolleret et antal biler indtil der er noteret $k = 2$ ulovlige forhold. Antallet af kontrollerede biler for hver patrulje er noteret som $\{y_i\}_{i=1}^n$. Figuren nedenfor viser den empiriske fordelingen af antal kontrollerede biler for hver patrulje, og det samlede antal er $\sum_{i=1}^{60} y_i = 748$.

For at modellere antal af kontrollerede biler antages $y_i$ at være en realisation af en stokastisk variabel $Y_i$, med udfaldsrum givet ved $Y_i \in \mathbb{Y} = \{2, 3, 4, ...\}$. Som model for det antal biler der skal kontrolleres før der er noteret to ulovlige forhold anvendes en negativ binomial fordeling med antalsparameter $k = 2$, så

$$Y_i \overset{d}{=} \mathrm{NB}(2, \theta), \tag{1}$$

hvor sandsynlighedsparameteren, $\theta$, måler sandsynligheden for at hver tilfældig bil $i$ skal noteres for et ulovligt forhold. Den negative binomialfordeling har en sandsynlighedsfunktion givet ved

$$f_{Y_i}(y_i \mid \theta) = (y_i - 1)\,\theta^2 (1-\theta)^{(y_i - 2)}, \tag{2}$$

for $y_i \in \mathbb{Y}$ og $\theta$ liggende i parameterrummet $\Theta$. Det oplyses at den forventede værdi er givet som

$$E(Y_i) = 2 + \frac{2\,(1-\theta)}{\theta}, \tag{3}$$

og det antages at alle $\{Y_i\}_{i=1}^n$ har samme fordeling og er uafhængige.

1. Angiv det relevante parameterrum $\Theta$, sådan at $\theta \in \Theta$.

   Opskriv sample likelihood-bidraget, $\ell(\theta \mid y_i)$, sample likelihood funktionen, $L(\theta \mid y_1, ..., y_n)$, og de tilsvarende log-likelihood bidrag og log-likelihood funktion. Angiv de antagelser du anvender undervejs.

2. Angiv første-ordens betingelsen for maksimum af log-likelihood funktionen og udled maksimum likelihood estimatoren, $\hat{\theta}_n$, for den givne model. Brug informationen i teksten til at bestemme maksimum likelihood estimatet i det givne tilfælde.

3. Udled bidraget til Hesse-matricen:

$$H_i(\theta) = \frac{\partial^2 \log \ell\,(\theta \mid Y_i)}{\partial\theta\partial\theta},$$

   og brug det til at finde variansen på estimatoren.

   Udregn $V(\hat{\theta}_n)$ og $\mathsf{se}(\hat{\theta})$ i det konkrete tilfælde med observationer for de $n = 60$ politi-patruljer.

4. Det oplyses nu at $\hat{\theta}_n = 0.160$ og $\mathsf{se}(\hat{\theta}_n) = 0.0134$. Udregn det forventede antal kontroller der skal til for en tilfældig patrulje før der er noteret $k = 2$ ulovligheder.

   I afdelingen er der kun budgetteret med tid til 10 kontroller per patrulje. Udregn hvilken sandsynlighedsparameter det svarer til.

5. Test med et hypotese test på $\theta$ om det forventede antal kontroller i modellen er signifikant forskelligt fra budgettet om 10 kontoller. Vær præcis om nul-hypotese, alternativ hypotese, test-størrelse og den relevante kritiske værdi.

6. OxMetrics koden til estimation af modellen ovenfor kan fx skrives som

```
actual = Y;
theta  = &1;
fitted = theta;
loglik = log(Y-1)+2*log(theta)+(Y-2)*log(1-theta);
&1     = 0.5;
```

   Vi vil nu modificere modellen så sandsynligheden for at der findes ulovligheder på bil $i$ afhænger af bilens alder, noteret som $\{x_i\}_{i=1}^n$. Den modificerede model for $(Y_i \mid X_i = x_i)$ skrives som

$$(Y_i \mid X_i = x_i) \overset{d}{=} \mathrm{NB}(2, \theta_i), \tag{4}$$

   med sandsynligheds-parameter

$$\theta_i = \alpha + \beta x_i. \tag{5}$$

Skriv likelihood-bidraget for den betingede model, dvs.

$$\ell(\alpha, \beta \mid y_i, x_i) = f_{Y_i|X_i}(y_i \mid x_i; \alpha, \beta).$$

Modificer OxMetrics koden ovenfor til at estimere den betingede model.