Institutt for matematiske fag

Eksamensoppgave i **TMA4240 Statistikk - Solution sketch**

Faglig kontakt under eksamen: [a], [b]
Tlf: [a] , [b]

Eksamensdato: 09. desember 2021
Eksamenstid (fra–til): 09:00–13:00
Hjelpemiddelkode/Tillatte hjelpemidler:

Målform/språk: bokmål
Antall sider: ??
Antall sider vedlegg: 0

Kontrollert av:

Informasjon om trykking av eksamensoppgave	
Originalen er:	
1-sidig ☐	**2-sidig** ☒
sort/hvit ☒	**farger** ☐
skal ha flervalgskjema ☒	

Dato   Sign

Oppgave 1 Automatically corrected exercises

1a) We have that $X \sim N(2,1)$ and $Y \sim N(-2,3^2)$. Let $Z \sim N(0,1)$, then:

- $P(X \geq 3) = P(\frac{X-1}{1} \geq \frac{3-1}{1}) = P(Z \geq 1) = 0.158$
- We find that

$$E(2X - Y) = 4 + 1 = 5$$
$$\text{Var}(2X - Y) = 4\text{Var}(X) + \text{Var}(Y) = 13$$

then

$$P(2X - Y \leq 0) = P(Z \leq \frac{0-5}{\sqrt{13}}) = 0.083$$

- $$P(X^2 \leq \frac{1}{4}) = P(-\frac{1}{2} \leq X \leq \frac{1}{2}) = P(X \leq \frac{1}{2}) - P(X \leq -\frac{1}{2}) = 0.06$$

1b)

- The number of possible reference groups consisting of 5 students can be found as

$$\binom{55}{5} = 3478761$$

- If we require that at each group contains at least 2 girls and two boys we have that:

$$\binom{25}{3}\binom{30}{2} + \binom{25}{2}\binom{30}{3} = 2218500$$

1c) $X \sim \text{Bin}(25, 0.33)$ then

$$E(X) = np = 8.25$$
$$\text{Var}(X) = np(1-p) = 5.5275$$

then we hve

$$Z = \frac{X - 8.25}{\sqrt{5.5275}} \approx N(0,1)$$

- $P(X < 5) \approx P(X < 4.5) = 0.055$
- $P(X > 10) \approx P(X > 10.5) = 0.169$
- $P(X > 10 | X \geq 4) = \frac{P(X>10, X\geq 4)}{P(X\geq 4)} = \frac{P(X>10)}{P(X\geq 4)} \approx 0.171$

1d) The table gives the joint probabilities $P(X = x, Y = y)$. The given quantities can be computed as:

- $P(X = 1) = P(X = 1, Y = 0) + P(X = 1, Y = 1) + P(X = 1, Y = 2) = 0.32 + 0.15 = 0.47$
- $P(Y = 0|X = 0) = \frac{P(Y=0,X=0)}{P(X=0)} = \frac{0.22}{0.22+0.31} = 0.415$
- We have $g(X, Y) = 2XY$ then $E(g(X, Y)) = \sum_{x,y} g(x, y)P(X = x, Y = y) = 2 * 0.15 = 0.3$

Oppgave 2 Probability in continuous distributions

- $P(X \geq 2) = \int_2^\infty \frac{2x}{10}e^{-x^2/10}dx = -e^{-x^2/10}|_2^\infty = e^{-4/10} = 0.67$
- $P(X^2 - 4X \leq -1) = P(X^2 - 4X - 1 \leq 0) = \int_{2-\sqrt{3}}^{2+\sqrt{3}} \frac{2x}{10}e^{-x^2/10}dx = -e^{-x^2/10}|_{2-\sqrt{3}}^{2+\sqrt{3}} = 0.744$
- The median m is defined as

$$\int_0^m \frac{2x}{10}e^{-x^2/10}dx = 0.5$$

Then:

$$\begin{aligned}\int_0^m \frac{2x}{10}e^{-x^2/10}dx &= -e^{-x^2/10}|_0^m \\ &= 1 - e^{-m^2/10} = 0.5\end{aligned}$$

which gives:

$$\begin{aligned}m^2 &= 10\log(2) \\ m &= \sqrt{10\log(2)} = 2.63\end{aligned}$$

Oppgave 3 Lifetime

3a)
- We have that $P(T < t) = 1 - \exp(-\frac{t}{\lambda})$, then

$$P(\frac{2T}{\lambda} < t) = P(T < \frac{t\lambda}{2}) = 1 - \exp(-t/2)$$

and the pdf of $\frac{2T}{\lambda}$ is $\frac{1}{2}\exp(-t/2)$ which is the pdf of a gamma distributed SV with $\alpha = 1$, $\beta = 1$. This is equivalent to a χ^2 distribution with 2 degrees of freedom.

- We can write
$$V = \frac{2\sum_{i=1}^n T_i}{\lambda} = \sum \frac{2T_i}{\lambda}$$
The T_i's are mutually independent and $\frac{2T_i}{\lambda}$ has a χ^2 distribution with 2 degrees of freedom. So V is the sum of n independent and χ^2 distributed SV, therefore is also V χ^2 distributed with $2n$ degrees of freedom.

3b) To determine the 90% confidence interval we start from
$$V = \frac{2\sum_{i=1}^n T_i}{\lambda} \sim \chi^2_{2n}$$
We have that n = 20 and $\alpha = 0.05$,then:
$$P(\chi^2_{0.95,40} < V < \chi^2_{0.05,40}) = 0.9$$
From the table we get that $\chi^2_{0.95,40} = 26.509$ and $\chi^2_{0.05,40} = 55.758$, which gives the 90% CI:
$$[\frac{2\sum T_i}{55.758}; \frac{2\sum T_i}{26.509}]$$
setting in the numerical value for $\bar{T}$ we get: $[7.53, 15.84]$.

Oppgave 4 Accidents

4a)
- $P(X_i \geq 1) = 1 - P(X = 0) = 1 - e^{-\lambda m_i} = 1 - e^{-0.02\ 50} = 0.632$
- $P(X_i > 1|X_i \geq 1) = \frac{P(X_i>1, X_i \geq 1)}{P(X_i \geq 1)} = \frac{P(X_i>1)}{P(X_i \geq 1)} = \frac{1-P(X_i=0)pP(X_i=1)}{0.632} = 0.418$
- $P(X_i \leq 3|X_i \geq 1) = \frac{P(X_i \leq 3, X_i \geq 1)}{P(X_i \geq 1)} = \frac{P(X_i=1)+P(X_i=2)+P(X_i=3)}{P(X_i \geq 1)} = \frac{0.613}{0.632} = 0.97$

4b) The likelihood function is
$$L(\lambda; x_1, \ldots, x_n) = \prod_{i=1}^n \frac{(\lambda m_i)^{x_i}}{x_i!} e^{-\lambda m_i} = \prod_{i=1}^n \frac{m_i^{x_i}}{x_i!} \lambda^{\sum x_i} e^{-\lambda \sum m_i}$$
and the log-likelihood
$$l(\lambda; x_1, \ldots, x_n) = \log(\prod_{i=1}^n \frac{m_i^{x_i}}{x_i!}) + \sum x_i \log(\lambda) - \lambda \sum m_i$$

Deriving wrt λ we get:

$$\frac{dl(\lambda)}{d\lambda} = \frac{\sum x_i}{\lambda} - \sum m_i = 0$$

and we get $\hat{\lambda} = \frac{\sum X_i}{\sum m_i}$
Moreover we have that:

$$E(\hat{\lambda}) = E(\frac{\sum X_i}{\sum m_i}) = \frac{\sum E(X_i)}{\sum m_i} = \lambda$$

and

$$\text{Var}(\hat{\lambda}) = \frac{\sum \text{Var}(X_i)}{(\sum m_i)^2} = \frac{\lambda \sum m_i}{(\sum m_i)^2} = \frac{\lambda}{\sum m_i}$$

4c) We have that

$$H_0 : \lambda = 0.02,$$
$$H_1 : \lambda \leq 0.02$$

Under H_0 we have that

$$Z = \frac{\hat{\lambda} - 0.02}{\sqrt{\frac{0.02}{1784.4}}} \sim N(0, 1)$$

A decision rule is the that we reject H_0 if $Z < k$ or equivalently $\hat{\lambda} < k\sqrt{\frac{0.02}{1784.4}} + 0.02$
The p-value is the probabilty, under H_0 of observing a value more extreme than the observed one. In our case we have that $\hat{\lambda} = \frac{\sum x_i}{\sum m_i} = \frac{33}{1784.4} \approx 0.0185$.
The p-value is then:

$$P(\hat{\lambda} < 0.0185|H_0) = P(Z < \frac{0.0185 - 0.02}{0.003}|H_0) = P(Z < -0.5) = 0.31$$

The p-value is so big that we would not reject the null hypothesis at all reasonable levels.

4d) When we observe the process for t years the MLE for λ is

$$\hat{\lambda} = \frac{\sum X_i}{t \sum m_i}$$

this is unbiased and with variance $\text{Var}(\hat{\lambda}) = \frac{\lambda}{t\sum m_i}$. This means that under H_0:

$$\frac{\hat{\lambda} - 0.02}{\sqrt{\frac{0.02}{t\sum m_i}}} \approx N(0,1)$$

when and at a $\alpha = 0.05$ significance level we reject the null hypothesis if $Z < -1.645$ or equivalently if

$$\hat{\lambda} < -1.645\sqrt{\frac{0.02}{1784.4t}} + 0.02$$

The power of the test is then:

$$P(\hat{\lambda} < -1.645\sqrt{\frac{0.02}{1784.4t}} + 0.02 | \lambda = 0.018) =$$

$$P(\frac{\hat{\lambda} - 0.018}{\sqrt{\frac{0.018}{1784.4t}}} < \frac{-1.645\sqrt{\frac{0.02}{1784.4t}} + 0.02 - 0.018}{\sqrt{\frac{0.018}{1784.4t}}} | \lambda = 0.018) =$$

$$P(Z < -1.645\sqrt{\frac{0.02}{0.018}} + 0.002\frac{\sqrt{1784.4t}}{\sqrt{0.018}} | \lambda = 0.018)$$

We want this probability to be at least 0.2, therefore we need:

$$-1.645\sqrt{\frac{0.02}{0.018}} + 0.002\frac{\sqrt{1784.4t}}{\sqrt{0.018}} > -0.842$$

$$\sqrt{t} > 1.416$$

which gives $t > 2.005$ so we need to observe the process for at least 3 years.

4e) When $t = 1$ we have that

$$E(\tilde{\lambda}) = \lambda\sum_{i=1}^{n} b_i m_i \text{ and } \text{Var}(\tilde{\lambda}) = \lambda\sum_{i=1}^{n} b_i^2 m_i$$

we want $\tilde{\lambda}$ to be unbiased, therefore we require

$$E(\tilde{\lambda}) = \lambda$$

$$\sum_{i=1}^{n} b_i m_i = 1$$

$$b_1 = \frac{1}{z_1}(1 - \sum_{i=2}^{n} b_i m_i)$$

Moreover, we want $\text{Var}(\tilde{\lambda})$ to be minimal under the condition that $b_1 = \frac{1}{z_1}(1 - \sum_{i=2}^n b_i m_i)$. We want to minimize

$$\begin{aligned}\sum_{i=1}^n b_i^2 m_i &= b_1^2 z_1 + \cdots + b_n^2 z_n \\ &= \frac{1}{z_1^2}(1 - \sum_{i=2}^n b_i m_i)^2 z_1 + \cdots + b_n^2 z_n \\ &= \frac{1}{z_1}(1 - \sum_{i=2}^n b_i m_i)^2 + \sum_{j=2}^n b_j^2 z_j\end{aligned}$$

The partial derivatives are

$$\begin{aligned}\frac{\partial \sum_{i=1}^n b_i^2 m_i}{\partial b_2} &= \frac{2}{z_1}(1 - \sum_{j=2}^n b_j z_j)(-z_2) + 2b_2 z_2 &= -\frac{1}{z_1}(1 - \sum_{j=2}^n b_j z_j) + b_2 = 0 \\ \frac{\partial \sum_{i=1}^n b_i^2 m_i}{\partial b_3} &= \frac{2}{z_1}(1 - \sum_{j=2}^n b_j z_j)(-z_3) + 2b_3 z_3 &= -\frac{1}{z_1}(1 - \sum_{j=2}^n b_j z_j) + b_3 = 0 \\ &\cdots \\ \frac{\partial \sum_{i=1}^n b_i^2 m_i}{\partial b_n} &= \frac{2}{z_1}(1 - \sum_{j=2}^n b_j z_j)(-z_n) + 2b_n z_n &= -\frac{1}{z_1}(1 - \sum_{j=2}^n b_j z_j) + b_n = 0\end{aligned}$$

This can be written as:

$$\begin{aligned}-b_1 + b_n &= 0 \\ -b_2 + b_n &= 0 \\ &\cdots \\ -b_n + b_n &= 0\end{aligned}$$

So we need to have $b_1 = b_2 = \cdots = b_n = b$. In addition we want that

$$1 = \sum_{i=1}^n b_i m_i = b \sum_{i=1}^n m_i$$

which gives $b = \frac{1}{\sum_{i=1}^n m_i}$.