Kunnskap for en bedre verden

Institutt for matematiske fag

Eksamensoppgave i **TMA4240 Statistikk**

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

Tlf: [a] , [b]

Eksamensdato: 09. desember 2021

Eksamenstid (fra–til): 13:00–17:00

Hjelpemiddelkode/Tillatte hjelpemidler:

Annen informasjon:

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

- $P(X \geq 0) = P(\frac{X+2}{2} \geq \frac{0+2}{2}) = P(Z \geq 1) = 0.158$
- We find that

$$E(2X - Y) = -4 - 1 = -5$$
$$\text{Var}(2X - Y) = 4\text{Var}(X) + \text{Var}(Y) = 17$$

  then

$$P(2X - Y \geq -1) = P(Z \geq \frac{-1+5}{\sqrt{17}}) = 0.166$$

- We have that $E(Y - X) = 1 + 2 = 3$ and $\text{Var}(Y - X) = 1 + 4 = 5$

$$P(Y > X) = P(Y - X > 0) = P(Z > -3/\sqrt{5}) = 0.91$$

1b)

- We can choose 6 out of 35 balls in

$$\binom{35}{6} = 1623160$$

  different ways.
- We can choose 6 out of 35 balls, such that there are at least 2 blue and 2 red in

$$\binom{20}{4}\binom{15}{2} + \binom{20}{3}\binom{15}{3} + \binom{20}{2}\binom{15}{4} = 1286775$$

1c) We know that $\bar{X}$ is approximately normally distributed with $E(\bar{X}) = 3.5$ and $\text{Var}(\bar{X}) = 1.5/50$. We then have that:

- $P(\bar{X} < 4) = 0.998$
- $P\left(\sum_{i=1}^{50} X_i \leq 185\right) = P\left(\bar{X} \leq 185/50\right) = 0.875$
-

$$\begin{aligned}
P(\bar{X} \leq 4 | \sum_{i=1}^{10} = 48.3) &= P(\sum_{i=1}^{50} X_i \leq 200 | \sum_{i=1}^{10} X_i = 48.3) \\
&= P(\sum_{i=1}^{40} X_i \leq 151.8) \\
&= P(\frac{\frac{1}{40}\sum_{i=1}^{40} X_i - 3.5}{\sqrt{\frac{1.5}{40}}} \leq \frac{\frac{151.8}{40} - 3.5}{\sqrt{\frac{1.5}{40}}}) \\
&= P(Z \leq 1.523) = 0.934
\end{aligned}$$

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

- $P(Y = 1) = P(X = 0, Y = 1) + P(X = 1, Y = 1) + P(X = 2, Y = 1) = 0.1 + 0.2 = 0.3$
- $P(Y = 0|X = 0) = \frac{P(Y=0,X=0)}{P(X=0)} = \frac{0.4}{0.4+0.1} = 0.8$
- We have $g(X, Y) = X(Y - 1)$ then $E(g(X, Y)) = \sum_{x,y} g(x, y)P(X = x, Y = y) = -1\ 0.3 = -0.3$

Oppgave 2 Probability in continuous distributions

Using the integration by part rule, we have that:

$$\int 4xe^{-2x}dx = -\frac{1}{2}xe^{-2x}\ 4x - \int(-\frac{1}{2}e^{-2x}\ 4)\ dx = -e^{-2x}(2x + 1)$$

then

- $P(X \leq 2) = -e^{-4}(5) + e^{0} = 0.908$
- $P(X^2 - 4X < -3) = P(1 < X < 3) = 0.389$
- We need to compute

$$E[\sqrt{x}] = \int_0^\infty \sqrt{x}4xe^{-2x}dx = 4\int_0^\infty x^{3/2}e^{-2x}dx$$

  The formula inside the integral is the core of a gamma distribution with parameters $\alpha = \frac{5}{2}$ and $\beta = \frac{1}{2}$. We know therefore that

$$\int_0^\infty \frac{1}{\frac{1}{2}^{5/2}\Gamma(\frac{5}{2})} x^{3/2}e^{-2x}dx = 1$$

  and we can use this to conclude that

$$E[\sqrt{x}] = 4 * \frac{1}{2}^{5/2}\Gamma(\frac{5}{2}) = 0.94$$

Oppgave 3 Virus

3a)

- Let 'Pos' indicate the event that the test outcome is positive, and 'Neg' the event that the test outcome is negative. Moreover let S indicate the event that a random person is infected and $\bar{S}$ the event that a random person is not infected. We have that:

$$P(S) = p$$
$$P(\text{Neg}|S) = 0.02$$

  The outcome of each person in the test is independent from all others, moreover we have that the probability for each person to test positively is

$$P(\text{Pos}) = P(\text{Pos}|S)p(S) + P(\text{Pos}|\bar{S})P(\bar{S}) = 0.98p + 0$$

  The stochastic variable X is then binomially distributed with parameters N and $0.98p$

- We have that

$$E(\frac{X}{0.98N}) = \frac{1}{0.98N}E(X) = \frac{1}{0.98N}N0.98p = p$$

  so $\hat{p}$ is unbiased. Moreover

$$\text{Var}(\frac{X}{0.98N}) = \frac{1}{0.98^2N^2}\text{Var}(X) = \frac{1}{0.98^2N^2}N0.98p(1-0.98p) = \frac{p(1-0.98p)}{0.98N}$$

  so the standard deviation of $\hat{p}$ is $\sqrt{\frac{p(1-0.98p)}{0.98N}}$

3b) Since N is large we can assume that $\hat{p}$ is approximately normal distributed. To build a confidence interval we can therefore start from the statistics

$$\frac{\hat{p}-p}{\sqrt{\frac{p(1-0.98p)}{0.98N}}} \approx N(0,1)$$

We can approximate the sd of $\hat{p}$ as $\sqrt{\frac{\hat{p}(1-0.98\hat{p})}{0.98N}}$ and a 95% confidence interval is then

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-0.98\hat{p})}{0.98N}}$$

Using the observed data we get $\hat{p} = 0.0561$ and the CI is $[0.0417, 0.0705]$

Oppgave 4 Levetiden

4a)

- $P(X > 15) = 1 - F(15) = 1 - 1 + \exp(-\sqrt(15)/2.5) = 0.212$
- $P(X > 20|X > 15) = \frac{P(X>20, X>15)}{P(X>15)} = \frac{P(X>20)}{P(X>15)} = \frac{\exp(-\sqrt(15)/2.5)}{0.212} = 0.787$
-

$$
\begin{aligned}
E(x) = &\int_0^\infty x \frac{1}{2\theta} \frac{1}{\sqrt{x}} \exp\left\{-\frac{\sqrt{x}}{\theta}\right\} dx \\
=&\frac{1}{2\theta} \int_0^\infty \sqrt{x} \exp\left\{-\frac{\sqrt{x}}{\theta}\right\} dx \\
=&\frac{1}{2\theta} [2\theta^{1+2}\Gamma(1+2)] \\
=&\theta^2\Gamma(3) = 2\theta^2
\end{aligned}
$$

4b) The likelihood is:

$$
\begin{aligned}
L(\theta) &= \prod \frac{1}{2\theta} \frac{1}{\sqrt{x_i}} \exp(-\frac{\sqrt{x_i}}{\theta}) \\
&= \frac{1}{2^n\theta^n} \frac{1}{\prod \sqrt{x_i}} \exp(-\frac{\sum \sqrt{x_i}}{\theta})
\end{aligned}
$$

and the log-likelihood

$$
l(\theta) = -n\log(2) - n\log(\theta) - \log \prod \sqrt{x_i} - \frac{\sum \sqrt{x_i}}{\theta}
$$

Deriving wrt θ we get

$$
\frac{dl(\theta)}{d\theta} = -\frac{n}{\theta} + \frac{\sum \sqrt{x_i}}{\theta^2}
$$

which gives taht the MLE estimator is $\hat{\theta} = \frac{\sum \sqrt{x_i}}{n}$

We now have to show that this is an unbiased estimator.

We have that

$$
E(\hat{\theta}) = \frac{1}{n} \sum E(\sqrt{X_i})
$$

and, using the given formula with $\alpha = 0$ we have that

$$E(\sqrt{X}) = \int_0^\infty \sqrt{x}\frac{1}{2\theta}\frac{1}{\sqrt{x}}\exp\left\{-\frac{\sqrt{x}}{\theta}\right\}dx$$
$$=\frac{1}{2\theta}\int_0^\infty \exp\left\{-\frac{\sqrt{x}}{\theta}\right\}dx = \frac{1}{2\theta}\theta^2\Gamma(2) = \theta$$

Then

$$E(\hat{\theta}) = \frac{1}{n}\sum E(\sqrt{X_i}) = \theta$$

4c) We have that

$$P(Z_i < z) = P(\frac{2}{\theta}\sqrt{X_i} < z) = P(X_i < (\frac{z\theta}{2})^2) = 1-\exp(-\frac{z\theta}{2\theta}) = 1-\exp(-\frac{z}{2})$$

so the pdf of Z_i is

$$\frac{1}{2}e^{-\frac{z}{2}}$$

which is the pdf of a χ^2 distributed SV with 2 degrees of freedom.

We can write $\frac{2n\widehat{\theta}}{\theta}$ as $\sum \frac{2X_i}{\theta}$ which is the sum of n independent SV each χ^2 distributed SV with 2 degrees of freedom, therefore $\frac{2n\widehat{\theta}}{\theta}$ is χ^2 distributed SV with $2n$ degrees of freedom

4d) We want to test

$$H_0 : \theta = 2.6,$$
$$H_1 : \theta \geq 2.6$$

A decision rule is given by $Z > k$ where, under H_0, $Z = \frac{2n\hat{\theta}}{2.6}$ is χ^2 distributed with $2n = 50$ degrees of freedom.

The observed value for the test statistics is $\frac{2\ 70.493}{2.6} = 54.22$ and the p value is:

$$P(Z > 54.22|\theta = 2.6) \approx 1 - P(Z < 54.22|\theta = 2.6) = 1 - 0.6927 = 0.3073$$

and we would not reject the null hypotheses under any plausible significance level.

4e) With $\alpha = 0.05$ we reject the null hypothesis if $V = \frac{2n\hat{\theta}}{2.6} > \chi^2_{2n,0.05}$. We want to find n such that the probability to reject H_0 when $\theta = 3.4$ is at least 50%.

Then

$$\begin{aligned} P(\text{Reject } H_0|\theta = 3.4) &= P(\frac{2n\hat{\theta}}{2.6} > \chi^2_{2n,0.05}|\theta = 3.4) \\ &= P(\frac{2n\hat{\theta}}{3.4} > \frac{2.6}{3.4}\chi^2_{2n,0.05}|\theta = 3.4) \\ &= P(\chi^2_{2n} > 0.765\chi^2_{2n,0.05}|\theta = 3.4) \geq 0.5 \end{aligned}$$

We then need that $0.765\chi^2_{2n,0.05} \leq \chi^2_{2n,0.5}$ or equivalently $\frac{\chi^2_{2n,0.5}}{\chi^2_{2n,0.05}} \geq 0.765$. From the table we have that:

ν	$\chi^2_{\nu,0.5}$	$\chi^2_{\nu,0.05}$	$\frac{\chi^2_{\nu,0.5}}{\chi^2_{\nu,0.05}}$
60	59.33467	79.08194	0.7502935
62	61.33462	81.38102	0.7536724
64	63.33458	83.67526	0.7569093
66	65.33454	85.96491	0.7600141
68	67.33451	88.25016	0.7629958
70	69.33447	90.53123	0.7658625
72	71.33444	92.80827	0.7686216
74	73.33441	95.08147	0.7712798
76	75.33438	97.35097	0.7738432
78	77.33436	99.61693	0.7763174
80	79.33433	101.87947	0.7787077

So we get that we need to observe at least $\frac{70}{2} = 35$ components.