Kunnskap for en bedre verden

Institutt for matematiske fag

Eksamensoppgave i **TMA4240 Statistikk**

Faglig kontakt under eksamen:

Tlf:

Eksamensdato: 30.11.2019

Eksamenstid (fra–til): 09:00–13:00

Hjelpemiddelkode/Tillatte hjelpemidler:

Annen informasjon:

Målform/språk: bokmål

Antall sider: 8

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

Merk! Studenter finner sensur i Studentweb. Har du spørsmål om din sensur må du kontakte instituttet ditt. Eksamenskontoret vil ikke kunne svare på slike spørsmål.

Oppgave 1 Correct answers:

1a) B

1b) B

1c) D

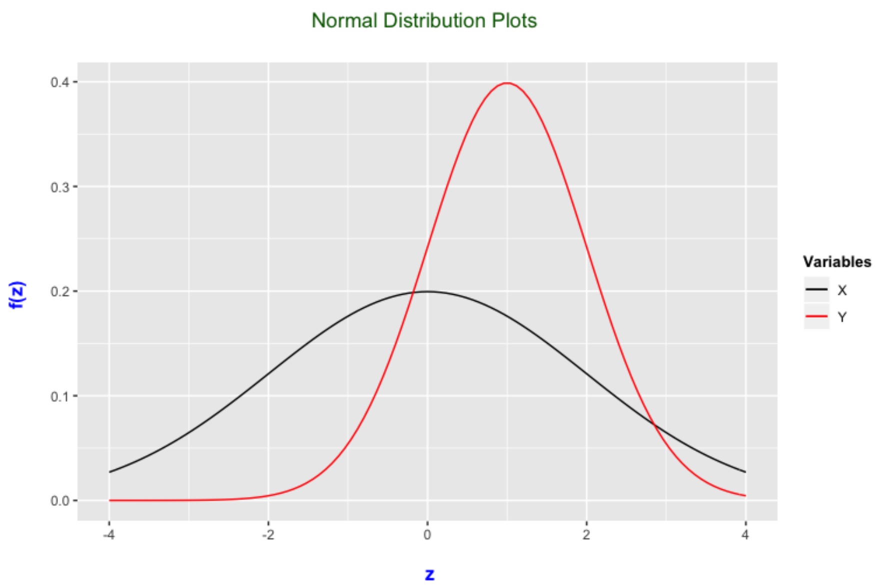
1d) B

1e) A

Oppgave 2

We have that $X \sim N(0,2)$ and $Y \sim (1,1)$ and X and Y are mutually independent.

Drawing of the pdfs:



Compute the probabilities:

$$P(X \leq 1) = P(\frac{X-0}{2} \leq \frac{1-0}{2}) = P(Z \leq 0.5) \approx 0.69$$

$$P(Y \geq -1) = P(\frac{Y-1}{1} \geq \frac{-1-1}{1}) = 1 - P(Z < -2) \approx 1 - 0.022 = 0.978.$$

In addition we have that $(X - Y) \sim N(-1, \sqrt{5})$ so

$$P(X - Y \leq 0) = P(\frac{X - Y + 1}{\sqrt{5}} \leq \frac{0 + 1}{\sqrt{5}}) = P(Z \leq \frac{1}{\sqrt{5}}) \approx 0.67$$

Oppgave 3

Define the events:

M: The mann is carrier of sikle cell anemia

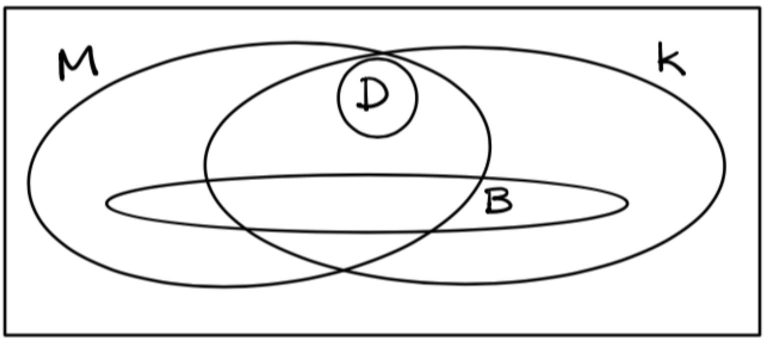
K: The woman is carrier if sikle cell anemia

D: The first born has sikle cell anemia

B: The first born is carrier of sikle cell anemia

Let indicate with $\bar{A}$ the complement of event A

a) Draw the events in a Venn diagram



We have that:

$$
\begin{aligned}
P(M) &= 0.08 \\
P(K) &= 0.08 \\
P(D|M,K) &= 0.25 \\
P(B|M,K) &= 0.5 \\
P(D|\bar{M},K) &= 0 \\
P(D|M,\bar{K}) &= 0 \\
P(D|\bar{M},\bar{K}) &= 0 \\
P(B|\bar{M},K) &= 0.5 \\
P(B|M,\bar{K}) &= 0.5 \\
P(B|\bar{M},\bar{K}) &= 0
\end{aligned}
$$

A child can get sikle cell anemia only if both mother and father are carrier, so the probability that the first born has sikle cell anemia is:

$$P(D) = P(D|K,M)P(K)P(M) = 0.08 \cdot 0.08 \cdot 0.25 = 0.0016$$

To compute the probability that the first born is carrier of sikle cell anemia, consider that the event space S can be decomposed as following:

$$S = (K \cap M) \cup (\bar{K} \cap M) \cup (K \cap \bar{M}) \cup (\bar{K} \cap \bar{M})$$

Using the law of total probability we can write that the probability of the event B as:

$$
\begin{aligned}
P(B) = \; & P(B|K,M)P(K)P(M) + \\
& P(B|\bar{K},M)P(\bar{K})P(M) + \\
& P(B|K,\bar{M})P(K)P(\bar{M}]) + \\
& P(B|\bar{K},\bar{M})P(\bar{K})P(\bar{M}]) = \\
& 0.5 \cdot 0.08 \cdot 0.08 + \\
& 0.5 \cdot (1 - 0.08) \cdot 0.08 + \\
& 0.5 \cdot 0.08 \cdot (1 - 0.08) + \\
& 0 \cdot (1 - 0.08)^2 = 0.0768
\end{aligned}
$$

b) Define the events:

D_1: the first born has sikle cell anemia

D_2: the second born has sikle cell anemia

B_1: the first born is carrier of sikle cell anemia

B_2: the second born is carrier of sikle cell anemia

We want to compute the probability that the second born has sikle cell anemia given that the first born does not have it:

$$P(D_2|\bar{D_1}) = \frac{P(D_2, \bar{D_1})}{P(\bar{D_1})}$$

We can derive the denominator from the result in a)

$$P(\bar{D_1}) = 1 - P(D_1) = 1 - 0.0016 = 0.9984$$

To compute the numerator we note that, both given M and K the events $\bar{D_1}$ and D_2 are independent and that the event D_2 can happen only if both mother and father are carrier of sikle cell anemia:

$$\begin{aligned} PP(D_2, \bar{D_1}) &= P(D_2, \bar{D_1}|K, M)P(K)P(M) \\ &= P(D_2|K, M)P(\bar{D_1}|K, M)P(K)P(M) \\ &= 0.25 \cdot 0.75 \cdot 0.08^2 = 0.0012 \end{aligned}$$

Putting things together we get:

$$P(D_2|\bar{D_1}) = \frac{0.0012}{0.9984} = 0.0012012$$

The probability that the second born is carrier given that first one does not have sikle cell anemia:

$$P(B_2|\bar{D_1}) = \frac{P(B_2, \bar{D_1})}{P(\bar{D_1})}$$

The numerator can be computed again conditioning on weather the father and mother are carrier or not:

$$\begin{aligned} P(B_2, \bar{D_1}) &= P(B_2, \bar{D_1}|K, M)P(K)P(M) + \\ &\quad P(B_2, \bar{D_1}|\bar{K}, M)P(\bar{K})P(M) + \\ &\quad P(B_2, \bar{D_1}|K, \bar{M})P(K)P(\bar{M}) + 0 \\ &= 0.5 \cdot 0.75 \cdot 0.08^2 + \\ &\quad 0.5 \cdot 1 \cdot 0.08 \cdot (1 - 0.08) + \\ &\quad 0.5 \cdot 1 \cdot 0.08 \cdot (1 - 0.08) = 0.076 \end{aligned}$$

Putting things together we get that:

$$P(B_2|\bar{D_1}) = \frac{0.0768}{0.9984} \approx 0.07$$

Oppgave 4

a) Let $n = 12$ and $p = 0.2$. To find the first of the three probabilities we can use the table about the binomial distribution in the booklet Tabeller og formler i statistikk"

We have that:

$$P(X \leq 3) = 0.795$$

To find the second probability, we note that X and Y are independent, therefore:

$$P(Y \geq 4|X \leq 3) = PY(\geq 4) = 1 - P(Y \leq 3) = 1 - 0.225 = 0.775$$

where $P(Y \leq 3) = 0.225$ is taken from the table about binomial distribution with $p = 0.4$.

Finally we have that:

$$\begin{aligned}
P(X + Y \leq 1) &= P(X = 0 \cap Y = 0) + P(X = 1 \cap Y = 0) + P(X = 0 \cap Y = 1) \\
&= P(X = 0)P(Y = 0) + P(X = 1)P(Y = 0) + P(X = 0)P(Y = 1) \\
&= P(X \leq 0)P(Y \leq 0) + [P(X \leq 1) - P(X \leq 0)]P(Y \leq 0) + \\
&\quad P(X \leq 0)[P(Y \leq 1) - P(Y \leq 0)] \\
&= 0.069\ 0.002 + (0.275 - 0.069)\ 0.002 + 0.069\ (0.02 - 0.002) \\
&= 0.001792
\end{aligned}$$

Where the probabilities are again taken from the relative binomial table.

b) To compare the estimators we first have to check if they are biased or not, that is we have to compute the $E(\hat{p})$, $E(\hat{\hat{p}})$ and $E(p^\star)$:

$$\begin{aligned}
E(\hat{p}) &= E(\frac{X+Y}{2n}) = \frac{1}{2n}(E(X) + E(Y)) \\
&= \frac{1}{2n}(np + 2np) = \frac{3np}{2n} = \frac{3p}{2} \neq p
\end{aligned}$$

$$\begin{aligned}
E(\hat{\hat{p}}) &= E(\frac{X+Y}{3n}) = \frac{1}{3n}(E(X) + E(Y)) \\
&= \frac{1}{3n}(np + 2np) = \frac{3np}{3n} = p
\end{aligned}$$

$$
\begin{aligned}
E(p^\star) &= E(\frac{X}{2n} + \frac{Y}{4n}) = \frac{1}{2n}E(X) + \frac{1}{4n}E(Y) \\
&= \frac{np}{2n} + \frac{2np}{4n} = p
\end{aligned}
$$

That is $\hat{\hat{p}}$ and $E(p^\star)$ are unbiased while $\hat{p}$ is biased.

We then check the variance of the unbiased estimators, we look for the unbiased estimator with minimum variance.

$$
\begin{aligned}
\mathrm{Var}(\hat{\hat{p}}) &= \mathrm{Var}(\frac{X+Y}{3n}) = \frac{1}{9n^2}(\mathrm{Var}(X) + \mathrm{Var}(Y)) \\
&= \frac{1}{9n^2}(np(1-p) + 2np(1-2p)) = \frac{3-5p}{9n}p
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(p^\star) &= \mathrm{Var}(\frac{X}{2n} + \frac{Y}{4n}) = \frac{1}{4n^2}\mathrm{Var}(X) + \frac{1}{16n^2}\mathrm{Var}(Y) \\
&= \frac{np(1-p)}{4n^2} + \frac{n2p(1-2p)}{16n^2} = \\
&= \frac{3-4p}{8n}p
\end{aligned}
$$

We check the difference between the two variances:

$$
\begin{aligned}
\mathrm{Var}(\hat{\hat{p}}) - \mathrm{Var}(p^\star) &= \frac{3-5p}{9n}p - \frac{3-4p}{8n}p \\
&= \frac{-4p-3}{73n}p < 0 \text{ for } p \in [0, 1/2]
\end{aligned}
$$

We would then choose $\hat{\hat{p}}$ as it is unbiased and has minimum variance.

c) The likelihood function for p is given by:

$$
\begin{aligned}
L(p; x, y) &= P(X = x; p)p(Y = y; p) \\
&= \binom{n}{x} p^x (1-p)^{n-x} \binom{n}{y} (2p)^y (1-2p)^{n-y}
\end{aligned}
$$

and the log-likelihood is

$$
\begin{aligned}
l(p; x, y) &= \log \binom{n}{x} + \log \binom{n}{y} \\
&+ x \log p + (n-x) \log(1-p) \\
&+ y \log(2p) + (n-y) \log(1-2p)
\end{aligned}
$$

To find the MLE estimator for p we need to derive $l(p)$ wrt p and set it to 0

$$l'(p; x, y) = \frac{x}{p} - \frac{n-x}{1-p} + \frac{y}{p} - 2\frac{n-y}{1-2p} = 0$$

Setting in the data $n = 25$, $x = 3$, $y = 8$ we get:

$$\begin{aligned} \frac{11}{p} - \frac{22}{1-p} - \frac{34}{1-2p} &= 0 \Rightarrow \\ 100\,p^2 - 89\,p + 11 &= 0 \end{aligned}$$

The last equation has two possible solutions $p_1 \approx 0.15$ and $p_2 \approx 0.74$. Since $p \in [0, 0.5]$ our maximum likelihood estimate for p is 0.15.

d) The test to perform is

$$H_0 : p = 0.2 \text{ against } H_1 : p > 0.2$$

We consider the estimator $\hat{\hat{p}} = \frac{X+Y}{3n}$. We assume that X and Y are approximately normally distributed. Since $\hat{\hat{p}}$ is a linear combination of X and Y then $\hat{\hat{p}}$ is also normally distributed with mean $E(\hat{\hat{p}}) = p$ and variance $\text{Var}\hat{\hat{p}} = p\frac{3-5p}{9n}$ as computed in b).

To perform the test we define the test statistics:

$$Z = \frac{\hat{\hat{p}} - p_0}{\sqrt{p_0\frac{3-5p_0}{9n}}}$$

which is $N(0,1)$ under H_0.

With a significance of $\alpha = 0.05$ we reject the null hypotheses if $Z > z_{0.05} = 1.645$.

e) We want to find n such that:

$$P(\text{Reject } H_0 | p = 0.25) \geq 0.9$$

we have then that:

$$P(\frac{\hat{\hat{p}} - p_0}{\sqrt{p_0\frac{3-5p_0}{9n}}} > z_\alpha | p = 0.25) \geq 0.9$$

$$P(\hat{\hat{p}} > p_0 + z_\alpha\sqrt{p_0\frac{3-5p_0}{9n}} | p = 0.25) \geq 0.9$$

$$P(\frac{\hat{\hat{p}} - p}{\sqrt{p\frac{3-5p}{9n}}} > \frac{p_0 + z_\alpha\sqrt{p_0\frac{3-5p_0}{9n}} - p}{\sqrt{p\frac{3-5p}{9n}}} | p = 0.25) \geq 0.9$$

where $Z = \frac{\hat{\hat{p}}-p}{\sqrt{p\frac{3-5p}{9n}}} \sim N(0,1)$. We have then:

$$P(Z > \frac{0.2 + 1.645\sqrt{0.2\frac{3-5\ 0.2}{9n}} - 0.25}{\sqrt{0.25\frac{3-5\ 0.25}{9n}}} | p = 0.25) \geq 0.9$$

which is true if

$$\frac{0.2 + 1.645\sqrt{0.2\frac{3-5\ 0.2}{9n}} - 0.25}{\sqrt{0.25\frac{3-5\ 0.25}{9n}}} \leq -z_{0.1} = -1.282.$$

We can now solve the disequality wrt n to find the answer:

$$\begin{aligned}
0.2 + 1.645\sqrt{\frac{0.4}{9n}} - 0.25 &\leq -1.282\sqrt{\frac{0.4375}{9n}} \\
\frac{1}{\sqrt{n}}\left(1.645\sqrt{\frac{0.4}{9}} + 1.282\sqrt{\frac{0.4375}{9}}\right) &\leq 0.25 - 0.2 \\
0.6294509 &\leq 0.05\sqrt{n} \\
n &\geq (\frac{0.6294509}{0.05})^2 = 158.4834
\end{aligned}$$

Therefore our answer is that n has to be larger than 159