

Institutt for matematiske fag

Eksamensoppgave i TMA4240 Statistikk

Faglig kontakt under eksamen: Håkon Tjelmeland

Tlf: 48 22 18 96

Eksamensdato: 11. august 2022 – løsningskisse

Eksamenstid (fra–til): 09.00–13.00

Hjelpemiddelkode/Tillatte hjelpemidler: C

Tabeller og formler i statistikk, Akademika,

Bestemt, enkel kalkulator,

Gult stemplet A5-ark med egne håndskrevne notater.

Annen informasjon:

Alle svar skal begrunnes og besvarelsen skal inneholde naturlig mellomregning.

Målform/språk: bokmål

Antall sider: 13

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

For å løse denne oppgaven anbefales det å tegne hendelsene A , B og C inn i en venndiagram. Da vil man få en oversikt over situasjonen som gjør det enklere å se hvordan sannsynlighetene det spørres etter kan regnes ut.

For å finne den første sannsynligheten kan vi benytte den generelle regneregelen $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Denne gir

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.500 + 0.400 - 0.300 = 0.600. \end{aligned}$$

Så svaret blir $P(A \cup B) = 0.600$.

Den neste sannsynligheten finner vi ved å benytte definisjonen av betinget sannsynlighet,

$$\begin{aligned} P(B|A) &= \frac{P(B \cap A)}{P(A)} \\ &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{0.300}{0.500} = 0.600. \end{aligned}$$

Så svaret blir $P(B|A) = 0.600$.

Den siste sannsynligheten er noe vanskeligere å bestemme. Her kan et venndiagram være svært nyttig. Det er ulike måter man kan regne ut svaret på. Her velger vi å starte med $P(C)$ og benytte setningen om total sannsynlighet med hensyn på partisjonen $\{A \cup B, (A \cup B)'\}$ av S . Dette gir

$$P(C) = P(C|A \cup B)P(A \cup B) + P(C|(A \cup B)')P((A \cup B)').$$

Ved å løse denne ligningen ved hensyn på $P(C|A \cup B)$ og bruke at $P((A \cup B)') = 1 - P(A \cup B)$ får vi

$$\begin{aligned} P(C|A \cup B) &= \frac{P(C) - P(C|(A \cup B)')P((A \cup B)')}{P(A \cup B)} \\ &= \frac{P(C) - P(C|(A \cup B)')(1 - P(A \cup B))}{P(A \cup B)} \\ &= \frac{0.300 - 0.200 \cdot (1 - 0.600)}{0.600} \\ &= 0.367. \end{aligned}$$

Så den siste sannsynligheten blir altså $P(C|A \cup B) = 0.367$.

Oppgave 2

Da $X \sim N(1, 4^2)$ får vi ved å standardisere og bruke tabell over sannsynligheter i en standard normalfordeling at

$$\begin{aligned} P(X > 4) &= 1 - P(X \leq 4) \\ &= 1 - \Phi\left(\frac{4 - 1}{4}\right) \\ &= 1 - \Phi(0.75) \\ &= 1 - 0.7734 \\ &= 0.2266. \end{aligned}$$

Vi har altså at $P(X \geq 4) = 0.2266$.

For å finne den neste sannsynligheten må vi først finne fordelingen til $X + 2Y$. Da $X + 2Y$ er en lineærkombinasjon av uavhengige og normalfordelte variabler har vi at også $X + 2Y$ er normalfordelt. Dessuten, ved å benytte regneregler for forventingsverdi og varians får vi at

$$\begin{aligned} E[X + 2Y] &= E[X] + E[2Y] \\ &= E[X] + 2E[Y] \\ &= 1 + 2 \cdot (-2) \\ &= -3, \\ \text{Var}[X + 2Y] &= \text{Var}[X] + \text{Var}[2Y] \\ &= \text{Var}[X] + 2^2 \text{Var}[Y] \\ &= 4^2 + 2^2 \cdot 2^2 \\ &= 32. \end{aligned}$$

Vi har altså at $X + 2Y \sim N(-3, 32)$ og ved å standardisere og bruke tabell som vi gjorde for å finne den første sannsynligheten får vi

$$\begin{aligned} P(X + 2Y < 0) &= P(X + 2Y \leq 0) \\ &= \Phi\left(\frac{0 - (-3)}{\sqrt{32}}\right) \\ &= \Phi(0.53) \\ &= 0.7019. \end{aligned}$$

Så svaret blir altså at $P(X + 2Y < 0) = 0.7019$.

For å finne den siste sannsynligheten må vi først finne fordelingen for $Y - X$. Også $Y - X$ er en lineærkombinasjon av uavhengige normalfordelte variabler så $Y - X$

blir normalfordelt. Forventningsverdi og varians finner vi ved å bruke de samme regnereglene som over,

$$\begin{aligned}
 E[Y - X] &= E[Y + (-1)X] \\
 &= E[Y] + E[(-1)X] \\
 &= E[Y] + (-1)E[X] \\
 &= -2 + (-1) \cdot 1 \\
 &= -3, \\
 \text{Var}[Y - X] &= \text{Var}[Y + (-1)X] \\
 &= \text{Var}[Y] + \text{Var}[(-1)X] \\
 &= \text{Var}[Y] + (-1)^2 \text{Var}[X] \\
 &= 2^2 + 4^2 \\
 &= 20.
 \end{aligned}$$

Vi har altså at $Y - X \sim N(-3, 20)$ og ved å bruke definisjonen av betinget sannsynlighet og deretter standardisere får vi

$$\begin{aligned}
 P(Y - X \leq 4 | Y - X > -3) &= \frac{P(Y - X \leq 4 \cap Y - X > -3)}{P(Y - X > -3)} \\
 &= \frac{P(-3 < Y - X \leq 4)}{P(Y - X > -3)} \\
 &= \frac{P(Y - X \leq 4) - P(Y - X \leq -3)}{1 - P(Y - X \leq -3)} \\
 &= \frac{\Phi\left(\frac{4 - (-3)}{\sqrt{20}}\right) - \Phi\left(\frac{-3 - (-3)}{\sqrt{20}}\right)}{1 - \Phi\left(\frac{-3 - (-3)}{\sqrt{20}}\right)} \\
 &= \frac{\Phi(1.57) - \Phi(0)}{1 - \Phi(0)} \\
 &= \frac{0.9418 - 0.5}{1 - 0.5} \\
 &= 0.8836.
 \end{aligned}$$

Svaret er altså $P(Y - X \leq 4 | Y - X > -3) = 0.8836$.

Oppgave 3

Ved å benytte regneregler for varians for lineære funksjoner av uavhengige stokas-

tiske variabler får vi

$$\begin{aligned}
 \text{Var}[U] &= \text{Var} \left[\sum_{i=1}^n a_i (X_i - \mu) \right] \\
 &= \sum_{i=1}^n \text{Var}[a_i (X_i - \mu)] \\
 &= \sum_{i=1}^n a_i^2 \text{Var}[X_i - \mu] \\
 &= \sum_{i=1}^n a_i^2 \text{Var}[X_i] \\
 &= \sum_{i=1}^n a_i^2 \sigma^2 \\
 &= \sigma^2 \sum_{i=1}^n a_i^2
 \end{aligned}$$

og

$$\begin{aligned}
 \text{Var}[V] &= \text{Var} \left[\frac{\sum_{i=1}^n a_i X_i}{\sum_{i=1}^n a_i} \right] \\
 &= \left(\frac{1}{\sum_{i=1}^n a_i} \right)^2 \text{Var} \left[\sum_{i=1}^n a_i X_i \right] \\
 &= \frac{1}{(\sum_{i=1}^n a_i)^2} \sum_{i=1}^n \text{Var}[a_i X_i] \\
 &= \frac{\sum_{i=1}^n a_i^2 \text{Var}[X_i]}{(\sum_{i=1}^n a_i)^2} \\
 &= \frac{\sum_{i=1}^n a_i^2 \sigma^2}{(\sum_{i=1}^n a_i)^2} \\
 &= \sigma^2 \cdot \frac{\sum_{i=1}^n a_i^2}{(\sum_{i=1}^n a_i)^2}.
 \end{aligned}$$

De riktige uttrykkene for variansene er altså

$$\underline{\underline{\text{Var}[U] = \sigma^2 \sum_{i=1}^n a_i^2}} \quad \text{og} \quad \underline{\underline{\text{Var}[V] = \sigma^2 \cdot \frac{\sum_{i=1}^n a_i^2}{(\sum_{i=1}^n a_i)^2}}}.$$

Oppgave 4

For å regne ut $f_Y(y)$ kan man enten starte med å finne $f_X(x)$ ved å derivere $F_X(x)$ og deretter bruke transformasjonsformelen til å bestemme $f_Y(y)$ fra $f_X(x)$, eller man kan først finne $F_Y(y)$ fra $F_X(x)$ og deretter derivere $F_Y(y)$ for å finne $f_Y(y)$. Her velger vi å følge den siste av disse to prosedyrene. Ved å benytte at $X \geq 0$ får vi at for $y \geq 0$ har vi

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X^2 \leq y) \\ &= P(X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) \\ &= 1 - e^{-(\sqrt{y})^2} \\ &= 1 - e^{-y}. \end{aligned}$$

Vi finner så $f_Y(y)$ ved å derivere $F_Y(y)$,

$$\begin{aligned} f_Y(y) &= F_Y'(y) \\ &= -(e^{-y}) \cdot (-1) \\ &= e^{-y}. \end{aligned}$$

For $y \geq 0$ har vi dermed at $f_Y(y) = e^{-y}$.

Oppgave 5

- a) Når vi kaster terninger er det rimelig å benytte en uniform sannsynlighetsmodell så alle sannsynlighetene $P(X = x, Y = Y)$ blir gitt som antall gunstige utfall delt på antall mulige utfall. Antall mulige utfall er lik for alle verdier av x og y . Vi har 6 mulige utfall for den ene terningen og 6 mulige utfall på den andre terningen. Antall mulige utfall blir dermed $m = 6 \cdot 6 = 36$.

Når $x = 0$ og $y = 0$ må man på begge terningene få 4 eller 5 øyne. Man har altså to gunstige resultater for hver terning slik at $g = 2 \cdot 2 = 4$ og man får

$$P(X = 0, Y = 0) = \frac{4}{36} = \frac{1}{9}.$$

Når $x = 1$ og $y = 0$ må man få sekser på en terning og 4 eller 5 på den andre. Sekser på en terning kan man få på 2 måter, og for hver av disse kan man få 4 eller 5 på den andre terningen på 2 måter. Dermed blir $g = 2 \cdot 2$ og

$$P(X = 1, Y = 0) = \frac{4}{36} = \frac{1}{9}.$$

Når $x = 2$ og $y = 0$ må man ha seksere på begge terningene. Dette kan bare skje på en måte så $g = 1$ og

$$P(X = 2, Y = 0) = \frac{1}{36}.$$

Når $x = 0$ og $y = 1$ må man få 1, 2 eller 3 på en terning og 4 eller 5 på den andre terningen. Det er 2 valg man kan gjøre for å velge hvilke terning som skal få 1, 2 eller 3. For hver av disse 2 valgene finnes det 3 måter den valgte terning skal få 1, 2 eller 3, og for hver av disse igjen finnes de 2 måter den andre terningen kan få 4 eller 5. Dermed blir $g = 2 \cdot 3 \cdot 2 = 12$ og

$$P(X = 0, Y = 1) = \frac{12}{36} = \frac{1}{3}.$$

Når $x = 1$ og $y = 1$ må man ha sekser på en terning og 1, 2 eller 3 på den andre. Sekser på en terning kan man få på 2 måter og for hver av disse kan man få 1, 2 eller 3 på den andre terningen på 3 måter. Dermed blir $g = 2 \cdot 3 = 6$ og

$$P(X = 1, Y = 1) = \frac{6}{36} = \frac{1}{6}.$$

Når $x = 0$ og $y = 2$ må man ha 1, 2 eller 3 øyne på begge terningene. Det er 3 måter å få dette på den ene terningen, og hver av disse er det også 3 måter å få det på den andre terningen. Dermed blir $g = 3 \cdot 3 = 9$ og

$$P(X = 0, Y = 2) = \frac{9}{36} = \frac{1}{4}.$$

Siden vi kaster bare to terninger må vi åpenbart ha at $X + Y \leq 2$. For de resterende tre kombinasjoner av $x \in \{0, 1, 2\}$ og $y \in \{0, 1, 2\}$ er dermed $g = 0$, og følgelig blir de tilhørende sannsynlighetene også lik 0.

b) Ved å benytte punktsannsynlighetene oppgitt i oppgave 5(a) får vi

$$\begin{aligned} P(X + Y = 2) &= P(X = 0, Y = 2) + P(X = 1, Y = 1) + P(X = 2, Y = 0) \\ &= \frac{1}{4} + \frac{1}{6} + \frac{1}{36} \\ &= \frac{9 + 6 + 1}{36} \\ &= \frac{16}{36} \\ &= \frac{4}{9}. \end{aligned}$$

Vi har altså $\underline{\underline{P(X + Y = 2) = \frac{4}{9} \approx 0.4444.}}$

Marginal punktsannsynlighet for X finner vi ved å summere hver kolonne i tabellen i oppgave 5(a). Dette gir

$$\begin{aligned} P(X = 0) &= \frac{1}{9} + \frac{1}{3} + \frac{1}{4} = \frac{4 + 12 + 9}{36} = \frac{25}{36}, \\ P(X = 1) &= \frac{1}{9} + \frac{1}{6} + 0 = \frac{4 + 6}{36} = \frac{10}{36}, \\ P(X = 2) &= \frac{1}{36} + 0 + 0 = \frac{1}{36}. \end{aligned}$$

Marginal punktsannsynlighet for Y finner vi tilsvarende ved å summere hver linje i den samme tabellen,

$$\begin{aligned} P(Y = 0) &= \frac{1}{9} + \frac{1}{9} + \frac{1}{36} = \frac{4 + 4 + 1}{36} = \frac{9}{36} = \frac{1}{4}, \\ P(Y = 1) &= \frac{1}{3} + \frac{1}{6} + 0 = \frac{12 + 6}{36} = \frac{18}{36} = \frac{1}{2}, \\ P(Y = 2) &= \frac{1}{4} + 0 + 0 = \frac{1}{4} \end{aligned}$$

For at X og Y skal være uavhengige må man ha at $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$ for alle verdier av x og y . Vi har at $P(X = 0, Y = 0) = \frac{1}{9} \approx 0.1111$ og $P(X = 0) \cdot P(Y = 0) = \frac{25}{36} \cdot \frac{1}{4} = \frac{25}{144} \approx 0.1736 \neq 0.1111$. De stokastiske variablene X og Y er følgelig ikke uavhengige.

Oppgave 6

a) Siden X_i -ene er uavhengige og $X_i \sim N(0, \sigma^2)$ blir rimelighetsfunksjonen

$$\begin{aligned} L(\sigma^2) &= \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\sigma^2}} e^{-\frac{1}{2\sigma^2} x_i^2} \right] \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 \right\}. \end{aligned}$$

Log-rimelighetsfunksjon blir dermed

$$\ell(\sigma^2) = \ln L(\sigma^2) = -\frac{n}{2}(\ln(2\pi) + \ln(\sigma^2)) - \frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2.$$

Fra denne finner vi sannsynlighetsmaksimeringsestimatoren for σ^2 ved å sette den deriverte lik null. Den deriverte av $\ell(\sigma^2)$ med hensyn på σ^2 blir

$$\begin{aligned}\ell'(\sigma^2) &= -\frac{n}{2} \cdot \frac{1}{\sigma^2} - \frac{1}{2} \cdot \left(-\frac{1}{(\sigma^2)^2} \right) \sum_{i=1}^n x_i^2 \\ &= -\frac{1}{2\sigma^2} \left(n - \frac{1}{\sigma^2} \sum_{i=1}^n x_i^2 \right).\end{aligned}$$

Ved å sette $\ell'(\sigma^2) = 0$ og løse med hensyn på σ^2 får vi

$$n - \frac{1}{\sigma^2} \sum_{i=1}^n x_i^2 = 0 \Rightarrow \sigma^2 = \frac{1}{n} \sum_{i=1}^n x_i^2.$$

Sannsynlighetsmaksimeringsestimatoren for σ^2 blir dermed

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

Bruker nå at $\text{Var}[X_i] = \text{E}[X_i^2] - \text{E}[X_i]^2$, løser denne med hensyn på $\text{E}[X_i^2]$ og setter inn $\text{E}[X_i] = 0$,

$$\begin{aligned}\text{E}[X_i^2] &= \text{Var}[X_i] - \text{E}[X_i]^2 \\ &= \sigma^2 + 0 \\ &= \sigma^2.\end{aligned}$$

Forventningsverdien til $\hat{\sigma}^2$ blir dermed

$$\begin{aligned}\text{E}[\hat{\sigma}^2] &= \text{E} \left[\frac{1}{n} \sum_{i=1}^n X_i^2 \right] \\ &= \frac{1}{n} \text{E} \left[\sum_{i=1}^n X_i^2 \right] \\ &= \frac{1}{n} \sum_{i=1}^n \text{E}[X_i^2] \\ &= \frac{1}{n} \sum_{i=1}^n \sigma^2 \\ &= \frac{1}{n} \cdot n\sigma^2 \\ &= \sigma^2.\end{aligned}$$

Følgelig har vi at $\hat{\sigma}^2$ er forventningsrett.

For variansen til $\widehat{\sigma^2}$ får vi at

$$\begin{aligned}\text{Var}[\widehat{\sigma^2}] &= \text{Var}\left[\frac{1}{n} \sum_{i=1}^n X_i^2\right] \\ &= \left(\frac{1}{n}\right)^2 \text{Var}\left[\sum_{i=1}^n X_i^2\right] \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i^2] \\ &= \frac{1}{n^2} \sum_{i=1}^n 3\sigma^4 \\ &= \frac{1}{n^2} \cdot n \cdot 3\sigma^4 \\ &= \frac{3\sigma^4}{n}.\end{aligned}$$

Vi har altså vist at $\underline{\underline{\text{Var}[\widehat{\sigma^2}] = \frac{3\sigma^4}{n}}}$.

b) Vi observerer først at

$$\begin{aligned}\frac{n\widehat{\sigma^2}}{\sigma^2} &= \frac{n}{\sigma^2} \cdot \frac{1}{n} \sum_{i=1}^n X_i^2 \\ &= \sum_{i=1}^n \frac{X_i^2}{\sigma^2} \\ &= \sum_{i=1}^n \left(\frac{X_i}{\sigma}\right)^2.\end{aligned}$$

Ved å standardisere $X_i \sim N(0, \sigma^2)$ får vi en standard normalfordelt variabel.

$$Z_i = \frac{X_i - \text{E}[X_i]}{\sqrt{\text{Var}[X_i]}} = \frac{X_i - 0}{\sqrt{\sigma^2}} = \frac{X_i}{\sigma} \sim N(0, 1).$$

Når $Z_i \sim N(0, 1)$ vet vi at $Z_i^2 \sim \chi_1^2$. Dessuten vet vi at en sum av uavhengige χ^2 -fordelte variabler er χ^2 -fordelt og at antall frihetsgrader til summen er lik summen av antall frihetsgrader. Siden X_i -ene er uavhengige blir også Z_i -ene uavhengige slik at vi her får

$$\sum_{i=1}^n Z_i^2 \sim \chi_n^2.$$

Ved å sette inn $Z_i = \frac{X_i}{\sigma}$ har vi dermed vist at

$$\frac{n\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i}{\sigma} \right)^2 \underset{\sim}{=} \chi_n^2.$$

Dette gir at

$$P\left(\chi_{1-\frac{\alpha}{2}}^2 \leq \frac{n\hat{\sigma}^2}{\sigma^2} \leq \chi_{\frac{\alpha}{2}}^2\right) = 1 - \alpha.$$

Løser så hver av ulikhetene inni dette sannsynlighetsuttrykket med hensyn på σ^2 . Dette gir

$$\chi_{1-\frac{\alpha}{2}}^2 \leq \frac{n\hat{\sigma}^2}{\sigma^2} \Leftrightarrow \sigma^2 \leq \frac{n\hat{\sigma}^2}{\chi_{1-\frac{\alpha}{2}}^2}$$

og

$$\frac{n\hat{\sigma}^2}{\sigma^2} \leq \chi_{\frac{\alpha}{2}}^2 \Leftrightarrow \frac{n\hat{\sigma}^2}{\chi_{\frac{\alpha}{2}}^2} \leq \sigma^2.$$

Ved å sette de to ulikhetene sammen igjen inne i sannsynlighetsuttrykket, med σ^2 i midten, får vi

$$P\left(\frac{n\hat{\sigma}^2}{\chi_{\frac{\alpha}{2}}^2} \leq \sigma^2 \leq \frac{n\hat{\sigma}^2}{\chi_{1-\frac{\alpha}{2}}^2}\right) = 1 - \alpha.$$

Et $(1 - \alpha) \cdot 100\%$ -konfidensintervall for σ^2 er dermed

$$\left[\frac{n\hat{\sigma}^2}{\chi_{\frac{\alpha}{2}}^2}, \frac{n\hat{\sigma}^2}{\chi_{1-\frac{\alpha}{2}}^2} \right].$$

- c) Sentralgrenseteoremt antar at man har sier at dersom man har at $Y_1, Y_2, \dots, Y_n$ er uavhengige stokastiske variabler med samme sannsynlighetsfordeling $f(x)$ hvor $E[Y_i] = \mu$ og $\text{Var}[Y_i] = \tau^2$, og man definerer

$$Z = \frac{\bar{Y} - \mu}{\sqrt{\frac{\tau^2}{n}}}$$

vil sannsynlighetsfordelingen til Z konvergere mot en standard normalfordeling når $n \rightarrow \infty$.

Rent praktisk betyr dette at når n er stor nok vil Z være tilnærmet standard normalfordelt. Ved å velge $Y_i = X_i^2$ og bruke at da $\mu = E[Y_i] = E[X_i^2] = \sigma^2$ og $\tau^2 = \text{Var}[Y_i] = \text{Var}[X_i^2] = 3\sigma^4$ gir dette at

$$Z = \frac{\frac{1}{n} \sum_{i=1}^n X_i^2 - \sigma^2}{\sqrt{\frac{3\sigma^4}{n}}} = \frac{\hat{\sigma}^2 - \sigma^2}{\sigma^2 \sqrt{\frac{3}{n}}} \underset{\approx}{=} N(0, 1).$$

- d) Siden Z definert i besvarelsen av forrige punkt er tilnærmet standard normalfordelt har vi at

$$P\left(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}\right) \approx 1 - \alpha.$$

Ved å sette inn uttrykket vi har for Z fra besvarelsen av forrige punkt betyr dette at

$$P\left(-z_{\frac{\alpha}{2}} \leq \frac{\hat{\sigma}^2 - \sigma^2}{\sigma^2 \sqrt{\frac{3}{n}}} \leq z_{\frac{\alpha}{2}}\right) \approx 1 - \alpha.$$

Løser så hver av ulikhetene inne i dette sannsynlighetsuttrykket med hensyn på σ^2 . Den første ulikheten gir

$$-z_{\frac{\alpha}{2}} \leq \frac{\hat{\sigma}^2 - \sigma^2}{\sigma^2 \sqrt{\frac{3}{n}}} \Leftrightarrow \sigma^2 \left(1 - z_{\frac{\alpha}{2}} \sqrt{\frac{3}{n}}\right) \leq \hat{\sigma}^2 \Leftrightarrow \sigma^2 \leq \frac{\hat{\sigma}^2}{1 - z_{\frac{\alpha}{2}} \sqrt{\frac{3}{n}}},$$

hvor vi benytter at $1 - z_{\frac{\alpha}{2}} \sqrt{\frac{3}{n}} > 0$ når n er stor. Den andre ulikheten gir og

$$\frac{\hat{\sigma}^2 - \sigma^2}{\sigma^2 \sqrt{\frac{3}{n}}} \leq z_{\frac{\alpha}{2}} \Leftrightarrow \hat{\sigma}^2 \leq \sigma^2 \left(1 + z_{\frac{\alpha}{2}} \sqrt{\frac{3}{n}}\right) \Leftrightarrow \frac{\hat{\sigma}^2}{1 + z_{\frac{\alpha}{2}} \sqrt{\frac{3}{n}}} \leq \sigma^2.$$

Dermed har vi at

$$P\left(\frac{\hat{\sigma}^2}{1 + z_{\frac{\alpha}{2}} \sqrt{\frac{3}{n}}} \leq \sigma^2 \leq \frac{\hat{\sigma}^2}{1 - z_{\frac{\alpha}{2}} \sqrt{\frac{3}{n}}}\right) \approx 1 - \alpha$$

og et tilnærmet $(1 - \alpha) \cdot 100\%$ -konfidensintervall for σ^2 blir

$$\left[\frac{\hat{\sigma}^2}{1 + z_{\frac{\alpha}{2}} \sqrt{\frac{3}{n}}}, \frac{\hat{\sigma}^2}{1 - z_{\frac{\alpha}{2}} \sqrt{\frac{3}{n}}} \right].$$

- e) Vi bestemmer k ut fra kravet

$$\begin{aligned} P(\text{Forkast } H_0 | H_0 \text{ er riktig}) &= \alpha \\ P(\hat{\sigma}^2 \leq k\sigma_0^2 | \sigma^2 = \sigma_0^2) &= \alpha \end{aligned}$$

Vi ønsker så å benytte at vi fra (b) vet at $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi_n^2$. Vi omskriver derfor sannsynligheten over og får kravet

$$\begin{aligned} P(\hat{\sigma}^2 \leq k\sigma_0^2 | \sigma^2 = \sigma_0^2) &= P\left(\frac{n\hat{\sigma}^2}{\sigma^2} \leq \frac{nk\sigma_0^2}{\sigma^2} \middle| \sigma^2 = \sigma_0^2\right) \\ &= P\left(\frac{n\hat{\sigma}^2}{\sigma^2} \leq \frac{nk\sigma_0^2}{\sigma_0^2} \middle| \sigma^2 = \sigma_0^2\right) \\ &= P\left(\frac{n\hat{\sigma}^2}{\sigma^2} \leq nk\right) = \alpha, \end{aligned}$$

som er oppfylt dersom

$$nk = \chi_{1-\alpha, n}^2.$$

Vi må derfor ha

$$\underline{\underline{k = \frac{\chi_{1-\alpha, n}^2}{n}}}$$

Type III-feil er ikke å forkaste H_0 når H_0 er feil. Sannsynligheten det spørres etter er følgelig

$$\begin{aligned}\beta &= P(\text{Ikke forkast } H_0 | \sigma^2 = \gamma\sigma_0^2) \\ &= P\left(\hat{\sigma}^2 > k\sigma_0^2 \mid \sigma^2 = \gamma\sigma_0^2\right) \\ &= P\left(\hat{\sigma}^2 > \frac{\chi_{1-\alpha, n}^2 \sigma_0^2}{n} \mid \sigma^2 = \gamma\sigma_0^2\right)\end{aligned}$$

Vi ønsker så å benytte resultatet i (c), så vi standardisere derfor $\hat{\sigma}^2$,

$$\begin{aligned}\beta &= P\left(\frac{\hat{\sigma}^2 - \sigma^2}{\sigma^2 \sqrt{\frac{3}{n}}} > \frac{\frac{\chi_{1-\alpha, n}^2 \sigma_0^2}{n} - \sigma^2}{\sigma^2 \sqrt{\frac{3}{n}}} \mid \sigma^2 = \gamma\sigma_0^2\right) \\ &= P\left(\frac{\hat{\sigma}^2 - \sigma^2}{\sigma^2 \sqrt{\frac{3}{n}}} > \frac{\frac{\chi_{1-\alpha, n}^2 \sigma_0^2}{n} - \gamma\sigma_0^2}{\gamma\sigma_0^2 \sqrt{\frac{3}{n}}} \mid \sigma^2 = \gamma\sigma_0^2\right) \\ &= 1 - P\left(\frac{\hat{\sigma}^2 - \sigma^2}{\sigma^2 \sqrt{\frac{3}{n}}} \leq \frac{\frac{\chi_{1-\alpha, n}^2 \sigma_0^2}{n} - \gamma\sigma_0^2}{\gamma\sigma_0^2 \sqrt{\frac{3}{n}}}\right) \\ &\approx \underline{\underline{1 - \Phi\left(\frac{\frac{\chi_{1-\alpha, n}^2 \sigma_0^2}{n} - \gamma\sigma_0^2}{\gamma\sigma_0^2 \sqrt{\frac{3}{n}}}\right)}}.\end{aligned}$$

Oppgave 7

For å sjekke om $\hat{\theta}$ er forventningsrett trenger vi å regne ut $E[\hat{\theta}]$,

$$\begin{aligned}E[\hat{\theta}] &= E[n\hat{p}(1 - \hat{p})] \\ &= E\left[n \cdot \frac{X}{n} \cdot \left(1 - \frac{X}{n}\right)\right] \\ &= E\left[X - \frac{X^2}{n}\right] \\ &= E[X] - \frac{1}{n}E[X^2].\end{aligned}$$

Siden X er binomisk fordelt med n forsøk og sannsynlighet for suksess lik p vet vi at $E[X] = np$ og $\text{Var}[X] = np(1-p)$. For å finne $E[X^2]$ kan vi bruke den generelle sammenhengen

$$\text{Var}[X] = E[X^2] - E[X]^2.$$

Denne gir her at

$$E[X^2] = \text{Var}[X] + E[X]^2 = np(1-p) + (np)^2 = np(1-p+np).$$

Setter så dette inn i uttrykket vi fant for $E[\hat{\theta}]$,

$$\begin{aligned} E[\hat{\theta}] &= np - \frac{1}{n} \cdot np(1-p+np) \\ &= np - p(1-p+np) \\ &= np - p + p^2 - np^2 \\ &= (np - np^2) - (p - p^2) \\ &= np(1-p) - p(1-p) \\ &= (n-1)p(1-p) \\ &= \frac{n-1}{n} \cdot np(1-p) \\ &= \frac{n-1}{n} \cdot \theta. \end{aligned}$$

Vi har altså at $E[\hat{\theta}] \neq \theta$ og har dermed vist at $\hat{\theta}$ er forventningsskjev. Men vi ser også at ved å gange $\hat{\theta}$ med faktoren $\frac{n}{n-1}$ får vi en forventningsrett estimator for θ . En forventningsrett estimator for θ er altså,

$$\underline{\underline{\tilde{\theta} = \frac{n}{n-1}\hat{\theta} = \frac{n^2}{n-1} \cdot \hat{p}(1-\hat{p}).}}$$