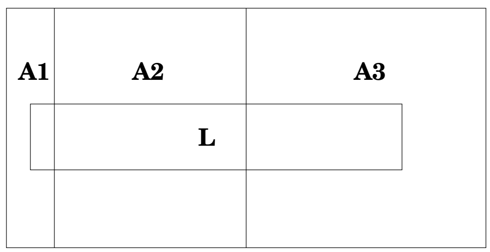
Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

LØSNINGSFORSLAG TIL EKSAMEN I FAG TMA4245 STATISTIKK

Lørdag 5. juni 2004

Oppgave 1 Atle, du lyver!

a) Venndiagram for de fire hendelsene:



For å regne ut $P(L'|A_2)$ benytter vi regelen for sannsynlighet for komplementære hendelser:

$$\begin{aligned} P(L'|A_2) + P(L|A_2) &= 1 \\ P(L'|A_2) &= = 1 - P(L|A_2) = 1 - 0.2 = \underline{\underline{0.8}} \end{aligned}$$

For å regne ut $P(L)$ bruker vi setningen om total sannsynlighet. Vi vet at A_1, A_2, A_3 er en partisjon av utfallsrommet (det ser vi lett av venndiagrammet).

$$\begin{aligned} P(L) &= P(L \cap A_1) + P(L \cap A_2) + P(L \cap A_3) \\ &= P(L|A_1) \cdot P(A_1) + P(L|A_2) \cdot P(A_2) + P(L|A_3) \cdot P(A_3) \\ &= 0.05 \cdot 0.1 + 0.2 \cdot 0.4 + 0.6 \cdot 0.5 = \underline{\underline{0.385}} \end{aligned}$$

b) Betingelser for at X er binomisk fordelt:

- Vi spør n personer.
- For hver person registerer vi om personen lyver eller ikke lyver (to komplmentære hendelser).
- Sannsynligheten for at en tilfeldig valgt person lyver er p, og denne er den samme for alle de n personene vi spør.

- De n personene vi spør svarer uavhengig av hverandre (n uavhengige forsøk).

Under disse 4 betingelsene er X="antall personer som lyver" binomisk fordelt med parametere n og p. Dermed er sannsynlighetsfordelingen til X gitt ved punktsannsynligheten $f(x)$,

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, ..., n$$

Vi vet at da er forventningen til X $E(X) = np$ og variansen $\mathrm{Var}(X) = np(1-p)$.

Videre: vi har at $p = 0.2$, og $n = 20$. $P(X = 4)$ finner vi ved å sette inn $X = 4$ i punktsannsynligheten $f(x)$ over.

$$P(X = 4) = f(4) = \binom{20}{4} 0.2^4 (1 - 0.2)^{20-4} = \underline{\underline{0.218}}$$

Det er også mulig å finne $P(X = 4)$ ved tabelloppslag (s 17 i formelsamlingen),

$$P(X = 4) = P(X \leq 4) - P(X \leq 3) = 0.630 - 0.411 = 0.219$$

Sannsynligheten $P[(X \leq 2) \cup (X > 5)]$ finner vi enklest ved tabelloppslag (s 17 i formelsamlingen),

$$\begin{aligned} P[(X \leq 2) \cup (X > 5) = P(X \leq 2) + P(X > 5) &= (X \leq 2) + 1 - P(X \leq 5) \\ &= 0.206 + 1 - 0.804 = \underline{\underline{0.402}} \end{aligned}$$

c) Nå er p ukjent.

Først forventning:

$$\begin{aligned} E(\hat{p}) &= E(\frac{X}{n}) = \frac{1}{n} E(X) = \frac{1}{n} np = p \\ E(p^*) &= E(\frac{X}{n-1}) = \frac{1}{n-1} E(X) = \frac{1}{n-1} np = \frac{n}{n-1} p \end{aligned}$$

Vi ser videre på varians:

$$\begin{aligned} \mathrm{Var}(\hat{p}) &= \mathrm{Var}(\frac{X}{n}) = \frac{1}{n^2} \mathrm{Var}(X) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n} \\ \mathrm{Var}(p^*) &= \mathrm{Var}(\frac{X}{n-1}) = \frac{1}{(n-1)^2} \mathrm{Var}(X) = \frac{1}{(n-1)^2} np(1-p) = \frac{np(1-p)}{(n-1)^2} \end{aligned}$$

En god estimator $\hat{p}$ er en estimator som er

- forventningsrett, dvs. $E(\hat{p}) = p$, og

- har liten varians, dvs. $\text{Var}(\hat{p})$ er liten.

Vi liker veldig godt hvis variansen minker når antall observasjoner som estimatoren er basert på øker.

Sammenligner vi to estimatorer som begge er forventningsrette velger vi estimatoren med minst varians. Sammenlinger vi to estimatorer der kun den ene er forventningsrett, velger vi gjerne den estimatoren som er forventningsrett (ofte sjekker vi også at det ikke er veldig stor forskjell på variansene).

For å velge mellom $\hat{p}$ og p^* ser vi på uttrykkene for forventning og varians til begge estimatorene.

Vi ser at $\hat{p}$ er forventningsrett, men det er ikke p^*. I prinsippet kan vi stoppe her og konkluere med at vi foretrekker den forventningsrette estimatoren $\hat{p}$. Men, det kan være fint å sjekke at det ikke er stor forskjell på variansen til de to estimatorene (hva hvis den ene hadde hatt to ganger så stor varians?).

Vi ser at $\text{Var}(\hat{p}) = (\frac{(n-1)}{n})^2\text{Var}(p^*)$, dvs. $\text{Var}(\hat{p}) < \text{Var}(p^*)$ med en faktor $(\frac{n-1}{n})^2$ i forskjell. For $n = 20$ er denne faktoren $(\frac{19}{20})^2 = 0.95^2 = 0.9$, dvs. $\text{Var}(\hat{p}) = 0.9 \cdot \text{Var}(p^*)$. Dermed har estimatoren $\text{Var}(\hat{p})$ både minst varians og er forventningsrett. Vi velger derfor estimatoren $\hat{p}$.

Kommentarer: Asymptotisk (når $n \to \infty$) vil de to estimatorene være like gode. Vi har i vårt pensum ikke snakket om begrepet konsistente estimatorer, men begge disse estimatorene er konsistente.

d) Vi velger den konservative og hittil "gjeldene" antakelsen om at $p = 0.2$ som nullhypotese og hypotesen som vi ønsker å teste, $p > 0.2$ som alternativ hypotese.

$$H_0 : p = 0.2 \qquad vs. \qquad H_1 : p > 0.2$$

Vi bruker $\hat{p}$ som estimator for p og vi vil forkaste H_0 når $\hat{p}$ er stor. Det betyr at vi forkaster H_0 når $Z_0 = \frac{\hat{p}-0.2}{\sqrt{\frac{p(1-p)}{n}}}$ er større en en konstant k. Vi skriver "Forkast H_0 når $Z_0 \geq k$".

Videre bestemmer vi k slik at

$$P(\text{type I feil}) = P(\text{forkaste } H_0 | H_0 \text{ er sann}) \leq \alpha.$$

Innsatt $Z_0 \geq k$ for hendelsen "forkaste H_0" og $p = 0.2$ for hendelsen "H_0 sann":

$$\begin{aligned} P(\text{forkaste } H_0 | H_0 \text{ er sann}) &\leq \alpha \\ P(Z_0 \geq k | p = 0.2) &= \alpha \\ P(\frac{\hat{p} - 0.2}{\sqrt{\frac{0.2(1-0.2)}{n}}} \geq k | p = 0.2)) &= \alpha \end{aligned}$$

Her er $\frac{\hat{p}-0.2}{\sqrt{\frac{0.2(1-0.2)}{n}}}$ tilnærmet standard normalfordelt under H_0 og tallet k som har areal α til høyre i standard normalfordelingen er kvantilen z_α, dvs. $k = z_\alpha$.

i) Dvs. vi forkaster H_0 når

$$\underline{\underline{\frac{\hat{p}-0.2}{\sqrt{\frac{0.2(1-0.2)}{n}}} \geq z_\alpha}}$$

ii) Alternativt kan vi løse ut forkastningsområdet over som

$$\underline{\underline{\hat{p} \geq 0.2 + z_\alpha \sqrt{\frac{0.2(1-0.2)}{n}}}}$$

For $\alpha = 0.01$ er $z_{0.01} = 2.326$. Videre har vi $n = 200$ og $x = 55$. Vi kan bruke begge måtene for å skrive opp forkastningsområdet:

i) $\frac{\hat{p}-0.2}{\sqrt{\frac{0.2(1-0.2)}{n}}} = 2.65$ som er større enn 2.326, og gir forkastning.

ii) $\hat{p} = \frac{55}{200} = 0.275$ og forkastningsområdet $\hat{p} \geq 0.2 + 2.326\sqrt{\frac{0.2(1-0.2)}{200}} = 0.27$. Her er $0.275 > 0.27$ og vi forkaster H_0.

Konklusjonen er at det vi har observert (eller noe verre) er lite sannsynlig (mindre sannsynlig enn 0.01) når H_0 er sann, og vi forkaster dermed H_0 og konkluderer med at p er større enn 0.2.

$P-$verdien angir sannsynligheten for det vi har observert eller noe verre gitt at H_0 er sann (der verre henspeiler på den alternative hypotesen). Vi har forkastet H_0 med signifikansnivå 0.01, det betyr at det vi har observert eller noe verre har mindre sannsynlighet enn 0.01 når H_0 er sann. Det betyr at p-verdien til testen vil være *mindre* enn 0.01.

Vi kan raskt regne ut p-verdien (men det er ikke krevd i oppgave). Vi har observert at $z_0 = \frac{\hat{p}-0.2}{\sqrt{\frac{0.2(1-0.2)}{n}}} = 2.65$. Under H_0 er det mer ekstremt hvis denne verdien hadde vært enda større.

$$\begin{aligned}
P(\text{det vi har observert eller noe verre}|H_0 \text{ er sann}) &= P(Z_0 \geq 2.65|p = 0.2) \\
&= 1 - \Phi(2.65) \\
&= 1 - 0.9960 = 0.004
\end{aligned}$$

Oppgave 2 Aksjekurser

a) X er normal($\mu_X = 0.15$ kroner, $\sigma_X = 0.6$ kroner)

$$\begin{aligned} P(X < 0) &= P(X \leq 0) = P(\frac{X - 0.15}{0.6} \leq \frac{0 - 0.15}{0.6}) \\ &= \Phi(-0.25) = \underline{\underline{0.401}} \end{aligned}$$

$$\begin{aligned} P(0 \leq X \leq 0.15) &= P(X \leq 0.15) - P(X \leq 0) \\ &= P(\frac{X - 0.15}{0.6} \leq \frac{0.15 - 0.15}{0.6}) - \Phi(-0.25) \\ &= \Phi(0) - \Phi(-0.25) = 0.5 - 0.401 = \underline{\underline{0.1}} \end{aligned}$$

Vi ser på $10 \cdot X$ og finner forventning og varians:

$$\begin{aligned} E(10 \cdot X) &= 10 \cdot E(X) = 10 \cdot \mu_X = 10 \cdot 0.15 = \underline{\underline{1.5}} \\ \text{Var}(10 \cdot X) &= 10^2 \cdot \text{Var}(X) = 10^2 \cdot \sigma_X^2 = 100 \cdot 0.6^2 = \underline{\underline{36}} \end{aligned}$$

b) X er normal($\mu_X = 0.15$ kroner, $\sigma_X = 0.6$) og Y er normal($\mu_X = 0.15$ kroner, $\sigma_X = 0.8$).

Tre strategier skal sammenlignes, og den med minst risiko velges.

Minst risiko har man hvis man har høy forventningsverdi og liten varians. Alternativt kan man si at minst risiko har den strategien der vi har minst sannsynlighet for å tape penger. (Du behøver bare se på forventning og varians, eller alternativt på sannsynlighet for tap – ikke på begge deler).

Vi studerer de tre strategiene:

i) Kjøp to aksjer i Agderfrukt, fortjenesten er da $2 \cdot X$.

$$\begin{aligned} E(2 \cdot X) &= 2 \cdot E(X) = 2 \cdot \mu_X = 2 \cdot 0.15 = \underline{\underline{0.3}} \\ \text{Var}(2 \cdot X) &= 2^2 \cdot \text{Var}(X) = 2^2 \cdot \sigma_X^2 = 4 \cdot 0.6^2 = \underline{\underline{1.44}} \\ P(2 \cdot X < 0) &= P(2 \cdot X \leq 0) \\ &= P(\frac{2 \cdot X - 0.3}{\sqrt{1.44}} \leq \frac{0 - 0.3}{\sqrt{1.44}}) \\ &= \Phi(-0.25) = 0.40 \end{aligned}$$

Vi kunne sett at $P(2 \cdot X < 0) = P(X < 0)$ uten regning.

ii) Kjøp en aksje i Agderfrukt og en aksje i Trønderfrukt, fortjenesten er da $X+Y$ og X og Y er uavhengige.

$$\begin{aligned}
E(X+Y) &= E(X)+E(Y) = \mu_X + \mu_Y = 0.15 + 0.15 = \underline{\underline{0.3}} \\
\mathrm{Var}(X+Y) &= \mathrm{Var}(X)+\mathrm{Var}(Y) = \sigma_X^2 + \sigma_Y^2 = 0.6^2 + 0.8^2 = \underline{\underline{1.0}} \\
P(X+Y<0) &= P(X+Y \leq 0) \\
&= P(\frac{X+Y-0.3}{\sqrt{1}} \leq \frac{0-0.3}{\sqrt{1}}) \\
&= \Phi(-0.3) = 0.38
\end{aligned}$$

iii) Kjøp to aksjer i Trønderfrukt, fortjenesten er da $2 \cdot Y$.

$$\begin{aligned}
E(2 \cdot Y) &= 2 \cdot E(Y) = 2 \cdot \mu_Y = 2 \cdot 0.15 = \underline{\underline{0.3}} \\
\mathrm{Var}(2 \cdot Y) &= 2^2 \cdot \mathrm{Var}(Y) = 2^2 \cdot \sigma_Y^2 = 4 \cdot 0.8^2 = \underline{\underline{2.56}} \\
P(2 \cdot Y < 0) &= P(2 \cdot Y \leq 0) \\
&= P(\frac{2 \cdot Y - 0.3}{\sqrt{2.56}} \leq \frac{0-0.3}{\sqrt{2.56}}) \\
&= \Phi(-0.19) = 0.43
\end{aligned}$$

Her ser vi at alle tre strategier har samme forventningsverdi, mens strategi ii) har lavest varians. Vi vil derfor foretrekke å kjøpe en aksje i Agderfrukt og en aksje i Trønderfrukt.

Vi ser at vi får samme konklusjon, dvs. at det er minst sannsynlighet for å tape penger når vi kjøper en aksje i Agderfrukt og en aksje i Trønderfrukt.

c) Situasjonen er en videreføring av situasjonen i 2a og 2b, der vi nå innfører kursendring for flere dager. Vi har for Agderfrukt at X_i er normalfordelt med forventning $\mu_X = 0.15$ kr og standardavvik $\sigma_X = 0.60$ kr. Tilsvarende for Trønderfrukt med $\mu_Y = 0.15$ kr og standardavvik $\sigma_Y = 0.80$ kr. Videre er kursendringen på ulike dager uavhengige. Dette betyr at X_i og X_j er uavhengige for $i \neq j$, at Y_i og Y_j er uavhengige for $i \neq j$ og at X_i og Y_j er uavhengige for $i \neq j$.

Ser vi på figuren over utviklingen av de to aksjekursene, så ser vi at de to kurvene følger hverandre og at en positiv endring for Agderfrukt ofte skjer sammen med en positiv endring for Trønderfrukt. Det betyr at X_i og Y_i, *ikke* er uavhengige, men at vi ser en *positiv* korrelasjon. Vi ser at X_i og Y_i er *positivt korrelerte.*

Korrelasjonen mellom X_i og Y_i for disse to selskapene, $\rho(X_i, Y_i)$, må da være 0.5 siden det er det eneste av de tre foreslåtte verdiene for korrelasjonen som er positiv. Til sammenligning så betyr en korrelasjon på 0 at vi ser svært liten sammenheng mellom endringen i aksjeverdien til Agderfrukt og endringen i aksjeverdien til Trønderfrukt (hvis X_i og Y_i er uavhengige fører det til at korrelasjonen er 0). Korrelasjon med negativ verdi, f.eks. -0.5, betyr at en positiv endring i Agderfrukt ofte skjer sammen med en negativ endring for Trønderfrukt.

Vi ser så på gjennomsnittet over 10 dager av differansene i endringen i aksjekursen til Agderfrukt og endringen i aksjekursen til Trønderfrukt,$\overline{D}$.

$$\begin{aligned} D_i &= X_i - Y_i \\ \overline{D} &= \frac{1}{10}\sum_{i=1}^{10} D_i \end{aligned}$$

Først forventingsverdi:

$$\begin{aligned} E(\overline{D}) &= E(\frac{1}{10}\sum_{i=1}^{10} D_i) = \frac{1}{10}\sum_{i=1}^{10} E(X_i - Y_i) \\ &= \frac{1}{10}\sum_{i=1}^{10}[E(X_i) - E(Y_i)] \\ &= \frac{1}{10}\sum_{i=1}^{10}[\mu_X - \mu_Y] \\ &= \mu_X - \mu_Y = 0.15 - 0.15 = \underline{\underline{0}} \end{aligned}$$

Før vi beregner variansen til $\overline{D}$ ser vi først på variansen til D_i. Vi bruker vi formelen for variansen til en lineærkombinasjon på side 34 i formelsamlingen (men vår sum har bare to ledd).

$$\begin{aligned} \text{Var}(D_i) &= \text{Var}(X_i - Y_i) \\ &= \text{Var}(X_i) + (-1)^2\text{Var}(Y_i) + 2\cdot(-1)\cdot\text{Cov}(X_i, Y_i) \end{aligned}$$

Vi trenger kovariansen mellom X_i og Y_i, dvs. for endringen i aksjeverdien til Agderfrukt for dag i og endringen i aksjeverdien til Trønderfrukt for dag i (samme dag). Denne er *ikke* 0. Vi vet at $\rho(X_i, Y_i) = 0.5$, og vi har følgende sammenheng mellom korrelasjon og kovarians (formelsamlingen side 33).

$$\rho(X_i, Y_i) = \frac{\text{Cov}(X_i, Y_i)}{\sqrt{\text{Var}(X_i)}\sqrt{\text{Var}(Y_i)}}$$

Dermed blir $\text{Cov}(X_i, Y_i)$

$$\begin{aligned}
\text{Cov}(X_i, Y_i) &= \rho(X_i, Y_i)\sqrt{\text{Var}(X_i)}\sqrt{\text{Var}(Y_i)} \\
&= \rho(X_i, Y_i)\sigma_X \sigma_Y \\
&= 0.5 \cdot \sigma_X \sigma_Y
\end{aligned}$$

Regner vi videre på $\text{Var}(D_i)$ får vi da:

$$\begin{aligned}
\text{Var}(D_i) &= \text{Var}(X_i - Y_i) \\
&= \text{Var}(X_i) + (-1)^2\text{Var}(Y_i) + 2 \cdot (-1) \cdot \text{Cov}(X_i, Y_i) \\
&= \sigma_X^2 + \sigma_Y^2 - 2 \cdot \rho(X_i, Y_i)\sigma_X \sigma_Y
\end{aligned}$$

Vi har videre at D_i og D_j er uavhengige for $i \neq j$. Dette siden kursendringer for ulike dager er uavhengige.

$$\begin{aligned}
\text{Var}(\overline{D}) &= \text{Var}(\frac{1}{10}\sum_{i=1}^{10} D_i) \\
&= (\frac{1}{10})^2 \sum_{i=1}^{10} \text{Var}(D_i) \\
&= (\frac{1}{10})^2 \sum_{i=1}^{10} (\sigma_X^2 + \sigma_Y^2 - 2 \cdot \rho(X_i, Y_i)\sigma_X \cdot \sigma_Y) \\
&= \frac{1}{10}(\sigma_X^2 + \sigma_Y^2 - 2 \cdot 0.5 \cdot \sigma_X \cdot \sigma_Y) \\
&= \frac{1}{10}(0.6^2 + 0.8^2 - 2 \cdot 0.5 \cdot 0.6 \cdot 0.8) \\
&= \underline{\underline{0.052}}
\end{aligned}$$

Oppgave 3 Bølgehøyde

a) Kumulativ fordeling:

Bruk substitusjonen $u = s^2$ med $du = 2sds$:

$$\begin{aligned}
P(X \leq x) &= \int_0^x f(s)ds = \int_0^x \frac{2s}{\theta} e^{-\frac{s^2}{\theta}} ds \\
&= \int_0^{x^2} \frac{1}{\theta} e^{-\frac{u}{\theta}} du = \left[-e^{-\frac{u}{\theta}}\right]_0^{x^2} \\
&= \underline{\underline{1 - e^{-\frac{x^2}{\theta}}}}
\end{aligned}$$

Betinget fordeling:
Vi har at $P(X > x) = 1 - P(X \leq x) = e^{-\frac{x^2}{\theta}}$.

$$\begin{aligned} P(X > 15 | X > 10) &= \frac{P(X > 15 \cap X > 10)}{P(X > 10)} = \frac{P(X > 15)}{P(X > 10)} \\ &= \frac{e^{-\frac{15^2}{25}}}{e^{-\frac{10^2}{25}}} \\ &= \frac{0.000123}{0.0183} \approx \underline{\underline{0.0067}}. \end{aligned}$$

b) Rimelighetsfunksjonen er gitt ved:

$$\begin{aligned} L(\theta; x_1, ..., x_n) &= f(x_1, ..., x_n; \theta) \\ &= f(x_1; \theta) \cdots f(x_n; \theta) \\ &= \frac{2x_1}{\theta} e^{-\frac{x_1^2}{\theta}} \cdots \frac{2x_n}{\theta} e^{-\frac{x_n^2}{\theta}} \\ &= \left(\frac{2}{\theta}\right)^n \cdot (x_1 \cdots x_n) \cdot e^{-\frac{1}{\theta}\sum_{i=1}^n x_i^2}. \end{aligned}$$

Tar logaritmen:

$$\begin{aligned} l(\theta; x_1, ..., x_n) &= \ln\left[L(\theta)\right] = n \ln\left(\frac{2}{\theta}\right) + \ln(x_1 \cdots x_n) - \frac{1}{\theta}\sum_{i=1}^n x_i^2 \\ &= n\ln(2) - n\ln(\theta) + \ln(x_1 \cdots x_n) - \frac{1}{\theta}\sum_{i=1}^n x_i^2. \end{aligned}$$

Finn maksimumspunkt ved å derivere ln-rimelighetsfunksjonen og sette lik 0.

$$\begin{aligned} \frac{\partial l}{\partial \theta}\Big|_{\theta=\hat{\theta}} &= 0 - n\frac{1}{\theta} + 0 + \frac{1}{\theta^2}\sum_{i=1}^n x_i^2 = 0 \\ \hat{\theta} &= \frac{1}{n}\sum_{i=1}^n x_i^2. \end{aligned}$$

SME for θ blir da $\underline{\underline{\hat{\theta} = \frac{1}{n}\sum_{i=1}^n X_i^2}}$.

Viser at $\hat{\theta}$ er forventningsrett:

$$\mathrm{E}[\hat{\theta}] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^n X_i^2\right] = \frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_i^2] = \frac{1}{n}\sum_{i=1}^n \theta = \underline{\underline{\theta}}.$$

For å regne ut variansen til $\hat{\theta}$ trenger vi først variansen til X^2. La $Y = X^2$. Bruker at

$\mathrm{Var}[Y] = \mathrm{E}[Y^2] - \mathrm{E}[Y]^2$. Dermed er

$$\begin{aligned} \mathrm{Var}[X^2] &= \mathrm{Var}[Y] = \mathrm{E}[Y^2] - \mathrm{E}[Y]^2 \\ &= \mathrm{E}[X^4] - \mathrm{E}[X^2]^2 = 2\theta^2 - \theta^2 = \underline{\underline{\theta^2}}. \end{aligned}$$

Da får vi:

$$\mathrm{Var}[\hat{\theta}] = \mathrm{Var}[\frac{1}{n}\sum_{i=1}^{n} X_i^2] = \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{\mathrm{Var}[X_i^2]}_{=\theta^2} = \frac{\theta^2}{n}.$$

c) Sentralgrenseteoremet sier at hvis $Y_1, ..., Y_n$ er uavhengige og identisk fordelte, så er gjennomsnittet tilnærmet normalfordelt. Mer presist er

$$Z = \frac{\bar{Y} - \mathrm{E}[\bar{Y}]}{\sqrt{\mathrm{Var}[\bar{Y}]}}$$

tilnærmet standard normalfordelt når n er stor (>30). I vårt tilfelle er $Y_i = X_i^2$ uavhengige og $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \hat{\theta}$. Videre er som funnet i forrige punkt (3b), $E(\bar{Y}) = \theta$ og $\mathrm{Var}(\bar{Y}) = \frac{\theta^2}{n}$. Dermed blir

$$Z = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \theta}{\sqrt{\frac{\theta^2}{n}}}$$

tilnærmet normalfordelt når n er stor.

Vi finner så et 95% konfidensintervall med $\alpha = 0.05$. Sett $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$. Bruker at

$$\begin{aligned} 1-\alpha &\approx P\left(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}\right) \\ &= P\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{Y} - \theta}{\sqrt{\frac{\theta^2}{n}}} \leq z_{\frac{\alpha}{2}}\right) \\ &= P\left(-\frac{z_{\frac{\alpha}{2}}}{\sqrt{n}} \leq \frac{1}{\theta}\bar{Y} - 1 \leq \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}\right) \\ &= P\left(1 - \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}} \leq \frac{1}{\theta}\bar{Y} \leq 1 + \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}\right) \\ &= P\left(\frac{1}{1 + \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}} \leq \frac{\theta}{\bar{Y}} \leq \frac{1}{1 - \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}}\right) \\ &= P\left(\frac{\bar{Y}}{1 + \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}} \leq \theta \leq \frac{\bar{Y}}{1 - \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}}\right) \end{aligned}$$

Et tilnærmet 95% konfidensintervall for θ blir da

$$\underline{\underline{\left[\frac{\frac{1}{n}\sum_{i=1}^n X_i^2}{1+\frac{z_{0.025}}{\sqrt{n}}}, \quad \frac{\frac{1}{n}\sum_{i=1}^n X_i^2}{1-\frac{z_{0.025}}{\sqrt{n}}}\right]}}$$

Siden $p = e^{-\frac{100}{\theta}}$ er en monotont stigende funksjon av θ, har vi at

$$P\left(\underbrace{\frac{\frac{1}{n}\sum_{i=1}^n X_i^2}{1+\frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}}}_{=\hat{\theta}_L} \quad \le \quad \theta \quad \le \quad \underbrace{\frac{\frac{1}{n}\sum_{i=1}^n X_i^2}{1-\frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}}}_{=\hat{\theta}_U}\right) \quad \approx \quad 1-\alpha$$

$$\Updownarrow$$

$$P\left(e^{-\frac{100}{\hat{\theta}_L}} \quad \le \quad e^{-\frac{100}{\theta}} \quad \le \quad e^{-\frac{100}{\hat{\theta}_U}}\right) \quad \approx \quad 1-\alpha$$

Et tilnærmet 95% konfidensintervall for $e^{-\frac{100}{\theta}}$ blir da $\underline{\underline{\left[e^{-\frac{100}{\hat{\theta}_L}}, \quad e^{-\frac{100}{\hat{\theta}_U}}\right]}}$