



Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

TMA4240 Statistikk
Høst 2018

Anbefalt øving 10

Oppgavene i denne anbefalte øvingen har det til felles at de omhandler ulike typer konfidensintervall. Andre aktuelle tema er punktestimering og sannsynlighetsmaksimering.

Oppgave 1

Miljøkonsulenten i en kommune ønsker å undersøke den ukjente pH-verdien i et vann. Betegn den sanne pH-verdien for μ . Konsulenten har tilgjengelig to målemetoder. Metode I er rask, men måleresultatene er beheftet med betydelig måleusikkerhet. Metode II er mye mer tidkrevende, men gir mer nøyaktige målinger. Begge målemetodene er velbrukte og variansen i målingene er derfor kjent. Miljøkonsulenten velger å gjøre en observasjon med hver metode. La X betegne observasjonen ved bruk av metode I og Y observasjonen ved metode II. Vi antar at X og Y uavhengige og normalfordelt med

$$E(X) = \mu, \quad \text{Var}(X) = \sigma_0^2, \quad E(Y) = \mu, \quad \text{Var}(Y) = \tau_0^2$$

der σ_0^2 og τ_0^2 er kjente størrelser.

Det oppgis at en forventningsrett estimator (som forøvrig også er sannsynlighetsmaksimerings-estimator) for μ i denne situasjonen er

$$\hat{\mu} = \frac{\tau_0^2 X + \sigma_0^2 Y}{\tau_0^2 + \sigma_0^2}.$$

Ta utgangspunkt i estimatoren $\hat{\mu}$ og utled et $(1 - \alpha)100\%$ konfidensintervall for μ .

Oppgave 2

Levetiden, T , for en bestemt type elektroniske komponenter har vist seg å være eksponensialfordelt med parameter λ , dvs.

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & , t > 0 \\ 0 & , \text{ellers} \end{cases}$$

a) Vis at $Z = 2\lambda T$ er χ^2 -fordelt med 2 frihetsgrader, dvs. at

$$f_Z(z) = \begin{cases} \frac{1}{2} e^{-\frac{z}{2}} & , z > 0 \\ 0 & , \text{ellers} \end{cases}$$

b) Bruk de 10 målte levetidene som er angitt nedenfor til å finne et 90% konfidensintervall for λ .

Komponent i	1	2	3	4	5	6	7	8	9	10
t_i	941.3	465.2	479.1	1071.9	504.7	933.4	1137.3	343.7	165.2	388.4

Oppgave 3

Et ingeniørfirma har fått anbudet på å bygge en bro over en fjord. Ingeniørene har slått ned en påle på hver side av fjorden der brokarene skal konstrueres. En ønsker å måle avstanden, a , mellom pålene.

Firmaet har noe gammelt avstandsmåleutstyr som måler avstanden med standardavvik σ_G . En beslutter å ta n uavhengige målinger X_1, \dots, X_n med det gamle utstyret og antar at hver måling er normalfordelt med forventning a og varians σ_G^2 . Avstanden mellom pålene, a , er selvsagt ukjent og det er også σ_G^2 .

- a) En benytter følgende estimator for avstanden a :

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

Utled uttrykk for forventning og varians til estimatoren \hat{a} .

Hvilken sannsynlighetsfordeling har estimatoren \hat{a} ? Begrunn svaret.

- b) En benytter følgende estimator for målevariansen σ_G^2 :

$$S_G^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{a})^2$$

Utled et 95% konfidensintervall for avstanden mellom pålene, a , basert på estimatorene over.

Oppgave 4

For å kunne dimensjonere en oljeplattform er det viktig å vite hvor store bølgene kan bli i området der plattformen skal plasseres. Det settes derfor ut en bølgehøydemåler. La X være største bølgehøyde en tilfeldig valgt dag. Vi antar at sannsynlighetstettheten til X er gitt ved

$$f(x; \theta) = \frac{2x}{\theta} e^{-\frac{x^2}{\theta}}, \quad x \geq 0, \quad \theta > 0.$$

Det oppgis at $E[X^2] = \theta$ og $E[X^4] = 2\theta^2$.

- a) Vis at den kumulative fordelingsfunksjonen, $F(x) = P(X \leq x)$, er $F(x) = 1 - e^{-\frac{x^2}{\theta}}$. (Hint: Bruk substitusjon med $u = x^2$).

Gitt at største bølgehøyde er større enn 10 meter, finn sannsynligheten for at den er større enn 15 meter hvis $\theta = 25$, dvs. $P(X > 15 | X > 10)$?

I resten av oppgaven regnes θ som ukjent.

Vi har observert største bølgehøyde i n dager. La X_i være største bølgehøyde på dag i . Vi antar at X_1, \dots, X_n er uavhengige og identisk fordelte med sannsynlighetstetthet $f(x; \theta)$.

b) Finn sannsynlighetsmaksimeringsestimatorens (SME) $\hat{\theta}$ for θ .

Er estimatoren $\hat{\theta}$ forventningsrett?

Finn også variansen til $\hat{\theta}$.

c) Bruk sentralgrenseteoremet til å argumentere for at

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

er tilnærmet standard normalfordelt.

Bruk Z til å finne et tilnærmet 95% konfidensintervall for θ .

Sannsynligheten for at største bølgehøyde en tilfeldig valgt dag overskrider 10 meter er $P(X > 10) = e^{-\frac{100}{\theta}}$. Bruk det tilnærmede konfidensintervallet for θ til å finne et tilnærmet 95% konfidensintervall for $e^{-\frac{100}{\theta}}$.

Oppgave 5

La tiden X (målt i uker) mellom to påfølgende feil i et mobilnett være en kontinuerlig stokastisk variabel med sannsynlighetstetthet

$$f(x) = \beta x^{-\beta-1}, \quad x > 1, \beta > 1.$$

a) Vis at den kumulative fordelingsfunksjonen, $F(x)$, for X er $F(x) = 1 - x^{-\beta}$, for $x > 1$.

Anta i resten av dette punktet at $\beta = 3$.

Hva er sannsynligheten for at det tar mer enn 2 uker mellom to påfølgende feil?

Dersom det er gått 2 uker siden sist det var en feil på nettet, hva er sannsynligheten for at det svikter innen det er gått 3.5 uker fra siste feil?

Vi vil estimere parameteren β basert på data for tidligere tilfeller av feil på nettet. La X_i , $i = 1, \dots, n$ være lengden på n tidsintervaller (målt i uker) mellom to påfølgende feil. Vi antar at X_1, X_2, \dots, X_n er uavhengige og identisk fordelte stokastiske variabler, med sannsynlighetstetthet $f(x)$ som gitt i starten av oppgaven.

Tre alternative estimatorene for β er

$$\hat{\beta}_1 = \frac{n}{\sum_{i=1}^n \ln(X_i)}, \quad \hat{\beta}_2 = \frac{n-1}{\sum_{i=1}^n \ln(X_i)} \quad \text{og} \quad \hat{\beta}_3 = \frac{\sum_{i=1}^n \ln(X_i)}{n},$$

der \ln er den naturlige logaritmen.

b) Hvilken av estimatorene over er sannsynlighetsmaksimeringsestimatorens (SME)? Begrunn svaret ved å utlede SME. Beregn estimatet når $n = 10$ og de observerte verdiene er som følger:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1.23	2.04	1.27	1.79	1.10	1.29	2.74	1.15	1.10	1.06

Det oppgis at $\sum_{i=1}^{10} \ln(x_i) = 3.39$.

c) Vis at $2\beta \ln(X_i)$ er kjikvadratfordelt med 2 frihetsgrader, og videre at $2\beta \sum_{i=1}^n \ln(X_i)$ er kjikvadratfordelt med $2n$ frihetsgrader.

Utled et 95% konfidensintervall for β . Hva blir intervallet når dataene er som i punkt b)?

Fasit

2. b) $(8.438 \cdot 10^{-4}, 24.424 \cdot 10^{-4})$

4. a) 0.0067

5. a) 0.125, 0.813 b) 2.95 c) [1.41, 5.04]