TMA4240 Statistikk
Høst 2017

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

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

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

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

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

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

Ta utgangspunkt i estimatoren $\widehat{\mu}$ 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) Benytt resultatet i oppgave **a)** til å utlede et 90% konfidensintervall for λ.

I filen `levetider.txt` på hjemmesiden er 500 observerte levetider $T_1, \ldots, T_{500}$ som kan antas å ha tetthet $f_T(t)$ med ukjent parameter λ.

c) Det kan antas som kjent at sannsynlighetsmaksimeringsestimatoren (SME) til λ er $\hat{\lambda} = 500/\sum_{i=1}^{500} T_i$. Benytt Matlab til å beregne estimatet.

Bruk Matlab til å finne et 90% konfidensintervall for λ basert på uttrykket du utledet i oppgave **b)**.

Oppgave 3

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

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

Det oppgis at $\mathrm{E}[X^2] = \theta$ 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 $X_1, ..., X_n$ er uavhengige og identisk fordelte med sannsynlighetstetthet $f(x;\theta)$.

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

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

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

c) Bruk sentralgrenseteoremet til å argumentere for at

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

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 4

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 $X_i,\ i = 1, \ldots, n$ være lengden på n tidsintervaller (målt i uker) mellom to påfølgende feil. Vi antar at $X_1, X_2, \ldots, X_n$ er uavhengige og identisk fordelte stokastiske variabler, med sannsynlighetstetthet $f(x)$ som gitt i starten av oppgaven.

Tre alternative estimatorer 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 sannsynlighetsmaksimeringsestimatoren (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)?

Oppgave 5 Dekningssannsynlighet for konfidensintervall

La $X_1, X_2, \ldots, X_n$ være uavhengig og identisk normalfordelte tilfeldige variabler med ukjent forventningsverdi μ og ukjent varians σ^2.

a) Vis at et 90% konfidensintervall for μ er $[L, U]$ hvor

$$L(X_1, X_2, \ldots, X_n) = \overline{X} - t_{n-1,0.05} \cdot \sqrt{\frac{S^2}{n}}$$

og

$$U(X_1, X_2, \ldots, X_n) = \overline{X} + t_{n-1,0.05} \cdot \sqrt{\frac{S^2}{n}}$$

med

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i, \quad \text{og} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2,$$

og hvor $t_{n-1,0.05}$ er 0.95-kvantilen i t-fordelingen med $n-1$ frihetsgrader.

b) Sett $n = 10$, $\mu = 3$ og $\sigma^2 = 2^2$. Trekk realisasjoner av $X_1, X_2, \ldots, X_{10}$ fra en $N(\mu, \sigma^2)$-fordeling. Regn ut L og U. Dekker intervallet den sanne forventningsverdien, det vil si holder dobbeltulikheten $L \leq \mu \leq U$?

c) Gjenta prosedyren i **b)** 10 000 ganger. Finn den empiriske sannsynligheten for at konfidensintervallet dekker den sanne forventningsverdien μ. Kommenter svaret.

Fasit

2. **c**) 0.0097, $[0.0090, 0.0104]$

3. **a**) 0.0067

4. **a**) 0.125,0.813 **b**) 2.95 **c**) $[1.41, 5.04]$