TMA4240 Statistikk
Høst 2019

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

**Anbefalt øving 9**

Denne øvingen er basert på pensum forelest i 9. forelesningsuke. Oppgavene i denne øvingen handler blant annet om sannsynlighetsmaksimeringsestimator (SME), egenskaper til estimatorer og momentgenererende funksjoner.

### Oppgave 1

Miljøkonsulenten i en kommune ønsker å undersøke den ukjente pH-verdien i et vann. Betegn den sanne pH-verdien for $\mu$. 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 $\sigma_0^2$ og $\tau_0^2$ er kjente størrelser.

Det oppgis at en forventningsrett estimator (som forøvrig også er sannsynlighetsmaksimeringsestimator) for $\mu$ 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 $\mu$.

### Oppgave 2

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 $\theta$ 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 $\theta$.

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

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

**c)** Bruk sentralgrenseteoremet til å argumentere for at

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

er tilnærmet standard normalfordelt.

Bruk $Z$ til å finne et tilnærmet 95% konfidensintervall for $\theta$.

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 $\theta$ til å finne et tilnærmet 95% konfidensintervall for $e^{-\frac{100}{\theta}}$.

## Oppgave 3

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 $\beta$ 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 $\beta$ 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.

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

**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:

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 $\beta$. Hva blir intervallet når dataene er som i punkt b)?

## Oppgave 4

Når lys av vilkårlig retning treffer en kule så absorberes en viss andel av lyset. Vi betegner denne andelen med $X$. Det kan vises at det er rimelig å oppfatte $X$ som en stokastisk variabel med sannsynlighetstetthet

$$f(x) = \begin{cases} \frac{1}{\theta} & \text{for } 0 \leq x \leq \theta, \\ 0 & \text{ellers,} \end{cases}$$

der $\theta$ er en parameter som avhenger av kulas radius og type overflate.

**a)** Bestem kumulativ fordelingsfunksjon for $X$, F(x).

Skisser $f(x)$ og $F(x)$.

For $\theta = 2.0$, finn sannsynligheten $\mathrm{P}(X \leq 0.4)$.

Anta at $\theta$ for en bestemt kule er ukjent. For å skaffe informasjon om $\theta$ for denne kula gjøres $n$ målinger $X_1, X_2, \ldots, X_n$, der $X_1, X_2, \ldots, X_n$ er uavhengige stokastiske variable, alle med sannsynlighetsfordeling som gitt over.

**b)** Vis at sannsynlighetsmaksimeringsestimatoren (SME), $\widehat{\theta}$, for $\theta$ er

$$\widehat{\theta} = \max(X_1, X_2, \ldots, X_n).$$

Videre i oppgaven innfører vi notasjonen $Y = \max(X_1, X_2, \ldots, X_n)$ slik at $\widehat{\theta} = Y$.

**c)** Bestem kumulativ fordelingsfunksjon for $Y$, $G(y) = \mathrm{P}(Y \leq y)$.

Benytt så dette til å vise at sannsynlighetstettheten for $Y$ er

$$g(y) = \begin{cases} \frac{ny^{n-1}}{\theta^n} & \text{for } 0 \leq y \leq \theta, \\ 0 & \text{ellers.} \end{cases}$$

**d)** Benytt sannsynlighetstettheten for $Y$ gitt i punkt **c)** til å vise at estimatoren $\widehat{\theta} = Y$ **ikke** er forventningsrett.

Bestem $k$ slik at $\widetilde{\theta} = kY$ blir en forventningsrett estimator for $\theta$.

**e**) Finn et 95% konfidensintervall for $\theta$ basert på $\widetilde{\theta}$. (Hint: Bestem fordelingen til $Z = \widetilde{\theta}/(k\theta)$ og ta utgangspunkt i $Z$ for å bestemme konfidensintervallet.)

Beregn også intervallet numerisk når $n = 10$ og den største målte verdien er 0.46.

## Fasit

**2**. **a**) 0.0067

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

**4**. **a**) 0.2 **d**) $k = (n+1)/n$ **e**) $[0.461, 0.665]$