TMA4240 Statistikk
Høst 2017

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Skriftlig innlevering 5 (blokk 2)

Dette er den nest siste skriftlige innleveringen. Den er basert på det som er diskutert i forelesningene frem til og med uke 44. Spesielt er det i denne innleveringen fokus på parameterestimering og konfidensintervall. Innleveringen inneholder dessuten en oppgave som omhandler hypotesetestingssituasjonen. Alle deloppgaver teller like mye.

Oppgave 1

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 $X_1, \ldots, X_n$ 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 en 95% intervallestimator for avstanden mellom pålene, a, basert på estimatorene foran.

Oppgave 2

Levetiden (målt i måneder), X, til en del typer elektroniske komponenter har vist seg å følge en fordeling med sannsynlighetstetthet

$$f(x) = \begin{cases} \frac{1}{2\theta}\frac{1}{\sqrt{x}}e^{-\frac{\sqrt{x}}{\theta}} & \text{for } x > 0 \\ 0 & \text{ellers,} \end{cases} \tag{2.1}$$

der θ er en parameter som beskriver kvaliteten til komponentene. Tilhørende kumulative fordelingsfunksjon er gitt ved

$$F(x) = \begin{cases} 1 - e^{-\frac{\sqrt{x}}{\theta}} & \text{for } x > 0 \\ 0 & \text{ellers.} \end{cases}$$

I hele denne oppgaven antar vi at levetider til ulike komponenter er uavhengige. Det oppgis dessuten at $\mathrm{E}[X] = 2\theta^2$.

For å undersøke kvaliteten, θ, til en ny type elektronisk komponent, har man undersøkt $n = 20$ komponenter. La $X_1, X_2, \ldots, X_n$ betegne de tilhørende levetider. De observerte levetider er

1.27	0.40	11.37	1.91	0.083	2.15	4.68	5.54	0.88	23.45
7.12	0.18	7.75	1.41	2.33	0.020	3.00	0.37	0.0088	0.029

Det oppgis at $\sum_{i=1}^{n} \sqrt{x_i} = 29.902$.

a) Vis at sannsynlighetsmaksimeringsestimatoren (SME) for θ er

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \sqrt{X_i}.$$

Er $\widehat{\theta}$ forventningsrett ?

La $Z_i = 2\sqrt{X_i}/\theta$ for $i = 1, 2, \ldots, n$.

b) Vis ved hjelp av transformasjonsformelen at Z_i er χ^2-fordelt med 2 frihetsgrader.

Bruk så dette til å begrunne at

$$\frac{2n\widehat{\theta}}{\theta}$$

er χ^2-fordelt med $2n$ frihetsgrader.

c) Benytt resultatet i punkt **b)** til å utlede et $(1 - \alpha) \cdot 100\%$ konfidensintervall for θ.

Regn også ut intervallet numerisk når $\alpha = 0.05$ og dataene er som gitt over.

Oppgave 3

I situasjoner der det er uklart hvem som er den biologiske faren til et barn kan farskapet avklares ved å sammenligne DNA-prøver fra barnet med mulige fedre. For en mulig far gjøres dette ved å sammenligne n ulike deler av DNA-strukturen til mannen med de samme n deler av DNA-strukturen hos barnet. De n undersøkte delene av DNA-strukturen antas uavhengige.

Hos et barn og en tilfeldig valgt mann (som ikke er biologisk far) er det for hver enkel del av DNA-strukturen som undersøkes en sannsynlighet $p = 0.15$ for at delen er sammenfallende hos barnet og mannen. Anta videre at en biologisk far alltid har alle de undersøkte delene av DNA-strukturen sammenfallende med barnets (dvs. vi ser bort fra mutasjoner o.l.), slik at hver undersøkte del av DNA-strukturen hos biologisk far og barn er sammenfallende med sannsynlighet $p = 1$.

La X være antall sammenfallende deler i DNA-strukturen hos et barn og en tilfeldig valgt mann (som ikke er biologisk far).

a) Begrunn at X er binomisk fordelt med parametre n og $p = 0.15$.

Dersom $n = 5$, beregn sannsynlighetene $P(X = 2)$, $P(X \geq 2)$ og $P(X = 2|X \geq 2)$.

I en farsskapssak blir en mann erklært å være biologisk far dersom alle undersøkte deler av DNA-strukturen er sammenfallende hos mannen og barnet. Dette kan vi se på som en hypotesetest der vi tester

$$H_0: \quad p = 0.15 \text{ (ikke far)} \qquad \text{mot} \qquad H_1: \quad p = 1.0 \text{ (far)}$$

der H_0 forkastes (dvs. mannen erklæres som far til barnet) dersom $X = n$.

b) For $n = 5$, finn sannsynligheten for å begå type 1 feil i testen over.

For $n = 5$, finn sannsynligheten for å begå type 2 feil i testen over.

Hvor mange ulike deler, n, av DNA-strukturen må man minst sammenligne dersom man ønsker at sannsynligheten for feilaktig å erklære en mann som far skal være mindre enn 0.000001?

Fasit

2. **a**) $\widehat{\theta}$ er forventningsrett **c**) $[1.01, 2.45]$

3. **a**) $0.138, 0.165, 0.836$ **b**) $0.000076, 0, n = 8$