TMA4245 Statistikk
Vår 2016

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Anbefalte oppgaver 9, blokk II

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

Oppgave 1

Et problem med vindmøller til kraftproduksjon er at fugler kan kollidere med rotorbladene og dø. Som et prøveprosjekt monteres ei mølle (A) på et sted langs norskekysten, og det registreres kontinuerlig hvor mange fugler som blir funnet døde av kollisjonsskader rundt mølla. Erfaring fra Danmark med lignende møller tilsier at forventet antall døde fugler per uke er $\lambda = 1$. Anta at den tilfeldige variabelen Y, antall fugler som kolliderer og dør med mølle A per uke, er poissonfordelt.

a) Hvis vi antar like forhold i Norge som i Danmark og at etterfølgende uker er uavhengige, finn sannsynligheten for at det i løpet av de første 5 uker skal kollidere mer enn 10 fugler med vindmølle A.

Gitt at det kolliderer mindre enn 5 fugler, hva er sannsynligheten for at ingen fugler kolliderte?

b) I løpet av fire år blir det funnet 261 fugler rundt den norske mølla. Estimer parameteren λ med sannsynlighetsmaksimering (maximum likelihood) fra disse data.

Ei anna mølle (B) ble plassert litt lenger inn i landet, på en plass med mindre fugletetthet. Anta at X, antall fugler som kolliderer og dør med mølle B per uke, er poissonfordelt med parameter ν og uavhengig av Y. La så $Z = X + Y$ være summen av døde fugler ved de to møllene

c) Finn den momentgenererende funksjonen $M_Z(t)$ til den tilfeldige variabelen Z.

Bruk den momentgenererende funskjonen til å si hvilken fordeling Z har.

Oppgave 2

Turid skal pusse opp kjøkkenet og vurderer å leie inn snekkere til å gjøre jobben. Det er mye å gjøre i byggebransjen, og Turid vil undersøke hvor lang ventetid hun må regne med før hun kan få satt i gang arbeidet. Hun har fått anbefalt byggmestrene A&B av venner som har gjennomført tilsvarende prosjekter.

Anta at ventetiden X, i antall uker, hos A&B er en stokastisk variabel med kumulativ fordeling

$$F(x) = \begin{cases} 1 - \exp(-\alpha x^2) & x \geq 0, \\ 0 & x < 0. \end{cases}$$

a) Anta at $\alpha = 0.04$.

Hva er sannsynligheten for at Turid må vente lenger enn 2 uker?

Dersom Turid må får beskjed om at ventetiden er minst 2 uker, hva er da sannsynligheten for at hun må vente i minst 5 uker?

Vis at sannsynlighetstettheten til X (for generell α) er

$$f(x) = 2\alpha x \exp(-\alpha x^2) \quad \text{for} \quad x \geq 0.$$

I resten av oppgaven skal vi anta at α er ukjent.

Før hun bestemmer seg for å engasjere A&B, ønsker Turid å estimere forventet ventetid, μ, hos A&B. Denne er gitt ved

$$\mu = \mathrm{E}(X) = \frac{\sqrt{\pi}}{2\sqrt{\alpha}}.$$

Turid innhenter informasjon om ventetidene for n tilfeldig valgte venner som har bruke A&B. La $X_1, \ldots, X_n$ representere ventetidene for de n vennene. Vi antar at $X_1, \ldots, X_n$ er uavhengige stokastiske variabler som alle har samme sannsynlighetstetthet $f(x)$.

En estimator for α basert på $X_1, \ldots, X_n$ er

$$\hat{\alpha} = \frac{n-1}{\sum_{i=1}^n X_i^2}$$

b) Undersøk om $\hat{\alpha}$ er sannsynlighetsmaksimeringsestimatoren (SME) for α (ved å finne SME).

Foreslå en estimator for forventet ventetid $\mu = \mathrm{E}(X)$ basert på $\hat{\alpha}$.

Hva blir estimatet for μ dersom $n = 6$ og de oppgitte ventetidene er 3, 4.5, 5, 7, 6.5 og 5 uker?

c) La $Y = X^2$. Vis at Y er eksponensialfordelt med forventning $1/\alpha$.

Bruk dette resultatet til å undersøke om $\hat{\alpha}$ er forventningsrett.

(Du kan bruke uten bevis at dersom $T_1, \ldots, T_n$ er uavhengige og eksponensialfordelte stokastiske variabler med lik forventning β, så er $E\left(\frac{1}{\sum_{i=1}^n T_i}\right) = \frac{1}{\beta(n-1)}$).

Oppgave 3

Anta vi har data $x_1, \ldots, x_n$ fra en normalfordeling med forventningsverdi μ og varians σ^2. Da vil gjennomsnittet $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i$ være en forventningsrett estimator av μ. I denne oppgaven skal vi simulere syntetiske normalfordelte data i Matlab og se om teorien holder.

1. Kjør koden *run_confds.m* i Matlab. Dette scriptet bruker funksjonen *confds.m*, som er en funksjon som beregner et 95% konfidensinterval for forventningsverdien μ i en normalfordeling. Begge Matlabfilene er tilgjengelige fra emnets hjemmeside.

2. Tolk hva koden gjør og det resulterende plottet som blir laget når du kjører koden. Prøv å kjøre koden for andre datastørrelser n, og tolk resultatet.

Tips: Den innebygde Matlabfunksjonen 'hist()' plotter et histogram av dataene. For hjelp med f.eks. denne funksjonen, skriv 'help hist' i kommandovinduet, eller gå til ***Help*** $\rightarrow$ ***Product help*** *og søk etter funksjonen.*

Oppgave 4

En bestemt målemetode for bestemmelse av pH-verdien i en løsning gir måleresultater som antas å være uavhengige og normalfordelte, med forventningsverdi μ lik virkelig pH og varians $\sigma^2 = 0.060^2$. La $X_1, \ldots, X_n$ være uavhengige målinger av pH i en bestemt løsning.

a) Anta (bare i dette punktet) at den virkelige pH-verdien i en løsning er 6.8.

Hva er sannsynligheten for at en bestemt måling gir et resultat som er under 6.74?

Hva er sannsynligheten for at en bestemt måling gir et resultat som er mellom 6.74 og 6.86?

Hva er sannsynligheten for at en bestemt måling, X, gir et resultat som avviker mer enn 0.06 fra μ, dvs bestem $P(|X - \mu| > 0.06)$?

Du skal estimere pH i en løsning, og bruker gjennomsnittet av 5 uavhengige målinger som estimat. La Y være gjennomsnittet av 5 uavhengige målinger.

b) Hva er sannsynligheten for at Y avviker mer enn 0.06 fra μ?

Utled et 95% konfidensintervall for μ. Hva blir konfidensintervallet når gjennomsnittet av fem uavhengige målinger ble 6.76?

Fasit

1. **a**) 0.014,0.015 **b**) 1.25

4. **a**) $0.159, 0.682, 0.318$ **b**) $0.026, [6.707, 6.813]$