TMA4240 Statistikk
Høst 2018

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Innlevering 4

Dette er første av tre innleveringer i blokk 2. Denne øvingen skal oppsummere pensum forelest frem til og med uke 42. Spesielt er det i denne innleveringen fokus på funksjoner av stokastiske variabler og parmeterestimering. For å få godkjent innleveringen kreves det at minimum 40% av svarene er riktige, og at man har gjort et ordentlig forsøk på å løse alle oppgavene. Alle deloppgaver teller like mye.

Oppgave 1

La X være gammafordelt med sannsynlighetstetthet f gitt ved

$$f(x) = \frac{\theta^4}{6} x^3 e^{-\theta x}, \;\; x > 0.$$

Skriv opp uttrykket for den momentgenererende funksjonen (MGF) til en tilfeldig variabel. Vis at MGF til X er gitt ved

$$M_X(t) = \left(1 - \frac{t}{\theta}\right)^{-4}, \;\; t < \theta.$$

(HINT: Integranden kan omskrives til en kjent form.)

Finn forventningen til X ved bruk av M_X.

Oppgave 2

Anta at den stokastiske variabelen X har sannsynlighetstetthet

$$f(x;\theta) = c\exp\{-(x-\theta)\}\,, \;\; x \geq \theta,$$

mens $f(x;\theta) = 0$ for $x < \theta$. θ betegner en konstant parameter og c er en positiv konstant.

a) Bestem konstanten c, og finn sannsynligheten for at $X > \theta + 1$.

b) La $Y = 2X + 3$. Bestem sannsynlighetstettheten for den stokastiske variabelen Y. Regn ut sannsynligheten for at $Y > \theta + 1$.

c) La $X_1, X_2, \ldots, X_{10}$ betegne 10 uavhengige stokastiske variabler som hver har samme sannsynlighetstetthet som X. Bestem sannsynlighetstettheten for den stokastiske variabelen $W = \min\{X_1, X_2, \ldots, X_{10}\}$. Regn ut sannsynligheten for at $W > \theta + 1$.

Oppgave 3

Man skal undersøke pH-verdien, μ, i et ferskvann. Til dette formål benyttes to ulike målemetoder, metode A og metode B. La X betegne målt verdi ved metode A og la Y tilsvarende betegne målt verdi med metode B. Anta at X og Y er uavhengige og at begge er normalfordelt med forventning μ. Målenøyaktigheten for de to målemetodene er forskjellige. Anta at variansen til X er $\sigma_A^2 = 0.2^2$, mens variansen til Y er $\sigma_B^2 = 0.1^2$. For å estimere μ er det foreslått tre estimatorer

$$\widehat{\mu} = Y \quad , \quad \widetilde{\mu} = \frac{1}{2}X + \frac{1}{2}Y \quad \text{og} \quad \mu^* = \frac{1}{5}X + \frac{4}{5}Y.$$

Hvilke to egenskaper kjennetegner en god estimator?

Hvilken av estimatorene $\widehat{\mu}$, $\widetilde{\mu}$ og μ^* vil du foretrekke? Begrunn svaret.

Oppgave 4

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 5

Kjell Inge drar ofte og fisker en times tid i nærheten av der han bor.

a) Han mener at vekten på en fisk er normalfordelt med forventningsverdi 800 gram og standardavvik 100 gram. Anta at Kjell Inge har rett i det.

Hva er sannsynligheten for at en fisk veier mer enn 1000 gram?

Hva er sannsynligheten for at en fisk veier mellom 500 gram og 1000 gram?

Kjell Inge tenker at antallet fisk er viktigere enn vekten. Han grubler på sannsynlighetsfordelingen til antall fisk per tur.

b) La X være antall fisk på en tur. Alle turer tar 1 time. Vi antar at X er Poisson-fordelt med parameter (forventningsverdi) $\mu = 3$.

Hva er sannsynligheten for at Kjell Inge ikke får fisk på en tur?

Gitt at han får fisk, hva er sannsynligheten for at han får flere enn 3 fisk?

En alternativ sannsynlighetsmodell er som følger: Med sannsynlighet θ får han helt sikkert 0 fisk. Med sannsynlighet $1 - \theta$ er antall fisk Poisson-fordelt med parameter μ. Punktsannsynligheten til antall fisk er da:

$$P(X = x) = \theta I(X = 0) + (1 - \theta)\frac{\mu^x}{x!}e^{-\mu}, \quad x = 0, 1, \ldots$$

der $I(A) = 1$ dersom hendelsen A inntreffer, og $I(A) = 0$ ellers.

c) Anta at $\theta = 0.5$ og $\mu = 4$.

Hva er sannsynligheten for at Kjell Inge får fisk på en tur?

Bruk sannsynlighetsmodellen til å regne ut forventet antall fisk per tur.

Vi antar nå at vi har data fra $n = 20$ uavhengige fisketurer. Av disse endte $r = 8$ med ingen fisk. I de resterende tolv turene ble det tilsammen 40 fisk. Vi bruker dette tallmaterialet til å estimere modellparametrene θ og μ.

d) Sett opp rimelighetsfunksjonen (likelihood-funksjonen) for μ og θ.

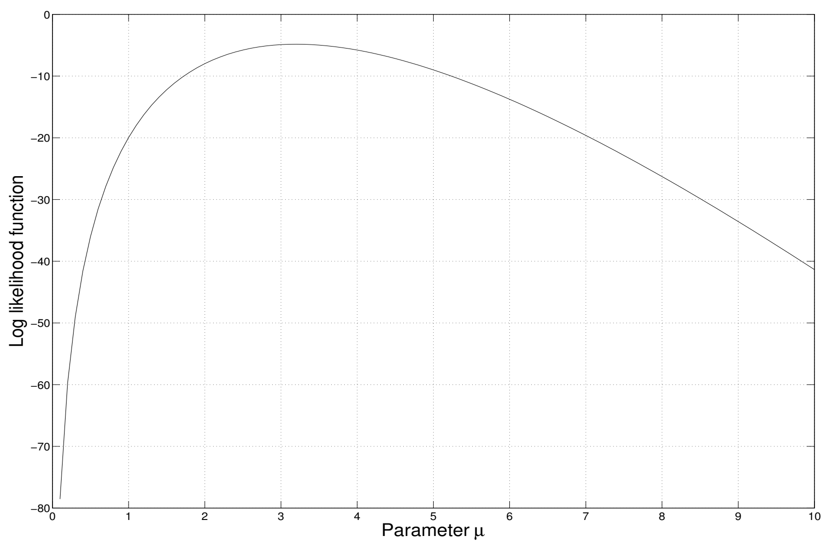
Anta kjent μ. Vis at sannsynlighetsmaksimeringsestimatoren (maximum likelihood estimator) for θ er gitt som

$$\hat{\theta} = \hat{\theta}(\mu) = \frac{r - ne^{-\mu}}{n(1 - e^{-\mu})}$$

Ved å sette $\hat{\theta}(\mu)$ inn i rimelighetsfunksjonen, kan vi studere rimelighetsfunksjonen kun som en funksjon av μ. Figuren nedenfor viser denne rimelighetsfunksjonen.

Bruk plottet til å finne sannsynlighetsmaksimeringsestimatet for μ.

Bruk resultatet til å regne ut sannsynlighetsmaksimeringsestimatet for θ.



Oppgave 6

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.

Ledelsen i ingeniørfirmaet synes anslaget på avstanden a er for upresist og leier derfor inn mer moderne måleutstyr. Det er kjent at standardavviket for dette moderne utstyret, σ_M, er en firedel av standardavviket for det gamle utstyret. En tar m uavhengige målinger $Y_1, \ldots, Y_m$ med det nye utstyret og antar at hver måling er normalfordelt med forventning a og varians σ_M^2. En har altså at σ_M er ukjent, men at $\sigma_M = \sigma_G/4$.

b) Utled sannsynlighetsmaksimeringsestimatorene (*maximum likelihood estimators*) for avstanden a og variansen σ_M^2, basert på den samlede måleserien $X_1, \ldots, X_n, Y_1, \ldots, Y_m$.

Fasit

1. $\mathrm{E}[X] = 4/\theta$

2. **a**) $1, 0.368$ **b**) 1 **c**) e^{-10}

4. **a**) 0.8521, 0.4317 **b**) 5.2 uker

5. **a**) 0.023,0.976 **b**) 0.05,0.37 **c**) 0.49,2 **d**) 0.37