Obligatorisk øving for STK1110, Høsten 2021

Øving 1 av 2

Innleveringsfrist

Torsdag 16. september 2020, klokken 14:30 i Canvas (`canvas.uio.no`).

Instruksjoner

Du velger selv om du skriver besvarelsen for hånd og scanner besvarelsen eller om du skriver løsningen direkte inn på datamaskin (for eksempel ved bruk av Latex). Besvarelsen skal leveres som **én PDF-fil**. Scannede ark må være godt lesbare. Besvarelsen skal inneholde navn, emne og oblignummer.

Det forventes at man har en klar og ryddig besvarelse med tydelige begrunnelser. Husk å inkludere alle relevante plott og figurer. Studenter som ikke får sin opprinnelige besvarelse godkjent, men som har gjort et reelt forsøk på å løse oppgavene, vil få én mulighet til å levere en revidert besvarelse. Samarbeid og alle slags hjelpemidler er tillatt, men den innleverte besvarelsen skal være skrevet av deg og reflektere din forståelse av stoffet. Er vi i tvil om du virkelig har forstått det du har levert inn, kan vi be deg om en muntlig redegjørelse. I oppgaver der du blir bedt om å programmere må du legge ved programkoden og levere den sammen med resten av besvarelsen.

Søknad om utsettelse av innleveringsfrist

Hvis du blir syk eller av andre grunner trenger å søke om utsettelse av innleveringsfristen, må du ta kontakt med studieadministrasjonen ved Matematisk institutt (e-post: studieinfo@math.uio.no) i god tid før innleveringsfristen. For å få adgang til avsluttende eksamen i dette emnet, må man bestå alle obligatoriske oppgaver i ett og samme semester.

Spesielt om dette oppgavesettet

Du skal bruke programpakken R til å gjøre beregninger i oppgavene, og du må angi hvilke kommandoer du har brukt for å komme fram til svarene dine. For å få godkjent besvarelsen, må du ha minst 65% riktig på hver av de to oppgavene (Oppgave 1 og 2).

For fullstendige retningslinjer for innlevering av obligatoriske oppgaver, se her:

`www.uio.no/studier/admin/obligatoriske-aktiviteter/mn-math-oblig.html`

LYKKE TIL!

Oppgave 0

Denne oppgaven teller ikke med, men faglærer spør av ren nysgjerrighet, så svar gjerne :-)

a) Hvorfor tar du STK1110?

b) Hva forventer du av STK1110?

Oppgave 1

Fila `https://www.uio.no/studier/emner/matnat/math/STK1110/data/forsikringskrav.txt` inneholder 6377 bilforsikringskrav til et norsk forsikringsselskap et gitt år. Øvre rad av Figur 1 nedenfor viser et histogram og et kvantilplott av disse dataene. Vi ser at de åpenbart ikke er normalfordelt. Raden under viser samme plott av de log-transformerte dataene. For disse ser normalfordelingsantakelsen ut til å være grei.

Anta nå at $Y_1, \dots, Y_n \overset{uif}{\sim} N(\mu, \sigma^2)$, og at $X_i = e^{Y_i}$, $i = 1, \dots, n$. Da er $X_1, \dots, X_n$ uavhengige, identisk log-normalfordelte variabler med parametere (μ, σ^2), $log - N(\mu, \sigma^2)$. Vi antar heretter at forsikringskravene er observasjoner fra en log-normal fordeling.

a) Vis at $\mathrm{E}(X_i) = e^{\mu + \frac{1}{2}\sigma^2}$ og at $\mathrm{E}(X_i^2) = e^{2\mu + 2\sigma^2}$ (*Hint: du kan få bruk for momentgenererende funksjon for normalfordelingen* $N(\mu, \sigma^2)$, *som er* $M_Y(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$).

b) Finn momentestimatorene $\hat{\mu}_{mom}$ og $\widehat{\sigma^2}_{mom}$ for μ og σ^2, og beregn de tilsvarende estimatene for bilforsikringskravene.

c) Finn maksimum likelihood-estimatorene $\hat{\mu}_{mle}$ og $\widehat{\sigma^2}_{mle}$ for μ og σ^2 på vanlig måte, ved å

   1. spesifisere likelihood-funksjonen
   2. ta logaritmen til denne og derivere den
   3. sette den deriverte lik 0 og løse med hensyn på μ og σ^2. Beregn de tilsvarende estimatene for bilforsikringskravene, og sammenlign med resultatene fra b).

d) Husk at $Y_i = \log(X_i) \sim N(\mu, \sigma^2)$, $i = 1, \dots, n$. Bruk dette, samt maksimum likelihood-estimatorene for parameterne i normalfordelingen (disse behøver du ikke å utlede) til å finne $\hat{\mu}_{mle}$ og $\widehat{\sigma^2}_{mle}$ på en alternativ (og mye enklere) måte.

e) Vis at informasjonsmatrisa for én observasjon er gitt ved

$$I(\mu, \sigma) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

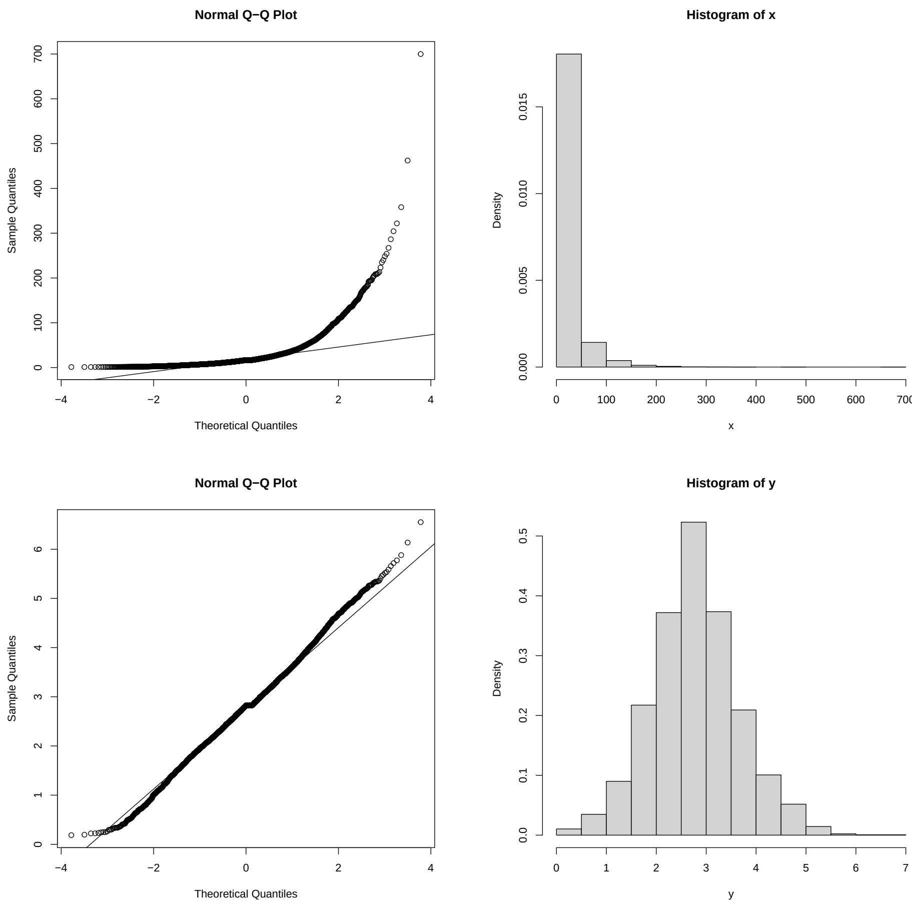

Figure 1: Øvre rad viser et kvantilplott (til venstre) og et histogram (til høyre) av forsikringskravene $x_1, \ldots, x_n$. Nedre rad viser et kvantilplott (til venstre) og et histogram (til høyre) av de loggede forsikringskravene $y_1, \ldots, y_n$, med $y_i = \log(x_i)$.

og bruk denne til å finne et estimat for standardfeilen til $\hat{\mu}_{mle}$ og $\widehat{\sigma^2}_{mle}$.

f) Estimér nå standardfeilen til $\hat{\mu}_{mle}$ og $\widehat{\sigma^2}_{mle}$ for foriskringsdataene ved hjelp av ikke-parametrisk bootstrapping. Sammenlign med estimatene du fikk i e) og kommentér.

g) Fra a) vet vi at $\phi = \mathrm{E}(X_i) = e^{\mu+\frac{1}{2}\sigma^2} = \phi(\mu, \sigma^2)$. Hva er maksimum likelihood-estimatoren $\hat{\phi}_{mle}$ for ϕ? Beregn det estimatet for bilforsikringskravene. En alternativ estimator for ϕ er den forventningsrette $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Beregn estimatet $\bar{x}$ for forsikringsdataene og sammenlign med maksimum likelihood-estimatet.

Oppgave 2

Anta at $X_1, \ldots, X_n$ er uavhengige og uniformt fordelt på intervallet $[0, \theta]$, dvs. at de har tetthet

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

Her er θ en ukjent parameter som vi ønsker å estimere.

a) Bestem $\mathrm{E}(X_i)$ og $\mathrm{V}(X_i)$.

b) Vis at momentestimatoren for θ er $\hat{\theta}_{mom} = 2\bar{X}$. Er estimatoren forventningsrett?

c) Bestem standardfeilen til $\hat{\theta}_{mom}$.

d) Vis at likelihooden er gitt ved

$$f(x_1, \ldots, x_n; \theta) = \begin{cases} \frac{1}{\theta^n}, & 0 \leq x_1, \ldots, x_n \leq \theta \\ 0, & \text{ellers.} \end{cases}$$

og forklar at maksimum likelihood-estimatoren for θ er $\hat{\theta}_{mle} = \max_i(X_i)$ (*Hint: her nytter det ikke å derivere log-likelihood-funksjonen, du må se direkte på likelihood-funksjonen for å finne maksimum*).

e) Det kan vises at forventningen og andremomentet til $\hat{\theta}_{mle}$ er henholdsvis

$$\mathrm{E}(\hat{\theta}_{mle}) = \frac{n\theta}{n+1} \text{ og } \mathrm{E}(\hat{\theta}^2_{mle}) = \frac{n\theta^2}{n+2},$$

hvilket betyr at $\hat{\theta}_{mle}$ ikke er forventningsrett for θ. Bruk resultatene over til å vise at estimatoren $\tilde{\theta} = \frac{n+1}{n}\hat{\theta}_{mle}$ er forventningsrett og til å finne standardfeilen til $\tilde{\theta}$.

f) Sammenlign momentestimatoren $\hat{\theta}_{mom}$ fra b) med $\tilde{\theta}$. Hvilken estimator velger du? Svaret skal begrunnes.

g) Sjekk hvordan teorien virker i praksis ved å simulere $n = 20$ observasjoner fra uniform fordeling på $[0, 1]$ (du kan bruke `data = runif(20)` i R). I dette tilfellet er den sanne verdien for θ lik 1. Hvor nær kommer du med de to estimatorene du sammenlignet i forrige punkt? Gjenta prosedyren over mange ganger (f.eks. 1000 til sammen) for å illustrere poenget i f), og presentér resultatene grafisk (f.eks. i et boksplott eller et histogram).