UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	STK1100 — Sannsynlighetsregning og statistisk modellering
Eksamensdag:	Onsdag 31. mai 2023
Tid for eksamen:	09.00 – 13.00

Oppgavesettet er på 4 sider.

Vedlegg:	Ingen
Tillatte hjelpemidler:	Godkjent kalkulator Formelsamling for STK1100

Kontroller at oppgavesettet er komplett før du begynner å besvare spørsmålene.

Oppgave 1

Heldigvis kan vi bruke antibiotika til å bekjempe bakterielle infeksjoner. Imidlertid har vi bare et begrenset antall antibiotikamedisiner, og bakterier utvikler seg til å bli resistente mot disse. I dette eksemplet skal vi bruke sannsynlighetsregning for å forstå utviklingen av antibiotikaresistens. Tenk deg at du har en stor mengde sykdomsbringende bakterier i tarmen, hvorav 10% har en mutasjon som gjør dem litt mer motstandsdyktige mot antibiotika. Du tar en antibiotika-kur. Sannsynligheten for at bakterier med mutasjonen overlever, er 20%. Sannsynligheten for at bakterier uten mutasjonen overlever, er 1%.

a

Hva er sannsynligheten for at en tilfeldig valgt bakterie i tarmen din overlever antibiotika-kuren?

b

Hva er sannsynligheten for at en overlevende bakterie har mutasjonen som gjør den litt mer motstandsdyktig mot antibiotika?

c

Dersom de gjenværende bakteriene får formere seg videre, og du tar en ny kur med samme antibiotika, hva blir svarene på a) og b) i denne nye runden?

Oppgave 2

La X og Y være to kontinuerlige stokastiske variabler med simultantetthet

(Fortsettes på side 2.)

$$f(x,y) = \begin{cases} \frac{\alpha}{\mu} y^{\alpha-1} e^{-\frac{x}{\mu} - y^{\alpha}}, & x > 0, y > 0 \\ 0, & ellers \end{cases},$$

der $\mu > 0$ og $\alpha > 0$ er parametere.

a

Vis at X har marginaltetthet

$$f_X(x) = \begin{cases} \frac{1}{\mu} e^{-\frac{x}{\mu}}, & x > 0 \\ 0, & ellers \end{cases}$$

ved å integrere ut Y fra simultantettheten, og finn også marginaltettheten til Y.

b

Er X og Y uavhengige? Begrunn svaret.

c

Vis at X har momentgenererende funksjon $M_X(t) = \frac{1}{1-t\mu}$ for $t < \frac{1}{\mu}$.

d

Bruk resultatet fra Oppgave c) til å vise at $\mathrm{E}(X) = \mu$ og $\mathrm{V}(X) = \mu^2$.

Oppgave 3

En Geigerteller brukes til å detektere radioaktive partikler i omgivelsene. Det er rimelig å anta at partiklene treffer Geigertelleren som en Poisson-prosess, og at tiden mellom to påfølgende partikler derfor følger en eksponensialfordeling. La $X_1, \ldots, X_n$ være n slike ventetider mellom to påfølgende partikler. Disse X_i-ene vil være uavhengige og altså eksponensialfordelte. Vi skal estimere *forventet ventetid* mellom to påfølgende partikler, som vi kaller for μ, fra tilhørende observasjoner $x_1, \ldots, x_n$.

a

Vis at tettheten til hver X_i kan skrives som

$$f_{X_i}(x) = \begin{cases} \frac{1}{\mu} e^{-\frac{x}{\mu}}, & x > 0 \\ 0, & ellers, \end{cases}$$

der μ er forventningen til X_i (merk at dette er marginaltettheten til X i Oppgave 2 b), og vis at log-likelihood-funksjonen blir

$$l(\mu) = -n \log \mu - \frac{1}{\mu} \sum_{i=1}^{n} x_i$$

(Fortsettes på side 3.)

I et eksperiment ble det detektert 30 ventetider mellom påfølgende partikler, og den samlede tiden dette tok, var 1200 sekunder.

b

Vis at Maximum Likelihood Estimator (MLE) for μ blir $\hat{\mu} = \bar{X}$. Beregn estimatet $\hat{\mu}$ basert på tallene over. Hva blir momentestimatoren for μ?

c

Bruk resultatene fra Oppgave 2 d) til å vise at $\mathrm{E}(\hat{\mu}) = \mu$ og $\mathrm{V}(\hat{\mu}) = \mu^2/n$. Begrunn (ved hjelp av sentralgrenseteoremet) at $\hat{\mu}$ er tilnærmet normalfordelt, dvs. $N(\mu, \mu^2/n)$, når n er stor. Finn et estimat for standardfeilen til $\hat{\mu}$ basert på tallene fra Oppgave b).

d

I dette tilfellet kan vi også finne den eksakte fordelingen til $\hat{\mu}$. Vis at $\hat{\mu} \sim gamma\left(n, \frac{\mu}{n}\right)$.
(*Hint*: Du kan få bruk for svaret fra Oppgave 2 c).

Oppgave 4

Vi ønsker nå å lage et konfidensintervall for forventningen μ til levetidene $X_1, \ldots, X_n$ fra Oppgave 3.

a

La $Y_i = \frac{2}{\mu} X_i$. Vis at Y_i er $gamma(1,2)$-fordelt, dvs. at Y_i har tetthet

$$f_Y(y) = \begin{cases} \frac{1}{2} e^{-\frac{y}{2}}, & y > 0 \\ 0, & ellers, \end{cases}.$$

For å lage et konfidensintervall for μ vil vi først ta utgangspunkt i maksimum likelihood-estimatoren $\hat{\mu} = \bar{X}$ fra Oppgave 3 b).

b

Vis at en kan skrive $\hat{\mu} = \frac{\mu}{2n} \sum_{i=1}^{n} Y_i$, og bruk det til å begrunne at $\frac{2n\hat{\mu}}{\mu}$ er $gamma(n, 2)$-fordelt, som er det samme som kjikvadratfordelingen med $2n$ frihetsgrader.
(*Hint*: Husk at en sum av uavhengige $gamma(\alpha_i, \beta)$-fordelte variabler er $gamma(\sum_{i=1}^{n} \alpha_i, \beta)$-fordelt).

(Fortsettes på side 4.)

La $\chi^2_{2n,0.975}$ og $\chi^2_{2n,0.025}$ være henholdvis $\alpha/2 \cdot 100\%$- og $(1-\alpha/2) \cdot 100\%$-persentilen i kjikvadratfordelingen med $2n$ frihetsgrader.

c

Utled et 95% konfidensintervall for μ basert på resultatene over. Regn ut intervallet når du får vite at $\chi^2_{60,0.975} = 40.48$, $\chi^2_{60,0.025} = 83.30$ og $\hat{\mu}$ er som i Oppgave 3 b).

Anta nå at antallet observasjoner n er så stort at vi kan tilnærme med normalfordelingen.

d

Bruk tilnærming til normalfordeling til å utlede et tilnærmet 95% konfidensintervall for forventningen μ. Du kan bruke at $z_{0.025} = 1.96$.