Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Faglig kontakt under eksamen:
Håkon Tjelmeland 73 59 35 20

EKSAMEN I FAG 75510/75515 STATISTIKK 1

Onsdag 5. august 1998
Tid: 09:00–14:00

Hjelpemidler:
Godkjent lommekalkulator, utdelt ordliste, Statistiske tabeller og formler (Tapir forlag).

Oppgave 1

La X være en diskret fordelt stokastisk variable med punktsannsynlighet gitt i følgende tabell

x	-2	-1	0	1	2
$f(x)$	0.1	0.1	0.5	0.2	0.1

Regn ut forventningsverdien til X.

Bestem sannsynlighetene

$$\mathrm{P}(X \geq 0) \qquad \text{og} \qquad \mathrm{P}(X \geq 0 | X \leq 1).$$

Oppgave 2 Kabelproduksjon

En fabrikk produserer kabel og tid om annet oppstår det feil på den produserte kabelen. La Z betegne lengden (i kilometer) på kabelen mellom to etterfølgende feil. Vi skal anta at feilene oppstår uavhengig av hverandre, dvs. at påfølgende observasjoner av Z langs kabelen, $Z_1, Z_2, Z_3, \ldots$, er uavhengige stokastiske variable.

Av erfaring vet en at lengden mellom to etterfølgende feil er eksponensialfordelt med parameter λ, dvs. Z har sannsynlighetstetthet

$$f(z;\lambda) = \begin{cases} \lambda e^{-\lambda z} & \text{for } z > 0, \\ 0 & \text{ellers} \end{cases}$$

og kumulativ fordelingsfunksjon

$$F(z;\lambda) = \begin{cases} 1 - e^{-\lambda z} & \text{for } z > 0, \\ 0 & \text{ellers.} \end{cases}$$

a) Anta i dette punktet at $\lambda = 0.05$.

Hva er sannsynligheten for at lengden mellom to etterfølgende feil er mer enn 10 kilometer?

Dersom man har observert at de første 10 kilometrene er feilfrie, hva er da sannsynligheten for at også de neste 10 kilometrene er feilfrie?

I resten av oppgaven skal vi anta at λ er en ukjent parameter. Ved hjelp av fabrikkens opptegnelser over tidligere feil på kabelen ønsker vi å estimere λ. Men det viser seg dessverre at fabrikken ikke har notert nøyaktig lengde på kabelen mellom hver feil, i stedet er det kun notert antall hele kilometer, M, med kabel mellom hver feil. Dvs, dersom $Z < 1.0$ har fabrikken notert seg $M = 0$, dersom $1.0 \leq Z < 2.0$ har fabrikken notert seg $M = 1$, dersom $2.0 \leq Z < 3.0$ har fabrikken notert seg $M = 2$, osv.

b) Vis at punktsannsynligheten for M blir

$$\mathrm{P}(M = m) = (1 - e^{-\lambda})e^{-\lambda m} \quad \text{for} \quad m = 0, 1, 2, \ldots.$$

c) Finn sannsynlighetsmaksimeringsestimatoren (SME) for λ basert på n observasjoner $M_1, M_2, \ldots, M_n$.

Oppgave 3 Bensinforbruk

Kari har nylig kjøpt seg en ny bil. Nå ønsker hun å undersøke bilens bensinforbruk ved landeveiskjøring. La x være lengden av en tur (i mil) og Y tilhørende bensinforbruk (i liter). Kari forutsetter at hun selv velger lengden på turene og betrakter derfor ikke x som en stokastisk variabel, mens hun antar at Y er en normalfordelt stokastisk variabel med

$$\mathrm{E}(Y) = \beta x \quad \text{og} \quad \mathrm{Var}(Y) = x\sigma^2.$$

Dessuten antar hun at bensinforbruk på forskjellig turer er uavhengige stokastiske variable. Du skal i hele oppgaven forutsette det kjent at $\sigma^2 = 0.1^2$.

a) Hvilken tolkning har parameteren β i modellen?

Som et alternativ til antagelsen om $\mathrm{E}(Y)$ gitt over, kunne en satt $\mathrm{E}(Y) = \alpha + \beta x$. Hvorfor er det, slik Kari har gjort, mest rimelig å velge $\alpha = 0$.

Hvorfor er det rimelig å anta at variansen til Y er proporsjonal med x?

b) Anta i dette punktet at $\beta = 0.75$.

Hva er sannsynligheten for at Kari på en 5 mil lang tur vil bruke mer enn 4 liter bensin?

Betrakt to kjøreturer på henholdsvis $x_1 = 5$ og $x_2 = 10$ mil. Hva er sannsynligheten for at totalt bensinforbruk på de to turene er mindre enn 12 liter?

Betrakt igjen to kjøreturer på henholdsvis $x_1 = 5$ og $x_2 = 10$ mil og la Y_1 og Y_2 være tilhørende bensinforbruk på de to turene. Hva er sannsynligheten for at bensinforbruket på turen på 10 mil er mer enn dobbelt så stor som bensinforbruket på turen på 5 mil? (dvs. finn $\mathrm{P}(Y_2 - 2Y_1 > 0)$)

For å undersøke bilens bensinforbruk, kjører Kari $n = 6$ turer av forskjellig lengde og måler bensinforbruket for hver tur. Målingene hennes gir følgende resultat

Lengde (mil)	5	10	20	50	100	150
Bensinforbruk (liter)	2.73	5.97	11.64	30.20	59.16	85.92

For å estimere β, betrakter Kari to estimatorer,

$$\widehat{\beta} = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i} \qquad \text{og} \qquad \widetilde{\beta} = \frac{1}{n}\sum_{i=1}^n \frac{Y_i}{x_i},$$

der x_i og Y_i er henholdsvis lengde av og bensinforbruk på tur nr i.

c) Vis at

$$\mathrm{E}(\widehat{\beta}) = \beta \qquad \text{og} \qquad \mathrm{Var}(\widehat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^n x_i}$$

Finn også forventningsverdi og varians for estimatoren $\widetilde{\beta}$.

Hvilken av de to estimatorene vil du foretrekke? (Begrunn svaret)

Selgeren som solgte bilen til Kari opplyste at bilens bensinforbruk ved landeveiskjøring var 0.56 liter/mil. Kari ønsker å benytte sine observasjoner til å sjekke om det er grunnlag for å påstå at bilens bensinforbruk er høyere enn hva selgeren opplyste.

d) Formuler dette som et hypotesetestingsproblem. Ta utgangspunkt i estimatoren $\widehat{\beta}$ og lag en test med signifikansnivå 5%. Hva blir konklusjonen på testen med observasjonene gitt over?

e) Ta utgangspunkt i estimatoren $\widehat{\beta}$ og utled et 95%-konfidensintervall for β. Hva blir intervallet med observasjoner som gitt over?

Oppgave 4 Levetid av elektronisk komponent

Anta at levetiden til en bestemt type elektroniske komponenter er eksponensialfordelt med forventningsverdi lik $1/\lambda$. Det finnes mange produsenter av denne typen elektroniske komponenter og kvaliteten på produktet varierer fra produsent til produsent. Dvs. de forskjellige produsentene har forskjellig parameterverdi λ og verdien på λ beskriver dermed gjennomgående kvalitet på komponenter fra den enkelte produsent. Anta videre at dersom en tilfeldig velger en produsent så kan en betrakte tilhørende λ som en kontinuerlig fordelt stokastisk variabel som er eksponensialfordelt med forventningsverdi $1/\theta$.

Anta at en kunde, som skal kjøpe en elektronisk komponent, går frem på følgende måte. Først velger han tilfeldig en produsent og deretter går han og kjøper en komponent produsert av denne produsenten. La T betegne levetiden for den komponenten kunden kjøper. Finn sannsynlighetsfordelingen for T.