



Faglig kontakt under eksamen:
Mette Langaas 98847649

EKSAMEN I FAG TMA4240/TMA4245 STATISTIKK

10. august 2005

Tid: 09:00–13:00

Tillatte hjelpemidler:

Gult A5-ark med egne håndskrevne notater.

Tabeller og formel i statistikk (Tapir Forlag).

K. Rottmann: *Matematisk formelsamling*.

Kalkulator: HP30S.

BOKMÅL

Sensur: 1. september 2005.

Oppgave 1 Smeltepunktsbestemmelse

En metallurg har vært med på å utvikle en ny legering, og skal presentere ulike egenskaper ved legeringen til sine kolleger. Vi skal her se på bestemmelse av smeltepunktet til legeringen.

På grunn av mindre variasjoner i sammensetningen av legeringen, målefeil og lignende, kan gjentatte målinger av smeltepunktet til legeringen antas å være realisasjoner av uavhengige og normalfordelte variabler med forventningsverdi μ og kjent standardavvik $\sigma = 2^\circ \text{C}$.

a) Anta først (kun i dette punktet) at forventningsverdien er $\mu = 1468^\circ \text{C}$.

Metallurgen tar en måling av smeltepunktet. Hva er sannsynligheten for at observert smeltepunkt er lavere enn 1467°C ?

Anta så at metallurgen tar åtte uavhengige målinger av smeltepunktet, X_1, X_2, \dots, X_8 . Hva er sannsynligheten for at gjennomsnittet av de åtte målingene av smeltepunktet, $\bar{X} = \frac{1}{8} \sum_{i=1}^8 X_i$, ligger mellom 1467°C og 1469°C ?

Anta i resten av oppgaven at forventningsverdien til legeringens smeltepunkt, μ , er ukjent, men at standardavviket er kjent og lik $\sigma = 2^\circ \text{C}$.

- b) Utled et 90% konfidensintervall for μ basert på n uavhengige målinger av smeltepunktet, X_1, X_2, \dots, X_n .

Hvor stor må n minst være for at lengden på intervallet ikke overstiger 3°C ? Kall dette antallet n_0 og bruk de n_0 første observasjonene gitt i tabell 1 til å bestemme intervallet numerisk.

Metallurgen har kommet frem til at forventningsverdien til legeringens smeltepunkt er 1468°C .

En av kollegene, Professor Stål, er ikke enig i metallurgens analyser og mener at smeltepunktet for legeringen er høyere enn 1468°C .

- c) Formulér Professor Ståls utsagn i en hypotesetest ved å definere nullhypotese og alternativ hypotese.

Sett opp en testobservator og finn forkastingsområdet. Hva blir konklusjonen på testen, med data gitt i tabell 1, når signifikansnivået er $\alpha = 0.05$?

Hvor stor må n velges for at vi med en sannsynlighet på minst 0.99 skal konkludere med at $\mu > 1468^\circ \text{C}$, når i virkeligheten $\mu = 1470^\circ \text{C}$? Signifikansnivået skal fortsatt være $\alpha = 0.05$.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1467.4	1468.0	1471.6	1468.6	1468.8	1471.9	1469.4	1466.0

Tabell 1: Data over smeltepunkt i $^\circ \text{C}$. Det oppgis at gjennomsnittet av de 8 observasjonene er 1469.0°C .

Oppgave 2 Ventetid på snekkertjenester

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 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

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

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

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

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

En estimator for α basert på X_1, \dots, 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 $\mu = E(X)$ 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, \dots, 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 Ulykker

Hyppigheten av større ulykker ved oljeboring fra flytende plattformer i Nordsjøen skal undersøkes. La λ være en parameter som angir gjennomsnittlig antall ulykker pr. timeverk, der λ antas ukjent (men $\lambda > 0$). Vi ser på n plattformer, og lar X_i betegne antall store ulykker på plattform nr. i , $i = 1, \dots, n$.

Anta at X_i er Poisson-fordelt med $E(X_i) = a_i \lambda$, hvor a_i er totalt antall utførte timeverk på plattform nr. i . X_1, \dots, X_n antas å være uavhengige tilfeldige variabler.

- a) To estimatorer for λ basert på data fra de n plattformene er $\hat{\Lambda}$ og Λ^* , gitt ved

$$\begin{aligned}\hat{\Lambda} &= \frac{1}{n} \sum_{i=1}^n \frac{X_i}{a_i}, \text{ og} \\ \Lambda^* &= \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n a_i}.\end{aligned}$$

Finn forventingsverdien og variansen til hver av de to estimatorene $\hat{\Lambda}$ og Λ^* .

Beregn numeriske verdier for $\hat{\Lambda}$ og Λ^* når $n = 5$ og de innsamlede data er gitt i tabell 2.

Hvilken av de to estimatorene ville du foretrekke i denne situasjonen? Begrunn svaret.

Plattform nr. i	a_i	x_i
1	$1.5 \cdot 10^5$	1
2	$2.72 \cdot 10^5$	4
3	$0.87 \cdot 10^5$	0
4	$3.28 \cdot 10^5$	2
5	$0.63 \cdot 10^5$	1

Tabell 2: Data over antall ulykker og totalt antall utførte timeverk på fem plattformer. Følgende størrelser oppgis: $\sum_{i=1}^5 a_i = 9.0 \cdot 10^5$, $\sum_{i=1}^5 a_i^2 = 2.16 \cdot 10^{11}$, $\sum_{i=1}^5 \frac{1}{a_i} = 4.08 \cdot 10^{-5}$,

$$\sum_{i=1}^5 \frac{1}{a_i^2} = 4.51 \cdot 10^{-10}, \quad \sum_{i=1}^5 \frac{x_i}{a_i} = 4.33 \cdot 10^{-5}.$$

- b) Anta at data blir innsamlet bare fra plattformer der det har vært minst én stor ulykke. La Y betegne antall store ulykker på en slik plattform, og la a være antall utførte timeverk på plattformen. Vis at

$$P(Y = y) = \frac{(a\lambda)^y}{y!} \frac{e^{-a\lambda}}{1 - e^{-a\lambda}}, \quad y = 1, 2, \dots$$

Beregn $E[Y]$.