**Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag**

Faglige kontakter under eksamen:
Bo Lindqvist 73593532
Henning Omre 73593531
John Tyssedal 73593534

EKSAMEN I FAG TMA4240 STATISTIKK

Fredag 12. desember 2003

Tid: 09:00–14:00

Tillatte hjelpemidler:
Tabeller og formler i statistikk (Tapir Forlag)
K. Rottmann: Matematisk formelsamling
Kalkulator: HP30S.

BOKMÅL

Sensur: 12. januar 2004.

Oppgave 1 TOGFORSINKELSEN

I denne oppgaven kan du bruke uten å vise det at

$$\int_0^\infty x^r e^{-ax} dx = \frac{r!}{a^{r+1}} \text{ når } a > 0 \text{ og } r \text{ er et heltall} \geq 0$$

Vi betrakter ankomst- og oppholdstider for et bestemt lokaltog på en jernbanestasjon. Toget skal etter rutetabellen ankomme hver hverdag klokka 8:00, men kommer alltid etter dette tidspunktet.

La X (minutter) betegne togets forsinkelse på en tilfeldig valgt hverdag. Vi antar at X er en stokastisk variabel med sannsynlighetstetthet

$$g(x) = \begin{cases} kxe^{-2x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

der $k > 0$ er en konstant.

a) Vis at $k = 4$.

Hva er den forventede forsinkelse for toget?

Vis at sannsynligheten for at toget er mer enn 2 minutter forsinket er tilnærmet lik 0.09.

b) La V være antall ganger i løpet av en måned (= 22 hverdager) at toget er mer enn 2 minutter forsinket. Foreslå en sannsynlighetsfordeling for V og sett opp de forutsetninger som ligger til grunn for denne.

Hva er sannsynligheten for at toget er mer enn 2 minutter forsinket minst 2 ganger i løpet av en måned (= 22 hverdager)?

Hva er (tilnærmet) sannsynligheten for at toget er mer enn 2 minutter forsinket mer enn 30 ganger i løpet av 220 hverdager?

La Y (minutter) være den tiden toget står på stasjonen. Oppholdstiden Y vil være influert av forsinkelsen, og vi antar at den betingede sannsynlighetstetthet $f(y|x)$ for Y, gitt at forsinkelsen X er lik x (> 0), er gitt ved

$$f(y|x) = \begin{cases} (x/2)\, e^{-xy/2} & \text{for } y > 0 \\ 0 & \text{for } y \leq 0 \end{cases}$$

c) Hvilken fordeling har oppholdstiden Y når det er gitt at forsinkelsen er 2 minutter?

Hva er forventet oppholdstid når forsinkelsen er 2 minutter?

Sett opp simultantettheten $f(x, y)$ for X og Y.

Finn sannsynlighetstettheten $h(y)$ for oppholdstiden Y.

Oppgave 2 SØLVÅREN

Når prøver fra ett og samme sted i en sølvåre analyseres med hensyn på sølvinnholdet, fåes analyseresultater som vi skal anta er uavhengige og normalfordelte med forventning μ (g/tonn) og varians σ^2.

a) Anta i dette punktet at $\mu = 500$ og at $\sigma = 80$.

La Y betegne en slik måling. Hva blir sannsynligheten for at Y skal overskride 550?

La Y_1 og Y_2 være 2 slike målinger. Hvor stor er sannsynligheten for at disse skal avvike fra hverandre med minst 80 g/tonn.

Den nevnte sølvåren er 40 meter lang og rettlinjet og går fra vest mot øst. Det er av interesse å anslå hvor mye sølv som finnes i sølvåren. Erfaringer fra andre sølvårer av tilsvarende type tilsier at sølvinnholdet i store trekk endrer seg lineært fra den ene enden av sølvåren til den andre.

La Y_j betegne målt sølvinnhold i en prøve som er tatt x_j meter fra den vestlige enden, $j = 1, 2, \ldots, n$. Vi skal anta at $Y_1, \ldots, Y_n$ er uavhengige og normalfordelte med samme ukjente varians σ^2 og forventningsverdi

$$E(Y_j) = \alpha + \beta x_j$$

der α og β er ukjente konstanter. Minste kvadratsums-estimatorene for α og β er da gitt ved, henholdsvis (du skal ikke vise dette):

$$\begin{aligned} B &= \frac{\sum_{j=1}^{n}(x_j - \bar{x})Y_j}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \\ A &= \bar{Y} - B\bar{x} \end{aligned}$$

der $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$ og $\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$.

Av praktiske årsaker bestemmer en seg for å ta 5 prøver av sølvinnholdet i hver ende av sølv åren. La $Y_1, \ldots, Y_5$ betegne målt sølvinnhold i den vestre enden ($x_i = 0$) og $Y_6, \ldots, Y_{10}$ betegne målt sølvinnhold i den østre enden ($x_i = 40$).

b) Vis at da er

$$B = \frac{\sum_{j=6}^{10} Y_j - \sum_{j=1}^{5} Y_j}{200}$$

Hva blir det tilsvarende uttrykket for A?

Finn variansen til B uttrykt ved σ^2.

Resultatet av de 10 målingene i g/tonn er gitt nedenfor:

x_i	0	0	0	0	0	40	40	40	40	40
y_i	248	176	52	98	76	682	854	870	838	806

La $\bar{y}_V$ være gjennomsnittet av målingene i den vestre enden og $\bar{y}_E$ være gjennomsnittet av målingene i den østre enden. Til hjelp for videre utregninger får du opplyst at

$$\bar{y}_V = 130,\ \bar{y}_E = 810,\ \sum_{j=1}^{5}(y_j - \bar{y}_V)^2 = 26064,\ \sum_{j=6}^{10}(y_j - \bar{y}_E)^2 = 22720$$

c) Finn et estimat for σ^2 basert på disse dataene.

Fra erfaring med andre sølvårer med relativt lite sølv i den ene enden, blir det fra økonomisk hold uttalt at β nok må være større enn 12 for at sølvåren skal være lønnsom. Gir dataene grunnlag for å påstå at $\beta > 12$? Formuler spørsmålsstillingen som en hypotesetest og utfør testingen. Hva blir konklusjonen når signifikansnivået settes til 5%?

d) En av personene i ledelsen for firmaet som eier sølvåren, hevder at det hadde blitt et sikrere estimat for β om de 10 prøvene hadde blitt tatt med noenlunde jevne mellomrom langs sølvåren. Anta at dette hadde blitt gjort, og at en fremdeles hadde $\bar{x} = 20$. Ville variansen til estimatoren for β i den gitte modellen da blitt mindre? Begrunn svaret.

Personen insisterte på at det måtte tas en ekstra prøve midt i sølvåren, dvs. for $x = 20$. Resultatet av denne ble 600 g/tonn. Vurder om denne verdien er rimelig ut i fra modellen ved å se om den er inneholdt i et 95% prediksjonsintervall for en slik prøve. Du får opplyst at

$$Var(Y_0 - \hat{Y}_0) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{j=1}^{n} (x_j - \bar{x})^2} \right)$$

der Y_0 er en ny observasjon for sølvinnholdet i et punkt x_0, $\hat{Y}_0 = A + Bx_0$, og n er antall observasjonspunkter som estimeringen er basert på.

Oppgave 3 BROSPENNET

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.

b) En benytter følgende estimator for målevariansen σ_G^2:

$$S_G^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{a})^2$$

Utled en 95% intervallestimator for avstanden mellom pålene, a, basert på estimatorene foran.

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 $\sigma_M = \sigma_G/4$.

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