Norges teknisk–naturvitenskapelige universitet
Institutt for matematiske fag

Bokmål

Faglig kontakt under eksamen:
Håkon Tjelmeland 73 59 35 38

EKSAMEN I EMNE TMA4240 STATISTIKK

Onsdag 2. desember 2009
Tid: 09:00–13:00

Hjelpemidler: Kalkulator HP30S eller Citizen SR-270X med tomt minne.
Statistiske tabeller og formler, Tapir forlag.
K. Rottman: Matematisk formelsamling.
Ett gult ark (A5 med stempel) med egne håndskrevne formler og notater.

Sensur er ferdig: 23. desember 2009.

Oppgave 1

La X være en kontinuerlig fordelt stokastisk variabel med sannsynlighetstetthet

$$f(x) = \begin{cases} k\,(4x+1) & \text{for } x \in [0,1], \\ 0 & \text{ellers,} \end{cases}$$

der k er en konstant.

a) Regn ut hvilken verdi konstanten k må ha. Skisser $f(x)$.
Finn den kumulative fordelingsfunksjonen $F(x) = \mathrm{P}(X \leq x)$.
Finn sannsynlighetene

$$\mathrm{P}\left(X > \frac{1}{4}\right) \quad \text{og} \quad \mathrm{P}\left(X > \frac{1}{2} \,\middle|\, X > \frac{1}{4}\right).$$

La også Y være en kontinuerlig fordelt stokastisk variabel, og la betinget sannsynlighetstetthet for Y gitt $X = x$ være gitt ved

$$f(y|x) = \begin{cases} \frac{4x+2y}{4x+1} & \text{for } y \in [0,1], \\ 0 & \text{ellers.} \end{cases}$$

b) Finn simultan sannsynlighetstetthet for X og Y.

Finn marginal sannsynlighetstetthet for Y.

Finn sannsynligheten

$$\mathrm{P}\left(Y \leq X\right).$$

Oppgave 2 Telefonproblemer

En gruppe bestående av n studenter sitter en dag og diskuterer at de har store problemer med å komme gjennom på telefon til et bestemt firma. Firmaet har ikke et system med telefonkø slik at når man ringer firmaet får man enten opptattsignal eller man får kontakt med firmaets sentralbord. Flere av studentene klager på at de tilsynelatende alltid får opptattsignal. Studentene bestemmer seg for å gjøre en del forsøk for å få bedre oversikt over situasjonen. Hver av de n studentene skal flere ganger ringe til firmaet og notere opp om de får opptattsignal eller kontakt med sentralbordet. Hver student skal fortsette med å ringe inntil han/hun har fått snakket med sentralbordet k ganger og så rapportere tilbake til gruppen totalt antall ganger han/hun har ringt til firmaet.

Anta at vi nummererer studentene fra 1 til n og la X_i være antall telefonoppringninger student nummer i rapporterer tilbake til gruppen. La p betegne sannsynligheten for å få snakke med sentralbordet hvis man ringer en gang til firmaet.

Studentene, som ikke er så veldig gode i statistikk, konkluderer raskt med at $X_1, X_2, \ldots, X_n$ er et tilfeldig utvalg fra en negativt binomisk fordeling med parametre k og p, dvs.

$$f(x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}, \;\; x = k, k+1, \ldots.$$

a) Hvilke forutsetninger, i tillegg til de som er spesifisert over, må være oppfylt for at studentenes konklusjon om at $X_1, X_2, \ldots, X_n$ er et tilfeldig utvalg fra en negativt binomisk fordeling skal være korrekt?

Anta nå at disse forutsetningene er oppfylt, og dessuten at $k = 2$ og $p = 0.1$. Finn da sannsynlighetene

$$\mathrm{P}(X_1 = 2) \text{ og } \mathrm{P}(X_1 \leq 4 | X_1 > 2).$$

I resten av denne oppgaven skal vi anta at forutsetningene spesifisert i punkt **a**) er oppfylt slik at studentenes konklusjon er korrekt. Videre skal vi forutsette at k er et kjent tall, mens p er ukjent og skal estimeres.

b) Utled sannsynlighetsmaksimeringsestimatoren (SME) for p ut fra dataene $X_1, X_2, \ldots, X_n$ og vis at den kan skrives på formen

$$\widehat{p} = \frac{nk}{\sum_{i=1}^{n} X_i}.$$

En tid senere ser studentene en reportasje på fjernsyn der en representant for firmaet blir intervjuet angående nettopp den situasjonen studentene var opptatt av. Representanten for firmaet innrømmer at det kan være noe vanskelig å komme gjennom til dem på telefon, men hevder bestemt at $p \geq 0.1$. Studentene bestemmer seg for å teste om dataene de har samlet inn gir grunnlag for å hevde at firmaets påstand er feil.

c) Formuler studentenes problemstilling som et hypotesetestingsproblem. Dvs, velg nullhypotese og alternativ hypotese, velg testobservator, og bestem forkastningskriterium.

Regn ut (tilnærmet) p-verdi for testen når $k = 2$, $n = 50$ og studentenes observasjoner gav $\sum_{i=1}^{n} x_i = 779$. Vil du konkludere med at firmaets påstand er feil? (Grunngi svaret)

Oppgave 3 Smeltepunktbestemmelse

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.

Til å bestemme smeltepunktet til legeringen har metallurgen to målemetoder. Vi kaller disse henholdsvis målemetode A og målemetode B. På grunn av målefeil kan gjentatte målinger av smeltepunktet ved målemetode A antas å være realisasjoner av uavhengige og normalfordelte variabler med forventningsverdi μ og varians σ_A^2. Metallurgen ønsker å estimere forventningsverdien μ. Tilsvarende kan gjentatte målinger av smeltepunktet ved målemetode B antas å være realisasjoner av uavhengige og normalfordelte variabler med forventningsverdi μ og varians σ_B^2.

Vi skal i denne oppgaven anta at variansene σ_A^2 og σ_B^2 har kjente verdier, mens den felles forventningsverdien μ er ukjent og skal estimeres. Til dette formål gjør metallurgen n målinger med målemetode A og m målinger med målemetode B. La $X_1, X_2, \ldots, X_n$ betegne resultatene av målingene ved metode A og la $Y_1, Y_2, \ldots, Y_m$ betegne resultatene av målingene ved metode B. Vi skal også anta at $X_1, X_2, \ldots, X_n$ er uavhengig av $Y_1, Y_2, \ldots, Y_m$.

Som estimator for μ benytter metallurgen

$$\widehat{\mu} = a\bar{X} + b\bar{Y} = \frac{a}{n}\sum_{i=1}^{n} X_i + \frac{b}{m}\sum_{i=1}^{m} Y_i,$$

der a og b er to konstanter.

a) Benytt regneregler for forventningsverdi og varians til å vise at

$$\mathrm{E}\left(\widehat{\mu}\right) = (a+b)\mu \quad \text{og} \quad \mathrm{Var}\left(\widehat{\mu}\right) = \frac{a^2\sigma_A^2}{n} + \frac{b^2\sigma_B^2}{m}.$$

b) Bestem verdier for konstantene a og b slik at $\widehat{\mu}$ blir en best mulig estimator for μ.

Videre i oppgaven skal du fremdeles ta utgangspunkt i $\widehat{\mu} = a\bar{X} + b\bar{Y}$, dvs du skal **ikke** benytte de optimale verdiene for a og b du fant i punkt **b**).

c) Hvilken type sannsynlighetsfordeling har $\widehat{\mu}$? Begrunn svaret.

Utled et $(1-\alpha)\cdot 100\%$ konfidensintervall for μ.

d) Bestem hvordan konstantene a og b bør velges for at konfidensintervallet du utledet i punkt **c**) skal bli kortest mulig. Sammenlign med resultatet i punkt **b**) og kommenter.