Noregs teknisk–naturvitskaplege universitet
Institutt for matematiske fag

Nynorsk

Fagleg kontakt under eksamen:
Håkon Tjelmeland 73 59 35 38

EKSAMEN I EMNE TMA4240 STATISTIKK

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

Hjelpemiddel: Kalkulator HP30S eller Citizen SR-270X med tomt minne.
Statistiske tabeller og formler, Tapir forlag.
K. Rottman: Matematisk formelsamling.
Eitt gult ark (A5 med stempel) med eigne handskrevne formlar og notat.

Sensur er ferdig: 23. desember 2009.

Oppgåve 1

La X vere ein kontinuerleg fordelt stokastisk variabel med sannsynstettleik

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

der k er ein konstant.

a) Rekn ut kva for eit verd konstanten k må ha. Skisser $f(x)$.
Finn kumulativ fordelingsfunksjon $F(x) = \mathrm{P}(X \leq x)$.
Finn sannsyna

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

La óg Y vere ein kontinuerleg fordelt stokastisk variabel, og la betinga sannsynstettleik for Y gitt $X = x$ vere gitt ved

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

b) Finn simultan sannsynstettleik for X og Y.

Finn marginal sannsynstettleik for Y.

Finn sannsynet

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

Oppgåve 2 Telefonproblem

Ei gruppe av n studentar sit ein dag og drøftar at dei har store problem med å kome gjennom på telefon til eit bestemt firma. Firmaet har ikkje eit system med telefonkø slik at ein som ringjer firmaet får anten opptattsignal eller kontakt med sentralbordet til firmaet. Fleire av studantene klagar på at dei tilsynelatande alltid får opptattsignal. Studentane bestemmer seg for å gjere ein del forsøk for å få betre oversikt over situasjonen. Kvar av dei n studentane skal fleire gonger ringje til firmaet og skrive ned om dei får opptattsignal eller kontakt med sentralbordet. Kvar student skal fortsetje med å ringje inntil han/ho har fått snakka med sentralbordet k gonger og så rapportere attende til gruppa talet på oppringingar til firmaet.

Nummerer studentane frå 1 til n og la X_i vere talet på oppringingar student nummer i rapporterer attende til gruppa. La p vere sannsynet for å få snakke med sentralbordet viss ein ringjer ei gong til firmaet.

Studentane, som ikkje er så veldig gode i statistikk, konkluderer raskt med at $X_1, X_2, \ldots, X_n$ er eit tilfeldig utval frå ei negativ binomisk fordeling med parametrar k og p, dvs.

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

a) Kva vilkår, i tillegg til dei gitt over, må vere oppfylte for at konklusjonen til studentane om at $X_1, X_2, \ldots, X_n$ er eit tilfeldig utval frå ei negativ binomisk fordeling skal vere korrekt?

Gå ut frå at desse vilkåra er oppfylte, og at $k = 2$ og $p = 0.1$. Finn då sannsyna

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

I resten av denne oppgåva skal vi gå ut frå at vilkåra gitt i punkt **a**) er oppfylte slik at konklusjonen til studentane er korrekt. Vidare skal vi gå ut frå at k er eit kjent tal, medan p er ukjent og skal estimerast.

b) Utlei sannsynsmaksimeringsestimatoren (SME) for p ut frå dataene $X_1, X_2, \ldots, X_n$ og vis at han kan skrivast på forma

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

Ei tid seinare ser studentane ei reportasje på fjernsyn der ein representant for firmaet blir intervjua om nettopp situasjonen studentane var opptekne av. Representanten for firmaet vedgår at det kan vere noko vanskeleg å kome gjennom til dei på telefon, men hevder bestemt at $p \geq 0.1$. Studentane bestemmer seg for å teste om dataene dei har samla inn gir grunnlag for å hevde at firmaets påstand er feil.

c) Formuler studentanes problemstilling som eit hypotesetestingsproblem. Dvs, velj nullhypotese og alternativ hypotese, velj testobservator, og bestem forkastningskriterium.

Rekn ut (tilnærma) p-verdet for testen når $k = 2$, $n = 50$ og studentanes observasjonar gav $\sum_{i=1}^{n} x_i = 779$. Vil du konkludere med at firmaets påstand er feil? (Gi grunn for svaret)

Oppgåve 3 Smeltepunktbestemmelse

Ein metallurg har vore med på å utvikle ei ny legering og skal presentere ulike eigenskapar ved legeringa til sine kolleger. Vi skal her sjå på smeltepunktet for legeringa.

Til å bestemme smeltepunktet til legeringa har metallurgen to målemetodar. Vi kallar desse høvesvis målemetode A og målemetode B. På grunn av målefeil kan gjentatte målinger av smeltepunktet ved målemetode A antekes å vere realisasjonar av uavhengige og normalfordelte variablar med forventningsverd μ og varians σ_A^2. Metallurgen ynskjer å estimere forventningsverdet μ. Tilsvarande kan gjentatte målinger av smeltepunktet ved målemetode B antekes å vere realisasjonar av uavhengige og normalfordelte variablar med forventningsverd μ og varians σ_B^2.

Vi skal i denne oppgåva gå ut frå at variansane σ_A^2 og σ_B^2 har kjende verd, medan det felles forventningsverdet μ er ukjent og skal estimerast. Til dette føremål gjer metallurgen n målinger med målemetode A og m målinger med målemetode B. La $X_1, X_2, \ldots, X_n$ vere resultata av målingene ved metode A og la $Y_1, Y_2, \ldots, Y_m$ vere resultata av målingene ved metode B. Vi skal óg gå ut frå at $X_1, X_2, \ldots, X_n$ er uavhengig av $Y_1, Y_2, \ldots, Y_m$.

Som estimator for μ nyttar 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 konstantar.

a) Nytt reknereglar for forventningsverd 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 verd for konstantane a og b slik at $\widehat{\mu}$ blir ein best mogleg estimator for μ.

Vidare i oppgåva skal du framleis ta utgangspunkt i $\widehat{\mu} = a\bar{X} + b\bar{Y}$, dvs du skal **ikkje** nytte dei optimale verda for a og b du fann i punkt **b**).

c) Kva for ein type sannsynsfordeling har $\widehat{\mu}$? Gi grunn for svaret.

Utlei eit $(1-\alpha)\cdot 100\%$ konfidensintervall for μ.

d) Finn korleis konstantane a og b bør vere for at konfidensintervallet du utleidde i punkt **c**) skal bli som kort som mogleg. Samanlikn med resultatet i punkt **b**) og kommenter.