Noregs teknisk–
naturvitskaplege universitet
Institutt for matematiske fag

Nynorsk

Fagleg kontakt under eksamen:
Håkon Tjelmeland 7359 3538

EKSAMEN I EMNE ST1201/ST6201 STATISTISKE METODAR

Fredag 8. desember 2008
Tid: 09:00–13:00

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

Sensur er ferdig: 29. desember 2008.

Oppgåve 1

Gå ut ifrå at $X_1, X_2, \ldots, X_n$ er eit tilfeldig utval frå ei laplacefordeling, dvs. sannsynstettleiken for X_i'ane er gitt som

$$f(x) = \frac{1}{2\theta} e^{-\frac{|x|}{\theta}}, \ -\infty < x < \infty.$$

Parametaren $\theta > 0$ er ukjent og skal estimeres.

a) Finn sannsynsmaksimeringsestimatoren, $\widehat{\theta}$, for θ.

Er $\widehat{\theta}$ forventingsrett?

Oppgåve 2

Gå ut ifrå at $X_1, X_2, \ldots, X_n$ er eit tilfeldig utval frå ei normalfordeling med forventing μ_x og varians σ^2, og at $Y_1, Y_2, \ldots, Y_m$ er eit tilfeldig utval frå ei normalfordeling med forventing μ_y og varians $k\sigma^2$. Gå dessutan ut ifrå at X_i'ane og Y_i'ane alle er uavhengige av kvarandre. Parametrane μ_x og μ_y er ukjente og vi skal i denne oppgåva se på korleis vi kan estimere og lage konfidensintervall for differansen $\mu_x - \mu_y$.

Vi brukar følgjande estimatorar for μ_x og μ_y:

$$\widehat{\mu}_x = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{og} \quad \widehat{\mu}_y = \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i.$$

a) Grunngi at $\widehat{\mu}_x - \widehat{\mu}_y$ er normalfordelt.

Vis ved å nytte kjente regnereglar for forventingsverdi og varians, samt føresetnadene gitt over, at

$$\mathrm{E}\left[\widehat{\mu}_x - \widehat{\mu}_y\right] = \mu_x - \mu_y \quad \text{og} \quad \mathrm{Var}\left[\widehat{\mu}_x - \widehat{\mu}_y\right] = \sigma^2\left(\frac{1}{n} + \frac{k}{m}\right).$$

b) Gå i dette punktet ut ifrå at begge parametrane σ^2 og k er kjente. Utlei då eit $(1-\alpha)\cdot 100\%$-konfidensintervall for differansen $\mu_x - \mu_y$.

I resten av oppgåva skal vi gå ut ifrå at parametaren σ^2 er ukjent, mens parametaren k er kjent.

La S_x^2 vere empirisk varians for X_i'ane og la S_y^2 vere empirisk varians for Y_i'ene, dvs.

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \quad \text{og} \quad S_y^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(Y_i - \bar{Y}\right)^2.$$

Då har vi som kjent at

$$\frac{(n-1)S_x^2}{\sigma^2} \sim \chi^2_{n-1} \quad \text{og} \quad \frac{(m-1)S_y^2}{k\sigma^2} \sim \chi^2_{m-1},$$

samt at S_x^2 og $\bar{X}$ er uavhengige stokastiske variablar og at S_y^2 og $\bar{Y}$ er uavhengige stokastiske variablar.

c) Vis at

$$S_p^2 = \frac{n-1}{n+m-2} S_x^2 + \frac{m-1}{k(n+m-2)} S_y^2$$

er ein forventingsrett estimator for σ^2 og at

$$\frac{(n+m-2)S_p^2}{\sigma^2} \sim \chi^2_{n+m-2}.$$

Vis vidare at

$$T = \frac{(\widehat{\mu}_x - \widehat{\mu}_y) - (\mu_x - \mu_y)}{\sqrt{S_p^2 \left(\frac{1}{n} + \frac{k}{m}\right)}}$$

er Student t-fordelt med $n+m-2$ fridomsgrader. Når du gir svara for dette punktet, spesifiser spesielt kva for kjente eigenskapar og samanhengar mellom ulike typar fordelinger du nytter, og grunngi kvifor vilkåra for desse eigenskapane er oppfylte i den aktuelle situasjonen.

d) Utlei eit $(1-\alpha) \cdot 100\%$-konfidensintervall for differansen $\mu_x - \mu_y$.

Gå ut ifrå at du får vere med i planlegginga av forsøket. Det totale talet på målinger som kan utføres, $N = n + m$, er gitt ut frå økonomien i prosjektet, men kan du avgjere talet på målinger av kvar type, n og m. Korleis vil du velje n og m for at forventa lengd på konfidensintervallet skal bli minst mogleg?

Oppgåve 3

La X vere inntekta til ein tilfeldig valt lønnsmottakar i ei befolkningsgruppe. Det er då mye vanleg å gå ut ifrå at X er paretofordelt, dvs at X har sannsynstettleik

$$f(x) = \begin{cases} \frac{\theta c^\theta}{x^{\theta+1}} & \text{når } c < x < \infty, \\ 0 & \text{elles,} \end{cases}$$

kor c er minsteinntekta i denne befolkningsgruppa og $\theta > 1$ er ein parametar som skildrar lønnsskilnadene i gruppa. I denne oppgåva skal vi gå ut ifrå at minsteinntekta c er kjent, mens θ er ukjent og skal estimeres frå observerte data.

a) La $Y = 2\theta(\ln X - \ln c)$. Vis at Y er kji-kvadrat fordelt med to fridomsgrader.

For å estimere parametaren θ gjer ein n observasjonar $X_1, X_2, \ldots, X_n$ av inntekter i den aktuelle befolkningsgruppa. Vi skal gå ut ifrå at desse utgjer eit tilfeldig utval frå paretofordelinga gitt over. Sannsynsmaksimeringsestimatoren (SME) for θ blir då (du treng **ikkje** å vise dette)

$$\widehat{\theta} = \frac{n}{\sum_{i=1}^{n} \ln X_i - n \ln c}.$$

Vi ønsker så å nytte observasjonane $X_1, X_2, \ldots, x_n$ til å avgjere om det er grunnlag for å påstå at $\theta > \theta_0$, der θ_0 er et gitt tal. Vi har altså ein hypotesetestingssituasjon med hypotesene

$$H_0 : \theta = \theta_0 \quad \text{mot} \quad H_1 : \theta > \theta_0.$$

Gå ut ifrå at vi vil nytte avgjerdsregelen at ein skal forkaste H_0 dersom $\widehat{\theta} \geq k$ der k er slik at sannsynet for type I feil er lik eit gitt signifikansnivå α, dvs

$$\mathrm{P}\left(\widehat{\theta} \geq k \,\middle|\, H_0 \text{ er riktig}\right) = \alpha.$$

b) Grunngi at $2n\theta/\widehat{\theta}$ er kji-kvadrat fordelt med $2n$ fridomsgrader.

Nytt dette til å utleie ein formel for k som funksjon av n, θ_0 og α. Rekn deretter ut ein numerisk verdi for k når $n = 20$, $\theta_0 = 2.0$ og $\alpha = 0.01$.

c) Når $n = 20$, $\theta_0 = 2.0$ og $\alpha = 0.01$, avgjer for kva verdi av θ sannsynet for type II feil er lik 0.05.