Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Bokmål

Faglig kontakt under eksamen:

Håkon Tjelmeland	73 59 35 20
Oddvar Kristian Husby	73 59 35 20
Odd Kolbjørnsen	73 59 35 20

EKSAMEN I EMNE SIF5060/SIF5505 STATISTIKK

Torsdag 30. november 2000
Tid: 09:00–14:00

Hjelpemidler: Godkjent lommekalkulator med tomt minne.
Tabeller og formler i statistikk, Tapir forlag.
K. Rottman: Matematisk formelsamling.
Ordliste utdelt på forelesning.

Oppgave 1 Fabrizio Frizzi

På italiensk fjernsyn vises det et søndagsshow med Fabrizio Frizzi som programleder, der Frizzi sitter ved siden av en stor safe som inneholder 250 000 000 lire. Safen har en hemmelig firesifret kode. Seerne ringer inn og foreslår koder, og den første som gjetter riktig, får pengene som safen inneholder. Hvert av sifrene i den firesifrede koden er altså et av tallene $0, 1, 2, 3, 4, 5, 6, 7, 8$ eller 9. Anta foreløpig at ingen av seerne er i stand til å huske koder som er foreslått av tidligere innringere og forkastet, slik at hvert gjett kan betraktes som et tilfeldig forslag av en firesifret kode.

a) Hva er sannsynligheten for at en tilfeldig innringer gjetter riktig kode?

La X være antall innringinger til (og med) den personen som vinner pengene. Gjør kort rede for hvorfor X er geometrisk fordelt.

Bestem sannsynligheten for at akkurat innringer nummer 300 er den første som gjetter riktig.

Anta så at det i begynnelsen av programmet opplyses at sifferet 7 forekommer nøyaktig to ganger i den hemmelige koden (det er i denne formen showet faktisk blir vist) og la m betegne antall koder innringerne nå kan velge blant.

b) Vis at $m = 486$ (dvs. forklar hvordan denne verdien fremkommer).

Dersom vi antar at Frizzi snakker så fort at han rekker å ta i mot to forslag per minutt, hva blir nå forventet tid til noen vinner de 250 000 000 lire?

Anta så at seerne skriver ned alle koder som blir foreslått slik at ingen innringere foreslår en kode som tidligere er blitt foreslått. La Y betegne antall innringinger til førstemann gjetter riktig kode i dette tilfellet. Anta fremdeles at det opplyses at sifferet 7 opptrer nøyaktig to ganger i koden.

c) Bestem svarene på følgende spørsmål uttrykt som funksjon av m:

Hva er utfallsrommet til Y?

Utled sannsynlighetsfordelingen til Y.

Hva blir nå forventet tid til noen vinner de 250 000 000 lire (hvis Frizzi fremdeles rekker å ta i mot to forslag per minutt)?

Oppgave 2 Exit polls

I forbindelse med valg og folkeavstemninger er det i dag vanlig at meningsmålingsinstitutter foretar såkalte “exit polls”. Dette utføres ved at et tilfeldig utvalg av de som har avgitt stemme, i det de kommer ut fra stemmelokalet, blir spurt om hva de stemte. I denne oppgaven skal vi regne litt på denne situasjonen for et valg mellom kun to kandidater, som vi skal benevne henholdsvis G og B. Vi skal altså se bort fra at noen kan stemme blankt og at stemmer kan bli forkastet. Vi skal også se bort fra muligheten av at noen ikke vil svare hva de har stemt, eller at noen svarer usant.

La N betegne antall personer som avgir stemme og la p betegne andelen av disse som stemte på G. La videre n betegne antall personer som ble spurt av meningsmålingsinstituttet om sin stemmegivning og la X betegne antall av disse n som stemte på G.

a) Hvilke(n) betingelse(r) må være oppfylt for at X skal være tilnærmet binomisk fordelt ?

Gitt at denne betingelsen er oppfylt, bestem følgende sannsynligheter for $n = 20$ og $p = 0.50$:

$$\mathrm{P}(X = 9) \quad , \quad \mathrm{P}(X > 9) \quad \text{og} \quad \mathrm{P}(X > 9 | X \leq 12)$$

I resten av oppgaven skal vi forutsette at X er (tilnærmet) binomisk fordelt.

b) Utled sannsynlighetsmaksimeringsestimatoren (SME) for p.

Bestem estimatorens forventning og varians.

Som kjent kan en binomisk fordeling tilnærmes med en normalfordeling dersom n er stor nok. I resten av oppgaven skal vi forutsette at dette er tilfelle slik at vi har at

$$\frac{X - np}{\sqrt{np(1-p)}}$$

er (tilnærmet) standard-normalfordelt.

Anta at meningsmålingen utføres på oppdrag fra et fjernsynsselskap. Dersom et tilstrekkelig stort flertall av de spurte stemte på en av kandidatene, vil fjernesynsselskapet gå ut og erklære denne kandidaten som vinner av valget.

c) Formuler dette som en hypotesetest. Spesifiser nullhypotese og alternativ hypotese og bestem forkastningskriteriet når signifikansnivået er α.

Hva blir konklusjonen på hypotesetesten dersom 5 000 ble spurt, 2562 av disse hadde stemt på G, og en benytter signifikansnivå $\alpha = 0.10$?

Oppgave 3 SAR-målinger

SAR (Synthetic Aperture Radar) er en målemetode som benyttes for å kartlegge jordoverflaten fra satelitt. Teknikken går kort fortalt ut på at man sender ut radarstråler fra satelitten og observerer hvor mye av denne strålingen som reflekteres tilbake. Hvor mye av radarstrålingen som reflekteres, avhenger av egenskapene til jordoverflaten på den aktuelle posisjonen og dermed kan man skille mellom ulike overflatetyper. En observasjon gjøres egentlig ved at man tar flere målinger (såkalte "looks") og summerer disse. Ut fra fysiske lover for radarstråler er det kjent at en observasjon, X, vil være gammafordelt med parametre a og $b = r/a$, dvs. sannsynlighetstettheten er gitt ved

$$f(x) = \frac{a^a}{r^a \Gamma(a)} x^{a-1} \exp\left\{-\frac{ax}{r}\right\},$$

der a er antall "looks" og refleksivitetsparameteren r er en størrelse som beskriver refleksjonsegenskapene til jordoverflaten der observasjonen gjøres. Ut fra kjente formler for forventningsverdi og varians for gammafordelingen vet vi dermed også at tilhørende forventningsverdi og varians er gitt ved

$$\mathrm{E}(X) = r \qquad \text{og} \qquad \mathrm{Var}(X) = \frac{r^2}{a}.$$

Vi skal nå anta at vi har n observasjoner fra et homogent område (dvs. størrelsen r er den samme for alle n observasjonene). La $X_1, X_2, \ldots, X_n$ betegne de n observasjonene og anta at de er uavhengige. Fra disse observasjonene er vi interessert i å estimere r. Antall "looks", a, antar vi kjent. Som estimator for r skal vi benytte

$$\widehat{r} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

a) Vis at $\widehat{r}$ er forventningsrett og bestem estimatorens varians.

Redegjør kort for det generelle resultatet i sentralgrenseteoremet. Forklar deretter hvordan vi i vår situasjon kan benytte sentralgrenseteoremet til å konkludere at

$$\sqrt{na}\,\frac{\widehat{r} - r}{r} \tag{1}$$

er tilnærmet standard-normalfordelt når n er stor.

b) Benytt resultatet oppgitt i **a)** til å utlede et tilnærmet $(1-\alpha)\cdot 100\%$ konfidensintervall for r når n er stor. Uttrykk svaret som funksjon av α, a, n og $\overline{X}$.

Regn også ut numerisk verdi for konfidensintervallet når $\alpha = 0.05$, $a = 5$, $n = 20$ og de observerte verdier for $X_i; i = 1, \ldots, 20$ er

7.98	10.82	15.88	17.00	24.22	12.20	8.17	16.53	7.46	14.31
34.55	19.46	20.21	13.58	10.98	4.42	24.92	30.29	23.45	23.36

Det oppgis at $\overline{x} = 16.99$.

Anta så at man har observasjoner fra to ulike områder, område A og område B. Fra område A har man n observasjoner, $X_1, X_2, \ldots, X_n$, og alle disse antas å ha samme refleksjonsparameter r_A. Fra område B har man også n observasjoner, $Y_1, Y_2, \ldots, Y_n$, og alle disse antas å ha refleksjonsparameter r_B. Vi skal anta at alle $2n$ observasjonene er uavhengige og at alle er tatt med samme antall "looks", a. Man ønsker å benytte de totalt $2n$ observasjonene til å undersøke om det er grunn til å påstå at de to refleksjonsparametrene r_A og r_B er ulike, dvs. nullhypotese og alternativ hypotese er

$$\mathrm{H}_0 : r_A = r_B \qquad \text{mot} \qquad \mathrm{H}_1 : r_A \neq r_B.$$

c) Når n fremdeles antas stor, finn en passende testobservator for å teste hypotesene gitt over og bestem forkastningskriteriet slik at testen får (tilnærmet) signifikansnivå α. Om nødvendig kan du her gjøre flere approksimasjoner, man disse skal i så fall begrunnes.

Oppgave 4

Anta at X er standard-normalfordelt og la $Y = X^2$. Vis at da er Y χ^2-fordelt med 1 frihetsgrad.