Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Nynorsk

Fagleg 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

Hjelpemidel: Godkjend 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 blir det vist eit søndagsshow med Fabrizio Frizzi som programleidar, der Frizzi sit ved sida av ein stor safe som inneheld 250 000 000 lire. Safen har ein hemmeleg firesifra kode. Sjåarane ringjer inn og foreslår kodar, og den første som gjetter rett, får pengane i safen. Kvart av sifra i den firesifra koden er altså eit av tala $0, 1, 2, 3, 4, 5, 6, 7, 8$ eller 9. Anta foreløpig at ingen av sjåarane er i stand til å hugse kodane som er foreslått av tidligare innringjarar og blitt forkasta, slik at kvart gjett kan bli sett på som eit tilfeldig forslag av ein firesifra kode.

a) Hva er sannsynet for at ein tilfeldig innringjar gjetter rett kode?

La X vere talet på innringjarar til (og med) den personen som vinn pengane. Gjer kort greie for kvifor X er geometrisk fordelt?

Bestem sannsynet for at akkurat innringjar nummer 300 er den første som gjettar rett.

Anta så at det i byrjinga av programmet blir opplyst at sifferet 7 forekjem nøyaktig to gongar i den hemmelege koden (det er i denne formen showet faktisk blir vist) og la m vere talet på kodar innringjarane no kan velgje blant.

b) Vis at $m = 486$ (dvs. forklar korleis denne verdien framkjem).

Dersom vi antar at Frizzi snakkar så fort at han rekk å ta i mot to forslag per minutt, hva blir no forventa tid til nokre vinn dei 250 000 000 lirene ?

Anta så at sjåarane skriv ned alle kodane som blir foreslått slik at ingen innringjarar foreslår ein kode som tidligare er blitt foreslått. La Y vere talet på innringingar til førstemann gjetter rett kode i dette tilfellet. Anta framleis at det blir opplyst at sifferet 7 opptrer nøyaktig to gongar i koden.

c) Bestem svara på følgjande spørsmål uttrykt som funksjon av m:

Kva er utfallsrommet til Y ?

Utlei sannsynsfordelinga til Y.

Hva blir no forventa tid til nokre vinn dei 250 000 000 lirene (dersom Frizzi framleis rekk å ta i mot to forslag per minutt) ?

Oppgave 2 Exit polls

I samband med val og folkerøystingar er det i dag vanleg at meningsmålingsinstitutt foretar såkalte "exit polls". Dette blir utført ved at eit tilfeldig utval av dei som har avgjeve stemme, i det dei kjem ut frå stemmelokalet, blir spurt om kva dei stemte. I denne oppgåva skal vi rekne litt på denne situasjonen for eit val mellom kun to kandidatar. Vi kallar dei G og B. Vi skal altså sjå bort frå at nokre kan stemme blankt og at stemmer kan bli forkasta. Vi skal og sjå bort frå moglegheita av at nokre ikkje vil svare kva dei har stemt, eller at nokre svarer usant.

La N vere talet på personar som avgjer stemme og la p vere delen av desse som stemte på G. La vidare n vere talet på personar som blei spurde av meningsmålingsinstituttet om si stemmegjevning og la X vere talet på dei av desse n som stemte på G.

a) Kva føresetning(ar) må vere oppfylt for at X skal vere tilnærma binomisk fordelt ?

Gjeve at denne føresettnaden er oppfylt, bestem følgjande sannsyn 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 oppgåva skal vi forutsettje at X er (tilnærma) binomisk fordelt.

b) Utlei sannsynsmaksimeringsestimatoren (SME) for p.

Bestem estimatoren si forventning og varians.

Som kjend kan ei binomisk fordeling tilnærmast med ei normalfordeling dersom n er stor nok. I resten av oppgåva skal vi forutsettje at dette er tilfelle slik at vi har at

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

er (tilnærma) standard-normalfordelt.

Anta at meningsmålinga blir utført på oppdrag frå eit fjernsynsselskap. Dersom eit tilstrekkeleg stort fleirtal av dei spurde stemte på ein av kandidatane, vil fjernesynsselskapet gå ut og erklære denne kandidaten som vinnar av valet.

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

Hva blir konklusjonen på hypotesetesten dersom 5 000 ble spurde, 2562 av desse hadde stemt på G, og ein nyttar signifikansnivå $\alpha = 0.10$?

Oppgave 3 SAR-målingar

SAR (Synthetic Aperture Radar) er ein målemetode som blir nytta for å kartleggje jordoverflata frå satelitt. Teknikken går kort fortalt ut på at ein sender ut radarstråler frå satelitten og observerer kor mykje av denne strålinga som blir reflektert tilbake. Kor mykje av radarstrålinga som blir reflektert, avheng av eigenskapane til jordoverflata på den aktuelle posisjonen og dermed kan ein skille mellom ulike overflatetypar. En observasjon blir eigentleg gjort ved at ein tar fleire målingar (såkalte "looks") og summerer desse. Ut frå fysiske lover for radarstråler er det kjend at ein observasjon, X, vil vere gammafordelt med parameter a og $b = r/a$, dvs. sannsynstettleiken er gjeve ved

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

der a er talet på "looks" og refleksivitetsparameteren r er ein storleik som skildrar refleksjonseigenskapane til jordoverflata der observasjonen blir gjort. Ut frå kjende formlar for forventningsverdi og varians for gammafordelingen veit vi dermed også at tilhøyrande forventningsverdi og varians er gjeve ved

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

Vi skal no anta at vi har n observasjonar frå eit homogent område (dvs. storleiken r er den samme for alle n observasjonane). La $X_1, X_2, \ldots, X_n$ vere dei n observasjonane og anta at dei er uavhengige. Frå desse observasjonane er vi interessert i å estimere r. Tallet på "looks", a, antar vi kjend. Som estimator for r skal vi nytte

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

a) Vis at $\widehat{r}$ er forventningsrett og bestem variensen til estimatoren.

Gjer kort greie for det generelle resultatet i sentralgrenseteoremet. Forklar deretter korleis vi i vår situasjon kan nytte sentralgrenseteoremet til å konkludere at

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

er tilnærma standard-normalfordelt når n er stor.

b) Bruk resultatet oppgjeve i **a**) til å utleie eit tilnærma $(1-\alpha)\cdot 100\%$ konfidensintervall for r når n er stor. Uttrykk svaret som funksjon av α, a, n og $\overline{X}$.

Rekn og ut numerisk verdi for konfidensintervallet når $\alpha = 0.05$, $a = 5$, $n = 20$ og dei observerte verdiane 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 blir oppgjeve at $\overline{x} = 16.99$.

Anta så at ein har observasjonar frå to ulike område, område A og område B. Frå område A har ein n observasjonar, $X_1, X_2, \ldots, X_n$, og alle desse blir antatt å ha same refleksjonsparameter r_A. Frå område B har ein og n observasjonar, $Y_1, Y_2, \ldots, Y_n$, og alle desse antar ein har refleksjonsparameter r_B. Vi skal anta at alle $2n$ observasjonane er uavhengige og at alle er tatt med samme talet på "looks", a. Ein ønsker å nytte dei totalt $2n$ observasjonene til å undersøke om det er grunn til å påstå at de to refleksjonsparametera 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 framleis forutsettjast å vere stor, finn ein passande testobservator for å teste hypotesene gjeve over og bestem forkastningskriteriet slik at testen får (tilnærma) signifikansnivå α. Om nødvendig kan du her gjere fleire approksimasjonar, men desse skal i så fall grunngjevast.

Oppgave 4

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