**Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag**

Faglig kontakt under eksamen:
Turid Follestad (98 06 68 80/73 59 35 37)
Hugo Hammer (45 21 01 84/73 59 77 74)
Eirik Mo (41 10 66 33/73 55 02 39)
Henning Omre (90 93 78 48/73 59 35 31)

EKSAMEN I TMA4245 Statistikk

Torsdag 7. juni 2007
Tid: 09:00 – 13:00

Tillatte hjelpemidler:
Gult A5-ark med egne håndskrevne notater (stemplet ved Institutt for matematiske fag)
Tabeller og formler i statistikk (Tapir forlag)
K. Rottmann: Matematisk formelsamling
Kalkulator: HP30S

BOKMÅL

Sensur: 28. juni 2007

Oppgave 1 Pengespillet

I et TV-program får et visst antall deltakere sjansen til å vinne et større pengebeløp. For hver deltaker består spillet av en serie påfølgende runder, der deltakeren i hver runde får presentert en oppgave. For hver oppgave deltakeren klarer, får han/hun et gitt beløp. Spillet avsluttes når deltakeren første gang ikke greier oppgaven, og deltakeren får da med seg beløpet vunnet i de øvrige rundene. Vi antar at ingen deltaker trekker seg frivillig underveis.

La p være sannsynligheten for IKKE å klare oppgaven i hver enkelt runde, og la videre X være antall runder for en tilfeldig valgt deltaker. Antall runder X defineres her slik at deltakeren går ut etter å ha klart oppgavene i de $X-1$ første rundene, men ikke oppgaven i runde X. Vi antar at sannsynligheten p er lik for hver runde og for hver deltaker, og at resultatene for hver runde er uavhengige.

I denne situasjonen er X geometrisk fordelt med parameter p, slik at punktsannsynligheten $f(x;p)$ og den kumulative fordelingsfunksjonen $F(x;p) = P(X \leq x)$ for X er

$$\begin{aligned} f(x;p) &= p(1-p)^{x-1}, \quad x = 1, 2, \ldots \\ F(x;p) &= 1-(1-p)^x, \quad x = 1, 2, \ldots \end{aligned}$$

a) Anta bare i dette punktet at $p = 0.10$.

Forklar hvorfor X er geometrisk fordelt med parameter p i denne situasjonen.

Beregn sannsynligheten for at deltakeren går ut i første runde.

Beregn sannsynligheten for at deltakeren er med i spillet når det er gått fem runder.

Hva er sannsynligheten for at han/hun kommer videre til niende runde men ikke klarer oppgaven i niende runde, gitt at deltakeren var med i spillet når det var gått fem runder?

Det viser seg at deltakerne jevnt over henger med lenger enn forventet, og det begynner å bli dyrt for TV-selskapet. De vil undersøke om de har feilvurdert vanskelighetsgraden, og vil beregne et anslag for p, som nå ansees som ukjent.

La $X_1, X_2, \ldots, X_n$ være antall runder for hver av n tilfeldig valgte deltakere, der $X_i, \ i = 1, \ldots, n$ er uavhengige og geometrisk fordelte med samme parameter p.

b) Utled et uttrykk for sannsynlighetsmaksimeringsestimatoren (maximum likelihood estimator) for p basert på det tilfeldige utvalget.

Hva blir estimatet dersom $n = 8$, og antall runder for de 8 deltakerne er 4, 22, 9, 11, 15, 5, 26 og 17?

TV-selskapet bruker to personer, A og B, til å lage oppgavene. Selskapet ønsker å undersøke om vanskelighetsgraden er avhengig av hvem av dem som lager oppgavene.

De ser på resultatene fra n_1 tilfeldig valgte deltakere som har oppgaver fra oppgavelager A, og n_2 fra oppgavelager B. La Z_1 og Z_2 være antallet blant disse som klarer færre enn fem oppgaver fra henholdsvis oppgavelager A og B. Vi antar at Z_1 og Z_2 er uavhengige.

c) Forklar hvorfor Z_1 og Z_2 er binomisk fordelte med parametre (n_1, q_1) og (n_2, q_2), der q_1 og q_2 er sannsynligheten for å klare færre enn fem oppgaver i de to gruppene.

Som estimatorer for q_1 og q_2 skal vi bruke

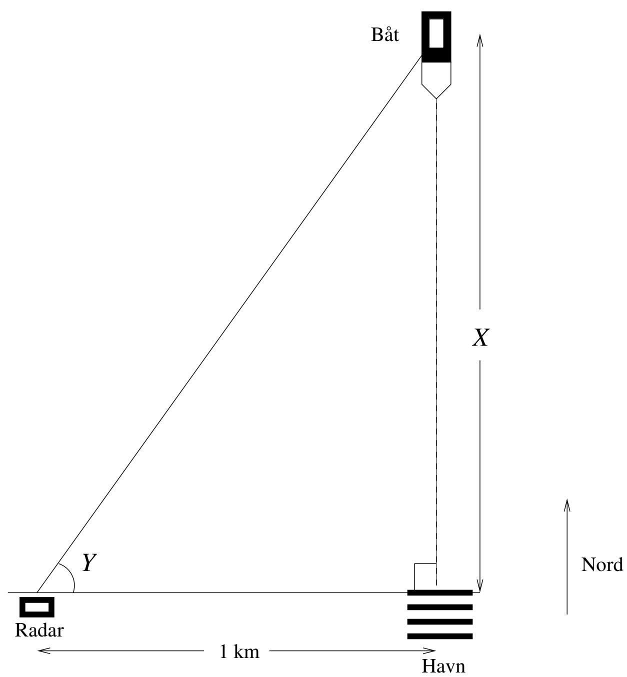
$$\hat{q_1} = \frac{Z_1}{n_1} \quad \text{og} \quad \hat{q_2} = \frac{Z_2}{n_2}.$$

Utled et tilnærmet 95% konfidensintervall for $q_1 - q_2$ basert på normaltilnærming til binomisk fordeling. Beregn intervallet numerisk når $n_1 = n_2 = 64$, og observerte verdier for Z_1 og Z_2 er $z_1 = 34$ og $z_2 = 18$.

Gir det estimerte konfidensintervallet TV-selskapet grunnlag for å si at oppgavene fra A og B har ulik vanskelighetsgrad? Begrunn svaret.

Oppgave 2 Radar

En havneby observerer ankommende skip ved å bruke radar. Vi antar for enkelhets skyld at skipene alltid ankommer fra nord. Radaren er plassert 1 kilometer vest for havna. Det er ønskelig å oppdage ankommende skip så tidlig som mulig av praktiske og sikkerhetsmessige årsaker. Når skipet første gang fanges inn på radaren, observerer radaren vinkelen $Y \in [0, \pi/2)$,



Figur 1: Illustrasjon til oppgave 2.

som vist i Figur 1. Radaren observerer kun vinkelen Y, og ikke avstanden til skipet. Vinkelen Y varierer fra skip til skip av mange årsaker.

La den kumulative fordelingsfunksjonen til Y være

$$F(y; \beta) = P(Y \leq y) = \frac{1 - \exp\{-y/\beta\}}{1 - \exp\{-\pi/(2\beta)\}}, \quad y \in [0, \pi/2),$$

der $\beta > 0$ er en parameter.

a) Anta bare i dette punktet at $\beta = \pi/8$.

Regn ut $\mathrm{P}\Big(Y > \dfrac{\pi}{4}\Big)$, $\mathrm{P}\Big(\dfrac{\pi}{4} < Y < \dfrac{\pi}{3}\Big)$ og $\mathrm{P}\Big(Y > \dfrac{\pi}{4} \mid Y < \dfrac{\pi}{3}\Big)$.

b) Vis at sannsynlighetstettheten $f(y;\beta)$ til Y er

$$f(y;\beta) = \frac{1}{\beta - \beta \exp\{-\pi/(2\beta)\}} \exp\{-y/\beta\}, \quad y \in [0, \pi/2).$$

Havnebyen er mer interessert i avstanden til havna når skipet først oppdages enn vinkelen Y som radaren observerer. La X betegne denne avstanden, som vist i Figur 1.

Utled et uttrykk for sannsynlighetstettheten til X.

Det oppgis at $\dfrac{d}{dx}(\tan(x)) = \dfrac{1}{\cos^2(x)}$ og $\dfrac{d}{dx}(\text{arc}\tan(x)) = \dfrac{d}{dx}(\tan^{-1}(x)) = \dfrac{1}{1+x^2}$.

Oppgave 3 Ultralydavbildning med kontrastmiddel

En måte å oppdage tidlig utvikling av kreftceller i leveren på er å studere tettheten av blodkar. Mikroskopiske gassbobler tilsettes blodet, og sees ved å sende ultralyd mot forskjellige deler av leveren. Mange gassbobler et sted gir kraftig høyfrekvent (annenharmonisk) ekko, som indikerer mange blodkar og mulig kreft.

La Y_i, $i = 1, \ldots, n$ være styrken på det høyfrekvente ekkoet i desibel ($20 \log_{10}$ av amplituden) som apparatet registrerer for n målinger. Vi vil anta at målingene Y_i er uavhengige og identisk normalfordelte, med varians σ^2.

a) Anta bare i dette punktet at $\sigma^2 = 0.01^2$ og at alle ekkodataene skaleres så forventningsverdien til Y_i er eksakt 1.0.

Regn ut sannsynligheten for at en enkelt ultralydmåling er større enn 1.0.

Regn ut sannsynligheten for at avviket i absoluttverdi fra 1.0 i en enkelt måling er større enn 0.02.

Hvis en tar to uavhengige målinger fra samme sted i leveren, hva er da sannsynligheten for at gjennomsnittet avviker i absoluttverdi mer enn 0.02 fra 1.0?

Absorbsjon og spredning gjør ultralydsignalet svakere fra punkter i leveren langt fra måleapparatet. Vi setter opp en lineær regresjonsmodell $Y = \alpha + \beta x + E$, med n par av målinger (x_i, Y_i), $i = 1, \ldots, n$, der Y_i er målt som før, og x_i er en kjent forklaringsvariabel basert på dybden i kroppen ekkoet kommer fra. Feilene E_i er uavhengig normalfordelte med forventningsverdi 0.0 og ukjent varians σ^2. Parametrene α og β avhenger av fysiologien til pasienten, og er ukjente.

Vi vil bruke følgende estimatorer for α og β:

$$\begin{aligned} A &= \widehat{\alpha} = \bar{Y} - B\bar{x}, \\ B &= \widehat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}, \end{aligned}$$

der $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ og $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$.

Det oppgis at

$$\mathrm{E}(B) = \beta \qquad \text{og} \qquad \mathrm{Var}(B) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

b) Utled uttrykk for $\mathrm{E}(A)$, $\mathrm{Var}(A)$ og $\mathrm{Cov}(A, B)$.

Ta med mellomregning/omforminger du bruker for å komme fram til svarene på enkleste form og bruk uten bevis at B og $\bar{Y}$ er uavhengige.

Sannsynlighetsmaksimeringsestimatoren (maximum likelihood estimator) for σ^2 basert på n målinger med forskjellige x_i er

$$\widetilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - A - Bx_i)^2.$$

Metoden er på prøvestadiet og testes bare på friske testpersoner. For en frisk pasient, skal variansen ikke overstige 0.01^2. (For disse pasientene er fordelingen av blodkar jevn/homogen, og variansen i målingene skyldes da bare tilfeldig variasjon i tettheten av gassbobler i blodet.) Hvis resultatet av $n = 30$ uavhengige målinger på en testperson gir grunnlag for å forkaste hypotesen om at $\sigma^2 = 0.01^2$ og hevde at $\sigma^2 > 0.01^2$, sendes testpersonen til en dyr kreftundersøkelse.

c) Utled et uttrykk for forventningsverdien til $\widetilde{\sigma}^2$.

Formulér problemstillingen over som en hypotesetest, og beregn kritisk område (forkastningsområde) for $\widetilde{\sigma}^2$ for testen. La signifikansnivået være 0.01.