Noregs teknisk–
naturvitskaplege universitet
Institutt for matematiske fag

Fagleg 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

Tillatne hjelpemiddel:
Gult A5-ark med eigne handskrivne notatar (stempla ved Institutt for matematiske fag)
Tabeller og formler i statistikk (Tapir forlag)
K. Rottmann: Matematisk formelsamling
Kalkulator: HP30S

NYNORSK

Sensur: 28. juni 2007

Oppgåve 1 Pengespelet

I eit TV-program får eit visst antal deltakarar sjansen til å vinne eit større pengebeløp. For kvar deltakar består spelet av ein serie påfølgjande runder, der deltakaren i kvar runde får presentert ei oppgåve. For kvar oppgåve deltakaren klarer, får han/ho eit gitt beløp. Spelet blir avslutta når deltakaren første gong ikkje klarer oppgåva, og deltakaren får då med seg beløpet vunne i dei øvrige rundene. Vi føreset at ingen deltakar trekker seg frivillig undervegs.

La p vere sannsynet for IKKJE å klare oppgåva i kvar enkelt runde, og la vidare X vere talet på runder for ein tilfeldig valt deltakar. Talet på runder X blir her definert slik at deltakaren går ut etter å ha klart oppgåvene i dei $X-1$ første rundene, men ikkje oppgåva i runde X. Vi føreset at sannsynet p er lik for kvar runde og for kvar deltakar, og at resultata for kvar runde er uavhengige.

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

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

a) Føreset berre i dette punktet at $p = 0.10$.

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

Rekn ut sannsynet for at deltakaren går ut i første runde.

Rekn ut sannsynet for at deltakaren er med i spelet når det er gått fem runder.

Kva er sannsynet for at han/ho kjem vidare til niande runde men ikkje klarer oppgåva i niande runde, gitt at deltakaren var med i spelet når det var gått fem runder?

Det viser seg at deltakarane jamnt over heng med lenger enn forventa, og det begynner å bli dyrt for TV-selskapet. Dei vil undersøke om dei har feilvurdert vanskegraden, og vil berekne eit anslag for p, som vi no ser på som ukjent.

La $X_1, X_2, \ldots, X_n$ vere talet på runder for kvar av n tilfeldig valde deltakarar, der $X_i,\ i = 1, \ldots, n$ er uavhengige og geometrisk fordelte med same parameter p.

b) Utlei eit uttrykk for sannsynsmaksimeringsestimatoren (maximum likelihood estimator) for p basert på det tilfeldige utvalet.

Kva blir estimatet dersom $n = 8$, og talet på runder for dei 8 deltakarne er 4, 22, 9, 11, 15, 5, 26 og 17?

TV-selskapet bruker to personar, A og B, til å lage oppgåvene. Selskapet ønskjer å undersøke om vanskegraden er avhengig av kven av dei som lagar oppgåvene.

Dei ser på resultata frå n_1 tilfeldig valde deltakarar som har oppgåver frå oppgåvelagar A, og n_2 frå oppgåvelagar B. La Z_1 og Z_2 vere talet på deltakarar blant desse som klarer færre enn fem oppgåver frå høvesvis oppgåvelagar A og B. Vi føreset at Z_1 og Z_2 er uavhengige.

c) Forklar kvifor Z_1 og Z_2 er binomisk fordelte med parametrar (n_1, q_1) og (n_2, q_2), der q_1 og q_2 er sannsynet for å klare færre enn fem oppgåver i dei to gruppene.

Som estimatorar for q_1 og q_2 skal vi bruke

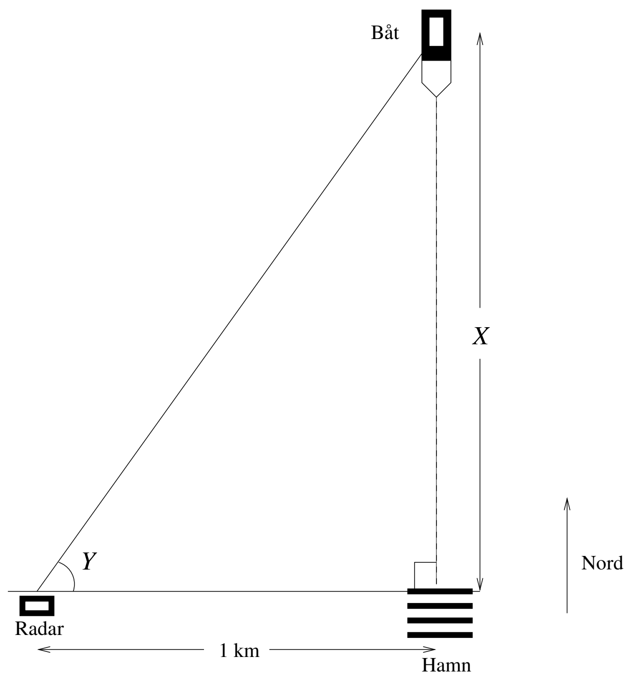
$$\hat{q}_1 = \frac{Z_1}{n_1} \quad \text{og} \quad \hat{q}_2 = \frac{Z_2}{n_2}.$$

Utlei eit tilnærma 95% konfidensintervall for $q_1 - q_2$ basert på normaltilnærming til binomisk fordeling. Rekn ut intervallet numerisk når $n_1 = n_2 = 64$, og observerte verdiar for Z_1 og Z_2 er $z_1 = 34$ og $z_2 = 18$.

Gir det estimerte konfidensintervallet TV-selskapet grunnlag for å seie at oppgåvene frå A og B har ulik vanskegrad? Grunngi svaret.

Oppgåve 2 Radar

Ein hamneby observerer skip som kjem inn mot hamna ved å bruke radar. Vi føreset for enkelheits skuld at skipa alltid kjem frå nord. Radaren er plassert 1 kilometer vest for hamna. Det er ønskjeleg å oppdage skip som kjem inn mot hamna så tidleg som mogleg av praktiske og tryggleiksmessige årsaker. Når skipet første gong blir fanga inn på radaren, observerer



Figur 1: Illustrasjon til oppgave 2.

radaren vinkelen $Y \in [0, \pi/2)$, som vist i Figur 1. Radaren observerer kun vinkelen Y, og ikkje avstanden til skipet. Vinkelen Y varierer frå skip til skip av mange årsaker.

La den kumulative fordelingsfunksjonen til Y vere

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

der $\beta > 0$ er ein parameter.

a) Føreset berre i dette punktet at $\beta = \pi/8$.

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

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

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

Hamnebyen er meir interessert i avstanden til hamna når skipet først blir oppdaga enn vinkelen Y som radaren observerer. La X vere denne avstanden, som vist i Figur 1.

Utlei eit uttrykk for sannsynstettleiken til X.

Det blir oppgitt at $\frac{d}{dx}(\tan(x)) = \frac{1}{\cos^2(x)}$ og $\frac{d}{dx}(\arc\tan(x)) = \frac{d}{dx}(\tan^{-1}(x)) = \frac{1}{1+x^2}$.

Oppgåve 3 Ultralydavbilding med kontrastmiddel

Ein måte å oppdage tidleg utvikling av kreftceller i levra på er å studere tettleiken av blodkar. Mikroskopiske gassbobler blir tilsett blodet, og blir sett ved å sende ultralyd mot forskjellige delar av levra. Mange gassbobler ein stad gir kraftig høgfrekvent (andreharmonisk) ekko, som indikerer mange blodkar og mogleg kreft.

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

a) Føreset berre i dette punktet at $\sigma^2 = 0.01^2$ og at alle ekkodataene blir skalerte slik at forventningsverdien til Y_i er eksakt 1.0.

Rekn ut sannsynet for at ei enkelt ultralydmåling er større enn 1.0.

Rekn ut sannsynet for at avviket i absoluttverdi frå 1.0 i ei enkelt måling er større enn 0.02.

Dersom ein tar to uavhengige målingar frå same stad i levra, kva er då sannsynet for at gjennomsnittet avviker i absoluttverdi meir enn 0.02 frå 1.0?

Absorbsjon og spreiiing gjer ultralydsignalet svakare frå punkt i levra langt frå måleapparatet. Vi set opp ein lineær regresjonsmodell $Y = \alpha + \beta x + E$, med n par av målingar (x_i, Y_i), $i = 1, \ldots, n$, der Y_i er målt som før, og x_i er ein kjent forklaringsvariabel basert på djupna i kroppen ekkoet kjem frå. Feila E_i er uavhengig normalfordelte med forventningsverdi 0.0 og ukjent varians σ^2. Parametrane α og β avhenger av fysiologien til pasienten, og er ukjente.

Vi vil bruke følgjande estimatorar 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 blir oppgitt 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) Utlei uttrykk for $\mathrm{E}(A)$, $\mathrm{Var}(A)$ og $\mathrm{Cov}(A, B)$.

Ta med mellomrekning/omformingar du bruker for å kome fram til svara på enklaste form og bruk utan bevis at B og $\bar{Y}$ er uavhengige.

Sannsynsmaksimeringsestimatoren (maximum likelihood estimator) for σ^2 basert på n målingar 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 blir berre testa på friske testpersonar. For ein frisk pasient, skal variansen ikkje overstige 0.01^2. (For desse personane er fordelinga av blodkar jamn/homogen, og variansen i målingane skuldast då berre tilfeldig variasjon i tettleiken av gassbobler i blodet.) Dersom resultatet av $n = 30$ uavhengige målingar på ein testperson gir grunnlag for å forkaste hypotesen om at $\sigma^2 = 0.01^2$ og hevde at $\sigma^2 > 0.01^2$, blir testpersonen sendt til ei dyr kreftundersøking.

c) Utlei eit uttrykk for forventningsverdien til $\widetilde{\sigma}^2$.

Formulér problemstillinga over som ein hypotesetest, og berekn kritisk område (forkastingsområde) for $\widetilde{\sigma}^2$ for testen. La signifikansnivået vere 0.01.