TMA4240 Statistikk
Høst 2020

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Anbefalte oppgaver 8

Oppgave 1 For flyselskapene er det et velkjent problem at en viss andel av de som har reservert plass, unnlater å møte opp til den flyavgang de har bestilt plass på. Selskapene har derfor innført en spesiell overbookingsteknikk, basert på at booking-situasjonen på hver enkelt flyging sammenliknes med tidligere avganger på samme ukedag og klokkeslett. På grunnlag av denne "no-show"-statistikken gjøres bevisste overbookinger, dvs. at det foretas flere reservasjoner enn det er plasser i flyet. Dette fører til at selskapene fra tid til annen må avvise passasjerer med bekreftet billett. (Disse blir imidlertid gitt en erstatning, såkalt "denied boarding compensation".) Vi studerer estimeringen av sannsynligheten p for at en person med reservert plass virkelig møter opp til flyavgangen. Flyselskapets "no-show"-statistikk for de siste 6 årene, for en bestemt avgang antas å være som følger:

År (i)	Reservert (n_i)	Møtt (X_i)
1	111	105
2	125	120
3	125	124
4	124	119
5	120	107
6	125	119
Totalt	730	694

Det antas at X_i ($i = 1, 2, \ldots, 6$) er binomisk fordelt med parametre n_i og p. Videre antas at X_1, $X_2, \ldots, X_6$ er stokastisk uavhengige.

Bestem konstanten k slik at estimatoren

$$\hat{p} = k \sum_{i=1}^{6} X_i$$

blir en forventningsrett estimator for p. Bestem estimatorens varians. Beregn estimatet for p fra de observerte verdier.

Oppgave 2

Sikkerhet er en av de høyest prioriterte oppgavene i norske alpinanlegg. Vi antar at antall alpinulykker som krever legebehandling i alpinanlegget "Alpinfjellet" i en periode på t skidager, X, er Poisson-fordelt med forventningsverdi $\mu = \lambda t$. Her er λ skadefrekvens pr skidag og t er eksponeringstid i antall skidager. En skidag er definert som “en person i alpinanlegget en hel dag”.

Det følger altså at punktsannsynligheten for X er

$$f(x) = \frac{(\lambda t)^x}{x!} \exp(-\lambda t); \quad \text{for} \quad x = 0, 1, 2, \ldots$$

a) Alpinanleggenes Landsforening har, basert på data fra noen av de største skianleggene i Norge, anslått at $\lambda = 1/1000$. Vi antar i dette punktet at det er kjent at $\lambda = 1/1000$ for “Alpinfjellet”.

Hvis vi ser på $t = 2000$ skidager, hva er da sannsynligheten for at det skjer akkurat én ulykke, $P(X = 1)$?

Hvis du tilbringer 10 skidager i “Alpinfjellet”, hva er sannsynligheten for at du utsettes for en eller flere ulykker?

Hvor mange skidager må du tilbringe i “Alpinfjellet” for at din sannsynlighet for minst en ulykke skal bli større enn 0.1?

Ved “Alpinfjellet” har man aktivt registert samhørende verdier av antall ulykker, X_i, og eksponering i antall skidager, t_i $(i = 1, ..., n)$, for n tilfeldig valgte dager anlegget var åpent. Vi antar at $X_1, \ldots, X_n$ er uavhengige.

b) Finn sannsynlighetsmaksimeringsestimatoren (SME) for λ basert på de n observasjonsparene $(X_1, t_1), (X_2, t_2), \ldots, (X_n, t_n)$.

Er estimatoren forventningsrett?

Oppgave 3

Densiteten (massetettheten) av snø er høyst 1 kg/dm^3 (veldig våt snø). Anta at sannsynlighetstettheten for densiteten i kg/dm^3 av en tilfeldig valgt snøprøve er gitt ved $f(x) = \beta(\beta+1)x(1-x)^{\beta-1}, \ 0 \leq x \leq 1$, der β er en positiv parameter.

a) Anta, bare i dette punktet, at $\beta = 2$.

Hva er sannsynligheten for at en tilfeldig valgt snøprøve har densitet mellom 0,5 og 0,9 kg/dm^3?

b) Finn sannsynlighetsmaksimeringsestimatoren for β basert på et tilfeldig utvalg X_1, X_2, $\ldots$, X_n av snødensiteter.

Hva blir estimatet hvis $n = 100$ og $\sum_{i=1}^{n} \ln(1 - x_i) = -104{,}0$?

Oppgave 4

Kjell Inge drar ofte og fisker en times tid i nærheten av der han bor.

a) Han mener at vekten på en fisk er normalfordelt med forventningsverdi 800 gram og standardavvik 100 gram. Anta at Kjell Inge har rett i det.

Hva er sannsynligheten for at en fisk veier mer enn 1000 gram?

Hva er sannsynligheten for at en fisk veier mellom 500 gram og 1000 gram?

Kjell Inge tenker at antallet fisk er viktigere enn vekten. Han grubler på sannsynlighetsfordelingen til antall fisk per tur.

b) La X være antall fisk på en tur. Alle turer tar 1 time. Vi antar at X er Poisson-fordelt med parameter (forventningsverdi) $\mu = 3$.

Hva er sannsynligheten for at Kjell Inge ikke får fisk på en tur?

Gitt at han får fisk, hva er sannsynligheten for at han får flere enn 3 fisk?

En alternativ sannsynlighetsmodell er som følger: Med sannsynlighet θ får han helt sikkert 0 fisk. Med sannsynlighet $1-\theta$ er antall fisk Poisson-fordelt med parameter μ. Punktsannsynligheten til antall fisk er da:

$$P(X = x) = \theta I(X = 0) + (1-\theta)\frac{\mu^x}{x!}e^{-\mu}, \quad x = 0, 1, \ldots$$

der $I(A) = 1$ dersom hendelsen A inntreffer, og $I(A) = 0$ ellers.

c) Anta at $\theta = 0.5$ og $\mu = 4$.

Hva er sannsynligheten for at Kjell Inge får fisk på en tur?

Bruk sannsynlighetsmodellen til å regne ut forventet antall fisk per tur.

Vi antar nå at vi har data fra $n = 20$ uavhengige fisketurer. Av disse endte $r = 8$ med ingen fisk. I de resterende tolv turene ble det tilsammen 40 fisk. Vi bruker dette tallmaterialet til å estimere modellparametrene θ og μ.

d) Sett opp rimelighetsfunksjonen (likelihood-funksjonen) for μ og θ.

Anta kjent μ. Vis at sannsynlighetsmaksimeringsestimatoren (maximum likelihood estimator) for θ er gitt som

$$\hat{\theta} = \hat{\theta}(\mu) = \frac{r - ne^{-\mu}}{n(1-e^{-\mu})}$$

Ved å sette $\hat{\theta}(\mu)$ inn i rimelighetsfunksjonen, kan vi studere rimelighetsfunksjonen kun som en funksjon av μ. Figuren nedenfor viser denne rimelighetsfunksjonen.

Bruk plottet til å finne sannsynlighetsmaksimeringsestimatet for μ.

Bruk resultatet til å regne ut sannsynlighetsmaksimeringsestimatet for θ.

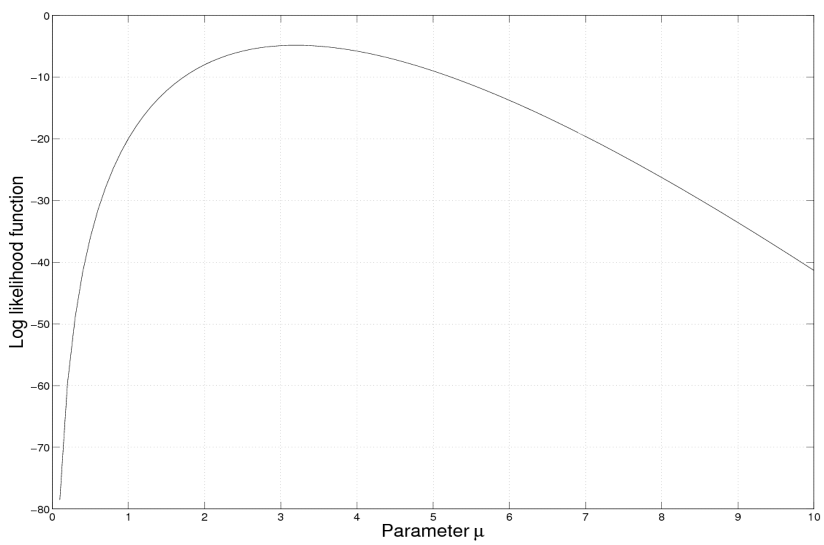
Fasit

1. $k = 1/730$, 0.9507

2. **a**) 0.27, 0.01, 106

3. **a**) 0.472 **b**) 1.545

4. **a**) 0.023,0.976 **b**) 0.05,0.37 **c**) 0.49,2 **d**) 0.37



Log likelihood function
Parameter μ
0
−10
−20
−30
−40
−50
−60
−70
−80
0
1
2
3
4
5
6
7
8
9
10