TMA4245 Statistikk
Eksamen 20. desember 2012

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

Vi skal i denne oppgaven se for oss at vi kaster en mynt flere ganger. Mynten har 0.5 sannsynlighet for utfall «mynt» og 0.5 sannsynlighet for «kron». Vi antar at utfallene av ulike myntkast er uavhengige.

a) Anta at vi kaster mynten 5 ganger.

Hva er sannsynligheten for å få 5 «kron»?

Hva er sannsynligheten for å få 3 «kron»?

Hva er sannsynligheten for å få minst 4 «kron» på rad, det vil si en sekvens (*streak*) av bare «kron» utfall som minst er av lengde 4?

b) Vi kaster mynten 30 ganger. Fordelingen til lengste sekvens med «kron» er vanskelig å regne ut. I stedet kan vi få en datamaskin til å generere 30 stokastiske og uavhengige myntkast. Vi registrerer lengste sekvens av «kron». Prosedyren gjentas B ganger, og resultatet er representativt for fordelingen for lengste sekvens av «kron», se Figur 1.

Anslå sannsynligheten for å få en lengste sekvens på 5 eller 6.

Miriam har fått hjemmelekse å kaste en mynt 30 ganger. Resultatet er som følger, der 0 betyr «mynt» og 1 betyr «kron»:
$(0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1)$.

Læreren mistenker at Miriam har jukset og bare funnet på tallene istedenfor å faktisk kaste en mynt, og læreren vil undersøke dette. Formuler dette som en hypotesetest om lengste sekvens av «kron». Bruk histogrammet i Figur 1 til å svare.

Oppgave 2

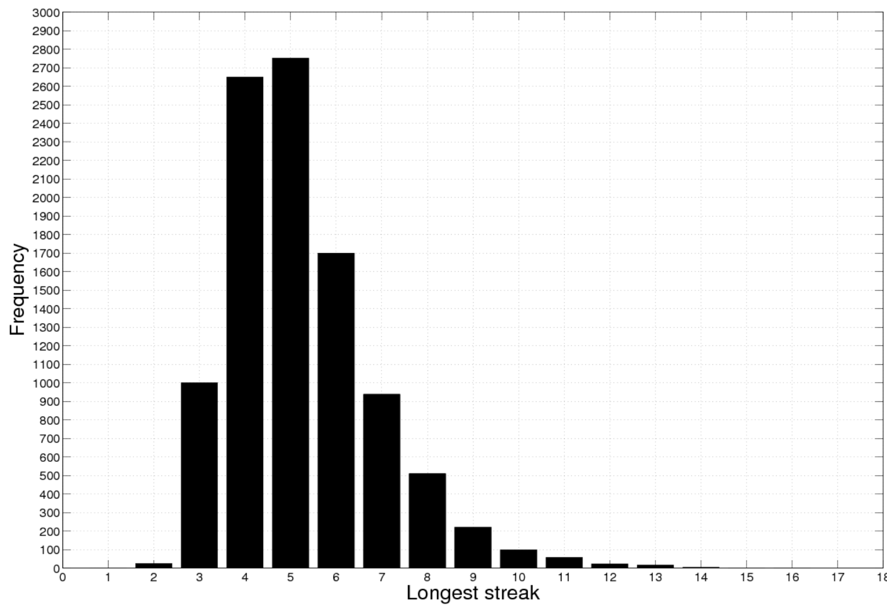
Kjell Inge drar ofte og fisker en times tid i nærheten av der han bor.

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.



Figur 1: Figuren viser et histogram av sekvenser. Disse er resultat av $B = 10000$ gjentak av 30 myntkast. Høyden på stolpene angir i hvor mange av de 10000 forsøkene lengste sekvens av «kron» var en bestemt lengde

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.

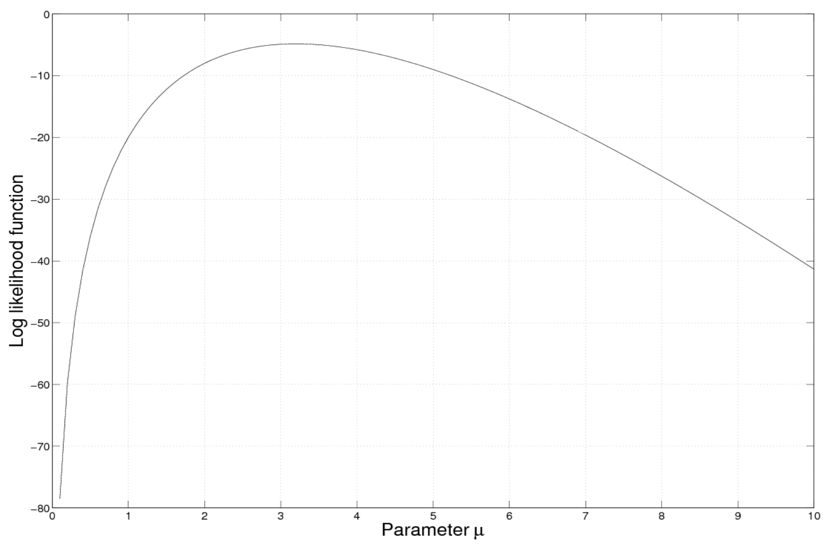
Anta kjent μ. Vis at sannsynlighetsmaksimeringsestimatoren (maximum likelihood estimator) for θ er gitt som

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

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.



Oppgave 3

Figuren viser vinnertidene på 800 m løping for menn i alle Olympiske Leker (OL).

Totalt er det $n = 28$ vinnertider. Vi lar Y_i være vinnertiden i OL nummer i, og x_i årstallet for OL nummer i. Vi antar følgende regresjonsmodell for vinnertidene:

$$Y_i = \alpha + \beta x_i + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma^2).$$

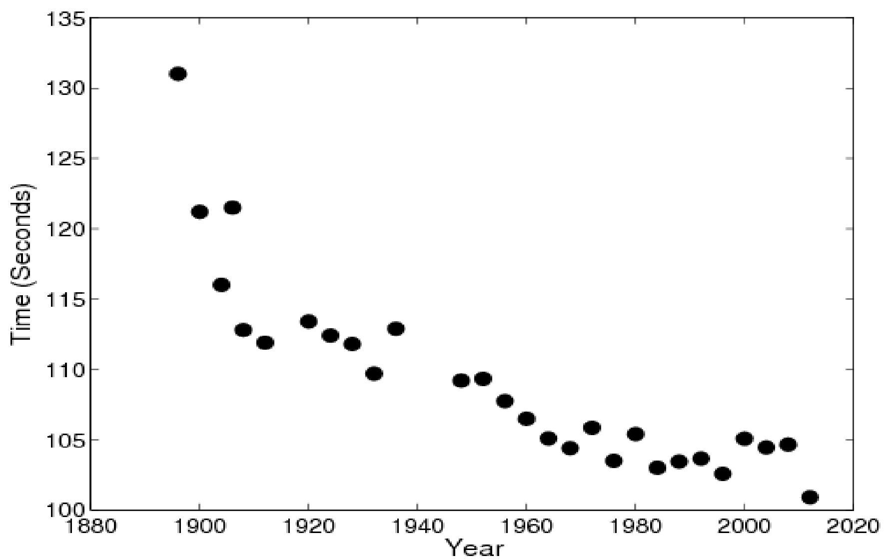
I tillegg antas at støyleddene $\epsilon_1, \ldots, \epsilon_n$ er uavhengige.

a) Gi en kort forklaring av minste kvadraters metode (også kalt minste kvadratsums metode eller method of least squares) for linjetilpasning.

Vis at denne metoden gir estimatorer:

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}, \quad \hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i - n\bar{x}\bar{Y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

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



En alternativ skrivemåte for estimatoren av stigningstallet er $\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})Y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$.

Det oppgis at $\bar{Y} = 109.26$, $\bar{x} = 1954.5$, $\sum_{i=1}^{n}(x_i - \bar{x})Y_i = -5942$ og $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 36517$. Et estimat på variansen til støyleddene er $s^2 = 3.40^2$.

b) Det kan vises at

$$T = \frac{(\hat{\beta} - \beta)}{s/\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \sim t_{n-2}$$

Bruk dette resultatet til å utlede et 95 prosent konfidensintervall for β.

Regn ut konfidensintervallet ved bruk av tall oppgitt over.

Vi vil predikere vinnertiden i neste OL: 2016 i Brasil.

c) Regn ut predikert vinnertid i 2016.

Finn et 95 prosent prediksjonsintervall for vinnertiden i 2016.

d) Bruk modellen til å anslå året da 90-sekundersgrensen brytes, altså det første året da vinnertiden er under 90 sekunder.

Vurder modellantakelsene som gjøres. Hvilke metoder kan brukes for å undersøke antakelsene?

Fasit

1. **a)** 0.031,0.313,0.09 **b)** 0.44, forkast H_0

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

3. **b)** $[-0.199, -0.126]$ **c)** 99.3,$[91.8, 106.7]$ **d)** 2073