TMA4245 Statistikk
Eksamen 20. desember 2012

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

Vi skal no sjå for oss at vi kastar ein mynt fleire gonger. Mynten har 0.5 sannsyn for utfall «mynt» og 0.5 sannsyn for «kron». Vi antar at utfall av ulike myntkast er uavhengige.

a) Anta at vi kastar mynten 5 gonger.

Kva er sannsynet for å få 5 «kron»?

Kva er sannsynet for å få 3 «kron»?

Kva er sannsynet for å få minst 4 «kron» på rad, det vil si ein sekvens (*streak*) av berre «kron» utfall som minst er av lengde 4?

b) Vi kastar mynten 30 gonger. Fordelinga til lengste sekvens med «kron» er vanskelig å rekne ut. I staden instruerer vi datamaskinen til å generere 30 stokastiske og uavhengige myntkast. Vi registrerer lengste sekvens av «kron». Prosedyren vert gjenteke B gonger, og resultata er representative for fordelinga for lengste sekvens av «kron».

Figuren viser eit histogram av sekvensar. Desse er resultat av $B = 10000$ gjentak av 30 myntkast. Høgda på stolpene angir i kor mange av dei 10000 forsøka lengste sekvens av «kron» var ei bestemt lengde.

Anslå sannsynet for å få ein lengste sekvens på 5 eller 6.

Miriam har fått heimelekse å kaste ein mynt 30 gonger. Resultata er som følgjer, 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 mistenkjer at Miriam har juksa og berre funne på talla istaden for å faktisk kaste ein mynt, og læreren vil undersøkje dette. Formuler dette som ein hypotesetest om lengste sekvens av «kron». Bruk histogrammet over til å svare.

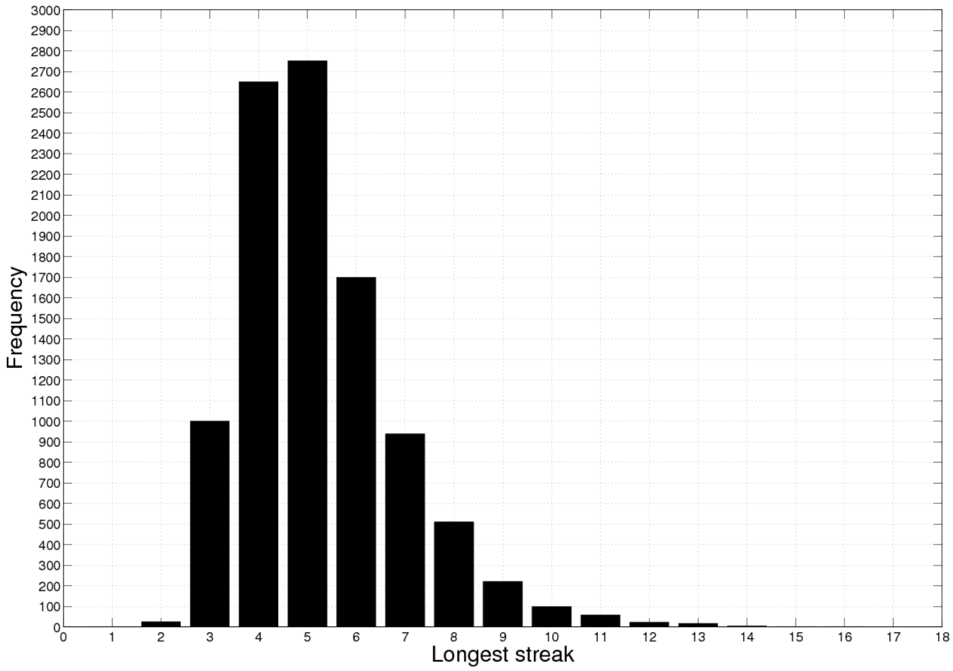
Oppgave 2

Kjell Inge drar ofte og fiskar ein times tid i nærleiken av der han bur.

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

Kva er sannsynet for at ein fisk er meir enn 1000 gram?

Kva er sannsynet for at ein fisk er mellom 500 gram og 1000 gram?



Kjell Inge tenkjer at antal fisk er viktigare enn vekten. Han grubler på fordelinga til antal fisk per tur.

b) La X være antal fisk på ein tur. Alle turar tek 1 time. Vi antar at X er Poisson-fordelt med parameter (forventningsverdi) $\mu = 3$.

Kva er sannsynet for at Kjell Inge ikkje får fisk på ein tur?

Gitt at han får fisk, kva er sannsynet for at han får fleire enn 3 fisk?

Ein alternativ sannsynmodell er som følgjer: Med sannsyn θ får han heilt sikkert 0 fisk. Med sannsyn $1-\theta$ er antal fisk Poisson-fordelt med parameter μ. Punktsannsyn til antal fisk vert:

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

der $I(A) = 1$ om hendinga A inntreff, og $I(A) = 0$ ellers.

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

Kva er sannsynet for at Kjell Inge får fisk på ein tur?

Bruk modellen til å rekne ut forventa antal fisk per tur.

Vi antar no at vi har data frå $n = 20$ uavhengige fisketurar. Av disse vert det $r = 8$ turar med ingen fisk. I dei resterande tolv turane vert det tilsaman 40 fisk. Vi bruker dette talmaterialet til å estimere modellparametrane θ og μ.

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

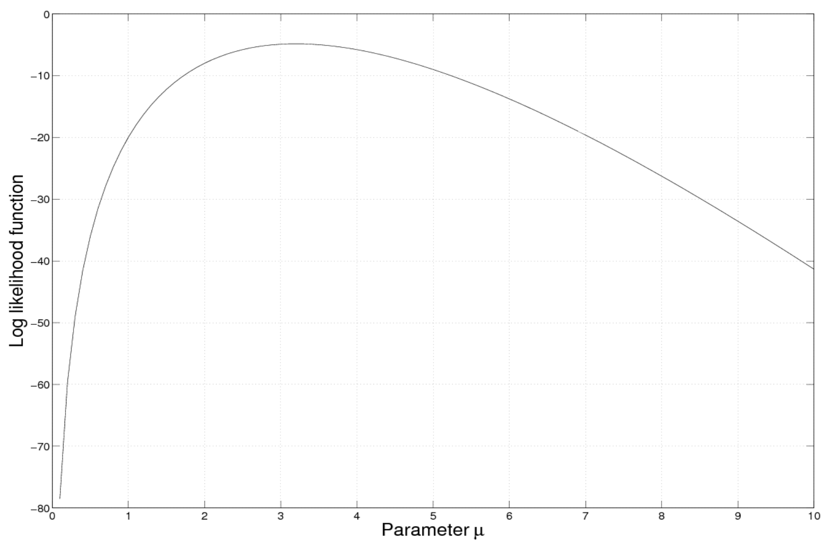
Anta kjent μ. Vis at sannsynmaksimeringsestimatoren (maximum likelihood estimator) for θ er

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

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

Bruk plottet til å finne sannsynmaksimeringsestimatet for μ.

Bruk resultatet til å rekne ut sannsynmaksimeringsestimatet for θ.



Oppgave 3

Figuren viser vinnartidene på 800 m løping for menn i alle Olympiske Leikar (OL).

Totalt er det $n = 28$ vinnartider. Y_i er vinnartiden i OL nummer i, og x_i årstalet for OL nummer i. Vi antar følgjande regresjonsmodell for vinnartidene:

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

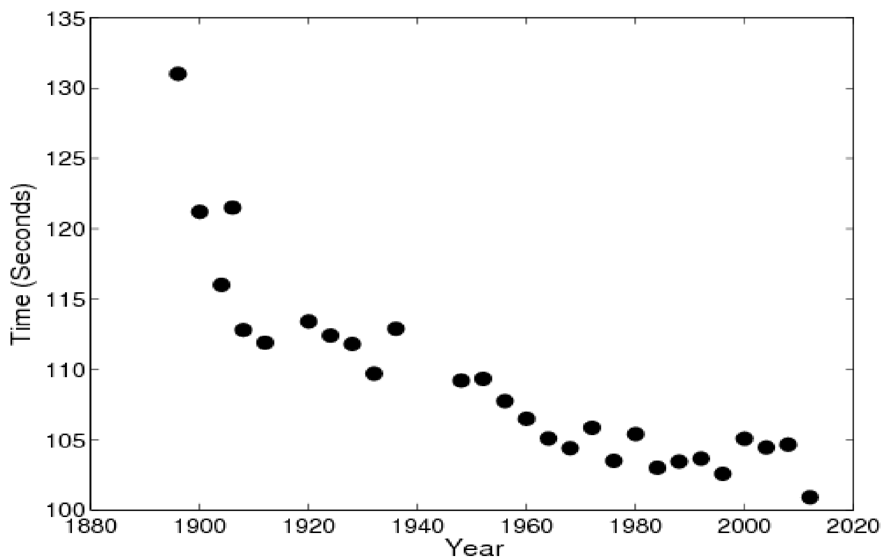
I tillegg antar vi at støyledda $\epsilon_1, \ldots, \epsilon_n$ er uavhengige.

a) Gi ein kort forklaring av minste kvadraters metode (og kalt minste kvadratsums metode eller method of least squares) for linjetilpassing.

Vis at denne metoden gir estimatorar:

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

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



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

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

b) No gjeld 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 utleie eit 95 prosent konfidensintervall for β.

Rekn ut konfidensintervallet ved bruk av tala over.

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

c) Rekn ut predikert vinnartid i 2016.

Finn eit 95 prosent prediksjonsintervall for vinnartida i 2016.

d) Bruk modellen til å anslå året når 90-sekundersgrensen brytast.

Vurder modellantakelsene for regresjonsmodellen. Kva for metoder kan brukast for å undersøkje 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