

Løsningsforslag øving 6, ST1301

Oppgave 1 Løse Euler-Lotka ligningen ved bruk av Newton's metode.

Anta at vi har en organisme med maksimal alder lik n år. Vi ser kun på hunnene i populasjonen. La m_i være antall avkom en hunn i gjennomsnitt får ved alder i (såkalt fekunditet) og la l_i være gjennomsnittlig overlevelse til et avkom opp til alder i . Det kan da vises at populasjonen etter hvert vil vokse tilnærmet eksponentielt med rate r hvor r er løsningen av Euler-Lotka ligningen

$$1 = \sum_{i=1}^n e^{-ri} l_i m_i. \quad (1)$$

Betrakt en organisme hvor m_i og l_i er gitt ved følgende tabell:

Alder i	Fekunditet m_i	Overlevelse l_i
1	0	0.75
2	1.20	0.60
3	1.40	0.48
4	1.03	0.36
5	0.96	0.18

Programmer en funksjon som løser ligning (1) med hensyn på r ved bruk av Newton's metode og som returnerer vekstraten r som funksjonsverdi for gitte fekunditeter og overlevelseshparametere. Hva blir vekstraten til en populasjon med parametere som i tabellen over?

Vi kan skrive ligning (1) på formen

$$f(r) = \sum_{i=1}^n e^{-ri} l_i m_i - 1 = 0. \quad (2)$$

Når vi skal anvende Newton's metode trenger vi den deriverte,

$$f'(r) = - \sum_{i=1}^n i e^{-ri} l_i m_i. \quad (3)$$

Iterasjonsligningen vår blir dermed

$$r' = r + \frac{\sum_{i=1}^n e^{-ri} l_i m_i - 1}{\sum_{i=1}^n i e^{-ri} l_i m_i}. \quad (4)$$

Uttrykket som inngår i iterasjonsligningen kan enkelt beregnes ved bruk av elementvis operasjoner på vektorer og funksjonen `sum` i R:

```

vekstrate <- function(m,l,r=0,tol=1e-6) {
  n <- length(l) # Maksimal alder
  i <- 1:n      # Trenger denne nedenfor
  repeat {
    forrigger <- r # Tar var på denne for å kunne sammenligne
    r <- r + (sum(exp(-r*i)*l*m)-1)/sum(i*exp(-r*i)*l*m)
    if (abs(r-forrigger)<tol) break()
  }
  return(r)
}
> m<-c(0,1.2,1.4,1.03,.96)
> l<-c(.75,.6,.48,.36,.18)
> vekstrate(m,l)
[1] 0.2278773

```

Øk fekunditeten m_i ved alder $i = 2$ med 1% og undersøk hvordan dette endrer vekstraten. Sammenlign dette med hva du får om øker fekunditeten m_i ved alder $i = 5$ tilsvarende. Er vekstraten mest sensitiv for endringer av fekunditetene i lave eller i høye aldersklasser? Om vi betrakter de to endringene som to genetiske varianter, hvilken variant vil da bli mest tallrik om noen generasjoner?

Endring i r som følge av 1% økning i m_2 blir

```

> r<-vekstrate(m,l)
> m[2]<-1.01*m[2]
> vekstrate(m,l)-r
[1] 0.001625896

```

Tilsvarende endring i r om vi øker m_5 blir

```

> m<-c(0,1.2,1.4,1.03,.96)
> m[5]<-1.01*m[5]
> vekstrate(m,l)-r
[1] 0.0001971374

```

altså langt mindre. Vekstraten er m.a.o. langt mer sensitiv til endringer i parametere som bestemmer fekunditet i lave aldersklasser. Genotyper med høy reproduksjon ved lav alder vil altså øke raskere enn genotyper med høy reproduksjon ved høy alder.

Oppgave 2 Vis at

$$\frac{d}{dx}y^x = y^x \ln y. \quad (5)$$

Oppgave 3 Bruk av Newton's metode for å løse estimeringsligning.

Weibull-fordelingen er en mye brukt levetidsfordeling. Sannsynlighetstettheten er gitt ved

$$f_T(t) = \frac{a}{b} \left(\frac{t}{b}\right)^{a-1} \exp\left(-\left(\frac{t}{b}\right)^a\right). \quad (6)$$

Legg merke til at parametriseringen er litt annerledes enn i brukerkurset i statistikk.

Om vi ser på dødsraten definert ved

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(T < t + \Delta t | T > t)}{\Delta t}, \quad (7)$$

altså sannsynlighet per tid for at et individ dør i et lite intervall av lengde Δt , gitt at det er i live ved tidspunkt t , kan det generelt vises at

$$\lambda(t) = \frac{f_T(t)}{1 - F_T(t)}. \quad (8)$$

For Weibull-fordelingen får vi dermed

$$\lambda(t) = \frac{a}{b} \left(\frac{t}{b}\right)^{a-1}. \quad (9)$$

Om dødsraten øker med tiden svarer dette altså til at formparameteren a i Weibullfordelingen er større enn 1. For $a = 1$ ser vi at modellen svarer til vanlig eksponentiell fordeling og for $a < 1$ avtar dødsraten med tiden.

Hos menneske finner en typisk at dødsraten $\lambda(t)$ øker ved høy alder ($a > 1$), såkalt aldring (senescence), i motsetning til hos mange naturlig organismer som typisk lever så kort at aldring sjeldnere forekommer.

Last ned datasettet `spurv.dat` fra hjemmesiden til faget. Last dette inn i R ved å skrive

```
t <- scan("spurv.dat")
```

Datasettet inneholder levetider fra ett års alder til 870 spurv ringmerket på Helgelandskysten i perioden 1993-1997. Lag et histogram over levetidene (med funksjonen `hist`).

Sett opp log-likelihoodfunksjonen til modellen og utled en estimeringsligning for formparameteren a . Anta at skalaparameteren er kjent og lik $b = 2$.

Programmer en funksjon som beregner et estimatet av formparameteren a ved å løse estimeringsligningen ved hjelp av Newton's metode. Husk at $\ln L$ skal deriveres med hensyn på a .

Kontroller svaret ved å lage et plot av $f_T(t)$ basert på estimerte parameterverdier. Sammenlign med histogrammet av de observerte dataene.

Lag også et plot av dødsraten $\lambda(t)$ basert på $b = 2$ og estimatet av a . Hvordan avhenger dødsraten med alder. Aldres spurven?

Beregn til slutt forventet levealder, $E(T)$, basert på $b = 2$ og estimatet av a .

La t_1, t_2, \dots, t_n være de observerte dataene. Likelihoodfunksjonen blir da

$$L(a) = \prod_{i=1}^n \frac{a}{b} \left(\frac{t_i}{b}\right)^{a-1} \exp\left(-\left(\frac{t_i}{b}\right)^a\right), \quad (10)$$

og

$$\ln L(a) = n \ln a - n \ln b + (a-1) \left(\sum_{i=1}^n \ln t_i - n \ln b\right) - \sum_{i=1}^n \left(\frac{t_i}{b}\right)^a. \quad (11)$$

I funksjonens maksimum er

$$\frac{\partial}{\partial a} \ln L(a) = 0, \quad (12)$$

som gir oss estimeringsligningen

$$\frac{n}{a} + \sum_{i=1}^n \ln t_i - n \ln b - \sum_{i=1}^n \ln \left(\frac{t_i}{b}\right) \left(\frac{t_i}{b}\right)^a = 0. \quad (13)$$

Når vi skal løse denne ligningen (som er på formen $f(a) = 0$) ved bruk av Newton's metode trenger vi $f'(a)$, som vi etter litt regning finner at er lik

$$-\frac{n}{a^2} - \sum_{i=1}^n \ln^2 \left(\frac{t_i}{b}\right) \left(\frac{t_i}{b}\right)^a \quad (14)$$

Bruker vi dette for å programmere estimatoren i R får vi

```
ahat <- function(t,b=2,a0=1,tol=1e-6) {
  n <- length(t)
  a <- a0
  forrigea <- a0-1
  while (abs(a-forrigea)>tol) {
```

```

    forrigea <- a
    a <- a - (n/a + sum(log(t)) - n*log(b) -
              sum(log(t/b)*(t/b)^a)
              ) /
              (-n/a^2 - sum(log(t/b)^2*(t/b)^a))
    print(a)
  }
  return(a)
}

```

Estimatet av formparameteren a blir:

```

> ahat(t)
[1] 1.295519
[1] 1.365822
[1] 1.367989
[1] 1.367991
[1] 1.367991
[1] 1.367991

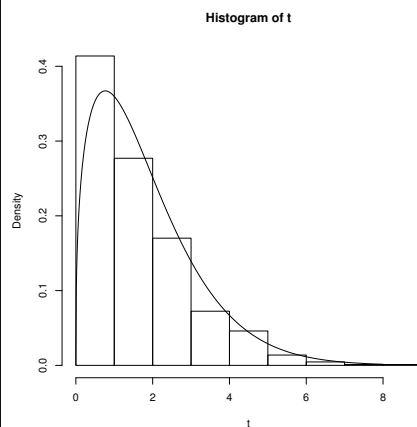
```

Histogram over dataene og tilpasset modell for $f_T(t)$ kan lages slik:

```

> hist(t,prob=T)
> tt <- seq(0,8,.01)
> lines(tt,dweibull(tt,shape=1.367,scale=2))

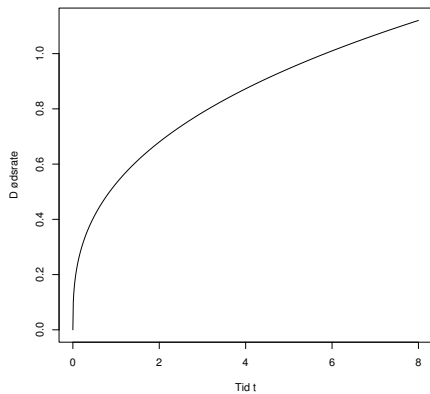
```



Argumentet `prob=T` i kallet til `hist` gjør at arealet under søylene i histogrammet blir 1 slik at histogrammet kan tolkes som et estimat av sannsynlighetstettheten.

Plotter vi dødsraten $\lambda(t)$ basert på estimerte parameter verdier får vi:

```
> plot(tt,1.36/2*(tt/2)^(1.36-1),type="l",  
       xlab="Tid t",ylab="Dødsrate")
```



Forventingen til en Weibullfordelt variabel (for parameteriseringen brukt her) er gitt ved

$$E(X) = b\Gamma\left(1 + \frac{1}{a}\right), \quad (15)$$

som innsatt estimerte verdier i vårt tilfelle blir

```
> 2*gamma(1+1/1.36)  
[1] 1.831600
```

d.v.s. 1.83 år etter fylte 1 år.