Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Bioberegninger, ST1301

Mandag 24. mai 2004

Oppgave 1

a) `qnorm(p=0.5,mean=10)`
Dette gir oss 0.5-kvantilen i en normalfordeling med forventning $\mu = 10$. Symmetrien medfører at denne blir lik μ, altså 10.

b)
```
a <- 1:10
mean(a[a>=4 & a<9])
```
Elementene 4,5,6,7, og 8 blir valgt ut og disse har snitt lik 6.

c) `sum(dbinom(x=0:5,size=5,prob=0.2))`
Dette gir oss alle punktsannsynligheten $P(X = x)$ for $x = 0, 2, \ldots, 5$ for en binomisk fordeling med parameter $n = 5$. Summen av disse blir følgelig 1.

Oppgave 2

a) `minfunksjon(6)`
Default er å ikke multiplisere. De to første elementene i `v` settes lik 1 og 2. De neste elementene blir lik $1 + 2 = 3$, $2 + 3 = 5$, $3 + 5 = 8$, $5 + 8 = 13$. Dette siste elementet returneres som funksjonsverdi.

b) `minfunksjon(n=5,multipliser=T)`
Nå blir de neste elementene $1 \cdot 2 = 2$, $2 \cdot 2 = 4$, og $2 \cdot 4 = 8$ som returneres som funskjonsverdi.

Oppgave 3

a) Vi har at

$$X = F_X^{-1}(U), \tag{1}$$

altså at

$$U = F_X(X) = \frac{1}{1 + \exp(-\frac{X-a}{b})}. \tag{2}$$

Løser vi m.h.p. X får vi

$$\begin{aligned} -\frac{X-a}{b} &= \ln\frac{1-U}{U} \\ X &= a + b\ln\frac{U}{1-U} \end{aligned} \tag{3}$$

b)

```
rlogistisk <- function(n,a,b) {
    u <- runif(n)
    return(a+b*log(u/(1-u))
}
```

c) Vi trenger først sannsynlighetstetthetsfunksjonen. Fra sannsynlighetsregningskurset vet vi at denne er den deriverte av den kumulative tettheten. Bruk av kjerneregel to ganger gir

$$\begin{aligned} f(x) &= \frac{d}{dx}\left(\frac{1}{1+\exp(-\frac{x-a}{b})}\right) \\ &= \frac{\exp(-\frac{x-a}{b})}{b(1+\exp(-\frac{x-a}{b}))^2} \end{aligned} \tag{4}$$

Observerer vi $x_1, x_2, \ldots, x_n$ får vi likelihoodfunksjonen

$$\begin{aligned} L(a,b) &= \prod_{i=1}^{n} \frac{\exp(-\frac{x_i-a}{b})}{b(1+\exp(-\frac{x_i-a}{b}))^2}, \\ &= \frac{\exp(-\frac{1}{b}\sum_{i=1}^{n}(x_i-a))}{b^n\prod_{i=1}^{n}(1+\exp(-\frac{x_i-a}{b}))^2} \end{aligned} \tag{5}$$

og log-likelihood

$$\ln L(a,b) = -\frac{1}{b}\sum_{i=1}^{n}(x_i-a) - n\ln b - 2\sum_{i=1}^{n}\ln\Big(1+\exp(-\frac{x_i-a}{b})\Big) \tag{6}$$

d) F.eks.:

```
lnL <- function(p,x) {
  a <- p[1]
  b <- p[2]
  n <- length(x)
  l <- -1/b*sum(x-a)-n*log(b)-2*sum(1+exp(-(x-a)/b))
  return(l)
}
```

Oppgave 4

a) Vi har at $f(x) = e^x - 2$. Deriverer vi får vi

$$f'(x) = e^{2x}, \tag{7}$$

slik at iterasjonsformelen blir

$$\begin{aligned} x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} \\ &= x_n - \frac{e^{x_n} - 2}{e^{x_n}} \\ &= x_n - 1 + 2e^{-x_n}. \end{aligned} \tag{8}$$

b) Lar vi $x_0 = 2$ og bruker iterasjonsformelen over får vi $x_1 = 0.6931894$,
$x_2 = 1.270671$,
$x_3 = 0.8319573$,
$x_4 = 0.7023506$,
$x_5 = 0.6931894$,
$x_6 = 0.6931472$,
altså svaret med 4 desimalers nøyaktighet etter 6 iterasjoner. Dette stemmer bra med den analytiske løsningen $x = \ln 2 = 0.6931472$

c) For $x_0 = -5$ får vi
$x_1 = 290.8263$,
$x_2 = 289.8263$,
$x_3 = 288.8263$,
$x_4 = 287.8263$.
Vi ser altså at vi først havner langt til høyre for løsningen. Dette skyldes at tangenten til $f(x)$ i punktet $x = x_0$ har et svært lite stingingstall slik at algoritmen fører oss lenger vekk

og til motsatt side av løsningen. Deretter blir konvergensen svært langssom; fordi e^{-x} er tilnærmet lik null store verdier av x ser vi fra (8) at x_n bare reduseres med tilnærmet 1 så lenge x er stor.

Konklusjonen er at valg av startverdi i dette tilfelle har stor betydning for hvor raskt algoritmen vil finne løsningen.