Demo 1

25 October 2023 20:31

1. Sievennä laskennallisesti kätevään muotoon, joka sisältää mahdollisimman vähän kertolaskuja:

   a) $\log\left(\prod_{i=1}^{n} \alpha z_i \exp(-\alpha z_i x_i)\right)$

   b) $\log\left(\theta^n \left(\prod_{i=1}^{n} x_i\right)^{-(\theta+1)}\right)$

   c) $\log\left(\prod_{i=1}^{6}\left[\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right)\right](1-F(c;\mu,\sigma^2))^4\right)$

a) $$\log\left(\prod_{i=1}^{n} \alpha z_i e^{-\alpha z_i x_i}\right) = \sum_{i=1}^{n} \log(\alpha z_i e^{-\alpha z_i x_i})$$
$$= \sum_{i=1}^{n}\left(\log\alpha + \log z_i + \log(e^{-\alpha z_i x_i})\right)$$
$$= n\log\alpha + \sum_{i=1}^{n}\log z_i - \sum_{i=1}^{n} \alpha z_i x_i$$

b) $$\log\left(\theta^n\left(\prod_{i=1}^{n} x_i\right)^{-(\theta+1)}\right) = \log\theta^n + \log\left(\prod_{i=1}^{n} x_i\right)^{-(\theta+1)} = n\log\theta - (\theta+1)\log\left(\prod_{i=1}^{n} x_i\right)$$
$$= n\log\theta - (\theta+1)\sum_{i=1}^{n}\log x_i$$

c) $$\log\Bigg(\underbrace{\prod_{i=1}^{6}\left[\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right)\right]}_{A}(1-F(c;\mu,\sigma^2))^4\Bigg) = \log A + \log\left(1-F(c;\mu,\sigma^2)\right)^4$$
$$= \sum_{i=1}^{6}\log\left[\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right)\right] + 4\log(1-F(c;\mu,\sigma^2))$$
$$= -6\log(\sigma\sqrt{2\pi}) - \sum_{i=1}^{6}\frac{(y_i-\mu)^2}{2\sigma^2} + 4\log(1-F(c;\mu,\sigma^2))$$
$$= -6\left(\log\sigma + \tfrac{1}{2}(\log 2 + \log\pi)\right) - \sum_{i=1}^{6}\frac{(y_i-\mu)^2}{2\sigma^2} + 4\log(1-F(c;\mu,\sigma^2))$$

2. Derivoi seuraavat lausekkeet parametrin θ suhteen ja ratkaise θ, kun derivaatan arvo on 0.

   a) $n\log\theta - (\theta+1)\sum_{i=1}^{n}\log(x_i)$

   b) $\log\theta\sum_{i=1}^{n} x_i - n\theta - \sum_{i=1}^{n}\log(x_i!)$

   c) $k\log\theta + (n-k)\log(1-\theta)$

a) $$\frac{d}{d\theta}\left(n\log\theta - (\theta+1)\sum_{i=1}^{n}\log(x_i)\right) = \frac{n}{\theta} - \sum_{i=1}^{n}\log(x_i)$$

$$\frac{n}{\theta} - \sum_{i=1}^{n}\log(x_i) = 0$$
$$\frac{n}{\theta} = \sum_{i=1}^{n}\log(x_i)$$
$$\theta = \frac{n}{\sum_{i=1}^{n}\log(x_i)}$$

b) $$\frac{d}{d\theta}\left(\log\theta\sum_{i=1}^{n} x_i - n\theta - \sum_{i=1}^{n}\log(x_i!)\right) = \frac{1}{\theta}\sum_{i=1}^{n} x_i - n$$

$$\frac{1}{\theta}\sum_{i=1}^{n} x_i - n = 0$$

$$\frac{1}{\theta}\sum_{i=1}^{n} x_i = n$$

$$\theta = \frac{1}{n}\sum_{i=1}^{n} x_i$$

c) $$\frac{d}{d\theta}\left(k\log\theta + (n-k)\log(1-\theta)\right) = \frac{k}{\theta} - (n-k)\frac{1}{1-\theta} = \frac{k}{\theta} - \frac{n-k}{1-\theta}$$

$$\frac{k}{\theta} - \frac{n-k}{1-\theta} = 0$$

$$\frac{k}{\theta} = \frac{n-k}{1-\theta}$$

$$\theta n - \theta k = k - k\theta$$

$$\theta n = k$$

$$\theta = \frac{k}{n}$$

3. Presidentinvaalien lähestyessä ehdokkaiden kannatusta selvitetään vaaligallupeilla.
   a) Mikä on kiinnostuksen kohteena oleva parametri? Määrittele mahdollisimman tarkasti.
   b) Millaisia tutkimusasetelmia vaaligallupeissa käytetään?
   c) Millaisia vaikeuksia ja epävarmuustekijöitä estimointiin liittyy?
   d) Mitä kaikkea vastaajilta kannattaisi kysyä?

a) Ehdokkaan suhteellinen kannatus. (Ehdokasta äänestäneet / kaikki äänet)
b) Poikkileikkaustutkimuksen asetelmia koska tutkimuksessa kerätään dataa vain yhtenä ajanjaksona.
c) Paikkakuntien välillä kannatukset voivat vaihdella paljonkin joten gallupia tulisi toteuttaa useissa erilaisissa kaupungeissa/kunnissa. Gallupien tulokset voivat myös riippua siitä missä gallup toteutetaan ja mistä ihmiset saavat tietää gallupista. Lisäksi ihmisten suosima ehdokas saattaa vaihtua.
d) Ketä aiot äänestää? Kuka on varavaihtoehtosi?

4. Olkoon $x_1, \dots, x_n$ riippumaton otos Laplace-jakaumasta, jonka odotusarvo on tuntematon parametri μ ja skaalaparametri on 1. Laplace-jakauman tiheysfunktio on tällöin

$$f(x;\mu) = \frac{1}{2}e^{-|x-\mu|}, \quad x \in \mathbb{R}.$$

Kirjoita
   a) uskottavuusfunktio
   b) log-uskottavuusfunktio

a) $$f(x_1, \dots, x_n ; \mu) = \prod_{i=1}^{n} \frac{1}{2} e^{-|x_i - \mu|}$$

$$L(\mu; x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{1}{2} e^{-|x_i - \mu|}$$

b) $$\ell(\mu; x_1, \dots, x_n) = \log \prod_{i=1}^{n} \frac{1}{2} e^{-|x_i-\mu|} = \frac{1}{2}\sum_{i=1}^{n} \log e^{-|x_i - \mu|}$$

$$= -\frac{1}{2}\sum_{i=1}^{n} |x_i - \mu|$$

5. Luo aineisto, joka sisältää 2010 riviä ja muuttujat x ja y, jotka generoidaan seuraavasti:
   - muuttuja x saa arvoja $-1, -0.99, -0.98, \ldots, 0.98, 0.99, 1$ siten, että kukin näistä arvoista esiintyy 10 kertaa,
   - muuttujan y odotusarvo määräytyy mallista $\mathrm{E}(Y_i \mid x_i) = 0.2 + 3x_i$ ja
   - muuttujaan y liittyvä virhetermi on normaalijakautunut siten, että odotusarvo on 0 ja varianssi on 0.8.

https://github.com/JussiKoo/Stat_demos/blob/main/Demo1_T5.R

```
#Here seq-function creates a sequence from -1 to 1 with
step 0.01.
#rep-function repeats each value 10 times.
x <- rep(seq(-1, 1, by=0.01), each=10)

#Here normally distributed error terms are created
#with mean=0 and variance=0.8 (standard deviation is
square root of this)
error <- rnorm(length(x), mean=0, sd=sqrt(0.8))

#Here values for variable y are created
y <- 0.2+3*x+error

#Here values of x and y are added to a dataframe.
t5_data <- data.frame(x,y)
t5_data
```

6. Laadi R-funktio uskottavuusfunktion kuvaajan piirtämiseen. R-funktio saa syötteenään CSV-tiedoston, jonka muoto ja muuttujanimet vastaavat Moodlessa olevaa tiedostoa `phone_talking.csv`. Funktio piirtää uskottavuusfunktion, kun $\theta \in (0, 1)$. Testaa funktiotasi tiedostolla `phone_talking.csv`. (Tällainen kuva esiteltiin luennolla.)

https://github.com/JussiKoo/Stat_demos/blob/main/Demo1_T6.R

```
#function to evaluate likelihood.
L <- function(theta, n, k) {
  choose(n,k)*theta^k*(1-theta)^(n-k)
}

#function that draws the curve
drawLfunction <- function(file) {
  phone_talking <- read.csv(file)

  #n is number of observations.
  #k is the number of occurences.
  n <- length(phone_talking$obs)
  k <- length(which(phone_talking$talking == 1))

  curve(L(x,n,k), from=0, to=1, xlab="theta", ylab="Likelihood", lwd=2)
}

drawLfunction("C:\\kurssit\\TILA141\\phone_talking.csv")
```