Harjoitustehtävät 6

1. Tee luentorungon esimerkin 3 (s. 68) mukainen analyysi eri puolueiden kannatuksista tuoreimman Ylen kannatusmittauksen perusteella, ks.

   https://yle.fi/uutiset/3-11000815

   Voit olettaa, että kyselyssä on käytetty yksinkertaista satunnaisotantaa ja että kannatusjakauma on sama niillä, jotka ovat ilmoittaneet puoluekantansa, kuin niillä, jotka eivät ole. Olettaen priorijakaumaksi Dirichlet(1,1,...,1), määritä posteriorijakauma. Määritä lisäksi 95% posteriorivälit kaikkien puolueiden kannatukselle.

   (Ohje: tee posteriorisimulointia käyttäen R:n rdirichlet-funktiota.)

2. (jatkoa edelliseen) Määritä posterioritodennäköisyydet, että a) Kokoomuksen kannatus on suurempi kuin SDP:n b) Perussuomalaisten kannatus on kaikkein suurin c) Vihreiden kannatus on neljän suurimman puolueen joukossa.

3. (jatkoa edelliseen) Määritä kullekin puolueelle posterioritodennäköisyys, että puolueen kannatus on parantunut verrattuna edelliseen kannatusmittaukseen. Ks. edellinen kannatusmittaus

   https://yle.fi/uutiset/3-10954150

4. Hepatiitti B:tä vastaan kehitetyn rokotteen tehoa tutkittiin Gambiassa. Mitattava vaste y_{ij} on logaritminen vasta-ainemäärä j:nnessä mittauksessa i:nneltä henkilöltä. Malli on seuraava:

$$\begin{aligned}
y_{ij} &\sim \mathrm{N}(\theta_{ij}, \phi) \\
\theta_{ij} &= \alpha_i + \beta_i(\log t_{ij} - \log 730) \\
\alpha_i &\sim \mathrm{N}(\alpha_0, \phi_\alpha) \\
\beta_i &\sim \mathrm{N}(\beta_0, \phi_\beta),
\end{aligned}$$

   missä t_{ij} on henkilöstä i tehdyn j:nnen mittauksen ajankohta rokotuksesta laskien (0–730 vrk). Lisäksi oletetaan, että $p(\alpha_0, \beta_0, \phi_\alpha, \phi_\beta) \propto 1/\phi_\alpha\phi_\beta$.

   a) Selvitä, mitkä tässä hierarkkisessa mallissa ovat parametreja, hyperparametreja ja hyper-hyperparametreja.

b) Piirrä ei-syklinen graafi (DAG).

c) Tulkitse malli. [Keskity mallin osaan $\theta_{ij} = \alpha_i + \beta_i(\log t_{ij} - \log 730)$. Selvitä myös, mitkä ovat henkilövaikutuksia ja mitkä populaatiovaikutuksia (koko mallissa).]

5. Seuraava aineisto kuvaa vuosina 1971–1980 mitattuja sademääriä [mm] marraskuussa ja joulukuussa Yorkissa.

vuosi	marraskuu x	joulukuu y
1971	23.9	41.0
1972	43.3	52.0
1973	36.3	18.7
1974	40.6	55.0
1975	57.0	40.0
1976	52.5	29.2
1977	46.1	51.0
1978	142.0	17.6
1979	112.0	46.6
1980	23.7	57.0

Oletetaan, että sademäärä y noudattaa normaalijakaumaa. Tehtävänä on laatia ennuste tulevan joulukuun sademääräksi, kun juuri havaittu marraskuun sademäärä on $x_{\text{new}} = 46.1$. Määritä posterioriennustejakauman odotusarvo ja 90% ennusteväli. Tutki myös, miten marraskuun sademäärä vaikuttaa joulukuun sademäärään (regressiokerroin).

Aineisto on tiedostona `york_sademaara.txt`.

(Vihje: ks. Esim. 1 luvussa 10.1 ja esimerkkikoodi regExample.r)

6. Toista luvun 10.2 esimerkkiä 2 vastaava analyysi (ks. esimerkkikoodi regExample.r).

a) Säädä priorijakaumia niin, että LD:n posterioriodotusarvosta tulee positiivinen.

b) Kokeile logit-linkin sijasta probit-linkkiä ja katso, muuttuvatko tulokset.

7. Oletetaan, että $y_i|\theta \sim \mathrm{Bin}(1,\theta)$, $i = 1,2,\ldots,n$ ja a priori $\theta \sim$ Tas(0,1). Mikäli havainnot ovat vaihdannaisia, posteriori on

$$p(\theta|y) \propto \theta^s(1-\theta)^{n-s}, \quad s = \sum_{i=1}^{n} y_i\,.$$

Oletetaan, että aineisto ($n = 20$) on

```
1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0
```

Aineistoa tarkasteltaessa herää epäilys, että havainnot ovat autokorreloituneita (pitkä jono peräkkäisiä ykkösiä/nollia).

Testataan riippumattomuus bayesläisittäin. Valitaan testisuureeksi $T(y)$ peräkkäisten $0-1$ ja $1-0$ parien määrät. (Positiivisen autokorrelaation tapauksessa $T(y)$ on pieni, negatiivisen suuri.) Aineistossa $T(y) = 3$.

Laske Bayes-P-arvo riippumattomuushypoteesille. (Esimerkkikoodi tiedostossa testExample.r.)