TAMPEREEN YLIOPISTO Harj. 9
Tilastotiede
Tilastollinen päättely 2 28.11.2014

1. Osoita, että Metropolis-Hastings-algoritmi tuottaa Markovin ketjun, jonka stationaarinen jakauma on $p(x)$ (vrt. luentodiat, s. 98).

2. Estimoi valon nopeus Newcomb-aineiston avulla käyttäen jakaumana t-jakaumaa, jonka vapausasteluku on tuntematon. Tee mallintarkistuksia testisuureiden avulla.

3. Tarkastellaan hierarkista mallia aineistolle, jossa on J ryhmää ja niissä kussakin I havaintoa. (Esim. J koulusta valitaan kustakin I oppilasta ja tehdään heille koulusuoritustesti). Merkitään j. koulun i. havaintoa Y_{ij} ja oletetaan, että $\{Y_{ij}|\theta_j, \tau_1\} \sim N(\theta_j, \tau_1^{-1})$ ja $\{\theta_j|\mu, \tau_2\} \sim N(\mu, \tau_2^{-1})$. Oletetaan parametreille μ, τ_1 ja τ_2 riippumattomat priorijakaumat $\mu \sim N(\mu_0, \tau_0^{-1})$, $\tau_1 \sim \Gamma(\alpha_1, \beta_1)$ ja $\tau_2 \sim \Gamma(\alpha_2, \beta_2)$.

   (a) Kirjoita JAGS-koodi mallin estimoimiseksi

   (b) Estimoi malli käyttäen alla olevaa aineistoa sekä kymmentä 1000 posterioirisimuloinnin ketjua. Tarkista suppeneminen käyttäen aikasarjakuvioita sekä Gelmanin ja Rubinin diagnostiikkoja.

   (1) 3.1, -0.7, 15.0, -1.0, 0.2 (2) 14.9, -4.5, 7.8, -19.1, -8.9
   (3) 4.1, -1.2, 14.9, -1.4, 1.9 (4) 6.8, 7.5, 8.7, 4.6, 1.3

   (c) Estimoi vielä lisäksi malli, joissa ryhmien välillä ei ole vaihteluja ($\theta_j = \theta$, $j = 1, ..., J$) sekä ei-hierarkinen malli, jossa θ_i:t ovat riippumattomia ($\tau_2 = 0$). Määritä kaikille malleille $\bar{D}$, p_D ja DIC, sekä näiden avulla paras malli.

4. Hae aineisto, jossa on annettu kangasrullien pituudet ja vikojen lukumäärät:

   data<-read.table("http://www.stat.columbia.edu/ gelman/book/data/fabric.asc", skip=4,header=TRUE)

   Estimoi JAGS:lla seuraavat mallit: 1) Tavallinen Poissonin regressio log-linkillä. 2) Tavallinen Poissonin malli, jossa vikojen lukumäärän odotusarvo on suoraan verrannollinen rullan pituuteen. 3) Poissonin regressio log-linkillä, kun otetaan ylihajonta (rullien välinen vaihtelu) huomioon. 4) Yleistetty lineaarinen malli log-linkillä käyttäen negatiivista binomijakaumaa. Vertaile mallien hyvyyttä suureiden $\bar{D}$, p_D ja DIC avulla.

5. Dobson (1983) analysoi binomiaalista annos-vaste-dataa, joka sisältää kuolleiden kovakuoriasten lukumäärät, sen jälkeen kun ne on altistettu 5 tunniksi hiilidisulfidille (rikkihiilelle) $n = 8$ eri konsentraatiolla, ks. aineisto

   http://mathstat.helsinki.fi/openbugs/Examples/Beetles.html

   a) Estimoi binomiaalinen regressiomalli, jossa kuoleman todennäköisyyttä selitetään konsentraation avulla, käyttäen logit-, probit- ja komplementääristä log-log-linkkiä. Koska posteriorijakaumassa on voimakas korrelaatio parametrien välillä, kannattaa käyttää selittäjänä keskistettyä muuttujaa x[j]-mean(x[]).

   b) Vertaa malleja suureiden $\bar{D}$ ja DIC avulla.

   c) Piirrä kussakin tapauksessa annos-vaste-käyrä ennusteväleineen ja raakaestimaatteineen $\hat{\theta}_j = y_j/n_j$, vrt. luentodiojen esimerkki.