Bayes-päättely, 2. harjoitukset (1.–2.2.2017)

1. Jatkoa edellisen viikon tehtävälle 3. Multinomijakauma Dirichlet-priorilla.

(a) Osoita, että kun

$$p(\mathbf{y}|\boldsymbol{\theta}) = \binom{n}{y_1 \dots y_d} \prod_{i=1}^{d} \theta_i^{y_i},$$

ja

$$p(\boldsymbol{\theta}) = \frac{\Gamma(\sum_{i=1}^{d} \alpha_i)}{\prod_{i=1}^{d} \Gamma(\alpha_i)} \prod_{i=1}^{d} \theta_i^{\alpha_i - 1},$$

satunnaisvektorin $\mathbf{Y}$ prioriennustejakauma $p(\mathbf{y})$ on ns. Dirichlet-multinomi-jakauma parametreilla n ja $\boldsymbol{\alpha}$, eli että

$$p(\mathbf{y}) = \frac{n!\,\Gamma(\alpha_0)}{\Gamma(\alpha_0 + n)} \prod_{i=1}^{d} \frac{\Gamma(\alpha_i + y_i)}{y_i!\,\Gamma(\alpha_i)},$$

missä

$$\alpha_0 = \sum_{i=1}^{d} \alpha_i.$$

(b) Perustele, mikä on posterioriennustejakauma $p(\tilde{\mathbf{y}}|\mathbf{y})$ havainnolle $\tilde{\mathbf{Y}} \sim \text{Multinom}(m, \boldsymbol{\theta})$ multinomijakaumasta samalla parametrilla, mutta eri otoskoolla m, eli jonka jakauma ehdolla parametri on

$$p(\tilde{\mathbf{y}}|\boldsymbol{\theta}) = \binom{m}{\tilde{y}_1 \dots \tilde{y}_d} \prod_{i=1}^{d} \theta_i^{\tilde{y}_i},$$

kun oletetaan, että

$$\tilde{\mathbf{Y}} \perp\!\!\!\perp \mathbf{Y} \,|\, \boldsymbol{\theta}.$$

2. Jatkoa luentojen esimerkkiin 1.5. Ennen kuin meillä on havaintoja nastanheitosta, voisi olla järkevää olettaa, että kumpikin vaihtoehto on yhtä todennäköinen, eli että nasta voi pudota yhtä hyvin kanta ylöspäin kuin kanta alaspäin. Nyt pitää vielä määrittää priorijakauman ”vahvuus”, joka kertoo, kuinka paljon priori vaikuttaa posteriorijakaumaan.

Kokeillaan kolmea erilaista priorijakaumaa parametrille θ:

- ”Heikko” priori $\text{Beta}(0.01, 0.01)$,
- Tasajakauma $\text{Beta}(1, 1)$,
- ”Vahva” priori $\text{Beta}(5, 5)$.

(a) Piirrä `R`:llä näiden priorijakaumien tiheysfunktoiden kuvaajat samaan kuvaan eri väreillä.

(b) Etsi nasta. Heitä sitä 3 kertaa, ja kirjaa ylös montako kertaa nasta laskeutui kanta alaspäin. Nyt sinulla on siis havainto jakaumasta $\text{Bin}(3, \theta)$.

Piirrä uuteen kuvaan parametrin θ posteriorijakaumien tiheysfunktioiden kuvaajat edellämainituilla prioreilla ja havaitsemallasi aineistolla.

Laske myös aineistosi perusteella kustakin posteriorijakaumasta todennäköisyys, että nasta laskeutuu todennäköisemmin kanta ylös- kuin alaspäin, eli $P(\theta < 0.5 \,|\, Y = y)$.

Mitä eroja huomaat tuloksissa eri priorijakaumien välillä, ja miten selittäisit eroja?

(c) Heitä nastaa vielä 22 kertaa lisää, ja kirjaa tulokset ylös. Nyt sinulla on 25 heiton aineisto, eli havainto jakaumasta $\mathrm{Bin}(25, \theta)$. Toista b-kohdan tarkastelut. Miten tuloksesi eroavat b-kohdan tuloksista, ja miten selittäisit niitä?

(d) Kirjaa 25 nastanheiton tuloksesi laskuharjoituslistan erilliseen sarakkeeseen!

3. Jatkoa edelliseen tehtävään.

(a) Piirrä prioriennustejakaumien[1] pistetodennäköisyysfunktioiden kuvaajat[2] satunnaismuuttujalle $\tilde{Y} \sim \mathrm{Bin}(10, \theta)$ edellisen tehtävän prioreilla samaan kuvaan. Miten eri priorit painottavat $\tilde{Y}$:n eri arvoja?

(b) Piirrä $\tilde{Y}$:n posterioriennustejakaumien $p(\tilde{y}|y)$ kuvaajat edellisen tehtävän prioreilla ja edellisen tehtävän (b)-kohdassa havaitsemallasi aineistolla uuteen kuvaan. Piirrä lisäksi binomijakauman $\mathrm{Bin}(10, \hat{\theta})$, missä $\hat{\theta} = \frac{y}{3}$, on parametrin θ suurimman uskottavuuden estimaatti, jakauman pistetodennäköisyysjakauman kuvaaja. Millä priorijakaumista posterioriennustejakauma on lähimpänä binomijakaumaa, jonka parametriksi on sijoitettu SU-estimaatti?

(c) Toista edellinen kohta edellisen tehtävän (c)-kohdassa havaitsemallasi aineistolla (nyt SU-estimaatti on tietenkin $\frac{y}{25}$). Miten suurempi aineisto vaikuttaa posterioriennustejakaumaan?

4. Jatkoa luentojen esimerkkiin 2.3. Sama jakauma, mutta n havaintoa. Oletetaan riippumattomasti normaalijakautuneet satunnaismuuttujat

$$Y_1, \ldots, Y_n \sim N(\theta, \sigma_0^2), \quad Y_1, \ldots, Y_n \perp\!\!\!\perp \mid \theta,$$

missä varianssi $\sigma_0^2 \in (0, \infty)$ on tunnettu vakio.

Osoita samankaltaisella laskulla kuin luentojen esimerkissä, että priorijakaumalla $\theta \sim N(\mu_0, \tau_0^2)$ parametrin θ posteriorijakauma on $N(\mu_n, \tau_n^2)$, missä

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma_0^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma_0^2}},$$

ja

$$\frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma_0^2}.$$

[1] Beta-binomijakauman pistetodennäköisyysfunktiota ei löydy valmiina `R`:stä, mutta voit kirjoittaa sen helposti itse.

[2] Pistetodennäköisyysfunktioita piirrettäessä kannattaa käyttää `plot`-funktiolle argumenttia `type = 'b'`, joka yhdistää pisteet viivoilla.