# Harjoitustehtävät 3

1. Oletetaan, että otokselle $y_1, ..., y_n$ kukin $(y_i|\theta) \sim Tas(0, \theta)$ ja $y_i$:t ovat riippumattomia ehdolla $\theta$. Tällöin otoksen uskottavuus

$$p(y_1, ..., y_n|\theta) = 1/\theta^n,$$

kun $t_n = \max\{y_1, ..., y_n\} < \theta$ ja 0, muutoin. Olkoon priori $p(\theta) \propto 1/\theta, \theta > 0$ (epäaito priori, jota voi käyttää, jos posteriori on aito).

a) Laske posteriori $p(\theta|y_1, ..., y_n)$.

b) Oletetaan, että data on $(n = 6)$ 1.11,1.86,1.96,1.62,1.28,1.82. Laske posterioritodennäköisyys $P(\theta > 2.5|y)$.

2. Oletetaan, että $y_i|\theta \sim N(\theta, v),\ i = 1, ..., n$ ja havainnot ovat ehdollisesti riippumattomat ehdolla $\theta$. Varianssi $v$ oletetaan tunnetuksi. Olkoon priorijakauma $N(m_0, w_0)$. Tällöin posteriorijakauma on $N(m_1, v_1)$ (ks. parametrit luentomateriaalista). Laske jakauman $p(\tilde{y}|y)$ odotusarvo ja varianssi käyttäen hyväksi kaavoja

$$\begin{aligned}
\mathrm{E}(Y|X) &= \mathrm{E}[\mathrm{E}(Y|X, Z)|X], \\
\mathrm{Var}(Y|X) &= \mathrm{Var}[\mathrm{E}(Y|X, Z)|X] + \mathrm{E}[\mathrm{Var}(Y|X, Z)|X].
\end{aligned}$$

3. Puolueen kannatuksen kunnallisvaaleissa on kyselytutkimuksen perusteella arvioitu olevan 19.0% keskivirheen (estimaattorin hajonnan) ollessa 1.5%.

   a) Määritä parametrit $\alpha$ ja $\beta$ betajakaumalle, joka kuvaa puolueen ääniosuuden $\theta$ priorijakaumaa kyselyn antaman tiedon mukaisesti.

   b) Mikä on kyselyn otoskoko?

   c) Toisessa pari päivää myöhemmin tehdyssä kyselyssä 220 ihmistä 1000 vastanneesta ilmoitti äänestävänsä puoluetta. Käytä ensimmäistä kyselyä priorina ja määritä posteriorijakauma puolueen ääniosuudelle. Mikä on kannatuksen odotusarvo ja keskihajonta?

   d) Muuttuuko posteriorijakauma, jos kyselyjen järjestys on päinvastainen?

   e) Valitaan vielä satunnaisesti uusi vastaaja. Mikä on posterioriennustetodennäköisyys, että hän on puolueen kannattaja?

4. Bakteerinäyte on jaettu kuuteen viljelyyn, joista havaitaan seuraavat lukumäärät: 74, 20, 40, 58, 63, 56 (=311). Oletetaan prioriksi Gamma$(300, 5)$. Laske sen odotusarvo ja varianssi. Oletetaan, että $y_i \sim$ Poisson$(\theta)$ ja havainnot ovat ehdollisesti riippumattomia ehdolla $\theta$. Laske posteriori ja kuvaa sitä sopivilla tunnusluvuilla. Laske todennäköisyys $P(\theta \geq 60|y)$. Voit laskea todennäköisyyden R:llä.

5. Laske 95% posterioriennusteväli kuolemaan johtavien liikenneonnettomuuksien määrälle Helsingissä aikavälillä 1.10.2019–30.9.2020. Tehtävään sisältyy priorijakauman ja mallin määrittäminen, aineiston etsiminen ja posteriorijakauman laskeminen.

6. Oletetaan, että $\theta \sim \text{Beta}(\alpha, \beta)$. Laske $\text{Var}(\theta)$.

7. Johda negatiivisen binomijakauman odotusarvo ja varianssi käyttäen hyväksi kaavoja

$$\begin{aligned} \text{E}(Y) &= \text{E}[\text{E}(Y|X)] \\ \text{Var}(Y) &= \text{Var}[\text{E}(Y|X)] + \text{E}[\text{Var}(Y|X)] \end{aligned}$$

sekä tietoa, että jakauma on gamma-jakaumalla painotettu Poisson-jakaumien sekoitusjakauma.

8. Oletetaan, että havainnot $y_1, ..., y_n$ noudattavat Poisson$(\theta)$-jakaumaa ja ovat riippumattomia ehdolla $\theta$. Lisäksi oletetaan, että $\theta \sim \text{Gamma}(\alpha, \beta)$. Osoita, että uuden havainnon $\tilde{y}$ posterioriennustejakauma on

$$\tilde{y}|y \sim \text{NegBin}(\alpha + s, \beta + n),$$

missä $s = \sum y_i$.