5B Bayesläinen päättely

Tuntitehtävät

5B1 (Kadonnut lentokone.) Seuraavantyyppistä tilastollista hakumenetelmää on käytetty mm. yritettäessä paikantaa 8.3.2014 kadonnutta Malaysia Airlinesin lentoa MH370. Etsintöjen kohteena oleva alue on jaettu neljään suureen ruutuun. Lennon taustatietojen pohjalta arvellaan lentokoneen sijaitsevan ruudussa 1 tn:llä 50%, ruudussa 2 tn:llä 30%, ruudussa 3 tn:llä 10% ja ruudussa 4 tn:llä 10%. Olosuhteista johtuen arvellaan, että yksittäiseen ruutuun kohdistettu etsintäyritys epäonnistuu löytämään ruudussa sijaitsevan koneen tn:llä 70%, aiemmista etsintäyrityksistä riippumatta.

(a) Mallinna lentokoneen sijaintia satunnaismuuttujana $\Theta \in \{1, 2, 3, 4\}$ ja määritä sijainnin priorijakauma.

(b) Etsintäryhmä tutkii ensiksi ruudun 1. Merkitään $X_1 = 1$ jos kone löytyy ja $X_1 = 0$ muuten. Määritä Θ:n uskottavuusfunktio havainnolle $X_1 = 0$.

(c) Konetta ei löydetä ensimmäisellä hakuyrityksellä. Määritä koneen sijainnin posteriorijakauma tämän havainnon suhteen. Kannattaako ruudusta 1 etsiä uudelleen?

(d) Etsintäryhmä päättää tutkia ruudun 1 uudelleen. Tämänkin haun lopputulos on, että konetta ei löytynyt. Määritä lentokoneen sijainnin posteriorijakauma tämän datan valossa. Mistä etsintäryhmän kannattaa seuraavaksi etsiä?

5B2 (Bussien vuoroväli.) Eri matkustuspäivien odotusajat (min) eräällä bussipysäkillä ovat toisistaan riippumattomia ja noudattavat jatkuvan välin $[0, \theta]$ tasajakaumaa tiheysfunktiona

$$f(t \,|\, \theta) \;=\; \begin{cases} \frac{1}{\theta}, & 0 < t < \theta, \\ 0, & \text{muuten.} \end{cases}$$

Tuntematon parametri θ voidaan tulkita bussien vuorovälinä kyseisellä pysäkillä.

Taustatietojensa pohjalta Tuomas olettaa, että bussien vuoroväli on vähintään 1 min. Hän tulkitsee tuntemattoman parametrin satunnaismuuttujaksi Θ ja valitsee priorijakauman tiheysfunktioksi

$$f_0(\theta) \;=\; \begin{cases} 2\theta^{-3}, & \theta \geq 1, \\ 0, & \text{muuten.} \end{cases}$$

Tuomas päättää käyttää estimaattina Θ:n posteriorijakauman odotusarvoa. Laske tämän estimaatin arvo, kun viiden matkustuspäivän aikana on havaittu odotusajat $x = (2, 6, 3, 4, 8)$.

Kotitehtävät

5B3 (Lehtien pituudet.) Botanisti arvelee, että tietyn kasvilajin lehtien pituus (senttimetreinä) noudattaa normaalijakaumaa odotusarvona θ ja keskihajontana $\sigma = 2$. Hän olettaa lisäksi, että tuntematon odotusarvo Θ noudattaa normaalijakaumaa odotusarvona $\mu_0 = 10$ keskihajontana $\sigma_0 = 1$. Botanisti mittasi viiden kyseisen lajin kasvin lehtien pituudet ja sai tulokseksi $x = (10, 14, 11, 17, 8)$ (lehtien pituudet voidaan olettaa toisistaan riippumattomiksi). Määritä tehtyjen havaintojen valossa:

(a) Satunnaismuuttujan Θ posteriorijakauman odotusarvo.

(b) Väli, joka sisältää tuntemattoman odotusarvon todellisen arvon 90% todennäköisyydellä.

(**Vihje:** Ylläolevin oletuksin tuntemattoman odotusarvon posteriorijakauma on normaalijakauma, jonka odotusarvo ja keskihajonta löytyvät luentomonisteen Lauseesta 9.5.)

5B4 (Viestin alkuperä.) Erään Twitter-viestin väitetään olevan Ullan kirjoittama. Viestin sisällön pohjalta on syytä epäillä, että 10% todennäköisyydellä viestin kirjoittaja onkin Ullan ystävä Venla.

Ullan viesteissä tiedetään esiintyvän kirjoitusvirheitä keskimäärin 2.1 per rivi ja Venlan viesteissä keskimäärin 0.3 per rivi. Kirjoitusvirheiden lukumäärän riviä kohden X oletetaan noudattavan Poisson-jakaumaa

$$f(k \mid \theta) = \frac{\theta^k}{k!} e^{-\theta}, \quad k = 0, 1, 2, \ldots,$$

tuntemattomalla odotusarvoparametrilla $\theta > 0$.

Olkoon $\Theta \in \{2.1, 0.3\}$ satunnaismuuttuja joka kuvaa mahdollisia odotusarvoja θ (eli epäsuorasti kyseessä olevia henkilöitä).

(a) Viestin ensimmäiseltä riviltä ei löydy kirjoitusvirheitä. Määritä satunnaismuuttujan Θ posteriorijakauma tämän datan valossa.

(b) Viestin toiselta riviltä löydetään 5 kirjoitusvirhettä. Määritä satunnaismuuttujan Θ posteriorijakauma molempien rivien havaintojen valossa.