Johdatus tilastolliseen päättelyyn, kesä 2015
Harjoitus 5 (11.6.)

Tehtävät 1–4 käsittelevät Bayes-päättelyä (monisteen luku 10). Tärkeimmät niistä ovat tehtävät 1–3. Tehtävä 4 liittyy liittojakaumien käsitteeseen (jakso 10.5), joka ei kuulu kurssin ydinsisältöön.

1. Palataan edellisten harjoitusten viimeiseen tehtävään, jossa mikropiirin viallisuuden perusteella tehtiin päätelmiä siitä, millä linjalla se on valmistettu. Sijoita tämä esimerkki bayesiläisen päättelyn yleiseen viitekehykseen siinä mielessä kuin se on kuvattu esim. monisteen sivulla 116. Mikä olisi parametri ja sen priorijakauma? Entä uskottavuusfunktio? Mikä on posteriorijakauma?

2. Kulhossa on 4 palloa, joista θ on valkoisia ja loput mustia. Pekalla ei ole mitään tietoa siitä, miten pallojen värit ovat määräytyneet, joten hänen ennakkokäsityksensä mukaan kaikki vaihtoehdot θ:n arvolle (0,1,2,3,4) ovat yhtä todennäköisiä. Hän nostaa korista umpimähkään ja palauttaen kaksi palloa: kumpikin niistä on musta. (Tätä satunnaiskoetta kuvaava malli on esitetty monisteen jaksossa 10.2.)

Esitä arvot luettelemalla ja halutessasi myös kaavalla i) Pekan priorijakauma, ii) uskottavuusfunktio, iii) Pekan posteriorijakauma. Mikä on Pekan mielestä havaintojen teon jälkeen todennäköisin θ:n arvo?

3. Jatkoa edelliseen tehtävään. Maija seuraa sivusta Pekan koetta. Samalla hän tietää, että pallot päätyivät kulhoon seuraavasti: Oli 4 samanlaista valkoista palloa. Kunkin pallon kohdalla heitettiin harhatonta lanttia, ja mikäli saatiin kruunu, pallo värjättiin mustaksi. Sitten pallot pantiin kulhoon. Mitkä ovat Maijan priori- ja posteriorijakauma (luettele arvot)? Mikä on hänen mielestään havaintojen teon jälkeen todennäköisin θ:n arvo?

4. Tilastollinen malli aineistolle $\mathbf{y} = (y_1, \ldots, y_n)$ on satunnaisotos normaalijakaumasta $N(\theta, \sigma^2)$ eli

$$f(\mathbf{y}|\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right\},$$

jossa $\theta \in \mathbb{R}$ on parametri ja $\sigma^2 > 0$ on jokin tunnettu luku. Priorijakaumaksi $p(\theta)$ valitaan normaalijakauma $N(0, \sigma_0^2)$, jossa $\sigma_0^2 > 0$ on tunnettu luku. Osoita, että posteriorijakauma $p(\theta|\mathbf{y})$ on eräs normaalijakauma $N(\mu_1, \sigma_1^2)$ ja lausu μ_1 sekä σ_1^2 lukujen σ_0^2, σ^2, n ja $\bar{y}$ avulla.

Tämä lasku osoittaa, että normaalijakauma on normaalimallin uskottavuuden *liittopriori* (vrt. monisteen jakso 10.5). Tulos on voimassa myös yleisemmälle priorille $N(\mu_0, \sigma_0^2)$; oletus $\mu_0 = 0$ tekee vain laskun hieman yksinkertaisemmaksi.

Apu. Tässä lienee kätevää edetä verrannollisuustarkastelun $p(\theta|\mathbf{y}) \propto p(\theta)f(\mathbf{y}|\theta)$ kautta (vrt. monisteen s. 123).

5. Kertaustehtävä suurimman uskottavuuden menetelmästä. Tilastollinen malli on n-kertainen satunnaisotos eksponenttijakaumasta $\mathrm{Exp}(\lambda)$, jonka tiheysfunktio on $g(x;\lambda) = \lambda e^{-\lambda x}$, $x > 0$, ja jossa $\lambda > 0$ on parametri. Lyhyesti kirjoitettuna siis $Y_1, \ldots, Y_n \sim \mathrm{Exp}(\lambda) \perp\!\!\!\perp$.

Muodosta havaintojen yhteistiheysfunktio ja aineistoa $\mathbf{y} = (y_1, \ldots, y_n)$ vastaava uskottavuusfunktio. Johda uskottavuusfunktion logaritmia tutkimalla ja huolellisesti perustellen λ:n suurimman uskottavuuden estimaatti.

6. Pohdittavaksi ja harjoituksissa keskusteltavaksi:

a) Sukulaisesi, joka ei ole opiskellut lainkaan tilastotiedettä, pyytää sinua kertomaan yhdellä virkkeellä, mistä tilastollisessa päättelyssä on kysymys. Miten vastaisit hänelle tämän JTP-kurssin pohjalta?

b) Luettele kolme mielestäsi tärkeintä (tai ainakin mielenkiintoisinta) tällä kurssilla opittua tilastollisen päättelyn käsitettä tai menetelmää.

Kurssikoe pidetään ma 15.6. klo 12.00–14.00 salissa CK112. Kokeessa kuulusteltavat asiat käyvät ilmi kurssin kotisivulla olevasta luentopäiväkirjasta. Kokeessa saa käyttää laskimen lisäksi käsinkirjoitettua kaksipuolista ja A4-kokoista ”lunttilappua”. Omia kaavakirjoja ja taulukoita ei saa käyttää.