**Johdatus tilastolliseen päättelyyn, kevät 2014**
**Harjoitus 6 (28. 4. – 2. 5.)**

1. Palataan edellisten harjoitusten viimeiseen tehtävään, jossa hotelliasiakkaan alushousujen värillisyyden perusteella tehtiin päätelmiä asiakkaan kansalaisuudesta. Sijoitetaan 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? (Voi olla järkevää koodata alushousujen ominaisuudet ”väritön” ja ”värillinen” sekä asiakkaan kansalaisuus joillakin lukuarvoilla.)

2. Kulhossa on 4 palloa, joista $\theta$ on valkoisia ja loput mustia. Pekalla ei ole mitään tietoa siitä, miten pallojen värit ovat määräytyneet, joten hänen ennakkokäsityksensä mukaan kaikki vaihtoehdot $\theta$:n arvolle (0,1,2,3,4) ovat yhtä todennäköisiä. Hän nostaa korista umpimähkään ja palauttaen kolme palloa ja saa tulokseksi värit valkoinen, musta ja valkoinen. (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 $\theta$: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 $\theta$:n arvo?

4. Tilastollinen malli aineistolle $\mathbf{y} = (y_1, \dots, y_n)$ 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 $\mu_1$ sekä $\sigma_1^2$ lukujen $\sigma_0^2$, $\sigma^2$, $n$ 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. 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.

KÄÄNNÄ!

**Vapunpäivänä** to 1. 5. ei pidetä harjoituksia. Torstain ryhmiin osallistuneita pyydetään ensisijaisesti vierailemaan muissa ryhmissä. Jos tämä ei ole mahdollista, voi ratkaisut lähettää emailissa harjoituksenpitäjälle (*henri.c.karttunen@helsinki.fi* tai *ville.k.makinen@helsinki.fi*) viimeistään pe 2. 5. aamupäivällä.

**Kurssikoe** pidetään ma 5. 5. klo 13.00–15.00 Exactumin auditorioissa. Tarkka koealue ja muuta kokeeseen liittyvää kerrotaan piakkoin kurssin www-sivulla.