Tilastollinen päättely, syksy 2014 – kevät 2015
Harjoitus 13 (24. ja 26. 2. 2015)

1. (Vrt. monisteen teht. 6.1.) Olkoot $Y_1, \ldots, Y_{25} \sim N(\mu, \sigma^2) \perp\!\!\!\perp$ ja $S^2 = \sum_{i=1}^{25}(Y_i - \overline{Y})^2/24$. Palauta todennäköisyyslaskennan kurssilta mieleen, miten muuttuja S^2/σ^2 on jakautunut; erityisesti sen jakauma ei riipu parametreista μ ja σ^2. Etsi tämän avulla keskihajonnalle σ ylemmät 95 %:n ja 99 %:n luottamusrajat eli luottamusvälit muotoa $]0, b[$, kun on havaittu $s = 10$.

2. (Monisteen teht. 6.2.) Olkoot $Y_1 \perp\!\!\!\perp Y_2$ ja $Y_1 \sim N(\mu_1, 1)$ sekä $Y_2 \sim N(\mu_2, 1)$. Etsi luvut $a, b > 0$ siten, että

$$P\{|Y_1 - \mu_1| \leq a,\ |Y_2 - \mu_2| \leq a\} = 0.95,$$
$$P\{(Y_1 - \mu_1)^2 + (Y_2 - \mu_2)^2 \leq b^2\} = 0.95.$$

Aineisto on $(y_1, y_2) = (1, 0.5)$. Mitkä kaksi 95 %:n luottamusjoukkoa saadaan yo. yhtälöiden perusteella parametriparille (μ_1, μ_2)? Piirrä kuva. Kumpi luottamusjoukoista on mielestäsi parempi? [*Ohje*. Tarvitset jakaumien $N(0, 1)$ ja χ^2_2 taulukoita.]

3. (Tämä tehtävä on monisteen esimerkin 6.1.5 diskreetti muunnelma ja havainnollistaa luottamusjoukon käsitteen tulkinnallista vaikeutta.)

Olkoot Y_1 ja Y_2 riippumattomia satunnaismuuttujia, joiden kummankin arvojoukko on $\{\theta, \theta + 1\}$ ja pistetodennäköisyydet $P(Y_i = \theta) = P(Y_i = \theta + 1) = \frac{1}{2}$, jossa $\theta \in \mathbb{R}$ on tuntematon parametri. Kun havainnot ovat y_1 ja y_2, määritellään $A(y_1, y_2) = \{y_{(1)}\}$, jossa $y_{(1)} = \min(y_1, y_2)$. Osoita:

a) Jos $Y_1 \neq Y_2$, niin $\theta \in A(Y_1, Y_2)$ varmasti:

$$P(Y_{(1)} = \theta \,|\, Y_1 \neq Y_2) = 1.$$

b) Jos $Y_1 = Y_2$, niin $\theta \in A(Y_1, Y_2)$ todennäköisyydellä 50 %:

$$P(Y_{(1)} = \theta \,|\, Y_1 = Y_2) = \tfrac{1}{2}.$$

c) $A(y_1, y_2)$ on 75 %:n luottamusjoukko θ:lle:

$$P(\theta \in A(Y_1, Y_2)) = P(Y_{(1)} = \theta) = \tfrac{3}{4}.$$

Pohdi lopuksi luottamusjoukkokäsitteen mielekkyyttä tässä tapauksessa empiirisen tutkijan kannalta, jolla on käsillä yksi ainutkertainen aineisto (y_1, y_2), jolle pätee joko $y_1 \neq y_2$ tai $y_1 = y_2$.

4. (Vanha koetehtävä.) Oletetaan, että $Y_1, \ldots, Y_n$ ovat riippumattomia ja noudattavat kukin jatkuvaa jakaumaa, jonka tiheysfunktio on

$$f(y; \lambda) = 2\lambda y \exp(-\lambda y^2), \qquad y > 0,$$

ja jossa λ on positiivinen parametri.

a) Muodosta tilastollinen malli $f_{\boldsymbol{Y}}$ ja johda parametrin λ suurimman uskottavuuden estimaattori $\hat{\lambda}$ sekä Fisherin informaatio $i(\lambda)$.

b) Mitä normaalijakaumaa $\hat{\lambda}$ approksimatiivisesti noudattaa, kun n on suuri?

c) Muodosta Waldin testiin eli yo. normaaliapproksimaatioon perustuva approksimatiivinen 95 %:n luottamusväli parametrille λ, kun aineisto on $\boldsymbol{y} = (y_1, \ldots, y_n)$.

KÄÄNNÄ!

5. Keskusteltavaksi:

   a) Mainitse pari kolme mielestäsi tärkeintä tai mielenkiintoisinta tilastollisen päättelyn piiriin kuuluvaa käsitettä tai menetelmää, jotka olet tällä kurssilla oppinut. Perustele!
   b) Miten käsityksesi tilastollisesta päättelystä on muuttunut tällä kurssilla verrattuna *Johdatus tilastolliseen päättelyyn* -kurssiin (tai vastaavaan)?

Ajankohtaista: Nämä ovat kurssin viimeiset harjoitustehtävät. Viimeiset luennot pidetään ti 24.2. ja to 26.2. Tällöin on mahdollisuus kysymyksiin ja keskusteluun kurssin aihepiireistä; torstaina myös luvassa pientä tarjoilua. Toinen kurssikoe pidetään ma 2.3. klo 13.00–16.00 salissa CK112. Siinä kuulustellaan muistiinpanojen jaksot 3.6–6.4 ja harjoitukset 7–13. Kurssikokeeseen saa ottaa mukaan käsinkirjoitetun, A4-kokoisen ja kaksipuolisen ”lunttilapun” mutta ei omia taulukoita tai kaavakirjoja. Lappu on pyynnöstä esitettävä valvojalle.