Tilastollinen päättely, syksy 2014 – kevät 2015
Harjoitus 3 (18. ja 20. 11. 2014)

1. a) Olkoon $f(y;\theta) = \theta y^{\theta-1}$, kun $0 \le y \le 1$ (ja $= 0$ muulloin). Varmista, että f on erään jatkuvan jakauman tiheysfunktio, kun $\theta > 0$. Mikä on tämän jakauman odotusarvo μ parametrin θ avulla ilmaistuna?

   b) Oletetaan, että havaintoja vastaavat satunnaismuuttujat $Y_1, \ldots, Y_n$ ovat riippumattomia ja noudattavat em. jakaumaa. Muodosta syntyvän tilastollisen mallin yhteistiheysfunktio. Ilmoita sen log-uskottavuusfunktio ja etsi parametrin θ suurimman uskottavuuden estimaatti, kun aineisto on $\boldsymbol{y} = (y_1, \ldots, y_n)$. Mikä on μ:n su-estimaatti?

2. (Monisteen teht. 2.11.) Jatkoa harjoituksen 2 tehtävään 3. Olkoot $Y_1, \ldots, Y_n \sim \mathit{Exp}(\lambda) \perp\!\!\!\perp$. Laske havaittu informaatio $j(\hat{\lambda}; \boldsymbol{y})$, Fisherin informaatio $i(\lambda)$ ja odotusarvo $E[l'(\lambda; \boldsymbol{Y})^2]$.

3. (Monisteen teht. 2.13.) Tarkastellaan mallia, jossa havaintoja vastaavat satunnaismuuttujat $Y_1, \ldots, Y_n$ ovat riippumattomat. Mallin parametri on θ. Totea, että mallin havaittu informaatio ja Fisherin informaatio ovat

$$j(\theta; \boldsymbol{y}) = j_1(\theta; y_1) + \cdots + j_n(\theta; y_n), \qquad i(\theta) = i_1(\theta) + \cdots + i_n(\theta),$$

   jossa $j_k(\theta; y_k)$ on pelkästään yhteen havaintoon y_k perustuva havaittu informaatio ja $i_k(\theta) = E[j_k(\theta; Y_k)]$ on vastaava Fisherin informaatio. Miten tulkitset tämän tuloksen?

4. (Monisteen teht. 2.18.) Eräällä järjestelmällä on kaksi mahdollista tilaa: 0 ja 1. Satunnaismuuttuja Y_i kertoo tilan hetkellä $i = 1, 2, 3, 4$. Se riippuu tilasta hetkellä $i-1$ (mutta ei aikaisemmista) seuraavien ehdollisten todennäköisyyksien kuvaamalla tavalla:

$$P\{Y_i = 1 \,|\, Y_{i-1} = 1\} = \alpha, \qquad P\{Y_i = 0 \,|\, Y_{i-1} = 1\} = 1 - \alpha,$$
$$P\{Y_i = 1 \,|\, Y_{i-1} = 0\} = 1 - \beta, \qquad P\{Y_i = 0 \,|\, Y_{i-1} = 0\} = \beta,$$

   jossa $0 < \alpha < 1$ ja $0 < \beta < 1$. Järjestelmä lähtee aina tilasta 0: $Y_0 = 0$.

   Aineisto on $\boldsymbol{y} = (y_1, y_2, y_3, y_4) = (0, 1, 1, 0)$. Muodosta vastaava log-uskottavuusfunktio $l(\alpha, \beta; \boldsymbol{y})$ ja määritä suurimman uskottavuuden estimaatti $(\hat{\alpha}, \hat{\beta})$.

5. Luennolla ja monisteessa on tarkasteltu kahta erilaista mallia toistokokeelle: Bernoulli-jakaumaan perustuvaa ja binomijakaumaan perustuvaa (monisteen esim. 2.1.2 ja 2.1.5).

   Kolmas vaihtoehto: Suoritetaan toistoja, kunnes ennalta päätetty määrä k onnistumisia on sattunut. Olkoon tarvittavien toistojen lukumäärä n, jolloin selvästi $n \ge k$. Johda n:ää vastaavan satunnaismuuttujan N pistetodennäköisyysfunktio, kun yhden toiston onnistumistodennäköisyys on $0 < \theta < 1$. Mikä on havaintoa n vastaava uskottavuusfunktio parametrille θ? Vertaa sitä binomijakaumamallin uskottavuusfunktioon. Eroavatko uskottavuuteen perustuvat päätelmät θ:sta toisistaan näissä malleissa?

   Pohdittavaksi: Keksitkö mitään sovellustilannetta, jossa toistokoetta olisi luontevaa lähestyä kolmannen mallin esittämällä tavalla? Oletetaan, että ystäväsi kertoo: ”Kyselin ostarilla mielipiteitä NATO-jäsenyydestä. Sadas tapaamani ihminen oli samalla kymmenes, joka ilmoittautui jäsenyyden kannattajaksi.” Millä tavalla luulet hänen ”mallittaneen” kyselytutkimuksensa: oliko hän ostarille mennessään päättänyt kysyä tarkalleen sadalta ihmiseltä vai kysyä, kunnes löytää kymmenen kannattajaa, vai ehkä ”kunhan kyselin niin pitkään kuin huvitti”? Saisiko vastaus tähän vaikuttaa jotenkin tilastotieteilijän tekemiin päätelmiin koskien NATO:n kannattajien suhteellista osuutta kaikista ostarilla kävijöistä?

   Ohje. Tapahtuma $\{N = n\}$ merkitsee, että $n-1$ ensimmäisessä toistossa on sattunut $k-1$ onnistumista ja $n-k$ epäonnistumista ja että n:s toisto on ollut onnistuminen. Toistot oletetaan riippumattomiksi. N:n jakaumalla on vakiintunut nimikin; mikähän se on?