Tilastollinen päättely II, syksy 2015 – kevät 2016
Harjoitus 5 (1. ja 3. 12. 2015)

1. Paluu toistokokeeseen. 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 $n \geq k$. 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? Voidaanko aina edellyttää toistokokeen tekijältä (esim. puhelinhaastattelija, katugallupin tekijä, lantinheittäjä tai lääkäri, joka haluaa selvittää, kuinka suuri osuus tietyn sydänleikkauksen läpikäyneistä potilaista kuolee ensimmäisen vuorokauden aikana), että hän osaa selvästi kertoa, mihin toistojen tekeminen päättyi: ennalta valitun toistojen määrän tultua täyteen, ennalta valitun "onnistumisten" lukumäärän tultua täyteen vai johonkin muuhun syyhyn (kaveri soitti ja pyysi tulemaan bileisiin, joten tutkimus keskeytyi)?

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?

Tehtävät 2–4 liittyvät harhattomuuteen. Monisteen jakso 3.2.

2. (Vrt. monisteen teht. 3.6.) Olkoot $x_1, x_2, \ldots$ annettuja nollasta eroavia reaalilukuja (selittävän muuttujan arvoja). Tarkastellaan regressiomallia $Y_1, \ldots, Y_n \perp\!\!\!\perp$, $Y_i \sim N(\beta x_i, \sigma^2)$, jossa β ja σ^2 ovat tuntemattomia parametreja.

a) Varmista, että parametrin β su-estimaattori on

$$\hat{\beta} = \sum_{i=1}^{n} x_i Y_i \Big/ \sum_{i=1}^{n} x_i^2.$$

Totea, että se on harhaton.

b) Totea, että myös β:n estimaattori

$$T = \sum_{i=1}^{n} Y_i \Big/ \sum_{i=1}^{n} x_i$$

on harhaton.

3. (Monisteen teht. 3.4.) Havainnoista $Y_1, \ldots, Y_n$ oletetaan samoin kuin harjoituksen 4 tehtävässä 4: ne ovat riippumattomia ja noudattavat jakaumaa, jolla on odotusarvo μ ja varianssi σ^2. Keksi jokin harhaton estimaattori odotusarvon neliölle μ^2.

KÄÄNNÄ!

4. (Vrt. monisteen teht. 3.10.) Mallissa $Y_1, \ldots, Y_n \sim \text{Tas}(0, \theta) \perp\!\!\!\perp$ on su-estimaattoriksi saatu $\hat{\theta} = \max(Y_1, \ldots, Y_n)$ (ks. kohta 2.2.8).

a) Muodosta $\hat{\theta}$:n kertymäfunktio F lähtien havainnosta

$$P\{\hat{\theta} \le t\} = P\{Y_1 \le t\} \cdots P\{Y_n \le t\}$$

ja derivoi siitä tiheysfunktio $f = F'$.

b) Laske $\hat{\theta}$:n odotusarvo ja totea, että $\hat{\theta}$ on harhainen mutta asymptoottisesti harhaton.

Tehtävää 5 varten tutustu alustavasti monisteen jakson 3.4 alkuun.

5. (Monisteen teht. 3.1.) Johda funktion $g(\theta)$ estimaattorin T keskineliövirheelle monisteen kohdassa 3.4.1 mainittu hajotelma

$$E_\theta[(T - g(\theta))^2] = \text{var}_\theta(T) + b(\theta)^2,$$

jossa $b(\theta)$ on estimaattorin harha.