Tilastollinen päättely I, kevät 2017
Harjoitus 3 – palautuspäivämäärä 4. 4.

Opastus: Tehtävät 1–3 liittyvät uskottavuusfunktioon ja suurimman uskottavuuden menetelmään (jaksot 4.1–4.4; jaksoa 4.5 emme käsittele). Tehtävät 4–6 liittyvät luottamusväleihin (jaksot 5.1–5.6.3). Tehtäviin 4 ja 6 liittyvä taulukko jaetaan erikseen Moodlessa.

1. Tarkastellaan binomikoemallia $X \sim \mathrm{Bin}(n, \theta)$, jossa havaitaan $X = x$ "onnistumista" n:ssä toistossa, parametrin ollessa $0 \leq \theta \leq 1$. Perustele huolellisesti uskottavuusfunktion lausekkesta ja su-estimaatin määritelmästä lähtien, että kaava $\hat{\theta} = x/n$ pätee myös tapauksissa $x = 0$ ja $x = n$.

2. Eksponenttijakauma $\mathrm{Exp}(\lambda)$ on jatkuva jakauma, jonka tiheysfunktio on $g(y; \lambda) = \lambda e^{-\lambda y}$, kun $y > 0$ (ja $= 0$ muulloin). Oletetaan, että havaintoja vastaavat satunnaismuuttujat $Y_1, \ldots, Y_n$ ovat riippumattomia ja noudattavat kukin $\mathrm{Exp}(\lambda)$-jakaumaa, jossa $\lambda > 0$ on parametri.

a) Muodosta tilastollisen mallin lauseke eli yhteistiheysfunktio satunnaisvektorille $\mathbf{Y} = (Y_1, \ldots, Y_n)$.

b) Oletetaan, että havainnot ovat $y_1, \ldots, y_n$. Muodosta näitä vastaava uskottavuusfunktio ja totea, että se voidaan saattaa sellaiseen muotoon, joka riippuu havainnoista vain niiden keskiarvon $\bar{y}$ kautta.

c) Johda huolellisesti perustellen su-estimaatti $\hat{\lambda}$. (Vihje: logaritmointi voi auttaa.)

3. Eksponenttijakauma (ks. edellä) voidaan parametroida myös omalla odotusarvollaan $\mu = 1/\lambda$ ns. intensiteettiparametrin λ sijasta.

a) Miltä edellisen tehtävän tilastollisen mallin lauseke ja uskottavuusfunktio näyttävät parametrin μ avulla lausuttuina?

b) Totea, että μ:n su-estimaatille pätee $\hat{\mu} = 1/\hat{\lambda}$.

Lisätieto. Tämä tehtävä on esimerkki tilastollisen mallin *uudelleenparametroinnista* ja siihen liittyvästä su-estimaatin *invarianssiominaisuudesta*. Niistä puhutaan yleisemmin kurssilla Tilastollinen päättely II.

4. Tilastollinen mallimme on satunnaisotos kokoa n normaalijakaumasta $N(\mu, \sigma^2)$, jossa $\sigma^2 > 0$ ajatellaan tunnetuksi luvuksi (halutessasi voit valita $\sigma^2 = 1$). Havaintoja vastaaville satunnaismuuttujille pätee siis $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2) \perp\!\!\!\perp$. Palauta mieleen, miten parametrille μ muodostettiin luottamusväli annetulla luottamustasolla $1 - \alpha$ (ns. z-luottamusväli, ks. monisteen kaava (5.10)).

a) Miten otoskoon n kasvu vaikuttaa luottamusvälin pituuteen? Jos luottamusvälin pituus halutaan kymmenenteen osaan aikaisemmasta eli halutaan "kymmenkertainen tarkkuus", mitä n:lle on tehtävä?

b) Tutki, miten luottamustason muuttaminen vaikuttaa luottamusvälin pituuteen. Jos luottamustaso halutaan nostaa 90 %:sta 95 %:een, kuinka moninkertaiseksi luottamusväli tulee pituudeltaan? Entä jos siirrytään vielä 95 %:n luottamustasosta 99 %:n luottamustasoon? Kuinka pitkä olisi 100 %:n luottamusväli?

c) Oletetaan, että $\sigma^2 = 1$. Kuinka suuri on otoskoon n likimain oltava, jotta 95 %:n luottamusväli μ:lle olisi pituudeltaan noin 0.1 eli muotoa $[\bar{y} - 0.05,\ \bar{y} + 0.05]$?

5. Oletetaan, että vastasyntyneen tyttölapsen päänympärys noudattaa likimain normaalijakaumaa. Eräässä väestössä on otantatutkimuksen perusteella johdettu keskimääräiselle päänympärykselle μ (senttimetriä) 95 %:n luottamusväli $[34.3, 34.9]$. Mitkä seuraavista tulkinnoista ovat oikein ja mitkä väärin?

a) μ:n todennäköisyysjakaumasta 95 % sijaitsee kyseisellä välillä.

b) 95 %:lla kaikista vastasyntyneistä tyttölapsista päänympärys on kyseisellä välillä.

c) Mikäli samanlainen otantatutkimus toistettaisiin useita kertoja ja joka kerta muodostettaisiin vastaava luottamusväli, kuuluisi μ:n arvo noin 95 %:iin saaduista väleistä.

Perustele erityisesti väärien väitteiden kohdalla, miksi ne ovat väärin.

6. Normaalijakaumaan liittyvien luottamusvälien ja testiteorian yhteydessä tarvitaan *Studentin t-jakaumia* t_ν. Niiden (eräs) täsmällinen määritelmä on esitetty monisteen sivulla 49 ja yhteys normaalijakaumamalliin käy ilmi sivun 50 tarkasteluista. t-jakaumien kvantiileja (ks. kvantiilin käsite monisteen jaksossa 5.4) ja "häntätodennäköisyyksiä" on perinteisesti koottu taulukoiden muotoon (ks. liite).

a) Etsi kvantiilit (likiarvot) a, b ja c siten, että $P(X < a) = 0.01$, $P(X > b) = 0.01$ ja $P(-c < X < c) = 0.99$, kun $X \sim t_{15}$.

Vapausasteluvun ν ollessa suuri t_ν-jakauman kvantiileja voidaan approksimoida vastaavilla standardinormaalijakauman kvantiileilla.

b) Kuinka monta prosenttia suurempi on yläkvantiili $t_{30}(0.025)$ kuin vastaava standardinormaalijakauman yläkvantiili $z_{0.025}$?

c) Tutki, kuinka suuri ν:n olisi oltava, jotta $t_\nu(0.025)$ poikkeaisi enintään yhden prosentin verran yläkvantiilista $z_{0.025}$. Voit tässä arvioida t-jakauman yläkvantiileja (eräästä asymptoottisesta kehitelmästä saatavalla) kaavalla

$$t_\nu(u) \approx z_u + \frac{z_u^3 + z_u}{4\nu}.$$

Huom. Normaalijakaumaan ja t-jakaumiin liittyviä kvantiileja ja todennäköisyyksiä voi kätevästi laskea tilastollisilla ohjelmistoilla (esim. R:llä) tai netistä löytyvillä online-laskimilla; ks. esim. http://surfstat.anu.edu.au/surfstat-home/tables/t.php. Taulukon käyttö on kuitenkin syytä harjoitella ainakin koetta varten.