Tilastollinen päättely II, syksy 2015 – kevät 2016
Harjoitus 4 (24. ja 26. 11. 2015)

Tehtävät 1–3 liittyvät su-estimointiin ja informaation käsitteeseen. Tehtävässä 1 tarvitaan myös säännöllisen mallin erityispiirteitä. Luentomonisteen jaksot 2.2–2.5.

1. (Monisteen teht. 2.14.) Olkoon $f_{\mathbf{Y}}(\mathbf{y};\theta)$ tilastollinen malli, jonka parametri θ on yksiulotteinen. Olkoon $\phi = \phi(\theta)$ kääntäen yksikäsitteinen parametrimuunnos, jonka käänteismuunnos on $\theta = \theta(\phi)$. Tarkastellaan uudelleenparametroitua mallia $f^*_{\mathbf{Y}}(\mathbf{y};\phi) = f_{\mathbf{Y}}(\mathbf{y};\theta(\phi))$. Näytä, että sen havaittu informaatio ja Fisherin informaatio saadaan alkuperäisen mallin informaatioista kaavoilla

$$j^*(\hat{\phi};\mathbf{y}) = j(\hat{\theta};\mathbf{y})\,\theta'(\hat{\phi})^2, \qquad i^*(\phi) = i(\theta(\phi))\,\theta'(\phi)^2.$$

Oletetaan, että malli täyttää kaikki tarpeelliset säännöllisyysehdot ja että parametrimuunnos on riittävän monta kertaa derivoituva. [*Apu.* $l'(\hat{\theta};\mathbf{y}) = 0$ ja $E[l'(\theta;\mathbf{Y})] = 0$.]

2. ”Sensuroidut” eksponenttihavainnot. Kurssilla on esitelty kaksi vaihtoehtoista tapaa tutkia elinaikojen jakaumaa: kaikkien otosyksiköiden (ihmisten, sähkölaitteiden, radioaktiivisten atomien, jne.) elinaikojen mittaus (ks. monisteen esim. 1.2.2) tai tietyllä ajanhetkellä tapahtuva elävien ja kuolleiden lukumäärien laskeminen, joka voitiin tulkita toistokokeena (ks. harjoituksen 1 tehtävä 3).

Käytännössä tavallisempi koejärjestely on seuraava: Ryhdytään seuraamaan n otosyksikköä ajanhetkellä 0 ja lopetetaan seuranta ennalta päätetyllä ajanhetkellä a. Niiden otosyksiköiden osalta, jotka kuolivat aikavälillä $(0, a]$, saadaan selville tarkka elinaika. Lisäksi havaitaan hetkellä a yhä elossa olevien yksiköiden lukumäärä; näiden yksiköiden elinajat ovat siis $> a$ mutta niitä ei mitata tarkasti. Tällaista menettelyä kutsutaan ”sensuroinniksi” (*censoring*).

Oletetaan, että tutkittavien otosyksiköiden elinajat satunnaismuuttujina ovat riippumattomia ja Exp(λ)-jakautuneita, ts. $Y_1, \ldots, Y_n \sim \text{Exp}(\lambda) \perp\!\!\!\perp$. Oletetaan, että havaituista elinajoista k kpl on välillä $(0, a]$; olkoot ne $y_1 \ldots, y_k$. Ajanhetkellä a on vielä elossa $n - k$ otosyksikköä. Koska $P(Y_i > a) = e^{-\lambda a}$, voidaan perustella, että aineistoa vastaava uskottavuusfunktio on

$$L(\lambda) = \lambda e^{-\lambda y_1} \cdots \lambda e^{-\lambda y_k} \cdot (e^{-\lambda a})^{n-k} = \lambda^k e^{-\lambda s},$$

jossa $s = \sum_{i=1}^k y_i + (n - k)a$.

a) Muodosta log-uskottavuusfunktio ja johda kaava λ:n su-estimaatille $\hat{\lambda}$ sekä keskimääräisen (eli odotettavissa olevan) elinajan $\mu = 1/\lambda$ su-estimaatille $\hat{\mu}$. Totea, että tapauksessa $k = n$ ne yhtyvät eksponenttimallista saataviin estimaatteihin (harjoituksen 2 tehtävä 2).

b) Oletetaan, että $n = 10$ ja otosyksiköiden elinajat suuruusjärjestyksessä (päivinä) ovat

$$4 \quad 5 \quad 8 \quad 11 \quad 20 \quad 29 \quad 35 \quad 40 \quad 66 \quad 70.$$

Mikä on $\hat{\mu}$ eksponenttimalliin perustuen (kaikkien elinaikojen tarkka mittaus)? Entä jos päätettiin suorittaa ”sensurointi” ajanhetkellä $a = 50$ (päivää)? Entä jos $a = 25$?

KÄÄNNÄ!

3. Jatkoa edellisen tehtävän a-kohtaan. Laske Fisherin informaatio $i_a(\lambda)$ "sensuroitujen" eksponenttihavaintojen mallissa (a = sensurointihetki eli seuranta-ajan pituus). Huomaa, että k on tulkittava erään satunnaismuuttujan K toteutuneeksi arvoksi: se on niiden otosyksiköiden lukumäärä, jotka ovat kuolleet viimeistään ajanhetkellä a, joten se noudattaa erästä binomijakaumaa (mitä?).

Totea, että $i_a(\lambda) \to n/\lambda^2$, kun $a \to \infty$. Tässä n/λ^2 on eksponenttimallin Fisherin informaatio (harjoituksen 3 tehtävä 4). Kuinka suureksi a olisi pyrittävä valitsemaan (λ:sta riippuen), jotta sensuroidun mallin informaatio $i_a(\lambda)$ olisi ainakin puolet eksponenttimallin informaatiosta?

Tehtävässä 4 tutustutaan estimaattorin harhattomuuden käsitteeseen. Opiskele sitä varten monisteen jaksoja 3.1 ja 3.2.

4. (Monisteen teht. 3.3.) Oletetaan, että havainnot $Y_1, \ldots, Y_n$ ovat riippumattomia ja noudattavat jakaumaa, jolla on odotusarvo μ ja varianssi σ^2; esimerkkinä $N(\mu, \sigma^2)$. Tarkastellaan parametrin σ^2 estimointia muotoa cV olevilla estimaattoreilla, kun $c > 0$ on vakio ja

$$V = \sum_{i=1}^{n} (Y_i - \overline{Y})^2.$$

Näytä, että cV on harhaton jos ja vain jos $c = 1/(n-1)$. [*Ehdotus.* Käytä harjoituksen 2 tehtävän 1 hajotelmaa valinnalla $a = \mu$.]