Ei-parametriset ja robustit menetelmät, kevät 2015

Harjoitus 3

1. Olkoon $X = (x_1, \ldots, x_n)$ havaintoaineisto, jossa kaikki havainnot ovat erisuuria. Voidaan osoittaa, että erilaisia bootstrap-otoksia on tällöin yhteensä $\binom{2n-1}{n}$ kappaletta. Käy läpi kaikki erilaiset bootstrap-otokset, kun $n = 2, 3, 4$ ja totea otosten lukumäärän kaavan paikkaansapitävyys näissä erikoistapauksissa.

2. Tarkastellaan aineistoa `kilndry.data`. Tutki bootstrap-estimoinnin avulla käsittelyn vaikutuksen Δ estimoinnin tarkkuutta, kun estimoinnissa on käytetty keskiarvoa, mediaania, HL-estimaattia ja MIDRANGE-estimaattia. Voit käyttää hyväksi luentomonisteen R-funktioita.

3. Etsi otosmediaanin ja midrange-estimaatin (eli pienimmän ja suurimman arvon keskiarvon) murtumispisteet.

4. Olkoon $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ järjestetty otos jakaumasta, jonka tuntematon symmetriapiste on Δ. Tarkastellaan seuraavanlaista Δ:n estimaattia:

$$\hat{\Delta} = \frac{1}{n-2k} \sum_{i=1}^{n} x_{(i)} I(k+1 \leq i \leq n-k).$$

   Kyseessä on niin sanottu **trimmattu keskiarvo** $\bar{x}_\alpha$ trimmaussuhteena $\alpha = k/n$. Etsi trimmatun keskiarvon murtumispiste. Osoita, että otosmediaani saadaan trimmatun keskiarvon erikoistapauksena.

5. Osoita, että otosmediaanin empiirinen influenssifunktio on

$$\begin{aligned} IF_n(x; F_n, T_n) &= (n+1)\Bigg[\frac{x_{(k)} - x_{(k+1)}}{2} I(x < x_{(k)}) \\ &\quad + \frac{x - x_{(k+1)}}{2} I(x_{(k)} \leq x \leq x_{(k+2)}) \\ &\quad + \frac{x_{(k+2)} - x_{(k+1)}}{2} I(x > x_{(k+2)})\Bigg] \end{aligned}$$

   kun otoskoko on $n = 2k+1$ ja

$$\begin{aligned} IF_n(x; F_n, T_n) &= (n+1)\Bigg[\frac{x_{(k)} - x_{(k+1)}}{2} I(x < x_{(k)}) \\ &\quad + \frac{2x - x_{(k)} - x_{(k+1)}}{2} I(x_{(k)} \leq x \leq x_{(k+1)}) \\ &\quad + \frac{x_{(k+1)} - x_{(k)}}{2} I(x > x_{(k+1)})\Bigg] \end{aligned}$$

   kun otoskoko on $n = 2k$.