**Tilastollinen päättely, syksy 2009 - kevät 2010**

Harjoitus 6, viikko 3

1. Olkoot $x_1, x_2, \ldots, x_n$ annettuja nollasta eroavia reaalilukuja. Tarkastellaan regressiomallia $Y_1, \ldots, Y_n \perp\!\!\!\perp$, $Y_i \sim N(\beta x_i, \sigma^2)$, jossa $\beta \in \mathbb{R}$ ja $\sigma^2 > 0$ ovat tuntemattomia parametreja.

   (a) Varmista, että parametrin $\beta$ su-estimaattori on
   $$\hat{\beta} = \sum_{i=1}^n x_i Y_i \Big/ \sum_{i=1}^n x_i^2$$
   ja totea, että se on harhaton ja laske sen varianssi.

   (b) Totea, että myös estimaattori
   $$T = \sum_{i=1}^n Y_i \Big/ \sum_{i=1}^n x_i$$
   on harhaton ja laske sen varianssi.

   (c) Kumpi estimaattori on tehokkaampi?

   *Ohje.* Epäyhtälö $\left(\sum_{i=1}^n x_i\right)^2 \leq n \sum_{i=1}^n x_i^2$ lienee hyödyksi. Siinä yhtäsuuruus pätee vain jos kaikki $x_i$:t ovat samoja.

2. Jatkoa tehtävään 1. Oletetaan, että $x_i \geq c$ kaikilla $i$, jossa $c > 0$ on vakio. Näytä, että $\hat{\beta}$ ja $T$ ovat tarkentuvia. Päteekö tarkentuvuus, jos $x_i = 1/i$? [Jälkimmäisessä kohdassa ei vaadita täsmällistä matemaattista perustelua].

3. Tarkastellaan ns. Poisson-regressiomallia: $Y_1, \ldots, Y_n \perp\!\!\!\perp$ ja $Y_i \sim P(\lambda x_i)$, jossa $x_1, \ldots, x_n$ ovat tunnettuja positiivisia lukuja (selittävän muuttujan arvoja). Muodosta tämän mallin log-uskottavuusfunktio ja johda parametrin $\lambda$ suurimman uskottavuuden estimaattorille lauseke
   $$\hat{\lambda} = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i}.$$
   Näytä, että $\hat{\lambda}$ on harhaton. Konkreettisena esimerkkinä voidaan ajatella, että $Y_i$ on johonkin tautiin kuolleiden lukumäärä populaatiossa, jonka koko on $x_i$.

4. Jatkoa edelliselle. Laske suurimman uskottavuuden estimaattorin $\hat{\lambda}$ varianssi. Onko $\hat{\lambda}$ tarkentuva, jos luvuille $x_1, x_2, \ldots, x_n$ pätee $x_i \geq c$, jossa $c > 0$ on vakio?

5. Olkoon $Y$ jatkuvasti jakautunut satunnaismuuttuja, jonka tiheysfunktio on $f$, ja olkoon $g$ jonkin toisen jakauman tiheysfunktio. Todista, että

$$E[\log f(Y)] > E[\log g(Y)].$$

Tätä tulosta käytettiin kohdassa 3.6.2. Miten?

*Ohje.* Todennäköisyyslaskennan Jensenin epäyhtälö kertoo, että jos $u$ on aidosti ylöspäin kupera (eli aidosti konkaavi) funktio ja $X$ on satunnaismuuttuja, niin $E(u(X)) \leq u(E(X))$ ja yhtäsuuruus pätee vain jos $X$ on vakio. Sovella tätä valinnoilla $u(x) = \log x$ ja $X = g(Y)/f(Y)$. Muista myös, että $\log a/b = \log a - \log b$.