Tilastollinen päättely III, kevät 2016
Laskuharjoitukset 6, palautus 2.5.2016 klo 23:55 mennessä kurssin Moodle-sivulle.

Viimeistä viedään, vielä jaksaa!

1. Tunnetusti satunnaismuuttuja Y noudattaa Bernoulli(θ)-jakaumaa, jossa $0 < \theta < 1$, mikäli $\Pr\{Y = 1\} = \theta$ ja $\Pr\{Y = 0\} = 1 - \theta$. Voidaan myös kirjoittaa $\Pr\{Y = y\} = \theta^y (1-\theta)^{1-y}$, kun $y = 0, 1$.

   Logistisessa regressiomallissa oletetaan, että $Y_1, \ldots, Y_n \perp\!\!\!\perp$, jossa $Y_i \sim \text{Bernoulli}(\theta_i)$ ja θ_i:n *logit-muunnos* toteuttaa lineaarisen regressioyhtälön

   $$\log \frac{\theta_i}{1-\theta_i} = \alpha + \beta x_i,$$

   jossa (α, β) on kaksiulotteinen parametri ja $x_1, \ldots, x_n \in \mathbb{R}$ ovat kiinteitä selittävän muuttujan arvoja. Yhtäpitävästi $\theta_i = \theta_i(\alpha, \beta) = (1 + e^{-\alpha-\beta x_i})^{-1}$.

   Muodosta tämän mallin logaritminen uskottavuusfunktio, pistemääräfunktio ja havaittu informaatiomatriisi.

2. Tarkastellaan tehtävän 1 mallia ja hypoteesia $H_0\colon \beta = 0$, jonka mukaan selittäjillä x_i ei ole merkitystä.

   Muodosta Waldin testisuureen lauseke hypoteesille H_0 ja kerro, mikä on sen asymptoottinen jakauma H_0:n pätiessä. Oletetaan tässä, että lauseen 3.2 tulos ja kaikki siitä seuraava asymptotiikka pätee tarkasteltavalle mallille (käytännössä tämä merkitsee joitakin rajoitteita selittäjien arvoille).

   Huomaa, että su-estimaattorin $(\hat{\alpha}, \hat{\beta})$ lauseketta ei voi ratkaista ”suljetussa muodossa”. Merkitse lyhyesti esim. $\hat{\theta}_i = \theta_i(\hat{\alpha}, \hat{\beta})$.

3. Sama malli ja hypoteesi kuin edellä. Johda rajoitettu (eli H_0:n pätiessä muodostettu) su-estimaattori $(\tilde{\alpha}, \tilde{\beta})$. Muodosta sitten Raon testisuureen lauseke ja kerro, mikä on sen asymptoottinen jakauma H_0:n pätiessä. Käytä testisuureen muodostamiseen kaavaa (3.15).

Vihjeitä seuraavalla sivulla!

Vihjeitä:

1. Log-uskottavuus on

$$l(\alpha, \beta; \boldsymbol{y}) = \sum_{i=1}^{n} y_i(\alpha + \beta x_i) - \sum_{i=1}^{n} \log(1 + e^{\alpha + \beta x_i})$$

   ja havaittu informaatio on

$$\mathcal{J}(\alpha, \beta; \boldsymbol{y}) = \begin{bmatrix} \sum \theta_i(1-\theta_i) & \sum x_i\theta_i(1-\theta_i) \\ \sum x_i\theta_i(1-\theta_i) & \sum x_i^2\theta_i(1-\theta_i) \end{bmatrix}.$$

2. Jos

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

   on matriisi, jolle $\det(M) = ad - bc \neq 0$, niin

$$M^{-1} = \frac{1}{\det(M)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

3. Rajoitettu malli on hyvin tuttu aineopintojen päättelyn kurssilta (uudelleenparametroituna). Testisuureen voi lopulta esittää melko siistinä lausekkeena esim. suureista $\bar{y} = \sum_i y_i/n$, $s_x^2 = \sum_i (x_i - \bar{x})^2/n$ ja $s_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y})/n$.