Testy pre regresné kvantily založené na metóde sedlového bodu

Radka Sabolová

KPMS MFF UK

11. september 2012

Obsah

Metóda sedlového bodu

- aproximácie hustoty odhadov metódou sedlového bodu predstavil Daniels (1954), neskôr rozpracoval Hampel (1973), Field a Hampel (1982)
- veľmi presné aproximácie hustoty odhadov
- testy pre M-odhady založené na výraze v exponente odhadu hustoty

Model

model

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + u_i, \qquad i = 1, \dots, n,$$

kde $u_i \sim \frac{1}{\sigma} g(.)$ a $(Y_i, \mathbf{X}_i)$ sú i.i.d. s hustotou $\frac{1}{\sigma} g\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right) k(\mathbf{x}_i)$, $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$

- regresný kvantil je riešením

$$\boldsymbol{\beta}_\alpha = \arg\min \left\{ \mathbb{E} \rho_\alpha (Y - \boldsymbol{X}\mathbf{t}) : \mathbf{t} \in \mathbb{R}^p \right\},$$

kde

$$\rho_\alpha(x) = |x| \left\{ (1-\alpha) \, \mathrm{I}[x < 0] + \alpha \, \mathrm{I}[x > 0] \right\}$$

- odhad regresného kvantilu $\hat{\boldsymbol{\beta}}_\alpha$ je riešením

$$\arg\min_{\mathbf{t}} \sum_{i=1}^{n} (\alpha - I[Y_i - \mathbf{X}_i^T t < 0]) \mathbf{X}_i$$

- $\hat{\boldsymbol{\beta}}_\alpha$ je konzistentným odhadom $(\beta_0 + G^{-1}(\alpha), \beta_1, \dots, \beta_p)$

Testujeme hypotézu

$$H_0 : \boldsymbol{\beta}_\alpha = \boldsymbol{\beta}_{\alpha 0}$$

Model

model

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + u_i, \qquad i = 1, \dots, n,$$

kde $u_i \sim \frac{1}{\sigma} g(.)$ a $(Y_i, \mathbf{X}_i)$ sú i.i.d. s hustotou $\frac{1}{\sigma} g\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right) k(\mathbf{x}_i)$, $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$

- regresný kvantil je riešením

$$\boldsymbol{\beta}_\alpha = \arg\min \left\{ \mathbb{E}\rho_\alpha(Y - \boldsymbol{X}\mathbf{t}) : \mathbf{t} \in \mathbb{R}^p \right\},$$

  kde

$$\rho_\alpha(x) = |x| \left\{ (1 - \alpha) \, \mathrm{I}[x < 0] + \alpha \, \mathrm{I}[x > 0] \right\}$$

- odhad regresného kvantilu $\hat{\boldsymbol{\beta}}_\alpha$ je riešením

$$\arg\min_{\mathbf{t}} \sum_{i=1}^{n} (\alpha - I[Y_i - \mathbf{X}_i^T t < 0]) \mathbf{X}_i$$

- $\hat{\boldsymbol{\beta}}_\alpha$ je konzistentným odhadom $(\beta_0 + G^{-1}(\alpha), \beta_1, \dots, \beta_p)$

Testujeme hypotézu

$$H_0 : \boldsymbol{\beta}_\alpha = \boldsymbol{\beta}_{\alpha 0}$$

Parametrický test

kumulatívna vytvárajúca funkcia skórov $\psi(Y_i, \boldsymbol{\beta}_\alpha)$

$$K_\psi(\boldsymbol{\lambda}, \boldsymbol{\beta}_\alpha) = \log \int \left\{ e^{\alpha \boldsymbol{\lambda}^T \mathbf{x}_i} k(\mathbf{x}_i) \left(e^{-\boldsymbol{\lambda}^T \mathbf{x}_i} G\left(\frac{\mathbf{x}_i^T (\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})}{\sigma} \right) + 1 - G\left(\frac{\mathbf{x}_i^T (\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})}{\sigma} \right) \right) \right\} d\mathbf{x}_i$$

testová štatistika

$$h(\hat{\boldsymbol{\beta}}_\alpha) = \log \int \left(\frac{G\left(\frac{\mathbf{x}_i^T (\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_0)}{\sigma} \right)}{\alpha} \right)^\alpha \left(\frac{1 - G\left(\frac{\mathbf{x}_i^T (\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_0)}{\sigma} \right)}{1 - \alpha} \right)^{1-\alpha} k(\mathbf{x}_i)\, d\mathbf{x}_i$$

za platnosti H_0

$$2nh(\hat{\boldsymbol{\beta}}_\alpha) \xrightarrow[n\to\infty]{\mathcal{D}} \chi^2_{p+1}$$

Parametrický test

kumulatívna vytvárajúca funkcia skórov $\psi(Y_i, \boldsymbol{\beta}_\alpha)$

$$K_\psi(\boldsymbol{\lambda}, \boldsymbol{\beta}_\alpha) = \log \int \left\{ e^{\alpha \boldsymbol{\lambda}^T \mathbf{x}_i} k(\mathbf{x}_i) \left(e^{-\boldsymbol{\lambda}^T \mathbf{x}_i} G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})}{\sigma} \right) + 1 - G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})}{\sigma} \right) \right) \right\} d\mathbf{x}_i$$

testová štatistika

$$h(\hat{\boldsymbol{\beta}}_\alpha) = \log \int \left(\frac{G\left(\frac{\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_0)}{\sigma} \right)}{\alpha} \right)^{\alpha} \left(\frac{1 - G\left(\frac{\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_0)}{\sigma} \right)}{1-\alpha} \right)^{1-\alpha} k(\mathbf{x}_i)\, d\mathbf{x}_i$$

za platnosti H_0

$$2nh(\hat{\boldsymbol{\beta}}_\alpha) \xrightarrow[n\to\infty]{\mathcal{D}} \chi^2_{p+1}$$

Neparametrický test

- reziduá a ich empirická distribučná funkcia

$$
\begin{aligned}
r_i &= y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_\alpha \\
I_i &= I[r_i < 0] \\
I_{ij}(\boldsymbol{\beta}) &= I[r_i + \mathbf{x}_j^T (\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}) < 0] \\
G_n^j(\boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^{n} I_{ij}(\boldsymbol{\beta})
\end{aligned}
$$

- empirická hypotetická distribučná funkcia

$$
w_{ij} = \frac{\left(\frac{\alpha}{1-\alpha} \frac{1-G_n^j(\boldsymbol{\beta}_{\alpha 0})}{G_n^j(\boldsymbol{\beta}_{\alpha 0})} \right)^{I_{ij}(\boldsymbol{\beta}_{\alpha 0})}}{\sum_{k=1}^n \left(\frac{\alpha}{1-\alpha} \frac{1-G_n^j(\boldsymbol{\beta}_{\alpha 0})}{G_n^j(\boldsymbol{\beta}_{\alpha 0})} \right)^{I_{kj}(\boldsymbol{\beta}_{\alpha 0})}}, i,j = 1, \dots, n
$$

Neparametrický test

- reziduá a ich empirická distribučná funkcia

$$\begin{aligned} r_i &= y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_\alpha \\ I_i &= I[r_i < 0] \\ I_{ij}(\boldsymbol{\beta}) &= I[r_i + \mathbf{x}_j^T (\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}) < 0] \\ G_n^j(\boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^{n} I_{ij}(\boldsymbol{\beta}) \end{aligned}$$

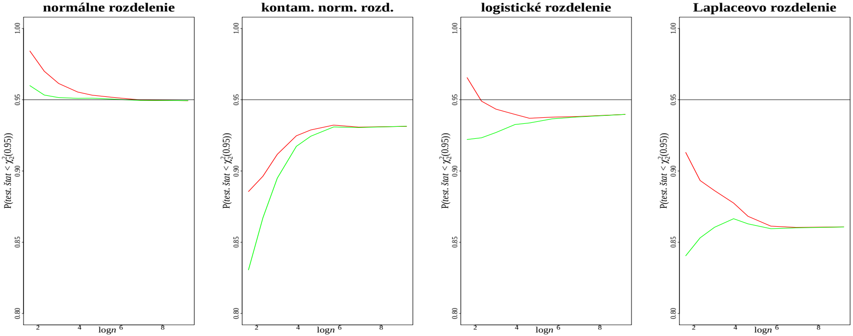
- empirická hypotetická distribučná funkcia

$$w_{ij} = \frac{\left(\frac{\alpha}{1-\alpha} \frac{1-G_n^j(\boldsymbol{\beta}_{\alpha 0})}{G_n^j(\boldsymbol{\beta}_{\alpha 0})} \right)^{I_{ij}(\boldsymbol{\beta}_{\alpha 0})}}{\sum_{k=1}^{n} \left(\frac{\alpha}{1-\alpha} \frac{1-G_n^j(\boldsymbol{\beta}_{\alpha 0})}{G_n^j(\boldsymbol{\beta}_{\alpha 0})} \right)^{I_{kj}(\boldsymbol{\beta}_{\alpha 0})}}, i, j = 1, \ldots, n$$

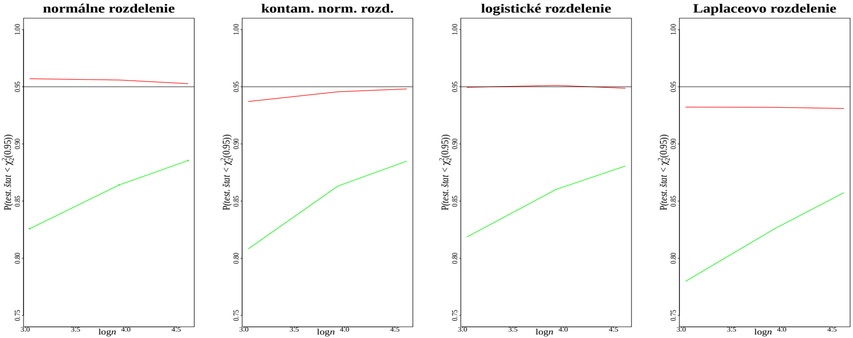
Potom za platnosti H_0

$$-2\sum_{j=1}^{n}\log\left\{\left(\frac{\sum_{i=1}^{n} w_{ij} l_i}{\alpha}\right)^{\alpha}\left(\frac{\sum_{i=1}^{n} w_{ij}(1-l_i)}{1-\alpha}\right)^{1-\alpha}\right\}\xrightarrow[n\to\infty]{\mathcal{D}}\chi^2_{p+1}$$

Parametrický test - výsledky



Neparametrický test - výsledky



Literatúra

- Daniels, HE
  Saddlepoint approximations on statistics
  Ann. Math. Statist. 25, 631–650.
- Field, CA and Hampel, FR (1982)
  Small-sample asymptotic distribution of M-estimators of location.
  Biometrika 69, 29–46.
- Hampel, F. R. (1973). Some small-sample asymptotics. In: Proc. Prague Symposium on Asymptotic Statistics (J. Hájek, ed.), pp. 109–126. Charles University in Prague.
- Ronchetti E, Robinson J, Young GA (2003)
  Saddlepoint Approximations and Tests Based on Multivariate M-estimates
  The Annals of Statistics 31: 1154–1169
- Ronchetti E, Welsh AH (1994)
  Empirical Saddlepoint Approximations for Multivariate M-estimators
  Journal of the Royal Statistical Society 56, no.2: 313–326