0.1 Квантилна регресия

Нека е даден класическия модел на линейна множествена регресия

$$y_i = x_i^T\beta + \varepsilon_i \quad \text{for} \quad i = 1, \ldots, n,$$

където $y_i \in R^1$ и $x_i \in R^p$ са наблюденията на зависимата и предикторни променливи, $\beta \in R^p$ е вектор от неизвестни параметри, ε_i са независими еднакво разпределени случайни грешки с очакване $\mathrm{E}(\varepsilon_i) = 0$ и константна дисперсия $\mathrm{var}(\varepsilon_i)$=$\sigma^2 > 0$.

Регресионните остатъци на модела се дефинират като

$$r_i(\theta) := y_i - x_i^T\beta \quad \text{за} \quad i = 1, \ldots, n.$$

Нека ε_i е асиметричното Лапласово разпределена. **Дефиниция:** Асиметричното Лапласово разпределение се дефинира като

$$\begin{aligned} f(y_i, x_i^T\beta, \tau, \sigma) &= \frac{\tau(1-\tau)}{\sigma}\exp\left(-\frac{\rho_\tau(y_i - x_i^T\beta)}{\sigma}\right) \\ &= \frac{\tau(1-\tau)}{\sigma}\begin{cases} \exp\left(-\frac{(\tau-1)r(\beta)}{\sigma}\right) & \text{if } r(\beta) < 0, \\ \exp\left(-\frac{\tau r(\beta)}{\sigma}\right) & \text{if } r(\beta) \geq 0, \end{cases} \end{aligned}$$

Дефиниция: Максимално правдоподобната оценка на β за всеки зададен квантил $0 < \tau < 1$ води до минимизирането на

$$\min_{\beta \in R^p} \sum_{i=1}^{n} \rho_\tau(r_i(\beta)), \tag{1}$$

където

$$\begin{aligned} \rho_\tau(r(\beta)) &= |r(\beta)|\left[\tau 1_{\{r(\beta)\geq 0\}} + (1-\tau)1_{\{r(\beta)<0\}}\right] \\ &= \begin{cases} (\tau-1)r(\beta) & \text{if} \quad r(\beta) < 0, \\ \tau r(\beta) & \text{if} \quad r(\beta) \geq 0, \end{cases} \end{aligned}$$

При $\tau = \frac{1}{2}$, $\widehat{\beta}(\tau) = \hat{\beta}_{LAD}$, т.е., 0.5 квантилната регресионна оценка съвпада с оценката по МММ.

Процедурата по оценяването на регресионния квантил се свежда до проблем от линейното програмирано. Разработена е библиотека "quantreg"от програмната среда R, Koenker (2005), която провежда квантилен регресионен анализ по подобие на класическия регресионен анализ.

Ще отбележим, че чрез класическия регресионен модел по МНК и МММ се моделират очакването и медината на предиктанта, докато квантилната регресия позволява моделирането на всички квантили на предиктанта. В този смисъл, чрез методологията на квантилната регресия получаваме възможност за провеждането на пълен анализ на данните в сравнение с МНК и МММ.