Distributional Modelling in R

Quantile Regression

Thomas Kneib, Nikolaus Umlauf

https://nikum.org/dmr.html

Implied Quantiles of the Linear Model

- Consider the model specification

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i.$$

- Assuming $\mathbb{E}(\varepsilon_i) = 0$ implies that the regression coefficients relate to expected differences in the response variable since

$$\mathbb{E}(y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\beta}.$$

- If $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$, the model also implies

$$Q_\tau(y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\beta} + \sigma z_\tau$$

  where z_τ is the τ-quantile of the standard normal and $Q_\tau(y_i|\boldsymbol{x}_i)$ denotes the conditional quantile function with quantile level $0 < \tau < 1$.

Idea of Quantile Regression

- Make the conditional quantiles covariate-dependent beyond a simple shift effect.
- Avoid parametric assumptions such as $\varepsilon_i \sim \mathrm{N}(0, \exp(\boldsymbol{x}_i'\gamma)^2)$ for which

$$Q_\tau(y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i'\beta + \exp(\boldsymbol{x}_i'\gamma)z_\tau.$$

- Rather assume

$$Q_\tau(y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\beta}_\tau$$

  with quantile-specific regression coefficients $\boldsymbol{\beta}_\tau$.

Quantile Regression

- Assume the model
$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta}_\tau + \varepsilon_{i,\tau}$$
with quantile-specific regression coefficients $\boldsymbol{\beta}_\tau$ and $F_{\varepsilon_{i,\tau}}(0) = \tau$ i.e. the τ-quantile of the error term $\varepsilon_{i,\tau}$ is zero.
- Then the τ-quantile of the response y_i is given by
$$Q_\tau(y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\beta}_\tau \quad \text{or} \quad F_{y_i}(\boldsymbol{x}_i'\boldsymbol{\beta}_\tau) = \tau$$
where $Q_\tau(y_i|\boldsymbol{x}_i) = F_{y_i}^{-1}(\tau)$ is the quantile function of the distribution of y_i

Pinball Loss

- Estimation of the quantile regression coefficients can be achieved by minimizing the pinball loss

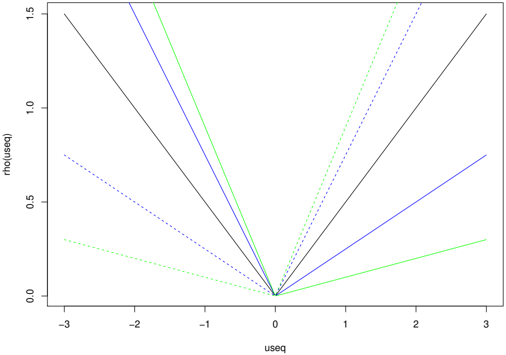
$$\hat{\boldsymbol{\beta}}_\tau = \underset{\boldsymbol{\beta}_\tau}{\arg\min} \sum_{i=1}^{n} w_\tau(y_i) |y_i - \boldsymbol{x}_i' \boldsymbol{\beta}_\tau|$$

with

$$w_\tau(y_i) = \begin{cases} \tau & y_i > \boldsymbol{x}_i' \boldsymbol{\beta}_\tau \\ 0 & y_i = \boldsymbol{x}_i' \boldsymbol{\beta}_\tau \\ 1 - \tau & y_i < \boldsymbol{x}_i' \boldsymbol{\beta}_\tau . \end{cases}$$

Asymmetric weights

- Visualization of the pinball loss $w_\tau(y)|y - q|$ for different values of τ:



Minimizing Weighted Absolute Errors

- Linear programming can be used to determine the quantile regression estimates.
- Introduce $2n$ auxiliary variables

$$\begin{aligned} u_i &= (y_i - \boldsymbol{x}_i' \boldsymbol{\beta}_\tau)_+ \\ v_i &= (\boldsymbol{x}_i' \boldsymbol{\beta}_\tau - y_i)_+ \end{aligned}$$

- Then the minimisation problem can be rewritten as

$$\min_{\boldsymbol{\beta}_\tau, \boldsymbol{u}, \boldsymbol{v}} \{\tau \mathbb{1}' \boldsymbol{u} + (1-\tau) \mathbb{1}' \boldsymbol{v} | \boldsymbol{X} \boldsymbol{\beta}_\tau + \boldsymbol{u} - \boldsymbol{v} = \boldsymbol{y}\}$$

  which defines a linear minimization problem subject to a polyhedric constraint.

Quantile Crossing

- For the theoretical quantiles we have

$$Q_{\tau_1}(y) \leq Q_{\tau_2}(y) \quad \text{if} \quad \tau_1 \leq \tau_2.$$

- This does not necessarily hold for the estimated quantiles $\boldsymbol{x}'\hat{\beta}_{\tau_1}$ and $\boldsymbol{x}'\hat{\beta}_{\tau_2}$. However,

$$\bar{\boldsymbol{x}}'\hat{\beta}_{\tau_1} \leq \bar{\boldsymbol{x}}'\hat{\beta}_{\tau_2} \quad \text{for} \quad \tau_1 \leq \tau_2$$

  is always fulfilled.

Additive Quantile Regression

- Conceptually, it is straightforward to replace the linear predictor $\boldsymbol{x}_i'\boldsymbol{\beta}_\tau$ in quantile regression with an additive predictor.
- However, quadratic penalties do not fit well with the absolute error criterion such that linear programming can no longer be used.
- There have been attempts to define quantile smoothing splines based on L_1-penalties but the resulting curves are often very wiggly.

Bayesian Quantile Regression

- A Bayesian variant of quantile regression can be obtained by assuming

$$y_i = \eta_{i\tau} + \varepsilon_{i\tau}, \quad i = 1, \dots, n,$$

with independent and identically distributed errors following an asymmetric Laplace distribution, i.e., $\varepsilon_{i\tau} \mid \sigma^2$ i.i.d. $\mathrm{ALD}(0, \sigma^2, \tau)$ with density

$$f(\varepsilon_{i\tau} \mid \sigma^2) = \frac{\tau(1-\tau)}{\sigma^2} \exp\left(-w_\tau(\varepsilon_{i\tau}, 0)\frac{|\varepsilon_{i\tau}|}{\sigma^2}\right)$$

and $w_\tau(y_i, \eta_{i\tau})$ as in frequentist quantile regression.

Bayesian Quantile Regression

- For the responses, this implies $y_i \mid \eta_{i\tau}, \sigma^2 \sim \mathrm{ALD}(\eta_{i\tau}, \sigma^2, \tau)$, such that the density of the responses is given by

$$f(y_i \mid \eta_{i\tau}, \sigma^2) = \frac{\tau(1-\tau)}{\sigma^2} \exp\left(-w_\tau(y_i, \eta_{i\tau}) \frac{|y_i - \eta_{i\tau}|}{\sigma^2}\right).$$

- The asymmetric Laplace distribution has a latent Gaussian scale-mixture representation that enables the construction of a Gibbs sampler.
- It is then also easy to consider structured additive predictor structures.