Scalar-on-Function Bayesian Quantile Regression with Heteroskedastic Data

Nicolas Escobar

Background

Linear quantile regression

Simple setting: $P[Y_i \leq y] = F(y|X_i)$, scalars:

- Choose a quantile $0 < \tau < 1$
- Assume $Q_\tau(Y_i|X_i) = \beta_0 + \beta_1 X_i$ where $P[Y_i \leq Q_\tau(Y_i|X_i)|X_i] = \tau$
- Estimate β_0, β_1 by

$$(\hat{\beta}_0, \hat{\beta}_1) = \operatorname{argmin}_{\beta_0,\beta_1} \sum_i \rho_\tau(y_i - \beta_0 + \beta_1 x_i)$$

where $\rho_\tau(u) = (\tau - I(u < 0))u$

Bayesian Quantile Regression

- At first sight, a Bayesian approach to quantile regression is puzzling, because there is no likelihood
- Asymmetric Laplace (AL) distribution:

$$\log f_\tau(y|\mu, \sigma) \propto -\rho_\tau \left(\frac{y - \mu}{\sigma} \right)$$

- Bayesian Quantile Regression (BQR): We don't know F, but we use the model $f_Y(y) = f_\tau(y|\beta_{0,\tau} + \beta_{1,\tau} x_i, \sigma)$
- Equivalently $Y_i = \beta_{0,\tau} + \beta_{1,\tau} x_i + \epsilon_i$, with $\epsilon_i \sim f_\tau(\cdot|0, \sigma)$
- We perform Bayesian inference on $\beta_{0,\tau}, \beta_{1,\tau}, \sigma$, usually with MCMC

GAL distribution

- AL distribution is too rigid. For instance, it's always symmetric for median regression
- AL admits the mixture representation

$$\epsilon = A\nu + u\sqrt{\sigma B\nu}$$

  where A, B, σ are constants, $\nu \sim \mathrm{Exp}(1)$ and $u \sim \mathrm{N}(0, 1)$
- Yan and Kottas (2017) propose the following generalization:

$$\epsilon = \alpha s + A\nu + u\sqrt{\sigma B\nu}$$

  where α is a constant, $s \sim \mathrm{N}^{+}(0, 1)$
- This is called the Generalized Asymmetric Laplace (GAL) distribution. It can be reparametrized in terms of $\tau, \mu, \sigma, \gamma$.
- Thus, the model becomes $Y_i = \beta_{0,\tau} + \beta_{1,\tau} x_i + \epsilon_i$ with $\epsilon_i \sim \mathrm{GAL}_\tau(0, \sigma, \gamma)$

Hamiltonian Monte Carlo

- HMC has gained popularity as an alternative to Gibbs, Metropolis
- Stan programming language
- rstan implements Stan in R
- brms uses rstan to implement formula syntax (as in lme4) and nonparametric capabilities (with mgcv)

Our problem

Low dimensional ilustration

- Consider data generated according to

$$y_i = x_i^2 + \delta_i$$

where $\delta_i \sim N(0, \exp(4x_i))$

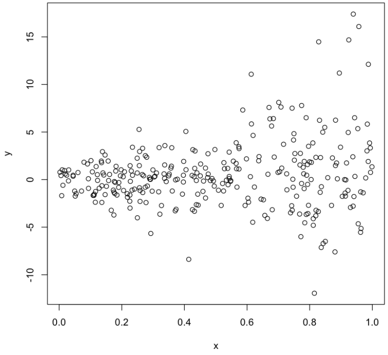
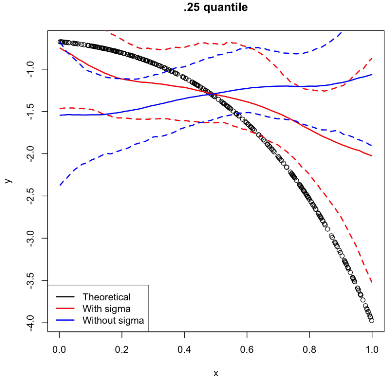


Figure 1: Heteroskedastic data

Warning

- Quantile regression should be able to handle this kind of heteroskedasticity
- It turns out you have to be careful:

```
model1 <- brm(y ~ s(x), family = 'GAL')
model2 <- brm(y ~ s(x), log sigma ~ s(x),
                          family = 'GAL')
```



Literature

- The failure of model1 to model heteroskedastic data has gone unnoticed in the literature
- It has been recognized that GAL is still not flexible enough
  - Dirichlet's mixtures of GAL distributions have become popular: $\mathrm{MGAL}_\tau = \sum_k \pi_k \mathrm{GAL}_\tau(0, \sigma_k, \gamma_k)$
- Two problems:
  - Computationally challenging
  - Blunt

Framework

- Consider data generated as

$$Y_i = \beta_0 + \mathbf{Z}_i^T \boldsymbol{\beta}_z + \int_I \mathbf{X}_i(t)^T \boldsymbol{\beta}_1(t)dt \\ + \exp\left(\eta_0 + \mathbf{Z}_i^T \boldsymbol{\eta}_z + \int_I \mathbf{X}_i(t)^T \boldsymbol{\eta}_1(t)dt\right)\delta_i$$

- Exponential heteroskedasticity.
- No distributional assumption on δ_i, other than iid.
- Easy to see that

$$Q_\tau(Y_i|\mathbf{Z}_i, X_i) = \beta_0 + \mathbf{Z}_i^T \boldsymbol{\beta}_z + \int_I \mathbf{X}_i(t)^T \boldsymbol{\beta}_1(t)dt \\ + \exp\left(\eta_0 + \mathbf{Z}_i^T \boldsymbol{\eta}_z + \int_I \mathbf{X}_i(t)^T \boldsymbol{\eta}_1(t)dt\right) q_\tau$$

q_τ being the τ quantile of δ_i

Basis expansion

- Write

$$\mathbf{X}_i = \sum_{k=1}^{K} X_{i,k}\mathbf{b}_k$$

where the $\mathbf{b}_k$'s are some set of basis functions

- Denote

$$\tilde{\mathbf{X}}_i = (X_{i,1}, \ldots, X_{i,K})^T$$
$$\tilde{\beta}_{1,k} = \int_I \mathbf{b}_k(t)^T \boldsymbol{\beta}_1(t)dt$$
$$\tilde{\boldsymbol{\beta}}_1 = (\tilde{\beta}_{1,1}, \ldots, \tilde{\beta}_{1,K})^T$$

Define $\tilde{\boldsymbol{\eta}}_1$ similarly

Approach

- We try to estimate

$$Q_\tau(Y_i|\mathbf{Z}_i, X_i) = \beta_0 + \mathbf{Z}_i^T \boldsymbol{\beta}_z + \tilde{\mathbf{X}}_i^T \tilde{\beta}_1 + \exp\left(\eta_0 + \mathbf{Z}_i^T \boldsymbol{\eta}_z + \tilde{\mathbf{X}}_i^T \tilde{\eta}_1\right) q_\tau$$

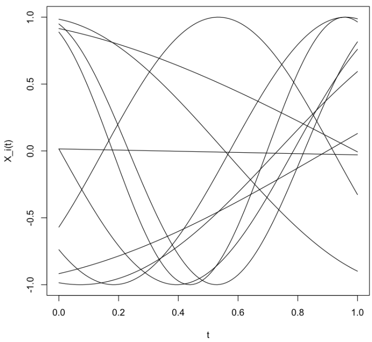
- Model:

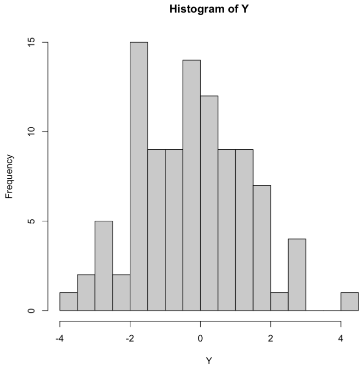
$$Y_i = \lambda_{0,\tau} + \mathbf{Z}_i^T \boldsymbol{\lambda}_{Z,\tau} + \tilde{\mathbf{X}}_i^T \tilde{\boldsymbol{\lambda}}_{1,\tau} + \epsilon_i$$

- $\epsilon_i \sim \mathrm{GAL}_\tau(0, \sigma_i, \gamma)$ where $\log \sigma_i = \kappa_0 + \mathbf{Z}_i^T \boldsymbol{\kappa}_Z + \tilde{\mathbf{X}}_i^T \tilde{\boldsymbol{\kappa}}_1$

Simulations (Preliminary)

- Choose $D = 1$, $I = [0, 1]$, $\delta_i = \mathrm{N}(0, 1)$, $N = 100$
- Generate $\omega_i, \phi_i \sim U(0, 1)$
- Set $X_i = \sin(2\pi(\omega_i t + \phi_i))$
- Generate $Z_i \sim N(0, 1)$
- $\beta_0 = 0$, $\beta_z = 1$, $\beta_1(t) = \cos(2\pi t)$
- $\eta_0 = 0$, $\eta_z = -.1$, $\eta_1(t) = .1\cos(2\pi(t + 1/8))$





Results

- Fit using brms
- Normal priors
- 2000 iterations
- Divergent transitions

Table 1: Results

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	-0.875	0.129	-1.133	-0.634
sigma_Intercept	-0.999	0.151	-1.326	-0.728
Z	1.049	0.119	0.805	1.282

Diagnostics

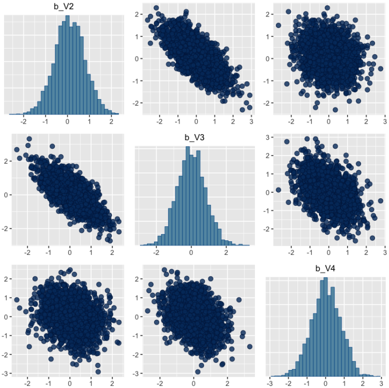


Figure 5: Pairs

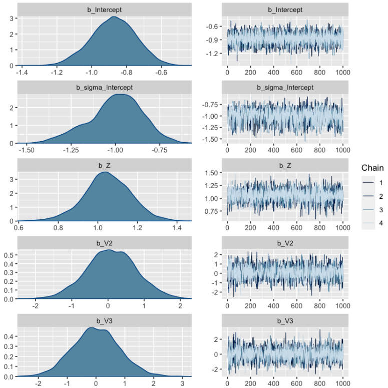


Figure 6: Trace

- Challenging to get posterior predictive checks