Bayesian Quantile Regression

Keming Yu
Brunel University London

Outline

- Quantile regression (QR)
- Bayesian inference quantile regression (BQR)
- Methods/algorithms
- Conclusions

Quantile regression

- Regression or regression model usually refers to modelling the conditional expectation of Y given $\boldsymbol{X}$

$$E(Y|\boldsymbol{X} = \boldsymbol{x}).$$

- Regression model typically measures the relationship between Y and $\boldsymbol{X}$,

- it could be used to identify the dependency or effect,

- it can also be used for prediction.

Quantile regression

- **QR model is to modelling the conditional quantile of Y given $\boldsymbol{X}$**

$$Q_\tau(Y|\boldsymbol{x}),$$

  where $0 < \tau < 1$ stands for the τth quantile of Y.

- **For example, median regression with $\tau = 0.5$.**

- **If the conditional distribution of Y given $\boldsymbol{X}$ is symmetric, then the median regression is identical to the mean regression, otherwise, they are different.**

- **In general,** $E(y|\boldsymbol{x}) = \int_0^1 Q_\tau(Y|\boldsymbol{x}) d\tau$

  $\Longrightarrow$ **mean regression model can be viewed as a summary of all the quantile effects.**

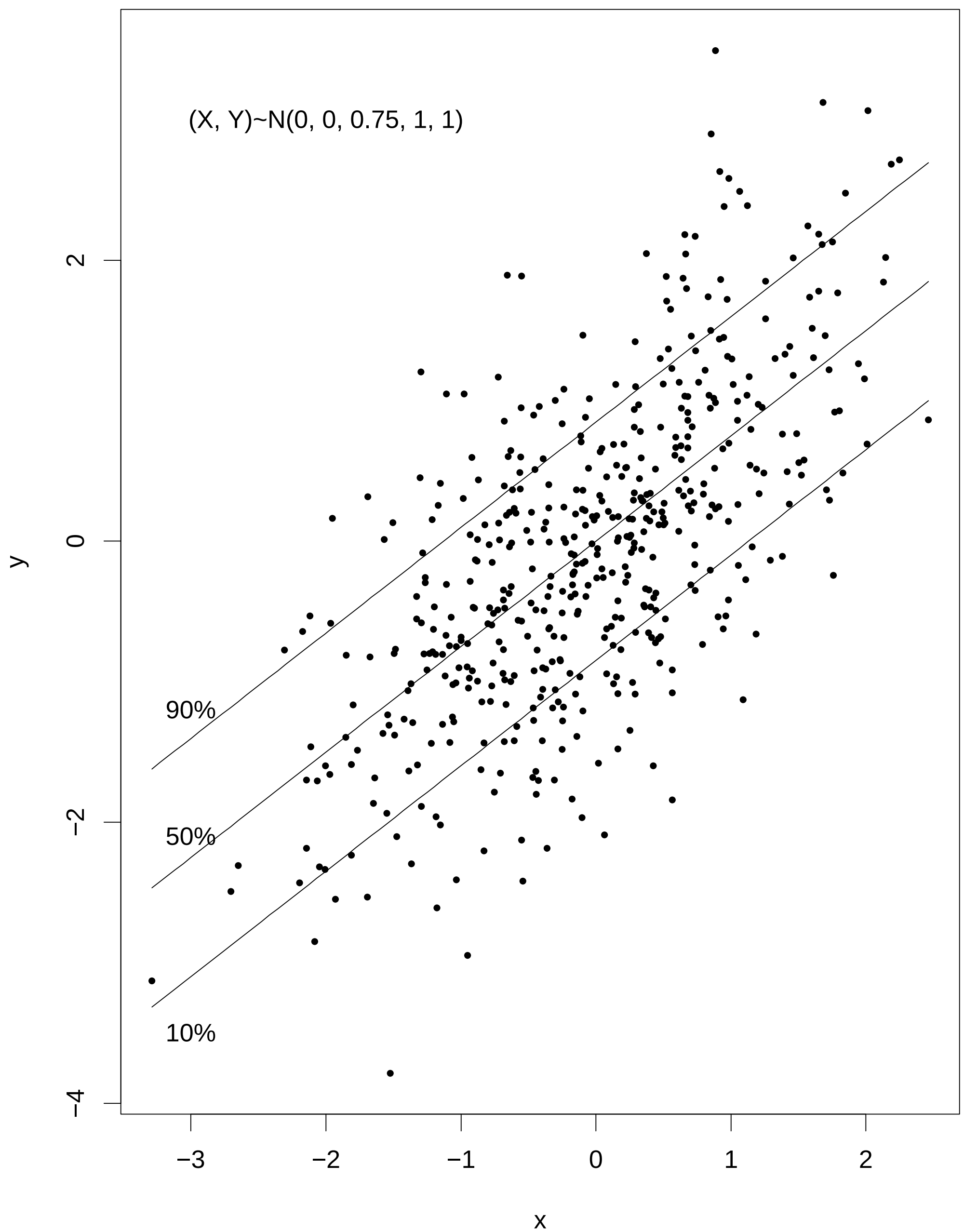
  $\Longrightarrow$ **QR gives a deep analysis of the way that** Y **and** X **are related.**

An example of a simple linear quantile regression: when $(X, Y) \sim$ bivariate normal distribution $N(0, 0, r, 1, 1)$,

$$Q_\tau(x) = rx + (1 - r^2)\Phi^{-1}(\tau);$$

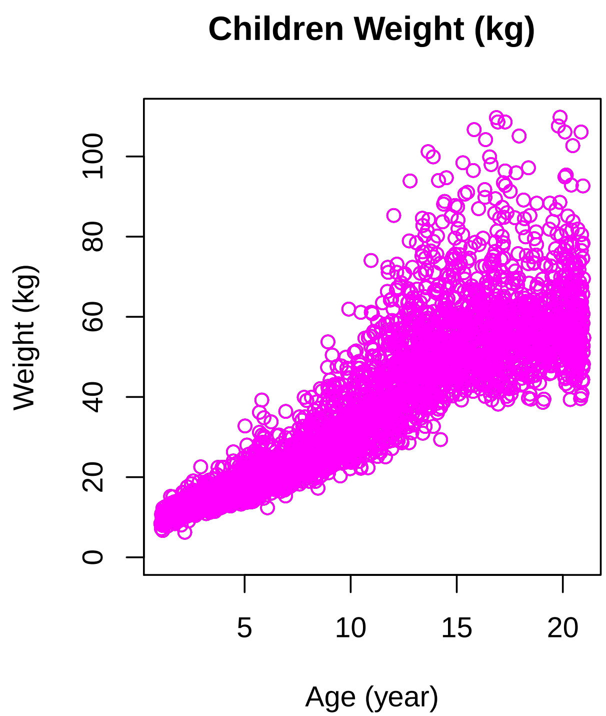
where $\Phi^{-1}(\tau)$ denotes the inverse of standard normal distribution.

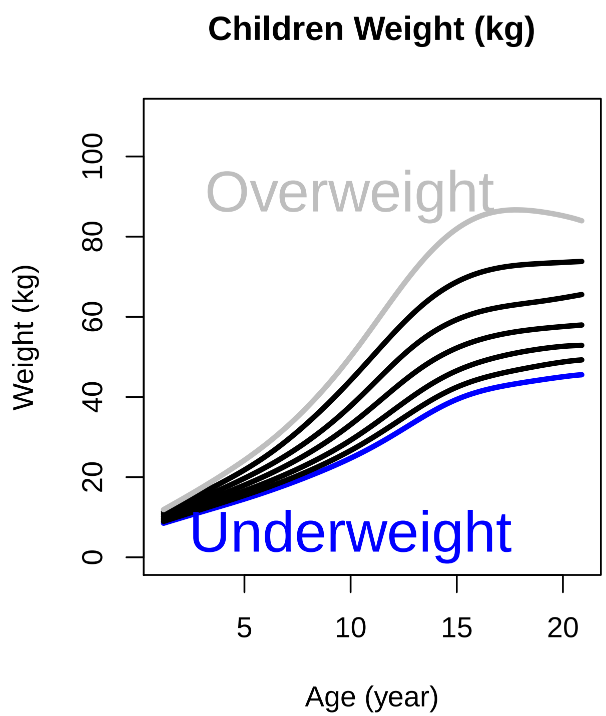
Due to the symmetry of the conditional distribution, all quantile curves $q_{0.1}(x)$, $q_{0.5}(x)$ and $q_{0.9}(x)$ from $N(0, 0, 0.75, 1, 1)$ are parallel.



(X, Y)~N(0, 0, 0.75, 1, 1)
90%
50%
10%
y
x
2
0
−2
−4
−3
−2
−1
0
1
2

Otherwise, as see from the 4011 US girl weights against ages:





Applications of quantile regression include

- **risk measurement:**
- **the commonly used risk measurers in finance is called VaR (value at risk), which actually corresponding a tail value of an asset price.**
- **People are more concerned the upper pollution level than an average one.**

Applications of quantile regression include

- analysis of fat-tailed distribution:

- does cheaper food contribute to children obesity?

- does a specific diet make difference on children's weights?

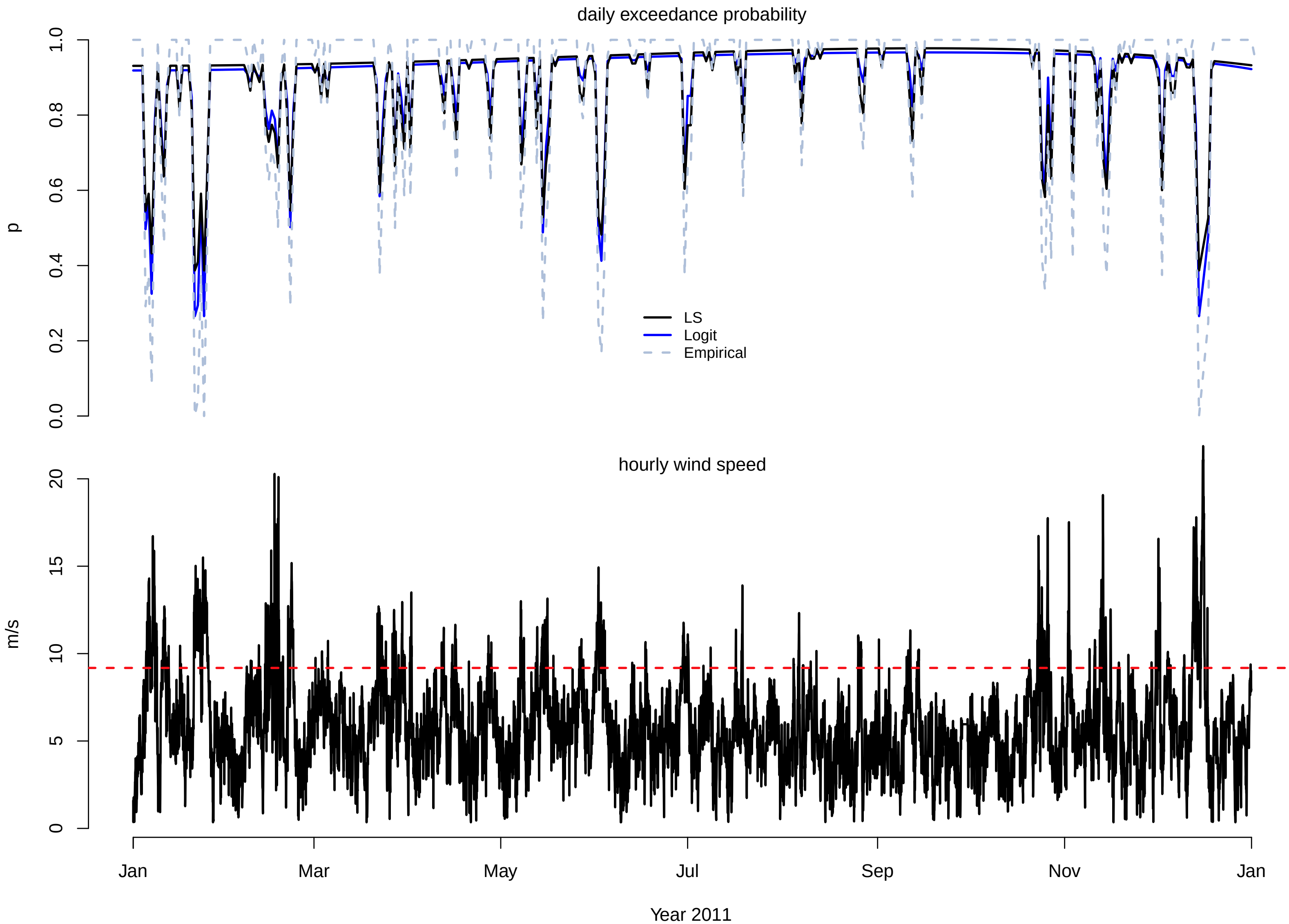
- **The probability of Y exceeding a threshold Q:**

- **given Q, find τ: $\tau = Pr[Y \leq Q|X]$.**

- **forecast of the time-varying probability of a financial return exceeding a given threshold;**

- **In wind-farm to predict/check if the wind speed exceeds a certain level;**

- **flooding prediction and energy demand prediction.**



daily exceedance probability
p
1.0
0.8
0.6
0.4
0.2
0.0
LS
Logit
Empirical
hourly wind speed
m/s
20
15
10
5
0
Jan
Mar
May
Jul
Sep
Nov
Jan
Year 2011

Basic fitting method of QR

- **mean regression satisfies**

- $E(Y|X = x) = argmin_a E[(Y - a)^2|X]$.

- **an QR $q_\tau(x)$ satisfies**

- $\tau = Pr[Y \leq q_\tau(x)|X = x]$.

where the 'Check function'

- $\rho_\tau(u) = [\tau - I(u < 0)]u = \{ \begin{array}{cc} \tau u, & u > 0 \\ -(1-\tau)u, & u \leq 0. \end{array}$

- **Given the data** $\{Y_i, X_i\}_{i=1}^n$ **to fit a linear QR model** $Y = \boldsymbol{x}\boldsymbol{\beta} + \epsilon$ **then**

- $\widehat{\beta}(\tau) = argmin_b \sum_{i=1}^n \rho_\tau(Y_i - \boldsymbol{X}_i' b)$.

‘Working’ likelihood function for QR

- **Mean regression corresponds to** *loss* **function** $L(u) = u^2$ **and normal likelihood with pdf** $f_N(u) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(y-\mu)}{2\sigma^2})$.

- **QR corresponds to** *check* **function** $\rho_\tau(u)$ **and the asymmetric Laplace likelihood (ALD) with pdf**

$$f_{ALD}(u) = \frac{\tau(1-\tau)}{\sigma} \exp(-\rho_\tau(\frac{y-\mu}{\sigma})).$$

- **As minimizing the ‘check function’** $\Longleftrightarrow$ **maximizing an ALD-based likelihood function.**

Bayesian Quantile Regression (BQR)

- Regarding the regression parameters β from a linear QR model $Q_\tau(Y|\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta}$ as a random variable, aiming at its posterior distribution according to Bayes Theorem:

$$Pr(\boldsymbol{\beta}|.) \approx Pr(\boldsymbol{\beta}) \times Pr(.|\boldsymbol{\beta}).$$

- Bayesian inference depends on prior and likelihood function.

- All selected likelihoods may not be true but some are working.

- **ALD has good performance on data generated from many error distributions (Ji et al. 2012; Li et al. 2010; among others) and theoretic justification:**

- **Komunjer (2005) Quasi-Maximum Likelihood Estimation for Conditional Quantiles,** *J. Econometrics***.**
  Sriram, Ramamoorthi and Ghosh (2013). Posterior consistency of Bayesian quantile Regression under a mis-specified likelihood based on asymmetric Laplace density, *Bayesian Analysis***.**

BQR

- **Yu and Moyeed(2001) used Metropolis Hastings (M-H) on joint distribution** $Pr(\boldsymbol{\beta}|...)$.

- **R packages:**

  `http://cran.r-project.org/web/packages/bayesQR/bayesQR.pdf`

  `http://hosho.ees.hokudai.ac.jp/~kubo/Rdoc/library/MCMCpack/html/MCMCquantreg.html`

Gibbs sampling for BQR

- The advantage of Gibbs sampler over Metropolis Hastings (MH) algorithm:

- Gibbs sampling doesn't have the convergence issue as MH algorithm has, and unlike M-H algorithm which has proposal distribution selection, the proposal distribution of Gibbs sampling is simply taken to be the conditional distributions of the target distribution.

- Each density to be sampling is of low dimension (often one dimensional) and thus it is very easy and efficient to sample from it.

Gibbs sampling for BQR

- **ALD as a Mixture of Normals**

$$
\begin{aligned}
f_\tau(z;\mu,\sigma) &= \frac{\tau\,(1-\tau)}{\sigma}\,\exp\left\{-\rho_\tau\left(\frac{z-\mu}{\sigma}\right)\right\} \\
&= \int_0^\infty \frac{1}{2\sqrt{\pi\sigma w}}\,\exp\left\{-\frac{1}{4\sigma w}\{z-\mu-(1-2\tau)w\}^2\right\} \\
&\times \frac{\tau(1-\tau)}{\sigma}\exp\left\{-\frac{\tau(1-\tau)}{\sigma}w\right\}\,\mathrm{d}w
\end{aligned}
$$

- **This extends the results (Laplace distribution with $\tau = 0.5$ here) of West (1987) and Andrews and Mallows (1974) to the ALD.**

Gibbs sampling for BQR

- **That is,**

$$\epsilon = \mu\, v + \delta \sqrt{\sigma\, v}\, u,$$

where $\mu = \frac{1-2\tau}{\tau(1-\tau)}$ **and** $\delta^2 = \frac{2}{\tau(1-\tau)}$**,**

- v **and** u **are independent and follow standard exponential distribution and normal distribution respectively:**

- $v \sim Exp(1/\sigma)$ **and** $u \sim N(0, 1)$**.**

- **Therefore the conditional distribution** $f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma, \boldsymbol{v})$ **under** $y = \boldsymbol{\beta}'\boldsymbol{X} + \epsilon$ **consists of independent normal distribution** $N(\boldsymbol{\beta}'\boldsymbol{X}_t + \mu\sigma z_t,\ \delta^2\sigma^2 z_t)$.

- **Once the normal-Gamma conjugate prior of a normal distribution is provided for** $(\boldsymbol{\beta}, \sigma)$**, we can construct a Gibbs sampler for inference of the posterior conditional densities of all quantities.**

- **To this end we assume that the prior** $\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0,\ B_0)$ **and** $\sigma \sim IG(n_0/2,\ s_0/2)$**, an inverse Gamma distribution** $IG(a, b)$ **with parameters** $a = n_0/2$ **and** $b = s_0/2$**.**

- **Then the posterior of β still follows normal distribution**

$$\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{v},\sigma \sim N(\boldsymbol{\beta}_p, B_p)$$

with $B_p^{-1} = \sum_{i=1}^{n}(\boldsymbol{X}_i\boldsymbol{X}_i'/\delta^2\sigma v_i) + B_0^{-1}$ **and** $\boldsymbol{\beta}_p = B_p(\sum_{i=1}^{n}(\boldsymbol{X}_i(Y_i - \mu v_i)/\delta^2\sigma v_i) + B_p^{-1}\boldsymbol{\beta}_0)$.

- **The posterior distribution for v_i follows a generalized inverse Gaussian distribution**

$$v_i|\boldsymbol{y},\boldsymbol{\beta},\sigma \sim GIG(\frac{1}{2},\alpha_i,\gamma_i),$$

with $\alpha_i^2 = (Y_i - \boldsymbol{\beta}'\boldsymbol{X}_i)^2/\delta^2\sigma$ **and** $\gamma_i^2 = 2/\sigma + \mu^2/\delta^2\sigma$.

- **The posterior distribution for σ follows an inverse Gamma distribution**

$$\sigma|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{v} \sim IG(\frac{n^*}{2},\frac{s^*}{2}),$$

with $n^* = n_0 + 3n$ **and** $s^* = s_0 + 2\sum_{i=1}^{n} v_i + \sum_{i=1}^{n}(Y_i - \boldsymbol{\beta}'\boldsymbol{X}_i - \mu\, v_i)^2/\delta^2 v_i$.

- **A Gibbs sampler successively sampling from**

$$\sigma|\boldsymbol{\beta},\boldsymbol{v},\boldsymbol{y}$$
$$\boldsymbol{v}|\boldsymbol{\beta},\sigma,\boldsymbol{y}$$
$$\boldsymbol{\beta}|\sigma,\boldsymbol{v},\boldsymbol{y}$$

converges to $p(\boldsymbol{\beta},\sigma,\boldsymbol{v}|\boldsymbol{y})$**.**

R packages:

R-Package lqmm (cran.r-project.org/package=lqmm)
for panel (longitudinal) data.

R-package 'bayesQR'
cran.r-project.org/web/packages/bayesQR/bayesQR.pdf.

R-package Package 'Brq'
cran.r-project.org/web/packages/Brq/Brq.pdf
for BQR with cross-section data.

R-package BSquare is also for BQR with panel data.

Mathematics Justification of ALD-likelihood

- Komunjer (2005): Quasi-Maximum Likelihood Estimation for Conditional Quantiles," Journal of Econometrics, 128, 137–164.

- Komunjer (2005): likelihood should belong a tick-exponential family of densities – family whose role in QR is analog to the role of the linear-exponential family in mean regression estimation.

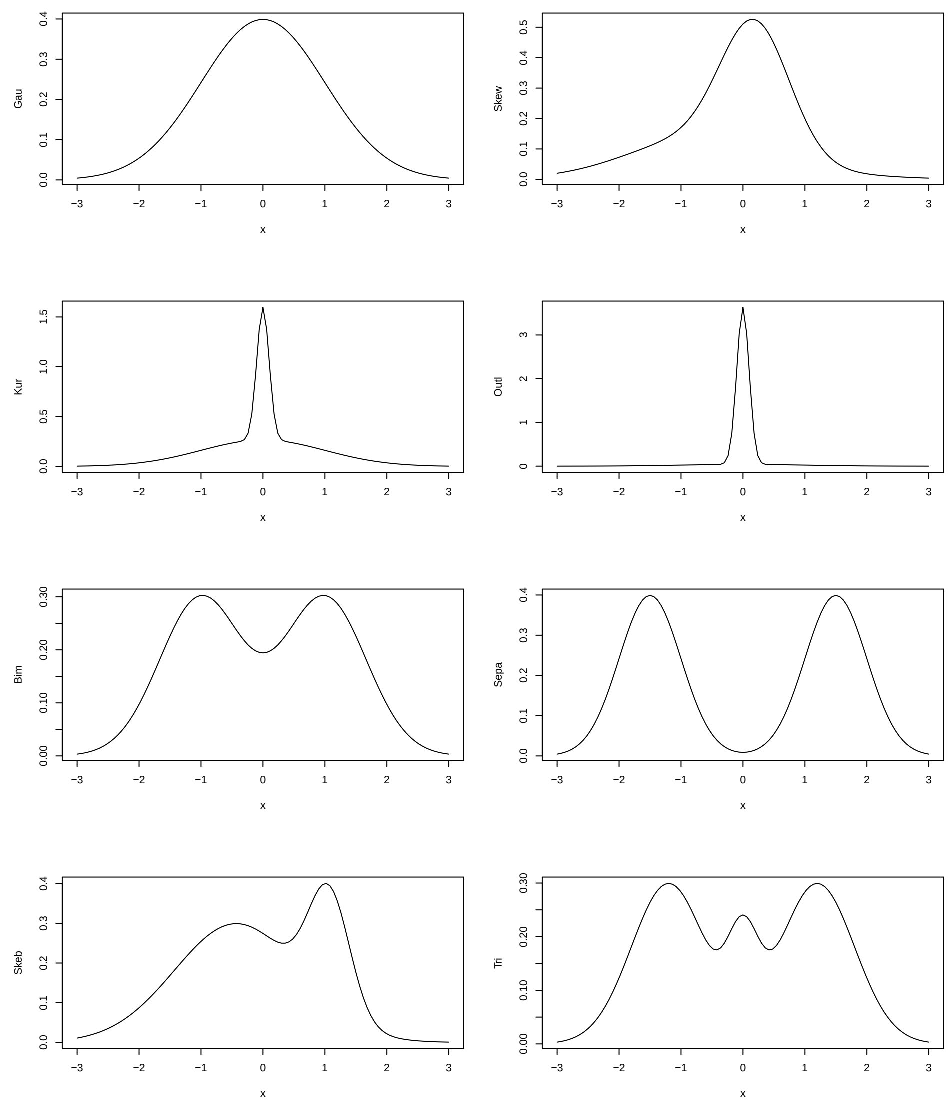
- The "well-known member of the tick-exponential family is the asymmetric Laplace density..."

How accurate of Gibbs sampler with ALD-likelihood?

Consider fitting a QR model $y = Q_\tau(y|x) + \epsilon$ with BQR and eight different model errors.

The 8 error distributions chosen to represent a wide variety of characteristics that a true error distribution may possess.

- **Gaussian (Gau):** $N(0, 1^2)$,
- **Skewed (Skew):** $\frac{1}{5}N(\frac{-22}{25}, 1^2) + \frac{1}{5}N(-\frac{49}{125}, \frac{3}{2}^2) + \frac{3}{5}N(\frac{49}{250}, \frac{5}{9}^2)$,
- **Kurtotic (Kur):** $\frac{2}{3}N(0, 1^2) + \frac{1}{3}N(0, \frac{1}{10}^2)$,
- **Outlier (Outl):** $\frac{1}{10}N(0, 1^2) + \frac{9}{10}N(0, \frac{1}{10}^2)$,
- **Bimodal (Bim):** $\frac{1}{2}N(-1, \frac{2}{3}^2) + \frac{1}{2}N(1, (\frac{2}{3})^2)$,
- **Bimodal, separate modes (Sepa):** $\frac{1}{2}N(-\frac{3}{2}, \frac{1}{2}^2) + \frac{1}{2}N(\frac{3}{2}, \frac{1}{2}^2)$,
- **Skewed bimodal (Skeb):** $\frac{3}{4}N(-\frac{43}{100}, 1^2) + \frac{1}{4}N(\frac{107}{100}, \frac{1}{3}^2)$,
- **Trimodal (Tri):** $\frac{9}{20}N(-\frac{6}{5}, \frac{3}{5}^2) + \frac{9}{20}N(\frac{6}{5}, \frac{3}{5}^2) + \frac{1}{10}N(0, \frac{1}{4}^2)$.



Gau
Skew
Kur
Outl
Bim
Sepa
Skeb
Tri
x

Error distributions. Reading from left to right, top to bottom, the densities are Gaussian (Gau), Skew (Skew), Kurtotic (Kur), Outlier (Outl), Bimodal (Bim), Bimodal, separate modes (Sepa), Skew Bimodal (Skeb) and Trimodal (Tri).

- **Generate** y_i**,** $i = 1, \ldots, 50$ **from the model** $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i$**.**

- **For this model, the conditional quantiles are given by**

$$Q_\tau(y_i|\boldsymbol{x_i}) = \alpha_0(\tau) + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i},$$

- **where** $\alpha_0(\tau) = \beta_0 + F_\epsilon^{-1}(\tau)$**. We set** $\beta_0 = 0$**,** $\beta_1 = \beta_2 = \beta_3 = 1$ **and generated 100 replications from each of the 8 models, giving a total of 800 datasets.**

Under improper prior for regression parameter, we ran the Gibbs sampler for 11,000 iterations with 1,000 of those discarded as burn in.

For each of the 100 simulated datasets corresponding to a particular error distribution, we recorded the posterior means.

Table 1 presents the average of the posterior means together with the standard error in parentheses. We analysed both $\tau = 0.5$ and $\tau = 0.1$.

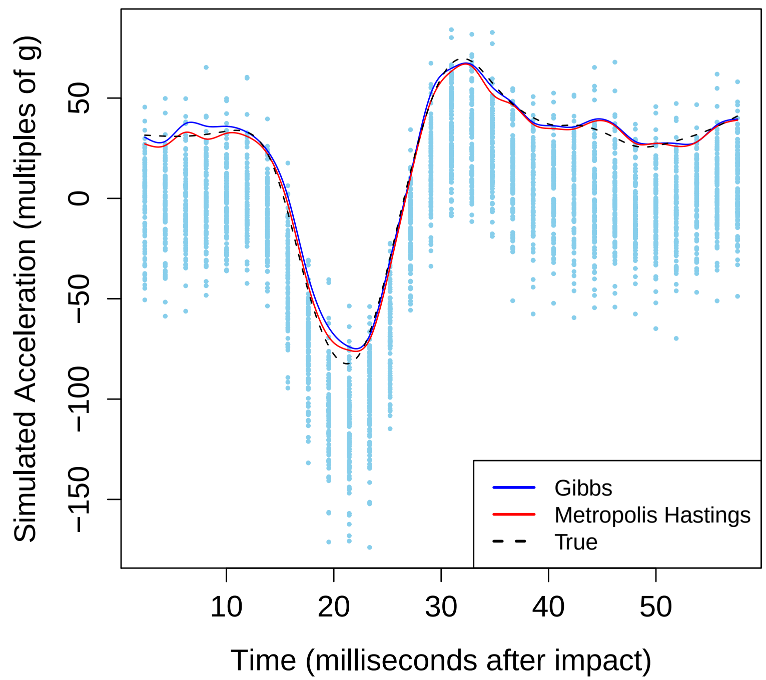
	$\alpha_0(\tau)$	β_1	β_2	β_3
$\tau = 0.1$				
Gau	**-1.540 (0.202)**	**1.001 (0.178)**	**0.964 (0.166)**	**1.015 (0.202)**
Skew	**-1.782 (0.289)**	**0.991 (0.227)**	**1.009 (0.259)**	**0.974 (0.261)**
Kur	**-1.339 (0.205)**	**0.991 (0.170)**	**1.013 (0.164)**	**0.995 (0.160)**
Outl	**-0.770 (0.114)**	**1.001 (0.071)**	**0.993 (0.074)**	**1.009 (0.076)**
Bim	**-1.819 (0.170)**	**0.975 (0.200)**	**1.017 (0.197)**	**1.015 (0.187)**
Sepa	**-2.172 (0.164)**	**1.000 (0.203)**	**0.982 (0.176)**	**1.018 (0.197)**
Skeb	**-1.820 (0.228)**	**1.015 (0.233)**	**1.014 (0.194)**	**0.997 (0.216)**
Tri	**-1.861 (0.192)**	**0.986 (0.210)**	**1.006 (0.196)**	**0.999 (0.203)**
$\tau = 0.5$				
Gau	**0.010 (0.176)**	**0.970 (0.164)**	**1.009 (0.168)**	**0.987 (0.141)**
Skew	**-0.054 (0.142)**	**0.996 (0.150)**	**1.006 (0.166)**	**1.001 (0.137)**
Kur	**-0.003 (0.082)**	**1.011 (0.080)**	**0.996 (0.101)**	**0.984 (0.101)**
Outl	**-0.003 (0.022)**	**0.999 (0.028)**	**0.995 (0.030)**	**0.996 (0.026)**
Bim	**-0.004 (0.241)**	**1.026 (0.212)**	**1.020 (0.223)**	**0.974 (0.249)**
Sepa	**0.093 (0.415)**	**1.039 (0.350)**	**1.050 (0.369)**	**0.978 (0.344)**
Skeb	**0.010 (0.196)**	**1.031 (0.201)**	**1.001 (0.199)**	**0.999 (0.201)**
Tri	**0.015 (0.247)**	**0.983 (0.203)**	**0.971 (0.268)**	**1.001 (0.272)**

Summary statistics based on calculating the posterior mean of the 100 datasets for each of the 8 diffrerent error distributions. The true values of $\alpha_0(\tau)$ **for** $\tau = 0.5$ **are 0 and for** $\tau = 0.1$ **are -1.282(Gau), -1.502(Skew), -1.036(Kur), -0.154(Outl), -1.561(Bim), -1.921(Sepa), -1.541(Skeb) and -1.659(Tri).**

Bayesian inference quantile curves using natural cubic splines

The response variable y of the motorcycle data is a record of the head acceleration, measured in multiples of the acceleration due to gravity g. The explanatory variable x is the times, measured in milliseconds, after a simulated motorcycle accident.

Thompson *et al.* (2010) ran the MH algorithm for 250,000 iterations discarding 50,000 as burn in and retaining every 10th iteration to reduce autocorrelation and for storage purposes. We plot the NCS obtained by the MH algorithm and that obtained by the Gibbs sampler using 11,000 iterations, 1,000 of which were discarded as burn in. We analysed $\tau = 0.95$.



Bayesian inference (both the MH algorithm and the Gibbs sampler) produce curves that can accurately reconstruct the true underlying curve and are very similar to each other.

ALD-likelihood extension

- **It can provide many extensions of BQR.**
- **For example:**
- **Bayesain spatial QR (Lum and Gelfand, 2012) in location s with spatial model error:**
- $\epsilon_p(\boldsymbol{s}) = \sqrt{\frac{2\xi(\boldsymbol{s})}{\sigma p(1-p)}} Z(\boldsymbol{s}) + \frac{1-2p}{p(1-p)}\xi(\boldsymbol{s})$
- **and** $Z(\boldsymbol{s}) \sim GP(0,\ r(\boldsymbol{s}, \boldsymbol{s}'; \theta))$;

Summary

- Bayesian Quantile regression is a sensible alternative to classical Quantile regression when making inference on underlying quantile regression functions.

- There are both mathematica and numerical justifications behind.

- Both M-H algorithm and Gibbs sampling are applicable.

- **The mixture of normals representation of the ALD allows efficient Gibbs sampling.**

- **A few R packages provide user-friendly softwares.o.**

- **Future work: expert judgement for BQR via prior.**

- **Thank you**