

Monte Carlo Experiments and a Few Observations on Bayesian Quantile Regression

Justin L. Tobias

Department of Economics, Purdue University, West Lafayette, USA

Email: jl Tobias@purdue.edu

How to cite this paper: Tobias, J. L. (2024). Monte Carlo Experiments and a Few Observations on Bayesian Quantile Regression. *Theoretical Economics Letters*, 14, 263-272.

<https://doi.org/10.4236/tel.2024.141015>

Received: November 25, 2023

Accepted: February 26, 2024

Published: February 29, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In many simulation-based Bayesian approaches to quantile regression, Markov Chain Monte Carlo techniques are employed to generate draws from a posterior distribution based on an asymmetric Laplace “working” likelihood. Under flat improper priors, the mode of this posterior distribution is coincident with the desired quantile function. However, simulation-based approaches for estimation and inference commonly report a posterior mean as a point estimate and interpret that mean synonymously with the quantile. In this note, we analytically derive the exact posterior distribution of a quantile regression parameter in a simple univariate setting free of covariates. We note the non-uniqueness of the posterior mode in some cases and conduct a series of Monte Carlo experiments to compare the sampling performances of posterior means and modes. Interestingly, and perhaps surprisingly, the mean performs similarly to, if not favorably to, the mode under several standard metrics, even in very small samples.

Keywords

Bayesian, Quantile Regression, Gibbs Sampling, Posterior Inference

1. Introduction

Quantile regression, as first described in Koenker and Bassett (Koenker & Bassett, 1978), is a powerful statistical tool that allows researchers to examine the influence of covariates across different points of the conditional outcome distribution. The quantile regression estimator is produced by minimizing a sum of potentially asymmetric absolute deviations of the form:

$$\hat{\beta}_p \equiv \arg \min_{\beta_p} \sum_{i=1}^n \rho_p(y_i - \mathbf{x}_i \beta_p), \quad (1)$$

where ρ_p is the so-called “check function,” defined as

$$\rho_p(z_i) \equiv z_i [p - I(z_i < 0)].$$

While quantile regression has a long and rich history in the frequentist literature, Bayesian approaches to quantile regression are more recent and, arguably, fewer in comparison. Yu and Moyeed (Yu & Moyeed, 2001), however, provide a likelihood-based interpretation of quantile regression that gives rise to a Bayesian solution. They note that if one assumes the errors of a regression model follow an asymmetric Laplace distribution, then a posterior kernel is produced that is coincident with the check function defining the quantile regression estimator.

Specifically, consider a regression problem of the form:¹

$$y_i = \mathbf{x}_i \boldsymbol{\beta}_p + \sigma \varepsilon_i$$

where $\varepsilon_i = \varepsilon_i(p)$ follows an asymmetric Laplace distribution

$$f_p(\varepsilon) = p(1-p) \exp[-\varepsilon_i(p - I(\varepsilon_i < 0))],$$

yielding the likelihood function:

$$L(\boldsymbol{\beta}_p) = \sigma^{-n} p^n (1-p)^n \exp\left(-\sigma^{-1} \sum_{i=1}^n \rho_p(y_i - \mathbf{x}_i \boldsymbol{\beta}_p)\right). \tag{2}$$

Given the likelihood in (2), Yu and Moyeed (Yu & Moyeed, 2001) establish propriety of the posterior under an improper uniform prior for $\boldsymbol{\beta}_p$ and provide several applications and experiments. Kozumi and Kobayashi (Kozumi & Kobayashi, 2011) describe an alternate approach to the Metropolis-Hastings based scheme of Yu and Moyeed (Yu & Moyeed, 2001) that is computationally appealing.² They exploit an additive mixture representation of the asymmetric Laplace distribution that admits a standard Gibbs sampling-based solution for Bayesian quantile regression. Specifically, Kozumi and Kobayashi (Kozumi & Kobayashi, 2011) note that the model can be equivalently expressed as

$$y_i = \mathbf{x}_i \boldsymbol{\beta}_p + \eta v_i + \delta \sqrt{\sigma v_i} \tilde{\varepsilon}_i, \quad i = 1, \dots, n, \tag{3}$$

with

$$(\tilde{\boldsymbol{\varepsilon}} | \mathbf{X}, \boldsymbol{\nu}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n), \quad \eta \equiv \frac{1-2p}{p(1-p)} \quad \text{and} \quad \delta^2 \equiv \frac{2}{p(1-p)}.$$

Furthermore, the v_i are *iid* exponential variates:

$$p(\boldsymbol{\nu} | \sigma) = \prod_{i=1}^n (\sigma^{-1} \exp(-\sigma^{-1} v_i)).$$

This particular mixture formulation is quite useful since (a) conditioned on $\boldsymbol{\nu}$, the model is a Gaussian linear model, and thus a Gibbs implementation is relatively easy to conduct and (b) when marginalizing (3) over $\boldsymbol{\nu}$, the asymmetric Laplace distribution is produced as the sampling model. Thus, one can employ a da-

¹We consider the more general asymmetric Laplace distribution with scale parameter σ here, although σ could be fixed at one (or another value) if desired (see, e.g., Yang et al. (Yang et al., 2016)).

²Khare and Hobert (Khare & Hobert, 2012) also establish geometric ergodicity of this Gibbs algorithm.

ta-augmented Gibbs sampling algorithm to recover the quantile function $\mathbf{x}\boldsymbol{\beta}_p$. Applications and extensions of this methodology can be found in, e.g., (Alhamzawi, 2016; Rahman, 2016; Alhamzawi & Ali, 2018; Mostafaei et al., 2019; Xu & Chen, 2019; Bresson et al., 2021; Kedia et al., 2023; Zhao & Xu, 2023), among many others.

An issue that merits further investigation—an investigation we conduct briefly in section 3—is the following: Simulation-based recipes for Bayesian posterior inference, such as the Gibbs sampler of Kozumi and Kobayashi (Kozumi & Kobayashi, 2011), generate a series of draws which converge to the desired target distribution, $p(\boldsymbol{\beta}_p | \mathbf{y})$ or $p(\boldsymbol{\beta}_p, \sigma | \mathbf{y})$. These draws can then be used to calculate various features of that posterior or to plot its entire distribution. The sample average of those post-convergence draws—an estimate of the posterior *mean*—is commonly, if not exclusively, used as a point estimate of $\boldsymbol{\beta}_p$. However, the motivation for the asymmetric Laplace “working” likelihood, or its expanded representation in (3), is that the posterior *mode* (under a flat prior) coincides with the minimizer of the check function in (1). To the extent that the posterior distribution for the model parameters is asymmetric, the mode and mean can depart significantly.

In the following sections of this paper, we offer several observations and contributions. We first derive the exact posterior distribution of the location parameter of an asymmetric Laplace distribution in an environment free of covariates. We do so while allowing a scale parameter to enter the asymmetric Laplace likelihood and while employing an inverse gamma prior for that scale parameter. This setting is then used to perform a series of generated data experiments, those designed to characterize the sampling distributions of the posterior mean and posterior mode under various sample sizes. Although the posterior mean is commonly reported as a point estimate in applied work of this type, under flat priors the posterior mode is formally coincident with the desired quantile. The mode, however, is often prohibitively difficult to calculate in high-dimensional settings, while an estimate of the mean is almost immediately available given a set of draws from the posterior distribution. Our experiments are intended to shed light on the performance of the widely-used posterior mean and how its use may suffer (if at all) relative to the mode, as both features of the posterior distribution can be easily calculated in our simplified univariate setting.

2. An Intercept-Only Model

Suppose our object of interest is the p^{th} quantile, $p \in (0, 1)$, and denote the associated parameter as θ_p . Employing the flat improper prior for θ_p : $p(\theta_p) \propto c$ and the following prior for the Laplace scale parameter: $\sigma \sim IG(a_\sigma, b_\sigma)$, where IG denotes an inverse gamma distribution,³ we obtain the following joint posterior under the asymmetric Laplace likelihood:

³We parameterize the inverse gamma distribution such that

$x \sim IG(a, b) \Rightarrow p(x) \propto x^{-(a+1)} \exp(-[xb]^{-1})$. See, for example, Chan et al. (Chan et al., 2019), page 442, for further discussion.

$$p(\sigma, \theta_p | \mathbf{y}) \propto p(\sigma) \exp\left(-\sigma^{-1} \left[np(\bar{y} - \theta_p) - \sum_{i=1}^n (y_i - \theta_p) \mathbb{I}(y_i < \theta_p) \right]\right). \quad (4)$$

In (4), n denotes the number of observations, $\{y_i\}_{i=1}^n$ is our data, and \bar{y} denotes the sample mean. Integrating σ out of the equation above, we obtain

$$p(\theta_p | \mathbf{y}) \propto \left[b_\sigma^{-1} + np(\bar{y} - \theta_p) - \sum_{i=1}^n (y_i - \theta_p) \mathbb{I}(y_i < \theta_p) \right]^{-(n+a_\sigma)}. \quad (5)$$

With a bit of additional work, we can determine the normalizing constant of (5) and thus provide an exact analytical expression of the marginal posterior distribution θ_p . To this end, let $y_{(j)}$ denote the j^{th} order statistic of the sample data \mathbf{y} , and for simplicity assume there are no ties so that $y_{(1)} < y_{(2)} < \dots < y_{(n)}$. In addition, define

$$\mathcal{Y}_{(j)} \equiv y_{(1)} + y_{(2)} + \dots + y_{(j)}$$

$$\omega_{(j)} \equiv j - np$$

$$T_{(0)} = \left[-\omega_{(0)}(n + a_\sigma - 1) \right]^{-1} \left[b_\sigma^{-1} + np(\bar{y} - y_{(1)}) \right]^{-(n+a_\sigma-1)}$$

$$T_{(n)} = \left[\omega_{(n)}(n + a_\sigma - 1) \right]^{-1} \left[b_\sigma^{-1} + n(p - 1)(\bar{y} - y_{(n)}) \right]^{-(n+a_\sigma-1)}$$

and, for $j = 1, 2, \dots, n - 1$,

$$T_{(ja)} = \left[-\omega_{(j)}(n + a_\sigma - 1) \right]^{-1} \left[\left(b_\sigma^{-1} + np\bar{y} - \mathcal{Y}_{(j)} + y_{(j+1)}\omega_{(j)} \right)^{-(n+a_\sigma-1)} - \left(b_\sigma^{-1} + np\bar{y} - \mathcal{Y}_{(j)} + y_{(j)}\omega_{(j)} \right)^{-(n+a_\sigma-1)} \right]$$

$$T_{(jb)} = \left[b_\sigma^{-1} + np\bar{y} - \mathcal{Y}_{(j)} \right]^{-(n+a_\sigma)} \left[y_{(j+1)} - y_{(j)} \right].$$

Finally, letting

$$T_{(j)} = \begin{cases} T_{(ja)} & \text{if } np - j \neq 0 \\ T_{(jb)} & \text{if } np - j = 0 \end{cases} \quad \text{and } c^* = c^*(n, p, a_\sigma, b_\sigma, \mathbf{y}) = \left[\sum_{j=0}^n T_{(j)} \right]^{-1},$$

we have

$$p(\theta_p | \mathbf{y}) = c^* \left[b_\sigma^{-1} + np(\bar{y} - \theta_p) - \sum_{i=1}^n (y_i - \theta_p) \mathbb{I}(y_i < \theta_p) \right]^{-(n+a_\sigma)}. \quad (6)$$

As noted by Chan et al. (Chan et al., 2019), pages 315-317, the mode of the posterior equals $y_{(\lceil np \rceil)}$, with “ $\lceil \cdot \rceil$ ” denoting the traditional ceiling operator. Interestingly, when np is an integer, the posterior mode is not unique, and the posterior distribution has a flat ridge around the p^{th} quantile of the sample data.⁴ To see this we note, again for integer-valued np :

⁴This point is certainly not new but may be underappreciated. See, for example, Yang et al. (Yang et al., 2016) for a related discussion.

$$\begin{aligned} & \sum_{i=1}^n (y_i - y_{(np)}) \mathbb{1}(y_i < y_{(np)}) \\ &= \left(\sum_{i=1}^{np-1} y_{(i)} \right) - (np-1)y_{(np)} = \mathcal{Y}_{(np-1)} - (np-1)y_{(np)} = \mathcal{Y}_{(np)} - np y_{(np)}. \end{aligned}$$

Therefore, from (6),

$$\begin{aligned} p(\theta_p = y_{(np)} | \mathbf{y}) &= c^* \left[b_\sigma^{-1} + np(\bar{y} - y_{(np)}) - \mathcal{Y}_{(np)} + np y_{(np)} \right]^{-(n+a_\sigma)} \\ &= c^* \left[b_\sigma^{-1} + np\bar{y} - \mathcal{Y}_{(np)} \right]^{-(n+a_\sigma)}. \end{aligned}$$

Likewise,

$$\sum_{i=1}^n (y_i - y_{(np+1)}) \mathbb{1}(y_i < y_{(np+1)}) = \left(\sum_{i=1}^{np} y_{(i)} \right) - (np)y_{(np+1)} = \mathcal{Y}_{(np)} - (np)y_{(np+1)}$$

so that

$$\begin{aligned} p(\theta_p = y_{(np+1)} | \mathbf{y}) &= c^* \left[b_\sigma^{-1} + np(\bar{y} - y_{(np+1)}) - \mathcal{Y}_{(np)} + (np)y_{(np+1)} \right]^{-(n+a_\sigma)} \\ &= c^* \left[b_\sigma^{-1} + np\bar{y} - \mathcal{Y}_{(np)} \right]^{-(n+a_\sigma)} \\ &= p(\theta_p = y_{(np)} | \mathbf{y}). \end{aligned}$$

Identical reasoning can be used to establish that the posterior distribution $p(\theta_p | \mathbf{y})$ is flat and reaches a maximum over the interval $\theta_p \in [y_{(np)}, y_{(np+1)}]$ when np is an integer.

Figure 1 offers some very brief graphical support of this result, plotting $p(\theta_p | \mathbf{y})$ when $a_\sigma = 2$, $b_\sigma = 1$, $p = 0.1$, and for three different *iid* samples

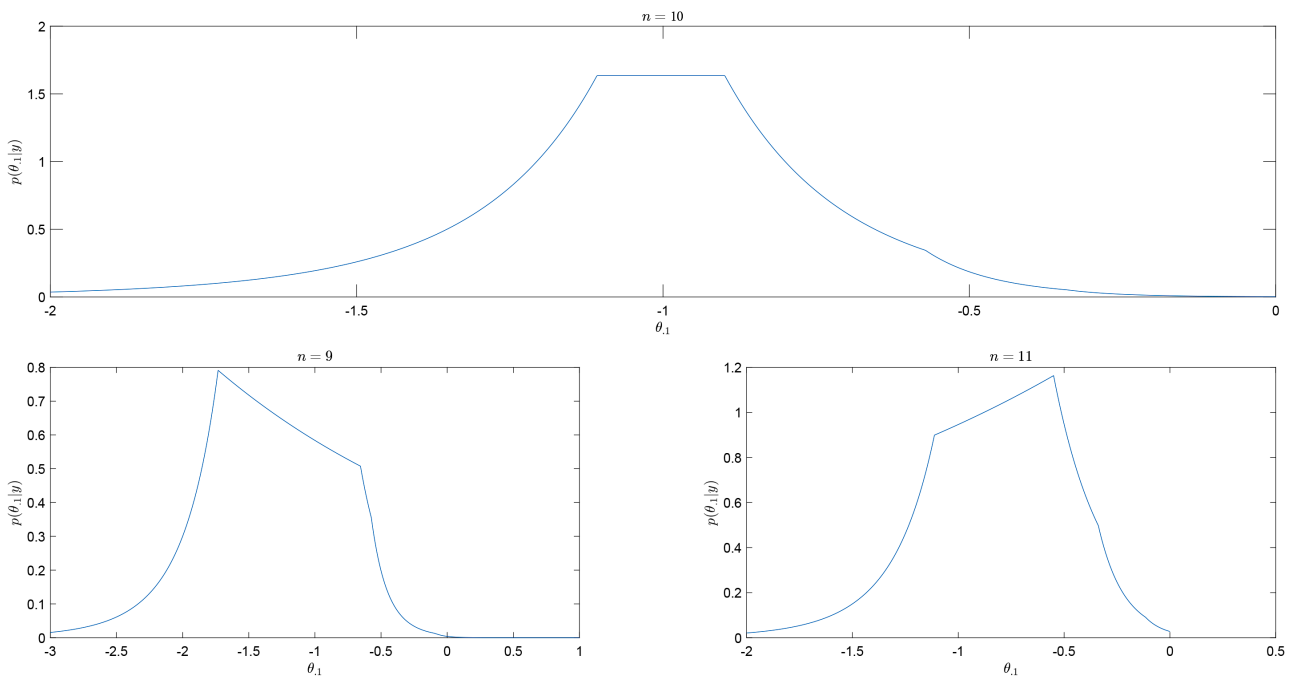


Figure 1. Plots of $p(\theta_{0.1} | \mathbf{y})$ with $y_i \sim \mathcal{N}^{iid}(0,1)$, $n = 9, 10, 11$.

from a $\mathcal{N}(0,1)$ distribution with $n = 10$ (upper panel), $n = 9$ (lower left) and $n = 11$ (lower right). For the $n = 10$ case, $np = 1$, and thus the posterior reaches a constant peak over $[y_{(1)}, y_{(2)}]$.

3. Generated Data Experiments

In this section, we perform a series of generated data experiments. We do so noting that when a posterior simulator is applied to generate draws from (6) or its generalization when including covariates, the mean is reported as a point estimate while the mode technically coincides with the desired quantile. We hope these experiments may shed light on whether or not use of the mean will lead us astray, and how far we might veer off course when doing so.

To sidestep issues surrounding the potential non-uniqueness of the mode we focus on cases where np is not an integer. We also focus on performance in small samples and consider cases where $n \in \{11, 21, 151\}$, fixing $p = 0.1$ throughout. For each sample size, we consider 5 different designs: (a) $y_i \sim \mathcal{N}(0,1)$, (b) $y_i \sim U(0,1)$, (c) $y_i \sim \text{Exp}(0.5)$, (d) $y_i \sim t(0,1,5)$ and (e) $y_i \sim 0.7\mathcal{N}(-1,1) + 0.3\mathcal{N}(1.5,0.75^2)$, corresponding to standard normal, uniform, exponential (with mean 1/2), standardized Student- t (with 5 degrees of freedom) and two-component normal mixture cases, respectively. Within each sample size/design cell, we generate 500,000 data sets, each time calculating the posterior mode and using (6) to calculate the posterior mean.⁵ To assess how well the means and modes perform, we calculate each estimator's average percentage error, defined as

$$\frac{1}{M} \sum_{i=1}^M \frac{\hat{\theta}_{0.1}^{(m)} - F^{-1}(0.1)}{|F^{-1}(0.1)|}$$

and also calculate the sampling probabilities that the posterior mean and mode lie within a given threshold of the quantile:

$$\frac{1}{M} \sum_{i=1}^M \mathbb{1}(\hat{\theta}_{0.1}^{(m)} \in [F^{-1}(0.1) - \sigma/\sqrt{n}, F^{-1}(0.1) + \sigma/\sqrt{n}]),$$

where $\hat{\theta}_{0.1}^{(m)}$ is either the posterior mean or posterior mode from the m^{th} experiment, $\sigma \equiv \sqrt{\text{Var}(y_i)}$ whose value changes in each of our 5 designs, $M = 500000$ and $F^{-1}(0.1)$ is the true 0.1 quantile for the given design.

Several interesting patterns emerge from **Table 1**. First and not surprisingly, estimates become more accurate (in the sense that average percentage errors tend to zero) as the sample size increases. For example, the average percentage error of the posterior mode in the Gaussian sampling experiment falls from 0.17 to 0.09 to 0.01 for $n = 11, 21$ and 151 (respectively). Second, posterior mean estimates tend to understate the quantile, as suggested by the signs of the average percentage errors, while posterior modes consistently overstate the quantile. Finally, and perhaps surprisingly, posterior means tend to have smaller average

⁵In these experiments, we set $a_\sigma = 2$ and $b_\sigma = 1$. Similar results emerged, however, when fixing $\sigma = 1$ and instead evaluating the posterior in (4) at this restriction.

Table 1. Performances of posterior mean and posterior mode across different sampling experiments.

Design	Posterior Mean					
	Avg. Percent Error			Pr $\hat{\theta}_{0.1} \in [F^{-1}(0.1) \pm \sigma/\sqrt{n}]$		
	$n=11$	$n=21$	$n=151$	$n=11$	$n=21$	$n=151$
$N(0,1)$	-0.0288	-0.0106	-0.0009	0.4995	0.4867	0.4586
$U(0,1)$	-0.0921	0.1260	0.0267	0.6919	0.7611	0.6967
$Exp(0.5)$	-0.6011	0.0588	0.0349	0.9903	0.9882	0.9948
$t(0,1,5)$	-0.0970	-0.0460	-0.0058	0.4674	0.4573	0.4349
$0.7\mathcal{N}(-1,1)+0.3\mathcal{N}(1.5,0.75^2)$	-0.0137	-0.0040	-0.0006	0.6278	0.6141	0.5848
Design	Posterior Mode					
	Avg. Percent Error			Pr $\hat{\theta}_{0.1} \in [F^{-1}(0.1) \pm \sigma/\sqrt{n}]$		
	$n=11$	$n=21$	$n=151$	$n=11$	$n=21$	$n=151$
$N(0,1)$	0.1714	0.0940	0.0138	0.4313	0.4360	0.4396
$U(0,1)$	0.6635	0.3626	0.0524	0.6303	0.6472	0.6623
$Exp(0.5)$	0.8081	0.4234	0.0585	0.9251	0.9564	0.9900
$t(0,1,5)$	0.1626	0.0882	0.0128	0.4173	0.4179	0.4183
$0.7\mathcal{N}(-1,1)+0.3\mathcal{N}(1.5,0.75^2)$	0.1269	0.0696	0.0098	0.5496	0.5582	0.5623

percentage errors than modes, despite their lack of formal connection to the quantile. This preference also exists under our second metric which calculates the (sampling) percentages that the posterior mean and posterior mode will fall within σ/\sqrt{n} of the true quantile. For this second metric, improvement is always seen for the posterior mode as the sample size grows, while the performance of the posterior mean sometimes degrades as the sample size increases.

Figure 2 provides some additional information about these experiments for the uniform and exponential cases as selected examples, each for $n=11$. The figure reveals: (a) the (sampling) average of the posterior distribution of $\theta_{p=0.1}$ places considerable mass over negative values, even though the quantiles must clearly be positive in both the uniform and exponential examples. This underscores the use of the asymmetric Laplace distribution as only an approximate or working likelihood, to be used for the narrow purpose of recovering the quantile,⁶ and (b) the posterior mean often yields a negative point estimate of the quantile which, for these sampling models, we know cannot be the case. This never occurs when using the posterior mode, of course, which equals $y_{(2)}$ in these examples.

⁶See, e.g., Yang et al. (Yang et al., 2016) for additional discussion.

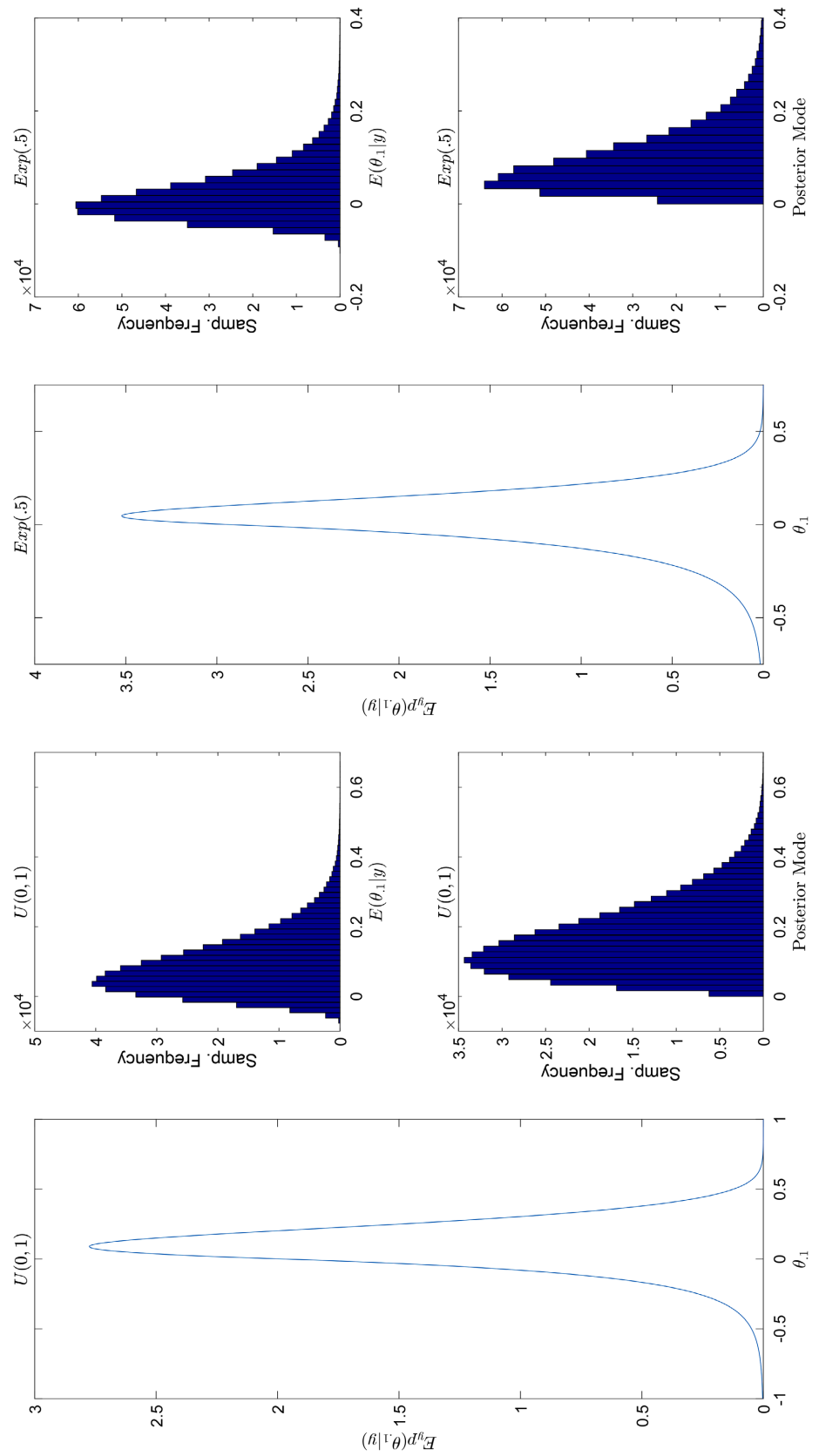


Figure 2. Results from uniform and exponential sampling experiments, $n = 11$.

The overall preference for the mean relative to the mode, according to our two metrics considered, stems from the fact that the posterior mode's sampling distribution places more mass over relatively large positive deviations from the true quantile. This preference, however, remains sensitive to the metric employed. The mode of the sampling distribution of the posterior mode, for example, is much closer to the true quantile than the comparable mode of the posterior mean: Those modes for the {posterior mean, posterior mode (and actual quantile)} are $\{-0.106, 0.046, (0.053)\}$ and $\{-0.08, 0.12, (0.1)\}$ for the exponential and uniform cases, respectively. Furthermore, when samples are very small, the posterior mean sometimes provides a point estimate that we are certain is wrong, as it lies outside of the support of the random variable. In terms of our two commonly-used measures of average performance, the posterior mean is certainly comparable to, if not preferred over, the posterior mode, despite the latter's formal connection with the desired quantile.

4. Conclusion

This note investigated a few issues surrounding Bayesian quantile regression. First, under commonly-used priors, we analytically derived the exact posterior distribution of the location parameter in an asymmetric Laplace likelihood, a likelihood that is often used for Bayesian quantile regression. We documented the non-uniqueness of the mode of that distribution in certain cases. Second, we performed a series of generated data experiments to investigate how well the posterior mean performs in recovering the quantile, even when the quantile itself is coincident with the mode rather than the mean. Although the posterior mean can sometimes produce estimates that are known to be wrong, in terms of certain measures of average performance, the posterior mean compares similarly to—and perhaps favorably to—the posterior mode. Finally, additional work can be conducted to investigate these issues in cases beyond the univariate example considered here, where the quantile functions are also covariate-dependent.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Alhamzawi, R. (2016). Bayesian Model Selection in Ordinal Quantile Regression. *Computational Statistics and Data Analysis*, 103, 68-78.
<https://doi.org/10.1016/j.csda.2016.04.014>
- Alhamzawi, R., & Ali, H. T. M. (2018). Bayesian Quantile Regression for Ordinal Longitudinal Data. *Journal of Applied Statistics*, 45, 815-828.
<https://doi.org/10.1080/02664763.2017.1315059>
- Bresson, G., Lacroix, G., & Rahman, M. A. (2021). Bayesian Panel Quantile Regression for Binary Outcomes with Correlated Random Effects: An Application on Crime Recidivism in Canada. *Empirical Economics*, 60, 227-259.

<https://doi.org/10.1007/s00181-020-01893-5>

Chan, J., Koop, G., Poirier, D. J., & Tobias, J. L. (2019). *Bayesian Econometric Methods* (2nd ed., 486 p.). Cambridge University Press.

<https://doi.org/10.1017/9781108525947>

Kedia, P., Kundu, D., & Das, K. (2023). A Bayesian Variable Selection Approach to Longitudinal Quantile Regression. *Statistical Methods & Applications*, 32, 149-168.

<https://doi.org/10.1007/s10260-022-00645-2>

Khare, K., & Hobert, J. P. (2012). Geometric Ergodicity of the Gibbs Sampler for Bayesian Quantile Regression. *Journal of Multivariate Analysis*, 112, 108-116.

<https://doi.org/10.1016/j.jmva.2012.05.004>

Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46, 33-50.

Kozumi, H., & Kobayashi, G. (2011). Gibbs Sampling Methods for Bayesian Quantile Regression. *Journal of Statistical Computation and Simulation*, 81, 1565-1578.

<https://doi.org/10.1080/00949655.2010.496117>

Mostafaei, S., Kabir, K., Kazemnejad, A., Feizi, A., Mansourian, M., Keshteli, A. H., Afshar, H., Arzaghi, S. M., Dehkordi, S. R., Adibi, P., & Ghadirian, F. (2019). Explanation of Somatic Symptoms by Mental Health and Personality Traits: Application of Bayesian Regularized Quantile Regression in a Large Population Study. *BMC Psychiatry*, 19, 1-8. <https://doi.org/10.1186/s12888-019-2189-1>

Rahman, M. A. (2016). Bayesian Quantile Regression for Ordinal Models. *Bayesian Analysis*, 11, 1-24. <https://doi.org/10.1214/15-BA939>

Xu, X., & Chen, L. (2019). Influencing Factors of Disability among the Elderly in China, 2003-2016: Application of Bayesian Quantile Regression. *Journal of Medical Economics*, 22, 605-611. <https://doi.org/10.1080/13696998.2019.1600525>

Yang, Y., Wang, H. J., & He, X. (2016). Posterior Inference in Bayesian Quantile Regression with Asymmetric Laplace Likelihood. *International Statistical Review*, 84, 327-344.

<https://doi.org/10.1111/insr.12114>

Yu, K., & Moyeed, R. A. (2001). Bayesian Quantile Regression. *Statistics & Probability Letters*, 54, 437-447. [https://doi.org/10.1016/S0167-7152\(01\)00124-9](https://doi.org/10.1016/S0167-7152(01)00124-9)

Zhao, Y., & Xu, D. (2023). A Bayesian Variable Selection Method for Spatial Autoregressive Quantile Models. *Mathematics*, 11, 1-19. <https://doi.org/10.3390/math11040987>