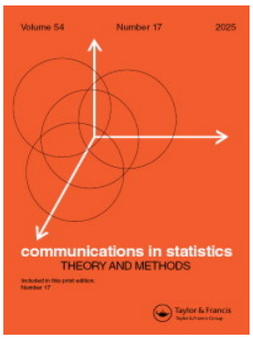

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: www.tandfonline.com/journals/lsta20

Posterior rates of convergence for composite quantile regression

Lukas Arnroth & Shaobo Jin

To cite this article: Lukas Arnroth & Shaobo Jin (2025) Posterior rates of convergence for composite quantile regression, Communications in Statistics - Theory and Methods, 54:17, 5460-5469, DOI: 10.1080/03610926.2024.2438312

To link to this article: https://doi.org/10.1080/03610926.2024.2438312

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2025, VOL. 54, NO. 17, 5460–5469
https://doi.org/10.1080/03610926.2024.2438312

Taylor & Francis
Taylor & Francis Group

Posterior rates of convergence for composite quantile regression

Lukas Arnroth[a] and Shaobo Jin[b]

[a]Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; [b]Department of Mathematics, Uppsala University, Uppsala, Sweden

ABSTRACT

Composite quantile regression is based on the convex combination of single quantile quantile loss functions and enjoys many advantages over single quantile regression. The Bayesian extension is based on the finite mixture of asymmetric Laplace densities. This article mainly aims to contribute to the theoretical justification of Bayesian composite quantile regression from the perspective of Bayesian density estimation. As such, we further show that the asymmetric Laplace distribution can be used for Bayesian density estimation in general. We obtain upper bounds on rates of convergence for mixtures of asymmetric Laplace densities. For finite mixtures, we obtain the parametric rate up to a logarithmic factor, and a slower rate for infinite mixture.

ARTICLE HISTORY
3 January 2024
30 November 2024

KEYWORDS
Convergence rates; composite quantile regression; Dirichlet process

1. Introduction

As an alternative to linear regression that models the conditional mean of a response variable given a vector of covariates, quantile regression (Koenker 2005) is a popular method to model the conditional quantile. Let $0 < \tau < 1$ be the quantile of interest and $\rho_\tau(x) = \big|\tau - 1(x \leq 0)\big||x|$ be the check-loss function. The quantile regression solves

$$\min_{\theta,\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau(y_i - \theta - \boldsymbol{x}_i^T\boldsymbol{\beta}), \tag{1}$$

where y_i is the response variable and $\boldsymbol{x}_i$ a is $p \times 1$ covariate vector. As a special case of quantile regression, the median regression ($\tau = 0.5$) is expected to be more robust against outliers than the least squares (LS) estimator in linear regression. However, the relative efficiency of the quantile regression estimator with respect to the LS estimator in linear regression can be arbitrarily small (Zou and Yuan 2008).

The composite quatile regression (CQR) addresses the limitation of the quantile regression by combining multiple models of different quantiles of the dependent variable as a function of covariates (Koenker 1984; Zou and Yuan 2008; Zhao and Xiao 2014). To be more specific, CQR proposed solves the convex combination of K quantile regression problems

CONTACT Lukas Arnroth lukas.arnroth@igp.uu.se Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden.

$$\min_{\theta_1,\ldots,\theta_K,\boldsymbol{\beta}} \sum_{j=1}^{K} w_j \left\{ \sum_{i=1}^{n} \rho_{\tau_j}(y_i - \theta_j - \boldsymbol{x}_i^T \boldsymbol{\beta}) \right\}, \tag{2}$$

where $0 < \tau_1 < \ldots < \tau_K < 1$ are the quantiles and $\sum_{j=1}^{K} w_j = 1$ with $w_j \in [0, 1]$. It is worth mentioning that the regression slopes are the same across all quantiles and the conditional quantiles differ mainly in their intercepts θ_j. The interpretation of $\boldsymbol{\beta}$ in CQR is similar to the interpretation of the LS estimator while also providing a more complete characterization of the conditional distribution of the dependent variable. Zou and Yuan (2008) showed that the relative efficiency of the CQR estimator relative to the LS estimator is at least 70% if we set $w_j = 1/K$ for all j, since the strengths of multiple quantile regression models are combined. The lower bound of the relative efficiency prevents uncontrollable loss of efficiency as in the quantile regression. See also Zhao and Lian (2016) for further discussions on the efficiency of CQR. Lately, CQR has gained much popularity. Examples include extensions to neural networks (Xu et al. 2017; Cannon 2018; Moon et al. 2021), local polynomial regression (Kai et al. 2010; R. Jiang et al. 2012), extreme value analysis (Wang, Li, and He 2012), and big data (Jiang et al. 2018).

In the frequentist setting, asymptotic properties of CQR have been carefully considered in various studies. For example, Zou and Yuan (2008) established asymptotic normality and relative efficiency of the CQR estimator if $w_j = 1/K$ for all j. Jiang, Jiang, and Song (2012) established asymptotic normality of the CQR estimator for non equal weights. Zhao and Lian (2016) established asymptotic normality after relaxing the assumption $w_j \geq 0$.

Despite CQR was initially proposed from a frequentist perspective, it can also be developed from a Bayesian perspective. Huang and Chen (2015) pointed out that the minimization problem (2) can be formulated as maximum likelihood estimation by assuming that the response y_i is a mixture of asymmetric Laplace distributions (ALD), i.e., y_i has density

$$\sum_{j=1}^{K} w_j \psi_{\sigma,\tau_j}(y_i - \theta_j - \boldsymbol{x}_i^T \boldsymbol{\beta}), \tag{3}$$

where

$$\psi_{\sigma,\tau}(y - \mu) := \frac{\tau(1-\tau)}{\sigma} \exp\left\{ -\rho_\tau \left(\frac{y-\mu}{\sigma} \right) \right\}, \quad \text{for } y \in \mathbb{R},$$

is the ALD density with location parameter $\mu \in \mathbb{R}$, asymmetry parameter $\tau \in (0, 1)$, and scale parameter $\sigma > 0$. Hence, they proposed the Bayesian composite quantile regression (BCQR) and developed an MCMC sampling procedure to obtain the posterior distribution of τ and $\boldsymbol{\beta}$. BCQR has been extended to handle various modeling scenarios (Alhusseini and Georgescu 2018; Adlouni, Salaou, and St-Hilaire 2018; Alsaadi and Alhamzawi 2022) and applied in various contexts, such as spatio-temporal modeling (Yu et al. 2022), linear mixed models (Tian, Lian, and Tian 2017), control law for stochastic systems (Ma et al. 2022), and extreme values (Hu et al. 2021).

However, to the best of our knowledge, no theoretical justification has been presented for CQR from the Bayesian perspective. We aim to bridge such gap in this article. Sriram, Ramamoorthi, and Ghosh (2013) proved posterior consistency of the Bayesian quantile regression estimator and derive the rate of convergence. However, they only considered one quantile, instead of combining multiple quantiles as in CQR. Our result can be viewed as an extension of their study to BCQR. One important addition to BCQR in this article is the

addition of a prior on the number of components, i.e., K in (3). As such, the extension of the procedure outlined in Huang and Chen (2015) to a data-driven selection of number of mixture components is theoretically justified. The mixture of finite mixture approach outlined in Miller and Harrison (2018) makes such an extension particularly easy and attractive for practitioners.

To facilitate our proofs, we will treat BCQR as Bayesian density estimation based on location mixtures of ALD, since minimizing (2) can be viewed as fitting the mixture (3). A benefit of treating it as a density estimation problem is that density estimation is of independent interest. Finite normal mixture is a popular model for density estimation, not only in the frequentist context but also in the Bayesian context. Evans, Guttman, and Olkin (1992) used a two-component normal mixture to model contaminated data and developed a Gibbs sampler to sample from the posterior distribution. Diebolt and Robert (1994) also used Gibbs sampling to fit a finite mixture with fixed number of components. Instead of assuming the number of components is fixed, Phillips et al. (1995) and Richardson and Green (1997) assumed a Poisson prior on K. The reader is directed to MacLahlan and Peel (2000) and Frühwirth-Schnatter (2006) for Bayesian finite mixture models and reference therein. Beyond the finite mixture, the mixtures of Dirichlet processes can be viewed as a mixture of infinite components. Ferguson (1983) proposed the framework of normal mixtures of Dirichlet processes. Escobar and West (1995) developed a corresponding sampling procedure to sample from the posterior distribution such that Dirichlet mixture became feasible in practice. For a review of various sampling methods, see Neal (2000). Since then, the mixtures of Dirichlet processes gained much popularity in various applications. For example, Da Silva (2007) used the Dirichlet process mixture to classify brain magnetic resonance images. Nguyen, Boyd-Graber, and Altschul (2013) used the Dirichlet process mixture for protein analysis. Shen, Tokdar, and Ghosal (2013) used the Dirichlet mixture to estimate multivariate densities. Ngan, Yung, and Yeh (2015) applied the Dirichlet mixture to detect outliers in traffic data.

Another benefit of treating BCQR as density estimation is that fruitful results and techniques, such as Gao and van der Vaart (2016); Ghosal and van der Vaart (2001); Ghosal and Van Der Vaart (2007); Ghosal, Ghosh, and Ramamoorthi (1999); Scricciolo (2011), in the Bayesian density estimation literature can be used. For the Dirichlet mixture of normal densities, Ghosal, Ghosh, and Ramamoorthi (1999) proved posterior consistency. Ghosal and van der Vaart (2001) derived the rate of convergence when the true density is a location or location-scale mixture of normal densities. Ghosal and Van Der Vaart (2007) derived the rate of convergence when the true density is twice continuously differentiable. Scricciolo (2011) derived the rate of convergence for finite mixture and Dirichlet mixture of exponential power densities. Dirichlet process mixtures of asymmetric Laplace distributions have been considered by Yu et al. (2019); Kobayashi et al. (2021).

The article is organized as follows. Section 2 describes the set-up and introduces some notation. Section 3 considers the case of finite mixtures, with Section 3.2 presenting the main result for BCQR. Section 4 presents the result on infinte mixtures. Section 5 concludes the article with discussion.

2. Preliminaries

Suppose that we have observations $Y_1, \ldots, Y_n$ from an unknown density function $f_{F,\sigma}$ on $\mathbb{R}$ which is a location mixture of ALD distributions,

$$f_{F,\sigma}(y) := \int \psi_{\sigma,\tau}(y-\theta)dF(\theta),$$

where F denotes the mixing distribution. Hence, $f_{F,\sigma}$ is the convolution of $\psi_{\sigma,\tau}$ and F,

$$f_{F,\sigma} = F * \psi_{\sigma,\tau}.$$

We consider the cases where F is either a continuous distribution with infinite support points or discrete with a finite number of support points. The discrete mixture distribution can be represented as $F = \sum_{j=1}^{k} w_j \delta_{\theta_j}$ where δ_{θ_j} is the Dirac's delta, denoting a point mass at θ_j. So, (3) can be expressed as

$$f_{F,\sigma,\boldsymbol{\beta}}(y) := f_{F,\sigma}(y - \boldsymbol{x}^T\boldsymbol{\beta}) = \sum_{j=1}^{k} w_j \psi_{\sigma,\tau_j}(y - \theta_j - \boldsymbol{x}^T\boldsymbol{\beta}), \tag{4}$$

where we consider $0 < \tau_1 < \ldots < \tau_k < 1$ to be fixed in the sequel. Note that in the sequel, with some abuse of notation, we refer to $f_{F,\sigma,\boldsymbol{\beta}}(y)$ as $f_{F,\sigma,\boldsymbol{\beta}}$ to simplify notation. The scale parameter σ and regression coefficient $\boldsymbol{\beta}$ are assumed to be distributed independently of F on $(0,\infty) \times \mathbb{R}^p$. Let π denote the overall prior on $\mathfrak{M}(\mathbb{R}) \times (0,\infty) \times \mathbb{R}^p$, where $\mathfrak{M}(\mathbb{R})$ is the set of all probability measures on $\mathbb{R}$. Then, π induces a prior on $\mathscr{F} := \{f_{F,\sigma,\beta} : (F,\sigma,\boldsymbol{\beta}) \in \mathfrak{M}(\mathbb{R}) \times (0,\infty) \times \mathbb{R}^p\}$, via the mapping $(F,\sigma,\boldsymbol{\beta}) \mapsto f_{F,\sigma}(\cdot - \boldsymbol{x}^T\boldsymbol{\beta})$. We assume that the space $\mathscr{F}$ is equipped with a metric d, either the Hellinger $d_H(f,g) = \left(\int (f^{1/2} - g^{1/2})^2 d\lambda\right)^{1/2}$, where λ is the Lebesgue measure on $\mathbb{R}$, or the one induced by the L_1-norm, $||f-g||_1 := \int |f - g| d\lambda$. We want to assess the rate of convergence for the posterior distribution corresponding to a prior π under the assumption that $Y_1, \ldots, Y_n$ are i.i.d observations from $f_0 = F_{0,\sigma_0,\beta_0}$. Then, a sequence $\varepsilon_n \to 0$ as $n \to \infty$ is defined to be an upper bound on the posterior rate of convergence relative to a metric d on $\mathscr{F}$ if, for some constant $M > 0$, the posterior probability

$$\Pi\{f_{F,\sigma,\beta} : d(f_{F,\sigma,\beta}, f_{F_0,\sigma_0,\beta_0}) > M\varepsilon_n | (Y_1,\boldsymbol{x}_1), \ldots, (Y_n,\boldsymbol{x}_n)\} \to 0 \text{ as } n \to \infty,$$

P_0^∞-almost surely or in P_0^n-probability, where P_0 denotes the probability measure corresponding to f_0.

Proofs for deriving the rates of convergence for ALD mixtures involves the packing number. Let $\mathscr{F}'$ be a class of densities on $\mathbb{R}$ equipped with (semi-)metric d, the ε-packing number $D(\varepsilon, \mathscr{F}', d)$ is the maximum number of points in $\mathscr{F}'$ such that the distance between each pair is ε. The ε-covering number $N(\varepsilon, \mathscr{F}', d)$ is defined to be the minimum number of balls of radius ε needed to cover $\mathscr{F}'$. The packing and covering numbers are related by the inequalities

$$N(\varepsilon, \mathscr{F}', d) \le D(\varepsilon, \mathscr{F}', d) \le N(\varepsilon/2, \mathscr{F}', d). \tag{5}$$

Furthermore, for distances $||\cdot||_1$ and d_H, we have

$$N(\varepsilon, \mathscr{F}', d_H) \le N(\varepsilon^2, \mathscr{F}', ||\cdot||_1). \tag{6}$$

See, e.g., Appendix C of Ghosal and van der Vaart (2016) for more details.

Note that the main theorems of this article closely follow those of Scricciolo (2011), both in terms of proofs and assumptions. However, we acknowledge that adapting the proofs to incorporate an asymmetry parameter is an extensive process. One obvious contributing factor is that the behavior of an asymmetric density function requires proofs to account for both sides of the mode separately. However, one particular issue, which leads to extensive

and intricate calculations, is that there is no counterpart to Lemma A.4 in Scricciolo (2011) for the asymmetry parameter of the ALD. Consequently, the triangle inequality, which generally leads to significantly easier computations, cannot be applied with respect to the asymmetry parameters. Thus, proofs involving, for example, the covering numbers, represent an important contribution of this paper.

3. Finite mixtures

3.1. Fixed asymmetry

In this section, we consider finite mixtures where all the components have the same value of the asymmetry parameter, which is assumed to be known. The density takes the form

$$f_{F,\sigma,\beta} = \sum_{j=1}^{k} w_j \psi_{\sigma,\tau}(\cdot - \theta_j - \boldsymbol{x}^T\boldsymbol{\beta}), \tag{7}$$

where $F = \sum_{j=1}^{k} w_j \delta_{\theta_j}$ is a discrete measure with finite atoms where δ_{θ_j} denotes a point mass at θ_j. The number of mixture components is a random variable K with pmf $\rho(\cdot)$. Given $K = k$, the vector $\boldsymbol{w}_k = (w_1, \ldots, w_k)^T$ of mixing weights has conditional distribution π_k on the $(k-1)$-dimensional simplex $\triangle_k := \{\boldsymbol{w}_k \in [0,1]^k : \sum_{j=1}^{k} w_j = 1\}$. Note that if we remove the part of the location that is not in the mixture distribution in (7), i.e., $\boldsymbol{x}^T\boldsymbol{\beta}$, we would have the same setting as in Section 3.1 of Scricciolo (2011) with the ALD replacing the EPD. Subsequently, the hierarchical model can be described as

- $K \sim \rho$, $\sigma \sim G$ and $\beta_j \sim W$, $j = 1, \ldots, p$, independently;
- Given $(k, \sigma, \boldsymbol{\beta})$ the r.v.'s $\theta_1, \ldots, \theta_k \overset{iid}{\sim} \Pi$ and $\boldsymbol{w}_k \sim \pi_k$ independently;
- Given $(k, \sigma, \boldsymbol{\beta}, \boldsymbol{\theta}_k, \boldsymbol{w}_k)$, the r.v.'s $Y_1, \ldots, Y_n$ are conditionally independent and identically distributed with density $f_{F,\sigma,\beta}$.

We make the following assumptions on the prior distributions of σ and $\boldsymbol{\beta}$.

(A) The prior G for σ has a continuous and positive Lebesgue density g on an interval containing σ_0 and, for constants $d > 0, \gamma, \varpi \in (0, \infty]$, satisfies

$$G(s) \lesssim e^{-ds^{-\gamma}} \text{ as } s \to 0 \quad \text{and} \quad 1 - G(s) \lesssim s^{-\varpi} \text{ as } s \to \infty,$$

where $a \lesssim b$ denotes $a \leq Lb$ for some constant L. The inverse Gamma distribution satisfies the above conditions.

(B) The common prior W for $\boldsymbol{\beta}$ is compactly supported on $[-B, B]^p$ for some constant $B > 0$, i.e., $W([-B, B]^p) = 1$.

The rate of convergence for finite mixtures with a known asymmetry parameter is stated in the next theorem.

Theorem 1. *Suppose that $f_0 = F_0 * \psi_{\sigma_0,\tau}(\cdot - \boldsymbol{x}^T\boldsymbol{\beta}_0)$ is a finite mixture of ALD with k_0 components where τ is a known constant on the compact set $[\underline{\tau}, \overline{\tau}] \subset (0,1)$. The covariates are assumed to satisfy $\max_{j\in\{1,\ldots,p\}} |x_j| \leq C_x$. Assume that*

(i) the priors G and W satisfy (A) and (B), respectively;

(ii) *the prior ρ for the number of components is such that, for constants $b, C_\rho > 0$, $0 < \rho(k) < C_\rho e^{-bk}$ for all $k \in \mathbb{N}$;*

(iii) *for each $k \in \mathbb{N}$, the prior π_k for the mixing weights is a Dirichlet distribution $Dir(\alpha_1, \ldots, \alpha_k)$ where $\alpha_1, \ldots, \alpha_k$ are such that, for constants $a, A > 0$, $D > 1$ and for $0 < \varepsilon < 1/(Dk)$,*

$$A\varepsilon^a \leq \alpha_j \leq D, \quad j = 1, \ldots, k; \tag{8}$$

(iv) *The common prior Π for the quantile specific intercepts has a continuous and positive density on a interval containing the support of F_0 and, for constants $c, \vartheta > 0$, satisfies*

$$\Pi(\{\theta : |\theta| > t\}) \lesssim e^{-ct^\vartheta}, \quad \text{for large } t > 0. \tag{9}$$

Then, the posterior rate of convergence in terms of d_H is $\varepsilon_n = n^{-1/2} \log n$.

Proof is given in Supplemental, Section 2.1.

3.2. Structured asymmetry

In this section, we consider BCQR specifically, which is defined as an extension of the finite mixture considered in the previous section in terms of the asymmetry parameter. The asymmetry parameter is now considered a vector with a unique element for each mixture component, misspecified, in a finite sample, relative to the assumed true density. We thus consider the mixture

$$f_{F,\sigma,\beta,\tau}(y) = \sum_{j=1}^{k} w_j \psi_{\sigma,\tau_k}(y - \theta_j - \boldsymbol{x}^T\boldsymbol{\beta}).$$

The hierarchical model can be described as

- $K \sim \rho, \sigma \sim G$ and $\beta_j \sim W, j = 1, \ldots, p$, independently;
- Given $K = k$, the asymmetry parameters $\tau_1 < \ldots < \tau_k$ are set as equidistant points on the compact interval $[\underline{\tau}, \overline{\tau}]$ where $0 < \underline{\tau} < \overline{\tau} < 1$ are known constants;
- Given $(k, \sigma, \boldsymbol{\beta})$ the r.v.'s $\theta_1, \ldots, \theta_k \overset{iid}{\sim} \Pi$ and $\boldsymbol{w}_k \sim \pi_k$ independently;
- Given $(k, \sigma, \boldsymbol{\beta}, \boldsymbol{\theta}_k, \boldsymbol{w}_k)$, the r.v.'s $Y_1, \ldots, Y_n$ are conditionally independent and identically distributed with density $f_{F,\sigma,\beta}$.

Theorem 2. *Suppose that $f_0(y) = F_0 * \psi_{\sigma_0,\boldsymbol{\tau}_0}(y - \boldsymbol{x}^T\boldsymbol{\beta}_0)$ is a finite mixture of ALD with k_0 components where τ_0 is a k_0-dimensional vector with equidistant elements on $[\underline{\tau}, \overline{\tau}]$. The covariates are assumed to satisfy $\max_{j\in\{1,\ldots,p\}} |x_j| \leq C_x$. Assume that*

(i) *the priors G and W satisfy (A) and (B), respectively;*

(ii) *the prior ρ for the number of components is such that, for constants $b, C_\rho > 0$, $0 < \rho(k) < C_\rho e^{-bk}$ for all $k \in \mathbb{N}$;*

(iii) *for each $k \in \mathbb{N}$, the prior π_k for the mixing weights is a Dirichlet distribution $Dir(\alpha_1, \ldots, \alpha_k)$ where $\alpha_1, \ldots, \alpha_k$ are such that, for constants $a, A > 0$, $D > 1$ and for $0 < \varepsilon < 1/(Dk)$,*

$$A\varepsilon^a \leq \alpha_j \leq D, \quad j = 1, \ldots, k;$$

(iv) *The common prior Π for the quantile specific intercepts has a continuous and positive density, on an interval containing the support of F_0 and, for constants $c, \vartheta > 0$, satisfies*

$$\Pi(\{\theta : |\theta| > t\}) \lesssim e^{-ct^\vartheta}, \quad \text{for large } t > 0.$$

Then, the posterior rate of convergence in terms of d_H is $\varepsilon_n = n^{-1/2}\log n$.

Proof is given in Supplemental, Section 2.2.

Remark. *We make stringent assumptions on the vector of asymmetry parameters of f_0 to simplify the proof. These can be modified by upper bounding $\left|\left|f_{F,\sigma,\beta,\tau} - f_{F,\sigma,\beta,\tau'}\right|\right|_1 \lesssim \left|\left|\tau - \tau'\right|\right|_{\ell_1}$, where the inequality be shown via the mean value theorem.*

4. Infinite mixtures

We now consider the case where the underlying factor may be infinitely many, i.e., the mixing distribution F can be any distribution on $\mathbb{R}$. As a prior for F, we use a Dirichlet process $\mathfrak{D}_\alpha$ with base measure α. We treat τ as scalar, known and defined on a compact set $[\underline{\tau}, \overline{\tau}] \subset (0,1)$. A hierarchical model can be described as follows

- $F \sim \mathfrak{D}_\alpha$, $\sigma \sim G$ and $\beta_j \sim W$, $j = 1, \ldots, p$, independently;
- Given $(F, \sigma, \boldsymbol{\beta})$, the r.v.'s $\theta_1, \ldots, \theta_n \sim F$;
- Given $(F, \sigma, \boldsymbol{\beta}, \theta_1, \ldots, \theta_n)$, the r.v. $e_i \sim \psi_{\sigma,\tau}(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})$;
- The observations are defined as $Y_i := \theta_i + e_i$ for $i = 1, \ldots, n$.

Theorem 3. *Let $f_0(y) = F_0 * \psi_{\sigma_0,\tau}(y - \boldsymbol{x}^T\boldsymbol{\beta})$ with the true mixing distribution having either compact support or sub-exponential tails, i.e.,*

$$F_0(\{\theta : |\theta| > t\}) \lesssim e^{-c_0 t}, \quad \text{for large } t > 0,$$

for some constant $c_0 > 0$. Further, assume that $\sigma_0 \in [\underline{\sigma}, \overline{\sigma}] \subset (0, \infty)$ and $\tau \in [\underline{\tau}, \overline{\tau}] \subset (0,1)$. If the base measure α has a continuous and positive density α' such that for constants $d > 0$ and $0 < \delta \leq 1$ satisfies

$$\alpha'(\theta) \gtrsim e^{-d|\theta|^\delta}, \quad \text{for large } |\theta|,$$

the prior G for σ is supported on $[\underline{\sigma}, \overline{\sigma}]$ and has a continuous and positive density on an interval containing σ_0 and the prior W satisfies condition (B) where $\boldsymbol{\beta}_0 \in [-B, B]^p$, then the posterior rate of convergence relative to d_H is

$$\varepsilon_n = n^{-\frac{1}{4}}(\log n)^{\frac{5}{4}}.$$

Proof is given in Supplemental, Section 2.3.

5. Discussion

In this article, we have studied asymptotic properties of mixtures of ALD with a specific interest in the composite quantile regression case, focusing on posterior convergence rates. We find that with the addition of an asymmetry parameter and the extension to part of the location being independent of the mixture, the convergence rates conform to those for the symmetric Laplace case as outlined in Scricciolo (2011). Most importantly, the setting considered most realistic in terms of quantile regression in Theorem 2 had the same convergence rate as the mixture considered in Theorem 1. For the finite mixture cases considered, we obtain

parametric convergence rates up to a logarithmic factor, $n^{-1/2}\log n$. For the infinite mixture, we obtain convergence rate $n^{-1/4}(\log n)^{5/4}$.

For future research in the direction of Bayesian density estimation, it is of interest to extend Theorems 1 and 3 in terms of the asymmetry parameter, e.g., making it unknown and extensions to mixing. For the finite mixture case, refinements outlined in Chapter 8 of Ghosal and van der Vaart (2016) could be considered to obtain parametric rates of convergence. Last, our work and that of Scricciolo (2011) can be combined by considering mixtures of skewed EPDs of which both the EPD and ALD are special cases. In the direction of BCQR, we consider two areas to be of particular importance. First, relaxing assumptions on the asymmetry parameters. Second, and most importantly for applications, is to scale back the scope and focus on posterior inference of the regression coefficient under more realistic data generating procedures. Such work should be closer in line with that of Sriram, Ramamoorthi, and Ghosh (2013) on single quantile regression, for example, by restricting attention to a fixed, possibly small, number of mixture components.

Acknowledgments

The authors are thankful to Silvelyn Zwanzig for suggestions for the improvement of the article.

Disclosure statement

The authors report there are no competing interests to declare.

References

Adlouni, S. E., G. Salaou, and A. St-Hilaire. 2018. Regularized Bayesian quantile regression. *Communications in Statistics- Simulation and Computation* 47 (1):277–93. doi: 10.1080/03610918.2017.1280830.

Alhusseini, F. H. H., and V. Georgescu. 2018. Bayesian composite Tobit quantile regression. *Journal of Applied Statistics* 45 (4):727–39. doi: 10.1080/02664763.2017.1299697.

Alsaadi, Z., and R. Alhamzawi. 2022. Bayesian bridge and reciprocal bridge composite quantile regression. *Communications in Statistics-Simulation and Computation* 53 (8):3571–88.

Cannon, A. J. 2018. Non-crossing nonlinear regression quantiles by monotone composite quantile regression neural network, with application to rainfall extremes. *Stochastic Environmental Research and Risk Assessment* 32 (11):3207–25. doi: 10.1007/s00477-018-1573-6.

Da Silva, A. R. F. 2007. A dirichlet process mixture model for brain mri tissue classification. *Medical Image Analysis* 11 (2):169–82.

Diebolt, J., and C. P. Robert. 1994. Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 56 (2):363–75. doi: 10.1111/j.2517-6161.1994.tb01985.x.

Escobar, M. D., and M. West. 1995. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 90 (430):577–88. doi: 10.1080/01621459.1995.10476550.

Evans, M., I. Guttman, and I. Olkin. 1992. Numerical aspects in estimating the parameters of a mixture of normal distributions. *Journal of Computational and Graphical Statistics* 1 (4):351–65. doi: 10.1080/10618600.1992.10474590.

Ferguson, T. S. 1983. Bayesian density estimation by mixtures of normal distributions. In *Recent advances in statistics*, ed. by H. Rizvi and J. Rustagi, 287–302. New York, NY: Academic Press.

Frühwirth-Schnatter, S. 2006. *Finite mixture and Markov switching models*. New York, NY: Springer.

Gao, F., and A. Van Der Vaart. 2016. Posterior contraction rates for deconvolution of dirichlet-laplace mixtures. *Electronic Journal of Statistics* 10 (1):608–27.
Ghosal, S., and A. Van Der Vaart. 2001. Entropies and rates of convergence for maximum likelihood and bayes estimation for mixtures of normal densities. *The Annals of Statistics* 29 (5):1233–63.
Ghosal, S., and A. W. Van Der Vaart. 2007. Posterior convergence rates of dirichlet mixtures at smooth densities. *The Annals of Statistics* 35:697–723.
Ghosal, S., and A. Van Der Vaart. 2016. *Fundamentals of nonparametric Bayesian inference*. Cambridge, UK: Cambridge University Press.
Ghosal, S., J. K. Ghosh, and R. V. Ramamoorthi. 1999. Posterior consistency of dirichlet mixtures in density estimation. *The Annals of Statistics* 27 (1):143–58.
Hu, Y., H. J. Wang, X. He, and J. Guo. 2021. Bayesian joint-quantile regression. *Computational Statistics* 36 (3):2033–53. doi: 10.1007/s00180-020-00998-w.
Huang, H., and Z. Chen. 2015. Bayesian composite quantile regression. *Journal of Statistical Computation and Simulation* 85 (18):3744–54. doi: 10.1080/00949655.2015.1014372.
Jiang, R., Z.-G. Zhou, W.-M. Qian, and W.-Q. Shao. 2012. Single-index composite quantile regression. *Journal of the Korean Statistical Society* 41 (3):323–32. doi: 10.1016/j.jkss.2011.11.001.
Jiang, R., X. Hu, K. Yu, and W. Qian. 2018. Composite quantile regression for massive datasets. *Statistics* 52 (5):980–1004. doi: 10.1080/02331888.2018.1500579.
Jiang, X., J. Jiang, and X. Song. 2012. Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statistica Sinica* 22 (4):1479–506.
Kai, B., R. Li, and H. Zou. 2010. Local CQR smoothing: An efficient and safe alternative to local polynomial regression. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 72 (1):49–69. doi: 10.1111/j.1467-9868.2009.00725.x. 20975930
Kobayashi, G., T. Roh, J. Lee, and T. Choi. 2021. Flexible Bayesian quantile curve fitting with shape restrictions under the Dirichlet process mixture of the generalized asymmetric Laplace distribution. *Canadian Journal of Statistics* 49 (3):698–730. doi: 10.1002/cjs.11582.
Koenker, R. 1984. A note on L-estimates for linear models. *Statistics & Probability Letters* 2 (6):323–5. doi: 10.1016/0167-7152(84)90040-3.
Koenker, R., 2005. *Quantile regression*. New York, NY: Cambridge University Press.
Ma, X., F. Qian, S. Zhang, and Li. Wu. 2022. Adaptive quantile control for stochastic system. *ISA Transactions* 123:110–21. doi: 10.1016/j.isatra.2021.05.032. 34090667
Maclahlan, G., and D. Peel. 2000. *Finite mixture models*. New York, NY: John Wiley & Sons.
Miller, J. W., and M. T. Harrison. 2018. Mixture models with a prior on the number of components. *Journal of the American Statistical Association* 113 (521):340–56. doi: 10.1080/01621459.2016.1255636. 29983475
Moon, S. J., J.-J. Jeon, J. S. H. Lee, and Y. Kim. 2021. Learning multiple quantiles with neural networks. *Journal of Computational and Graphical Statistics* 30 (4):1238–48. doi: 10.1080/10618600.2021.1909601.
Neal, R. M. 2000. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* 9 (2):249–65. doi: 10.1080/10618600.2000.10474879.
Ngan, H. Y., N. H. Yung, and A. G. Yeh. 2015. Outlier detection in traffic data based on the Dirichlet process mixture model. *IET Intelligent Transport Systems* 9 (7):773–81. doi: 10.1049/iet-its.2014.0063.
Nguyen, V.-A., J. Boyd-Graber, and S. F. Altschul. 2013. Dirichlet mixtures, the Dirichlet process, and the structure of protein space. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 20 (1):1–18. doi: 10.1089/cmb.2012.0244. 23294268
Phillips, D. B., and A. F. Smith. 1995. Bayesian model comparison via jump diffusions. In *Markov chain Monte Carlo in practice*, ed. by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, 215–39. London, UK: Chapman & Hill.
Richardson, S., and P. J. Green. 1997. On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society Series B: Statistical Methodology* 59 (4):731–92. doi: 10.1111/1467-9868.00095.
Scricciolo, C. 2011. Posterior rates of convergence for dirichlet mixtures of exponential power densities. *Electronic Journal of Statistics* 5:270–308.

Shen, W., S. T. Tokdar, and S. Ghosal. 2013. Adaptive Bayesian multivariate density estimation with Dirichlet mixtures. *Biometrika* 100 (3):623–40. doi: 10.1093/biomet/ast015.

Sriram, K., R. V. Ramamoorthi, and P. Ghosh. 2013. Posterior consistency of bayesian quantile regression based on the misspecified asymmetric laplace density. *Bayesian Analysis* 8 (2):479–504.

Tian, Y., H. Lian, and M. Tian. 2017. Bayesian composite quantile regression for linear mixed-effects models. *Communications in Statistics - Theory and Methods* 46 (15):7717–31. doi: 10.1080/03610926.2016.1161798.

Wang, H. J., D. Li, and X. He. 2012. Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association* 107 (500):1453–64. doi: 10.1080/01621459.2012.716382.

Xu, Q., K. Deng, C. Jiang, F. Sun, and X. Huang. 2017. Composite quantile regression neural network with applications. *Expert Systems with Applications* 76:129–39. doi: 10.1016/j.eswa.2017.01.054.

Yu, Y., Z. Zou, S. Wang, and R. Meyer. 2019. Bayesian nonparametric modelling of the link function in the single-index model using a Bernstein–Dirichlet process prior. *Journal of Statistical Computation and Simulation* 89 (17):3290–312. doi: 10.1080/00949655.2019.1663191.

Yu, Z., K. Yu, W. K. Härdle, X. Zhang, K. Wang, and M. Tian. 2022. Bayesian spatio-temporal modeling for the inpatient hospital costs of alcohol-related disorders. *Journal of the Royal Statistical Society Series A: Statistics in Society* 185 (Supplement_2):644–67.

Zhao, K., and H. Lian. 2016. A note on the efficiency of composite quantile regression. *Journal of Statistical Computation and Simulation* 86 (7):1334–41. doi: 10.1080/00949655.2015.1062096.

Zhao, Z., and Z. Xiao. 2014. Efficient regressions via optimally combining quantile information. *Econometric Theory* 30 (6):1272–314. doi: 10.1017/S0266466614000176.

Zou, H., and M. Yuan. 2008. Composite quantile regression and the oracle model selection theory. *The Annals of Statistics* 36 (3):1108–26.