POSTERIOR CONSISTENCY OF BAYESIAN QUANTILE REGRESSION UNDER A MIS-SPECIFIED LIKELIHOOD BASED ON ASYMMETRIC LAPLACE DENSITY

By Karthik Sriram*, R. V. Ramamoorthi**, Pulak Ghosh*

*Indian Institute of Management Bangalore, **Michigan State University*

Last revised on May 30, 2011

We provide a theoretical justification for the widely used and yet only empirically verified approach of using Asymmetric Laplace Density(ALD) in Bayesian Quantile Regression. We derive sufficient conditions for posterior consistency of the quantile regression parameters even if the true underlying likelihood is not ALD, by considering both the case of random as well as non-random covariates. While existing literature on misspecified models address more general models, our approach of exploiting the specific form of ALD allows for a more direct derivation. We verify that the conditions so derived are satisfied by a wide range of potential true underlying probability distributions. We also show that posterior consistency holds even in the case of improper priors as long as the posterior is well defined. We demonstrate the working of the method using simulations.

[0]*Key words and phrases.* Consistency, Asymmetric Laplace Density, Kullback-Leibler divergence, quantile regression, posterior distribution.

1. Introduction

Quantile regression has been popular as a simple, robust and distribution free modeling methodology since the seminal work by Koenker and Basset (1978). It provides a way to model different percentiles of the distribution of the response as a function of covariates. This makes it an indispensible tool for analyzing many important practical problems. For example, it can play a crucial role in helping understand the nature of tail events, which is an important problem in the financial services industry.

If Y_i is the dependent variable and X_i are the explanatory variables or covariates, then for a fixed $\tau \in (0,1)$, the quantile regression problem is to solve the following minimization problem over β

$$Minimize_{\beta}\ \rho_{\tau}(Y_k - X_k^T\beta) \tag{1.1}$$

$$\rho_{\tau}(u) = u(\tau - I_{u \leq 0}) \tag{1.2}$$

It is easy to see that the above minimization problem can be formulated as a maximum likelihood estimation problem by assuming $Y_i \sim ALD(., \mu_i^{\tau}, \sigma, \tau)$, the Asymmetric Laplace Distribution (ALD), whose density function is given by,

$$ALD(y; \mu^{\tau}, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} exp\left[-\frac{(y-\mu^{\tau})}{\sigma}(\tau - I_{(y \leq \mu^{\tau})})\right], for\ -\infty < y < \infty \tag{1.3}$$

The parameter μ^{τ} happens to be the τ^{th} quantile of the ALD distribution and hence in quantile regression one formulates the model as $\mu_i^{\tau} = X_i^T\beta$. ALD has therefore been a powerful tool for formulating both non-bayesian and bayesian quantile regression problems. This is interesting especially given the fact the the true underlying distribution in practical problems is almost never ALD. Koenker and Machado(1999) develop goodness of fit inference processes for quantile regression. They consider the Asymmetric Laplace Density and the corresponding likelihood ratio based inference for the quantile regression parameters and show that asymptotics work even if the underlying distribution is not ALD. Yu and Moyeed(2001) introduce the idea of bayesian methods in quantile regression by casting the problem as a generalized linear model using ALD for the response. They argue based on empirical results that even if the underlying distribution is not ALD, the results would be reasonable. This paper provides a theoretical justification for this phenomenon. We look at the problem where the likelihood is specified to be ALD with covariates(as described above), while the true underlying distribution may be different. We consider the case of random as well as non-random covariates and study posterior consistency of the parameters under mis-specification.

Posterior consistency of model parameters under mis-specification has been a problem of considerable interest. Prior work on this topic includes the early work of Berk (1966). Mis-specification under the parametric set up is investigated by Bunke and Milhaud (1988) and more general set up is considered by Kleijn and Vanderwaart(2006), Shalizi (2009). While these works address more general models, in our case, the specific form of the ALD likelihood allows for a more direct derivation. Our approach utilizes some of the key ideas and thought processes from the works of Shalizi (2009), Amewou-atisso et al (2003) and Ghosh and Rammoorti(2003).

In what follows, we will first present the main theorem in section 2, followed by a discussion of examples of kind of distributions that would satisfy the conditions of our theorems in section 3. In section the result is extended to the case of improper priors. We demonstrate the working of such a method through simulations in section 5 and then conclude with a discussion in section 6.

2. Theoretical Results

Let $\{Y_i\}, i = 1, 2, ...n$ be n independent observations of a univariate response. Let $\{X_i\}, i = 1, 2, .., n$ be p-dimensional covariates, whose components can either be random i.i.d realizations or non-random. Let P_{0i} denote the true (but unknown) probability distribution of (Y_i, X_i). Let $\tau \in (0, 1)$ be fixed. Our aim is to model the τ^{th} quantile of the conditional distribution of Y_i *given* X_i, denoted by $Q_\tau(Y_i/X_i)$. We will consider the most commonly used linear formulation for quantile regression and assume that the true conditional quantile is given by $Q_\tau(Y_i/X_i) = \alpha + X_i^T\beta$.

The specified model for the response is given to be $Y_i \sim ALD(., \mu_i^\tau, \sigma = 1, \tau)$, where $\mu_i^\tau = \alpha + X_i^T\beta$ and the bayesian specification if completed by putting a prior distribution on the parameters (α, β). Since the true underlying distribution of Y_i may not be ALD, the question we address is whether the procedure of using ALD is still good enough to ensure posterior consistency for (α, β). More precisely, we ask if (α_0, β_0) is the vector of true parameters and the set $U \subseteq \Re^{1+p}$ is an open ball in the (p+1) dimensional euclidean space, such that $(\alpha_0, \beta_0) \in U$ and $\Pi(\alpha, \beta)$ is a prior on the parameter space, then is it true that the posterior probability of U^c under the specified likelihood converges to zero almost surely, i.e. whether $\Pi(U^c/Y) \rightarrow 0 \; a.s. \; [P]$? , where P is the product measure $(P_{01} \times P_{02} \times ... \times P_{0n} \times ...)$.

Let $f_{(\alpha,\beta)i}(y_i)$ denote the density functions of $ALD(., \alpha + X_i^T\beta, \sigma = 1, \tau)$ at y_i. Then the posterior probability of the set U^c under the specified likelihood and prior is given by,

$$\Pi\left(U^c/(Y_1, X_1), (Y_2, X_2), ..., (Y_n, X_n)\right) = \frac{\int_{U^c} \prod_{i=1}^n f_{(\alpha,\beta)i}(Y_i) d\Pi(\alpha, \beta)}{\int_{\Theta} \prod_{i=1}^n f_{(\alpha,\beta)i}(Y_i) d\Pi(\alpha, \beta)}$$

Often, it is beneficial to modify the previous expression by dividing both the numerator and denominator by a suitable function of Y. A natural choice for such a function would be the true underlying density (as in Shalizi(2009)). Another choice is to divide both the numerator and denominator by $p^*(Y_i)$, which is the density function that minimizes the Kullback-Liebler divergence to the true density, among the specified family of densities. Berk(1966) allows for a more general choice of such a function. In our case, we find it useful to divide both the numerator and denominator by $f_{(\alpha_0,\beta_0)i}(Y_i)$, i.e. the ALD density with the true parameter values. We will show later that this particular ALD density function indeed minimizes the Kullback-Leibler divergence with the true underlying density. Accordingly, we write

$$\Pi(U^c/Y_1, Y_2, ..., Y_n) = \frac{\int_{U^c} \prod_{i=1}^n \frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)} d\Pi(\alpha, \beta)}{\int_{\Theta} \prod_{i=1}^n \frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)} d\Pi(\alpha, \beta)} = \frac{I_{1n}}{I_{2n}} \tag{2.1}$$

Our idea will be to show that under certain conditions, for some suitable $d_0 > 0$, $e^{nd_0} I_{1n} \rightarrow 0 \; a.s.[P]$ and $e^{nd_0} I_{2n} \rightarrow \infty \; a.s.[P]$. It is easy to see that the desired posterior consistency would then follow. To keep the discussion simple, we will look at the case of a single non-random covariate X_i (i.e p=1). The same approach is easily extendable to accomodate multiple covariates that include both random and non-random components. In the following subsection, we will derive the result for the more complicated case when X_i are non-random. In the subsequent section, we will remark on the case of multiple covariates that may include random components.

By way of notation, probabilites $P(\ .\ )$ and expectations $E[\ .\ ]$, will always be with respect to the true underlying product measure. $\Pi(\alpha,\beta)$ will denote the prior distribution for the parameters (α,β) on the parameter space $\Theta \subseteq \Re^{p+1}$. Also, recall that $f_{(\alpha,\beta)i}(Y_i)$ is the density function of $ALD(.,\alpha+\beta X_i,\sigma=1,\tau)$ at Y_i as given in equation (1.3). Without loss of generality, we consider open neighbourhoods of the form $U = \{(\alpha,\beta_1,\beta_2,...,\beta_p) : |\alpha-\alpha_0| < \Delta_1, |\beta-\beta_0| < \Delta_2, ...,\ |\beta_p-\beta_{0p}| < \Delta_p\}$ around the true parameter values $(\alpha_0,\beta_{01},...,\beta_{0p})$.

2..1 Posterior consistency in the case of univariate(p=1) non-random covariates

In the discussion that follows, we will often work with the log-ratio of ALD likelihood $\Big(\ i.e.\ log\frac{f_{(\alpha,\beta)i}}{f_{(\alpha_0,\beta_0)i}}(Y_i)\Big)$. The next lemma gives some identities and inequalities involving this ratio that are used through out the paper.

Lemma 2.1. *Let $f_{(\alpha,\beta)i}(Y_i)$ denote the density functions of $ALD(.,\alpha+X_i^T\beta,\sigma=1,\tau)$ at Y_i (as in (1.3)) and let $b_i = (\alpha-\alpha_0)+(\beta-\beta_0)X_i$. Then, the following hold.*

$$(a)\ log\left(\frac{f_{(\alpha,\beta)\ i}}{f_{(\alpha_0,\beta_0)\ i}}\right)(Y_i) = \begin{cases} -b_i(1-\tau) & ,if\ Y_i \le Min(\alpha+\beta X_i,\alpha_0+\beta_0 X_i) \\ (Y_i-\alpha_0-\beta_0 X_i)-b_i(1-\tau) & ,if\ \alpha_0+\beta_0 X_i < Y_i \le \alpha+\beta X_i \\ b_i\tau-(Y_i-\alpha_0-\beta_0 X_i) & ,if\ \alpha+\beta X_i < Y_i \le \alpha_0+\beta_0 X_i \\ b_i\tau & ,if\ Y_i \ge Max(\alpha+\beta X_i,\alpha_0+\beta_0 X_i) \end{cases}$$

$$(b)\ \left|log\left(\frac{f_{(\alpha_0,\beta_0)i}}{f_{(\alpha,\beta)\ i}}\right)(Y_i)\right| \le Max(\tau,(1-\tau))(|\alpha-\alpha_0|+|\beta-\beta_0||X_i|)$$

$$(c)\ log\left(\frac{f_{(\alpha_0,\beta_0)i}}{f_{(\alpha,\beta)\ i}}\right)(Y_i) \le |Y_i-\alpha_0-\beta_0 X_i|$$

$$(d)\ If\ |X_i| <= M_1\ then,\ E\left[log\left(\frac{f_{(\alpha_0,\beta_0)i}}{f_{(\alpha,\beta)\ i}}\right)(Y_i)\right] \le Max(\tau,(1-\tau))(|\alpha-\alpha_0|+|\beta-\beta_0|M_1)$$

$$(e)\qquad If\ b_i > 0,\ then\ log\left(\frac{f_{(\alpha,\beta)\ i}}{f_{(\alpha_0,\beta_0)\ i}}\right)(Y_i) = -b_i(1-\tau)+Min(Z_i^+,b_i)$$

$$If\ b_i <= 0,\ then\ log\left(\frac{f_{(\alpha,\beta)\ i}}{f_{(\alpha_0,\beta_0)\ i}}\right)(Y_i) = b_i\tau+Min(Z_i^-,-b_i)$$

where Z_i^+ and Z_i^- are positive and negative parts of the random variable $Z_i = Y_i-\alpha_0-\beta_0 X_i$

We skip the proof of the lemma, which follows by a little algebra. Note that the R.H.S in Lemma 2.1 (d) says that when the covariate is bounded, the log-ratio of the ALD likelihoods can be bounded by a continuous function in (α, β). This indeed turns out to be a very useful property and hence we introduce our first assumption.

(A.1) $\exists\ M_1 > 0, \textit{such that } |X_i| \leq M_1 \ \forall\ i \geq 1$

This is not an unreasonable assumption since X_i are non-random. For example, in a designed experiment for clinical trial, the X_i may be different levels of an administered drug. So, it is indeed reasonable to expect this. It turns out that if X_i are i.i.d random, we can modify this requirement to say $E|X_1| \leq M_1$.

We first work with the denominator of the posterior probability (i.e I_{2n}). In our quantile regression context, we consider Y_i's that are independent but not necessarily i.i.d, since their distribution depends on the non-random covariate X_i. Moreover, we are dealing with the case of mis-specified likelihood. To handle this case, we state the following proposition which is analogous to lemma 4.4.1, in Ghosh and Ramamoorthi(2003) used for proving consistency in a properly specified model under i.i.d assumption. The proof is omitted as it can be shown along the same lines and does not rely on the specific form of ALD.

Proposition 2.1. *If* $\forall\ \delta > 0,\ \Pi(V_\delta) > 0,$, *where,*

$$V_\delta = \left((\alpha,\beta) \in \Theta : \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E\left[\ log\frac{f_{(\alpha_0,\beta_0)i}(Y_i)}{f_{(\alpha,\beta)i}(Y_i)}\right] < \delta, \sum_{i=1}^\infty \frac{E\left[\left(log\frac{f_{(\alpha_0,\beta_0)i}(Y_i)}{f_{(\alpha,\beta)i}(Y_i)})\right)^2\right]}{i^2} < \infty \right)$$

then $\forall\ d > 0, \quad e^{nd} I_{2n} \to \infty\ a.s\ [P]$

Note that the condition required for the proposition 2.1 to hold is essentially a condition on the prior $\Pi(\alpha, \beta)$. Therefore, we introduce our next assumption on the prior.

(**A.2**) $\Pi(\{(\alpha,\beta) : |\alpha - \alpha_0| < \delta_1, |\beta - \beta_0| < \delta_2\}) > 0,\ \forall \delta_1 > 0, \delta_2 > 0$

Lemma 2.2. *If assumptions* **A.1 and A.2** *hold, then* $\forall\ d > 0, \quad e^{nd} I_{2n} \to \infty\ a.s\ [P]$

Proof. The idea is to verify that conditions of proposition 2.1 hold. By assumption **A.1** and part (d) of Lemma 2.1 it follows that

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E\left|log\left(\frac{f_{(\alpha_0,\beta_0)i}}{f_{(\alpha,\beta)\ i}}\right)(Y_i)\right| \quad \leq \quad Max(\tau,(1-\tau))(|\alpha-\alpha_0|+|\beta-\beta_0|M_1) \tag{2.2}$$

Since R.H.S of the above inequality is a continuous function in (α,β), for any $\delta > 0, \exists\ \delta_1, \delta_2$, such that $\forall\ (\alpha,\beta) : |\alpha-\alpha_0| < \delta_1, |\beta-\beta_0| < \delta_2$, $\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E\left|log\left(\frac{f_{(\alpha_0,\beta_0)i}}{f_{(\alpha,\beta)\ i}}\right)(Y_i)\right| < \delta$. Similarly, it follows that $E\left[log\left(\frac{f_{(\alpha_0,\beta_0)\ i}}{f_{(\alpha,\beta)\ i}}\right)(Y_i)\right]^2 \leq Max(\tau,(1-\tau))E(|\alpha-\alpha_0|+|\beta-\beta_0|M_1)^2 < \infty$, and hence $\sum_{i=1}^{\infty}\frac{E\left[log\left(\frac{f_{(\alpha_0,\beta_0)\ i}}{f_{(\alpha,\beta)\ i}}\right)(Y_i)\right]^2}{i^2} < \infty,\ \forall(\alpha,\beta)$. Now note that $\{(\alpha,\beta) : |\alpha-\alpha_0| < \delta_1, |\beta-\beta_0| < \delta_2\} \subseteq V_\delta$ and by assumption **A.2** $\Pi(V_\delta) > 0\ \forall\ \delta > 0$. This shows that conditions of proposition 2.1 are satisfied, thus completing the proof. □

It is interesting to note that among the family of ALD densities, $f_{(\alpha,\beta)i}$, *i.e*, the one with $\alpha = \alpha_0$ and $\beta = \beta_0$ minimizes the Kullback-leibler divergence with the true likelihood P_{0i}. This is indeed a consequence of our next lemma.

Lemma 2.3. *Let $f_{(\alpha,\beta)\ i}$ and p_{0i} be the probability density functions of ALD specification and the true underlying distribution of Y_i respectively, then the following identities and inequalities hold.*

(a) $E\left[log\left(\frac{f_{(\alpha,\beta)\ i}}{f_{(\alpha_0,\beta_0)\ i}}(Y_i)\right)\right] = E\left[(Y_i-\alpha-\beta X_i)I_{\alpha_0+\beta_0X_i<Y<\alpha+\beta X_i}\right] + E\left[(\alpha+\beta X_i-Y_i)I_{\alpha+\beta X_i<Y_i<\alpha_0+\beta_0X_i}\right]$

(b) $E\left[log\left(\frac{f_{(\alpha,\beta)\ i}}{f_{(\alpha_0,\beta_0)\ i}}\right)\right] \leq 0$

(c) $E\left[log\left(\frac{p_{0i}}{f_{(\alpha,\beta)\ i}}(Y_i)\right)\right] \geq E\left[log\left(\frac{p_{0i}}{f_{(\alpha_0,\beta_0)\ i}}(Y_i)\right)\right],\ \forall\ (\alpha,\beta) \in \Theta.$

Further, in (b) and (c), equality is achieved if $\alpha = \alpha_0$ and $\beta = \beta_0$

Proof. The proof of (a) just involves a bit of algebra and using the fact that $\alpha_0 + \beta_0 X_i$ is the τ^{th} quantile of $P(Y_i/X_i)$. (b) follows from (a) by noting that the expressions inside the expectation in both terms are negative. (c) is an immediate consequence of (b) by writing $E\left[log\left(\frac{p_{0i}}{f_{(\alpha,\beta)\ i}}(Y_i)\right)\right] = E\left[log\left(\frac{p_{0i}}{f_{(\alpha_0,\beta_0)\ i}}(Y_i)\right)\right] - E\left[log\left(\frac{f_{(\alpha,\beta)\ i}}{f_{(\alpha_0,\beta_0)\ i}}(Y_i)\right)\right]$ □

Now, we consider the numerator of equation (2.1), namely $I_{1n} = \int_\Theta \prod_{i=1}^n \frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)} d\Pi(\alpha,\beta)$. To lay out the idea of our approach, it helps to think through a slightly simpler case when $\Theta \subseteq \Re^2$ is compact. Recall, $U = \{(\alpha,\beta) \in \Theta : |\alpha - \alpha_0| < \Delta_1 \ and \ |\beta - \beta_0| < \Delta_2\}$. Then we can write

$$I_{1n} = \int_{U^c} \prod_{i=1}^n \frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)} d\Pi(\alpha,\beta) = \int_{U^c \cap \Theta} e^{\sum_{i=1}^n log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right)} d\Pi(\alpha,\beta)$$

We would like to get some $d_0 > 0$ such that $e^{nd_0}.I_{1n} \rightarrow \ 0$.

(i) Firstly, although the exponent in the integrand is not negative in general, we know from part (b) of Lemma 2.2 that $E\left[\sum_{i=1}^n log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right)\right] \leq 0$. Therefore, we would like to write the exponent in the integrand in terms of this expectation. A natural way to tackle this is via uniform SLLN over compact sets. Theorem 1.3.3 in Ghosh and Ramamoorti (2003) provides uniform strong law in the case when Y_i are i.i.d. However, in our case, we have Y_i independent but not identical. To our knowledge, this case has not recieved much attention in literature and hence we prove a version in the appendix (proposition A.1) and derive the specific result required for our problem in lemma 2.4.

(ii) Then our problem becomes that of showing $\int_{U^c \cap \Theta} e^{E\left[log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right)\right]} d\Pi(\alpha,\beta)$ decays to 0 at an exponential rate. We will address this in lemma 2.5.

(iii) Finally, when Θ is not necessarily compact, a natural idea would be to split the parameter space into two parts as $\Theta = G \cup G^c$, where G is a compact set such that the integral over G^c decays in a exponential manner.This will be covered in lemma 2.6.

We state these lemmas below. We defer the proofs to the appendix.

LEMMA 2.4. *Let $G \subset \Theta$ be compact. If assumption* **A.1** *holds, then*
$\sup_{(\alpha,\beta)\in\ G} \left| \frac{1}{n} \sum_{i=1}^n \left(log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right) - E\left[log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right)\right]\right)\right| \rightarrow \ 0 \ a.s \ [P]$

Since the objective of the model is to estimate the τ^{th} quantile, it is reasonable to assume that the quantile is unique. Otherwise, the linear function which we are trying to estimate (viz. $\alpha+\beta X_i$) will not be estimable. Another way to say this would be that for any $\Delta > 0$, $P(0 < Y_i - \alpha_0 - \beta_0 X_i < \Delta) \neq 0$. Similarly if the X_i's are all constant, then again the model will not be estimable. Therefore, it is reasonable to require that $\{X_i, \ for \ i \geq 1\}$ take on atleast two distinct values each infinitely many times. Without loss of generality (by adjusting the location of X_i's) this would mean that $\exists \ \epsilon_0 > 0 \ such \ that \ \liminf \frac{1}{n} \sum_{i=1}^{n} I_{X_i > \epsilon_0} > 0$ and $\liminf \frac{1}{n} \sum_{i=1}^{n} I_{X_i < -\epsilon_0} > 0$. Such a condition is indeed used by Amewou-atisso et al (2003). In our case, it so happens that we need the following assumption that is a combination of the two.

(**A.3**) $\exists \ \epsilon_0 > 0$ such that, for any $\Delta > 0$, the following conditions (i) to (iv) hold

(i) $\liminf \frac{1}{n} \sum_{i=1}^{n} P(0 < Y_i - \alpha_0 - \beta_0 X_i < \Delta) I_{X_i > \epsilon_0} > 0$

(ii) $\liminf \frac{1}{n} \sum_{i=1}^{n} P(0 < Y_i - \alpha_0 - \beta_0 X_i < \Delta) I_{X_i < -\epsilon_0} > 0$

(iii) $\liminf \frac{1}{n} \sum_{i=1}^{n} P(-\Delta < Y_i - \alpha_0 - \beta_0 X_i < 0) I_{X_i > \epsilon_0} > 0$

(iv) $\liminf \frac{1}{n} \sum_{i=1}^{n} P(-\Delta < Y_i - \alpha_0 - \beta_0 X_i < 0) I_{X_i < -\epsilon_0} > 0$

Note that this assumption in some sense is saying that the uniqueness of quantile is preserved in the limit. If the true model is a location shift model (i.e where $Y_i - \alpha_0 - \beta_0 X_i \ are \ i.i.d$ then this just reduces to saying $\liminf \frac{1}{n} \sum_{i=1}^{n} I_{X_i > \epsilon_0} > 0$ and $\liminf \frac{1}{n} \sum_{i=1}^{n} I_{X_i < -\epsilon_0} > 0$, which is similar to the condition specified in Amewou-Atisso et al (2003). This turns out to be the key assumption that helps prove the next two lemmas. Also, we find it more convenient to break up the set $U^c \ as \ \bigcup_{i=1}^{8} W_i$, where

$U = \{(\alpha, \beta) : |\alpha - \alpha_0| < \Delta_1 \ and \ |\beta - \beta_0| < \Delta_2\}$

$W_1 = \{(\alpha, \beta) : \alpha - \alpha_0 \geq \Delta_1, \beta \geq \beta_0\}$, $W_2 = \{(\alpha, \beta) : \alpha - \alpha_0 \geq \Delta_1, \beta < \beta_0\}$

$W_3 = \{(\alpha, \beta) : \alpha - \alpha_0 < -\Delta_1, \beta \geq \beta_0\}$, $W_4 = \{(\alpha, \beta) : \alpha - \alpha_0 < -\Delta_1, \beta < \beta_0\}$

$W_5 = \{(\alpha, \beta) : \alpha \geq \alpha_0, \beta - \beta_0 \geq \Delta_2\}$, $W_6 = \{(\alpha, \beta) : \alpha < \alpha_0, \beta - \beta_0 \geq \Delta_2\}$

$W_7 = \{(\alpha, \beta) : \alpha \geq \alpha_0, \beta - \beta_0 < -\Delta_2\}$, $W_8 = \{(\alpha, \beta) : \alpha < \alpha_0, \beta - \beta_0 < -\Delta_2\}$

Lemma 2.5. *If assumption* **A.3** *holds for some* $\epsilon_0 > 0$, *then for* $j = 1, 2, .., 8,\ \exists\ N_j^*$ *and constants* $K_j > 0$ *(depending on* $\Delta_1,\ \Delta_2,\ \epsilon_0$*), such that* $\forall\ n \geq N_j^*$,
$\int_{W_j} e^{\sum_{i=1}^{n} E\left[log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right)\right]} d\Pi(\alpha, \beta) \leq e^{-nK_j}$

As mentioned earlier, for the case when the parameter space Θ is compact, Lemma 2.4 and Lemma 2.5 would be sufficient to prove the desired convergence result for I_{1n}. In order to handle the case when the parameter space is not necessarily compact, the idea would be to find a compact set G, such that the conclusions of Lemma 2.4 still holds and the integral over G^c declines exponentially. To achieve this, we introduce the final two assumptions, whose need will become clear in the proofs.

(**A.4**) The prior Π on Θ is proper

The assumption on propriety of prior is in fact not needed. We will see in section 4 that as long as the posterior distribution is well defined, the results of this paper will hold.

(**A.5**) $\sum_{i=1}^{\infty} \frac{E\left[|Z_i|^2\right]}{i^2} < \infty$ and $S_1 = \limsup_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} E\left[|Z_i|\right] < \infty$, *where* $Z_i = Y_i - \alpha_0 - \beta_0 X_i$

The last assumption is for us to be able to apply Kolmogorov's SLLN for non-iid random variables $\{Z_n\}$.

Lemma 2.6. *If assumptions* **A.1**, **A.4** *and* **A.5** *hold, then for* $j = 1, 2, .., 8,\ \exists$ *a compact set* $G_j \subset W_j$ *and integer* $N_j^{**}(\omega)$ *such that* $\int_{G_j^c \cap W_j} e^{\sum_{i=1}^{n} log \frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}} d\Pi(\alpha, \beta) \leq e^{-b_j n}$ *for some* $b_j > 0,\ \forall\ n \geq N_j^{**}(\omega)$

The proof of this lemma again exploits the specific form of ALD to construct the set G. The proof is provided in the appendix. With this, we come to the main theorem of this paper.

Theorem 2.1. *Let* $\{Y_i\}, i = 1, 2, ...n$ *be* n *independent observations of a univariate response and let* $\{X_i\}, i = 1, 2, .., n$ *be 1-dimensional non-random covariates. Let* P_{0i} *denote the true (but unknown) probability distribution of* (Y_i, X_i), *with the true* τ^{th} *conditional quantile given by* $Q_\tau(Y_i/X_i) = \alpha_0 +$

$X_i^T\beta_0$. *Suppose however that the specified model for* Y_i *is* $ALD(.,\mu_i^\tau,\sigma=1,\tau)$, *where* $\mu_i^\tau=\alpha+X_i^T\beta$, *which is used to estimate* (α,β). *Let* $U=\{(\alpha,\beta):|\alpha-\alpha_0|<\Delta_1$ *and* $|\beta-\beta_0|<\Delta_2\}$ *for any arbitrary* $\Delta_1>0,\Delta_2>0$. *Then, under assumptions* **A.1 to A.5**, $\Pi(U^c/Y_1,Y_2,..,Y_n)\to 0$ *a.s.* $[P]$

Proof. As in equation (2.1) we write $\Pi(U^c/Y_1,Y_2,..,Y_n)=\frac{I_{1n}}{I_{2n}}$. Firstly, by Lemma 2.2, we have $\forall\ d>0,\ \ e^{nd}I_{2n}\to\infty$ *a.s* $[P]$. Therefore, it is enough to show that $\exists\ d_0>0$ *such that* $e^{nd_0}I_{1n}\to 0$ *a.s* $[P]$. Because then we would have $\Pi(U^c/Y_1,Y_2,..,Y_n)=\frac{I_{1n}}{I_{2n}}=\frac{e^{nd_0}I_{1n}}{e^{nd_0}I_{2n}}\to 0$ *a.s.* $[P]$. Note that $I_{1n}=\sum_{j=1}^8\int_{W_j}\prod_{i=1}^n\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}d\Pi(\alpha,\beta)$. Let us denote the j^{th} term by I_{1n}^j. Let $G_j,$ *for* $j=1,2..,8$ be the compact set as given by Lemma 2.6. Then,

$$I_{1n}^j\ =\ \int_{W_j\bigcap G_j}e^{\sum_{i=1}^n log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right)}d\Pi(\alpha,\beta)+\int_{W_j\bigcap G_j^c}e^{\sum_{i=1}^n log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right)}d\Pi(\alpha,\beta)$$

Let $\eta_j>0$. It's exact value will be determined later. By Lemma 2.4,
$\sum_{i=1}^n log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right)<\sum_{i=1}^n E\left[log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right)\right]+n\times\eta_j\ \forall\ n\geq$ *some* $k_0(\omega),\forall\ (\alpha,\beta)\in G_j\cap W_j$
It then follows by Lemma 2.5 that $\exists\ N_j^*>k_0(\omega)$ and $K_j>0$ such that $\forall\ n\geq N_j^*$

$$\begin{aligned}&\int_{W_j\bigcap G_j}e^{\sum_{i=1}^n log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right)}d\Pi(\alpha,\beta)\\ \leq\ &\int_{W_j\bigcap G_j}e^{\sum_{i=1}^n E\left[log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right)\right]+n\times\eta_j}d\Pi(\alpha,\beta)\\ \leq\ &e^{-nK_j}\times e^{n\times\eta_j}\\ \leq\ &e^{-n\frac{K_j}{2}}\\ &(by\ choosing\ \eta_j=\frac{K_j}{2})\end{aligned}$$

Also, by Lemma 2.6, the choice of the set G_j is such that, $\forall\ n\geq N_j^{**}$,

$$\int_{W_j\bigcap G_j^c}e^{\sum_{i=1}^n log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right)}d\Pi(\alpha,\beta)\leq e^{-nb_j}$$

The inequalities for the first and second term together imply that $\forall\ n\geq\ N_j=max(N_j^*(\omega),N_j^{**}(\omega))$

$$\begin{aligned}I_{1n}^j\ &\leq\ e^{-n\frac{K_j}{2}}+e^{-nb_j}\\ &\leq\ 2\times e^{-2nd_j}\end{aligned}$$

where, $2d_j = Min(b_j, \frac{K_j}{2})$. It follows then that

$I_{1n} = \sum_{j=1}^{8} I_{1n}^{j} \leq 16 \times e^{-2nd_0}\ \forall\ n \geq Max(N_1, ..., N_8)$, where $d_0 = Min(d_1, d_2, ..., d_8)$. This in turn implies that for $e^{nd_0} \times I_{1n} \to 0\ a.s.\ [P]$. □

2..2 Generalization to the case of multiple covariates

Generalizing the above results to accomodate multiple covariates that include random components is actually quite easy. We will denote the p-dimensional covariate for the i^{th} observation by $(X_{i1}, X_{i2}, ..., X_{ip})$. Correspondingly, we will denote the vector of β by $(\beta_1, ..., \beta_p)$. The prior Π will understandably be on the parameters $(\alpha, \beta_1, ..., \beta_p)$. The true (unknown) value of the parameter is denoted by $(\alpha_0, \beta_{01}, ..., \beta_{0p})$. It is easy to see that the analogous result of Theroem 1 will hold with the following analogous assumptions.

(A*.1) $\exists\, M_1 > 0, \textit{such that}\ \forall\, i \geq 1, |X_{ij}| \leq M_1$ for covariates $X_{.j}$ that are non-random and $E|X_{ij}| \leq M_1$ for covariates $X_{.j}$ that are random and i.i.d across observations i=1,2,3...,n.

(A*.2) $\Pi(\{(\alpha, \beta_1, ..., \beta_p) : |\alpha - \alpha_0| < \delta_1, |\beta_1 - \beta_{01}| < \delta_2, ..., |\beta_p - \beta_{0p}| < \delta_{p+1}\}) > 0,\ \forall \delta_1 > 0, \delta_2 > 0, ..., \delta_{p+1} > 0$

(A*.3) $\exists\, \epsilon_0 > 0$ such that, for any $\Delta > 0$, the following conditions (i) to (iv) hold

(i) $\liminf \frac{1}{n} \sum_{i=1}^{n} E\left[I_{\{0<Y_i-\alpha_0-X_i^T\beta_0<\Delta\}} \times I_{X_{ij}>\epsilon_0}\right] > 0,\ \forall\ j = 1, 2, ..., p$

(ii) $\liminf \frac{1}{n} \sum_{i=1}^{n} E\left[I_{\{0<Y_i-\alpha_0-X_i^T\beta_0<\Delta\}} \times I_{X_{ij}<-\epsilon_0}\right] > 0,\ \forall\ j = 1, 2, ..., p$

(iii) $\liminf \frac{1}{n} \sum_{i=1}^{n} E\left[I_{\{-\Delta<Y_i-\alpha_0-X_i^T\beta_0<0\}} \times I_{X_{ij}>\epsilon_0}\right] > 0,\ \forall\ j = 1, 2, ..., p$

(iv) $\liminf \frac{1}{n} \sum_{i=1}^{n} E\left[I_{\{-\Delta<Y_i-\alpha_0-X_i^T\beta_0<0\}} \times I_{X_{ij}<-\epsilon_0}\right] > 0,\ \forall\ j = 1, 2, ..., p$

(A*.4) The prior Π on Θ is proper

(A*.5) $\sum_{i=1}^{\infty} \frac{E\left[|Z_i|^2\right]}{i^2} < \infty$ and $S_1 = \limsup_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} E\left[|Z_i|\right] < \infty,\ \textit{where}\ Z_i = Y_i - \alpha_0 - X_i^T\beta_0$

3. Examples

In this section, we will demonstrate that bayesian analysis carried out with ALD will work for a wide range of possiblities for the true underlying likelihood. Basically, we will analyze the workings of assumptions $A^*.3, and\ A^*.5$. The other assumptions are either on the prior or the covariates, which we will assume to hold for the purpose of this discussion. It is worth noting that these assumptions are typically satisfied if the probabilities and expectations involved turn out to be bounded functions of the non random covariates. Also, if they happen to be continuous functions of non-random covariates then by the boundedness assumption $A^*.1$, one can argue the validity of the required assumptions.

3..1 Example 1: Location Models

Consider the case when conditional on X_i, $Y_i = \alpha_0 + X_{i1}^T\beta_{10} + X_{i2}^T\beta_{20} + e_i$ where the error terms $e_i, for\ i = 1, 2, ..., n$'s are i.i.d from true unknown P_0 (with density function p_0 w.r.t lebesgue measure), with it's τ^{th} quantile at 0. For clarity we split the covariate into two parts with X_{i1} representing the random part that is i.i.d for i=1,2,...n and X_{i2} the nonrandom part. Note that $Z_i = Y_i - \alpha_0 - X_{i1}^T\beta_{10} - X_{i2}^T\beta_{20} = e_i$ are i.i.d. Now note the following facts

(a) Assumption $A^*.3$ will be satified if $P(0 < Z_1 < \Delta) > 0,\ \forall\ \Delta > 0$ and if $\exists\ \epsilon_0 > 0$ such that $\liminf \frac{1}{n}\sum_{i=1}^{n} I_{X_{ij}<-\epsilon_0}\ > 0\ and\ \liminf \frac{1}{n}\sum_{i=1}^{n} I_{X_{ij}>\epsilon_0}\ > 0$, for non random covariates $X_{.j}$, and $P(X_{1j} > \epsilon_0) > 0\ and\ P(X_{1j} < -\epsilon_0) > 0$ for random covariates. Latter is a condition on the covariates. The former condition is satisfied by any distribution which is monotonic. in particular, normal distribution with location shifted so as to make τ^{th} quantile zero or even mixtures of such distributions would satisfy this condition. Similarly, one can consider location shifted gamma, beta etc

(b) $A^*.5$ is satisfied if Z_i has finite variance. e.g location shifted normal, gamma, beta

Proposition 3.1. *Suppose the true model for response conditional on X_i is given by $Y_i = \alpha_0 + X_{1i}^T\beta_{10} + X_{2i}^T\beta_{20} + e_i$ where the error terms $e_i, for\ i = 1, 2, ..., n$'s are i.i.d from unknown distribution*

P_0 (with density function p_0 w.r.t lebesgue measure), with it's τ^{th} quantile at 0. Suppose that along with $A^*.1, A^*.2, A^*.4$, the following conditions hold

(i) $P(0 < Z_1 < \Delta) > 0,\ \forall\ \Delta > 0$ *and* $\exists\ \epsilon_0 > 0$ *such that* $\liminf \frac{1}{n}\sum_{i=1}^{n} I_{X_{ij}<-\epsilon_0} > 0$ *and* $\liminf \frac{1}{n}\sum_{i=1}^{n} I_{X_{ij}>\epsilon_0} > 0$*, for non random covariates* $X_{.j}$*, and* $P(X_{1j} > \epsilon_0) > 0$ *and* $P(X_{1j} < -\epsilon_0) > 0$ *for random covariates*

(ii) The error distribution P_0 has finite variance

Then posterior consistency will hold for the quantile regression parameters estimated using ALD for the response.

In particular, when p_0 is normal, beta or a gamma distribution (with shape parameter$\geq$1) then such a result indeed holds.

3..2 Example 2: Scale Models

An important feature of our result is that it can cover mis-specifications other than location models. To demonstrate this let us consider the case where the density function of Y_i conditional on X_i is given by $p_0\left(\frac{y\ \mu_0^\tau}{l(X_i)}\right) \times \frac{y\ \mu_0^\tau}{l(X_i)}$, where p_0 is a probability density function on $(0,\infty)$ with τ^{th} quantile=μ_0^τ and $l(X_i) = \alpha_0 + X_{i1}^T\beta_{10} + X_{i2}^T\beta_{20}$, where $X_{.1}$ and $X_{.2}$ denote vectors of random and non-random covariates. Ofcourse, we assume that the covariates X_i's are such that $l(X_i) > 0$. Note that this means that the τ^{th} quantile of Y_i, *given* X_i is $l(X_i) = \alpha_0 + X_{i1}^T\beta_{10} + X_{i2}^T\beta_{20}$. A gamma density would be a special case of such a model. Again, we would like to study the working of the asumptions $A^*.3$ *and* $A^*.5$.

We will investigate assumption $A^*.3$ by considering one of the four sub conditions, since others would be similar. $E\left[I_{0<Y_i-l(X_i)<\Delta} \times I_{X_{ij}>\epsilon_0}\right] = E\left[P_0\left(\mu_0^\tau < U < \frac{\Delta\mu_0^\tau}{l(X_i)} + \mu_0^\tau/X_i\right) \times I_{X_{ij}>\epsilon_0}\right] = E\left[\left(G_p\left(\frac{\Delta\mu_0^\tau}{l(X_i)} + \mu_0^\tau\right) - G_p\left(\mu_0^\tau\right)\right) \times I_{X_{ij}>\epsilon_0}\right]$, where $U \sim G_p$ is the distribution function of p_0 and is independent of X_i. Again, the expectation in the last expresion is over the random covariates X_{1i}, thus resulting in a function of non-random covariates X_{i2}. Since G_p is a continuous function (because it has a density w.r.t lebesge measure), $\left(G_p\left(\frac{\Delta\mu_0^\tau}{l(X_i)} + \mu_0^\tau\right) - G_p\left(\mu_0^\tau\right)\right) \times I_{X_{ij}>\epsilon_0}$ is a continuous function in X_i on the set $X_{ij} > \epsilon_0$. So, (noting that $G_p \leq 1$ and applying DCT) the expectation will result in a continuous function of the non-random covariates. If the distribution function G_p is also a strictly increasing continuous function (which is certainly true if for example p_0 is the gamma density) then this will result in a strictly positive continuous function of nonrandom covariates. Since the non-random covariates are bounded, by compactness and continuity it follows that this expectation will be bounded away from zero for all possible values of the non-random covariate. Further, if $\exists\, \epsilon_0 > 0$ such that $\liminf \frac{1}{n}\sum_{i=1}^{n} I_{X_{ij}<-\epsilon_0} > 0$ *and* $\liminf \frac{1}{n}\sum_{i=1}^{n} I_{X_{ij}>\epsilon_0} > 0$, for non random covariates $X_{.j}$, and $P(X_{1j} > \epsilon_0) > 0$ *and* $P(X_{1j} < -\epsilon_0) > 0$ for random covariates then assumption $A^*.3$ will be satisfied. In particular, this is satisfied for P_0 having a gamma distribution.

For assumption $A^*.5$, we just note that $E\left[Z_i^2\right] = E\left[(U-\mu_0^\tau)^2 \times (l(X_i))^2\right] = E\left[(U-\mu_0^\tau)^2\right] \times E\left[(l(X_i))^2\right]$, where $U \sim P_0$. As long as the second moment of P_0 exists (which is satisfied by gamma) and second moments of the random covariates exist, then this expression is a bounded function in the non-random covariates.

Proposition 3.2. *Suppose the density function of* Y_i *conditional on* X_i *is given by* $p_0\left(\frac{y\ \mu_0^\tau}{l(X_i)}\right) \times \frac{y\ \mu_0^\tau}{l(X_i)}$, *where* p_0 *is a probability density function on* $(0,\infty)$ *with* τ^{th} *quantile*$=\mu_0^\tau$ *and* $l(X_i) = \alpha_0 + X_{i1}^T\beta_{10} + X_{i2}^T\beta_{20}$. *Suppose condition* $A^*.1, A^*.2, A^*.4$ *hold. Suppose further the following conditions hold*

(i) Each random covariate has a finite second moment

(iii) The distribution function of the density p_0 *is a strictly increasing continuous function with finite second moment.*

(iv) $\exists\, \epsilon_0 > 0$ *such that* $\liminf \frac{1}{n}\sum_{i=1}^{n} I_{X_{ij}<-\epsilon_0} > 0$ *and* $\liminf \frac{1}{n}\sum_{i=1}^{n} I_{X_{ij}>\epsilon_0} > 0$*, for non random covariates* $X_{.j}$*, and* $P(X_{1j} > \epsilon_0) > 0$ *and* $P(X_{1j} < -\epsilon_0) > 0$

Then posterior consistency will hold for the quantile regression parameters estimated using ALD.

In particular if p_0 is a gamma density with shape parameter, the covariates are bounded and $\exists\, \epsilon_0 > 0$ such that $\liminf \frac{1}{n}\sum_{i=1}^{n} I_{X_{ij}<-\epsilon_0} > 0$ *and* $\liminf \frac{1}{n}\sum_{i=1}^{n} I_{X_{ij}>\epsilon_0} > 0$, for non random covariates $X_{.j}$, and $P(X_{1j} > \epsilon_0) > 0$ *and* $P(X_{1j} < -\epsilon_0) > 0$ then posterior consistency will hold for the quantile regression model that is based on a mis-specified ALD model

3..3 Example 3: A Normal location scale model

Let $Y_i \sim N(l(X_i) - \rho_\tau \phi(X_i), \phi^2(X_i))$, where $l(X_i) = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2}$ and $\phi(X_i)$ where ρ_τ is the τ^{th} quantile of standard normal distribution. Let X_{i1} are random and X_{i2} are non-random covariates. Clearly, the τ^{th} quantile of Y_i given X_i is $l(X_i)$.

For $A^*.3$, we again argue with one of the four subconditions since the argument for the others is similar. $E\left[I_{0<Y_i-l(X_i)<\Delta} \times I_{X_{ij}>\epsilon_0}\right] = E\left[\left(\Phi\left(\frac{\Delta}{\phi(X_i)} + \rho_\tau\right) - \Phi\left(\rho_\tau\right)\right) I_{X_{ij}>\epsilon_0}\right]$, where $\Phi()$ is the standard normal distribution function. Clearly, if $\phi(X_i)$ is a continuous function in X_i then by DCT this is a continuous function of the non-random covariates on the set $X_{ij} > \epsilon_0$. If $\exists\ \epsilon_0 > 0$ such that $\liminf \frac{1}{n}\sum_{i=1}^n I_{X_{ij}<-\epsilon_0} > 0$ *and* $\liminf \frac{1}{n}\sum_{i=1}^n I_{X_{ij}>\epsilon_0} > 0$, for non random covariates $X_{.j}$, and $P(X_{1j} > \epsilon_0) > 0$ *and* $P(X_{1j} < -\epsilon_0) > 0$, then condition $A^*.4$ will be satisfied.

To check $A^*.5$, observe that $E\left[Z_i^2\right] = E\left[\left(S + \rho_\tau\right)^2 \phi^2(X_i)\right] = E\left[\left(S + \rho_\tau\right)^2\right] E\left[\phi^2(X_i)\right]$. Therefore if $E\left[\phi^2(X_i)\right]$ is a bounded function of the non-random covariates then the required assumption will be satisfied.

4. Extending the result for improper priors

We can indeed relax the assumption on the propriety of prior (**A.4**) as long as the posterior is well defined. Our argument is in the lines of Choi and Ramamoorthi(2003). If $\Pi(./Y_1)$ is proper, then we could just apply Theorem 2.1 with $\Pi(./Y_1)$ as the prior in place of Π. However, we also need to make sure that condition **A.2** is satisfied by $\Pi(./Y_1)$ in place of Π. The following proposition ensures that the required condition is indeed satisfied. This result is particularly interesting in view of theorem 1 of Yu and Moyeed(2001) where it is shown that the posterior based on ALD is always well defined for a flat prior(i.e. $\Pi(.) \propto 1$). Therefore, the following proposition would imply in particular that Theorem 2.1 would hold when the prior Π is flat.

We will state the result for the case of a single covariate. The same argument works for the multiple covariates

Proposition 4.1. *Let Y_1 be univariate response and X_1 be 1-dimensional non-random covariate.*

Let the specified model for Y_1 be $ALD(., \mu_1^\tau, \sigma = 1, \tau)$, where $\mu_1^\tau = \alpha + \beta X_1$. Let P_{01} be the true (but unknown) probability distribution of Y_1, with the true τ^{th} conditional quantile given by $Q_\tau(Y_1/X_1) = \alpha_0 + \beta_0 X_1$. Let Π be a prior on the parameter space Θ such that $0 < \int_\Theta f_{(\alpha,\beta)1}(Y_1) d\Pi(\alpha, \beta) < \infty$ a.e.$[P_{01}]$. If $E \subset \Theta$ is such that $\Pi(E) > 0$, then $\Pi(E/Y_1) > 0$ a.e.$[P_{01}]$

Proof. Firstly, $\exists\ \Omega_1$ such that $P_{01}(\Omega_1) = 1$ and $0 < \int_\Theta f_{(\alpha,\beta)1}(Y_1(\omega)) d\Pi(\alpha, \beta) < \infty,\ \forall\ \omega \in \Omega_1$.
We know that,

$$\Pi(E/Y_1(\omega)) = \frac{\int_E f_{(\alpha,\beta)1}(Y_1(\omega)) d\Pi(\alpha, \beta)}{\int_\Theta f_{(\alpha,\beta)1}(Y_1(\omega)) d\Pi(\alpha, \beta)}$$

By nature of the ALD density function (equation (1.3)) $f_{(\alpha,\beta)1}(y) > 0,\ \forall\ y \in \Re,\ \forall\ (\alpha, \beta)$. Therefore, since $\Pi(E) > 0$, we have, $\int_E f_{(\alpha,\beta)1}(Y_1(\omega)) d\Pi(\alpha, \beta) > 0$ and hence $\Pi(E/Y_1) > 0, \forall\ \omega \in \Omega_1$. □

5. Simulation

We empirically verify the results of this paper by simulating from four different models and checking whether ALD based quantile regression indeed leads to reasonable results. We include two covariates, X_1, X_2 simulated from a truncated(between 1 and 1000) normal distribution with mean 3 and variance 1. The other variable is a 0-1 values variable simulated from bernoulli distribution with mean 0.3. For each model, conditional on X_i the $\tau = 75^{th}$ percentile is given by $\alpha_0 + \beta_{01} X_1 + \beta_{02} X_2$ where $(\alpha_0, \beta_{01}, \beta_{02}) = (1, 2, 3)$. For the bayesian estimation, a normal prior with mean =0 and variance=100 is used for each of the quantile regression coefficients. This kind of a weakly informative prior is commonly used in practice. The four models conditioned on X_1, X_2 can be described as follows

1. **Location shifted normal** : $Y \sim N(\alpha_0 + \beta_{01} X_1 + \beta_{02} X_2 - \rho_\tau, 1)$ where $\rho_\tau = \rho_{.75}$ is the 75^{th} percentile of standard normal distribution.

2. **Location shifted gamma** : $Y = \alpha_0 + \beta_{01} X_1 + \beta_{02} X_2 - \rho_\tau + e$, where $e \sim Gamma(scale = 1, shape = 1)$, ρ_τ is the τ^{th} quantile of $Gamma(scale = 1, shape = 1)$

3. **Scaled gamma** : $Y \sim Gamma(scale = \frac{1}{\alpha_0 + \beta_{01} X_1 + \beta_{02} X_2}, shape = 2)$

4. **Location shifted and scaled normal** : $Y \sim N(\alpha_0 + \beta_{01}X_1 + \beta_{02}X_2 - \rho_\tau|\alpha_0 + \beta_{01}X_1 + \beta_{02}X_2|, |\alpha_0 + \beta_{01}X_1 + \beta_{02}X_2|^2)$

Bayesian estimation of the ALD model with the above mentioned prior can be done by formulating a gibbs sampling algorithm.To facilitate a simple formulation of the MCMC scheme, we use the representation of ALD as a scaled mixture of normals in the lines of Yue and Rue(2011). Figure 1 shows the plot of the 2.5^{th} percentiles, mean and the 97.5^{th} percentiles of the posterior distribution of the intercept term, as estimated using the MCMC samples. In order to get a feel for the convergence of the estimates to the true parameter value, the estimation is done for different data sizes starting from as small as 100 data points to 25000 data points. For each case, the estimation is based on 1000 MCMC simulations after the burnin period. For smaller data sizes, as we should expect, we see that the distance between the extreme percentiles is larger. However, we can clearly see that as the data size increases the distance between the extreme percentiles narrrow down towards the true parameter value. Similarly, figure 2 and 3 show the plots for the coefficient of X_1 and X_2 respectively.

6. Conclusion

The main contribution of this paper has been to provide a mathematical justification using Assymetric Laplace Distribution(ALD) for bayesian quantile regression, even if the true underlying distribution may not be ALD. The method is justfied under some reasonable conditions on the covariates and the underlying true distribution. This is significant given the fact that this approach has been used extensively since the work of Yu and Moyeed (2001), but has only been verified empirically in the past without a thorough mathematical investigation. We would like to emphasize the fact that one of the main reasons why this approach works is due to the nice property of ALD (Lemma 2.3) that the Kullback-leibler divergence happens to get minimized at the true regression parameter values. This is not in general true for any distribution. For example, if instead of ALD, we use a normal density function whose location is adjusted so that the τ^{th} quantile is $\alpha_0 + X^T\beta_0$, this condition is sometimes violated depending on the true underlying distribution. In fact, we empirically verified(results not shown) that using a location adjusted normal instead of ALD works when the true underlying

Figure 1: Bayesian estimation of intercept term

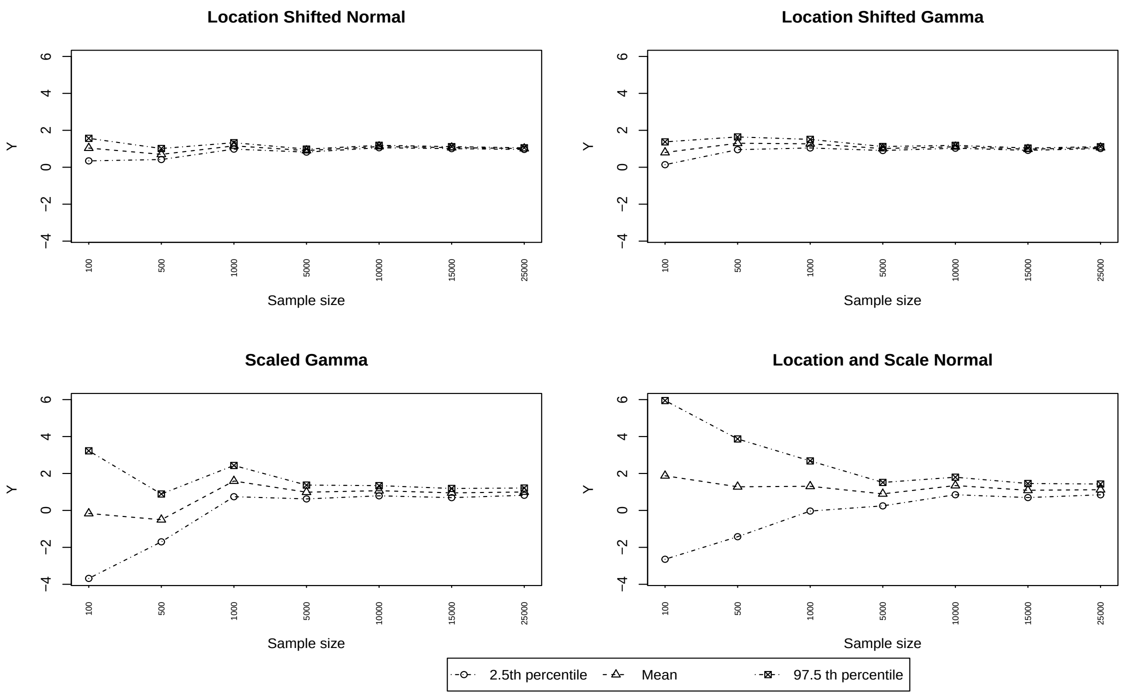


Figure 2: Bayesian estimation of co-efficient of X_1

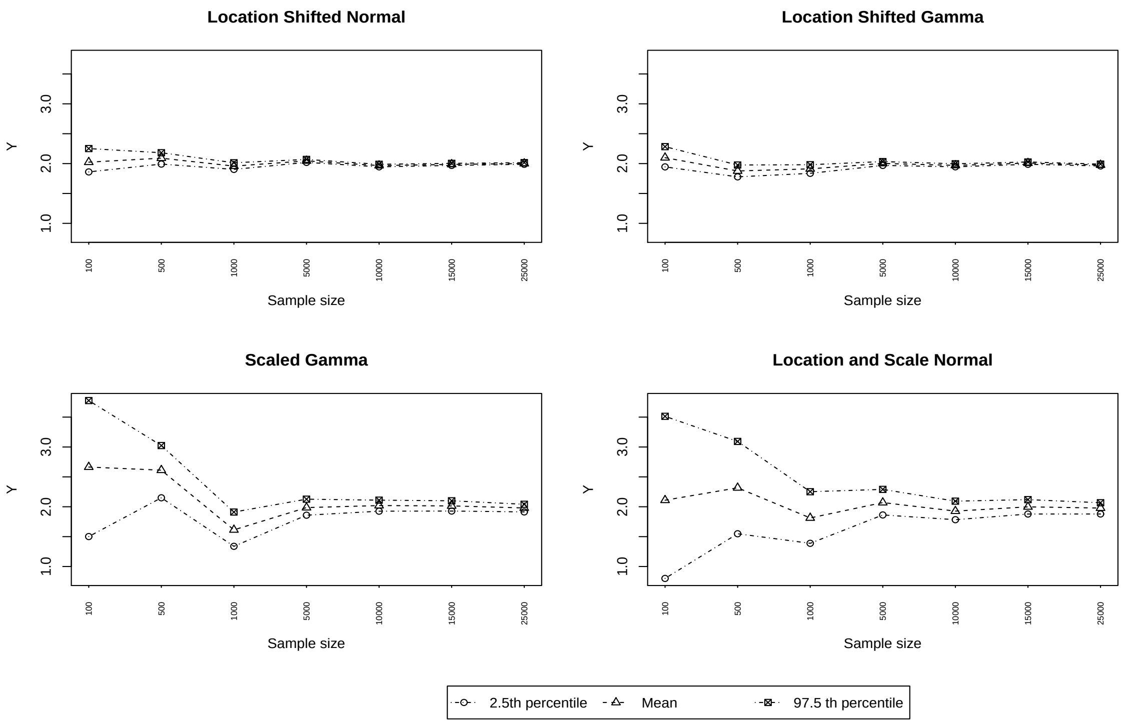
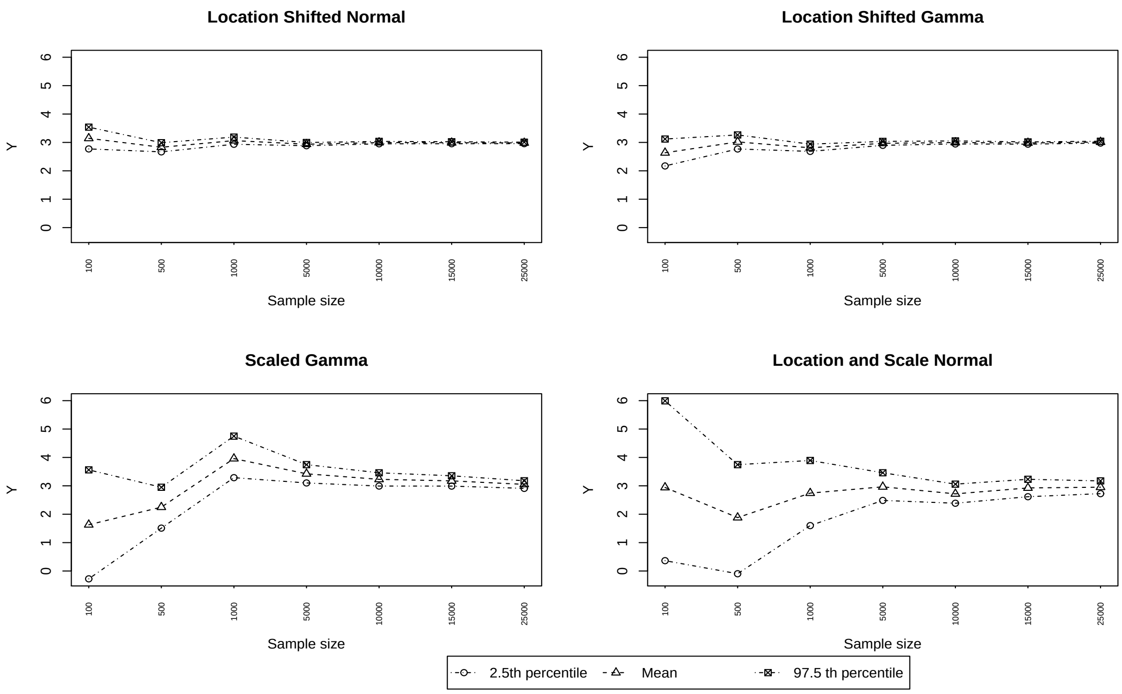


Figure 3:Bayesian estimation of co-efficient of X_2



distribution is a location shifted normal or location shifted gamma but breaks down when the true underlying distribution is more complicated (such as a scaled gamma or even a normal where the location and scale depend on the covariates). It may be worth investigating to see if the results can be generalized to encompass other potential "convenient" distributions apart from ALD for carrying out bayesian quantile regression.

In this paper, we focused on the case where the quantile of the response is modeled as a linear function in covariates. It is easy to see that the same result would apply to the case when we choose to use non-linear transformations of the covariates as long as the transformed covariates satisfy the conditions of the theorem. For instance, if the model is formulated as being linear in finitely many spline basis functions, all we would need the basis elements (taken as covariates) satisfy the conditions of the theorem.

Suppose we use a more general non-parametric formulation where the quantile is modeled as $\phi(X_i)$, by putting a prior on class of functions $\{\phi\}$. In this case, it is interesting to note that the conclusion of Lemma 2.2 (which corresponds to the required result for the denominator I_{2n} of (2.1)) still holds as long as the prior puts positive probability on kullback-liebler neighborhoods of the form(V_δ. However, the analogous results for the numerator are not as straight forward and will depend on the specific choice of prior chosen on ϕ. However, it is still worth noting that whatever be the prior chosen, as long as the conclusions of lemma 2.4, 2.5 and 2.6 hold then the result will hold for the non-parametric formulation as well. One may argue similarly for a semi-parametric formulation also.

We belive that the result of this paper will hold for the case of longitudinal data where the number of subjects are fixed but the number of observations increase with time. This is because the result can be applied to each subject separately. It may be interesting to extend these results to the case of longitudinal data where the number of subjects and time are both changing.

Finally, we would like to remark on the scale parameter σ of the ALD density in (1.3). In this paper we fixed σ at 1. It is easy to see that the same arguments go through if we fix a different value for $\sigma = \sigma_0$. We found empirically (results not shown) that a careful choice of this scale parameter can drastically improve the rate of convergence to the true parameter values. For instance if the support of the response Y is known to be between 0 and 1, then one can choose σ_0 so that the essential support of ALD to a large extent lies in $[0, 1]$ versus just using $\sigma = 1$ and this can drastically improve the convergence to the true parameter value. This observation is particularly interesting because the scale parameter almost plays no role in the frequentist analysis using ALD but seems to become useful for a bayesian analysis. Related to this, we also observe empirically that posterior consistency holds even in the case where we do not fix σ but put a prior on this parameter independent of the prior on (α, β). We continue to investigate these empirical observations to gain some theoretical insights.

APPENDIX

PROPOSITION A.1. *Let* $Y_i \sim P_{0i},\ i = 1, 2, ...,$ *be a sequence of independent random variables and* P *denote the corresponding product measure. Let*

(i) Θ *be a compact parameter space*

(ii) $T_i(\theta, Y_i)$ *be measurable and bounded i.e.* $\exists\ M > 0$ *such that* $|T_i(\theta, y)| \leq\ M\ a.s\ [P],\ \forall\ \theta \in \Theta$ *and* $\forall\ y$

(iii) For any $\theta_0 \in \Theta$,
$\lim_{\delta \to 0} sup_{n \geq 1} E\left[\frac{1}{n}\sum_{i=1}^{n} sup_{\{\theta : |\theta - \theta_0| < \delta\}} |T_i(\theta, Y_i) - E\left[T_i(\theta, Y_i)\right] - T_i(\theta_0, Y_i) + E\left[T_i(\theta_0, Y_i)\right]|\right] = 0.$

Then, $\lim_{n \to \infty} sup_{\theta \in \Theta}\left|\frac{1}{n}\sum_{i=1}^{n} T_i(\theta, Y_i) - \frac{1}{n}\sum_{i=1}^{n} E\left[T_i(\theta, Y_i)\right]\right| = 0\ a.s.[P]$

PROOF. Let $\epsilon > 0$ be arbitrary. By conditions (i) and (iii) $\exists\ \theta_1, .., \theta_k$ and $\delta_1, ..., \delta_k$ such that $sup_{n \geq 1} E\left[\frac{1}{n}\sum_{i=1}^{n} sup_{\{\theta : |\theta - \theta_j| < \delta_j\}} |T_i(\theta, Y_i) - E\left[T_i(\theta, Y_i)\right] - T_i(\theta_j, Y_i) + E\left[T_i(\theta_j, Y_i)\right]|\right] < \epsilon$ and $\bigcup_{j=1}^{k} B_j = \Theta,\ where\ B_j = \{\theta \in \Theta : |\theta - \theta_j| < \delta_j\},\ for\ j = 1, 2, ..., k$

Let $Z_{ij} = sup_{\{\theta : |\theta - \theta_j| < \delta_j\}} |T_i(\theta, Y_i) - E\left[T_i(\theta, Y_i)\right] - T_i(\theta_j, Y_i) + E\left[T_i(\theta_j, Y_i)\right]|$. By condition (ii), T_i are bounded and so are Z_{ij} (in fact by $4M$). Therefore, we can apply Kolmogorov's strong law to T_i as well as $Z_{ij},\ for\ i = 1, 2, 3, ..., n....$ So, for each j, we would have $\frac{1}{n}\sum_{i=1}^{n}(Z_{ij} - E\left[Z_{ij}\right]) \to\ 0\ a.s.[P]$ and $\frac{1}{n}\sum_{i=1}^{n} T_i(\theta_j, Y_i) - \frac{1}{n}\sum_{i=1}^{n} E\left[T_i(\theta_j, Y_i)\right]\ \to\ \ 0\ a.s.[P]$ .It follows that $\exists\ n_0\ such\ that\ \forall\ j = 1, 2, ..., k\ and\ \ n \geq n_0, |\frac{1}{n}\sum_{i=1}^{n} T_i(\theta_j, Y_i) - \frac{1}{n}\sum_{i=1}^{n} E\left[T_i(\theta_j, Y_i)\right]| < \epsilon$ and $|\frac{1}{n}\sum_{i=1}^{n}(Z_{ij} - E\left[Z_{ij}\right])| < \epsilon)$

Now, let $\theta \in \Theta,\ then\ \exists\ j\ such\ that\ \theta \in\ B_j$. So,

$$
\begin{aligned}
& \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}T_i(\theta,Y_i)-\frac{1}{n}\sum_{i=1}^{n}E\left[T_i(\theta,Y_i)\right]\right| \\
= & \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}(T_i(\theta,Y_i)-E\left[T_i(\theta,Y_i)\right]-T_i(\theta_j,Y_i)+E\left[T_i(\theta_j,Y_i)\right])+\frac{1}{n}\sum_{i=1}^{n}(T_i(\theta_j,Y_i)-E\left[T_i(\theta_j,Y_i)\right])\right| \\
\leq & \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}(T_i(\theta,Y_i)-E\left[T_i(\theta,Y_i)\right]-T_i(\theta_j,Y_i)+E\left[T_i(\theta_j,Y_i)\right])\right|+\left|\frac{1}{n}\sum_{i=1}^{n}(T_i(\theta_j,Y_i)-E\left[T_i(\theta_j,Y_i)\right])\right| \\
\leq & \frac{1}{n}\sum_{i=1}^{n}\sup_{\{\theta:|\theta-\theta_j|<\delta_j\}}\left|(T_i(\theta,Y_i)-E\left[T_i(\theta,Y_i)\right]-T_i(\theta_j,Y_i)E\left[T_i(\theta_j,Y_i)\right])\right| \\
+ & \left|\frac{1}{n}\sum_{i=1}^{n}(T_i(\theta_j,Y_i)-E\left[T_i(\theta_j,Y_i)\right])\right| \\
\leq & \frac{1}{n}\sum_{i=1}^{n}Z_{ij}+\left|\frac{1}{n}\sum_{i=1}^{n}(T_i(\theta_j,Y_i)-E\left[T_i(\theta_j,Y_i)\right])\right| \\
\leq & \frac{1}{n}\sum_{i=1}^{n}E\left[Z_{ij}\right]+\epsilon+\left|\frac{1}{n}\sum_{i=1}^{n}(T_i(\theta_j,Y_i)-E\left[T_i(\theta_j,Y_i)\right])\right|,\ \forall n\geq n_0 \\
\leq & \sup_{m\geq 1}\frac{1}{m}\sum_{i=1}^{m}E\left[Z_{ij}\right]+\epsilon+\left|\frac{1}{n}\sum_{i=1}^{n}(T_i(\theta_j,Y_i)-E\left[T_i(\theta_j,Y_i)\right])\right|,\ \forall n\geq n_0 \\
\leq & \epsilon+\epsilon+\epsilon=3\epsilon,\ \forall\ n\geq n_0
\end{aligned}
$$

□

Proof of Lemma 2.4. We will verify that conditions of Proposition A.1 are satisfied. Take $\theta = (\alpha,\beta),\ T_i(\theta,Y_i) = log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right)$. Firstly, condition (i) of Proposition A.1 has already been assumed and hence by compactness, $\exists\ R>0,\ such\ that\ \forall\ (\alpha,\beta)\in\Theta, |\alpha-\alpha_0|\leq R\ and |\beta-\beta_0|\leq R$

From part (d) of Lemma 2.1, we have $|T_i(\theta,Y_i)| \leq\leq Max(\tau,1-\tau)(1+M_1)*R$ which means that condition (ii) of Proposition A.1 holds.

To see condition (iii) of Proposition A.1 holds, we again use part (d) of Lemma 2.1 and observe that for $\theta_0 = (\alpha^{'}, \beta^{'})$

$$\begin{aligned}
& \sup_{n\geq 1} E\left[\frac{1}{n}\sum_{i=1}^{n} sup_{\{\theta:|\theta-\theta_0|<\delta\}}|T_i(\theta, Y_i) - E\left[T_i(\theta, Y_i)\right] - T_i(\theta_0, Y_i) + E\left[T_i(\theta_0, Y_i)\right]|\right] \\
= & \sup_{n\geq 1} E\left[\frac{1}{n}\sum_{i=1}^{n} sup_{\{(\alpha,\beta):|(\alpha',\beta')-(\alpha',\beta')|<\delta\}}\left|log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha',\beta')i}(Y_i)}\right) - E\left[log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha',\beta')i}(Y_i)}\right)\right]\right|\right] \\
\leq & \sup_{n\geq 1} \frac{1}{n}\sum_{i=1}^{n} sup_{\{(\alpha,\beta):|(\alpha,\beta)-(\alpha',\beta')|<\delta\}} Max(\tau, 1-\tau)\left[|\alpha-\alpha^{'}| + |\beta-\beta^{'}|M_1\right] \\
\leq & Max(\tau, 1-\tau)(1+M_1)\delta
\end{aligned}$$

Clearly, R.H.S of the last inequality goes to zero as δ goes to zero. □

Proof of Lemma 2.5. We will prove the result for the set W_1. The argument is similar for W_j for j=2,...,8. Suppose $(\alpha, \beta) \in W_1$. For any $X_i > \epsilon_0$, we have $(\alpha - \alpha_0) + (\beta - \beta_0)X_i > \Delta_1$. Using part (a) of Lemma 2.3, we have

$$\begin{aligned}
& E\left[log\left(\frac{f_{(\alpha,\beta)\ i}}{f_{(\alpha_0,\beta_0)\ i}}(Y_i)\right)\right] \\
= & E\left[(Y_i - \alpha - \beta X_i)I_{\alpha_0+\beta_0 X_i<Y_i<\alpha+\beta X_i}\right] \\
\leq & E\left[(Y_i - \alpha - \beta X_i)I_{\alpha_0+\beta_0 X_i<Y_i<\alpha+\beta X_i} \times I_{X_i>\epsilon_0}\right] \\
\leq & E\left[(\alpha_0 + \beta_0 X_i + \frac{\Delta_1}{2} - \alpha - \beta X_i)I_{\alpha_0+\beta_0 X_i<Y_i<\alpha_0+\beta_0 X_i+\frac{\Delta_1}{2}} \times I_{X_i>\epsilon_0}\right] \\
\leq & E\left[\left(\frac{\Delta_1}{2} - \Delta_1\right) I_{\alpha_0+\beta_0 X_i<Y_i<\alpha_0+\beta_0 X_i+\frac{\Delta_1}{2}} \times I_{X_i>\epsilon_0}\right] \\
\leq & -\frac{\Delta_1}{2} P\left(0 < Y_i - \alpha_0 - \beta_0 X_i < \frac{\Delta_1}{2}\right) \times I_{X_i>\epsilon_0}
\end{aligned}$$

This implies that

$$\limsup_{n\to\infty}\left(\int_{W_1} e^{\sum_{i=1}^n E\left[log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right)\right]} d\Pi(\alpha,\beta)\right)^{\frac{1}{n}} \leq e^{-\frac{\Delta_1}{2}\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n P\left(0<Y_i-\alpha_0-\beta_0 X_i<\frac{\Delta_1}{2}\right) I_{X_i>\epsilon_0}}$$

By assumption A.3, the R.H.S is well defined.

If we choose $K_1 = \frac{\Delta_1}{2} * \liminf \frac{1}{n}\sum_{i=1}^n P\left(0 < Y_i - \alpha_0 - \beta_0 X_i < \frac{\Delta_1}{2}\right) I_{X_i>\epsilon_0}$ then $\exists\ N_1^*$ such that $\forall\ n \geq N_1^*$, $\int_{W_1} e^{\sum_{i=1}^n E\left[log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right)\right]} d\Pi(\alpha,\beta) \leq e^{-nK_1}\Pi(W_1) \leq e^{-nK_1}$ (Note that the last step uses assumption **A.4** that Π is proper) □

Proof of Lemma 2.6. Again, we will prove the result for the set W_1. The argument is similar for other sets W_j for j=2,...,8. Recall that $W_1 = \{(\alpha,\beta) : \alpha - \alpha_0 \geq \Delta_1, \beta \geq \beta_0\}$. Let ϵ_0 be as in assumption **A.3** and $Z_i = Y_i - \alpha_0 - \beta_0 X_i$. Define

$$C_0 = \frac{2 \times \limsup_{m\to\infty} \frac{1}{m}\sum_{i=1}^m E[|Z_i|]}{(1-\tau)\liminf_{m\to\infty}\frac{1}{m}\sum_{i=1}^m I_{X_i>\epsilon_0}}$$

Note that assumption $A.3$ in particular implies that the denominator is well defined and assumption $A.5$ ensures that the numerator is well defined. Now let $A = B \times \epsilon_0 = 2C_0$ and define

$$\begin{aligned} G_1 &= \{(\alpha,\beta) \in W_1 : \alpha - \alpha_0 \leq A \ and \ \beta - \beta_0 \leq B\} \\ &= \{(\alpha,\beta) : (\alpha-\alpha_0, \beta-\beta_0) \in [\Delta_1, A] \times [0, B]\} \end{aligned}$$

Clearly G_1 is compact. Now if $(\alpha,\beta) \in G_1^c \cap W_1$ then either $(\alpha - \alpha_0) > A$ or $(\beta - \beta_0) > B$. Further if $X_i > \epsilon_0$ then in the former case we have $b_i = (\alpha-\alpha_0) + (\beta-\beta_0)X_i > A$ and in the latter case we would have $b_i > B \times \epsilon_0$. So, in either case when $X_i > \epsilon_0$, we have $b_i > 2C_0$. We can write

$$\sum_{i=1}^n log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right) = \sum_{i=1}^n log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right) I_{X_i>\epsilon_0} + \sum_{i=1}^n log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right) I_{X_i\leq\epsilon_0}$$

Now, applying part(e) of Lemma 2.1 to the first term in R.H.S and part (d) to the second term (for $(\alpha,\beta) \in G_1^c \cap W_1$) we have,

$$\begin{aligned} \sum_{i=1}^n log\left(\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}\right) &\leq -2C_0(1-\tau)\sum_{i=1}^n I_{X_i>\epsilon_0} + \sum_{i=1}^n Z_i^+ I_{X_i>\epsilon_0} + \sum_{i=1}^n |Z_i| I_{X_i\leq\epsilon_0} \\ &\leq -2C_0(1-\tau)\sum_{i=1}^n I_{X_i>\epsilon_0} + \sum_{i=1}^n |Z_i| \\ &\leq -2nC_0(1-\tau)\liminf_{m\to\infty}\frac{1}{m}\sum_{i=1}^m I_{X_i>\epsilon_0} + 2n\limsup_{m\to\infty}\frac{1}{m}\sum_{i=1}^m E\left[|Z_i|\right], \forall\ n \geq N_1(\omega) \\ &= -nC_0(1-\tau)\liminf_{m\to\infty}\frac{1}{m}\sum_{i=1}^m I_{X_i>\epsilon_0}, \forall\ n \geq N_1^{**}(\omega) \end{aligned}$$

The last but one nequality follows by using assumption **A.5**, which allows the application of SLLN on the sequence $\{|Z_n|\}$. Now, if we take $b_1 = C_0(1-\tau)\liminf_{m\to\infty}\frac{1}{m}\sum_{i=1}^m I_{X_i>\epsilon_0}$, it is easy to see that $\forall\ n \geq N_1^{**}(\omega)$, $\int_{G_j^c \cap W_j} e^{\sum_{i=1}^n log\frac{f_{(\alpha,\beta)i}(Y_i)}{f_{(\alpha_0,\beta_0)i}(Y_i)}} d\Pi(\alpha,\beta) \leq e^{-nb_1}\Pi(W_1 \cap G_1^c) \leq e^{-nb_1}$. (Note that the last step uses assumption **A.4** that Π is proper). ☐

REFERENCES

Amewou-Atisso, M., Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. V. (2003). Posterior consistency for semi-parametric regression problems. *Bernoulli*, **9** (2) 291312.

Berk, R.H.(1966), Limiting behaviour of posterior distributions when the model is incorrect.*Annals of Mathematical Statistics*, **37** 51-58.

Bunke, O. and Milhaud, X. (1998). Asymptotic behavior of bayes estimates under possibly incorrect models, *The Annals of Statistics* **26** 617–644.

Choi, T. and Ramamoorthi, R.V.(2003). Remarks on consistency of posterior distributions, *IMS Lecture Notes-Monograph Series*

Ghosh, J.K. and Ramamoorthi, R.V. (2003), Bayesian Nonparametrics, *Springer Verlag*

Kleijn,B.J.K. and Van der vaart, A.W.(2006), Misspecification in infinite dimensional bayesian statistics, *The Annals of Statistics*, **34** No.2, 837-877.

Koenker, R. and Bassett, G. Jr (1978), Regression quantiles, *Econometrica* **46** (1), 33-50.

Shalizi, C.R.(2009), Dynamics of bayesian updating with dependent data and misspecified models, *submitted to The Annals of Statistics*

Yue and Rue (2011), Bayesian inference for additive mixed quantile regression models, *Computational Statistics and Data Analysis* (available online)

Yu, K. and Moyeed, R.A(2001), Bayesian quantile regression, *Statistics and Probability Letters* **54** (4), 437-447