Discussion: Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood

James P. Hobert and Kshitij Khare
Department of Statistics
University of Florida

October 2015

We begin by congratulating Professors Yang, Wang & He (hereafter YW&H) for a top-notch contribution to the literature, and thanking Professor Hallin for giving us the opportunity to discuss their work. We found it quite satisfying to discover that the MCMC algorithms over which we have toiled can be used to make valid asymptotic inference! In this discussion, we focus on MCMC algorithms for exploring the posterior density defined in YW&H's equation (2.3) (when the prior on β is taken to be flat). Let $\{(y_i, x_i)\}_{i=1}^n$ represent the observed regression data. As usual, let X denote the $n \times p$ matrix whose ith row is x_i^T. Fix $\tau \in (0, 1)$, let $\mathbb{R}_+ = (0, \infty)$, and define a density $\pi : \mathbb{R}^p \to \mathbb{R}_+$ as follows

$$\pi(\beta) \propto \exp\left\{ -\frac{1}{\sigma} \sum_{i=1}^n \rho_\tau(y_i - x_i^T \beta) \right\}, \tag{1}$$

where $\rho_\tau(u) = u\big[\tau - I(u < 0)\big]$, and $\sigma > 0$ is a fixed scale parameter. We assume throughout that X has full column rank since this is a necessary and sufficient condition for $\int_{\mathbb{R}^p} \pi(\beta)\, d\beta < \infty$ (Choi and Hobert, 2013, Section 4).

There is a simple data augmentation (DA) algorithm for exploring $\pi(\beta)$. It is a modification of an algorithm developed in Kozumi and Kobayashi (2011) (who considered only proper priors for β). For the special case in which $\tau = 1/2$, Khare and Hobert (2011) (hereafter K&H) constructed an alternative *sandwich* algorithm, and showed that this algorithm is theoretically superior to the DA algorithm. Our main contribution in this discussion is to provide an extension of K&H's sandwich algorithm that can be used for any $\tau \in (0, 1)$, and to establish that it converges at least as fast as the DA algorithm. We begin by describing the DA algorithm.

As in YW&H, let $\theta_1 = \theta_1(\tau) = \frac{1-2\tau}{\tau(1-\tau)}$ and $\theta_2^2 = \theta_2^2(\tau) = \frac{2}{\tau(1-\tau)}$. For fixed $z = (z_1, \ldots, z_n)^T \in \mathbb{R}_+^n$, let D denote an $n \times n$ diagonal matrix whose ith diagonal element is $(\sigma\, \theta_2^2\, z_i)^{-1}$. Also, let $\Sigma = \Sigma(z) = (X^T D X)^{-1}$ and $\mu = \mu(z) = (X^T D X)^{-1} X^T D (y - \theta_1 z)$, where $y = (y_1, \ldots, y_n)^T$ denotes the observed responses. The algorithm requires draws from the generalized inverse Gaussian density, which we now formally define. We say $W \sim \text{GIG}(q, a, b)$ if its density is proportional to $w^{q-1} e^{-\frac{1}{2}(aw + \frac{b}{w})} I_{\mathbb{R}_+}(w)$, where $q, a > 0$ and $b \geq 0$. (Note that if $b = 0$, this is just a gamma density.) The basic DA algorithm is based on the Markov chain $\Phi = \{\beta_m\}_{m=0}^\infty$ whose state space is $\mathsf{X} = \mathbb{R}^p$, and whose dynamics are defined (implicitly) through the following two-step procedure for moving from the current state, $\beta_m = \beta$, to β_{m+1}.

Iteration $m+1$ of the DA algorithm:

1. Draw $Z_1, \ldots, Z_n$ independently with
$$Z_i \sim \text{GIG}\bigg(\frac{1}{2}, \frac{\theta_1^2}{\sigma\theta_2^2} + \frac{2}{\sigma}, \frac{(y_i - x_i^T\beta)^2}{\sigma\theta_2^2}\bigg) \ ,$$
and call the observed value $z = (z_1, \ldots, z_n)^T$

2. Draw $\beta_{m+1} \sim \text{N}_p\big(\mu(z), \Sigma(z)\big)$

We now briefly describe the latent data upon which this DA algorithm is based. Define a joint density for β and z as follows
$$\pi(\beta, z) \propto \prod_{i=1}^n \frac{1}{\sqrt{z_i}} \exp\Big\{ -\frac{1}{2\sigma\theta_2^2 z_i}\big(y_i - x_i^T\beta - \theta_1 z_i\big)^2 - \frac{z_i}{\sigma}\Big\} I_{\mathbb{R}_+}(z_i) \ .$$
The normal/exponential mixture representation of the asymmetric Laplace density (see equation (2.5) in YW&H) shows that $\int_{\mathbb{R}_+^n} \pi(\beta, z)\, dz \ \propto \ \pi(\beta)$, so the target density is the marginal of $\pi(\beta, z)$. The DA algorithm simply iterates between draws from $\pi(z|\beta)$ and $\pi(\beta|z)$ in the usual way.

K&H developed an alternative to the DA algorithm for the special case in which $\tau = 1/2$. (K&H actually took $\sigma = 1$, but the extension to arbitrary, fixed σ is trivial.) We now present an extension of K&H's sandwich algorithm that can be used for any $\tau \in (0,1)$. Define the matrix $M = D - DX(X^TDX)^{-1}X^TD$, which is clearly positive semi-definite. The algorithm is based on the Markov chain $\Phi^* = \{\beta_m^*\}_{m=0}^\infty$ whose state space is $\mathsf{X} = \mathbb{R}^p$, and whose dynamics are defined (implicitly) through the following three-step procedure for moving from the current state, $\beta_m^* = \beta$, to β_{m+1}^*.

Iteration $m+1$ of the sandwich algorithm:

1. Draw $Z_1, \ldots, Z_n$ independently with
$$Z_i \sim \text{GIG}\bigg(\frac{1}{2}, \frac{\theta_1^2}{\sigma\theta_2^2} + \frac{2}{\sigma}, \frac{(y_i - x_i^T\beta)^2}{\sigma\theta_2^2}\bigg) \ ,$$
and call the observed value $z = (z_1, \ldots, z_n)^T$

2. Draw
$$g \sim \text{GIG}\bigg(\frac{n+p}{2}, \frac{2\sum_{i=1}^n z_i}{\sigma} + \theta_1^2 z^T M z, y^T M y\bigg) \ ,$$
and set $z' = (gz_1, \ldots, gz_n)^T$

3. Draw $\beta_{m+1}^* \sim \text{N}_p\big(\mu(z'), \Sigma(z')\big)$

This sandwich algorithm is a direct generalization of the sandwich algorithm derived at the end of Section 5 in K&H. Moreover, it can be derived using the same arguments that appear there, along with the fact that the marginal density of z (for general $\tau \in (0,1)$) is given by

$$\pi(z) = \int_{\mathbb{R}^p} \pi(\beta, z)\, d\beta \propto \frac{1}{|X^T D X|^{\frac{1}{2}} \sqrt{\prod_{i=1}^n z_i}} \exp\left\{ -\frac{1}{2}(y - \theta_1 z)^T M (y - \theta_1 z) - \frac{\sum_{i=1}^n z_i}{\sigma} \right\} \prod_{i=1}^n I_{\mathbb{R}_+}(z_i)\ .$$

(We note that there is a typo in K&H: In the third line of page 2599, the term $g^{\frac{m-p-2}{2}}$ should, in fact, be $g^{\frac{m+p-2}{2}}$.)

The amount of computer time required to perform one iteration of the sandwich algorithm is only slightly larger than that required for the DA algorithm. Indeed, the difference is just a single generalized inverse Gaussian draw. To be fair, a small amount of matrix multiplication has to be performed before this extra GIG draw can be made, but some of the resulting quantities are required at the third step anyway. In any case, from a computational standpoint, the two algorithms are quite similar, so it is reasonable to compare their theoretical properties.

Let K and K^* denote the Markov operators corresponding to the DA and sandwich algorithms, respectively. These operators are both self-adjoint and positive. (See K&H for background and definitions.) K&H proved that, in the special case where $\tau = 1/2$, K and K^* are both *trace class* operators, which implies that the corresponding Markov chains are both geometrically ergodic. K&H also established that (when $\tau = 1/2$) the sandwich algorithm is better in the sense that, for each $i \in \{1, 2, 3, \dots\}$, the ith largest eigenvalue of the sandwich operator is less than or equal to the corresponding eigenvalue of the DA operator, with strict inequality for at least one i. (In general, self-adjoint, positive, trace class MCMC operators have pure eigenvalue spectra, the eigenvalues are all in $[0, 1)$, and smaller eigenvalues correspond to faster convergence.) Whether K&H's trace class results can be extended to the case where $\tau \neq 1/2$ is currently an open problem. Here we will establish something slightly weaker. Let P denote a generic self-adjoint Markov operator, and let $\|P\|$ denote its operator norm (which lives in $[0, 1]$). The smaller the norm, the faster the underlying Markov chain converges. Moreover, the chain is geometrically ergodic if and only if $\|P\| < 1$. Here is our result.

Proposition 1. *For any fixed $\tau \in (0, 1)$, the following inequalities hold: $\|K^*\| \leq \|K\| < 1$. Hence, the Markov chains underlying the sandwich and DA algorithms are both geometrically ergodic, and the former converges at least as fast as the latter.*

Proof. The inequality $\|K^*\| \leq \|K\|$ follows from the general theory of sandwich algorithms (see, e.g., K&H), so it suffices to prove that the DA algorithm is geometrically ergodic. Our proof will be indirect in the sense that we do not actually analyze the DA Markov chain, $\Phi = \{\beta_m\}_{m=0}^\infty$. Instead, we analyze the so-called conjugate chain, $\Delta = \{z_m\}_{m=0}^\infty$, which lives on $\mathbb{R}_+^n$, and has Markov transition density given by

$$k(z'|z) = \int_{\mathbb{R}^p} \pi(z'|\beta)\, \pi(\beta|z)\, d\beta\ . \tag{2}$$

Well-known results concerning two-block Gibbs samplers (see, e.g., Liu, Wong and Kong, 1994) imply that either both Φ and Δ are geometrically ergodic or neither of them is. To prove that Δ is geometrically ergodic, we will establish a geometric drift condition with a drift (Lyapunov) function that is unbounded off compact sets. (For a basic introduction to this technique, see Hobert (2011, Section 10.2.3).) Our drift function $V : \mathbb{R}^n_+ \to \mathbb{R}_+$ is given by

$$V(z) = \sum_{i=1}^n z_i + \sum_{i=1}^n z_i^{-1/4} \ .$$

We will prove that, for every $\lambda > \frac{|\theta_1|}{\sqrt{2\theta_2^2+\theta_1^2}}$, there exists a constant $K_\lambda > 0$ such that the following drift condition is satisfied

$$\int_{\mathbb{R}_n^+} V(z')\, k(z'|z)\, dz' \le \lambda V(z) + K_\lambda \ . \tag{3}$$

Geometric ergodicity follows since $\frac{|\theta_1|}{\sqrt{2\theta_2^2+\theta_1^2}} \in [0,1)$.

First, using (2) and Fubini's theorem, the left-hand side of (3) can be written as

$$\sum_{i=1}^n \int_{\mathbb{R}^p} \left[\int_{\mathbb{R}_n^+} z_i'\, \pi(z'|\beta)\, dz' \right] \pi(\beta|z)\, d\beta + \sum_{i=1}^n \int_{\mathbb{R}^p} \left[\int_{\mathbb{R}_n^+} (z_i')^{-1/4}\, \pi(z'|\beta)\, dz' \right] \pi(\beta|z)\, d\beta \ . \tag{4}$$

We begin by constructing an upper bound for the first term in (4). Using the formula for the mean of a generalized inverse Gaussian we have

$$\int_{\mathbb{R}_n^+} z_i'\, \pi(z'|\beta)\, dz = \frac{|y_i - x_i^T\beta|}{\sqrt{2\theta_2^2+\theta_1^2}} + \frac{\sigma\theta_2^2}{2\theta_2^2+\theta_1^2} \le \frac{|y_i|}{\sqrt{2\theta_2^2+\theta_1^2}} + \frac{|x_i^T\beta|}{\sqrt{2\theta_2^2+\theta_1^2}} + \frac{\sigma\theta_2^2}{2\theta_2^2+\theta_1^2} \ .$$

(Note that this formula still holds when $y_i - x_i^T\beta = 0$.) Hence,

$$\sum_{i=1}^n \int_{\mathbb{R}^p} \left[\int_{\mathbb{R}_n^+} z_i'\, \pi(z'|\beta)\, dz' \right] \pi(\beta|z)\, d\beta \le C_1 + \sum_{i=1}^n \int_{\mathbb{R}^p} \frac{|x_i^T\beta|}{\sqrt{2\theta_2^2+\theta_1^2}} \pi(\beta|z)\, d\beta \ ,$$

where, throughout the proof, $C_1, C_2, C_3, \dots$ represent constants. (This means constant in β and z.) Now define $\tilde{\mu} = \tilde{\mu}(z) = (X^TDX)^{-1}X^TDy$, $\mu^* = \mu^*(z) = -\frac{\theta_1}{\sigma\theta_2^2}(X^TDX)^{-1}X^Tl$, where l is an $n \times 1$ vector of ones, and note that $\mu = \tilde{\mu} + \mu^*$. (When $\tau = 1/2$, $\theta_1 = 0$ and $\mu = \tilde{\mu}$.) Now

$$|x_i^T\beta| = |x_i^T(\beta - \mu + \mu)| \le |x_i^T(\beta-\mu)| + |x_i^T\mu| \le |x_i^T(\beta-\mu)| + |x_i^T\tilde{\mu}| + |x_i^T\mu^*| \ ,$$

and it follows that

$$\sum_{i=1}^n \int_{\mathbb{R}^p} \left[\int_{\mathbb{R}_n^+} z_i'\, \pi(z'|\beta)\, dz' \right] \pi(\beta|z)\, d\beta \le$$
$$C_1 + \frac{1}{\sqrt{2\theta_2^2+\theta_1^2}} \left[\sum_{i=1}^n |x_i^T\tilde{\mu}| + \sum_{i=1}^n |x_i^T\mu^*| + \sum_{i=1}^n \int_{\mathbb{R}^p} |x_i^T(\beta-\mu)| \pi(\beta|z)\, d\beta \right] . \tag{5}$$

K&H show that $\|\tilde{\mu}\|$ is bounded above by a constant. (We are using $\|\cdot\|$ to denote the standard euclidean norm.) Hence, by Cauchy-Schwarz,

$$\sum_{i=1}^n |x_i^T\tilde{\mu}| \le \sum_{i=1}^n \|x_i^T\| \|\tilde{\mu}\| \le C_2 \ . \tag{6}$$

Let s be an $n \times 1$ vector such that $s_i = \text{sign}\{x_i^T \mu^*\}$ for $i = 1, 2 \ldots, n$. It follows that

$$\sum_{i=1}^n |x_i^T \mu^*| = \sum_{i=1}^n s_i x_i^T \mu^* = s^T X \mu^* = -\frac{\theta_1}{\sigma\theta_2^2} s^T X (X^T D X)^{-1} X^T l \le \frac{|\theta_1|}{\sigma\theta_2^2} |s^T X (X^T D X)^{-1} X^T l| \; . \quad (7)$$

Another application of Cauchy-Schwarz yields

$$|s^T X (X^T D X)^{-1} X^T l| \le \sqrt{s^T X (X^T D X)^{-1} X^T s} \sqrt{l^T X (X^T D X)^{-1} X^T l} \; . \quad (8)$$

Now combining (7) and (8), and using the fact that $D^{-1} - X(X^T D X)^{-1} X^T$ is a positive semi-definite matrix, we have

$$\sum_{i=1}^n |x_i^T \mu^*| \le \frac{|\theta_1|}{\sigma\theta_2^2} \sqrt{s^T D^{-1} s} \sqrt{l^T D^{-1} l} = \frac{|\theta_1|}{\sigma\theta_2^2} \sqrt{\sigma\theta_2^2 \sum_{i=1}^n s_i^2 z_i} \sqrt{\sigma\theta_2^2 \sum_{i=1}^n l_i^2 z_i} = |\theta_1| \sum_{i=1}^n z_i \; . \quad (9)$$

Now let $\alpha > 0$ be arbitrarily chosen. Using the fact that $|a| \le (\alpha a^2 + \alpha^{-1})/2$ for all real a, we have

$$\begin{aligned} \sum_{i=1}^n \int_{\mathbb{R}^p} |x_i^T (\beta - \mu)| \pi(\beta|z) \, d\beta &\le \frac{n}{2\alpha} + \frac{\alpha}{2} \sum_{i=1}^n \int_{\mathbb{R}^p} (\beta - \mu)^T x_i x_i^T (\beta - \mu) \, \pi(\beta|z) \, d\beta \\ &= \frac{n}{2\alpha} + \frac{\alpha}{2} \int_{\mathbb{R}^p} (\beta - \mu)^T X^T X (\beta - \mu) \, \pi(\beta|z) \, d\beta \\ &= \frac{n}{2\alpha} + \frac{\alpha}{2} \text{tr}\big(X (X^T D X)^{-1} X^T\big) \\ &\le \frac{n}{2\alpha} + \frac{\alpha}{2} \text{tr}(D^{-1}) \\ &= \frac{n}{2\alpha} + \frac{\alpha\sigma\theta_2^2}{2} \sum_{i=1}^n z_i \; , \end{aligned} \quad (10)$$

where the second equality follows from the fact that $\pi(\beta|z)$ is a p-variate normal with mean μ and covariance matrix $\Sigma = (X^T D X)^{-1}$. Combining (5), (6), (9) and (10) shows that

$$\sum_{i=1}^n \int_{\mathbb{R}^p} \left[\int_{\mathbb{R}_n^+} z_i' \, \pi(z'|\beta) \, dz' \right] \pi(\beta|z) \, d\beta \le C_3 + \left(\frac{|\theta_1|}{\sqrt{2\theta_2^2 + \theta_1^2}} + \frac{\alpha\sigma\theta_2^2}{2\sqrt{2\theta_2^2 + \theta_1^2}} \right) \sum_{i=1}^n z_i \; . \quad (11)$$

We now build an upper bound for the second term in (4). We begin by bounding $\int_{\mathbb{R}_n^+} (z_i')^{-1/4} \, \pi(z'|\beta) \, dz'$. First, if $y_i - x_i^T \beta = 0$, then the GIG becomes a gamma, and a simple calculation shows that

$$\int_{\mathbb{R}_n^+} (z_i')^{-1/4} \, \pi(z'|\beta) \, dz' = \frac{\Gamma\left(\frac{1}{4}\right)}{\Gamma\left(\frac{1}{2}\right)} \left(\frac{2\theta_2^2 + \theta_1^2}{\sigma\theta_2^2} \right)^{1/4} \; , \quad (12)$$

which is free of β. To do the calculation when $y_i - x_i^T \beta \ne 0$, we need the normalizing constant for the GIG density. For $q, a, b > 0$, $\int_0^\infty w^{q-1} e^{-\frac{1}{2}(aw + \frac{b}{w})} \, dw = 2(b/a)^{q/2} K_q(\sqrt{ab})$, where $K_\nu(\cdot)$ is the modified Bessel function of the second kind. This fact can be used to show that, if $y_i - x_i^T \beta \ne 0$, then

$$\int_{\mathbb{R}_n^+} (z_i')^{-1/4} \, \pi(z'|\beta) \, dz' = \left(\frac{\sqrt{2\theta_2^2 + \theta_1^2}}{|y_i - x_i^T \beta|} \right)^{1/4} \frac{K_{\frac{1}{4}}\left(\frac{\sqrt{2\tau^2 + \theta^2} \, |z_i - x_i^T \beta|}{\sigma\theta_2^2} \right)}{K_{\frac{1}{2}}\left(\frac{\sqrt{2\tau^2 + \theta^2} \, |z_i - x_i^T \beta|}{\sigma\theta_2^2} \right)} \le \left(\frac{\sqrt{2\theta_2^2 + \theta_1^2}}{|y_i - x_i^T \beta|} \right)^{1/4} \; , \quad (13)$$

where the inequality follows from the fact that, for $u > 0$, $K_\nu(u)$ is an increasing function of u (see, e.g., Laforgia, 1991, page 266).

Define three sets as follows: $S_1 = \{i : x_i = 0 \text{ and } y_i = 0\}$, $S_2 = \{i : x_i = 0 \text{ and } y_i \neq 0\}$ and $S_3 = \{i : x_i \neq 0\}$. Note that these three sets form a partition of the set $\{1, 2, \ldots, n\}$. If $i \in S_1 \cup S_2$, then we can use (12) and (13) (with $x_i = 0$) to conclude that

$$\int_{\mathbb{R}^p} \int_{\mathbb{R}_n^+} (z_i')^{-1/4} \, \pi(z'|\beta) \, \pi(\beta|z) \, dz' \, d\beta \leq C_4 \; . \tag{14}$$

Suppose now that $i \in S_3$. In this case, the set $\{\beta \in \mathbb{R}^p : y_i - x_i^T \beta = 0\}$ has measure zero in $\mathbb{R}^p$. Hence,

$$\int_{\mathbb{R}^p} \int_{\mathbb{R}_n^+} (z_i')^{-1/4} \, \pi(z'|\beta) \, \pi(\beta|z) \, dz' \, d\beta \leq (2\theta_2^2 + \theta_1^2)^{1/8} \int_{\mathbb{R}^p} \frac{1}{|y_i - x_i^T \beta|^{1/4}} \, \pi(\beta|z) \, d\beta \; . \tag{15}$$

Furthermore, conditional on z, $y_i - x_i^T \beta$ has a normal distribution with a finite mean, and variance $\gamma_i^2 = x_i^T (X^T D X)^{-1} x_i$. Letting $z_{(1)}$ denote the minimum element of z, we have $X^T D X \preceq \frac{1}{\sigma \, \theta_2^2 \, z_{(1)}} X^T X$ (where $A \preceq B$ means that $B - A$ is positive semi-definite). It follows that $\sigma \, \theta_2^2 \, z_{(1)} (X^T X)^{-1} \preceq (X^T D X)^{-1}$, so that $\sigma \, \theta_2^2 \, z_{(1)} x_i^T (X^T X)^{-1} x_i \leq x_i^T (X^T D X)^{-1} x_i = \gamma_i^2$. Using this fact after an application of Pal and Khare's (2014) Proposition A1 leads to

$$\int_{\mathbb{R}^p} \frac{1}{|y_i - x_i^T \beta|^{1/4}} \pi(\beta|z) \, d\beta \leq \frac{\Gamma\left(\frac{3}{8}\right) 2^{3/8}}{\sqrt{2\pi} \, \gamma_i^{1/4}} \leq \frac{\Gamma\left(\frac{3}{8}\right) 2^{3/8}}{\sqrt{2\pi} (\sigma \, \theta_2^2 \, z_{(1)} x_i^T (X^T X)^{-1} x_i)^{1/8}} \; . \tag{16}$$

Combining (14), (15) and (16), we have

$$\sum_{i=1}^{n} \int_{\mathbb{R}^p} \int_{\mathbb{R}_n^+} (z_i')^{-1/4} \, \pi(z'|\beta) \, \pi(\beta|z) \, dz' \, d\beta \leq \kappa \, z_{(1)}^{-1/8} + C_5 \; , \tag{17}$$

where

$$\kappa = \frac{(2\theta_2^2 + \theta_1^2)^{1/8} \, \Gamma\left(\frac{3}{8}\right) 2^{3/8}}{\sqrt{2\pi} (\sigma \, \theta_2^2)^{1/8}} \sum_{i \in S_3} \frac{1}{(x_i^T (X^T X)^{-1} x_i)^{1/8}} \; .$$

Let $\eta = \frac{|\theta_1|}{\sqrt{2\theta_2^2 + \theta_1^2}} + \frac{\alpha \sigma \theta_2^2}{2\sqrt{2\theta_2^2 + \theta_1^2}}$ as in (11). The inequality $|ab| \leq (a^2 + b^2)/2$ leads to $z_{(1)}^{-1/8} \leq \frac{\eta}{\kappa} z_{(1)}^{-1/4} + \frac{\kappa}{4\eta}$, and it follows that

$$\kappa \, z_{(1)}^{-1/8} + C_5 \leq \eta \, z_{(1)}^{-1/4} + \frac{\kappa^2}{4\eta} + C_5 \leq \eta \sum_{i=1}^{n} z_i^{-1/4} + C_7 \; . \tag{18}$$

Finally, putting together (3) (4), (11), (17) and (18), we have

$$\int_{\mathbb{R}_n^+} V(z') \, k(z'|z) \, dz' \leq \eta \sum_{i=1}^{n} z_i + \eta \sum_{i=1}^{n} z_i^{-1/4} + C_8 = \eta V(z) + C_8 \; .$$

The result now follows since $\alpha > 0$ was arbitrary. □

We end our discussion by correcting a minor mistake in YW&H. In Section 2.2, YW&H overstate the scope of Choi and Hobert's (2013) Theorem 1, which concerns Markov chain convergence rates. In particular, YW&H suggest that the result holds for all $\tau = (0, 1)$, when in fact it holds only in the case where $\tau = 1/2$. On the other hand, YW&H's description of Choi and Hobert's (2013) Proposition 2, which concerns posterior propriety, is exactly right. That result does indeed hold for all $\tau = (0, 1)$.

Acknowledgment. Both authors were supported by NSF Grant DMS-15-11945.

References

CHOI, H. M. and HOBERT, J. P. (2013). Analysis of MCMC algorithms for Bayesian linear regression with Laplace errors. *Journal of Multivariate Analysis*, **117** 32–40.

HOBERT, J. P. (2011). The data augmentation algorithm: Theory and methodology. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. Jones and X.-L. Meng, eds.). Chapman & Hall/CRC Press.

KHARE, K. and HOBERT, J. P. (2011). A spectral analytic comparison of trace-class data augmentation algorithms and their sandwich variants. *The Annals of Statistics*, **39** 2585–2606.

KOZUMI, H. and KOBAYASHI, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, **81** 1565–1578.

LAFORGIA, A. (1991). Bounds for modified Bessel functions. *Journal of Computational and Applied Mathematics*, **34** 263–267.

LIU, J. S., WONG, W. H. and KONG, A. (1994). Covariance structure of the Gibbs sampler with applications to comparisons of estimators and augmentation schemes. *Biometrika*, **81** 27–40.

PAL, S. and KHARE, K. (2014). Geometric ergodicity for Bayesian shrinkage models. *Electronic Journal of Statistics*, **8** 604–645.