Electronic Journal of Statistics
Vol. 8 (2014) 2448–2478
ISSN: 1935-7524
DOI: 10.1214/14-EJS957

Markov chain Monte Carlo estimation of quantiles

Charles R. Doss

School of Statistics
University of Minnesota
Minneapolis, MN 55455
e-mail: cdoss@umn.edu

James M. Flegal*

Department of Statistics
University of California, Riverside
Riverside, CA 92521
e-mail: jflegal@ucr.edu

Galin L. Jones†

School of Statistics
University of Minnesota
Minneapolis, MN 55455
e-mail: galin@umn.edu

and

Ronald C. Neath

Department of Statistics
Columbia University
New York, NY 10027
e-mail: rcn2112@columbia.edu

Abstract: We consider quantile estimation using Markov chain Monte Carlo and establish conditions under which the sampling distribution of the Monte Carlo error is approximately Normal. Further, we investigate techniques to estimate the associated asymptotic variance, which enables construction of an asymptotically valid interval estimator. Finally, we explore the finite sample properties of these methods through examples and provide some recommendations to practitioners.

MSC 2010 subject classifications: Primary 60J22, 62M05.
Keywords and phrases: Markov chain, Monte Carlo, quantile estimation, central limit theorem, regeneration, batch means.

Received July 2013.

*Research supported by the National Science Foundation.
†Research supported by the National Institutes of Health and the National Science Foundation.

1. Introduction

Let π denote a probability distribution having support $\mathsf{X} \subseteq \mathbb{R}^d$, $d \geq 1$. If $W \sim \pi$ and $g : \mathsf{X} \to \mathbb{R}$ is measurable, set $V = g(W)$. We consider estimation of quantiles of the distribution of V. Specifically, if $0 < q < 1$ and F_V denotes the distribution function of V, then our goal is to obtain

$$\xi_q := F_V^{-1}(q) = \inf\{v : F_V(v) \geq q\}.$$

We will assume throughout that $F_V(x)$ is absolutely continuous and has continuous density function $f_V(x)$ such that $0 < f_V(\xi_q) < \infty$. Notice that this means ξ_q is the unique solution y of $F_V(y-) \leq q \leq F_V(y)$.

Typically, it is not possible to calculate ξ_q directly. For example, a common goal in Bayesian inference is calculating the quantile of a marginal posterior distribution. In these settings, the quantile estimate is typically based upon Markov chain Monte Carlo (MCMC) simulation methods and is almost always reported without including any notion of the simulation error. Raftery and Lewis (1992) consider quantile estimation using MCMC, but their method is based on approximating the MCMC process with a two-state Markov chain, and does not produce an estimate of the simulation error; see also Brooks and Roberts (1999) and Cowles and Carlin (1996) who study the properties of the method proposed by Raftery and Lewis (1992). In contrast, our work enables practitioners to rigorously assess the simulation error, and hence increase the reliability of their inferences.

The basic MCMC method entails simulating a Markov chain $X = \{X_0, X_1, \ldots\}$ having invariant distribution π. Define $Y = \{Y_0, Y_1, \ldots\} = \{g(X_0), g(X_1), \ldots\}$. If we observe a realization of X of length n and let $Y_{n(j)}$ denote the jth order statistic of $\{Y_0, \ldots, Y_{n-1}\}$, then we estimate ξ_q with

$$\hat{\xi}_{n,q} := Y_{n(j)} \quad \text{where} \quad j - 1 < nq \leq j. \tag{1}$$

We will see that $\hat{\xi}_{n,q}$ is strongly consistent for ξ_q. While this justifies the use of $\hat{\xi}_{n,q}$, it will be more valuable if we can also assess the unknown Monte Carlo error, $\hat{\xi}_{n,q} - \xi_q$. We address this in two ways. The first is by finding a function $b : \mathbb{N} \times (0, \infty) \to [0, \infty)$ such that for all $\epsilon > 0$

$$\Pr\left(|\hat{\xi}_{n,q} - \xi_q| > \epsilon\right) \leq b(n, \epsilon). \tag{2}$$

We also assess the Monte Carlo error through its approximate sampling distribution. We will show that under a mixing condition on X, a quantile central limit theorem (CLT) will obtain; this mixing condition is much weaker than the mixing conditions required for a CLT for a sample mean (Jones, 2004). For now, assume there exists a constant $\gamma^2(\xi_q) > 0$ such that, as $n \to \infty$,

$$\sqrt{n}(\hat{\xi}_{n,q} - \xi_q) \xrightarrow{d} \mathrm{N}(0, \gamma^2(\xi_q)). \tag{3}$$

Note that $\gamma^2(\xi_q)$ must account for the serial dependence present in a non-trivial Markov chain and hence is more difficult to estimate well than when X is a random sample. However, if we can estimate $\gamma^2(\xi_q)$ with, say $\hat{\gamma}_n^2$, then an interval estimator of ξ_q is

$$\hat{\xi}_{n,q} \pm t_* \frac{\hat{\gamma}_n}{\sqrt{n}}$$

where t_* is an appropriate quantile. Such intervals, or at least, the *Monte Carlo standard error* (MCSE), $\hat{\gamma}_n/\sqrt{n}$, are useful in assessing the reliability of the simulation results as they explicitly describe the level of confidence we have in the reported number of significant figures in $\hat{\xi}_{n,q}$. For more on this approach see Flegal and Gong (2014), Flegal et al. (2008), Flegal and Jones (2011), Geyer (2011), Jones et al. (2006) and Jones and Hobert (2001).

We consider three methods for implementing this recipe, all of which produce effective interval estimators of ξ_q. The first two are based on the CLT at (3) where we consider using the method of batch means (BM) and the subsampling bootstrap method (SBM) to estimate $\gamma^2(\xi_q)$. Regenerative simulation (RS) is the third method, but it requires a slightly different quantile CLT than that in (3). Along the way we show that significantly weaker conditions are available for the RS-based expectation estimation case previously studied in Hobert et al. (2002) and Mykland et al. (1995).

The remainder is organized as follows. We begin in Section 2 with a brief introduction to some required Markov chain theory. In Section 3 we consider estimation of ξ_q with $\hat{\xi}_{n,q}$, establish a CLT for the Monte Carlo error, and consider how to obtain MCSEs using BM and SBM. In Section 4, we consider RS, establish an alternative CLT and show how an MCSE can be obtained. In Section 5, we illustrate the use of the methods presented here and investigate their finite-sample properties in three examples. Finally, in Section 6 we summarize our results and conclude with some practical recommendations.

2. Markov chain background

In this section we give some essential preliminary material. Recall that π has support X and let $\mathcal{B}(\mathsf{X})$ be the Borel σ-algebra. For $n \in \mathbb{N} = \{1, 2, 3, \ldots\}$, let the n-step Markov kernel associated with X be $P^n(x, dy)$. Then if $A \in \mathcal{B}(\mathsf{X})$ and $k \in \{0, 1, 2, \ldots\}$, $P^n(x, A) = \Pr(X_{k+n} \in A | X_k = x)$. Throughout we assume X is Harris ergodic (π-irreducible, aperiodic, and positive Harris recurrent–see Meyn and Tweedie (2009) for definitions) and has invariant distribution π.

Let $\|\cdot\|$ denote the total variation norm. Further, let $M : \mathsf{X} \mapsto \mathbb{R}^+$ with $E_\pi M < \infty$ and $\psi : \mathbb{N} \mapsto \mathbb{R}^+$ be decreasing such that

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq M(x)\psi(n). \tag{4}$$

Polynomial ergodicity of order m where $m > 0$ means (4) holds with $\psi(n) = n^{-m}$. *Geometric ergodicity* means (4) holds with $\psi(n) = t^n$ for some $0 < t < 1$. *Uniform ergodicity* means that X is geometrically ergodic and M is bounded.

An equivalent characterization of uniform ergodicity is often more convenient for applications. The Markov chain X is uniformly ergodic if and only if there exists a probability measure ϕ on X, $\lambda > 0$, and an integer $n_0 \geq 1$ such that

$$P^{n_0}(x, \cdot) \geq \lambda\phi(\cdot) \text{ for each } x \in \mathsf{X}. \tag{5}$$

When (5) holds we have that (Meyn and Tweedie, 2009, p. 392)

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq (1-\lambda)^{\lfloor n/n_0 \rfloor}. \tag{6}$$

See Jones and Hobert (2001) for an accessible introduction to methods for establishing (5) and further discussion of the methods for establishing (4).

3. Quantile estimation for Markov chains

Recall $Y = \{Y_0, Y_1, \ldots\} = \{g(X_0), g(X_1), \ldots\}$ and set $F_n(y) = n^{-1}\sum_{i=0}^{n-1} I(Y_i \leq y)$. By the Markov chain version of the strong law of large numbers (see e.g. Meyn and Tweedie, 2009) for each y, $F_n(y) \to F_V(y)$ with probability 1 as $n \to \infty$. Using this, the proof of the following result is similar to the proof for when Y is composed of independent and identically distributed random variables (see e.g. Serfling, 1981) and hence is omitted.

Theorem 1. *With probability 1, $\hat{\xi}_{n,q} \to \xi_q$ as $n \to \infty$.*

While this result justifies the use of $\hat{\xi}_{n,q}$ as an estimator of ξ_q, it does not allow one to assess the unknown Monte Carlo error $\hat{\xi}_{n,q} - \xi_q$ for any finite n. In Section 3.1 we establish conditions under which (2) holds, while in Section 3.2 we examine the approximate sampling distribution of the Monte Carlo error.

3.1. Monte Carlo error under stationarity

We will consider (in this subsection only) a best-case scenario where $X_0 \sim \pi$, that is, the Markov chain X is stationary. We begin with a refinement of a result due to Wang et al. (2011) to obtain a useful description of how the Monte Carlo error decreases with simulation sample size and the convergence rate of the Markov chain. The proof is given in Appendix B.1.

Proposition 1. *Suppose the Markov chain X is polynomially ergodic of order $m > 1$. If $\delta \in (9/(10+8m), 1/2)$, then, with probability 1, for sufficiently large n, there is a positive constant C_0 such that $\hat{\xi}_{n,q} \in [\, \xi_q - C_0 n^{-1/2+\delta}\sqrt{\log\log n},\, \xi_q + C_0 n^{-1/2+\delta}\sqrt{\log\log n}\,]$.*

For the rest of this section we consider finite sample properties of the Monte Carlo error in the sense that our goal is to find an explicit function $b : \mathbb{N} \times (0,\infty) \to [0,\infty)$ such that (2) holds. There has been some research on this in the context of estimating expectations using MCMC (e.g. Łatuszyński and Niemiro, 2011; Łatuszyński et al., 2012; Rudolf, 2012), but this has not been considered in the quantile case. The proofs of the remaining results in this section can be found in Appendix B.2.

Theorem 2. *If X satisfies (4), then for any integer $a \in [1, n/2]$ and any $\epsilon > 0$ and $0 < \delta < 1$*

$$\Pr\left(|\hat{\xi}_{n,q} - \xi_q| > \epsilon\right) \leq 8 \exp\left\{-\frac{a\gamma^2}{8}\right\} + 22a\left(1 + \frac{4}{\gamma}\right)^{1/2} \psi\left(\left\lfloor \frac{n}{2a} \right\rfloor\right) E_\pi M,$$

where $\gamma = \gamma(\delta, \epsilon) = \min\{F_V(\xi_q + \epsilon) - q, \delta(q - F_V(\xi_q - \epsilon))\}$.

To be useful Theorem 2 requires bounding $\psi(n)E_\pi M$. There has been a substantial amount of work in this area (see e.g. Baxendale, 2005; Fort and Moulines, 2003; Rosenthal, 1995), but these methods have been applied in only a few practically relevant settings (see e.g. Jones and Hobert, 2001, 2004). However, in the uniformly ergodic case we have the following easy corollary.

Corollary 1. *If X satisfies (5), then we have for any $a \in [1, n/2]$, any $\epsilon > 0$ and any $0 < \delta < 1$*

$$\Pr\left(|\hat{\xi}_{n,q} - \xi_q| > \epsilon\right) \leq 8 \exp\left\{-\frac{a\gamma^2}{8}\right\} + 22a\left(1 + \frac{4}{\gamma}\right)^{1/2} (1 - \lambda)^{\lfloor n/2an_0 \rfloor},$$

where $\gamma = \gamma(\delta, \epsilon) = \min\{F_V(\xi_q + \epsilon) - q, \delta(q - F_V(\xi_q - \epsilon))\}$.

Example 1. Let

$$\pi(x, y) = \frac{4}{\sqrt{2\pi}} y^{3/2} \exp\left\{-y\left(\frac{x^2}{2} + 2\right)\right\} I(0 < y < \infty). \tag{7}$$

Then $Y|X = x \sim \text{Gamma}(5/2, 2 + x^2/2)$ and marginally $X \sim t(4)$–Student's t with 4 degrees of freedom. Consider a linchpin variable sampler (Acosta et al., 2014) which first updates X with a Metropolis-Hastings independence sampler having the marginal of X as the invariant distribution using a $t(3)$ proposal distribution, then updates Y with a draw from the conditional of $Y|X$. Letting P denote the Markov kernel for this algorithm we show in Appendix B.3 that for any measurable set A

$$P((x, y), A) \geq \frac{\sqrt{9375}}{32\pi} \int_A \pi(x', y')\, dx' dy'$$

and hence the Markov chain satisfies (5) with $n_0 = 1$ and $\lambda = \sqrt{9375}/32\pi$.

Set $\delta = .99999$, $a = n/16$ and consider estimating the median of the marginal of X, i.e. $t(4)$. Then $q = 1/2$ and $\xi_{1/2} = 0$ so that $\gamma = 0.037422$. Suppose we want to find the Monte Carlo sample size required to ensure that the probability $\hat{\xi}_{n,1/2}$ is within .10 of the truth is approximately 0.9. Then Corollary 1 gives

$$\Pr\left(|\hat{\xi}_{4\times 10^5,1/2} - \xi_{1/2}| > .1\right) \leq 0.101.$$

We can improve upon the conclusion of Corollary 1.

TABLE 1
Simulation length for each of 500 independent replications, counts of sample medians more than .1 away from 0 in absolute value and, $\hat{\Pr}(|\hat{\xi}_{n,1/2} - \xi_{1/2}| > .1)$

Length	500	1000	4700
Count	60	9	0
$\hat{\Pr}$	.12	.018	0

Theorem 3. *If X satisfies (5), then for every $\epsilon > 0$ and $0 < \delta < 1$*

$$\Pr\left(|\hat{\xi}_{n,q} - \xi_q| > \epsilon\right) \le 2\exp\left\{-\frac{\lambda^2(n\gamma - 2n_0/\lambda)^2}{2nn_0^2}\right\},$$

for $n > 2n_0/(\lambda\gamma)$ where $\gamma = \min\{F_V(\xi_q + \epsilon) - q, \delta(q - F_V(\xi_q - \epsilon))\}$.

Example 2 (Continuation of Example 1)**.** Theorem 3 yields that

$$\Pr\left(|\hat{\xi}_{4700,1/2} - \xi_{1/2}| > .1\right) \le 0.101 \tag{8}$$

which clearly shows that the bound given in Example 1 is conservative.

We now compare the bound in (8) to the results of a simulation experiment. We performed 500 independent replications of this MCMC sampler for each of 3 simulation lengths and recorded the number of estimated medians for each that were more than .1 in absolute value away from the median of a $t(4)$ distribution. The results are presented in Table 1 and Figure 1. The results in Table 1 show that the estimated probability in (8) is somewhat conservative. On the other hand, from Figure 1 it is clear that the estimation procedure is not all that stable until $n = 4700$.

3.2. Central limit theorem

We consider the asymptotic distribution of the Monte Carlo error $\hat{\xi}_{n,q} - \xi_q$. Let

$$\sigma^2(y) := \mathrm{Var}_\pi I(Y_0 \le y) + 2\sum_{k=1}^{\infty} \mathrm{Cov}_\pi\left[I(Y_0 \le y), I(Y_k \le y)\right]. \tag{9}$$

The proof of the following result is in Appendix B.4.

Theorem 4. *If X is polynomially ergodic of order $m > 1$ and if $\sigma^2(\xi_q) > 0$, then, as $n \to \infty$,*

$$\sqrt{n}(\hat{\xi}_{n,q} - \xi_q) \xrightarrow{d} \mathrm{N}(0, \sigma^2(\xi_q)/[f_V(\xi_q)]^2). \tag{10}$$

To obtain an MCSE we need to estimate $\gamma^2(\xi_q) := \sigma^2(\xi_q)/[f_V(\xi_q)]^2$. We consider two methods for doing this–in Section 3.2.1 we consider the method of batch means while in Section 3.2.2 we consider subsampling.

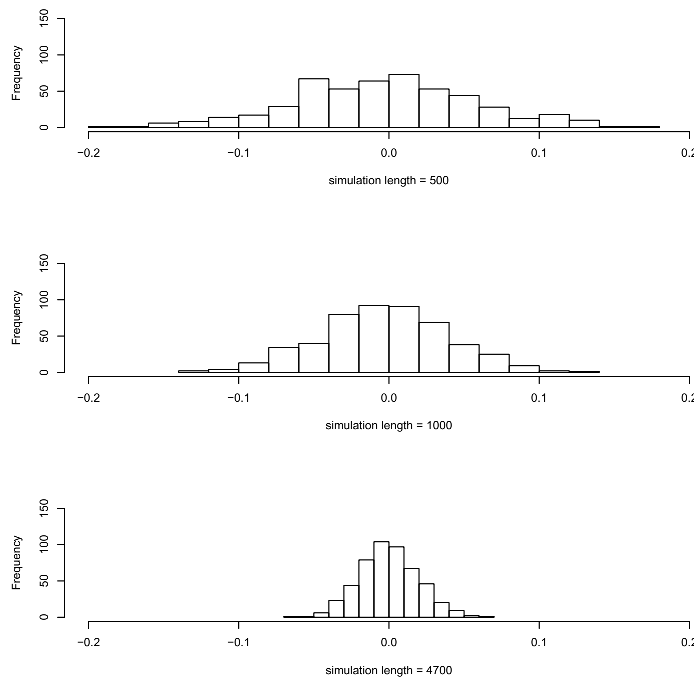

FIG 1. *Histograms of 500 sample medians for each of 3 simulation lengths.*

3.2.1. Batch Means

To estimate $\gamma^2(\xi_q)$, we substitute $\hat{\xi}_{n,q}$ for ξ_q and estimate $f_V(\hat{\xi}_{n,q})$ and $\sigma^2(\hat{\xi}_{n,q})$.

Consider estimating $f_V(\hat{\xi}_{n,q})$. Consistently estimating a density at a point has been studied extensively in the context of stationary time-series analysis (see e.g. Robinson, 1983) and many existing results are applicable since the Markov chains in MCMC are special cases of strong mixing processes. In our examples we use kernel density estimators with a Gaussian kernel to obtain $\hat{f}_V(\hat{\xi}_{n,q})$, an estimator of $f_V(\hat{\xi}_{n,q})$.

The quantity $\sigma^2(y)$, $y \in \mathbb{R}$ is familiar. Notice that

$$\sqrt{n}(F_n(y) - F_V(y)) \xrightarrow{d} \mathrm{N}(0, \sigma^2(y)) \quad \text{as } n \to \infty$$

by the usual Markov chain CLT for sample means (Jones, 2004). Moreover, we show in Corollary 4 that $\sigma^2(y)$ is continuous at ξ_q. In this context, estimating $\sigma^2(y)$ consistently is a well-studied problem and there are an array of methods for doing so; see Flegal et al. (2008), Flegal and Jones (2010), Flegal and Jones (2011) and Jones et al. (2006). Here we focus on the method of batch means for estimating $\sigma^2(\hat{\xi}_{n,q})$. For BM the output is split into batches of equal size. Suppose we obtain $n = a_n b_n$ iterations $\{X_0, \dots, X_{n-1}\}$ and for $k = 0, \dots, a_n - 1$ define $\bar{U}_k(\hat{\xi}_{n,q}) = b_n^{-1} \sum_{i=0}^{b_n - 1} I(Y_{kb_n+i} \le \hat{\xi}_{n,q})$. Then the BM estimator of

$\sigma^2(\hat{\xi}_{n,q})$ is

$$\hat{\sigma}^2_{BM}(\hat{\xi}_{n,q}) = \frac{b_n}{a_n - 1} \sum_{k=0}^{a_n-1} \left(\bar{U}_k(\hat{\xi}_{n,q}) - F_n(\hat{\xi}_{n,q}) \right)^2 . \tag{11}$$

Putting these two pieces together we estimate $\gamma^2(\xi_q)$ with

$$\hat{\gamma}^2(\hat{\xi}_{n,q}) := \frac{\hat{\sigma}^2_{BM}(\hat{\xi}_{n,q})}{[\hat{f}_V(\hat{\xi}_{n,q})]^2}$$

and we can obtain an approximate $100(1-\alpha)\%$ confidence interval for ξ_q by

$$\hat{\xi}_{n,q} \pm z_{\alpha/2} \frac{\hat{\gamma}(\hat{\xi}_{n,q})}{\sqrt{n}}, \tag{12}$$

where $z_{\alpha/2}$ is a standard Normal quantile.

3.2.2. Subsampling

It is natural to consider the utility of bootstrap methods for estimating quantiles and the Monte Carlo error. Indeed, there has been a substantial amount of work on using bootstrap methods for stationary time-series (e.g. Bertail and Clémençon, 2006; Bühlmann, 2002; Carlstein, 1986; Datta and McCormick, 1993; Politis, 2003). However, in our experience, MCMC simulations are typically sufficiently long so that standard bootstrap methods are prohibitively computationally expensive.

We focus on the subsampling bootstrap method (SBM) described in general by Politis et al. (1999) and, in the context of MCMC, by Flegal (2012) and Flegal and Jones (2011). The basic idea is to split X into $n-b+1$ overlapping blocks of length b. We then estimate ξ_q over each block resulting in $n-b+1$ estimates. Consider the ith subsample of Y, $\{Y_{i-1}, \dots, Y_{i+b-2}\}$. Define the corresponding ordered subsample as $\{Y^{i*}_{b(1)}, \dots, Y^{i*}_{b(b)}\}$ and quantile estimator as

$$\xi^*_i = Y^{i*}_{b(j)} \text{ where } j-1 < bq \le j \text{ for } i = 1, \dots, n-b+1.$$

If

$$\bar{\xi}^* = \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} \xi^*_i ,$$

then the SBM estimator of $\gamma^2(\xi_q)$ is given by

$$\hat{\gamma}^2_S = \frac{b}{n-b+1} \sum_{i=1}^{n-b+1} (\xi^*_i - \bar{\xi}^*)^2 .$$

Note that SBM avoids having to estimate the density $f_V(\hat{\xi}_{n,q})$. An approximate $100(1-\alpha)\%$ confidence interval for ξ_q is given by

$$\hat{\xi}_{n,q} \pm z_{\alpha/2} \frac{\hat{\gamma_S}(\hat{\xi}_{n,q})}{\sqrt{n}}, \tag{13}$$

where $z_{\alpha/2}$ is an appropriate standard Normal quantile.

4. Quantile estimation for regenerative Markov chains

Regenerative simulation (RS) provides an alternative estimation method for Markov chain simulations. RS is based on simulating an augmented Markov chain and Theorem 4 will not apply. We derive an alternative CLT based on RS and consider a natural estimator of the variance in the asymptotic Normal distribution.

Recall X has n-step Markov kernel $P^n(x, dy)$ and suppose there exists a function $s : \mathsf{X} \to [0, 1]$ with $E_\pi s > 0$ and a probability measure Q such that

$$P(x, A) \geq s(x)Q(A) \quad \text{for all } x \in \mathsf{X} \text{ and } A \in \mathcal{B}. \tag{14}$$

We call s the *small function* and Q the *small measure*. Define the *residual measure*

$$\mathsf{R}(x, dy) = \begin{cases} \dfrac{P(x, dy) - s(x)Q(dy)}{1 - s(x)} & s(x) < 1 \\ Q(dy) & s(x) = 1 \end{cases} \tag{15}$$

so that

$$P(x, dy) = s(x)Q(dy) + (1 - s(x))\mathsf{R}(x, dy). \tag{16}$$

We now have the ingredients for constructing the *split chain*,

$$X' = \{(X_0, \delta_0), (X_1, \delta_1), (X_2, \delta_2), \ldots\}$$

which lives on $\mathsf{X} \times \{0, 1\}$. Given $X_i = x$, then δ_i and X_{i+1} are found by

---

1. Simulate $\delta_i \sim \text{Bernoulli}(s(x))$
2. If $\delta_i = 1$, simulate $X_{i+1} \sim Q(\cdot)$; otherwise $X_{i+1} \sim \mathsf{R}(x, \cdot)$.

Two things are apparent from this construction. First, by (16) the marginal sequence $\{X_n\}$ has Markov transition kernel given by P. Second, the set of n for which $\delta_{n-1} = 1$, called *regeneration times*, represent times at which the chain probabilistically restarts itself in the sense that $X_n \sim Q(\cdot)$ does not depend on X_{n-1}.

The main practical impediment to the use of regenerative simulation would appear to be the means to simulate from the residual kernel $\mathsf{R}(\cdot, \cdot)$, defined at (15). Interestingly, as shown by Mykland et al. (1995), this is essentially a non-issue since there is an equivalent update rule for the split chain which does not depend on R. Given $X_k = x$, find X_{k+1} and δ_k by

---

1. Simulate $X_{k+1} \sim P(x, \cdot)$
2. Simulate $\delta_k \sim \text{Bernoulli}(r(X_k, X_{k+1}))$ where

$$r(x, y) = \frac{s(x)Q(dy)}{P(x, dy)}.$$

RS has received considerable attention in the case where either a Gibbs sampler or a full-dimensional Metropolis-Hastings sampler is employed. In particular, Mykland et al. (1995) give recipes for establishing minorization conditions as in (14), which have been implemented in several practically relevant statistical models; see e.g. Doss and Tan (2014); Gilks et al. (1998); Hobert et al. (2006); Jones et al. (2006); Jones and Hobert (2001); Roy and Hobert (2007).

Suppose we start X' with $X_0 \sim Q$; one can always discard the draws preceding the first regeneration to guarantee this, but it is frequently easy to draw directly from Q (Hobert et al., 2002; Mykland et al., 1995). We will write E_Q to denote expectation when the split chain is started with $X_0 \sim Q$. Let $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$ be the regeneration times so that $\tau_{t+1} = \min\{i > \tau_t : \delta_{i-1} = 1\}$. Assume X' is run for R tours so that the simulation is terminated the Rth time that a $\delta_i = 1$. Let τ_R be the total length of the simulation and $N_t = \tau_t - \tau_{t-1}$ be the length of the tth tour. Let $h : \mathsf{X} \to \mathbb{R}$, $V_i = h(X_i)$ and define

$$S_t = \sum_{i=\tau_{t-1}}^{\tau_t - 1} V_i \quad \text{for } t = 1, \dots, R.$$

The split chain construction ensures that the pairs (N_t, S_t) are independent and identically distributed. It is straightforward to show (Hobert et al., 2002; Meyn and Tweedie, 2009; Mykland et al., 1995) that if $E_Q N_t^2 < \infty$ and $E_Q S_t^2 < \infty$, then as $R \to \infty$,

$$\overline{h}_{\tau_R} = \frac{\sum_{t=1}^{R} S_t}{\sum_{t=1}^{R} N_t} = \frac{\overline{S}}{\overline{N}} \to E_\pi h \qquad \text{with probability 1} \tag{17}$$

and, if $\Gamma = E_Q[(S_1 - N_1 E_\pi h)^2]/[E_Q(N_1)]^2$, then

$$\sqrt{R}(\overline{h}_{\tau_R} - E_\pi h) \xrightarrow{d} \mathrm{N}(0, \Gamma). \tag{18}$$

Moreover, there is an easily calculated consistent estimator of Γ; see Hobert et al. (2002). However, the required moment conditions, $E_Q N_t^2 < \infty$ and $E_Q S_t^2 < \infty$, are difficult to check in practice. Hobert et al. (2002) showed that these moment conditions will hold if the Markov chain X is geometrically ergodic and there exists $\delta > 0$ such that $E_\pi |h|^{2+\delta} < \infty$. Our next result significantly weakens the required mixing conditions. The proof can be found in Appendix B.5.

Theorem 5. *If X is polynomially ergodic of order $m > 1$ and there exists $\delta > 2/(m-1)$ such that $E_\pi |h|^{2+\delta} < \infty$, then $E_Q N_t^2 < \infty$ and $E_Q S_t^2 < \infty$.*

Remark 1. If h is bounded, then $E_Q N_t^2 < \infty$ and $E_Q S_t^2 < \infty$ when X is polynomially ergodic of order $m > 1$

In the sequel we use Theorem 5 to develop an RS-based CLT for quantiles.

4.1. Quantile estimation

Recall $Y = \{Y_0, Y_1, \dots\} = \{g(X_0), g(X_1), \dots\}$ and define

$$S_t(y) = \sum_{i=\tau_{t-1}}^{\tau_t - 1} I(Y_i \le y) \quad \text{for } t = 1, \dots, R.$$

Note that $0 \le S_t(y) \le N_t$ for all $y \in \mathbb{R}$, and hence $E_Q(S_t(y))^2 \le E_Q(N_t)^2$. For each $y \in \mathbb{R}$ set

$$\Gamma(y) = E_Q\left[(S_1(y) - F_V(y)N_1)^2\right] / \left[E_Q(N_1)\right]^2,$$

which exists under the conditions of Theorem 5.

Let $j = \tau_R q + o(\sqrt{\tau_R})$ as $R \to \infty$ and consider estimating ξ_q with $Y_{\tau_R(j)}$, that is, the jth order statistic of $Y_1, \ldots, Y_{\tau_R}$. The proof of the following CLT is given in Appendix B.6.

Theorem 6. *If X is polynomially ergodic of order $m > 1$, then, as $R \to \infty$,*

$$\sqrt{R}\left(Y_{\tau_R(j)} - \xi_q\right) \xrightarrow{d} \mathrm{N}\left(0, \Gamma\left(\xi_q\right)/f_V^2\left(\xi_q\right)\right).$$

Since $\hat{\xi}_{\tau_R,q}$ requires j such that $0 \le j - \tau_R q < 1$ we have the following corollary.

Corollary 2. *If X is polynomially ergodic of order $m > 1$, then, as $R \to \infty$,*

$$\sqrt{R}(\hat{\xi}_{\tau_R,q} - \xi_q) \xrightarrow{d} \mathrm{N}\left(0, \Gamma\left(\xi_q\right)/f_V^2\left(\xi_q\right)\right).$$

To obtain an MCSE we need to estimate $\gamma_R^2(\xi_q) := \Gamma(\xi_q)/f_V^2(\xi_q)$. We substitute $\hat{\xi}_{\tau_R,q}$ for ξ_q and separately consider $\Gamma(\hat{\xi}_{\tau_R,q})$ and $f_V(\hat{\xi}_{\tau_R,q})$. Of course, we can handle estimating $f_V(\hat{\xi}_{\tau_R,q})$ exactly as before, so all we need to concern ourselves with is estimation of $\Gamma(\hat{\xi}_{\tau_R,q})$.

We can recognize $\Gamma(y)$ as the variance of an asymptotic Normal distribution. Let $\hat{F}_R(y) = \sum_{t=1}^R S_t(y)/\sum_{t=1}^R N_t$. Then, using (17), we have that, with probability 1, as $R \to \infty$, $\hat{F}_R(y) \to F_V(y)$ for each fixed y. Moreover, using (18), for each $y \in \mathbb{R}$, as $R \to \infty$,

$$\sqrt{R}(\hat{F}_R(y) - F_V(y)) \xrightarrow{d} \mathrm{N}\left(0, \Gamma(y)\right).$$

We can consistently estimate $\Gamma(y)$ for each y with

$$\hat{\Gamma}_R(y) = \frac{1}{R\bar{N}^2}\sum_{t=1}^{R}(S_t(y) - \hat{F}_R(y)N_t)^2.$$

Letting $\hat{f}_V(\hat{\xi}_{\tau_R,q})$ denote an estimator of $f_V(\hat{\xi}_{\tau_R,q})$ we estimate $\gamma_R^2(\xi_q)$ with

$$\hat{\gamma}_R^2(\hat{\xi}_{\tau_R,q}) := \frac{\hat{\Gamma}(\hat{\xi}_{\tau_R,q})}{\hat{f}_V(\hat{\xi}_{\tau_R,q})}.$$

Finally, if $t_{R-1,\alpha/2}$ is a quantile from a Student's t distribution with $R-1$ degrees of freedom, a $100(1-\alpha)\%$ confidence interval for ξ_q is

$$\hat{\xi}_{\tau_R,q} \pm t_{R-1,\alpha/2}\frac{\hat{\gamma}_R(\hat{\xi}_{\tau_R,q})}{\sqrt{R}}. \tag{19}$$

5. Examples

In this section, we investigate the finite-sample performance of the confidence intervals for ξ_q defined at (12), (13), and (19) corresponding to BM, SBM and RS, respectively. While each of our examples are quite different, the simulation studies were conducted using a common methodology. In each case we perform many independent replications of the MCMC sampler. Each replication was performed for a fixed number of regenerations, then confidence intervals were constructed on the same MCMC output. For the BM-based and SBM-based intervals we always used $b_n = \lfloor n^{1/2} \rfloor$, which has been found to work well in other settings (Jones et al., 2006; Flegal and Jones, 2010; Flegal, 2012). In order to estimate coverage probabilities we require the true values of the quantiles of interest. These are available in only one of our examples. In the other example we estimate the truth with an independent long run of the MCMC sampler. The details are described in the following sections.

5.1. Polynomial target distribution

Jarner and Roberts (2007) studied MCMC for heavy-tailed target distributions. A target distribution is said to be *polynomial of order r* if its density satisfies $f(x) = (l(|x|)/|x|)^{1+r}$, where $r > 0$ and l is a normalized slowly varying function—a particular example is Student's t-distribution. We consider estimating quantiles of Student's t-distribution $t(v)$ for degrees of freedom $v = 3$, 6, and 30; the $t(v)$ distribution is polynomial of order v. We use a Metropolis random walk algorithm with jump proposals drawn from a $\mathrm{N}(0, \sigma^2)$ distribution. By Proposition 3 of Jarner and Roberts (2007), a Metropolis random walk for a $t(v)$ target distribution using any proposal kernel with finite variance is polynomially ergodic of order $v/2$. Thus the conditions of Theorem 4 and Corollary 2 are satisfied for $v > 2$.

We tuned the scale parameter σ^2 in the proposal distribution in order to minimize autocorrelation in the resulting chain (second row of Table 2); the resulting acceptance rates varied from about 25% for $t(3)$ with $\sigma = 5.5$, the heaviest tailed target distribution, to about 40% for $t(30)$ with $\sigma = 2.5$. Regeneration times were identified using the retrospective method of Mykland et al. (1995); see Appendix C for implementation details, and the bottom rows of Table 2 for regeneration performance statistics (mean and SD of tour lengths). For

TABLE 2
Metropolis random walk on $t(v)$ target distribution with $N(0, \sigma^2)$ jump proposals, example of Section 5.1

	Target distribution		
	$t(30)$	$t(6)$	$t(3)$
Tuning parameter σ	2.5	3.5	5.5
Mean tour length	3.58	4.21	5.60
SD of tour lengths	3.14	3.80	5.23

Table 3

Empirical coverage rates for nominal 95% confidence intervals for ξ_q, the q-quantile of the $t(v)$ distribution. Based on $n = 10^4$ replications of 500 or 2000 regenerations of a Metropolis random walk with jump proposals drawn from a Normal distribution. The Monte Carlo standard errors for the observed sample proportions fall between 1.5E-3 and 3.2E-3

Estimating ξ_q of $t(v)$ distribution based on Normal Metropolis RW

		500 regenerations			2000 regenerations		
Quantile	Method	$t(30)$	$t(6)$	$t(3)$	$t(30)$	$t(6)$	$t(3)$
$q = 0.50$	BM	0.941	0.939	0.935	0.946	0.946	0.947
	SBM	0.946	0.945	0.947	0.948	0.949	0.950
	RS	0.952	0.951	0.946	0.951	0.950	0.952
$q = 0.75$	BM	0.935	0.931	0.932	0.946	0.939	0.945
	SBM	0.944	0.948	0.955	0.948	0.948	0.961
	RS	0.947	0.942	0.942	0.951	0.944	0.951
$q = 0.90$	BM	0.923	0.916	0.916	0.941	0.935	0.933
	SBM	0.926	0.942	0.957	0.948	0.955	0.976
	RS	0.933	0.928	0.927	0.945	0.940	0.940
$q = 0.95$	BM	0.906	0.898	0.895	0.934	0.930	0.931
	SBM	0.888	0.898	0.932	0.935	0.956	0.972
	RS	0.914	0.909	0.906	0.938	0.936	0.935

each of the 10^4 replications and using each of (12), (13), and (19) we computed a 95% confidence interval for ξ_q for $q = 0.50$, 0.75, 0.90, and 0.95.

Empirical coverage rates (percentage of the 10^4 intervals that indeed contain the true quantile ξ_q) are shown in Table 3. We first note that, as might be expected, agreement with the nominal coverage rate is closer for estimation of the median than for the tail quantiles $\xi_{.90}$ and $\xi_{.95}$. As for comparing the three approaches to MCSE estimation, we find that agreement with the nominal coverage rate is closest for SBM on average, but SBM also shows the greatest variability between cases considered, including a couple of instances ($\xi_{.90}$ and $\xi_{.95}$ for the $t(3)$ target distribution) where the method appears overly conservative. Results for BM and RS show less variability than those of SBM, with agreement with the nominal rate being slightly better for RS.

Table 4 shows the mean and standard deviation of interval half-widths for the three cases (defined by the quantile q and number of regenerations R) in which all empirical coverage rates were at least 0.935. The most striking result here is the huge variability in the standard errors as computed by SBM, particularly for the heaviest tailed target distribution. Results for BM and RS are comparable, with RS intervals being slightly wider and having slightly less variability. The SBM intervals are generally as wide or wider, demonstrating again the apparent conservatism of the method.

5.2. Probit regression

van Dyk and Meng (2001) report data which is concerned with the occurrence of latent membranous lupus nephritis. Let y_i be an indicator of the disease (1 for present), x_{i1} be the difference between IgG3 and IgG4 (immunoglobulin G), and x_{i2} be IgA (immunoglobulin A) where $i = 1, \ldots, 55$. Let Φ denote the standard

Table 4
Mean and standard deviation for half-widths of 95% confidence intervals for ξ_q, in 10^4 replications of Normal Metropolis random walk with R regenerations

$q = 0.50,\ R = 500$

MCSE Method	Target distribution $t(30)$	$t(6)$	$t(3)$
BM	0.120 (0.022)	0.127 (0.023)	0.134 (0.025)
SBM	0.121 (0.016)	0.129 (0.021)	0.146 (0.099)
RS	0.124 (0.015)	0.131 (0.017)	0.140 (0.020)

$q = 0.50,\ R = 2000$

MCSE Method	Target distribution $t(30)$	$t(6)$	$t(3)$
BM	0.061 (0.008)	0.064 (0.008)	0.068 (0.008)
SBM	0.060 (0.005)	0.064 (0.006)	0.072 (0.066)
RS	0.062 (0.004)	0.065 (0.005)	0.069 (0.006)

$q = 0.75,\ R = 2000$

MCSE Method	Target distribution $t(30)$	$t(6)$	$t(3)$
BM	0.066 (0.009)	0.072 (0.009)	0.080 (0.011)
SBM	0.066 (0.006)	0.074 (0.012)	0.094 (0.095)
RS	0.067 (0.005)	0.073 (0.006)	0.082 (0.008)

normal distribution function and suppose

$$\Pr(Y_i = 1) = \Phi\left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}\right)$$

and take the prior on $\beta := (\beta_0, \beta_1, \beta_2)$ to be Lebesgue measure on $\mathbb{R}^3$. Roy and Hobert (2007) show that the posterior $\pi(\beta|y)$ is proper. Our goal is to report a median and an 80% Bayesian credible region for each of the three marginal posterior distributions. Denote the qth quantile associated with the marginal for β_j as $\xi_q^{(j)}$ for $j = 0, 1, 2$. Then the vector of parameters to be estimated is

$$\Xi = \left(\xi_{.1}^{(0)}, \xi_{.5}^{(0)}, \xi_{.9}^{(0)}, \xi_{.1}^{(1)}, \xi_{.5}^{(1)}, \xi_{.9}^{(1)}, \xi_{.1}^{(2)}, \xi_{.5}^{(2)}, \xi_{.9}^{(2)}\right).$$

We will sample from the posterior using the PX-DA algorithm of Liu and Wu (1999), which Roy and Hobert (2007) prove is geometrically ergodic. For a full description of this algorithm in the context of this example see Flegal and Jones (2010) or Roy and Hobert (2007).

We now turn our attention to comparing coverage probabilities for estimating elements of Ξ based on the confidence intervals at (12), (13), and (19). We calculated a precise estimate from a long simulation of the PX-DA chain and declared the observed quantiles to be the truth–see Table 5. Roy and Hobert (2007) implement RS for this example and we use their settings exactly with 25 regenerations. This procedure was repeated for 1000 independent replications resulting in a mean simulation effort of 3.89E5 (2400). The resulting coverage probabilities can be found in Table 6. Notice that for the BM and SBM intervals all the coverage probabilities are within two MCSEs of the nominal 0.95 level.

Table 5

Summary for Probit regression example of calculated "truth". These calculations are based on 9E6 iterations where the MCSEs are calculated using a BM procedure

q	0.1	0.5	0.9
β_0	-5.348 (7.21E-03)	-2.692 (4.00E-03)	-1.150 (2.32E-03)
β_1	3.358 (4.79E-03)	6.294 (7.68E-03)	11.323 (1.34E-02)
β_2	1.649 (2.98E-03)	3.575 (5.02E-03)	6.884 (8.86E-03)

Table 6

Summary for estimated coverage probabilities and observed CI half-widths for Probit regression example. CIs reported have 0.95 nominal level with MCSEs equal ranging from 6.5E-3 to 7.9E-3

		Probability			Half-Width		
	q	0.1	0.5	0.9	0.1	0.5	0.9
β_0	BM	0.956	0.948	0.945	0.0671 (0.007)	0.0377 (0.004)	0.0222 (0.002)
	RS	0.942	0.936	0.934	0.0676 (0.015)	0.0384 (0.008)	0.0226 (0.005)
	SBM	0.952	0.947	0.955	0.0650 (0.006)	0.0375 (0.004)	0.0232 (0.003)
β_1	BM	0.948	0.943	0.948	0.0453 (0.005)	0.0720 (0.007)	0.1260 (0.013)
	RS	0.942	0.936	0.934	0.0459 (0.010)	0.0733 (0.016)	0.1270 (0.028)
	SBM	0.954	0.942	0.940	0.0464 (0.005)	0.0716 (0.007)	0.1230 (0.012)
β_2	BM	0.949	0.950	0.950	0.0287 (0.003)	0.0474 (0.005)	0.0825 (0.009)
	RS	0.938	0.940	0.937	0.0292 (0.006)	0.0481 (0.010)	0.0831 (0.018)
	SBM	0.955	0.948	0.948	0.0297 (0.003)	0.0470 (0.005)	0.0801 (0.008)

However, for RS only 7 of the 9 investigated settings are within two MCSEs of the nominal level. In addition, all of the results using RS are below the nominal 0.95 level.

Table 6 gives the empirical mean and standard deviation of the half-width of the BM-based, RS-based, and SBM-based confidence intervals. Notice the interval lengths are similar across the three methods, but the RS-based interval lengths are more variable. Further, the RS-based intervals are uniformly wider on average than the BM-based intervals even though they have uniformly lower empirical coverage probabilities.

5.3. A hierarchical random effects model

A well known data set first analyzed by Efron and Morris (1975) consists of the batting averages of 18 Major League Baseball players in their first 45 official at bats of the 1970 season. Let x_i denote the batting average of the ith player, and $y_i = \sqrt{45}\arcsin(2x_i - 1)$, for $i = 1, \ldots, K = 18$. Since this represents the variance stabilizing transformation of a binomial distribution, it is reasonable to suppose that

$$y_i|\theta_i \sim \mathrm{N}(\theta_i, 1) \quad \text{for } i = 1, \ldots, K.$$

Here we consider a hierarchical model proposed by Rosenthal (1996). Specifically we further assume that

$$\theta_1, \ldots, \theta_K \text{ are i.i.d. } \mathrm{N}(\mu, \lambda)$$

where

$$p(\mu, \lambda) \propto \lambda^{-(b+1)} e^{-c/\lambda} I(\lambda > 0)$$

TABLE 7

Monte Carlo estimates of posterior quantiles for θ_9 in example of Section 5.3, taken as the "truth" in subsequent analysis. Based on 2E7 independent draws

q	0.1	0.3	0.5	0.7	0.9
$\xi_q^{(9)}$	-4.278	-3.771	-3.428	-3.087	-2.590
MCSE	(2.6E-4)	(1.9E-4)	(1.8E-4)	(1.9E-4)	(2.5E-4)

TABLE 8

Empirical coverage rates of nominal 95% confidence intervals for $\xi_q^{(9)}$ in example of Section 5.3. Based on 5000 simulations, MCSEs range from 3.3E-3 to 3.6E-3

	q				
Method	0.1	0.3	0.5	0.7	0.9
BM	0.936	0.939	0.942	0.944	0.934
SBM	0.941	0.937	0.939	0.940	0.941
RS	0.932	0.938	0.940	0.940	0.931

with b and c known hyperparameters; thus μ has the flat prior and λ has an inverse gamma prior. This results in a proper posterior having dimension $K+2 = 20$. Rosenthal (1996) developed a block Gibbs sampler for simulating from the posterior distribution of $(\theta_1, \ldots, \theta_K, \mu, \lambda)$ and proved that the resulting Markov chain is geometrically ergodic. Jones et al. (2006) showed how to implement regenerative simulation.

Suppose we are interested in estimating the posterior quantiles $\xi_q^{(i)}$ of a particular θ_i, representing the "true" (transformed) batting average of a particular ballplayer. We conduct a simulation study to assess the performance of the confidence intervals at (12), (13), and (19), corresponding to BM, SBM, and RS, respectively.

Jones et al. (2006) showed how to simulate independent draws from the posterior distribution via rejection sampling. Setting hyperparameter values at $b = c = 2$, we generated 2E7 iterations of the rejection sampler to estimate the quantiles $\xi_q^{(9)}$—the 9th player in Efron and Morris's (1975) data set was Ron Santo of the Chicago Cubs—and obtained the quantiles summarized in Table 7. We then ran 5000 replications of Rosenthal's (1996) Gibbs sampler for 50 regenerations each. Using the regeneration recipe of Jones et al. (2006), the mean tour length was about 28 updates, with a standard deviation of approximately 28 as well. For each realized chain, we computed 95% confidence intervals for $\xi_q^{(9)}$ using each of (12), (13), and (19). Empirical coverage rates (with the values in Table 7 taken as the "truth") are reported in Table 8, and interval half-widths are summarized in Table 9.

6. Discussion

We have focused on assessing the Monte Carlo error for estimating quantiles in MCMC settings. In particular, we established quantile CLTs and considered using batch means, subsampling and regenerative simulation to estimate the variance of the asymptotic Normal distributions. We also studied the finite-

Table 9
Mean (and standard deviation) of CI half-widths for nominal 95% confidence intervals for $\xi_q^{(9)}$ in example of Section 5.3, based on 5000 replications

	Method		
q	BM	SBM	RS
0.1	0.0650 (0.010)	0.0656 (0.008)	0.0651 (0.011)
0.3	0.0514 (0.008)	0.0506 (0.006)	0.0519 (0.008)
0.5	0.0490 (0.007)	0.0479 (0.006)	0.0494 (0.008)
0.7	0.0507 (0.007)	0.0497 (0.006)	0.0511 (0.008)
0.9	0.0623 (0.009)	0.0631 (0.008)	0.0629 (0.011)

sample properties of the resulting confidence intervals in the context of three examples.

Overall, the finite-sample properties were comparable across the three variance estimation techniques considered. However, SBM required substantially more computational effort because it orders each of the $n - b + 1$ overlapping blocks to obtain the quantile estimates. For example, we ran a three dimensional probit regression Markov chain (Section 5.2) for 2×10^5 iterations and calculated an MCSE for the median of the three marginals. The BM calculation took 0.37 seconds while the SBM calculation took 84.04 seconds, or 227 times longer.

The conditions required in the CLT in Theorem 4 are the same as those required in the CLT of Theorem 6. However, RS requires stronger conditions in the sense that it requires the user to establish a *useful* minorization condition (14). Although minorization conditions are often nearly trivial to establish, they are seen as a substantial barrier by practitioners because they require a problem-specific approach. Alternatively, it is straightforward to implement the BM-based and SBM-based approaches in general software–see the recent `mcmcse` R package (Flegal and Hughes, 2012) which implements the methods of this paper.

Acknowledgments

The authors are grateful to two anonymous referees and an associate editor for their constructive comments which resulted in many improvements. The second author's work is partially supported by NSF DMS-13-08270. The third author's work is partially supported by NSF DMS-13-10096 and NIH NIBIB R01 EB012547.

Appendix A: Preliminaries: Markov chains as mixing processes

Let $S = \{S_n\}$ be a strictly stationary stochastic process on a probability space $(\Omega, \mathcal{F}, P)$ and set $\mathcal{F}_k^l = \sigma(S_k, \dots, S_l)$. Define the α-mixing coefficients for $n = 1, 2, 3, \dots$ as

$$\alpha(n) = \sup_{k \geq 1} \sup_{A \in \mathcal{F}_1^k,\, B \in \mathcal{F}_{k+n}^\infty} |P(A \cap B) - P(A)P(B)|.$$

Let $f : \Omega \to \mathbb{R}$ be Borel. Set $T = \{f(S_n)\}$ and let α_T and α_S be the α-mixing coefficients for T and S, respectively. Then by elementary properties of sigmaalgebras (cf. Chow and Teicher, 1978, p. 16) $\sigma(T_k,\ldots,T_l) \subseteq \sigma(S_k,\ldots,S_l) = \mathcal{F}_k^l$ and hence $\alpha_T(n) \le \alpha_S(n)$ for all n.

Define the β-mixing coefficients for $n = 1, 2, 3, \ldots$ as

$$\beta(n) = \sup_{\substack{m\in\mathbb{N} \\ A_1,\ldots,A_I \in \mathcal{F}_1^m \\ A_1,\ldots,A_I \text{ partition } \Omega \\ B_1,\ldots,B_J \in \mathcal{F}_{m+n}^\infty \\ B_1,\ldots,B_J \text{ partition } \Omega}} \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J} |P(A_i \cap B_j) - P(A_i)P(B_j)| .$$

If $\beta(n) \to 0$ as $n \to \infty$, we say that S is *β-mixing* while if $\alpha(n) \to 0$ as $n \to \infty$, we say that S is *α-mixing*. It is easy to prove that $2\alpha(n) \le \beta(n)$ (see Bradley, 1986, for discussion of this and other inequalities) for all n so that β-mixing implies α-mixing.

Let X be a stationary Harris ergodic Markov chain on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$, which has invariant distribution π. In this case the expressions for the α- and β-mixing coefficients can be simplified

$$\alpha(n) = \sup_{A,\, B\in\mathcal{B}} \left| \int_A \pi(dx) P^n(x, B) - \pi(A)\pi(B) \right|$$

while Davydov (1973) showed that

$$\beta(n) = \int_{\mathsf{X}} \|P^n(x,\cdot) - \pi(\cdot)\| \pi(dx). \tag{20}$$

Theorem 7. *A stationary Harris ergodic Markov chain is β-mixing, hence α-mixing. In addition, if (4) holds, then $\beta(n) \le \psi(n) E_\pi M$ for all n.*

Proof. The first part is Theorem 4.3 of Bradley (1986) while the second part can be found in the proof of Theorem 2 in Chan and Geyer (1994). □

Since $2\alpha(n) \le \beta(n)$ we observe that Theorem 7 ensures that if $p \ge 0$, then

$$\sum_{n=1}^{\infty} n^p \psi(n) < \infty \quad \text{implies} \quad \sum_{n=1}^{\infty} n^p \alpha(n) < \infty. \tag{21}$$

Appendix B: Proofs

B.1. Proof of Proposition 1

We begin by showing that we can weaken the conditions of Lemma 3.3 in Wang et al. (2011).

Lemma 1. *Let $S = \{S_n\}$ be a stationary α-mixing process such that $\alpha_S(n) \le Cn^{-\beta}$ for some $\beta > 1$ and positive finite constant C. Assume the common* *marginal distribution function F is absolutely continuous with continuous density function f such that $0 < f(\xi_q) < \infty$. For any $\theta > 0$ and $\delta \in (9/(10+8\beta), 1/2)$ there exists n_0 so that if $n \geq n_0$, then with probability 1*

$$|\hat{\xi}_{n,q} - \xi_q| \leq \frac{\theta(\log\log n)^{1/2}}{f(\xi_q)n^{1/2-\delta}}.$$

Proof. Let $\epsilon_n = \theta(\log\log n)^{1/2}/f_V(\xi_p)n^{1/2-\delta}$. Set $\delta_{n1} = F(\xi_q+\epsilon_n) - F(\xi_q)$ and note that by Taylor's expansion there exists $0 < h < 1$ such that

$$\delta_{n1} = \epsilon_n f(\xi_q)\frac{f(h\epsilon_n + \xi_q)}{f(\xi_q)}.$$

Also, note that

$$\frac{f(h\epsilon_n + \xi_q)}{f(\xi_q)} \to 1 \qquad n \to \infty$$

and hence for sufficiently large n

$$\frac{f(h\epsilon_n + \xi_q)}{f(\xi_q)} \geq \frac{1}{2}.$$

Then for sufficiently large n

$$\delta_{n1} \geq \frac{1}{2}\epsilon_n f(\xi_q) = \frac{\theta}{2}\frac{(\log\log n)^{1/2}}{n^{1/2-\delta}}.$$

A similar argument shows that for sufficiently large n

$$\delta_{n2} = F(\xi_q) - F(\xi_q - \epsilon_n) \geq \frac{\theta}{2}\frac{(\log\log n)^{1/2}}{n^{1/2-\delta}}.$$

The remainder exactly follows the proof of Lemma 3.3 in Wang et al. (2011) and hence is omitted. □

The proof of Proposition 1 will follow directly from the following Corollary.

Corollary 3. *Suppose the stationary Markov chain X is polynomially ergodic of order $m > 1$. For any $\theta > 0$ and $\delta \in (9/(10+8m), 1/2)$ with probability 1 for sufficiently large n*

$$|\hat{\xi}_{n,q} - \xi_q| \leq \frac{\theta(\log\log n)^{1/2}}{f_V(\xi_q)n^{1/2-\delta}}$$

and hence there is a positive constant C_0 such that $\hat{\xi}_{n,q} \in [\,\xi_q - C_0 n^{-1/2+\delta}\sqrt{\log\log n}, \xi_q + C_0 n^{-1/2+\delta}\sqrt{\log\log n}\,]$ with probability 1 for sufficiently large n.

Proof. Let $\alpha_Y(n)$ be the strong mixing coefficients for $Y = \{g(X_n)\}$ and note that $\alpha_Y(n) \leq n^{-m}E_\pi M$ by Theorem 7. The remainder follows from Lemma 1 and our basic assumptions on F_V and f_V. □

B.2. Proof of Theorems 2 and 3

We begin with some preliminary results.

Lemma 2. *Let X be stationary with β–mixing coefficients $\beta(n)$. Suppose $h : \mathsf{X} \to \mathbb{R}$ and set $W = \{h(X_n)\}$. If $||h|| := \sup_{x\in\mathsf{X}} |h(x)| < \infty$, then for any integer $a \in [1, n/2]$ and each $\epsilon > 0$,*

$$\Pr\left(\left|\sum_{i=0}^{n-1}(W_i - E_\pi W_i)\right| > n\epsilon\right) \leq 4\exp\left\{-\frac{a\epsilon^2}{8||h||^2}\right\} + 11a\left(1+\frac{4||h||}{\epsilon}\right)^{1/2}\beta\left(\left\lfloor\frac{n}{2a}\right\rfloor\right).$$

Proof. This follows easily by combining observations in Appendix A with Theorem 1.3 from Bosq (1998). □

Lemma 3 (Theorem 2, Glynn and Ormoneit, 2002)**.** *Suppose (5) holds, and $h : \mathsf{X} \to \mathbb{R}$ with $||h|| := \sup_{x\in\mathsf{X}} |h(x)| < \infty$. Set $W = \{h(X_n)\}$ and let $\epsilon > 0$, then for $n > 2||h||n_0/(\lambda\epsilon)$*

$$\Pr\left(\sum_{i=0}^{n-1} W_i - E\left(\sum_{i=0}^{n-1} W_i\right) \geq n\epsilon\right) \leq \exp\left\{-\frac{\lambda^2(n\epsilon - 2||h||n_0/\lambda)^2}{2n||h||^2 n_0^2}\right\}.$$

Lemma 4. *Suppose $X_0 \sim \pi$ and let $g : \mathsf{X} \to \mathbb{R}$ be Borel, $Y = \{g(X_n)\}$ and $\epsilon > 0$ If $W_n = I(Y_n > \xi_q + \epsilon)$ and $\delta_1 = F_V(\xi_q + \epsilon) - q$, then*

$$\Pr\left(\hat{\xi}_{n,q} > \xi_q + \epsilon\right) \leq \Pr\left(\left|\sum_{i=0}^{n-1}(W_i - E_\pi W_i)\right| > n\delta_1\right) \tag{22}$$

while if $V_n = I(Y_n \leq \xi_q - \epsilon)$ and $\delta_2 = q - F_V(\xi_q - \epsilon)$, then for $0 < \delta < 1$

$$\Pr\left(\hat{\xi}_{n,q} < \xi_q - \epsilon\right) \leq \Pr\left(\left|\sum_{i=0}^{n-1}(V_i - E_\pi V_i)\right| > n\delta_2\delta\right). \tag{23}$$

Proof. We compute

$$\begin{aligned}
\Pr\left(\hat{\xi}_{n,q} > \xi_q + \epsilon\right) &= \Pr\left(F_n(\hat{\xi}_{n,q}) > F_n(\xi_q + \epsilon)\right) \\
&= \Pr\left(q > F_n(\xi_q + \epsilon)\right) \\
&= \Pr\left(\sum_{i=0}^{n-1} I(Y_i > \xi_q + \epsilon) > n(1-q)\right) \\
&= \Pr\left(\sum_{i=0}^{n-1}(W_i - E_\pi W_i) > n\delta_1\right) \\
&\leq \Pr\left(\left|\sum_{i=0}^{n-1}(W_i - E_\pi W_i)\right| > n\delta_1\right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\Pr\left(\hat{\xi}_{n,q} < \xi_q - \epsilon\right) &\le \Pr\left(F_n(\hat{\xi}_{n,q}) \le F_n(\xi_q - \epsilon)\right) \\
&\le \Pr\left(q \le F_n(\xi_q - \epsilon)\right) \\
&= \Pr\left(\sum_{i=0}^{n-1} I(Y_i \le \xi_q - \epsilon) \ge nq\right) \\
&= \Pr\left(\sum_{i=0}^{n-1} (V_i - E_\pi V_i) \ge n\delta_2\right) \\
&\le \Pr\left(\left|\sum_{i=0}^{n-1} (V_i - E_\pi V_i)\right| > n\delta_2\delta\right).
\end{aligned}$$

□

Proof of Theorem 2. Let $\epsilon > 0$. Then

$$\Pr\left(\left|\hat{\xi}_{n,q} - \xi_q\right| > \epsilon\right) = \Pr\left(\hat{\xi}_{n,q} > \xi_q + \epsilon\right) + \Pr\left(\hat{\xi}_{n,q} < \xi_q - \epsilon\right).$$

From Lemmas 2 and 4, we have for any integer $a \in [1, n/2]$,

$$\Pr\left(\hat{\xi}_{n,q} > \xi_q + \epsilon\right) \le 4\exp\left\{-\frac{a\delta_1^2}{8}\right\} + 11a\left(1 + \frac{4}{\delta_1}\right)^{1/2} \beta\left(\left\lfloor \frac{n}{2a} \right\rfloor\right)$$

and

$$\Pr\left(\hat{\xi}_{n,q} < \xi_q - \epsilon\right) \le 4\exp\left\{-\frac{a(\delta_2\delta)^2}{8}\right\} + 11a\left(1 + \frac{4}{\delta_2\delta}\right)^{1/2} \beta\left(\left\lfloor \frac{n}{2a} \right\rfloor\right).$$

Suppose $\gamma = \min\{\delta_1, \delta_2\delta\}$, then

$$\Pr\left(\left|\hat{\xi}_{n,q} - \xi_q\right| > \epsilon\right) \le 8\exp\left\{-\frac{a\gamma^2}{8}\right\} + 22a\left(1 + \frac{4}{\gamma}\right)^{1/2} \beta\left(\left\lfloor \frac{n}{2a} \right\rfloor\right).$$

Finally note that by Theorem 7

$$\beta\left(\left\lfloor \frac{n}{2a} \right\rfloor\right) \le \psi\left(\left\lfloor \frac{n}{2a} \right\rfloor\right) E_\pi M.$$

□

Proof of Corollary 1. As in the proof of Theorem 2 we have

$$\Pr\left(\left|\hat{\xi}_{n,q} - \xi_q\right| > \epsilon\right) \le 8\exp\left\{-\frac{a\gamma^2}{8}\right\} + 22a\left(1 + \frac{4}{\gamma}\right)^{1/2} \beta\left(\left\lfloor \frac{n}{2a} \right\rfloor\right).$$

That

$$\beta\left(\left\lfloor \frac{n}{2a} \right\rfloor\right) \le (1 - \lambda)^{\left\lfloor \frac{n}{2an_0} \right\rfloor}$$

follows from (20) and that $\|P^n(x, \cdot) - \pi(\cdot)\| \le (1 - \lambda)^{\lfloor n/n_0 \rfloor}$ for all n. □

Proof of Theorem 3. First note that

$$\Pr\left(\left|\hat{\xi}_{n,q} - \xi_q\right| > \epsilon\right) = \Pr\left(\hat{\xi}_{n,q} > \xi_q + \epsilon\right) + \Pr\left(\hat{\xi}_{n,q} < \xi_q - \epsilon\right).$$

From Lemmas 3 and 4 we have for $n > 2n_0/(\lambda\delta_1)$

$$\Pr\left(\hat{\xi}_{n,q} > \xi_q + \epsilon\right) \leq \exp\left\{-\frac{\lambda^2(n\delta_1 - 2n_0/\lambda)^2}{2nn_0^2}\right\}$$

and for $n > 2n_0/(\lambda\delta\delta_2)$

$$\Pr\left(\hat{\xi}_{n,q} < \xi_q - \epsilon\right) \leq \exp\left\{-\frac{\lambda^2(n\delta\delta_2 - 2n_0/\lambda)^2}{2nn_0^2}\right\},$$

Suppose $\gamma = \min\{\delta_1, \delta\delta_2\}$, then for $n > 2n_0/(\lambda\gamma)$

$$\Pr\left(\left|\hat{\xi}_{n,q} - \xi_q\right| > \epsilon\right) \leq 2\exp\left\{-\frac{\lambda^2(n\gamma - 2n_0/\lambda)^2}{2nn_0^2}\right\}.$$

□

B.3. Proof for Example 1

Let $q(x)$ denote the density of a $t(3)$ distribution, $f_X(x)$ the density of a $t(4)$ distribution, $f_{Y|X}(y|x)$ the density of a Gamma$(5/2,\, 2+x^2/2)$ distribution and $\pi(x, y)$ the density at (7). Then the Markov chain has Markov transition density given by

$$k(x', y'|x, y) = f_{Y|X}(y'|x')k(x'|x)$$

where

$$k(x'|x) \geq q(x')\left\{1 \wedge \frac{f_X(x')q(x)}{f_X(x)q(x')}\right\} = f_X(x')\left\{\frac{q(x)}{f_X(x)} \wedge \frac{q(x')}{f_X(x')}\right\}.$$

Since for all x

$$\frac{q(x)}{f_X(x)} \geq \frac{\sqrt{9375}}{32\pi}$$

we have that for all x, y

$$k(x', y'|x, y) \geq \frac{\sqrt{9375}}{32\pi} f_{Y|X}(y'|x')f_X(x') = \frac{\sqrt{9375}}{32\pi}\pi(x', y')$$

and our claim follows immediately.

B.4. Proof of Theorem 4

We need some notation and few definitions before we begin; for more background on what follows the reader should consult van der Vaart and Wellner (1996). A class $\mathcal{T}$ of a set S is said to pick out a subset C of the set $\{x_1, \ldots, x_n\} \subset S$ if $C = T \cap \{x_1, \ldots, x_n\}$ for some $T \subset \mathcal{T}$. The class $\mathcal{T}$ is said to shatter $\{x_1, \ldots, x_n\}$

if it picks out all 2^n possible subsets. $\mathcal{T}$ is a V-C class if there is some $n < \infty$ such that no subset of size n is shattered by $\mathcal{T}$. The subgraph of a function $f : S \to \mathbb{R}$ is the set $\{(s,t) : 0 \le t \le f(s) \text{ or } f(s) \le t \le 0\}$. A class of functions $\mathcal{F}$ is a V-C subgraph class if the class of its subgraphs is a V-C class of sets (in $S \times \mathbb{R}$).

Let $\mathcal{F}$ be a class of functions and define for $f \in \mathcal{F}$

$$\mathbb{G}_n(f) := \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} (f(X_i) - E(f(X_i))).$$

If, considered as a process indexed by $\mathcal{F}$, $\mathbb{G}_n$ converges to a Gaussian limit process in the space

$$l_\infty(\mathcal{F}) := \left\{ g : \mathcal{F} \to \mathbb{R} \, : \, \sup_{f \in \mathcal{F}} |g(f)| < \infty \right\}$$

equipped with the supremum metric, then we say that $\{X_i\}$ satisfies a functional CLT.

We begin with a preliminary result.

Lemma 5. *Let $\mathcal{F}$ be a measurable uniformly bounded V-C subgraph class of functions. If X is stationary and polynomially ergodic of order $m > 1$, then there is a Gaussian process $\{G(f)\}_{f \in \mathcal{F}}$ which has a version with uniformly bounded and uniformly continuous paths with respect to the $L^2(\pi)$-norm such that*

$$\left\{ n^{-1/2} \sum_{i=1}^{n} (f(X_i) - E_\pi f) \right\}_{f \in \mathcal{F}} \Longrightarrow \{G(f)\}_{f \in \mathcal{F}} \quad \text{in} \quad l_\infty(\mathcal{F}). \tag{24}$$

Moreover,

$$\begin{aligned} \operatorname{Var}(G(f)) &= E[f(X_0) - E(f(X_0))]^2 \\ &\quad + 2 \sum_{i=1}^{\infty} E\left[(f(X_0) - E(f(X_0)))(f(X_i) - E(f(X_i)))\right]. \end{aligned} \tag{25}$$

Proof. In light of our Theorem 7, (24) follows from Corollary 2.1 in Arcones and Yu (1994) and (25) follows from Theorem 0 in Bradley (1985). ☐

Proof of Theorem 4. Let $\mathcal{I} = \{1_{(-\infty,t]}\}_{t \in \mathbb{R}}$ and set $\mathcal{F} = \mathcal{I} \circ g = \{1_{(-\infty,t]} \circ g\}_{t \in \mathbb{R}}$. The class of indicator functions $\mathcal{I} = \{1_{(-\infty,t]}\}_{t \in \mathbb{R}}$ is a uniformly bounded V-C class (see Example 2.6.1, page 135, and Problem 9, page 151, in van der Vaart and Wellner (1996)). By Lemma 2.6.18(vii), page 147, in van der Vaart and Wellner (1996) $\mathcal{I} \circ g$ is thus also a V-C class. Letting $\mathbb{F}_n(t) = (1/n) \sum_{i=0}^{n-1} 1_{(-\infty,t]}(Y_i)$ and using Lemma 5 shows that the empirical process $\sqrt{n}(\mathbb{F}_n - F_V)$, satisfies

$$\sqrt{n}(\mathbb{F}_n - F_V) \Longrightarrow \mathbb{G} \tag{26}$$

for a Gaussian process $\mathbb{G}$. Since F_V is continuously differentiable on $[a,b] = [F_V^{-1}(p) - \epsilon, F_V^{-1}(q) + \epsilon]$ for $0 < p < q < 1$ and some $\epsilon > 0$, with positive derivative f_V, it now follows from Theorem 3.9.4 and Lemma 3.9.23(i) in van der Vaart and Wellner (1996) that

$$\sqrt{n}\left(\mathbb{F}_n^{-1} - F_V^{-1}\right) \Longrightarrow -\frac{\mathbb{G}\circ F_V^{-1}}{f_V \circ F_V^{-1}}, \qquad \text{in } l^\infty[p,q], \tag{27}$$

Now, since the class $\mathcal{F} = \mathcal{I}\circ g$ is a (uniformly) bounded class by (27) we have that the variance of $\mathbb{G}(y)$ (which corresponds to evaluating G at $f = 1_{(-\infty,y]}\circ g$) is

$$\begin{aligned} &E\left(1_{(-\infty,y]}(Y_0) - F_V(y)\right)^2 \\ &+ 2\sum_{i=1}^{\infty} E\left((1_{(-\infty,y]}(Y_0) - F_V(y))(1_{(-\infty,y]}(Y_i) - F_V(y))\right) = \sigma^2(y), \end{aligned} \tag{28}$$

where $\sigma^2(y)$ is as defined in (9). To finish, we need to evaluate the processes in (27) at q so that (28) gives us that the variance of $\mathbb{G}(\xi_q)$ is $\sigma^2(\xi_q)$ and thus the variance of $-\mathbb{G}(\xi_q)/f_V(\xi_q) = \sigma^2(\xi_q)/f_V^2(\xi_q)$ as desired.

That the same conclusion holds for any initial distribution follows from the same argument as in Theorem 17.1.6 of Meyn and Tweedie (2009). □

B.5. Proof of Theorem 5

There exists $\epsilon > 0$ such that $m > 1 + \epsilon + 2/\delta$. Using (21) we have that

$$\sum_{n=1}^{\infty} n^{\epsilon+2/\delta}\alpha(n) < \infty.$$

Samur's (2004) Proposition 3.1 implies that $E_Q N_1^{2+\epsilon+2/\delta} < \infty$, and Samur's (2004) Corollary 3.5 says there exists $2 < p_1 < 2+\delta$ such that $E_Q(S_1)^{p_1} < \infty$.

B.6. Proof of Theorem 6

We require a preliminary result before proceeding with the rest of the proof.

Lemma 6. *If X is polynomially ergodic of order $m > 1$, then $\Gamma(y)$ is continuous at ξ_q.*

Proof. Denote the limit from the right and left as $\lim_{y\to x^+}$ and $\lim_{y\to x^-}$, respectively. From the assumption on F_V it is clear that

$$\lim_{y\to\xi_q^+} F_V(y) = \lim_{y\to\xi_q^-} F_V(y). \tag{29}$$

Recall that

$$S_1(y) = \sum_{i=0}^{\tau_1 - 1} I(Y_i \le y).$$

Let $Z_1(y) = S_1(y) - F_V(y)N_1$ and note $E_Q[Z_1(y)] = 0$ since Hobert et al. (2002) show

$$E_Q S_1(y) = F_V(y) E_Q N_1 \text{ for all } y \in \mathbb{R}. \tag{30}$$

Equations (29) and (30) yield $E_Q[\lim_{y\to\xi_q^+} S_1(y)] = E_Q[\lim_{y\to\xi_q^-} S_1(y)]$. The composition limit law and (29) result in

$$E_Q\left[\lim_{y\to\xi_q^+} Z_1(y)^2\right] = E_Q\left[\lim_{y\to\xi_q^-} Z_1(y)^2\right]. \tag{31}$$

What remains to show is that the limit of the expectation is the expectation of the limit. Notice that $0 < S_1(y) \le N_1$ for all $y \in \mathbb{R}$ and

$$|Z_1(y)| = |S_1(y) - F_V(y)N_1| \le S_1(y) + N_1 \le 2N_1,$$

which implies $E_Q[Z_1(y)^2] \le 4E_Q N_1^2$. By Theorem 5 $E_Q N_1^2 < \infty$ and the dominated convergence theorem gives, for any finite x,

$$\lim_{y\to x} E_Q\left[Z_1(y)^2\right] = E_Q\left[\lim_{y\to x} Z_1(y)^2\right].$$

Finally, from the above fact and (31) we have

$$\lim_{y\to\xi_q^+} E_Q\left[Z_1(y)^2\right] = \lim_{y\to\xi_q^-} E_Q\left[Z_1(y)^2\right],$$

and hence $E_Q[Z_1(y)^2]$ is continuous at ξ_q implying the desired result. □

Hobert et al. (2002) show that $\Gamma(y) = \sigma^2(y)E_\pi s$ where s is defined at (14), which yields the following corollary.

Corollary 4. *Under the conditions of Lemma 6, $\sigma^2(y)$ is continuous at ξ_q.*

Proof of Theorem 6. Notice

$$\begin{aligned}
\Pr\left(\sqrt{R}\big(Y_{\tau_R(j)} - \xi_q\big) \le y\right) &= \Pr\left(Y_{\tau_R(j)} \le \xi_q + y/\sqrt{R}\right) \\
&= \Pr\left(\sum_{k=0}^{\tau_R-1} I\{Y_k \le \xi_q + y/\sqrt{R}\} \ge j\right) \\
&= \Pr\left(\sum_{k=0}^{\tau_R-1} \left[I\{Y_k \le \xi_q + y/\sqrt{R}\} - F_V\big(\xi_q + y/\sqrt{R}\big)\right]\right. \\
&\qquad \left. \ge j - \tau_R F_V\left(\xi_q + y/\sqrt{R}\right)\right) \\
&= \Pr\left(\frac{\sqrt{R}}{\tau_R}\sum_{k=0}^{\tau_R-1} W_{R,k} \ge s_R\right),
\end{aligned}$$

where

$$W_{R,k} = I\{Y_k \le \xi_q + y/\sqrt{R}\} - F_V\left(\xi_q + y/\sqrt{R}\right), \quad k = 0,\ldots,\tau_R - 1,$$

and

$$s_R = \frac{\sqrt{R}}{\tau_R}\left(j - \tau_R F_V\left(\xi_q + y/\sqrt{R}\right)\right).$$

First, consider the s_R sequence. Let $h : \mathbb{R}^+ \to \mathbb{R}^+$ satisfy $\lim_{R\to\infty} h(\tau_R)/\sqrt{\tau_R} = 0$ and set $j = \tau_R q + h(\tau_R)$. Note that $q = F_V(\xi_q)$. For $y \neq 0$

$$\begin{aligned} s_R &= \frac{\sqrt{R}}{\tau_R}\left(j - \tau_R F_V(\xi_q + y/\sqrt{R})\right) \\ &= \frac{\sqrt{R}}{\tau_R}\left(\tau_R q + h(\tau_R) - \tau_R F_V(\xi_q + y/\sqrt{R})\right) \\ &= -\frac{y}{y}\frac{\sqrt{R}}{\tau_R}\left(\tau_R F_V(\xi_q + y/\sqrt{R}) - \tau_R q\right) + \frac{\sqrt{R}}{\tau_R} h(\tau_R) \\ &= -y\frac{\sqrt{R}}{y}\left(F_V(\xi_q + y/\sqrt{R}) - F_V(\xi_q)\right) + \frac{\sqrt{R}}{\tau_R} h(\tau_R) \\ &= -y\left(\frac{F_V(\xi_q + y/\sqrt{R}) - F_V(\xi_q)}{y/\sqrt{R}}\right) + \frac{h(\tau_R)}{\sqrt{\bar{N}}\sqrt{\tau_R}}, \end{aligned}$$

which, as $R \to \infty$, converges to $-y f_V(\xi_q)$ since $\bar{N} \to E(N_1)$ with probability 1 where $1 \leq E(N_1) < \infty$ by Kac's theorem. If $y = 0$, then $s_R = h(\tau_R)/\sqrt{\bar{N}}\sqrt{\tau_R}$ and hence $s_R \to 0$ as $R \to \infty$. Thus for all y we have $s_R \to -y f_V(\xi_q)$ as $R \to \infty$.

Second, consider $W_{R,k}$

$$\frac{\sqrt{R}}{\tau_R\left[\Gamma\left(\xi_q + y/\sqrt{R}\right)\right]^{1/2}} \sum_{k=0}^{\tau_R - 1} W_{R,k} \xrightarrow{d} \mathrm{N}(0,1).$$

Lemma 6 and the continuous mapping theorem imply

$$\frac{\sqrt{R}}{\tau_R\left[\Gamma\left(\xi_q\right)\right]^{1/2}} \sum_{k=0}^{\tau_R - 1} W_{R,k} \xrightarrow{d} \mathrm{N}(0,1). \tag{32}$$

Using $s_R \to -y f_V(\xi_Q)$ as $R \to \infty$, (32), and Slutsky's Theorem, we conclude that, as $R \to \infty$,

$$\begin{aligned} P\left(\sqrt{R}\left(Y_{\tau_R(j)} - \xi_q\right) \leq y\right) &= P\left(\frac{\sqrt{R}}{\tau_R\left[\Gamma\left(\xi_q\right)\right]^{1/2}} \sum_{k=0}^{\tau_R - 1} W_{R,k} \geq \frac{s_R}{\left[\Gamma\left(\xi_q\right)\right]^{1/2}}\right) \\ &\to 1 - \Phi\left\{\frac{-y f_V\left(\xi_q\right)}{\left[\Gamma\left(\xi_q\right)\right]^{1/2}}\right\} = \Phi\left\{\frac{y f_V\left(\xi_q\right)}{\left[\Gamma\left(\xi_q\right)\right]^{1/2}}\right\}, \end{aligned}$$

resulting in

$$\sqrt{R}\left(Y_{\tau_R(j)} - \xi_q\right) \xrightarrow{d} \mathrm{N}\left(0, \frac{\Gamma\left(\xi_q\right)}{f_V^2\left(\xi_q\right)}\right).$$

$\square$

Appendix C: Regenerative simulation in example of Section 5.1

The minorization condition necessary for RS is, at least in principle, quite straightforward for a Metropolis-Hastings algorithm. Let $q(x,y)$ denote the proposal kernel density, and $\alpha(x,y)$ the acceptance probability. Then $P(x,dy) \geq q(x,y)\alpha(x,y)dy$, since the right hand side only accounts for *accepted* jump proposals, and the minorization condition is established by finding s' and ν' such that $q(x,y)\alpha(x,y) \geq s'(x)\nu'(y)$. By Theorem 2 of Mykland et al. (1995), the probability of regeneration on an *accepted* jump from x to y is then given by

$$r_A(x,y) = \frac{s'(x)\nu'(y)}{q(x,y)\alpha(x,y)}.$$

Letting π denote the (possibly unnormalized) target density, we have for a Metropolis random walk

$$\alpha(x,y) = \min\left\{\frac{\pi(y)}{\pi(x)},\ 1\right\} \geq \min\left\{\frac{c}{\pi(x)},\ 1\right\} \min\left\{\frac{\pi(y)}{c},\ 1\right\}$$

for any positive constant c. Further, for any point $\tilde{x}$ and any set D we have

$$q(x,y) \geq \inf_{y\in D}\left\{\frac{q(x,y)}{q(\tilde{x},y)}\right\} q(\tilde{x},y) I_D(y).$$

Together, these inequalities suggest one possible choice of s' and ν', which results in

$$r_A(x,y) = I_D(y) \times \frac{\inf_{y\in D}\{q(x,y)/q(\tilde{x},y)\}}{q(x,y)/q(\tilde{x},y)} \times \frac{\min\{c/\pi(x),1\}\min\{\pi(y)/c,1\}}{\min\{\pi(y)/\pi(x),1\}}. \tag{33}$$

For a $t(v)$ target distribution, $\alpha(x,y)$ reduces to

$$\min\left\{\left(\frac{v+x^2}{v+y^2}\right)^{\frac{v+1}{2}},\ 1\right\} \geq \min\left\{\left(\frac{v+x^2}{c}\right)^{\frac{v+1}{2}},\ 1\right\} \times \min\left\{\left(\frac{c}{v+y^2}\right)^{\frac{v+1}{2}},\ 1\right\}$$

and the last component of (33) is given, up to the constant c, by

$$\left[\frac{\min\{v+x^2,c\}}{\min\{v+x^2,v+y^2\}} \times \frac{v+y^2}{\max\{v+y^2,c\}}\right]^{\frac{v+1}{2}}.$$

Since this piece of the acceptance probability takes the value 1 whenever $v+x^2 < c < v+y^2$ or $v+y^2 < c < v+x^2$, it makes sense to take c equal to the median value of $v+X^2$ under the target distribution.

The choice of $\tilde{x}$ and D, and the functional form of the middle component of (33), will of course depend on the proposal distribution. For the Metropolis random walk with Normally distributed jump proposals, $q(x,y) \propto \exp\{-\frac{1}{2\sigma^2}(y-x)^2\}$, taking $D = [\tilde{x}-d,\ \tilde{x}+d]$ for $d>0$ gives

$$\frac{\inf_{y\in D}\{q(x,y)/q(\tilde{x},y)\}}{q(x,y)/q(\tilde{x},y)} = \exp\left\{-\frac{1}{\sigma^2}\{(x-\tilde{x})(y-\tilde{x}) + d|x-\tilde{x}|\}\right\}.$$

For the $t(v)$ distributions we can take $\tilde{x} = 0$ in all cases, but the choice of d should depend on v. With the goal of maximizing regeneration frequency, we arrived at, by trial and error, $d = 2\sqrt{v/(v-2)}$, or two standard deviations in the target distribution.

References

Acosta, F., Huber, M. L., and Jones, G. L. (2014). Markov chain Monte Carlo with linchpin variables. *Preprint*.

Arcones, M. A. and Yu, B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *Journal of Theoretical Probability*, 7:47–71. MR1256391

Baxendale, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *The Annals of Applied Probability*, 15:700–738. MR2114987

Bertail, P. and Clémençon, S. (2006). Regenerative block-bootstrap for Markov chains. *Bernoulli*, 12:689–712. MR2248233

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer, New York. MR1640691

Bradley, R. C. (1985). On the central limit question under absolute regularity. *The Annals of Probability*, 13:1314–1325. MR0806228

Bradley, R. C. (1986). Basic properties of strong mixing conditions. In Eberlein, E. and Taqqu, M. S., editors, *Dependence in Probability and Statistics: A Survey of Recent Results*, pages 165–192. Birkhauser, Cambridge, MA. MR0899990

Brooks, S. P. and Roberts, G. O. (1999). On quantile estimation and Markov chain Monte Carlo convergence. *Biometrika*, 86:710–717. MR1723789

Bühlmann, P. (2002). Bootstraps for time series. *Statistical Science*, 17:52–72. MR1910074

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics*, 14:1171–1179. MR0856813

Chan, K. S. and Geyer, C. J. (1994). Comment on "Markov chains for exploring posterior distributions". *The Annals of Statistics*, 22:1747–1758. MR1329179

Chow, Y. S. and Teicher, H. (1978). *Probability Theory*. Springer-Verlag, New York. MR0513230

Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91:883–904. MR1395755

Datta, S. and McCormick, W. P. (1993). Regeneration-based bootstrap for Markov chains. *The Canadian Journal of Statistics*, 21:181–193. MR1234760

Davydov, Y. A. (1973). Mixing conditions for Markov chains. *Theory of Probability and Its Applications*, 27:312–328.

Doss, H. and Tan, A. (2014). Estimates and standard errors for ratios of normalizing constants from multiple Markov chains via regeneration. *Journal* *of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(4):683–712. MR3248672

Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, 70:311–319. MR0391403

Flegal, J. M. (2012). Applicability of subsampling bootstrap methods in Markov chain Monte Carlo. In Wozniakowski, H. and Plaskota, L., editors, *Monte Carlo and Quasi-Monte Carlo Methods 2010*, volume 23, pages 363–372. Springer Proceedings in Mathematics & Statistics. MR3173844

Flegal, J. M. and Gong, L. (2014). Relative fixed-width stopping rules for Markov chain Monte Carlo simulations. *Statistica Sinica* (to appear).

Flegal, J. M., Haran, M., and Jones, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, 23:250–260. MR2516823

Flegal, J. M. and Hughes, J. (2012). mcmcse: Monte Carlo standard errors for MCMC R package version 1.0-1. `http://cran.r-project.org/web/packages/mcmcse/index.html`.

Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, 38:1034–1070. MR2604704

Flegal, J. M. and Jones, G. L. (2011). Implementing Markov chain Monte Carlo: Estimating with confidence. In Brooks, S., Gelman, A., Jones, G., and Meng, X., editors, *Handbook of Markov Chain Monte Carlo*, pages 175–197. Chapman & Hall/CRC Press. MR2858449

Fort, G. and Moulines, E. (2003). Polynomial ergodicity of Markov transition kernels. *Stochastic Processes and Their Applications*, 103:57–99. MR1947960

Geyer, C. J. (2011). Introduction to Markov chain Monte Carlo. In *Handbook of Markov Chain Monte Carlo*. CRC, London.

Gilks, W. R., Roberts, G. O., and Sahu, S. K. (1998). Adaptive Markov chain Monte Carlo through regeneration. *Journal of the American Statistical Association*, 93:1045–1054. MR1649199

Glynn, P. and Ormoneit, D. (2002). Hoeffding's inequality for uniformly ergodic Markov chains. *Statistics & Probability Letters*, 56:143–146. MR1881167

Hobert, J. P., Jones, G. L., Presnell, B., and Rosenthal, J. S. (2002). On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika*, 89:731–743. MR1946508

Hobert, J. P., Jones, G. L., and Robert, C. P. (2006). Using a Markov chain to construct a tractable approximation of an intractable probability distribution. *Scandinavian Journal of Statistics*, 33:37–51. MR2255108

Jarner, S. F. and Roberts, G. O. (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scandinvian Journal of Statistics*, 24:101–121.

Jones, G. L. (2004). On the Markov chain central limit theorem. *Probability Surveys*, 1:299–320. MR2068475

Jones, G. L., Haran, M., Caffo, B. S., and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 101:1537–1547. MR2279478

Jones, G. L. and Hobert, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, 16:312–334. MR1888447

Jones, G. L. and Hobert, J. P. (2004). Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *The Annals of Statistics*, 32:784–817. MR2060178

Łatuszyński, K., Miasojedow, B., and Niemiro, W. (2012). Nonasymptotic bounds on the estimation error of MCMC algorithms. To appear in *Bernoulli*. MR3129043

Łatuszyński, K. and Niemiro, W. (2011). Rigorous confidence bounds for MCMC under a geometric drift condition. *Journal of Complexity*, 27:23–38. MR2745298

Liu, J. S. and Wu, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association*, 94:1264–1274. MR1731488

Meyn, S. and Tweedie, R. (2009). *Markov Chains and Stochastic Stability*, volume 2. Cambridge University Press Cambridge. MR2509253

Mykland, P., Tierney, L., and Yu, B. (1995). Regeneration in Markov chain samplers. *Journal of the American Statistical Association*, 90:233–241. MR1325131

Politis, D. N. (2003). The impact of bootstrap methods on time series analysis. *Statistical Science*, 18:219–230. MR2026081

Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling*. Springer-Verlag Inc. MR1707286

Raftery, A. E. and Lewis, S. M. (1992). How many iterations in the Gibbs sampler? In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting*, pages 763–773. Clarendon Press.

Robinson, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*, 4:185–207. MR0732897

Rosenthal, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 90:558–566. MR1340509

Rosenthal, J. S. (1996). Analysis of the Gibbs sampler for a model related to James-Stein estimators. *Statistics and Computing*, 6:269–275.

Roy, V. and Hobert, J. P. (2007). Convergence rates and asymptotic standard errors for Markov chain Monte Carlo algorithms for Bayesian probit regression. *Journal of the Royal Statistical Society, Series B*, 69:607–623. MR2370071

Rudolf, D. (2012). Explicit error bounds for Markov chain Monte Carlo. *Dissertationes Mathematicae*, 485. MR2977521

Samur, J. D. (2004). A regularity condition and a limit theorem for Harris ergodic Markov chains. *Stochastic Processes and Their Applications*, 111:207–235. MR2056537

Serfling, R. J. (1981). *Approximation Theorems of Mathematical Statistics*. Wiley-Interscience. MR0595165

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer-Verlag, New York. MR1385671

van Dyk, D. A. and Meng, X.-L. (2001). The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10:1–50. MR1936358

Wang, X., Hu, S., and Yang, W. (2011). The Bahadur representation for sample quantiles under strongly mixing sequence. *Journal of Statistical Planning and Inference*, 141:655–662. MR2732936