Statistica Sinica **25** (2015), 655-675
doi:http://dx.doi.org/10.5705/ss.2013.209

RELATIVE FIXED-WIDTH STOPPING RULES FOR MARKOV CHAIN MONTE CARLO SIMULATIONS

James M. Flegal and Lei Gong

University of California, Riverside

Abstract: Markov chain Monte Carlo (MCMC) simulations are commonly employed for estimating features of a target distribution, particularly for Bayesian inference. A fundamental challenge is determining when these simulations should stop. We consider a sequential stopping rule that terminates the simulation when the width of a confidence interval is sufficiently small relative to the size of the target parameter. Specifically, we propose relative magnitude and relative standard deviation stopping rules in the context of MCMC. In each setting, we develop conditions to ensure the simulation will terminate with probability one and the resulting confidence intervals will have the proper coverage probability. Our results are applicable in such MCMC estimation settings as expectation, quantile, or simultaneous multivariate estimation. We investigate the finite sample properties through a variety of examples, and provide some recommendations to practitioners.

Key words and phrases: Batch means, Bayesian computation, fixed-width confidence intervals, sequential estimation, sequential stopping rules, strong consistency.

1. Introduction

Markov chain Monte Carlo (MCMC) methods allow exploration of intractable probability distributions by constructing a Markov chain whose stationary distribution is the desired distribution. A major challenge for practitioners is determining how long to run an MCMC simulation. Many experiments employ a procedure that terminates after n iterations, where n is determined heuristically. While some simulations are so complex that this is the only practical approach, this is not so for most experiments.

Alternatively, some practitioners use convergence diagnostics to determine if n is sufficiently large (for a review see Cowles and Carlin (1996)). Although practical, these methods are mute about the quality of the resulting estimates (Flegal, Haran, and Jones (2008)) and they can introduce bias directly into the estimates (Cowles, Roberts, and Rosenthal (1999)).

We advocate terminating a simulation when, for the first time, a confidence interval width for a desired quantity is sufficiently small. We refer to such a procedure as a sequential fixed-width stopping rule, and note that the total simulation effort will be random.

Fixed-width methods are especially desirable because they are theoretically justified and constrained by few assumptions. The simplest fixed-width rule, first studied in MCMC by Jones et al. (2006), stops the simulation when the width of a confidence interval based on an ergodic average is less than a user-specified value, say ϵ. Flegal, Haran, and Jones (2008) and Jones et al. (2006) show this stopping rule is superior to using convergence diagnostics as a stopping criteria.

In this paper, we introduce relative fixed-width stopping rules that eliminate the need to specify an absolute value for ϵ. Specifically, the simulation is terminated the first time the width of a confidence interval is sufficiently small relative to the *size* of a target parameter. We consider two measures of size, magnitude and standard deviation. Further, we illustrate the utility of these rules for simultaneous estimation of multiple parameters.

We need some notation. Let π denote a probability distribution having support $\mathsf{X} \subseteq \mathbb{R}^d$, $d \geq 1$, about which we wish to make inference, typically based on parameters of π. For example, if $g : \mathsf{X} \to \mathbb{R}$, we may need to calculate

$$\mu_g := E_\pi[g(X)] = \int_{\mathsf{X}} g(x)\pi(dx) \ ,$$

or if $W \sim \pi$, then we might require quantiles of the distribution F_V of $V = h(W)$, where $h : \mathsf{X} \to \mathbb{R}$:

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

We take $\theta \in \mathbb{R}$ as an unknown target parameter of interest with respect to π. Then, given a probability distribution π, we want to estimate θ.

Frequently π is such that MCMC is the only viable technique for estimating θ. The basic MCMC method entails constructing a time-homogeneous Harris ergodic Markov chain $X = \left\{X^{(0)}, X^{(1)}, \ldots\right\}$ on state space X with σ-algebra $\mathcal{B} = \mathcal{B}(\mathsf{X})$ and invariant distribution π. The popularity of MCMC methods result from the ease with which X can be simulated (Robert and Casella (2004)).

Suppose we simulate X for n iterations. Let Z_n be an estimator of θ from the observed chain, with unknown Monte Carlo error, $Z_n - \theta$. We can obtain its approximate sampling distribution if a Markov chain central limit theorem (CLT) holds:

$$\sqrt{n}\,(Z_n - \theta) \xrightarrow{d} \mathrm{N}(0, \sigma_\theta^2) \tag{1.1}$$

as $n \to \infty$ where $\sigma_\theta^2 \in (0, \infty)$.

If $\hat{\sigma}_n^2$ is an estimator of σ_θ^2, an approximate $(1-\delta)100\%$ confidence interval for θ has width

$$w_\delta = \frac{2 z_{\delta/2} \hat{\sigma}_n}{\sqrt{n}} \tag{1.2}$$

where $z_{\delta/2}$ is a critical value from the standard Normal. The width at (1.2) allows analysts to report the uncertainty in their estimates and users to assess the practical reliability.

We use w_δ to construct sequential fixed-width stopping rules. If an i.i.d. sample from π is available, take λ_θ^2 as the asymptotic variance in the CLT associated with θ. Due to the correlation present in a Markov chain $\sigma_\theta^2 \neq \lambda_\theta^2$, except in trivial cases. For estimation of μ_g, $\lambda_\theta^2 = \text{Var}\,[g(X)]$. For estimation of ξ_q, we have $\lambda_\theta^2 = q(1-q)/\left(f_V(\xi_q)\right)^2$ where f_V is the density associated with F_V.

Our work advocates stopping the simulation the first time w_δ is sufficiently small. We consider three stopping rules: (i) an absolute precision rule that terminates when $w_\delta < \epsilon$, (ii) a relative magnitude rule that terminates when $w_\delta < \epsilon\,|\theta|$, and (iii) a relative standard deviation rule that terminates when $w_\delta < \epsilon\lambda_\theta$. Glynn and Whitt (1992) established conditions for the asymptotic validity of (i) and (ii). In this paper, we extend these results to establish asymptotic validity of the stopping rule (iii).

Flegal, Haran, and Jones (2008), Flegal and Jones (2010) and Jones et al. (2006) have previously investigated (i) for MCMC expectation estimation. We are not aware of any prior use of fixed-width methods for quantile estimation, or any use of (ii) or (iii) as a stopping rule in MCMC. The rule (iii) has significant promise in Bayesian applications since the simulation stops when an estimate of θ is sufficiently accurate relative to an associated posterior standard deviation. Another benefit of (iii) is that it is easy to implement in multivariate settings since ϵ can remain constant.

For asymptotic validity, we require a functional central limit theorem (FCLT) for the Monte Carlo error; Markov chains frequently enjoy a FCLT under identical conditions as those that ensure a CLT. We also require that $\hat{\sigma}_n^2 \to \sigma_\theta^2$ almost surely as $n \to \infty$. Many commonly used MCMC estimators of σ_θ^2 can satisfy this condition, see e.g. Doss et al. (2014), Flegal and Jones (2010), Hobert et al. (2002), and Jones et al. (2006).

We investigate the finite sample properties of relative fixed-width stopping rules through three examples. The first considers an independence Metropolis sampler to explore an exponential random variable; the second considers exploring a mixture of bivariate Normal distributions with Metropolis Hastings and Gibbs samplers. For these examples, we use true parameter values to illustrate the utility of our stopping rules. We also consider a Bayesian version of a logistic regression to model the presence or absence of the freshwater eel *Anguilla australis*.

For Bayesian practitioners, we advocate the stopping rule (iii) for its ease of implementation. In multivariate settings, one can terminate the first time the length of a confidence interval is sufficiently small for each parameter of interest.

Our examples indicate that, setting $\epsilon = 0.02$ provides excellent results in a wide variety of univariate and multivariate settings.

The rest of this paper is organized as follows. Section 2 formally introduces relative fixed-width stopping rules and establishes asymptotic validity. Section 3 investigates fixed-width stopping procedures when estimating expectations and quantiles. Section 4 studies the finite sample properties in numerical examples, and concludes with a discussion that provides some recommendations to practitioners.

2. Sequential Fixed-Width Procedures

In this section, we obtain conditions that ensure asymptotic validity of fixed-width procedures. The primary assumptions are that the limiting process satisfy a FCLT and $\hat{\sigma}_n^2 \to \sigma_\theta^2$ w.p.1 as $n \to \infty$. Section 3 outlines checkable sufficient conditions for the most common MCMC settings, estimating expectations and quantiles.

To estimate a parameter $\theta \in \mathbb{R}$, we assume there exists an $\mathbb{R}$-valued stochastic process $\{Z_n : n \geq 1\}$, the estimation process, for which $Z_n \to \theta$ in probability. Asymptotic validity requires the estimation process satisfies a FCLT as follows. For ease of exposition, we consider an $\mathbb{R}$-valued stochastic process $Z = \{Z(t) : t \geq 0\}$ for which $Z(t) \to \theta$ in probability as $t \to \infty$. Let $D(0, \infty)$ denote the space of right-continuous $\mathbb{R}$-valued functions with left limits on the open interval $(0, \infty)$. We assume that Z has sample paths in $D(0, \infty)$ and consider the family of scaled processes in $D(0, \infty)$ for $\epsilon > 0$

$$\mathcal{Z}_\epsilon(t) = \epsilon^{-1/2} \left(Z\left(\frac{t}{\epsilon}\right) - \theta \right) \text{ , where } t > 0.$$

A FCLT holds if there exists a constant $\sigma_\theta > 0$ such that as $\epsilon \to 0$

$$\mathcal{Z}_\epsilon(t) \xrightarrow{d} \frac{\sigma_\theta B(t)}{t} \ ,$$

in $D(0, \infty)$, where $B(t)$ denotes a standard Brownian motion process $\{B(t) : t \geq 0\}$. In many situations a FCLT holds under the same conditions as those required for an ordinary CLT.

Let $C[n] = \left(Z_n - z_{\delta/2}\hat{\sigma}_n/\sqrt{n} \ , \ Z_n + z_{\delta/2}\hat{\sigma}_n/\sqrt{n} \right)$. If a CLT at (1.1) holds and $\hat{\sigma}_n$ is weakly consistent for σ_θ, then $C[n]$ achieves the nominal coverage level as the sample size $n \to \infty$.

Consider a sequential procedure that terminates at the time the length of a confidence interval drops below a prescribed level ϵ:

$$\tilde{T}(\epsilon) = \inf \left\{ n \geq 0 : \frac{2 z_{\delta/2} \hat{\sigma}_n}{\sqrt{n}} \leq \epsilon \right\} \ .$$

Here $\tilde{T}(\epsilon)$ can be too small if $\hat{\sigma}_n$ is poorly behaved (Glynn and Whitt (1992)). Instead, suppose $p(n)$ is a positive function that decreases monotonically such that $p(n) = o(n^{-1/2})$ as $n \to \infty$ and let n^* be the desired minimum simulation effort (a reasonable default is $p(n) = \epsilon I(n \leq n^*) + n^{-1}$). Then an absolute precision stopping rule terminates the simulation at

$$T_1(\epsilon) = \inf \left\{ n \geq 0 : \frac{2z_{\delta/2}\hat{\sigma}_n}{\sqrt{n}} + p(n) \leq \epsilon \right\} .$$

Theorem 1 in Glynn and Whitt (1992) leads to the asymptotic validity of the sequential stopping rule $T_1(\epsilon)$, the desired coverage probability is obtained in an asymptotic sense as $\epsilon \to 0$.

Proposition 1. *Suppose a FCLT for the Monte Carlo error holds. If $\hat{\sigma}_n \to \sigma_\theta$ w.p.1 as $n \to \infty$, then as $\epsilon \to 0$ the simulation will terminate w.p.1 and $Pr\left(\theta \in C[T_1(\epsilon)]\right) \to 1 - \delta$.*

Remark 1. Glynn and Whitt (1992) show that the weak consistency of $\hat{\sigma}_n$ is not enough to ensure asymptotic validity .

The stopping rule $T_1(\epsilon)$ has previously been used for estimating expectations in MCMC (Flegal and Jones (2010); Flegal, Haran, and Jones (2008); Jones et al. (2006)). We will show that the rule works well for MCMC estimation of quantiles.

Using Z_n as an estimator of θ allows the relative magnitude stopping time

$$T_2(\epsilon) = \inf \left\{ n \geq 0 : \frac{2z_{\delta/2}\hat{\sigma}_n}{\sqrt{n}} + p(n) \leq \epsilon \left|Z_n\right| \right\} .$$

For large n, $T_2(\epsilon)$ behaves like $T_1(\epsilon|\theta|)$. The asymptotic validity of $T_2(\epsilon)$ is a direct consequence of Theorem 3 in Glynn and Whitt (1992).

Proposition 2. *Suppose a FCLT for the Monte Carlo error holds and $|\theta| > 0$. If $Z_n \to \theta$ w.p.1 and $\hat{\sigma}_n \to \sigma_\theta$ w.p.1 as $n \to \infty$, then as $\epsilon \to 0$ the simulation will terminate w.p.1 and $Pr\left(\theta \in C[T_2(\epsilon)]\right) \to 1 - \delta$.*

While $T_2(\epsilon)$ has some support in the operations research literature, it makes little intuitive sense in Bayesian settings. Specifically, if $\theta = 0$ then $T_2(\epsilon)$ is theoretically invalid and poorly behaved in finite simulations, and $T_2(\epsilon)$ can be problematic even when θ is not equal to zero.

We propose a stopping rule that terminates the simulation when the length of a confidence interval is less than the ϵth fraction of the magnitude of λ_θ, the posterior standard deviation of θ. Let $\hat{\lambda}_n$ be an estimator of λ_θ and consider the stopping time

$$T_3(\epsilon) = \inf \left\{ n \geq 0 : \frac{2z_{\delta/2}\hat{\sigma}_n}{\sqrt{n}} + p(n) \leq \epsilon \hat{\lambda}_n \right\} .$$

For large n, $T_3(\epsilon)$ behaves like $T_1(\epsilon\lambda_\theta)$. The next result establishes the asymptotic validity of $T_3(\epsilon)$; it is proved in Appendix A.

Theorem 1. *Suppose a FCLT for the Monte Carlo error holds and $\lambda_\theta > 0$. If $\hat{\lambda}_n \to \lambda_\theta$ w.p.1 and $\hat{\sigma}_n \to \sigma_\theta$ w.p.1 as $n \to \infty$, then as $\epsilon \to 0$ the simulation will terminate w.p.1 and $Pr\,(\theta \in C[T_3(\epsilon)]) \to 1 - \delta$.*

The additional condition required for Theorem 1 is a strongly consistent estimator of λ_θ. For expectations, an estimator is readily available via the Markov chain SLLN; for quantiles, we discuss a viable estimator in the next section.

The benefit of the stopping rule $T_3(\epsilon)$ is twofold: one only needs to specify ϵ; when estimating multiple parameters a single ϵ suffices to obtain estimates whose uncertainties are comparable relative to their standard deviations. In multivariate settings, one could address the issue of multiplicity by adjusting the critical value appropriately. We illustrate this procedure via examples in Section 4, and show the resulting simultaneous confidence regions obtain at least the nominal coverage probability.

Remark 2. Asymptotic validity of relative stopping rules can be established if a FCLT is replaced by a more general $\mathbb{R}$-valued stochastic process (Glynn and Whitt (1992)). The generalization enables consideration of θ that follow non-Normal asymptotic distributions.

Remark 3. A more general relative stopping rule that terminates when $w_\delta < \epsilon\nu_\theta$ can be established for any ν_θ such that $|\nu_\theta| > 0$ provided there exists an estimator $\hat{\nu}_n \to \nu_\theta$ w.p.1. Thus, one could consider relative stopping rules setting ν_θ as the interquartile range, the length of a Bayesian credible region, and so on.

3. Applications

This section demonstrates that fixed-width stopping rules are appropriate for MCMC estimation of expectations and quantiles. Raftery and Lewis (1992) propose a heuristic approach to terminating an MCMC simulation when the primary interest is quantile estimation, but Brooks and Roberts (1999) suggest caution in its use when quantiles themselves are not of interest.

We need some more notation to describe sufficient mixing conditions for a Markov chain CLT and consistent estimation of the asymptotic variance. See Meyn and Tweedie (2009) and Roberts and Rosenthal (2004) for more on Markov chain theory.

Let X be a Harris ergodic Markov chain on state space X with σ-algebra $\mathcal{B} = \mathcal{B}(\mathsf{X})$ and invariant distribution π. Denote the n-step Markov kernel associated with X as $P^n(x, dy)$ for $n \in \mathbb{N}$, so if $A \in \mathcal{B}(\mathsf{X})$ and $k \in \{0, 1, 2, \ldots\}$, $P^n(x, A) =$

$\Pr(X_{k+n} \in A | X_k = x)$. Let $\|\cdot\|$ denote the total variation norm. Let $M : \mathsf{X} \mapsto \mathbb{R}^+$ and $\gamma : \mathbb{N} \mapsto \mathbb{R}^+$ be decreasing such that

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq M(x)\gamma(n) \ . \tag{3.1}$$

Polynomial ergodicity of order m where $m \geq 0$ means (3.1) holds with $E_\pi M < \infty$ and $\gamma(n) = n^{-m}$ for all $X_0 = x$. *Geometrical ergodicity* means (3.1) holds with $\gamma(n) = t^n$ for some $0 < t < 1$ for all $X_0 = x$. *Uniform ergodicity* means (3.1) holds with M bounded and $\gamma(n) = t^n$ for some $0 < t < 1$.

Establishing (3.1) directly can be challenging, but some constructive techniques are available (Jarner and Roberts (2002); Meyn and Tweedie (2009)). Most literature on MCMC algorithms focuses on establishing geometric and uniform ergodicity, see e.g. Hobert (2011), Jones and Hobert (2001), Johnson, Jones, and Neath (2011), Mengersen and Tweedie (1996), Roberts and Tweedie (1996), and Tierney (1994). Less has been said concerning polynomial ergodicity, but see Douc et al. (2004), Fort and Moulines (2000, 2003), Jarner and Roberts (2002, 2007) and Jarner and Tweedie (2003).

3.1. Expectations

For $g : \mathsf{X} \to \mathbb{R}$, consider estimation of

$$\mu_g := E_\pi[g(X)] = \int_{\mathsf{X}} g(x)\pi(dx) \ .$$

Estimating μ_g is natural by appealing to a Markov chain SLLN, a special case of the Birkhoff Ergodic Theorem (Fristedt and Gray (1997)). Specifically, if $E_\pi|g| < \infty$ then w.p.1

$$Z_n := \bar{g}_n := \frac{1}{n}\sum_{i=0}^{n-1} g(X^{(i)}) \to \mu_g \text{ as } n \to \infty \ .$$

Hence the SLLN yields strongly consistent estimators of μ_g and $\lambda_\theta^2 = \mathrm{Var}\,[g(X)]$ (provided $E_\pi g^2 < \infty$).

We can obtain an approximate sampling distribution for the Monte Carlo error via a Markov chain CLT if

$$\sqrt{n}(\bar{g}_n - \mu_g) \stackrel{d}{\to} \mathrm{N}(0, \sigma_g^2) \tag{3.2}$$

as $n \to \infty$ where $\sigma_g^2 \in (0, \infty)$. Conditions that ensure (3.2) can be found in Chan and Geyer (1994), Jones (2004), Meyn and Tweedie (2009), Roberts and Rosenthal (2004), and Tierney (1994). For example, if X is geometrically ergodic and $E_\pi|g|^{2+\epsilon} < \infty$ for some $\epsilon > 0$, then (3.2) holds. Moreover, Markov chains frequently enjoy a FCLT under the same conditions (Jones et al. (2006); Oodaira and Yoshihara (1972); Ibragimov (1962)).

There are many strongly consistent variance estimation techniques applicable for σ_g^2 in MCMC settings including batch means (Flegal and Jones (2010); Jones et al. (2006)), spectral variance techniques (Flegal and Jones (2010)) and regenerative simulation (Hobert et al. (2002); Mykland, Tierney, and Yu (1995)). We consider only non-overlapping batch means (BM) because it is easy to implement and available in many software packages, e.g. the `mcmcse` package available on `CRAN`.

In BM the output is broken into a_n batches where each batch is b_n iterations in length. Suppose the algorithm is run for a total of $n = a_n b_n$ iterations and let

$$\bar{Y}_j := \frac{1}{b_n} \sum_{i=(j-1)b_n+1}^{jb_n} g(X_i) \qquad \text{for } j = 1, \ldots, a_n \ ,$$

and take

$$\hat{\sigma}_n^2 = \frac{b_n}{a_n - 1} \sum_{j=1}^{a_n} (\bar{Y}_j - \bar{g}_n)^2 \ . \tag{3.3}$$

Jones et al. (2006) established necessary conditions for $\hat{\sigma}_n^2 \to \sigma_g^2$ with probability 1 as $n \to \infty$ if the batch size and the number of batches are allowed to increase as the overall length of the simulation increases. Setting $b_n = \lfloor n^\tau \rfloor$ and $a_n = \lfloor n/b_n \rfloor$, the regularity conditions require that X be geometrically ergodic, $E_\pi |g|^{2+\epsilon_1+\epsilon_2} < \infty$ for some $\epsilon_1 > 0$, $\epsilon_2 > 0$ and $(1 + \epsilon_1/2)^{-1} < \tau < 1$. A common choice of $b_n = \lfloor \sqrt{n} \rfloor$ and $a_n = \lfloor n/b_n \rfloor$ has been shown to work well in applications (Jones et al. (2006); Flegal and Jones (2010); Flegal, Haran, and Jones (2008)).

Remark 4. Most sampling plans require storing the entire Markov chain to allow for recalculations as the batch size increases with n. If storage is a concern, one could consider increasing the batch size of the form $b_n \in \{2, 4, 8, ..., 2^k, ...\}$ in an effort to reduce memory usage. One can establish strong consistency for the BM variance estimator with such a sampling plan using results in Jones et al. (2006) and Bednorz and Latuszyński (2007).

3.2. Quantiles

It is common to estimate univariate quantiles associated with π, especially in Bayesian applications. With $W \sim \pi$ and $V = h(W)$, we consider estimation of the quantiles associated with the univariate distribution of V. If F_V is the distribution function of V, our goal is to obtain

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

We suppose that $F_V(x)$ is absolutely continuous and has continuous density function $f_V(x)$ such that $0 < f_V(\xi_q) < \infty$.

Here is the current state of understanding of the MCMC estimation of quantiles (for details see Doss et al. (2014)). Let

$$Z_n := \hat{\xi}_{n,q} = Y_{n(j)}, \quad \text{where} \quad j-1 < nq \leq j \;, \tag{3.4}$$

where $Y_{n(j)}$ denotes the jth order statistic of $\{Y_0, \dots, Y_{n-1}\} = \{h(X_0), \dots, h(X_{n-1}), \}$. If X is Harris recurrent, then $\hat{\xi}_{n,q} \to \xi_q$ w.p.1 as $n \to \infty$ (Doss et al. (2014)).

Under stronger mixing conditions on X, one can obtain a Markov chain CLT. Let

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

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

$$\sqrt{n}(\hat{\xi}_{n,q} - \xi_q) \xrightarrow{d} \text{N}(0, \gamma^2(\xi_q)) \;, \tag{3.5}$$

where $\gamma^2(\xi_q) = \sigma^2(\xi_q)/[f_V(\xi_q)]^2$ (Doss et al. (2014)). A FCLT holds for uniformly ergodic chains via sufficient conditions in Sen (1972) that can be verified with results in Jones (2004). As a direction of future work, it is likely a FCLT holds under polynomial ergodicity, following Doss et al. (2014) and Sen (1972).

Estimation of the variance from the asymptotic Normal distribution at (3.5) is broken into two parts. First, plug in $\hat{\xi}_{n,q}$ for ξ_q and separately consider estimating $f_V(\hat{\xi}_{n,q})$ and $\sigma^2(\hat{\xi}_{n,q})$. Estimate $f_V(\hat{\xi}_{n,q})$ using a kernel density approach with a gaussian kernel to get $\hat{f}_V(\hat{\xi}_{n,q})$. There are well-known conditions guaranteeing strongly consistent estimation of the density at a point (see e.g. Kim and Lee (2005); Yu (1993)).

We will use BM for estimating $\sigma^2(\hat{\xi}_{n,q})$. If we have $n = a_n b_n$ iterations, take, for $k = 0, \dots, a_n - 1$, $\bar{U}_k(\hat{\xi}_{n,q}) := b_n^{-1} \sum_{i=0}^{b_n-1} I(Y_{kb_n+i} \leq \hat{\xi}_{n,q})$. The BM estimate of $\sigma^2(\hat{\xi}_{n,q})$ is

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

Combining $\hat{f}_V(\hat{\xi}_{n,q})$ and $\hat{\sigma}^2_{BM}(\hat{\xi}_{n,q})$, we estimate $\gamma^2(\xi_q)$ with

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

This approach is implemented in the R package `mcmcse` which is used to perform the computations in our examples.

The relative standard deviation fixed-width stopping rule of Theorem 1 requires estimation of

$$\lambda_\theta = \frac{\sqrt{q(1-q)}}{f_V(\xi_q)} \;.$$

We use the same kernel density estimate resulting in

$$\hat{\lambda}_n = \frac{\sqrt{q(1-q)}}{\hat{f}_V(\hat{\xi}_{n,q})} .$$

4. Numerical Studies

This section investigates the finite sample properties of fixed-width stopping rules through a variety of simulations. In each example, we independently repeat the MCMC simulation to evaluate the resulting finite sample confidence intervals. In the first two examples, true values are readily available. In the third, the truth was estimated using an independent long run of the MCMC sampler. Overall, the empirical coverage probabilities obtained via fixed-width stopping rules are close to the nominal level.

Each simulation considered both expectations and quantiles with a common methodology. For a single replication, the same MCMC draws were used in applying the three stopping rules. We set $p(n) = \epsilon I(n < n^*) + n^{-1}$ and estimated σ_θ^2 via BM methods with $b_n = \lfloor \sqrt{n} \rfloor$ calculated with the `mcmcse` package. Standard errors for the empirical coverage probabilities were $\sqrt{\hat{p}(1-\hat{p})/r}$, where r is the number of replications.

4.1. Exponential distribution

Consider the exponential target $f(x) = e^{-x} I(x > 0)$. Here $E[X] = 1$ and $F^{-1}(q) = \log\left\{(1-q)^{-1}\right\}$, which we use to evaluate finite sample confidence intervals obtained via fixed-width methods. We sampled from $f(x)$ using an independence Metropolis sampler with an Exp(1/2) proposal, noting the chain is geometrically ergodic (Jones and Hobert (2001)).

We estimated of $E[X]$ using each combination of $T_i(\epsilon)$ for $i \in \{1, 2, 3\}$ and $\epsilon \in \{0.10, 0.05, 0.02\}$. The chain was started from 1 and run for a minimum of $n^* = 1,000$ iterations. If the stopping criteria was not met, an additional 500 iterations were added to the chain before checking again. The simulation was repeated for 2,000 replications to evaluate the resulting coverage probabilities.

Table 1 summarizes the mean and standard deviation of the number of iterations at termination along with the resulting coverage probabilities. All the coverage probabilities are close to the 0.90 nominal level suggesting all three stopping rules are preforming well.

Consider estimation of the median, $\xi_{.5}$, using the same simulation settings. Table 1 summarizes the results from 2,000 replications. The results are close to the 0.90 nominal level, though slightly lower than those for estimating the mean. Here we have $\xi_{.5} = 0.693$ and $\sqrt{0.5(1-0.5)}/e^{-\xi_{.5}} = 1$, hence for fixed ϵ we expect $T_1(\epsilon)$ and $T_3(\epsilon)$ to be similar and $T_2(\epsilon)$ to be larger.

Table 1. Summary of coverage probabilities for estimation of $E[X]$ and $\xi_{.5}$ based on 2,000 replications and 0.90 nominal level.

	Length (SD)	$E[X]$	Length (SD)	$\xi_{.5}$
$T_1(0.10)$	2.44E3 (4.9E2)	0.8840	2.70E3 (5.9E2)	0.8580
$T_1(0.05)$	8.89E3 (1.2E3)	0.8940	1.01E4 (1.5E3)	0.8805
$T_1(0.02)$	5.36E4 (4.7E3)	0.8875	6.17E4 (5.4E3)	0.8775
$T_2(0.10)$	2.44E3 (4.8E2)	0.8895	5.40E3 (9.4E2)	0.8800
$T_2(0.05)$	8.90E3 (1.2E3)	0.8910	2.07E4 (2.4E3)	0.8820
$T_2(0.02)$	5.35E4 (4.7E3)	0.8870	1.29E5 (9.1E3)	0.8830
$T_3(0.10)$	2.45E3 (4.7E2)	0.8885	2.79E3 (5.2E2)	0.8650
$T_3(0.05)$	8.90E3 (1.2E3)	0.8880	1.03E4 (1.3E3)	0.8820
$T_3(0.02)$	5.35E4 (4.6E3)	0.8895	6.23E4 (5.2E3)	0.8770

Table 2. Summary of coverage probabilities for estimation of Φ based on 2,000 replications. Individual confidence intervals have a 0.9655 nominal level, resulting in a 0.90 nominal level confidence region.

	Length (SD)	$E[X]$	$\xi_{.1}$	$\xi_{.9}$	Region
$T_1(0.10)$	2.88E4 (3.9E3)	0.963	0.989	0.963	0.930
$T_1(0.05)$	1.07E5 (9.7E3)	0.965	0.979	0.962	0.923
$T_1(0.02)$	6.53E5 (3.3E4)	0.965	0.967	0.968	0.917
$T_2(0.10)$	6.71E4 (5.9E3)	0.969	0.979	0.964	0.925
$T_2(0.05)$	2.29E5 (1.4E4)	0.966	0.974	0.963	0.920
$T_2(0.02)$	1.29E6 (5.0E4)	0.964	0.963	0.970	0.915
$T_3(0.10)$	1.00E4 (0)	0.962	0.991	0.955	0.927
$T_3(0.05)$	2.31E4 (2.9E3)	0.963	0.983	0.958	0.921
$T_3(0.02)$	1.30E5 (9.1E3)	0.961	0.970	0.965	0.914

Consider estimating the mean and an 80% Bayesian credible region simultaneously, which we denote as $\Phi = (E[X], \xi_{.1}, \xi_{.9})$. Due to increased computation time, each chain was run for a minimum of $n^* = 10,000$ iterations with an additional 5,000 added between checks. The simulation was terminated the first time the length of a confidence interval was sufficiently small for each parameter in Φ. To adjust for multiplicity, we applied a Bonferonni approach. Setting individual confidence intervals to have a coverage probability of $0.90^{1/3} = 0.9655$, resulting in a simultaneous confidence region with coverage probability of at least 0.90.

The simulation was repeated for 2,000 replications with each combination of $T_i(\epsilon)$ for $i \in \{1, 2, 3\}$ and $\epsilon \in \{0.10, 0.05, 0.02\}$. Table 2 summarizes the results. We can see the individual coverage probabilities improve as ϵ decreases, especially in the case of $\xi_{.1}$. For $\epsilon = 0.02$, all the individual coverage probabilities are close to the nominal level of 0.9655. The observed confidence region coverage probabilities are above the 0.90 nominal level; this is unsurprising due to correlation between parameters in Φ.

4.2. Mixture of bivariate normals

Consider a mixture of bivariate Normals $\mathbf{X} = [X_1, X_2]^T = p\mathbf{Y}_1 + (1-p)\mathbf{Y}_2$, where

$$\mathbf{Y}_1 = \begin{bmatrix} Y_{11} \\ Y_{12} \end{bmatrix} \sim \mathbb{N}_2 \left(\begin{bmatrix} \mu_{11} \\ \mu_{12} \end{bmatrix}, \begin{bmatrix} \sigma_{11}^2 & 0 \\ 0 & \sigma_{12}^2 \end{bmatrix} \right) \quad \text{and} \quad \mathbf{Y}_2 = \begin{bmatrix} Y_{21} \\ Y_{22} \end{bmatrix} \sim \mathbb{N}_2 \left(\begin{bmatrix} \mu_{21} \\ \mu_{22} \end{bmatrix}, \begin{bmatrix} \sigma_{21}^2 & 0 \\ 0 & \sigma_{22}^2 \end{bmatrix} \right).$$

In this example, we take $p = 0.25$, $\mu_{11} = 1$, $\mu_{12} = 10$, $\mu_{21} = 2.5$, $\mu_{22} = 25$, $\sigma_{11} = 0.5$, $\sigma_{12} = 5$, $\sigma_{21} = 0.7$ and $\sigma_{22} = 7$.

We first sampled from $f(\mathbf{X})$ with two different component-wise Metropolis random walk algorithms, one with Uniform proposals and another with Normal proposals. For the Uniform proposals, we applied a $Unif(-3, 3)$ and $Unif(-30, 30)$ random walk for the X_1 and X_2 dimensions, respectively. For the Normal proposals, we applied a $N(0, 3^2)$ and $N(0, 30^2)$ random walk for the X_1 and X_2 dimensions, respectively. It can be shown that these chains are geometrically ergodic (Jarner and Hansen (2000)).

Consider estimation of $\Phi = (E[X], \xi_{.1}, \xi_{.9})$ using fixed-width stopping rules $T_i(\epsilon)$ for $i \in \{1, 2, 3\}$ and $\epsilon \in \{0.10, 0.05, 0.02\}$. We ran the chain for a minimum of $n^* = 5,000$ iterations and added 1,000 iterations between checking the stopping criteria. This simulation was repeated for 1,000 independent replications.

Table 3 summarizes the mean and standard deviation of the number of iterations at termination along with empirical coverage probabilities from the Uniform and Normal proposals. For both samplers, the coverage probabilities improved as ϵ decreased and were close to the 0.95 nominal level once $\epsilon = 0.02$. It appears the Metropolis random walk with Normal proposals was mixing faster since the overall simulation effort was substantially lower than that of the Uniform proposals. This difference in simulation effort illustrates the importance of specifying a good proposal distribution in MCMC simulations.

Consider a Gibbs sampler using the full conditional densities,

$$f_{X_1|X_2}(x_1|x_2) = P_{X_2} Y_{11} + (1 - P_{X_2}) Y_{21} \text{ and}$$
$$f_{X_2|X_1}(x_2|x_1) = P_{X_1} Y_{12} + (1 - P_{X_1}) Y_{22} ,$$

where

$$P_{X_2} = \left(1 + \frac{(1-p)\sigma_{12}}{p\sigma_{22}} \exp \left\{ \frac{1}{2} \left(\left(\frac{x_2 - \mu_{12}}{\sigma_{12}} \right)^2 - \left(\frac{x_2 - \mu_{22}}{\sigma_{22}} \right)^2 \right) \right\} \right)^{-1},$$

$$P_{X_1} = \left(1 + \frac{(1-p)\sigma_{11}}{p\sigma_{21}} \exp \left\{ \frac{1}{2} \left(\left(\frac{x_1 - \mu_{11}}{\sigma_{11}} \right)^2 - \left(\frac{x_1 - \mu_{21}}{\sigma_{21}} \right)^2 \right) \right\} \right)^{-1}.$$

Table 3. Summary of coverage probabilities for estimations of Φ using a Metropolis random walk with Uniform and Normal proposals based on 1,000 replications and a 0.95 nominal level.

Uniform	Length (SD)	$E[X_1]$	$\xi_{.1,X_1}$	$\xi_{.9,X_1}$	$E[X_2]$	$\xi_{.1,X_2}$	$\xi_{.9,X_2}$
$T_1(0.10)$	14,658 (3.4E3)	0.930	0.932	0.917	0.936	0.945	0.937
$T_1(0.05)$	59,869 (9.1E3)	0.934	0.922	0.939	0.940	0.934	0.953
$T_1(0.02)$	391,566 (3.1E4)	0.956	0.944	0.945	0.956	0.948	0.953
$T_2(0.10)$	20,897 (5.0E3)	0.929	0.933	0.911	0.931	0.936	0.938
$T_2(0.05)$	85,401 (1.2E4)	0.950	0.926	0.934	0.929	0.925	0.942
$T_2(0.02)$	556,821 (3.9E4)	0.953	0.946	0.954	0.950	0.938	0.956
$T_3(0.10)$	8,827 (1.0E3)	0.926	0.928	0.899	0.920	0.922	0.920
$T_3(0.05)$	35,733 (2.9E3)	0.924	0.938	0.931	0.934	0.928	0.937
$T_3(0.02)$	233,312 (1.3E4)	0.954	0.955	0.959	0.948	0.958	0.956

Normal	Length (SD)	$E[X_1]$	$\xi_{.1,X_1}$	$\xi_{.9,X_1}$	$E[X_2]$	$\xi_{.1,X_2}$	$\xi_{.9,X_2}$
$T_1(0.10)$	8,028 (1.5E3)	0.946	0.939	0.939	0.934	0.943	0.937
$T_1(0.05)$	29,844 (3.7E3)	0.927	0.936	0.948	0.917	0.932	0.953
$T_1(0.02)$	186,061 (1.3E4)	0.952	0.936	0.952	0.943	0.946	0.938
$T_2(0.10)$	11,307 (2.1E3)	0.949	0.933	0.940	0.940	0.944	0.943
$T_2(0.05)$	42,338 (4.6E3)	0.911	0.943	0.956	0.937	0.934	0.951
$T_2(0.02)$	261,741 (1.6E4)	0.940	0.938	0.956	0.949	0.938	0.945
$T_3(0.10)$	5,114 (3.2E2)	0.944	0.950	0.933	0.936	0.936	0.924
$T_3(0.05)$	17,654 (1.8E3)	0.922	0.930	0.943	0.925	0.921	0.939
$T_3(0.02)$	112,626 (7.6E3)	0.933	0.946	0.941	0.941	0.930	0.940

Here $X_1|X_2 = x_2$ and $X_2|X_1 = x_1$ are easy to sample from since they are mixtures of Normal random variables.

Table 4 summarizes the results for the Gibbs sampler. The coverage probabilities do not improve uniformly as ϵ decreases. However, they are all close to the nominal 0.95 level using significantly fewer total iterations, suggesting that the Gibbs sampler mixes better than either of the Metropolis random walk samplers.

We performed additional simulations via i.i.d. sampling (not shown). The resulting empirical coverage probabilities were similar to these using the Gibbs sampler, albeit with slightly fewer iterations.

4.3. Bayesian logistic regression

The *Anguilla* eel data provided in the `dismo` R package (see e.g. Elith, Leathwick, and Hastie (2008); Hijmans et al. (2010)) consists of 1,000 observations from a New Zealand survey of site-level presence or absence for the short-finned eel (*Anguilla australis*). We selected six out of twelve covariates as in Leathwick et al. (2008). Five were continuous variables: SegSumT, DSDist, USNative, DSMaxSlope and DSSlope; Method was a categorical variable with levels Electric,

Table 4. Summary of coverage probabilities for estimations of Φ using a Gibbs sampler based on 1,000 replications and a 0.95 nominal level.

Gibbs	Length (SD)	$E[X_1]$	$\xi_{.1,X_1}$	$\xi_{.9,X_1}$	$E[X_2]$	$\xi_{.1,X_2}$	$\xi_{.9,X_2}$
$T_1(0.10)$	1,930 (3.7E2)	0.941	0.940	0.937	0.954	0.958	0.927
$T_1(0.05)$	5,727 (8.7E2)	0.946	0.958	0.941	0.942	0.945	0.940
$T_1(0.02)$	31,170 (2.8E3)	0.935	0.945	0.961	0.937	0.937	0.944
$T_2(0.10)$	2,465 (5.4E2)	0.935	0.939	0.939	0.954	0.950	0.937
$T_2(0.05)$	7,865 (1.1E3)	0.950	0.959	0.943	0.955	0.954	0.952
$T_2(0.02)$	43,756 (3.6E3)	0.933	0.936	0.959	0.936	0.959	0.946
$T_3(0.10)$	1,182 (3.9E2)	0.929	0.936	0.942	0.936	0.936	0.924
$T_3(0.05)$	3,786 (6.2E2)	0.956	0.951	0.944	0.940	0.940	0.935
$T_3(0.02)$	20,289 (2.0E3)	0.945	0.947	0.954	0.940	0.943	0.952

Spo, Trap, Net, and Mixture.

Let x_i be the regression vector of covariates for the ith observation of length k, and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_9)$ be the vector regression coefficients. For the ith observation, suppose $Y_i = 1$ denotes presence and $Y_i = 0$ denotes absence of *Anguilla australis*. The Bayesian logistic regression model is given by

$$\begin{aligned} Y_i &\sim Bernoulli(p_i) , \\ p_i &\sim \frac{\exp(x_i^T \boldsymbol{\beta})}{1+\exp(x_i^T \boldsymbol{\beta})} \text{ and,} \\ \boldsymbol{\beta} &\sim N(\mathbf{0}, \sigma_\beta^2 \mathbf{I}_k) , \end{aligned}$$

where $\mathbf{I}_k$ is the $k \times k$ identity matrix. For the analysis, $\sigma_\beta^2 = 100$ was chosen to represent a diffuse prior distribution on $\boldsymbol{\beta}$ (Boone, Merrick, and Krachey (2014)). We used the `MCMClogit` function in the `MCMCpack` package to sample from the target Markov chain.

Suppose interest is in estimating the posterior mean along with an 80% Bayesian credible interval for each regression coefficient in the model. True values are unknown. We ran 1,000 independent chains for 1E6 iterations to obtain an accurate estimate, which we treat as the truth (Table 5).

Consider estimating $\Phi_j = \left(\beta_j, \xi_{.1}^{(j)}, \xi_{.9}^{(j)}\right)$, $j = 0, \ldots, 9$, using fixed-width stopping rules $T_i(\epsilon)$ for $i \in \{1,2,3\}$. We specified an ϵ for $T_1(\epsilon)$ for each Φ_j with respect to its magnitude. We chose three simulation settings such that $\boldsymbol{\epsilon}_1 = $ (1, 0.01, 0.001, 0.1, 0.1, 0.1, 0.1, 0.1, 0.01, 0.01), $0.5\boldsymbol{\epsilon}_1$, and $0.2\boldsymbol{\epsilon}_1$.

We specified an ϵ for $T_1(\epsilon)$ for each Φ_j. In this case, we chose three simulation settings such that $\boldsymbol{\epsilon}_2 = $ (0.1, 0.1, 1, 0.1, 1, 0.1, 0.1, 0.1, 0.1, 1), $0.5\boldsymbol{\epsilon}_2$, and $0.2\boldsymbol{\epsilon}_2$.

For both $T_1(\epsilon)$ and $T_2(\epsilon)$, it is overwhelmingly tedious to specify appropriate ϵ vectors when the number of parameters is large. For the stopping rule $T_3(\epsilon)$

Table 5. Summary of estimated true values with standard errors for the Bayesian logistic regression example.

Variable	β_j		$\xi^{(j)}_{.1}$		$\xi^{(j)}_{.9}$	
Intercept	-10.463	(2.7E-5)	-12.224	(3.9E-4)	-8.730	(3.7E-4)
SegSumT	0.657	(1.5E-5)	0.559	(2.1E-5)	0.757	(2.2E-5)
DSDist	-4.02E-3	(3.3E-7)	-6.15E-3	(4.9E-7)	-1.93E-3	(4.4E-7)
USNative	-1.170	(7.1E-5)	-1.625	(9.9E-5)	-0.718	(1.0E-4)
MethodMixture	-0.468	(6.8E-5)	-0.910	(9.8E-5)	-0.028	(9.8E-5)
MethodNet	-1.525	(8.2E-5)	-2.026	(1.2E-4)	-1.035	(1.1E-4)
MethodSpo	-1.831	(1.3E-4)	-2.623	(2.2E-4)	-1.798	(1.4E-4)
MethodTrap	-2.594	(1.1E-4)	-3.285	(1.8E-4)	-1.937	(1.3E-4)
DSMaxSlope	-0.170	(1.1E-5)	-0.244	(1.7E-5)	-0.099	(1.5E-5)
USSlope	-0.052	(3.7E-6)	-0.076	(5.5E-6)	-0.028	(5.1E-6)

we used a single ϵ for the 30-dimensional target parameter vector. Specifically, we chose three simulation settings such that $\epsilon_3 \in \{0.10, 0.05, 0.02\}$.

For the two larger ϵ settings, we set $n^* = 10,000$ and added 1,000 iterations between checks. For the smallest ϵ setting, we set $n^* = 1\text{E}5$ and added 10,000 iterations between checks due to increased computational demands. Each simulation setting was independently repeated 1,000 times.

Table 6 summarizes the empirical coverage probabilities. Here the coverage probabilities for each stopping rule increase toward the nominal level of 0.95 as ϵ decreases, suggesting that all the stopping rules perform well. For high-dimensional settings such as this, $T_3(\epsilon)$ is advantageous since a practitioner can specify a single ϵ value.

To adjust for multiplicity, we again applied a Bonferonni approach. We set individual confidence intervals to have a nominal level of $0.80^{1/10} = 0.9779$ resulting in simultaneous confidence region with nominal level of at least 0.80. We only considered estimating the posterior mean of the 10-dimensional vector $\boldsymbol{\beta}$ using $T_3(\epsilon)$ with $\epsilon \in \{0.20, 0.10, 0.05, 0.02\}$. The minimum simulation effort was $n^* = 1E5$ iterations with an additional 1,000 added between checks. Again, for the smallest ϵ setting, we set $n^* = 1E6$ with an additional 10,000 added between checks. The simulation was terminated the first time $T_3(\epsilon)$ was met and independently repeated 1,000 times.

Table 7 summarizes the simulation results. We can see that, as ϵ decreases, all the individual coverage probabilities are close to the nominal level of 0.9779, and the observed confidence region coverage probabilities approach the nominal level of 0.80. The last was bit surprising in how close to the nominal 0.80 level this was given possible correlation among parameters. We investigated the correlation between pairs of target parameters and found that most pairs had low correlation, except for strong correlation between Intercept and SegSumT

Table 6. Summary of coverage probabilities for Bayesian logistic regression example with 1,000 independent replicates. The coverage probabilities have a 0.95 nominal level.

	$T_1(\boldsymbol{\epsilon}_1)$			$T_1(0.5\boldsymbol{\epsilon}_1)$			$T_1(0.2\boldsymbol{\epsilon}_1)$		
Variable	β_j	$\xi_{.1}^{(j)}$	$\xi_{.9}^{(j)}$	β_j	$\xi_{.1}^{(j)}$	$\xi_{.9}^{(j)}$	β_j	$\xi_{.1}^{(j)}$	$\xi_{.9}^{(j)}$
Intercept	0.936	0.933	0.912	0.937	0.942	0.942	0.946	0.946	0.930
SegSumT	0.932	0.922	0.916	0.942	0.941	0.934	0.953	0.944	0.936
DSDist	0.987	0.969	0.979	0.976	0.969	0.960	0.956	0.954	0.952
USNative	0.927	0.929	0.917	0.939	0.933	0.943	0.948	0.939	0.944
MethodMixture	0.930	0.928	0.920	0.946	0.948	0.938	0.935	0.953	0.940
MethodNet	0.946	0.922	0.936	0.941	0.948	0.932	0.943	0.939	0.935
MethodSpo	0.913	0.913	0.927	0.931	0.929	0.931	0.943	0.942	0.926
MethodTrap	0.928	0.906	0.937	0.938	0.930	0.927	0.941	0.947	0.947
DSMaxSlope	0.932	0.930	0.921	0.942	0.943	0.945	0.953	0.958	0.951
USSlope	0.921	0.928	0.935	0.951	0.927	0.954	0.957	0.952	0.962
Length (SD)	19,521 (3.8E3)			76,894 (9.5E3)			492,910 (3.4E4)		

	$T_2(\boldsymbol{\epsilon}_2)$			$T_2(0.5\boldsymbol{\epsilon}_2)$			$T_2(0.2\boldsymbol{\epsilon}_2)$		
Variable	β_j	$\xi_{.1}^{(j)}$	$\xi_{.9}^{(j)}$	β_j	$\xi_{.1}^{(j)}$	$\xi_{.9}^{(j)}$	β_j	$\xi_{.1}^{(j)}$	$\xi_{.9}^{(j)}$
Intercept	0.928	0.938	0.915	0.950	0.948	0.947	0.945	0.949	0.938
SegSumT	0.923	0.916	0.937	0.953	0.955	0.948	0.944	0.947	0.947
DSDist	0.985	0.968	0.975	0.970	0.958	0.958	0.956	0.955	0.947
USNative	0.921	0.936	0.921	0.946	0.933	0.945	0.940	0.956	0.941
MethodMixture	0.941	0.938	0.933	0.942	0.945	0.916	0.935	0.933	0.942
MethodNet	0.942	0.920	0.922	0.940	0.942	0.939	0.942	0.944	0.935
MethodSpo	0.919	0.901	0.924	0.936	0.923	0.937	0.947	0.956	0.947
MethodTrap	0.935	0.910	0.936	0.939	0.939	0.931	0.941	0.933	0.941
DSMaxSlope	0.937	0.942	0.916	0.948	0.942	0.950	0.942	0.954	0.955
USSlope	0.935	0.933	0.930	0.949	0.936	0.941	0.949	0.944	0.943
Length (SD)	37,667 (3.5E4)			151,276 (8.9E4)			1,161,400 (2.6E5)		

	$T_3(0.10)$			$T_3(0.05)$			$T_3(0.02)$		
Variable	β_j	$\xi_{.1}^{(j)}$	$\xi_{.9}^{(j)}$	β_j	$\xi_{.1}^{(j)}$	$\xi_{.9}^{(j)}$	β_j	$\xi_{.1}^{(j)}$	$\xi_{.9}^{(j)}$
Intercept	0.932	0.944	0.929	0.943	0.950	0.943	0.943	0.954	0.934
SegSumT	0.932	0.935	0.941	0.942	0.934	0.946	0.942	0.934	0.946
DSDist	0.981	0.969	0.969	0.968	0.966	0.955	0.957	0.954	0.950
USNative	0.939	0.942	0.923	0.941	0.948	0.954	0.942	0.943	0.940
MethodMixture	0.939	0.928	0.920	0.947	0.943	0.933	0.927	0.947	0.928
MethodNet	0.929	0.922	0.931	0.939	0.939	0.934	0.930	0.938	0.939
MethodSpo	0.915	0.902	0.925	0.924	0.933	0.926	0.948	0.946	0.935
MethodTrap	0.930	0.909	0.920	0.941	0.937	0.933	0.939	0.935	0.948
DSMaxSlope	0.941	0.932	0.930	0.940	0.950	0.943	0.958	0.955	0.951
USSlope	0.939	0.928	0.940	0.953	0.937	0.955	0.954	0.957	0.958
Length (SD)	24,404 (1.4E3)			78,886 (4.2E3)			439,260 (1.7E4)		

Table 7. Summary of coverage probabilities for $\boldsymbol{\beta}$ based on $T_3(\epsilon)$ with 1,000 replicates. The coverage probabilities have a 0.9779 nominal level, resulting in a 0.80 nominal level confidence region.

	$T_3(0.20)$	$T_3(0.10)$	$T_3(0.05)$	$T_3(0.02)$
Variable	β_j	β_j	β_j	β_j
Intercept	0.959	0.975	0.976	0.973
SegSumT	0.960	0.971	0.979	0.974
DSDist	0.995	0.989	0.993	0.979
USNative	0.948	0.978	0.970	0.973
MethodMixture	0.950	0.973	0.967	0.968
MethodNet	0.962	0.962	0.976	0.973
MethodSpo	0.946	0.954	0.968	0.979
MethodTrap	0.950	0.960	0.970	0.978
DSMaxSlope	0.966	0.971	0.977	0.974
USSlope	0.964	0.965	0.973	0.982
Region	0.693	0.763	0.792	0.805
Length (SD)	10,082(2.7E2)	29,729(1.8E3)	100,261(5.2E3)	583,488(1.9E4)

and moderate correlation between USNative and USSlope. Given the lack of correlation, the confidence region coverages are encouraging.

4.4. Discussion

We advocate use of a relative standard deviation stopping rule since it is easy to implement and applicable in multivariate settings without a priori knowledge of the target parameter size. The rule terminates an MCMC simulation when estimates of target parameters are sufficiently accurate relative to their associated posterior standard deviations, with estimates approximately ϵ^{-1} more accurate than their posterior standard deviations. Using $\epsilon = 0.02$ has provided excellent results in the wide variety of examples considered here. A smaller ϵ may be appropriate when the accuracy of estimation is critical.

When estimating multiple quantities simultaneously, we controlled the width of each of the marginal confidence intervals. Alternatively, one could consider multiple quantities jointly by controlling the volume of confidence region, a subject of ongoing research. In this setting, one should be able to establish asymptotic validity for a relative fixed-volume approach using techniques presented here and in Glynn and Whitt (1992).

In any MCMC simulation, a key component is choosing a Markov chain that mixes well while sufficiently exploring the state space. Moreover, the computational effort to achieve a given accuracy depends on the sampling scheme. We have offered limited guidance in this direction. In this regard, see Brooks et al. (2010) and the references therein.

Our examples considered only BM to estimate the asymptotic variance from a CLT. Improving the variance estimation step might be possible using such alternative methods as overlapping batch means, spectral variance, or subsampling bootstrap methods (Flegal and Jones (2010); Flegal (2012); Doss et al. (2014)) that are currently available in the `mcmcse` package.

Acknowledgements

The authors are grateful to Brian Caffo and Galin Jones for helpful conversations about this paper. The authors also thank the referees and an associate editor for their constructive comments which resulted in many improvements. The first author's work is partially supported by NSF grant DMS-13-08270.

Appendix A. Proof of Theorem 1

The proof follows techniques introduced in Glynn and Whitt (1992). Define $z = z_{\delta/2}$ and recall that $T_3(\epsilon) \to \infty$ w.p.1 as $\epsilon \to 0$. Since $\hat{\sigma}_n \to \sigma_\theta$ w.p.1 as $n \to \infty$, we have $\hat{\sigma}_{T_3(\epsilon)} \to \sigma_\theta$ w.p.1 as $\epsilon \to 0$; since $\hat{\lambda}_n \to \lambda_\theta$ w.p.1 as $n \to \infty$, we have $\hat{\lambda}_{T_3(\epsilon)} \to \lambda_\theta$ w.p.1 as $\epsilon \to 0$.

Take $V(n) = 2z\hat{\sigma}_n/\sqrt{n} + p(n)$, where $p(n) = o(n^{-1/2})$. Then $T_3(\epsilon)$ can be written as $T_3(\epsilon) = \inf\left\{n \geq 0 : V(n) \leq \epsilon\hat{\lambda}_n\right\}$. As $\sigma_\theta^2 \in (0, \infty)$, it is easy to verify that

$$n^{1/2}V(n) \to 2z\sigma_\theta > 0 \text{ w.p.1 as } \ n \to \infty. \tag{A.1}$$

By definition of $T_3(\epsilon)$, $V(T_3(\epsilon)-1) > \epsilon\hat{\lambda}_{T_3(\epsilon)-1}$ and $V(T_3(\epsilon)) \leq \epsilon\hat{\lambda}_{T_3(\epsilon)}$. Using (A.1) we have

$$\limsup_{\epsilon\to 0} \epsilon T_3(\epsilon)^{1/2} \leq \limsup_{\epsilon\to 0} \frac{T_3(\epsilon)^{1/2}V(T_3(\epsilon)-1)}{\hat{\lambda}_{T_3(\epsilon)-1}} = \frac{2z\sigma_\theta}{\lambda_\theta} \text{ w.p.1.}$$

Similarly,

$$\liminf_{\epsilon\to 0} \epsilon T_3(\epsilon)^{1/2} \geq \liminf_{\epsilon\to 0} \frac{T_3(\epsilon)^{1/2}V(T_3(\epsilon))}{\hat{\lambda}_{T_3(\epsilon)}} = \frac{2z\sigma_\theta}{\lambda_\theta} \text{ w.p.1.}$$

Thus, we have

$$\lim_{\epsilon\to 0} \epsilon T_3(\epsilon)^{1/2} = \frac{2z\sigma_\theta}{\lambda_\theta} \text{ w.p.1.} \tag{A.2}$$

Using (A.2) with properties of $\hat{\sigma}_{T_3(\epsilon)}$ and $\hat{\lambda}_{T_3(\epsilon)}$, we have

$$\lim_{\epsilon\to 0} \frac{\epsilon^{-1}T_3(\epsilon)^{-1/2}2z\hat{\sigma}_{T_3(\epsilon)}}{\hat{\lambda}_{T_3(\epsilon)}} = 1 \text{ w.p.1.} \tag{A.3}$$

Let $\beta = 2z\sigma_\theta/\lambda_\theta$ and set $\tau_\epsilon(t) = T_3(\epsilon)\epsilon^2\beta^{-2}t$ for $t \geq 0$. Note that $\tau_\epsilon \to e$ as $\epsilon \to 0$ w.p.1 pointwise, where $e(t) = t$. Then it follows from the FCLT and a standard random-time-change argument (p. 151 Billingsley (1999)) that

$$\mathcal{Z}_{\epsilon^2\beta^{-2}}(\tau_\epsilon(1)) \xrightarrow{d} \frac{\sigma_\theta B(e(1))}{e(1)} = \sigma_\theta B(1) \text{ as } \epsilon \to 0 \;, \tag{A.4}$$

where $\mathcal{Z}_{\epsilon^2\beta^{-2}}(\tau_\epsilon(1)) = \beta\epsilon^{-1}\left(Z_{T_3(\epsilon)} - \theta\right)$. Slutsky's theorem with (A.3) and (A.4) yield $T_3(\epsilon)^{1/2}/\hat{\sigma}_{T_3(\epsilon)}\left(Z_{T_3(\epsilon)} - \theta\right) \xrightarrow{d} B(1)$ as $\epsilon \to 0$. Finally, we have

$$\begin{aligned}\Pr\left(\theta \in C[T_3(\epsilon)]\right) &= \Pr\left(Z_{T_3(\epsilon)} - \theta \in \left(-\frac{z\hat{\sigma}_{T_3(\epsilon)}}{T_3(\epsilon)^{1/2}},\ \frac{z\hat{\sigma}_{T_3(\epsilon)}}{T_3(\epsilon)^{1/2}}\right)\right)\\ &= \Pr\left(\frac{T_3(\epsilon)^{1/2}}{\hat{\sigma}_{T_3(\epsilon)}}(Z_{T_3(\epsilon)} - \theta)) \in (-z,\ z)\right) \to 1-\delta \text{ as } \epsilon \to 0.\end{aligned}$$

References

Bednorz, W. and Latuszyński, K. (2007). A few remarks on 'Fixed-width output analysis for Markov chain Monte Carlo' by Jones et al. *J. Amer. Statist. Assoc.* **102**, 1485-1486.

Billingsley, P. (1999). *Convergence of Probability Measures*. Second Edition. Wiley, New York.

Boone, E. L., Merrick, J. R. and Krachey, M. J. (2014). A Hellinger distance approach to MCMC diagnostics. *J. Statist. Comput. Simulation* **84**, 833-849.

Brooks, S., Gelman, A., Jones, G. and Meng, X. (2010). *Handbook of Markov Chain Monte Carlo: Methods and Applications*. Chapman & Hall.

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

Chan, K. S. and Geyer, C. J. (1994). Comment on "Markov chains for exploring posterior distributions". *Ann. Statist.* **22**, 1747-1758.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *J. Amer. Statist. Assoc.* **91**, 883-904.

Cowles, M. K., Roberts, G. O. and Rosenthal, J. S. (1999). Possible biases induced by MCMC convergence diagnostics. *J. Statist. Comput. Simulation* **64**, 87-104.

Doss, C., Flegal, J. M., Jones, G. L. and Neath, R. C. (2014). Markov chain Monte Carlo estimation of quantiles. *Electronic J. Statist.* **8**, 2448-2478.

Douc, R., Fort, G., Moulines, E. and Soulier, P. (2004). Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.* **14**, 1353-1377.

Elith, J., Leathwick, J. and Hastie, T. (2008). A working guide to boosted regression trees. *J. Animal Ecology* **77**, 802-813.

Flegal, J. M. (2012). Applicability of subsampling bootstrap methods in Markov chain Monte Carlo. In *Monte Carlo and Quasi-Monte Carlo Methods* (Edited by H. Wozniakowski and L. Plaskota), 363-372. Springer Proceedings in Mathematics & Statistics.

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

Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.* **38**, 1034-1070.

Fort, G. and Moulines, E. (2000). V-subgeometric ergodicity for a Hastings-Metropolis algorithm. *Statist. Probab. Lett.* **49**, 401-410.

Fort, G. and Moulines, E. (2003). Polynomial ergodicity of Markov transition kernels. *Stochastic Process. Appl.* **103**, 57-99.

Fristedt, B. and Gray, L. F. (1997). *A Modern Approach to Probability Theory.* Birkhauser Verlag.

Glynn, P. W. and Whitt, W. (1992). The asymptotic validity of sequential stopping rules for stochastic simulations. *Ann. Appl. Probab.* **2**, 180-198.

Hijmans, R., Phillips, S., Leathwick, J. and Elith, J. (2010). dismo: species distribution modeling. r package version 0.5-4.

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

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.

Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory Probab. Appl.* **7**, 349-382.

Jarner, S. F. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85**, 341–361.

Jarner, S. F. and Roberts, G. O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* **12**, 224-247.

Jarner, S. F. and Roberts, G. O. (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Statist.* **24**, 101-121.

Jarner, S. F. and Tweedie, R. L. (2003). Necessary conditions for geometric and polynomial ergodicity of random-walk-type Markov chains. *Bernoulli* **9**, 559-578.

Johnson, A. A., Jones, G. L. and Neath, R. C. (2011). Component-wise Markov chain Monte Carlo. Preprint.

Jones, G. L. (2004). On the Markov chain central limit theorem. *Probab. Surveys* **1**, 299-320.

Jones, G. L., Haran, M., Caffo, B. S. and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **101**, 1537-1547.

Jones, G. L. and Hobert, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statist. Sci.* **16**, 312-334.

Kim, T. Y. and Lee, S. (2005). Kernel density estimator for strong mixing processes. *J. Statist. Plann. Inference* **133**, 273-284.

Leathwick, J., Elith, J., Chadderton, W., Rowe, D. and Hastie, T. (2008). Dispersal, disturbance and the contrasting biogeographies of New Zealands diadromous and non-diadromous fish species. *J. Biogeography* **35**, 1481-1497.

Mengersen, K. and Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24**, 101-121.

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

Mykland, P., Tierney, L. and Yu, B. (1995). Regeneration in Markov chain samplers. *J. Amer. Statist. Assoc.* **90**, 233-241.

Oodaira, H. and Yoshihara, K.-i. (1972). Functional central limit theorems for strictly stationary processes satisfying the strong mixing condition. *Kodai Math. Seminar Reports* **24**, 259-269.

Raftery, A. E. and Lewis, S. M. (1992). Comment on "The Gibbs sampler and Markov chain Monte Carlo". *Statist. Sci.* **7**, 493-497.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods.* Springer, New York, second edition.

Roberts, G. O. and Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surveys* **1**, 20-71.

Roberts, G. O. and Tweedie, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83**, 95-110.

Sen, P. K. (1972). On the Bahadur representation of sample quantiles for sequences of ϕ-mixing random variables. *J. Multivariate Anal.* **2**, 77-95.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.* **22**, 1701-1762.

Yu, B. (1993). Density estimation in the L^∞ norm for dependent data with applications to the Gibbs sampler. *Ann. Statist.* **21**, 711-735.

Department of Statistics, University of California Riverside, Riverside, CA 92521, USA.

E-mail: jflegal@ucr.edu

Department of Statistics, University of California Riverside, Riverside, CA 92521, USA.

E-mail: lei.gong@email.ucr.edu

(Received July 2013; accepted April 2014)