

3-2014

The Number of MCMC Draws Needed to Compute Bayesian Credible Bounds

Jia Liu
Pfizer

Daniel J. Nordman
Iowa State University, dnordman@iastate.edu

William Q. Meeker
Iowa State University, wqmeeker@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/stat_las_preprints

 Part of the [Statistics and Probability Commons](#)

Recommended Citation

Liu, Jia; Nordman, Daniel J.; and Meeker, William Q., "The Number of MCMC Draws Needed to Compute Bayesian Credible Bounds" (2014). *Statistics Preprints*. 130.
http://lib.dr.iastate.edu/stat_las_preprints/130

This Article is brought to you for free and open access by the Statistics at Iowa State University Digital Repository. It has been accepted for inclusion in Statistics Preprints by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

The Number of MCMC Draws Needed to Compute Bayesian Credible Bounds

Abstract

Bayesian credible bounds produced from Markov chain Monte Carlo (MCMC) procedures contain Monte Carlo error and thus may require a long chain in order to have a reasonable degree of repeatability. This is especially true when there is a substantial amount of autocorrelation in the chain realization. Repeatability would be important in some applications where it would be undesirable to report numerical values containing substantial Monte Carlo error in the least significant digits. The endpoints of a credible interval correspond to quantiles of the empirical distribution of the MCMC draws from the marginal posterior distribution of the quantity of interest. Our goal is to provide an algorithm to choose the number of MCMC draws that will provide, with high probability, a specified amount of precision (i.e., the number of correct significant digits) in the interval endpoints.

Keywords

Monte Carlo error, nonparametric quantile estimation, precision

Disciplines

Statistics and Probability

Comments

This preprint was published as Jia Liu, Daniel J. Nordman, and William Q. Meeker, "The Number of MCMC Draws Needed to Compute Bayesian Credible Bounds".

The Number of MCMC Draws Needed to Compute Bayesian Credible Bounds

Jia Liu* Daniel J. Nordman[†] William Q. Meeker[‡]

*Jia Liu is Statistician, Pfizer Inc., Groton, CT 06340 (Email: Jia.Liu2@pfizer.com)

[†]Daniel J. Nordman is Associate Professor, Department of Statistics, Iowa State University, Ames, IA 50011 (Email: dnordman@iastate.edu)

[‡]William Q. Meeker is Distinguished Professor, Department of Statistics, Iowa State University, Ames, IA 50011 (Email: wqmeeker@iastate.edu)

Abstract

Bayesian credible bounds produced from Markov chain Monte Carlo (MCMC) procedures contain Monte Carlo error and thus may require a long chain in order to have a reasonable degree of repeatability. This is especially true when there is a substantial amount of autocorrelation in the chain realization. Repeatability would be important in some applications where it would be undesirable to report numerical values containing substantial Monte Carlo error in the least significant digits. The endpoints of a credible interval correspond to quantiles of the empirical distribution of the MCMC draws from the marginal posterior distribution of the quantity of interest. Our goal is to provide an algorithm to choose the number of MCMC draws that will provide, with high probability, a specified amount of precision (i.e., the number of correct significant digits) in the interval endpoints.

Keywords: Monte Carlo error; nonparametric quantile estimation; precision

1 Introduction

1.1 Background and Motivation

Markov chain Monte Carlo (MCMC) methods are commonly used for Bayesian inference computations, for numerically obtaining draws (or approximating draws) from a posterior distribution of interest. Such computations are often efficient and easy to implement, even for complicated data and model combinations. Additionally, MCMC procedures are particularly helpful for sampling posterior distributions that are very complicated and sometimes of high dimension. Metropolis et al. (1953) first introduced the idea of MCMC methods for evaluating complex integrals arising from physical problems. The integrals were restated as expectations of random variables having a distribution function $f(\cdot)$ and then samples were generated from $f(\cdot)$ to estimate the expectations. Hastings (1970) generalized the method to solve statistical problems. Given a distribution function $f(\cdot)$ that needs to be evaluated, the MCMC procedure constructs a Markov chain having a stationary distribution $f(\cdot)$. After a large number of draws the chain is then used to estimate particular functions of the parameters of $f(\cdot)$. The sample draws can be used to make inferences for unknown quantities of interest, based on the data and prior distribution specification. The quantity of interest can be a parameter in a model or some function of parameters.

There are several different MCMC algorithms that can be used to obtain the sample draws. These include the Metropolis algorithm by Metropolis et al. (1953), the Metropolis-Hastings algorithm by Hastings (1970) and the Gibbs sampler algorithm described by Geman and Geman (1984) and Gelfand and Smith (1990).

A credible interval can be used to quantify the statistical uncertainty of

the unknown quantity of interest. The end points of the credible interval are defined by the quantiles of the empirical distribution of the MCMC draws from the marginal posterior distribution. Because MCMC methods involve random sampling, we would not expect to obtain the same set of MCMC draws each time that we run the chain. However, one may be interested in having a certain degree of precision that would provide a specified amount of repeatability for the quantile estimates in terms of Monte Carlo error. Usually there is more variability in the tail of the marginal posterior distribution, thus a large number of MCMC draws may be needed in order to achieve a desired degree of precision for the end points of Bayesian credible intervals. This suggests the need for a method to choose the number of draws required to estimate the quantile with certain degrees of precision. The main idea of the method is to obtain a pilot stretch of MCMC draws which can be treated as an approximately stationary realization, and then apply state-of-the-art techniques for quantile estimation of a stationary, weakly dependent time process. This leads to a procedure for estimating MCMC sample sizes for Bayesian credible intervals of desired precision. The method provided in this paper can be applied to any MCMC algorithm.

1.2 Literature Review

A number of books and papers have described different MCMC algorithms for Bayesian computations. Gelfand and Smith (1990) reviewed and compared three sampling approaches, which are stochastic substitution, the Gibbs sampler, and the sampling-importance-resampling algorithm, for different model structures in applications. Geyer (1992) suggested one long run of the Markov chain

and estimated variances based on MCMC output by using window estimators, batch means, and specialized Markov chain estimators. Smith and Roberts (1993) reviewed implementations of the Gibbs sampler with some examples and also briefly described other MCMC methods, such as the Metropolis-Hastings algorithm. Tierney (1994) outlined the Gibbs sampler and Metropolis algorithm for constructing Markov chains with some theoretical results and implementation issues, including how to determine the run length. Chib and Greenberg (1995) provided a detailed description of the Metropolis-Hastings algorithm, addressing some implementation issues and illustrated the method using two examples. Athreya and Lahiri (2006) provided an overview of the theory behind Markov Chains and MCMC methods.

There is also previous work that introduced various ways to obtain nonparametric confidence intervals for quantiles. Woodruff (1952) proposed a method of obtaining confidence intervals for medians and other position measures by inverting the end points of confidence intervals for the corresponding distribution function under any sampling scheme. Sitter and Wu (2001) assessed the performance of confidence intervals for quantiles by varying tail probabilities. Gilat and Hill (1996) derived distribution-free confidence intervals for quantiles from any distribution based on order statistics and i.i.d. assumptions. Chen and Hall (1993) used smoothed empirical likelihood confidence intervals for quantiles and showed the procedure had good coverage properties.

Most directly relevant to our work, Raftery and Lewis (1992) proposed a method for computing the total number of MCMC draws as well as the length of “burn-in” period when the tail probability of the posterior distribution of a function is to be estimated within certain degree of precision. They focus on the precision for estimating a tail probability, and not the quantile itself,

which is a different quantity and the scales of the two quantities (probabilities and quantiles) are difficult to relate in terms of precision. In fact, the approach of Raftery and Lewis (1992) translates an MCMC draw into binary (0-1) time series, which may be useful for estimating proportions but presents a loss of information for other quantities, such as quantiles. Our focus is on estimating a particular quantile with a specified degree of precision, which has practical interpretation for the final interval estimator itself that is often of direct interest.

1.3 Overview

The rest of this paper is organized as follows. Section 2 summarizes the structure of MCMC algorithm output. Section 3 describes the details of the quantile estimation for both i.i.d. and MCMC sequences. Section 4 presents some applications to illustrate the use of the method. Section 5 contains concluding remarks and suggests some extensions for further research work.

2 Structure of MCMC Output

An MCMC procedure is first used to randomly generate a sequence of draws from the joint posterior distribution of the unknown parameters in a model. After removal of initial “burn-in” draws, a sequence of draws can be stored in a matrix

$$[X] = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{S1} & x_{S2} & \cdots & x_{Sp} \end{pmatrix},$$

where S represents the number of draws, p represents number of unknown parameters $\boldsymbol{\theta}$ in the model, and (x_{i1}, \dots, x_{ip}) denotes the i th MCMC draw for $\boldsymbol{\theta}$. If additional quantities are of interest, like some function of parameters the $g(\boldsymbol{\theta})$, extra columns may be added to the output matrix with the same number of rows. Then the new output matrix would be

$$[\tilde{X}] = \begin{pmatrix} x_{11} & \cdots & x_{1p} & g(\mathbf{x})_1 \\ x_{21} & \cdots & x_{2p} & g(\mathbf{x})_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_{S1} & \cdots & x_{Sp} & g(\mathbf{x})_S \end{pmatrix}.$$

After generating MCMC draws, point estimates of $g(\boldsymbol{\theta})$ can be obtained. One possibility is the median of the draws from the marginal posterior distribution. All S draws for each parameter are placed in ascending order such that $g(\mathbf{x})_{i(1)} < g(\mathbf{x})_{i(2)} < \cdots < g(\mathbf{x})_{i(S)}$, $i = 1, \dots, p$, and then the posterior median of the parameter $g(\boldsymbol{\theta})$, which is denoted by M_i , is defined by

$$M_i = \begin{cases} g(\mathbf{x})_{i,((1+S)/2)} & \text{if } S \text{ is odd,} \\ (g(\mathbf{x})_{i,(S/2)} + g(\mathbf{x})_{i,(S/2+1)})/2 & \text{if } S \text{ is even.} \end{cases}$$

In addition to the point estimate, a credible interval quantifies statistical uncertainty. A $100(1-\alpha)\%$ Bayesian credible interval for the parameter $g(\boldsymbol{\theta})$ may be obtained from the $\alpha/2$ and $(1 - \alpha/2)$ sample quantiles of the empirical posterior distribution of MCMC draws, denoted by

$$[g(\mathbf{x})_{i,(S(\alpha/2))}, g(\mathbf{x})_{i,(S(1-\alpha/2))}].$$

It is the precision resulting from the use of MCMC draws to obtain such

intervals that interest us here and, in particular, how many MCMC draws S are needed to obtain a desired degree of precision.

After a “burn-in” period, it is common to assume that the Markov Chain has been initialized for obtaining draws that approximately follow the stationary distribution (i.e., the posterior distribution) of the Markov chain. This is supported by the theory of Harris recurrent Markov chains; see Athreya and Lahiri (2006) Ch. 14. Then, by the transition probability structure of the MCMC algorithm, a sample $g(\boldsymbol{x})_{i(1)}, \dots, g(\boldsymbol{x})_{i(S)}$ may be then treated as a realization of a stationary, time dependent sequence from the marginal posterior distribution. From this, techniques of quantile estimation for such a time series may then be applied for determining the number of draws S needed to achieve a desired precision in the endpoints of a Bayesian credible interval.

3 Quantile Estimation and Sample Size

Determination

Here we provide background for quantile estimation and a method for sample size determination with a stationary time series. We begin with reviewing this process for i.i.d. data in Section 3.1 and describe the methodology of interest for time series (i.e., MCMC samples) in Section 3.2.

3.1 Quantile Estimation for i.i.d. Sequences

Let F be the increasing cumulative distribution function (cdf) of the i.i.d. continuous random variables X_1, \dots, X_S so that $F(x) = \Pr(X_i \leq x)$, $i = 1, \dots, S$. Also let f be the corresponding probability density function given by

$f(x) = F'(x)$. For a particular $0 < p < 1$, the p quantile of the distribution function F is denoted by ξ_p and is defined such that $F(\xi_p) = p$. The quantile ξ_p can be estimated by the sample quantile $\widehat{\xi}_p$ of the distribution, which is defined as

$$\widehat{\xi}_p = \begin{cases} x_{(Sp)} & \text{if } Sp \text{ is an integer,} \\ x_{(\lfloor Sp \rfloor + 1)} & \text{otherwise,} \end{cases} \quad (1)$$

where $\lfloor \cdot \rfloor$ denotes the floor function and $x_{(i)}$ represents the i th order statistic among $x_{(1)} < x_{(2)} < \dots < x_{(S)}$.

We can also refer to the sample quantile $\widehat{\xi}_p$ as the p quantile of the empirical distribution function F_S , which is defined as

$$F_S = \frac{1}{S} \sum_{i=1}^S \mathbf{I}(X_i \leq x),$$

where $\mathbf{I}(\cdot)$ is the indicator function. By classical distributional results (cf. Serfling 1980), it holds that

$$\frac{\widehat{\xi}_p - \xi_p}{\sqrt{p(1-p)/[Sf^2(\xi_p)]}} \xrightarrow{d} N(0, 1), \quad \text{as } S \rightarrow \infty.$$

Hence, for large sample sizes $\widehat{\xi}_p$ is approximately normally distributed with mean equal to the population quantile ξ_p and variance

$$p(1-p)/[Sf^2(\xi_p)]. \quad (2)$$

Given a desired relative precision d and the confidence level $1 - \alpha$, one may wish to determine the number of draws S so that $\Pr(|\widehat{\xi}_p - \xi_p| \leq d\xi_p) = 1 - \alpha$ holds approximately. Note here d represents precision relative to the quantile ξ_p of interest in estimation and $d\xi_p$ represents absolute precision. For example,

if $d = 0.01$, then we are seeking the number of draws S at which the estimate $\widehat{\xi}_p$ matches the target quantile ξ_p to two decimal place accuracy (in the case that $1 - d \leq \widehat{\xi}_p/\xi_p \leq 1 + d$ with confidence $1 - \alpha$). We first run the procedure based on an initial or pilot sample of draws $\tilde{X}_1, \dots, \tilde{X}_{\tilde{S}}$ ($\tilde{X}_1, \dots, \tilde{X}_{\tilde{S}}$ may correspond to the first \tilde{S} draws in X_1, \dots, X_S), and get the estimates $\tilde{\xi}_p$ and $\widehat{f}(\tilde{\xi}_p)$ for ξ_p and $f(\xi_p)$. Substituting $\tilde{\xi}_p$ and $\widehat{f}(\tilde{\xi}_p)$ in (2) and using normal theory, one may equate $d\tilde{\xi}_p = Z_{1-\alpha/2} \sqrt{p(1-p)/[S\widehat{f}(\tilde{\xi}_p)^2]}$, where $Z_{1-\alpha/2}$ denotes the upper $\alpha/2$ quantile of the standard normal distribution. Then the approximate total number of draws that will be needed can be estimated as

$$S = \left\lceil \frac{Z_{1-\alpha/2}^2 [p(1-p)]}{[d\tilde{\xi}_p \widehat{f}(\tilde{\xi}_p)]^2} \right\rceil + 1.$$

3.2 Quantile Estimation for MCMC Sequences

MCMC draws generally have some dependence structure and may be treated as a stationary time series realization after a “burn-in” period (cf. Section 2). For stationary time processes, the degree of dependency can be quantified by the autocovariance function. Suppose X_1, \dots, X_S are random variables from a real-valued stationary time process $\{X_t\}$, having an increasing marginal cdf $F(\cdot)$ with associated pdf $f(\cdot)$, and p quantile ξ_p satisfying $F(\xi_p) = p \in (0, 1)$. Let $\widehat{\xi}_p$ denote the p sample quantile from X_1, \dots, X_S . Under mild regularity conditions on the time dependence, $\widehat{\xi}_p$ is approximately normal in large samples with mean ξ_p and large-sample variance given by

$$\text{Avar}(\widehat{\xi}_p) = \frac{\sigma^2(p; S)}{S f^2(\xi_p)}, \quad (3)$$

where

$$\sigma^2(p; S) = S \times \text{Var} \left(\frac{1}{S} \sum_{i=1}^S Y_i \right) = p(1-p) + 2 \sum_{j=1}^{S-1} \left(1 - \frac{j}{S} \right) r(j),$$

for $Y_i \equiv \mathbb{I}(X_i < \xi_p)$, $i = 1, \dots, S$, and $r(j) \equiv \text{cov}(Y_1, Y_{j+1})$ for any integer j . $\{Y_i\}_{i=1}^S$ are i.i.d. Bernoulli(p) random variables, so that (3) reduces to the large-sample variance of the sample quantile ξ_p in the i.i.d. setting (2) by using $r(0) = p(1-p)$ and $r(j) = 0$ for $j \neq 0$.

In order to formulate a sample size determination based on the large sample normality of $\widehat{\xi}_p$, we need to estimate both the variance $\sigma^2(p; S)$ and the pdf $f(\cdot)$ in (3). To do so, we use kernel estimation based on the flat-top lag window of Politis and Romano (1995), which is denoted by $\lambda^T(t)$. There are various choices for the family of flat-top kernels. The simplest form shown in Politis (2003) is defined as

$$\lambda^T(t) = \begin{cases} 1 & 0 \leq |t| \leq 0.5 \\ 2(1 - |t|) & 0.5 \leq |t| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Flat-top kernels are known to induce good asymptotic properties and fast convergence rates when applied to spectral or marginal density estimation (cf. Politis 2003).

An estimated spectral density from the series $\{\widehat{Y}_i \equiv \mathbb{I}(X_i < \widehat{\xi}_p)\}_{i=1}^S$ can be used to estimate $\sigma^2(p; S)$. In particular, if $\widehat{\phi}(0)$ denotes the estimated spectral density of $\{Y_i\}$ at the origin, based on $\{\widehat{Y}_i\}_{i=1}^S$, then $2\pi\widehat{\phi}(0)$ estimates $\sigma^2(p; S)$

using that $\lim_{S \rightarrow \infty} \sigma^2(p; S) = 2\pi\phi(0)$ when $\sum_{j=0}^{\infty} |r(j)| < \infty$, where

$$\phi(\omega) = \frac{1}{(2\pi)} \sum_{k=-\infty}^{\infty} e^{ik\omega} r(k), \quad -\pi \leq \omega \leq \pi,$$

denotes the spectral density function of the process $\{Y_i \equiv \mathbf{I}(X_i < \xi_p)\}$. The corresponding sample version of $\phi(\omega)$ is

$$\hat{\phi}(\omega) = \frac{1}{(2\pi)} \sum_{k=-H}^H \lambda^T(k/H) e^{ik\omega} \hat{r}(k),$$

based on the flat-top kernel $\lambda^T(\cdot)$ with a bandwidth parameter H and the lag- k sample autocovariance

$$\hat{r}(k) = \frac{1}{S} \sum_{i=1}^{S-|k|} (\hat{Y}_i - \bar{Y})(\hat{Y}_{i+|k|} - \bar{Y}), \quad |k| < S, \quad (4)$$

where $\bar{Y} = \sum_{i=1}^S \hat{Y}_i / S$. There is a simple empirical rule that can be used to choose the kernel bandwidth. The optimal value of H , which is denoted by H^{opt} , minimizes the MSE of $\hat{\phi}(\omega)$. From Politis (2003), an estimate of H^{opt} is given by $\hat{H} = 2h$, where h is the smallest positive integer such that

$$\max_{k=1, \dots, K} |\hat{\rho}(h+k)| < c\sqrt{\log S/S},$$

for $\hat{\rho}(k) = \hat{r}(k)/\hat{r}(0)$. Here c denotes a positive constant, recommended to be 2, and the value of K was recommended to be 5; see Politis (2003) for details.

We obtain an estimate $\hat{f}(\cdot)$ of the marginal pdf $f(\cdot)$ of $\{X_t\}$ as follows, also based on the flat-top kernel. Based on an inversion of the Fourier transform,

we have

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ixt} Q(t) dt, \quad x \in \mathbb{R},$$

where

$$Q(t) = \int_{-\infty}^{\infty} e^{-itx} f(x) dx, \quad t \in \mathbb{R}$$

is the characteristic function of f . We estimate $f(x)$, $x \in \mathbb{R}$, by

$$\hat{f}(x) = \frac{1}{2\pi} \int_{-M}^M \lambda^T(t/M) e^{itx} \hat{Q}(t) dt = \frac{1}{\pi S} \sum_{j=1}^S g(x, x_j, M),$$

$$g(x, x_j, M) = \begin{cases} \frac{2}{M(x-x_j)^2} [\cos(\frac{M}{2}(x-x_j)) - \cos(M(x-x_j))] & \text{if } x \neq x_j \\ \frac{3}{4}M & \text{if } x = x_j, \end{cases}$$

where M is a bandwidth and

$$\hat{Q}(t) = \frac{1}{S} \sum_{j=1}^S e^{-itX_j} \quad (5)$$

is the sample characteristic function of $\{X_t\}_{t=1}^S$. In Politis (2003), a bandwidth estimate is given by $\widehat{M} = 2m$, where m is the smallest positive real number such that

$$\max_{t \in (0, K)} |\widehat{Q}(m+t)| < c\sqrt{\log S/S}.$$

As before, values $c = 2$ and $K = 5$ were recommended.

Now suppose that we wish to choose the number of draws S such that $\Pr(|\widehat{\xi}_p - \xi_p| \leq d\xi_p) = 1 - \alpha$ holds for dependent draws X_1, \dots, X_S . Based on a pilot set of draws $\tilde{X}_1, \dots, \tilde{X}_{\tilde{S}}$ of length \tilde{S} ($\tilde{X}_1, \dots, \tilde{X}_{\tilde{S}}$ may correspond to the first \tilde{S} draws in X_1, \dots, X_S), we apply the kernel estimation above to obtain

estimates $\tilde{\xi}_p$, $2\pi\tilde{\phi}(0)$ and $\hat{f}(\tilde{\xi}_p)$ of ξ_p , $\sigma^2(p; S)$ and $f(\xi_p)$ for substitution in the large sample variance formula (3). Then we equate

$$d\tilde{\xi}_p = Z_{1-\alpha/2}\sqrt{2\pi\tilde{\phi}(0)/[S\hat{f}(\tilde{\xi}_p)^2]},$$

and solve for

$$S = \left\lceil \frac{Z_{1-\alpha/2}^2 2\pi\tilde{\phi}(0)}{(d\tilde{\xi}_p\hat{f}(\tilde{\xi}_p))^2} \right\rceil + 1. \quad (6)$$

Remark: In some cases, one may wish a sample size determination with respect to an absolute precision expressed as d , rather than an absolute precision $d\xi_p$ expressed in terms of the process quantile ξ_p . This corresponds to finding a sample size S so that the estimated quantile satisfies $P(|\hat{\xi}_p - \xi_p| < d) = 1 - \alpha$, with a certain confidence level $1 - \alpha$. In this case, the sample size formula (6) is easily modified by replacing $d\tilde{\xi}_p$ with d in the denominator of (6).

4 Application and Evaluation of the Algorithm

In this section we illustrate the use of the algorithm by choosing the number of MCMC draws in the Bayesian analysis for two applications, involving the fit of a linear model with mixed effects and a generalized linear model with a Poisson distribution and a log link. For each application we will also evaluate the algorithm by repeating the procedure 1,000 times and checking the average relative precision and coverage. First we need to determine a gold standard to serve as the true posterior quantile of interest in these applications. We run the MCMC procedure to generate 50,000,000 MCMC draws, and use the empirical

quantile of some posterior quantity of interest based on these 50,000,000 MCMC draws as the gold standard, or the true quantile ξ_p of the marginal posterior distribution (which depends on the sample and the parametric inference problem). In order to obtain pilot spectral and probability density estimates, a certain length of initial MCMC draws is needed. Hence, we run an initial 10,000 iterations and discard the first 2,000 draws as “burn-in”. After the “burn-in”, we assume these 8,000 initial draws are an approximately stationary realization from the marginal posterior distribution of interest. Based on these initial draws, we can estimate the sample autocovariance function in (4) and sample characteristic function in (5) in order to determine the number of draws S needed to estimate the marginal posterior quantile ξ_p of interest with a desired level of precision. That is, based on (6), the total number of draws needed after “burn-in” in the MCMC procedure can be calculated. For estimating the marginal posterior quantile of some parametric functions, the initial run with 8,000 MCMC draws is more than enough to obtain the desired precision and we then simply use the available 8,000 MCMC draws to estimate the marginal posterior quantile. However, for credible intervals with a desired precision regarding other parametric functions, more draws are required and we need to continue running the MCMC procedure to obtain the calculated number of draws. We apply the algorithm 1,000 times and the marginal posterior quantile of a corresponding parametric function is estimated each time based on actual MCMC draws used. We record the average number of MCMC draws needed as well as the average number of MCMC draws actually used over 1,000 runs. The mean absolute relative precision $|(\widehat{\xi}_p - \xi_p)/\xi_p|$ based on actual MCMC draws used and the proportion of 1,000 simulation runs for which $|\widehat{\xi}_p - \xi_p| \leq d\xi_p$ holds are also recorded.

There are several factors that can be expected to have an effect on the required number of MCMC draws.

1. The desired relative precision, denoted by d , is set as 0.01 and 0.005.
2. The confidence level, denoted by $1 - \alpha$, is set as 0.9 and 0.95.
3. The tail probability p , which is set at 0.8, 0.9, 0.95 to vary from moderate to extreme cases.

4.1 Application 1: Linear model with mixed effects

We consider an analysis application based on the 2002 Education Longitudinal Study (ELS) data from Hoff (2009), which involves 10th grade students from 100 different large urban public high schools. The linear mixed-effects model used here is

$$Y_{ij} = b_{0i} + b_{1i}x_{ij} + \epsilon_{ij}, \quad j = 1, \dots, n_i, i = 1, \dots, n; \quad (7)$$

where Y_{ij} denotes the normalized math scores for student j in school i , n denotes the total number of schools, and n_i denotes the total number of students in school i . In (7), b_{0i} and b_{1i} represent normally distributed random intercept and slope effects corresponding to school i . The variable x_{ij} is a regressor associated with b_{1i} , representing the centered socioeconomic status (SES) scores of j th student's family in school i . In particular, the distribution of random intercept and slope effects is given by

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim \text{MVN} \left[\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \Sigma_b = \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix} \right].$$

Here $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ is a normal error term for student j in school i . It is assumed that $[\{(b_{0i}, b_{0i})', i = 1, \dots, n\}]$ are i.i.d. random vectors, $(\{\epsilon_{ij} : i = 1, \dots, n; j = 1, \dots, m\})$ are i.i.d. random variables and the two collections are independent.

A Gibbs sampling algorithm with conjugate priors is used in this application.

The priors for β , Σ_b , and σ_ϵ are as follows:

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim \text{MVN}(\boldsymbol{\mu}_0, \Lambda_0),$$

$$\Sigma_b^{-1} \sim \text{Wishart}(S_0, \eta_0),$$

$$\sigma_\epsilon^2 \sim \text{inverse-gamma}(\alpha_0, \nu_0),$$

The details about the model and the initial values $(\boldsymbol{\mu}_0, \Lambda_0, \eta_0, S_0, \nu_0, \alpha_0)$ can be found in Hoff (2009, p. 200).

We consider estimating the marginal posterior quantile of two different parametric functions. The first is $g_1 = \beta_1$ which can be interpreted as the mean of the slopes that vary from school to school. The second, more complicated function is the ratio of standard deviations for the slope and error terms $g_2 = \sigma_1/\sigma_\epsilon$.

Table 1 shows results of the application of the linear mixed-effects model when $g_1 = \beta_1$ is the function of interest. In the table, the value S denotes the average number of MCMC draws needed over 1,000 different runs, for estimating the p posterior quantile based on the desired relative precision d and the confidence level $1 - \alpha$ after “burn-in”. The value S' denotes the actual number of MCMC draws used, averaging over 1,000 runs. For some cases, the

initial 8,000 draws are sufficient to achieve the required precision in quantile estimation. In this case, the actual number of MCMC draws used for the simulation is 8,000. For other cases, the initial 8,000 draws are not enough to achieve the required precision so that more draws are used. The reported value $1 - \hat{\alpha}$ denotes the proportion of 1,000 runs for which $|\hat{\xi}_p - \xi_p| \leq d\xi_p$ holds based on the actual number of MCMC draws used, and \bar{d} denotes the mean of the estimated absolute relative precisions over 1,000 runs based on actual number of MCMC draws used. Table 2 shows results of the application of linear mixed-effects model when $g_2 = \sigma_1/\sigma_\epsilon$ is the parametric function of interest.

Table 1: Results of the application of linear mixed-effects model when $g_1 = \beta_1$ is the function of interest.

Relative Precision d	Confidence Level $1-\alpha$	Quantile p	Number Needed S	Number Used S'	Estimated Coverage $1 - \hat{\alpha}$	Mean Precision \bar{d}
0.010	0.90	0.80	2,236	8,000	0.999	0.0027
0.010	0.90	0.90	2,501	8,000	0.995	0.0029
0.010	0.90	0.95	3,076	8,000	0.982	0.0032
0.010	0.95	0.80	3,158	8,000	0.997	0.0027
0.010	0.95	0.90	3,544	8,000	0.999	0.0027
0.010	0.95	0.95	4,366	8,000	0.995	0.0029
0.005	0.90	0.80	8,952	8,952	0.982	0.0032
0.005	0.90	0.90	10,021	10,021	0.907	0.0025
0.005	0.90	0.95	12,494	12,494	0.900	0.0024
0.005	0.95	0.80	12,673	12,673	0.944	0.0021
0.005	0.95	0.90	14,190	14,190	0.946	0.0020
0.005	0.95	0.95	17,364	17,364	0.932	0.0021

From Table 1 and Table 2 we find that the needed number of MCMC draws increases as the desired amount of precision in quantile estimation increases. Also, the number of MCMC draws is highly correlated with the quantile probability and the confidence level. More draws are needed if a larger quantile probability

Table 2: Results of the application of linear mixed-effects model when $g_2 = \sigma_1/\sigma_\epsilon$ is the function of interest.

Relative Precision d	Confidence Level $1-\alpha$	Quantile p	Number Needed S	Estimated Coverage $1 - \hat{\alpha}$	Mean Precision \bar{d}
0.010	0.90	0.80	12,065	0.904	0.0048
0.010	0.90	0.90	12,267	0.878	0.0051
0.010	0.90	0.95	13,940	0.886	0.0050
0.010	0.95	0.80	17,124	0.936	0.0042
0.010	0.95	0.90	17,490	0.935	0.0044
0.010	0.95	0.95	19,702	0.939	0.0043
0.005	0.90	0.80	48,268	0.880	0.0026
0.005	0.90	0.90	49,661	0.893	0.0026
0.005	0.90	0.95	55,197	0.885	0.0025
0.005	0.95	0.80	68,806	0.942	0.0020
0.005	0.95	0.90	69,878	0.934	0.0022
0.005	0.95	0.95	77,807	0.942	0.0021

or a larger confidence level is needed. Overall the estimated coverage probability $1 - \hat{\alpha}$ is very close to the nominal level $1 - \alpha$, and the mean of the estimated absolute relative precision \bar{d} is always about half of the proposed relative precision d . This latter behavior is expected to be seen in these tables when the sample size estimation procedure (6) is accurate. To see this, note that equation (3) uses the large-sample normality of the sample quantile under data dependence

$$(\hat{\xi}_p - \xi_p) \sim N\left(0, \frac{2\pi\phi(0)}{Sf^2(\xi_p)}\right)$$

so that, if the estimated number of draws follows equation (6) and the pilot estimators capture the unknown process quantities, it follows that

$$S \approx \frac{Z_{1-\alpha/2}^2 2\pi\phi(0)}{(\xi_p df(\xi_p))^2},$$

and

$$(\widehat{\xi}_p - \xi_p) \sim N\left(0, \frac{d^2 \xi_p^2}{Z_{1-\alpha/2}^2}\right).$$

The latter result implies

$$\widehat{d} = \frac{(\widehat{\xi}_p - \xi_p)}{\xi_p} \sim N\left(0, \frac{d^2}{Z_{1-\alpha/2}^2}\right),$$

and

$$E|\widehat{d}| = E\left|\frac{\widehat{\xi}_p - \xi_p}{\xi_p}\right| \approx \frac{d}{|Z_{1-\alpha/2}|} \sqrt{2/\pi}$$

by normal theory; for $1-\alpha = 0.9$ or 0.95 , we then have $E|\widehat{d}| \approx 0.49d$ or $0.41d$. Hence, as supporting evidence that the sample size estimation (6) procedure is effective, we expect observed relative precision \bar{d} in Tables 1-2 to be about half of d .

4.2 Application 2: Generalized linear model with a Poisson distribution

The application is from Whyte et al. (1987) and was presented in Dobson (1990). The number of deaths due to AIDS in Australia per 3 month period from January 1983 to June 1986 was recorded. When the response is a count, an appropriate model for linking the response and other explanatory variables is often a Poisson regression model. The Poisson model with a logarithm link function used here is

$$Y_i \sim \text{Poisson}(\lambda_i),$$

$$\log(\lambda_i) = \alpha + \beta \log(x_i).$$

Here Y_i denotes the number of deaths in Australia due to AIDS and $x_i = i$ denotes time point (measured in multiples of 3 months after January 1983). A random walk Metropolis algorithm is used here to generate draws from the posterior distribution. A multivariate normal prior is assumed on (α, β) . The first function we consider in this application is $g_1 = \beta$ which can be interpreted as the marginal effect of the time period on the logarithm of the expected number of deaths due to AIDS. The second, more complicated function is $g_2 = P(Y_3 = 0)$ which represents the probability of no deaths accrued in the third quarter in 1983. The results are shown in Table 3 and Table 4 when $g_1 = \beta$ and $g_2 = P(Y_3 = 0)$ are the functions of interests, respectively. The definitions of S , p , d , α , S' , $\hat{\alpha}$, and \bar{d} refer to those in Tables 1 and 2. And the results are consistent with the ones shown in Application 1.

Table 3: Results of the application of generalized linear model with Poisson distribution when $g_1 = \beta$ is the function of interest.

Relative Precision d	Confidence Level $1-\alpha$	Quantile p	Number Needed S	Number Used S'	Estimated Coverage $1 - \hat{\alpha}$	Mean Precision \bar{d}
0.010	0.90	0.80	3,224	8,000	0.989	0.0031
0.010	0.90	0.90	4,422	8,000	0.956	0.0037
0.010	0.90	0.95	6,420	8,000	0.930	0.0044
0.010	0.95	0.80	4,578	8,000	0.989	0.0031
0.010	0.95	0.90	6,278	8,000	0.956	0.0037
0.010	0.95	0.95	9,115	9,115	0.941	0.0041
0.005	0.90	0.80	12,895	12,895	0.889	0.0025
0.005	0.90	0.90	17,685	17,685	0.886	0.0025
0.005	0.90	0.95	25,680	25,680	0.877	0.0026
0.005	0.95	0.80	18,309	18,309	0.950	0.0021
0.005	0.95	0.90	25,110	25,110	0.929	0.0022
0.005	0.95	0.95	36,462	36,462	0.931	0.0022

Table 4: Results of the application of generalized linear model with Poisson distribution when $g_2 = P(Y_3 = 0)$ is the function of interest.

Relative Precision d	Confidence Level $1-\alpha$	Quantile p	Number Needed S	Estimated Coverage $1 - \hat{\alpha}$	Mean Precision \bar{d}
0.010	0.90	0.80	46,431	0.880	0.0051
0.010	0.90	0.90	53,309	0.878	0.0051
0.010	0.90	0.95	75,690	0.94	0.0042
0.010	0.95	0.80	65,924	0.943	0.0042
0.010	0.95	0.90	75,690	0.94	0.0042
0.010	0.95	0.95	93,855	0.930	0.0042
0.005	0.90	0.80	18,5721	0.884	0.0025
0.005	0.90	0.90	213,233	0.904	0.0025
0.005	0.90	0.95	264,400	0.899	0.0025
0.005	0.95	0.80	263,696	0.949	0.0021
0.005	0.95	0.90	302,758	0.944	0.0021
0.005	0.95	0.95	375,407	0.934	0.0021

5 Concluding Remarks and Areas for Further Research

This paper proposed a method to estimate the number of MCMC draws needed in order to obtain the desired precision for quantile estimates used in constructing Bayesian credible intervals. This procedure is especially useful when it is important to eliminate Monte Carlo error in reported results. The number of draws required will depend on the function of interest. Some functions may only require a moderate number of MCMC draws, while others may need a surprisingly large number. The number of draws is also positively related with the quantile probability, the desired precision, and the confidence level. If we want to have more confidence to get more precise estimates for a larger quantile

probability, then more MCMC draws are needed.

In this paper we initialize the Markov Chain by choosing appropriate “burn-in” period and then assume the rest of draws approximately follow the stationary distribution. If initialization issues are intended to be avoided, approaches based on regenerative simulation or batch means could be used. A regeneration occurs when the chain restarts itself, independently of its past, and the time excursions between successive regenerations provide i.i.d. observational stretches from the chain. The batch means approach breaks MCMC chains into batches of equal sizes and then these batches are assumed to be approximately i.i.d.. So both approaches potentially allow the variance of Monte Carlo estimates to be computed based on i.i.d. assumptions. Mykland et al. (1995) applied chain splitting techniques, first introduced by Athreya and Ney (1978) and Nummelin (1978), in regenerative simulation and explicitly stated how to incorporate this into some Markov chain samplers. Hobert et al. (2002) also discussed the use of regenerative simulation under appropriate assumptions. Doss and Tan (2013) developed a way to calculate standard errors for estimates of ratios of normalizing constants based on MCMC draws by using regenerative simulation. Jones et al. (2006) considered a stopping rule for MCMC procedures based on the width of a confidence interval, and used regenerative simulation and modified batch means methods to estimate the variance. Both regenerative simulation and batch means could be extended to estimate the variance for quantile estimates and then further determine the number of MCMC draws needed for accurate quantile estimates. However, a compounding issue related to techniques based on regenerations is that these are computationally difficult to automate in a general and simple way that would be immediately applicable to an arbitrary MCMC sampler. Furthermore, for sample size determinations

based on regeneration techniques, if the technique is general and not tied to algorithms steps of a specific MCMC sampler, then it may become necessary to estimate regenerations based on nonparametric estimators of the chain's transition density. These represent challenging estimation aspects, and such issues have been discussed by Bertail and Cl  men  on (2006) and Harari-Kermadec (2011) for regenerative block resampling methods for time series. As a remedy, the sample size determination approach proposed in this paper for estimating quantiles from MCMC output is simple to implement, is computationally fast, and has good accuracy properties.

In addition, the approach proposed in this paper can be further extended to other applications. If a certain degree of repeatability in bootstrap interval estimates is desirable, then a similar approach can be used to decide the number of bootstrap samples needed. Details on how to do this require further development.

Acknowledgement

We would like to thank Dr. Alyson G. Wilson from North Carolina State University for suggestions that helped us to improve our paper.

References

- Athreya, K. B. and S. S. N. Lahiri (2006). *Measure theory and probability theory*. Springer, New York.
- Athreya, K. B. and P. Ney (1978). A new approach to the limit theory

- of recurrent Markov chains. *Transactions of the American Mathematical Society* 245, 493–501.
- Bertail, P. and S. Cléménçon (2006). Regenerative block bootstrap for Markov chains. *Bernoulli* 12, 689–712.
- Chen, S. X. and P. Hall (1993). Smoothed empirical likelihood confidence intervals for quantiles. *The Annals of Statistics* 21, 1166–1181.
- Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician* 49, 327–335.
- Dobson, A. J. (1990). *An introduction to generalized linear models*. Chapman and Hall, London and New York.
- Doss, H. and A. Tan (2013). 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)*.
- Gelfand, A. E. and A. F. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85, 398–409.
- Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 721–741.
- Geyer, C. J. (1992). Practical Markov Chain Monte Carlo. *Statistical Science* 7, 473–483.

- Gilat, D. and T. P. Hill (1996). Strongly-consistent, distribution-free confidence intervals for quantiles. *Statistics & Probability Letters* 29, 45–53.
- Harari-Kermadec, H. (2011). Regenerative block empirical likelihood for Markov chains. *Journal of Nonparametric Statistics* 23, 781–802.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.
- Hobert, J. P., G. L. Jones, B. Presnell, and J. S. Rosenthal (2002). On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika* 89, 731–743.
- Hoff, P. D. (2009). *A first course in Bayesian statistical methods*. Springer, New York.
- Jones, G. L., M. Haran, B. S. Caffo, and R. Neath (2006). Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association* 101, 1537–1547.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21, 1087.
- Mykland, P., L. Tierney, and B. Yu (1995). Regeneration in Markov chain samplers. *Journal of the American Statistical Association* 90, 233–241.
- Nummelin, E. (1978). A splitting technique for harris recurrent Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 43, 309–318.

- Politis, D. N. (2003). Adaptive bandwidth choice. *Journal of Nonparametric Statistics* 15, 517–533.
- Politis, D. N. and J. P. Romano (1995). Bias-corrected nonparametric spectral estimation. *Journal of Time Series Analysis* 16, 67–103.
- Raftery, A. E. and S. Lewis (1992). How many iterations in the Gibbs sampler. *Bayesian Statistics* 4, 763–773.
- Serfling, R. (1980). *Approximation theorems of mathematical statistics*. Wiley, New York.
- Sitter, R. R. and C. Wu (2001). A note on Woodruff confidence intervals for quantiles. *Statistics & Probability Letters* 52, 353–358.
- Smith, A. F. and G. O. Roberts (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 3–23.
- Tierney, L. (1994). Markov chains for exploring posterior distributions. *the Annals of Statistics*, 1701–1728.
- Whyte, B., J. Gold, A. Dobson, and D. Cooper (1987). Epidemiology of acquired immunodeficiency syndrome in australia. *The Medical journal of Australia* 146, 65–69.
- Woodruff, R. S. (1952). Confidence intervals for medians and other position measures. *Journal of the American Statistical Association* 47, 635–646.