Regenerative Simulation in MCMC

Galin Jones[1]

School of Statistics
University of Minnesota

January 2010

[1] Joint work with James Flegal, Alicia Johnson and Ron Neath.

Introduction

Let F be a probability distribution.

We want to calculate some feature of F; for example, an expectation

$$E_F[g(X)] = \int g(x)F(dx)$$

or a quantile

$$F^{-1}(p) = \inf\{x : F(x) \geq p\} \qquad 0 < p < 1 \, .$$

Introduction

In principle estimation is simple: Observe a realization of a Harris ergodic Markov chain $\Phi = \{X_n\}$ having target F and use the sample quantities

$$\bar{g}_n = \frac{1}{n}\sum_{i=0}^{n-1} g(X_i) \qquad \text{and} \qquad Q_n(p) = \inf\{x : F_n(x) \geq p\}$$

The estimates are more valuable if we can also assess the Monte Carlo error

$$\bar{g}_n - E_F[g(X)] \qquad \text{and} \qquad Q_n(p) - F^{-1}(p)$$

This is usually done through a central limit theorem (CLT).

Regenerative Simulation

Basic Assumption: There exists $0 \leq s(x) \leq 1$ and a p.m. Q so that

$$P(x, A) \geq s(x)Q(A) \quad \forall\ x\ \ \forall\ A$$

Then

$$P(x, A) = s(x)Q(A) + (1 - s(x))\frac{P(x, A) - s(x)Q(A)}{1 - s(x)}$$

Resulting in

$$\Phi' = \{(X_0, \delta_0), (X_1, \delta_1), \dots\}$$

Retrospective RS:

$$\Pr(\delta_i = 1 \,|\, X_i = x, X_{i+1} = y) = \frac{s(x)Q(dy)}{P(x, dy)}$$

Estimating Expectations with RS

Regeneration times: $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$

Define $N_t = \tau_t - \tau_{t-1}$ and

$$S_t = \sum_{j=\tau_{t-1}}^{\tau_t - 1} g(X_j)$$

Then (S_t, N_t) are iid. Moreover,

$$\bar{g}_{\tau_R} := \frac{\overline{S}}{\overline{N}} \stackrel{a.s.}{\to} E_F[g(X)] \qquad R \to \infty \, .$$

<u>Thm</u> (Hobert, J, Presnell and Rosenthal (Bka, 2002)) If Φ is geometrically ergodic and $E_F|g(X)|^{2+\epsilon} < \infty$ some $\epsilon > 0$, then

$$\sqrt{R}(\bar{g}_{\tau_R} - E_F[g(X)]) \stackrel{d}{\to} \text{N}(0, \sigma_g^2) \qquad R \to \infty \, .$$

Estimating Quantiles with RS

For each $x \in \mathbb{R}$

$$F_{\tau_R}(x) = \frac{\overline{S}}{\overline{N}} = \frac{1}{\tau_R} \sum_{j=0}^{\tau_R - 1} I(X_j \leq x)$$

and it is easy to prove a Glivenko-Cantelli result

$$\sup_{-\infty < x < \infty} |F_{\tau_R}(x) - F(x)| \overset{a.s.}{\to} 0 \qquad R \to \infty \; .$$

Notice that a CLT holds for $F_{\tau_R}(x)$ by the earlier result on estimating expectations.

Estimating Quantiles with RS

Thm Assume Φ is geometrically ergodic and that $F'(Q(p))$ is positive and finite Then

$$\sqrt{R}(Q_{\tau_R}(p) - Q(p)) \xrightarrow{d} \mathrm{N}(0, \sigma_p^2) \qquad R \to \infty \, .$$

Remarks
There exists an easily computed consistent estimator of σ_p^2 but requires estimating F'.

We have simliar results for quantile estimators based on using importance sampling. This is useful when trying to estimate an extreme quantile.

Linear Mixed Models

HMO data (Hodges, JRSSB 1998): The response is the monthly plan premium and the regressors are average expenses per admission and whether the plan is in New England or not.

$$
\begin{aligned}
Y|\beta, \lambda_R &\sim \mathrm{N}_{341}(X\beta, \lambda_R^{-1} I_{341}) \\
\beta|\lambda_R &\sim \mathrm{N}_3(b, B^{-1}) \\
\lambda_R &\sim \mathrm{Gamma}(r_1, r_2)
\end{aligned}
$$

$\beta|\lambda_R, y =$ multivariate normal and $\lambda_R|\beta, y =$ Gamma

<u>THM</u> (Johnson and J, EJS, 2010) The deterministic and random scan Gibbs samplers are geometrically ergodic.

Gibbs for Linear Mixed Models

b, B, r_1, r_2 were chosen via an empirical Bayes-like approach.

Implement RS to estimate the Monte Carlo error in the estimate of mean of β_1 and the median. Used $R = 100$ regenerations.

	Estimate	MCSE
$E[\beta_1 \mid y]$	3.90	.007
median $F_{\beta_1 \mid y}$	3.85	.012

Where are we?

- Regenerative simulation provides an elegant method for estimating expectations and quantiles.
- We have CLTs for the Monte Carlo error for estimating expectations and quantiles.
- We have also shown that RS together with importance sampling can be used for both purposes. (CLT)
- RS has been shown to work well in Gibbs samplers and in simulated tempering.

BUT

- RS requires a minorization condition.
- RS has been shown to be impractical in high-dimensions.

Logit-Normal GLMM

For $i = 1, \ldots, q$ and $j = 1, \ldots, m_i$

$$Y_{ij}|u_i \overset{ind}{\sim} \text{Ber}(p_{ij}) \quad u_i \overset{iid}{\sim} N(0, \sigma^2) \quad logit(p_{ij}) = \beta x_{ij} + u_i$$

Goal: Maximum likelihood estimation of (β, σ^2)

Algorithms: Monte Carlo Newton-Raphson, Monte Carlo EM, Monte Carlo Maximum Likelihood–All require simulation from

$$\pi(u|y, \beta, \sigma) \propto \exp\left\{ \sum_{i=1}^{q} \left[u_i y_{i+} - \sum_{j=1}^{m_i} \log(1 + e^{\beta x_{ij} + u_i}) - \frac{u_i^2}{2\sigma^2} \right] \right\}$$

Logit-Normal GLMM

Use a full-dimensional random walk Metropolis having proposal

$$U \sim \mathrm{N}_q(u_{t-1}, \tau^2 I_q)$$

<u>Thm</u> The random walk sampler is geometrically ergodic.

Mykland, Tierney and Yu (JASA, 1995) give a formula for doing retrospective regeneration. A regeneration can only occur when

- the proposal is accepted and
- the chain is in a hypercube centered at some point $\tilde{u}$

The smaller the hypercube, the larger the probability of regeneration.

Logit-Normal GLMM

Numerical experiment:

- $q = 10$, $m_i = 5$
- Simulated a chain of length 10^6
- Hypercube centered at overall mean of simulated values
- Width of hypercube given by a multiple of the sample variance of the simulated values
- Simulated another chain of length 10^6 and observed 391,804 accepted proposals

We could find no acceptable trade-off between size of hypercube and probability of regeneration which was 2×10^{-6} or smaller.

Component-wise MCMC

Let $P_1, \ldots, P_d$ be Markov kernels having the same invariant distribution F. Then

$$P_{comp} = P_1 \cdots P_d$$

is a systematic scan component-wise sampler with invariant distribution F.

If each P_i is a Gibbs update, then P_{comp} is a Gibbs sampler.

If at least one of the P_i are Metropolis-Hastings updates, then P_{comp} is a Metropolis-Hastings-within-Gibbs sampler.

RS in Component-wise MCMC

We derive sufficient conditions on the proposals for the embedded MH updates that guarantee

$$P_{comp}(x, A) \geq s(x)Q(A)$$

but these formulas are technical and not reported here.

Conditional on accepting *every* component-wise proposal, the probability of regeneration is

$$\Pr(\delta_i = 1 \,|\, X_i = x, X_{i+1} = y) = \frac{s(x)Q(dy)}{P_{comp}(x, dy)}$$

Logit-Normal GLMM

For $i = 1, \ldots, q$ and $j = 1, \ldots, m_i$

$$Y_{ij}|u_i \overset{ind}{\sim} \text{Ber}(p_{ij}) \quad u_i \overset{iid}{\sim} N(0, \sigma^2) \quad logit(p_{ij}) = \beta x_{ij} + u_i$$

<u>THM</u> The component-wise independence sampler with target $\pi(u|y, \beta, \sigma)$ and component-wise proposals $u_i \sim N_1(0, \sigma^2)$ is uniformly ergodic.

Empirical example:

- Benchmark Example: $x_{ij} = j/15$, $q = 10$, $m_i = 15$.
- 1000 independent chains with $R = 100$, 50, 25 regenerations
- $R = 100$: Average chain length was 1,073,406 (105,303)
- Average tour length was $10,712$

Adaptive Component-wise MCMC

$$P_{comp,t} = P_{1,t} \cdots P_{d,t}$$

where each component update P_i can be Gibbs or MH but there is at least one MH.

- Make draws for Gibbs updates and draw candidates for each MH component.
- If every candidate is accepted, check for regeneration.
- If regeneration, update each proposal distribution.

Based on method of Gilks, Roberts and Sahu (JASA, 1998).

Bivariate Normal Example

$$.34\phi_1(x) + .33\phi_2(x) + .33\phi_3(x)$$

where ϕ_1 is a bivariate standard normal and ϕ_2 and ϕ_3 are bivariate normals with means $(-3,-3)^T$ and $(2,2)^T$ and variances

$$\begin{pmatrix} 1 & .9 \\ .9 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & -.9 \\ -.9 & 1 \end{pmatrix}$$

Consider using the component-wise independence sampler where the proposals are for $i = 1, 2, 3$

$$p_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$$

The adaptation will be to set each μ_i to the current overall mean for the ith component and setting σ_i to be 110% of the current standard deviation.

Bivariate Normal Example

Compare CWIS-Adapt with CWIS-RS in terms of the confidence intervals used to estimate the Monte Carlo error in the estimate of $E_F X_2 = -.33$. Ran 1000 replications, each of length 10^5

Method	Coverage	Half-width	Number of tours
CWIS-RS	.931 (.008)	.140 (.012)	2337 (458)
CWIS-Adapt	.954 (.007)	.168 (.015)	2236 (170)

Logit-normal example

Compare CWIS-Adapt with CWIS-RS in terms of the confidence intervals for mles of (β, σ)

Ascent-based Monte Carlo EM (Caffo, J, Jank, JRSSB 2005) adaptively chooses the number of regenerations within each iteration.

Benchmark Example: $x_{ij} = j/15$, $q = 10$, $m_i = 15$.

Ran 500 replications, each starting with $R = 25$, adaptation was performed in the same way as the bivariate normal example.

Method	Coverage	Half-width
CWIS-RS	.929 (.011)	.233 (.021)
CWIS-Adapt	.953 (.009)	.334 (.039)

Summary

- Quantiles are often of interest in MCMC.
- Developed methods for estimating the MC error of quantile estimates based on CLTs using RS and importance sampling.
- Considered using RS in component-wise MCMC samplers.
- RS can work in component-wise samplers when it fails in full-dimensional updates.
- Gave examples where component-wise adaptive methods using RS worked well.
- Currently working on incorporating quantile estimation into adaptive methods.