Introduction to Markov Chain Monte Carlo

Charles Geyer
University of Minnesota

http://www.stat.umn.edu/geyer/mcmc

Good Old-Fashioned Monte Carlo

Good old-fashioned Monte Carlo (GOFMC), also called *ordinary Monte Carlo* (OMC) or *independent and identically distributed Monte Carlo* (IIDMC) is the idea of using IID simulations X_1, ..., X_n of a random process to approximate expectations

$$\mu = E\{g(X_i)\}$$

by sample averages

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i)$$

The cutesy name "Monte Carlo" alludes to the most famous casino in the world (randomness = gambling), get it? Over time it has lost its color and become a boring technical term.

Monte Carlo Error

A Monte Carlo approximation $\widehat{\mu}_n$ is only an approximation to the true quantity μ we are trying to calculate. So we need a discussion of *Monte Carlo error* $\widehat{\mu}_n - \mu$.

Fortunately, we don't need any new theory. All we need is elementary statistics.

The law of large numbers (LLN) says

$$\widehat{\mu}_n \xrightarrow{a.\,s.} \mu, \qquad n \to \infty$$

and the central limit theorem (CLT) says

$$\sqrt{n}\big(\widehat{\mu}_n - \mu\big) \xrightarrow{\mathcal{D}} \text{Normal}(0, \sigma^2), \qquad n \to \infty$$

where

$$\sigma^2 = \text{var}\{g(X_i)\}$$

(assuming this variance exists).

Monte Carlo Standard Error

Of course, if we don't know μ then we probably don't know σ either. But elementary statistics has the answer to this too.

The *Monte Carlo standard error* (MCSE) is a consistent estimate of σ, such as

$$\widehat{\sigma}_n = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \Bigl(g(X_i) - \widehat{\mu}_n \Bigr)^2}$$

Then Slutsky's theorem says

$$\frac{\widehat{\mu}_n - \mu}{\widehat{\sigma}_n / \sqrt{n}} \xrightarrow{\mathcal{D}} \text{Normal}(0, 1), \qquad n \to \infty$$

or in sloppy notation

$$\widehat{\mu}_n \approx \text{Normal}\left(\mu, \frac{\widehat{\sigma}_n^2}{n} \right)$$

MCSE Continued

It is remarkably difficult to get people to take standard errors seriously. Generally estimates ($\widehat{\mu}_n$) are reported, but MCSE ($\widehat{\sigma}_n$) are not. So no one, neither authors or readers has a clue what the numbers mean.

Why is a mystery to me. A statistician who doesn't care about standard errors certainly doesn't think like a statistician. I wonder what they are thinking.

Monte Carlo estimates not accompanied by standard errors are worthless.

Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) is the idea of using simulations X_1, ..., X_n of a Markov chain to approximate expectations

$$\mu = E_\pi\{g(X_i)\}$$

by sample averages

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i)$$

where π is the *equilibrium distribution*, also called *invariant distribution*, *stationary distribution*, or *ergodic limit* of the Markov chain (assuming such exists).

In other words, MCMC is just like GOFMC, except replace IID in GOFMC by Markov chain to get MCMC and proceed *mutatis mutandis*.

MCMC History

MCMC is a remarkable *tour de force*.

It dates back to the dawn of the computer age (Metropolis, et al., 1953), but is highly non-obvious, even in its original incarnation, which was calculating ergodic limits for models of physical systems.

What is obvious is run the (model of the) physical system and average over time (that's what an ergodic limit is).

The *tour de force* is the realization that any other Markov process with the same ergodic limit will also do.

Metropolis, et al. (1953) realized this and provided a simple algorithm for constructing a Markov chain having a specified equilibrium distribution (the Metropolis algorithm).

MCMC History Continued

The Metropolis algorithm, as generalized by Hastings (1970) and Green (1995), called the Metropolis-Hastings-Green algorithm, is the only known method of MCMC.

Every MCMC-like method is either a special case of the MHG algorithm, or is bogus.

Many researchers have invented almost-but-not-quite MCMC algorithms. But there is no theory about almost-but-not-quite Markov chains or about Markov chains having almost-but-not-quite a specified equilibrium distribution.

If you're going to do MCMC, do real MCMC, not bogo-MCMC.

The first task in any MCMC project is to verify that your computer code actually implements a Markov chain having the specified equilibrium distribution.

MCMC Theory

MCMC theory is just like IIDMC theory (except MC replaces IID).

The Markov chain law of large numbers (LLN) says

$$\widehat{\mu}_n \xrightarrow{a.\,s.} \mu, \qquad n \to \infty$$

and the Markov chain central limit theorem (CLT) says

$$\sqrt{n}\Big(\widehat{\mu}_n - \mu\Big) \xrightarrow{\mathcal{D}} \text{Normal}(0, \sigma^2), \qquad n \to \infty$$

where

$$\sigma^2 = \text{var}\{g(X_i)\} + 2 \sum_{k=1}^{\infty} \text{cov}\{g(X_i), g(X_{i+k})\}$$

(see Chan and Geyer, 1994, for assumptions).

Markov Chain MCSE via Batch Means

Fortunately, we do not have to approximate the rather obnoxious formula for asymptotic variance on the preceding slide.

If b is large, then

$$\sqrt{b}\left(\widehat{\mu}_b - \mu\right) \approx \text{Normal}(0, \sigma^2)$$

and

$$b\left(\widehat{\mu}_b - \mu\right)^2 \approx \sigma^2$$

And if b is small compared to n, then

$$b\left(\widehat{\mu}_b - \widehat{\mu}_n\right)^2 \approx \sigma^2$$

Thus if $1 \ll b \ll n$ the sample average of $b\left(\widehat{\mu}_b - \widehat{\mu}_n\right)^2$ over *batches of length b* estimates σ^2.

Normal Random Walk Metropolis

Let h be any nonnegative function on $\mathbb{R}^d$ that integrates, in which case we call h an *unnormalized probability density*.

Let Q be any d-dimensional, mean-zero, multivariate normal distribution.

The following is one step of a Markov chain having equilibrium distribution with unnormalized density h.

1. **[Proposal]** Simulate $z \sim Q$.

2. **[Metropolis Rejection]** Simulate $u \sim \text{Unif}(0, 1)$. If
$$h(x + z)/h(x) < u$$
set $x = x + z$. (Otherwise, x is unchanged.)

Normal Random Walk Metropolis (Continued)

Points to note.

- Does any continuous distribution on $\mathbb{R}^d$ specified by unnormalized density.

- There are many normal random walk Metropolis algorithms for any particular h, one for each positive definite matrix specifying the variance of Q.

- There are many other Metropolis algorithms, not to mention Metropolis-Hastings-Green.

Demo

I use an R package `mcmc` in computing classes.

It is still in version 0.x and not yet submitted to CRAN. You can get it at

`http://www.stat.umn.edu/geyer/mcmc`

For a demo problem, we will do the following, which was a PhD take-home exam question here at Minnesota.

`http://www.stat.umn.edu/geyer/PhD/F03/q1.pdf`

Demo (Continued)

The problem is to do Bayesian logistic regression with normal prior on the parameters with four predictor variables plus constant (five parameters).

Literature Cited

Chan, K. S. and Geyer, C. J. (1994).
Discussion of the paper by Tierney.
Annals of Statistics, 22, 1747–1758.

Green, P. J. (1995).
Reversible jump Markov chain Monte Carlo computation and Bayesian model determination.
Biometrika, 82, 711–732.

Hastings, W. K. (1970),
Monte Carlo sampling methods using Markov chains and their applications.
Biometrika, 57, 97–109.

Literature Cited (Continued)

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953).
Equation of state calculations by fast computing machines.
Journal of Chemical Physics, 21, 1087–1092.