

# Stat 8054 Lecture Notes: Geometric Ergodicity of Metropolis-Hastings

Charles J. Geyer

December 23, 2023

## Contents

<b>1</b>	<b>License</b>	<b>1</b>
<b>2</b>	<b>R</b>	<b>1</b>
<b>3</b>	<b>Geometric Ergodicity and the Markov Chain CLT</b>	<b>1</b>
<b>4</b>	<b>Lessons from Mengersen and Tweedie</b>	<b>2</b>
4.1	Independence Metropolis-Hastings is Just Like Rejection Sampling Except for Minor Details .	2
4.2	OMC Importance Sampling is Maybe Better . . . . .	2
4.3	Random Walk Metropolis-Hastings . . . . .	2
<b>5</b>	<b>A Lesson from Roberts and Tweedie</b>	<b>3</b>
<b>6</b>	<b>A Toy Problem</b>	<b>3</b>
	<b>Bibliography</b>	<b>4</b>

## 1 License

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (<http://creativecommons.org/licenses/by-sa/4.0/>).

## 2 R

- The version of R used to make this document is 4.3.2.
- The version of the `rmarkdown` package used to make this document is 2.25.

## 3 Geometric Ergodicity and the Markov Chain CLT

This is not a theory course but some theory came up in discussion of MCMC.

A Markov chain is *geometrically ergodic* (Meyn and Tweedie, 2009, Chapter 15) if

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq M(x)\rho^n \tag{1}$$

for some function  $M$  and some constant  $\rho < 1$ , where the left-hand is total variation distance between the distribution of the Markov chain at time  $n$  when started at  $x$  and the equilibrium distribution. Thus this distance is decreasing exponentially fast (also called geometrically fast because of the geometric series).

Geometric ergodicity is important because it is part of the sufficient conditions for the Markov chain central limit theorem (CLT) to hold. Chan and Geyer (1994), Theorem 2, showed that if the functional of the Markov chain under discussion has  $2 + \varepsilon$  moments for some  $\varepsilon > 0$  and the Markov chain is geometrically ergodic, then the CLT holds. Roberts and Rosenthal (1997), Theorem 2.1, showed that if the functional of the Markov chain under discussion has second moments and the Markov chain is *reversible* and geometrically ergodic, then the CLT holds.

Proving geometric ergodicity is hard. That is why Mengersen and Tweedie (1996), Roberts and Tweedie (1996), and Jarner and Hansen (2000) exist. But even the theorems in those papers are hard to apply to non-toy problems.

## 4 Lessons from Mengersen and Tweedie

### 4.1 Independence Metropolis-Hastings is Just Like Rejection Sampling Except for Minor Details

Tierney (1994) coined the name *independence* Metropolis-Hastings for a Metropolis-Hastings sampler for which the proposal distribution does not depend on the current state. Thus it is just a probability density function (PDF)  $q$ . Mengersen and Tweedie (1996), Theorem 2.1, says such a Markov chain is geometrically ergodic only if the importance weights  $\pi(x)/q(x)$  are bounded, where  $\pi$  is the target distribution, in which case the sampler is actually uniformly ergodic (which means that  $M(x)$  in (1) does not actually depend on  $x$ ).

This is exactly the same condition needed for ordinary Monte Carlo (OMC) rejection sampling to work, where  $q$  is the proposal distribution for that. The only differences are that

- in OMC rejection sampling one needs to know the bound (it is used in in the OMC rejection step), but
- in Metropolis-Hastings rejection one does not need to know the bound (it is not used in the Metropolis-Hastings rejection step).

and, of course,

- OMC rejection sampling produces independent and identically distributed samples from the target distribution, but
- the Metropolis-Hastings algorithm produces a Markov chain having the target distribution as its equilibrium distribution.

### 4.2 OMC Importance Sampling is Maybe Better

For the CLT to hold for OMC Importance sampling, one only needs that the quantities being averaged

$$\frac{g(X)\pi(X)}{q(X)}$$

have finite variance and this is different from the importance weights being bounded. Although, if one wants this to hold for every  $g$  that has second moments with respect to  $\pi$ , then this is equivalent to the importance weights being bounded.

### 4.3 Random Walk Metropolis-Hastings

Tierney (1994) coined the name *random walk* Metropolis (RWM) for a Metropolis sampler for which the proposal distribution has the form

$$q(x, y) = q(y - x)$$

where  $y$  is the proposal and  $x$  is the current state.

Mengersen and Tweedie (1996) provide some counterintuitive results for RWM. Their theorem 3.3 says that provided the proposal distribution has absolute first moment the sample cannot be geometrically ergodic unless the target distribution has a moment generating function. Jarner and Tweedie (2003) considerably improve this result. They drop the condition on the proposal distribution and extend to the multivariate case. Their Theorem 2.2 says an RMW sampler is geometrically ergodic only if the target distribution has a moment generating function. If it does not, then no proposal distribution will work.

Then Mengersen and Tweedie (1996) provide a sufficient condition for geometric ergodicity. For the multivariate case Roberts and Tweedie (1996) and Jarner and Hansen (2000) prove some sufficient conditions. All of these sufficient conditions are far from necessary. A student asked in class about distributions falling in between these necessary conditions and sufficient conditions.

## 5 A Lesson from Roberts and Tweedie

Theorem 5.1 of Roberts and Tweedie (1996) says that no Metropolis-Hastings algorithm that has

$$\text{ess sup } P(x, \{x\}) = 1 \tag{2}$$

can be geometrically ergodic. Here  $P(x, \{x\})$  is the probability that the (continuous) proposal is (Metropolis-Hastings) rejected when the current position is  $x$ . If that is not bounded away from one, then we cannot have geometric ergodicity.

## 6 A Toy Problem

What that student asked about is a target distribution for with the PDF does not go to zero at infinity. The PDF is spiky.

Such a distribution satisfies none of the sufficient conditions of Mengersen and Tweedie (1996) or Roberts and Tweedie (1996) or Jarner and Hansen (2000). So we do look at one such toy problem.

The toy problem we will analyze has a continuous uniform distribution on a certain set but is really a smeared out geometric distribution. the PDF is

$$f(x) = \begin{cases} 1, & n \in \mathbb{N} \text{ and } n < x < n + p(1-p)^n \\ 0, & \text{otherwise} \end{cases}$$

(here  $\mathbb{N}$  is the set of natural numbers (nonnegative integers)). This PDF clearly integrates to one and the  $\limsup$  as  $x \rightarrow \infty$  is one.

We propose to sample with via an RWM with any continuous unimodal proposal distribution. Examples are mean-zero normal distribution or uniform distribution on an interval symmetric about zero.

Clearly, for  $x$  very far out in the tails of the distribution, the per  $x$  rejection probability, the probability in (2), goes to zero as  $x$  goes to infinity.

Thus the sampler is not geometrically ergodic.

It is hard to know what lessons to draw from a toy problem. It is clear that there is no way to modify the example so that it is geometrically ergodic unless

- we modify the proposal distribution so it “knows” where the spikes are or
- we modify the target PDF so it isn’t so spiky.

But this example also shows there is something weird about asymptotics. The only problem is way out in the tails. The farther out in the tails the sampler gets, the longer it takes to get back to the center of the distribution. But we only have a problem far out in the tails. So as long as we are interested in getting a good idea about the part of the distribution that is not very far out in the tails, we are OK. For example if

we conditioned on the event that  $X < 10^{100000}$ , then we would have a geometrically ergodic sampler and the Markov CLT would hold. But we would never see samples outside this event in any run we would actually do anyway! So Markov chain asymptotics can be somewhat misleading.

## Bibliography

- Chan, K. S. and Geyer, C. J. (1994) Discussion: Markov Chains for Exploring Posterior Distributions. *Annals of Statistics*, **22**, 1747–1758. DOI: [10.1214/aos/1176325754](https://doi.org/10.1214/aos/1176325754).
- Jarner, S. F. and Hansen, E. (2000) Geometric ergodicity of Metropolis algorithms. *Stochastic Processes and their Applications*, **85**, 341–361. DOI: [10.1016/S0304-4149\(99\)00082-4](https://doi.org/10.1016/S0304-4149(99)00082-4).
- Jarner, S. F. and Tweedie, R. L. (2003) Necessary conditions for geometric and polynomial ergodicity of random-walk-type. *Bernoulli*, **9**, 559–578. DOI: [10.3150/bj/1066223269](https://doi.org/10.3150/bj/1066223269).
- Mengersen, K. L. and Tweedie, R. L. (1996) Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*, **24**, 101–121. DOI: [10.1214/aos/1033066201](https://doi.org/10.1214/aos/1033066201).
- Meyn, S. and Tweedie, R. L. (2009) *Markov Chains and Stochastic Stability*. second. Cambridge: Cambridge University Press.
- Roberts, G. and Rosenthal, J. (1997) Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability*, **2**, 13–25. DOI: [10.1214/ECP.v2-981](https://doi.org/10.1214/ECP.v2-981).
- Roberts, G. O. and Tweedie, R. L. (1996) Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, **83**, 95–110. DOI: [10.1093/biomet/83.1.95](https://doi.org/10.1093/biomet/83.1.95).
- Tierney, L. (1994) Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics*, **22**, 1701–1762. DOI: [10.1214/aos/1176325750](https://doi.org/10.1214/aos/1176325750).