

Lecture 39: Simulation; the Metropolis Sampler

Charles A. Bouman

Purdue University

Monday, November 30, 2026

Outline

- 1 Generating a sample: inverse transform sampling
- 2 Where the inverse CDF has closed form, and where it does not
- 3 A bimodal Gibbs density, and the question it raises
- 4 The Metropolis setup: the Gibbs PMF and the goal
- 5 Proposals: $q(w|x)$ and the symmetry condition
- 6 Accept or reject: ΔE and the acceptance probability
- 7 The Metropolis algorithm in pseudocode
- 8 Convergence: a Markov chain that is reversible
- 9 Detailed balance, full balance, and the limit

Core idea

The Metropolis sampler generates samples from any Gibbs distribution, $p(x) = \frac{1}{Z} e^{-u(x)}$, using only **changes in the energy function**. The proof that it works is an application of the ergodic theory for Markov chains from Chapter 13.

Today's Big Ideas

- Inverse transform sampling generates a sample of any random variable from a uniform random number, but it needs the inverse CDF, which usually has no closed form.
- The Metropolis sampler draws samples from any Gibbs distribution, and any strictly positive PMF can be written as a Gibbs distribution.
- Each step proposes a new sample from a symmetric proposal distribution, always accepts a decrease in energy, and accepts an increase with a probability that falls exponentially with the size of the increase.
- The sequence of samples is a Markov chain that satisfies the detailed-balance equations with the Gibbs distribution, so the Gibbs distribution is its stationary distribution.
- If the chain is irreducible and aperiodic and the number of states is finite, then the distribution of the samples converges to the Gibbs distribution.

Generating a sample: inverse transform sampling

Consider a situation in which one would like to use a computer to generate a sample of a random variable, X , with a CDF of $F(x) = P\{X \leq x\}$.

The classic approach is to first generate a pseudorandom number which is uniformly distributed on the interval $[0, 1]$,

$$U \leftarrow \text{Rand}([0, 1]) ,$$

where $\text{Rand}([0, 1])$ denotes a process for generating such a random variable. Next, one calculates the inverse CDF defined by

$$F^{-1}(u) \triangleq \inf\{x | F(x) \geq u\} .$$

Then the desired random variable can be generated by applying the inverse CDF to the uniform random variable:

$$X \leftarrow F^{-1}(U) .$$

Where the method works, and where it fails

Inverse transform sampling is the method just described. It can be proved to work for *any* random variable.

With this result, one might think it would be easy enough to generate a random variable with any desired distribution. However, this is not always the case.

- For some distributions, such as exponential or Gaussian random variables, it is possible to calculate the inverse CDF in closed form, so the inverse transform method can be directly implemented.
- For a general random variable, the inverse CDF typically does not have any simple form, so direct implementation is not tractable.

Remark on the Gaussian case

The inverse CDF does not have closed form for the Gaussian distribution, but this problem can be circumvented by generating the distribution of pairs of independent Gaussian random variables since their vector norm is exponentially distributed.

An example: a bimodal density

Consider the random variable X , with a density function given by

$$p(x) = \frac{1}{z} \exp \left\{ - \left(4(x-1)^2(x+1)^2 + \frac{x}{2} \right) \right\} ,$$

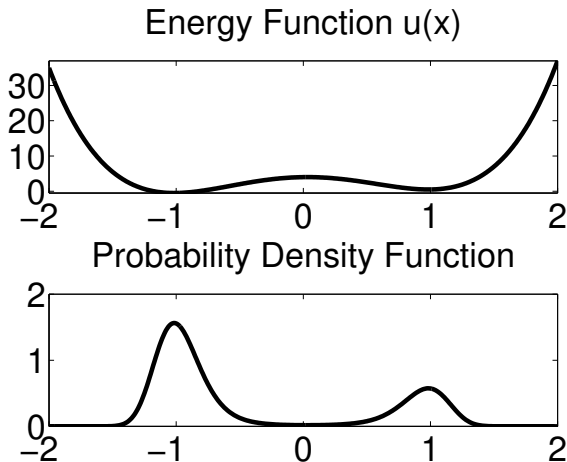
where z is a normalizing constant for the density.

This density has the form of a Gibbs distribution, $p(x) = \frac{1}{z} e^{-u(x)}$, where

$$u(x) = 4(x-1)^2(x+1)^2 + \frac{x}{2} .$$

- Both the energy function and its associated density are bimodal, with modes peaking at approximately $x = \pm 1$.
- The CDF of this density does not have closed analytical form, so neither does its inverse.

The energy function and the density



So how might one efficiently generate samples of this random variable on a computer?

Scope of the methods, and the plan

Stochastic simulation methods are a set of tools that generate samples from general distributions such as this one.

- While we motivate the method with the problem of generating a single random variable, the technique can be applied most usefully to generate samples from very complex high-dimensional distributions.
- The methods were first introduced to generate samples from chemical systems in thermodynamic equilibrium, so the approaches are quite generally useful.
- These methods also serve as a basic tool in stochastic optimization, stochastic integration, and stochastic EM parameter estimation.

The plan for the next three lectures

Today: the Metropolis sampler, perhaps the original example of a stochastic sampling method. Then the more general Hastings–Metropolis sampler, and finally the Gibbs sampler, which is in fact a special case.

Section 15.2: the Metropolis setup

The Metropolis sampler is a very clever method that allows samples to be generated from any Boltzmann or, equivalently, Gibbs distribution.

Consider the discrete distribution on the set Ω with a PMF of the form

$$p(x) = \frac{1}{z} \exp \{-u(x)\} ,$$

where $u(x)$ is an energy function and z is the associated partition function.

- Any strictly positive PMF can be put in this form, since it is always possible to take the energy function as $u(x) = -\log p(x) + \text{const.}$
- The random quantity, X , may be a random variable, or it may be a higher-dimensional random object such as a vector or random field.

The theory developed here assumes discrete random variables, but the methods equally apply to continuously valued random variables, since continuous random variables can be quantized to form discrete ones.

The goal, and the proposal

The objective of a simulation algorithm is to generate a sequence of random objects, denoted by $X^{(k)}$, whose distribution converges to the desired Gibbs distribution, $p(x)$.

The strategy of the Metropolis algorithm is to use the current sample, $X^{(k)}$, to generate a “proposal” for a new sample denoted by W .

- The proposal, W , is generated randomly, but its distribution depends on the current sample, $X^{(k)}$.
- More specifically, the proposed sample, W , is chosen from some conditional distribution, $q(w|X^{(k)})$.

However, this **proposal density**, $q(w|x^{(k)})$, must obey the symmetry property that for all $w, x \in \Omega$,

$$q(w|x) = q(x|w) .$$

Symmetric proposals, and how to generate them

While the symmetry condition is quite restrictive, there are reasonable methods for generating proposals that fulfill the condition.

- For example, we can set the proposal to be $W \leftarrow X^{(k)} + Z$, where Z is a random variable with a symmetric PDF that is independently generated at each new time, k .
- In this case, it is easy to show that the condition applies.

In practice, the proposal, W , can be generated with the inverse transform sampling method. In pseudocode, this can be expressed as

$$\begin{aligned}U &\leftarrow \text{Rand}([0, 1]) , \\ W &\leftarrow Q^{-1}(U|X^{(k)}) ,\end{aligned}$$

where $Q^{-1}(\cdot|x^{(k)})$ is the inverse CDF corresponding to the density, $q(w|x^{(k)})$.

- The proposal density is typically chosen to be relatively simple so that generating samples from this distribution is easy.

Accept or reject: the change in energy

The clever trick of the Metropolis algorithm is then to accept or reject these proposals in a very particular way. First compute the change in the energy given by $\Delta E \leftarrow u(W) - u(X^{(k)})$.

If the change in energy is negative, $\Delta E < 0$, then we immediately accept the proposal, so that

$$X^{(k+1)} \leftarrow W \quad \text{if } \Delta E < 0 .$$

However, if the change in energy is positive, $\Delta E \geq 0$, then we accept the proposal with a probability of $e^{-\Delta E}$ and reject it with probability $1 - e^{-\Delta E}$. In practice, this is done with the following two steps:

$$\begin{aligned} U &\leftarrow \text{Rand}([0, 1]) , \\ X^{(k+1)} &\leftarrow W \quad \text{if } \Delta E \geq 0 \text{ and } U < e^{-\Delta E} , \\ X^{(k+1)} &\leftarrow X^{(k)} \quad \text{if } \Delta E \geq 0 \text{ and } U \geq e^{-\Delta E} . \end{aligned}$$

The acceptance probability

We can simplify this acceptance decision by introducing a function that expresses the acceptance probability for each proposal:

$$\alpha(w|x^{(k)}) = \min \left\{ 1, e^{-[u(w) - u(x^{(k)})]} \right\} .$$

Using this function, the proposal, W , is generated according to the sampling distribution,

$$\begin{aligned} U &\leftarrow \text{Rand}([0, 1]) , \\ W &\leftarrow Q^{-1}(U|X^{(k)}) , \end{aligned}$$

and then the sample is either accepted or rejected according to the acceptance probability, $\alpha(W|X^{(k)})$:

$$\begin{aligned} U &\leftarrow \text{Rand}([0, 1]) , \\ X^{(k+1)} &\leftarrow W \quad \text{if } U < \alpha(W|X^{(k)}) , \\ X^{(k+1)} &\leftarrow X^{(k)} \quad \text{if } U \geq \alpha(W|X^{(k)}) . \end{aligned}$$

The Metropolis algorithm

Metropolis Algorithm:

Initialize $X^{(0)}$

For $k = 0$ to $K - 1$ {

 // Generate proposal

$W \sim q(w|X^{(k)})$

 // Accept or reject proposal

$\alpha \leftarrow \min \left\{ 1, e^{-[u(W) - u(X^{(k)})]} \right\}$

$X^{(k+1)} \leftarrow \begin{cases} W & \text{with probability } \alpha \\ X^{(k)} & \text{with probability } 1 - \alpha \end{cases}$

}

Intuition

Decreases in energy are always accepted, but the probability of accepting a proposal with an increasing energy decreases exponentially as the energy increases.

A Markov chain, and its transition probabilities

How do we know that the Metropolis algorithm converges to the desired Gibbs distribution? The proof of convergence depends on a clever use of ergodic theory for Markov chains. The sequence $\{X^{(k)}\}_{k=1}^N$ is a Markov chain, because each new sample, $X^{(k+1)}$, depends only on the previous sample, $X^{(k)}$, and other independently generated random variables. Next, we can show, perhaps surprisingly, that this Markov chain is **reversible**, as defined in Chapter 13.

Consider the transition probabilities denoted by $p(x^{(k+1)}|x^{(k)})$. With a little thought it is clear that the transition probability is given by

$$p(x^{(k+1)}|x^{(k)}) = \begin{cases} q(x^{(k+1)}|x^{(k)})\alpha(x^{(k+1)}|x^{(k)}) & \text{if } x^{(k+1)} \neq x^{(k)}, \\ 1 - \sum_{w \neq x^{(k)}} q(w|x^{(k)})\alpha(w|x^{(k)}) & \text{if } x^{(k+1)} = x^{(k)} \end{cases}.$$

When $x^{(k+1)} \neq x^{(k)}$, the probability of the transition is the product of the probability that you will generate the proposal and the probability that you will accept the proposal. Since the transition probabilities must sum to one, the probability that $x^{(k+1)} = x^{(k)}$ is just one minus the sum of the alternative cases.

What must be shown: detailed balance

In order to show that these transition probabilities are reversible, we must show that for all state combinations, (x, x') , the detailed-balance equations of Chapter 13 hold. For our problem, these are given by

$$p(x'|x) \frac{1}{Z} e^{-u(x)} = p(x|x') \frac{1}{Z} e^{-u(x')} ,$$

where $p(x) = \frac{1}{Z} e^{-u(x)}$ is the stationary distribution of the ergodic Markov chain.

To show this, we will consider three cases:

- Case 1, in which $u(x') \leq u(x)$ and $x' \neq x$;
- Case 2, in which $u(x') > u(x)$ and $x' \neq x$;
- Case 3, in which $x' = x$.

Case 1: $u(x') \leq u(x)$ and $x' \neq x$

$$\begin{aligned} p(x'|x) &= q(x'|x) \alpha(x'|x) \\ &= q(x'|x) \\ &= q(x|x') \\ &= q(x|x') e^{-[u(x)-u(x')]} e^{[u(x)-u(x')]} \\ &= q(x|x') \alpha(x|x') e^{[u(x)-u(x')]} \\ &= p(x|x') e^{[u(x)-u(x')]} . \end{aligned}$$

Rearranging terms and normalizing by the partition function, z , provides the desired result:

$$p(x'|x) \frac{1}{z} e^{-u(x)} = p(x|x') \frac{1}{z} e^{-u(x')} .$$

Case 1: which equation is used at each step

- Equality 1: the transition probability expression, in the case $x' \neq x$.
- Equality 2: the acceptance probability,
 $\alpha(x'|x) = \min\{1, e^{-[u(x')-u(x)]}\}$, equals one because $u(x') \leq u(x)$.
- Equality 3: the symmetry condition on the proposal density,
 $q(w|x) = q(x|w)$.
- Equality 4: multiplication by $e^{-[u(x)-u(x')]}e^{[u(x)-u(x')]} = 1$.
- Equality 5: the acceptance probability again. Here $u(x) \geq u(x')$, so
 $\alpha(x|x') = e^{-[u(x)-u(x')]}.$
- Equality 6: the transition probability expression again, now with the roles of x and x' exchanged.

Cases 2 and 3, and full balance

Now there are two additional cases to consider.

- Case 2 results from swapping the roles of x and x' . The symmetry of the detailed-balance equation then immediately yields the desired result.
- For case 3, substituting in $x = x'$ results in the two sides of the detailed-balance equation being identical, so the required equality holds trivially.

The detailed-balance equations then imply that the full-balance equations hold by simply summing over the variable x :

$$\sum_{x \in \Omega} p(x'|x) \frac{1}{z} e^{-u(x)} = \frac{1}{z} e^{-u(x')} .$$

So the Gibbs distribution, $p(x) = \frac{1}{z} e^{-u(x)}$, is in fact the stationary distribution of the Markov chain generated by the Metropolis algorithm.

The conditions, and the limit

By the limit theorem for Markov chains of Chapter 13, we now know that if

- the proposal distribution, $q(x^{(k+1)}|x^{(k)})$, is chosen so that all states can be accessed, that is, the Markov chain is **irreducible**;
- the probability of remaining in the same state is greater than zero, that is, the Markov chain is **aperiodic**; and
- the number of states is **finite**,

then the Metropolis algorithm must generate an ergodic Markov chain that converges to the desired Gibbs distribution. In this case, we have that

$$\lim_{k \rightarrow \infty} P\{X^{(k)} = x\} = \frac{1}{Z} \exp\{-u(x)\}.$$

Two closing remarks

The Metropolis algorithm is commonly used even when these technical conditions do not hold. For example, it is commonly used to generate samples from high-dimensional continuous distributions with densities, such as the non-Gaussian MRF models of Chapter 6. Also, much stronger convergence results exist for stochastic simulation methods and ergodic Markov chains, so our convergence result is simply provided to illustrate the basic concept.

Summary

- Inverse transform sampling, $X \leftarrow F^{-1}(U)$, works for any random variable, but the inverse CDF typically has no simple form.
- The Metropolis sampler draws samples from any Gibbs distribution, $p(x) = \frac{1}{Z} e^{-u(x)}$, and any strictly positive PMF can be written this way.
- Each step generates a proposal, $W \sim q(w|X^{(k)})$, with $q(w|x) = q(x|w)$, and accepts it with probability $\alpha(w|x^{(k)}) = \min\{1, e^{-[u(w)-u(x^{(k)})]}\}$.
- The sequence $X^{(k)}$ is a Markov chain. Its transition probabilities satisfy the detailed-balance equations with the Gibbs distribution, so the Gibbs distribution is its stationary distribution.
- If the chain is irreducible and aperiodic and the number of states is finite, then $\lim_{k \rightarrow \infty} P\{X^{(k)} = x\} = \frac{1}{Z} \exp\{-u(x)\}$.

Reading: Sections 15.1–15.2. Next lecture: Sections 15.3–15.4, the Hastings–Metropolis sampler and sampling of MRFs.