Ma 322: Biostatistics
Derivation of the Metropolis-Hastings Algorithm for Markov Chain Monte Carlo Sampling (MCMC)

Prof. Wickerhauser

Thursday, March 21st, 2024

The goal is to generate random samples of a variable X with probability density function $\pi = \pi(x) = Kf(x)$, where f is a given nonnegative integrable function but K is an unknown positive normalization constant. This problem arises in Bayesian analysis when the posterior pdf π is known to be proportional to the product of a nonnegative likelihood and a prior pdf, but the normalization constant is unknown so that neither inverse cdf sampling nor rejection sampling can be used.

Information about π may be extracted by randomly sampling X. For example, by the Law of Large Numbers, if X has finite expectation $\mu = E(X)$, then averages

$$\bar{X}_N \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} X_i$$

of N random samples $\{X_1, \ldots, X_N\}$ will converge to μ as $N \to \infty$. If, in addition, $E(X^2)$ is finite and the samples are independent, then the Central Limit Theorem gives a good rate of convergencevia the squared-error estimate

$$E(|\mu - \bar{X}_N|^2) = O(1/N),$$

as $N \to \infty$. Additionally, sampling π often enough to approximate it with a fine histogram is one way to locate a maximum likelihood estimator (MLE) within the most populated bin.

The MCMC method achieves this goal by finding a transition function

$$P(y|x) = \text{Prob}\,(x_{n+1} = y \mid x_n = x),$$

not dependent on the time step $n \in \{0, 1, 2, \ldots\}$, for a Markov chain $\{X_0, X_1, X_2, \ldots\}$ that has a stationary distribution π. This is to be done given only f where $\pi = Kf$ and the normalization constant K is unknown.

This problem is underdetermined so we will add the additional condition that $P(y|x)$ defines a **reversible** Markov chain, namely one satisfying the detailed balance equations

$$(\forall x, y)\ P(x|y)\pi(y) = P(y|x)\pi(x).$$

Note that any such P will have pdf π as a stationary distribution: compute its action on π by summing over all states x to get

$$(\forall y)\ \int_x P(y|x)\pi(x)\,dx = \int_x P(x|y)\pi(y)\,dx = \pi(y)\int_x P(x|y)\,dx = \pi(y),$$

since the conditional pdf $P(x|y)$, for any y, is a pdf in x whose integral must equal 1.

In addition, we will assume that $P(y|x) > 0$ for all x, y. This implies that P is **ergodic** and guarantees that it has a unique stationary distribution, which is π, so that we need not fear convergence to something else. It also guarantees that π is positive and justifies the algebra performed below.

Now we adapt an idea from rejection sampling: factor the transition probability into the product of a conditional **proposal** or **jumping** probability g and a joint **acceptance** probability A:

$$P(y|x) \stackrel{\text{def}}{=} g(y|x)A(y,x).$$

There is considerable freedom in choosing g, but then A must be adjusted to get P with the desired **target** pdf $\pi = Kf$. Combining this factorization with the detailed balance equation, we deduce that

$$(\forall x, y)\ \frac{f(y)}{f(x)} = \frac{\pi(y)}{\pi(x)} = \frac{P(y|x)}{P(x|y)} = \frac{g(y|x)A(y,x)}{g(x|y)A(x,y)},$$

which is equivalent to

$$(\forall x, y)\ \frac{A(y,x)}{A(x,y)} = \frac{g(x|y)f(y)}{g(y|x)f(x)}.$$

It is easily checked that the last equation holds if we take

$$A(y,x) \stackrel{\text{def}}{=} \min\left\{1, \frac{g(x|y)f(y)}{g(y|x)f(x)}\right\}.$$

We may now implement a transition $x \to y$ with probability $P(y|x)$ as follows:

Metropolis-Hastings Sampling for MCMC

0. Let $x = X_n$ be given

1. Generate a random sample y from the conditional pdf $g(y|x)$.

2. Compute the ratio
$$r \stackrel{\text{def}}{=} \frac{g(x|y)f(y)}{g(y|x)f(x)}$$
using the given function f and the chosen jumping pdf g.

3. Generate a random sample $u \in [0, 1]$ with distribution Uniform([0,1]).

4. If $u < r$ then jump to y by taking $X_{n+1} = y$. Otherwise, remain at x taking $X_{n+1} = x$.

This algorithm, starting at $n = 0$ with any initial value X_0, produces a sequence of samples $\{X_1, X_2, \ldots\}$ for a Markov chain with stationary pdf π.

Some computation may be avoided if the jumping pdf is **symmetric**, which means

$$(\forall x, y)\ g(x|y) = g(y|x).$$

This is true, for example, if g is the uniform conditional pdf since then both sides are the same constant. It is also true if $g(y|x)$ is a normal pdf for y with mean x, since then $g(x|y)$ is a normal pdf for x with mean y and the same variance, and the two will be equal at all pairs x, y.

In this "symmetric" case, the acceptance rule simplifies by cancellation:

$$A(y, x) \stackrel{\text{def}}{=} \min\left\{1, \frac{f(y)}{f(x)}\right\}.$$

The simplified implementation of transition $x \to y$ with probability $P(y|x)$ is the following:

Metropolis Sampling for MCMC

0. Let $x = X_n$ be given

1. Generate a random sample y from the symmetric conditional pdf $g(y|x)$.

2. Compute the ratio
$$r \overset{\text{def}}{=} \frac{f(y)}{g(x)}$$
using the known f and the chosen jumping pdf g.

3. Generate a random sample $u \in [0, 1]$ with distribution Uniform([0,1]).

4. If $u < r$ then jump to y by taking $X_{n+1} = y$. Otherwise, remain at x taking $X_{n+1} = x$.