Monte Carlo methods

Aaron A. King   Edward L. Ionides

2024-07-04

Context: Monte Carlo methods for POMP models

Let's consider a general POMP model. As before, let $y^*_{1:N}$ be the data, and let the model consist of a latent process $X_{0:N}$ and an observable process $Y_{1:N}$. Then the likelihood function is

$$\mathcal{L}(\theta) = f_{Y_{1:N}}(y^*_{1:N}; \theta)$$

$$= \int f_{X_{t_0}}(x_{t_0}; \theta) \prod_{n=1}^{N} f_{Y_n|X_{t_n}}(y^*_n | x_{t_n}; \theta)\, f_{X_{t_n}|X_{t_{n-1}}}(x_{t_n} | x_{t_{n-1}}; \theta)\, dx_{t_0} \dots dx_{t_N}.$$

i.e., computation of the likelihood requires integrating over all possible values of the unobserved latent process at each time point. This is very hard to do, in general.

Let's examine some *Monte Carlo* approaches for evaluating this and other difficult integrals. An excellent technical reference on Monte Carlo techniques is @Robert2004.

The fundamental theorem of Monte Carlo integration

The basic insight of Monte Carlo methods is that we can get a numerical approximation to a challenging integral,

$$H = \int h(x)\, f(x)\, dx,$$

if we can simulate (i.e., generate random draws) from the distribution with probability density function f.

Theorem 1 (Fundamental theorem of Monte Carlo integration). Let f_X be the probability density function for a random variable X, and let $X_{1:J} = \{X_1, \dots, X_J\}$ be an independent and identically distributed sample of size J from f_X. Let H_J be the sample average of $h(X_1) \dots, h(X_J)$,

$$H_J = \frac{1}{J} \sum_{j=1}^{J} h(X_j).$$

Then $H_J \to H$ as $J \to \infty$ with probability 1. Less formally, we write

$$H_J \approx \int h(x)\, f_X(x)\, dx.$$

Proof. This is the strong law of large numbers, together with the identity that

$$\mathbb{E}\left[h(X)\right] = \int h(x)\, f(x)\, \mathrm{d}x.$$

$\square$

We can estimate the error in this approximation, because the empirical variance

$$V_J = \frac{1}{J-1} \sum_{j=1}^{J} [h(X_j) - H_j]^2$$

approximates the true variance, $\mathrm{Var}\left[h(X)\right] = \mathbb{E}\left[\left(h(X) - \mathbb{E}\left[h(X)\right]\right)^2\right]$.

The standard error on the approximation $H_J \approx \mathbb{E}\left[h(X)\right]$ is therefore $\sqrt{V_J/J}$. From the central limit theorem, the error is approximately normally distributed:

$$H_J - \mathbb{E}\left[h(X)\right] \sim \mathrm{Normal}\left(0, \sqrt{\frac{V_J}{J}}\right).$$

Exercise 1. Use Monte Carlo integration to estimate

$$\int_{-1}^{1} \frac{\mathrm{d}x}{1+x^4}.$$

Take J large enough so that your estimate is accurate to 3 decimal places.

The fundamental theorem of Monte Carlo inspires us to give further thought to the problem of simulating from a desired density function f, which can itself be a challenging problem. However, we first consider a useful generalization of the fundamental theorem.

Importance sampling

Sometimes it is difficult to sample directly from the distribution of X. In this case, we can often make use of *importance sampling*, in which we generate random samples from another distribution (easier to simulate) and make the appropriate correction. Specifically, suppose we wish to compute $\mathbb{E}\left[h(X)\right]$, where $X \sim f$, but it is difficult or impossible to draw random samples from f. Suppose g is a probability distribution from which it is relatively easy to draw samples and let $Y_{1:J} \overset{iid}{\sim} g$. The observation that

$$\mathbb{E}\left[h(X)\right] = \int h(x)\, f(x)\, \mathrm{d}x = \int h(x)\, \frac{f(x)}{g(x)}\, g(x)\, \mathrm{d}x.$$

allows us to generalize the Monte Carlo integration theorem to give the **Monte Carlo importance sampling theorem**:

$$\mathbb{E}\left[h(X)\right] \approx \frac{1}{J} \sum_{j=1}^{J} h(Y_j)\, \frac{f(Y_j)}{g(Y_j)}.$$

If we call $w_j = f(Y_j)/g(Y_j)$ the *importance weights*, we can write

$$\mathbb{E}\left[h(X)\right] \approx \frac{1}{J} \sum_{j=1}^{J} w_j\, h(Y_j).$$

Since $\mathbb{E}\left[w_j\right] = \mathbb{E}\left[f(Y)/g(Y)\right] = 1$, we can modify this formula to give a *self-normalized importance sampling* estimate,

$$\mathbb{E}\left[h(X)\right] \approx \frac{\sum w_j\, h(Y_j)}{\sum w_j}.$$

The self-normalized estimate requires computation of w_j only up to a constant of proportionality. The Monte Carlo variance associated with this estimate is

$$\frac{\sum w_j\, (h(Y_j) - \overline{h})^2}{\sum w_j}.$$

Obtaining accurate estimates requires some thought to the importance distribution g. Specifically, if the tails of g are lighter than those of f, the Monte Carlo variance will be inflated and the estimates can be unusable.

Simulation techniques for general distributions

Simulation refers to the generation of random variables. The general problem of simulation is: given a probability distribution f, find a procedure that generates random draws from f. This is a very important problem in scientific computing and much thought and effort has gone into producing reliable simulators for many basic random variables.
There are two basic ways of solving this problem:

1. the transformation method.
2. the rejection method.

The transformation method

This method works for discrete or continuous scalar random variables. Let f be the probability distribution function we seek to draw from (known as the *target distribution*) and F be the corresponding cumulative distribution function, i.e., $F(x) = \int_{-\infty}^{x} f(v)\, dv$.
Let $F^{-1}(u) = \inf\{x : F(x) \geq u\}$ be the inverse of F.
A basic fact is that, if $X \sim f$, then $F(X) \sim \text{Uniform}\,(0,1)$.

Proof. If $f(X) > 0$, then

$$\mathbb{P}\left[F(X) \leq u\right] = \mathbb{P}\left[X < F^{-1}(u)\right] = F(F^{-1}(u)) = u.$$
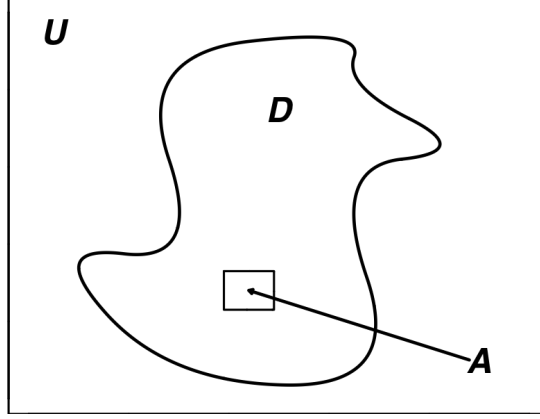
$\square$

This suggests that, if we can compute F^{-1}, we use the following algorithm to generate $X \sim f$:

1. Draw $U \sim \text{Uniform}\,(0,1)$.
2. Let $X = F^{-1}(U)$.

The rejection method

The transformation method is very efficient in that we are guaranteed to obtain a valid X from the density f for every $U \sim \text{Uniform}\,(0,1)$ we generate. Sometimes, however, we cannot compute the

inverse of the cumulative distribution function, as required by the transformation method. Under such circumstances, the rejection method offers a less efficient, but more flexible, alternative.

Rejection method. If it is difficult to sample a point from D directly, we can propose uniform draws X from a larger set U, and accept the proposals only if $X \in D$. {#fig:rejection}

We'll see how and why this method works in a simple case. Suppose a random variable X is *uniformly distributed* over a region $D \subset \mathbb{R}^d$. This means that, for any $A \subset D$,

$$\mathbb{P}\left[X \in A\right] = \frac{m(A)}{m(D)},$$

where m signifies the Lebesgue measure (i.e., ordinary Euclidean volume). To describe this situation, we say $X \sim \mathrm{Uniform}\,(D)$.

Let's suppose that we wish to simulate such an X. If D has a complicated shape, we might not know how to directly simulate a random draw from it. However, if we know D is a subset of some nicer region $U \subset \mathbb{R}^d$ and, in particular, if we know how to generate $Y \sim \mathrm{Uniform}\,(U)$, then we can simply do so until $Y \in D$, at which point we take $X = Y$. Since for any $A \subset D$,

$$\mathbb{P}\left[X \in A\right] = \mathbb{P}\left[Y \in A | Y \in D\right] = \frac{m(A)}{m(U)} \Big/ \frac{m(D)}{m(U)} = \frac{m(A)}{m(D)},$$

it follows that $Y \sim \mathrm{Uniform}\,(D)$. Fig.~@ref(fig:rejection) illustrates.

Consider an analogy to throwing darts. If the darts are thrown in such a way as to be equally likely to land anywhere in U, then those that do land in D are equally likely to land anywhere in D.

A useful little fact allows us to extend the rejection method from uniform distributions to arbitrary densities. @Robert2004 refer to this as the *fundamental theorem of simulation.*

Theorem 2 (Fundamental theorem of simulation)**.** Let $\mathbb{X} \subset \mathbb{R}^d$ and suppose $f : \mathbb{X} \to \mathbb{R}_+$ satisfies $\int f(x)\mathrm{d}x = 1$. Define
$$D = \{(x, u) \in \mathbb{X} \times \mathbb{R}_+ : x \in \mathbb{R}^d, 0 \leq u \leq f(x)\},$$
i.e., D is the region under the graph of f. If $(X, U) \sim \mathrm{Uniform}\,(D)$, then $X \sim f$.

Proof. The marginal distribution of X is

$$\int_0^{f(x)} \mathrm{d}u = f(x).$$

$\square$

This suggests the following rejection method for simulating an arbitrary random variable, illustrated in Fig.~@ref(fig:sim). Let f be the target distribution and g be another density from which it is easy to simulate. Choose M so large that $M\,g(x) \geq f(x)$ for all x. Then the following procedure simulates $X \sim f$:

1. Draw $Y \sim g$ and $U \sim \text{Uniform}\,(0, M)$. This is the *proposal*.
2. If $U \leq \frac{f(Y)}{g(Y)}$, *accept* the proposal, i.e., let $X = Y$, else repeat step 1.
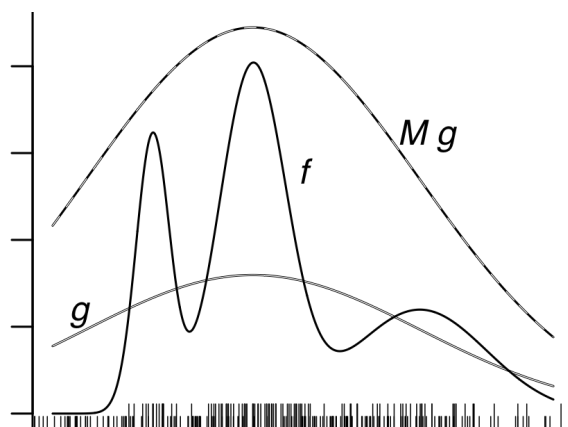
The rejection method for simulating an arbitrary random variable. The target density is f. We propose samples $X \sim g$ and accept the proposal only if $U \sim \text{Uniform}\,(0, M)$ satisfies $U \leq f(X)/g(X)$. The short ticks at bottom show the proposals X. The longer ticks show the accepted proposals, which are samples drawn from f.

Exercise 2. Use the rejection method to generate samples from the distribution with density

$$f(x) = \frac{\sqrt{2}}{\pi} \frac{1}{1 + x^4}.$$

Use Cauchy-distributed proposals, i.e., take

$$g(x) = \frac{1}{\pi} \frac{1}{1 + x^2}.$$

Use a histogram to verify that your samples are indeed drawn from the correct distribution.