Monte Carlo Integration Basics

Floyd Bullard

Teaching Contemporary Mathematics, NCSSM

25–26 January 2008
A Simple Integral

Consider the following definite integral $I \equiv \int_{-1}^{1} f(x)dx$, where $f(x) = 1 - x^2$.

Figure: $I = 4/3 \approx 1.3333$
A Simple Integral

Consider the following definite integral \( I \equiv \int_{-1}^{1} f(x) \, dx \), where \( f(x) = 1 - x^2 \).

**Figure:** Using 10 trapezoids: \( I \approx 1.32 \).
Let’s call $J$ the region bounded above by $y = f(x)$ and below by the $x$-axis. Its area is equal to the definite integral of interest: $I = A(J)$. 
Let’s call $J$ the region bounded above by $y = f(x)$ and below by the $x$-axis. Its area is equal to the definite integral of interest: $I = A(J)$.

Consider the rectangle $R = [-1, 1] \times [0, 1]$. We know that its area is $A(R) = 2$, and we know that it completely contains the region of interest, $J$. 
Let’s call $J$ the region bounded above by $y = f(x)$ and below by the $x$-axis. Its area is equal to the definite integral of interest: $I = A(J)$.

Consider the rectangle $R = [-1, 1] \times [0, 1]$. We know that its area is $A(R) = 2$, and we know that it completely contains the region of interest, $J$.

If we knew what fraction $p$ of $R$ was occupied by $J$, then we could compute the area of $J$ as simply $I = A(R) \times p = 2p$. 
Let’s call $J$ the region bounded above by $y = f(x)$ and below by the $x$-axis. Its area is equal to the definite integral of interest: $I = A(J)$.

Consider the rectangle $R = [-1, 1] \times [0, 1]$. We know that its area is $A(R) = 2$, and we know that it completely contains the region of interest, $J$.

If we knew what fraction $p$ of $R$ was occupied by $J$, then we could compute the area of $J$ as simply $I = A(R) \times p = 2p$.

Suggestions?
Let’s randomly sample points uniformly from $R$ and determine what fraction $\hat{p}$ of them lie in the region $J$. With a large number of sampled points, $\hat{p}$ will be a reasonably good approximation of $p$. 
Let’s randomly sample points uniformly from $R$ and determine what fraction $\hat{p}$ of them lie in the region $J$. With a large number of sampled points, $\hat{p}$ will be a reasonably good approximation of $p$.

Figure: In this simulation, $n=1000$, $\hat{p}=0.674$, and $\hat{I} = 1.348$. 
Using elementary statistics, we can not only estimate the value of $p$ with the sample proportion $\hat{p}$, but also determine a reasonable margin of error. Based on one simulation using $n = 100,000$ draws, a 99% confidence interval estimate of $p$ is $0.6656 \pm 0.0038$, resulting in an estimate of $I$ equal to $1.3312 \pm 0.0076$. 
Rejection sampling is occasionally useful in statistics, but it is inefficient for performing Monte Carlo integration: after drawing each $x_i$ and computing $f(x_i)$ (which may be computationally difficult), the information is reduced to a single bit, 0 or 1.

We can do better.
Recall that if $X$ is a random variable whose probability density function is $g$, and $h(X)$ is some function of $X$, then the expected value of $h(X)$ is given by

$$E[h(X)] = \int_{-\infty}^{+\infty} h(x)g(x)dx.$$
If $X$ has probability density function $g$ and $h(x) = \frac{f(x)}{g(x)}$, what is the expected value of $h(X)$?
If $X$ has probability density function $g$ and $h(x) = \frac{f(x)}{g(x)}$, what is the expected value of $h(X)$?

$$E[h(x)] = \int_{-\infty}^{+\infty} h(x)g(x)dx$$
If $X$ has probability density function $g$ and $h(x) = \frac{f(x)}{g(x)}$, what is the expected value of $h(X)$?

\[
E[h(x)] = \int_{-\infty}^{+\infty} h(x)g(x)dx
\]

\[
= \int_{-\infty}^{+\infty} \left[ \frac{f(x)}{g(x)} \right] g(x)dx
\]
If $X$ has probability density function $g$ and $h(x) = \frac{f(x)}{g(x)}$, what is the expected value of $h(X)$?

\[
E[h(x)] = \int_{-\infty}^{+\infty} h(x)g(x)dx
\]

\[
= \int_{-\infty}^{+\infty} \left[ \frac{f(x)}{g(x)} \right] g(x)dx
\]

\[
= \int_{-\infty}^{+\infty} f(x)dx
\]
If $X$ has probability density function $g$ and $h(x) = \frac{f(x)}{g(x)}$, what is the expected value of $h(X)$?

$$E[h(x)] = \int_{-\infty}^{+\infty} h(x)g(x)dx$$
$$= \int_{-\infty}^{+\infty} \left[ \frac{f(x)}{g(x)} \right] g(x)dx$$
$$= \int_{-\infty}^{+\infty} f(x)dx$$
$$= I$$
If $X$ has pdf $g$, then $E \left[ \frac{f(x)}{g(x)} \right] = I$. That means that we can sample an $x_i$ randomly from $g$, compute $h_i \equiv \frac{f(x_i)}{g(x_i)}$, and it will be an unbiased estimator of our integral $I$. 
If $X$ has pdf $g$, then $E \left[ \frac{f(x)}{g(x)} \right] = I$. That means that we can sample an $x_i$ randomly from $g$, compute $h_i \equiv \frac{f(x_i)}{g(x_i)}$, and it will be an unbiased estimator of our integral $I$.

We can do this many times and get a whole collection of unbiased estimates of $I$. These can be treated as a random sample from a population whose mean is $I$, and a confidence interval estimate of $I$ can be computed.
For example, consider again the definite integral 
\[ I \equiv \int_{-1}^{1} f(x)\,dx, \] where \( f(x) = 1 - x^2. \)
Let’s let $g$ be the standard normal distribution.
Let’s let $g$ be the standard normal distribution.

We’ll draw many values of $x_i$ from the distribution $g$, and for each one we’ll compute $h_i \equiv f(x_i)/g(x_i)$. Each $h_i$ should be an unbiased estimator of $I$. 
Figure: $f(x)$ and $g(x)$. We’ll draw $x_i$’s from $g$, then we’ll compute $h_i \equiv \frac{f(x)}{g(x)}$. Each $h_i$ will be an unbiased estimate of $I$. 
Figure: Here we sampled $n = 100,000$ points. A 99% confidence interval estimate of $I$ is $1.3311 \pm 0.0086$. 
When we used rejection sampling with $n=100,000$ draws, our margin of error was $\pm 0.0076$. Using importance sampling with $n=100,000$ draws, our margin of error was $\pm 0.0086$. Can we do better?
When we used rejection sampling with $n=100,000$ draws, our margin of error was $\pm 0.0076$. Using importance sampling with $n=100,000$ draws, our margin of error was $\pm 0.0086$. Can we do better?

How can we choose our importance function $g$ so that our margin of error will be smaller?
Let’s again consider the definite integral $I \equiv \int_{-1}^{1} f(x)dx$, where $f(x) = 1 - x^2$. This time our importance function will be the normal distribution with $\mu = 0$ and $\sigma = 0.4$.

Figure: $f(x)$ and $g(x)$.
Figure: Here we sampled $n = 100,000$ points. A 99% confidence interval estimate of $I$ is $1.3337 \pm 0.0037$. 
How do you choose a good function $g$ for an importance sampler?
How do you choose a good function $g$ for an importance sampler?

When we get our collection of estimates $h_i$, we’d like them to have as little variability as possible.
How do you choose a good function \( g \) for an importance sampler?

When we get our collection of estimates \( h_i \), we’d like them to have as little variability as possible.

Recall that each estimate \( h_i \) is defined by \( h_i \equiv \frac{f(x_i)}{g(x_i)} \). How do we choose \( g \) so as to minimize the variability in these estimates?
How do you choose a good function $g$ for an importance sampler?

When we get our collection of estimates $h_i$, we’d like them to have as little variability as possible.

Recall that each estimate $h_i$ is defined by $h_i \equiv \frac{f(x_i)}{g(x_i)}$. How do we choose $g$ so as to minimize the variability in these estimates?

We choose a function $g$ whose shape is close to the shape of $f$. 
What would happen if we let \( g \) be the normal distribution with \( \mu = 0 \) and \( \sigma = 0.2 \)?
In this non-standard histogram, the vertical scale is logarithmic.

**Figure:** Here we sampled $n = 100,000$ points. A 99% confidence interval estimate of $I$ is $1.3068 \pm 0.0857$. 
Summary of importance sampling

The steps of importance sampling are:

- Choose a probability density function $g$ that has a shape similar to $f$.
- Draw many $x_i$’s from $g$, and for each one compute the ratio $h_i = f(x_i)/g(x_i)$.
- The $h_i$’s are all unbiased estimators of $I$, so from them we may obtain an estimate of $I$ with a margin of error.
When would you want to use Monte Carlo integration? If an integral is hard, can’t you just use a deterministic method, such as trapezoids?
When would you want to use Monte Carlo integration? If an integral is hard, can’t you just use a deterministic method, such as trapezoids?

The main problem arises when the integral is over a multidimensional space. With even as few as three or four dimensions, (hyper-)trapezoids become too numerous to be computed efficiently. But Monte Carlo integration does not become any less efficient in higher dimensions.
Let \( f(w, x, y, z) = |\sin(wxz)| e^{-\sqrt{w^2+x^2+y^2+z^2}} \).

\[
\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(w, x, y, z) \, dw \, dx \, dy \, dz = ?
\]
Letting $g$ be the multivariate $t_{\nu=4}$ distribution with $\Sigma = 3I_{(4)}$, and taking $n = 100,000$ samples, we estimate the integral to be approximately $33.77 \pm 0.32$ (with 99% confidence).
Thank you for coming!
bullard@ncssm.edu