Computational Methods in Finance, Lecture 5, Importance sampling.

Jonathan Goodman *
Courant Institute of Mathematical Sciences, NYU

October 29, 1997

Importance sampling is a variance reduction technique. It is useful in situations where we want

$$\mathbf{E}[\phi(X)] = \int \phi(x) f(x) dx$$

($f(x)$ is the probability density function for X.) but the expectation is dominated by rare events. For example, suppose X is one of the integers 1, 2, ..., 100, each being equally likely. Now suppose the payout is 1000 if $X = 100$ but 0 otherwise. Then the expected payout is 10 but most of the samples have $X \neq 100$ and no payout. The event $X = 100$ is the rare event which, in this case, accounts for all the expected value.

The importance sampling method is to sample from a probability density other than f and correct for the discrepancy by measuring a quantity other than ϕ. Suppose $f(x)$ is the correct density and $g(x)$ is a different one. Then

$$\begin{aligned}
\mathbf{E}\left[\phi(X)\right] &= \int \phi(x) f(x) dx \\
&= \int \phi(x) \frac{f(x)}{g(x)} g(x) dx \quad .
\end{aligned}$$

*goodman@cims.nyu.edu, or http://www.math.nyu.edu/faculty/goodman,

This can be interpreted as

$$\mathbf{E}_f\left[\phi(X)\right] = \mathbf{E}_g\left[\tilde{\phi}(X)\right] \quad , \quad \text{where } \tilde{\phi}(x) = \phi(x)f(x)/g(x). \tag{1}$$

This formula says that you get the same expected value whether you sample from f and measure $\phi(x)$ or sample from g and measure $\tilde{g}$. The difference is that the (f,ϕ) method may larger (or smaller, if you mess up) than the $(g,\tilde{\phi})$ variance.

The art of importance sampling is to choose a density, g, that makes the important events less rare. This may be done systematically in some cases (for example, using the theory of large deviations) or using guesswork based on intuition. Intuitions are strong for one dimensional sampling problems, especially when aided by plots of f and $\phi(x)f(x)$. In high dimensions, or when solving stochastic differential equations, intuitions become harder to come by.

As an example, suppose X is a standard normal and we want $\mathbf{E}[e^{\lambda X}]$ with large λ. The mean value is

$$\begin{aligned} \mathbf{E}\left[e^{\lambda X}\right] &= \frac{1}{2\pi}\int e^{\lambda x}e_{-x^2/2}dx \\ &= \frac{1}{2\pi}\exp\left\{-\left(\frac{x^2}{2} - x\lambda + \frac{\lambda^2}{2} - \frac{\lambda^2}{2}\right)\right\}dx \\ &= e^{\lambda^2/2} \quad . \end{aligned}$$

The variance is

$$\begin{aligned} \operatorname{var}\left(e^{\lambda X}\right) &= \mathbf{E}\left[e^{2\lambda X}\right] - \mathbf{E}\left[e^{\lambda X}\right]^2 \\ &= e^{2\lambda^2} - e^{\lambda^2} \\ &\approx e^{2\lambda^2} \end{aligned}$$

for large λ. Therefore, the standard deviation is roughly e^{λ^2}, which is much larger than the mean. For $\lambda = 5$ the standard deviation is roughly $270,000$ time larger than the mean. This is because X is usually so small that $e^{\lambda X}$ is not large, but the occasional outliers contribute a lot.

one way to make outliers more common is to increase the variance of X. Let us take X to be gaussian with mean 0 but variance $\sigma^2 > 1$. This corresponds to

$$g(x) = \frac{1}{2\pi\sigma^2}e^{\frac{-x^2}{2\sigma^2}} \quad .$$

The general importance sampling formula, (1) specializes in this case to

$$\mathbf{E}_1\left[e^{\lambda X}\right] = \sigma \mathbf{E}_\sigma\left[e^{\lambda X} e^{-X^2/2} / x^{-x^2/2\sigma^2}\right] \quad .$$

(Where does the σ factor come from?) Here we use $\mathbf{E}_\sigma$ to be the expected value when X is normal with mean zero and variance σ^2. **Exercise:** find the formula for the variance as a function of σ and λ. The result is that the standard deviation is on the order of $\sigma e^{\lambda^2/2}$ when σ is large. For $\sigma = 10$, this is $27,000$ times smaller than the naive direct sampling approach.