Numerical Methods

Variance Reduction Techniques

Oleg Kozlovski

In the Monte-Carlo method the error is proportional to $\frac{\sigma}{\sqrt{n}}$ where n is the sample size. So, the amount of computational work required is proportional to the square of the desired precision. To reduce a amount of computations keeping the same precision one can try several techniques to decrease σ. These are referred to as variance reduction methods.

1 Antithetic reduction methods

Suppose that from i.i.d. random variables $X_1, \ldots, X_n$ one can easily obtain another sequence of i.i.d. random variables $Y_1, \ldots, Y_n$ but so that Y_i have the same distribution as X_i, but X_i and Y_i are actually dependent, and $Z_i = (X_i + Y_i)/2$ are i.i.d.

Then EX_i=EY_i=EZ_i, and the variance of Z_i is

$$\operatorname{var} Z_i = \frac{\operatorname{var}(X_i) + \operatorname{var}(Y_i) + 2\operatorname{cov}(X_i, Y_i)}{4}.$$

So if we can organise it so that $\operatorname{cov}(X_i, Y_i) < 0$ then one decreases the variance of the sample average $\bar{Z}^n$.

Suppose that x_i are i.i.d. uniformly distributed on $[0, 1]$ and $f : [0, 1] \to \mathbb{R}$ is a function. Let $X_i = f(x_i)$.

One common way of achieving a reduction in variance is by taking $Y_i = f(1 - x_i)$. This does NOT always work, but can give good improvement.

For example, take

$$\int_0^1 x \sin(x) \sqrt{1 + x^2}\, dx.$$

Then one can obtain a huge reduction in the variance estimator.

```
% the usual average
U=rand(1,1000);
X=svsfun(U);
Im=mean(X)
Sm=var(X)
```

```
%the antithetic average
U1=rand(1,1000);U2=1-U1;
X1=svsfun(U1);
X2=svsfun(U2);
X=(X1+X2)/2;
Ia=mean(X)
Sa=var(X)

% the integral computed by quadrature.
I=quad(@svsfun,0,1)
```

and where svsfun is defined in the separate file called svsfun.m

```
function y=svsfun(x)
y=x.*sin(x).*sqrt(1+x.^2);
```

In the first "naive" case we get $\sigma = 0.92$ and in the second "antithetic" case 0.21. The integral is 1.1272.

The reason this variance reduction works is because the function f is monotone. (It is not too hard to show that if $\int_0^1 f(t)\,dt = 0$ and f is monotone, then $\int_0^1 f(t) * f(1-t)\,dt < 0$.) If f is not monotone, then the above method certainly need NOT give variance reduction.

For example, consider

$$\int_0^1 x \sin(x) \sqrt{1 - x^2}\, dx.$$

In this case the integral is 0.5708, the σ is 0.3112 in the first "naive" case and 0.1415 in the second case when the antithetic method is applied.

1.1 Antithetic method for European calls

Let us recall that European call options can be priced by the Monte-Carlo method in the following way: its price is the expectation of the discounted payoff function

$$f(S) = \exp(-rT) \max(0, S - X),$$

where S is distributed as

$$S(T) = S(0) \exp((r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}\epsilon)$$

with $\epsilon \sim N(0, 1)$.

Notice that $S(T)(\epsilon)$ and $S(T)(-\epsilon)$ are negatively correlated and this can be used to reduce the variance. The algorithm would look like:

1. Compute n pseudo random $N(0, 1)$ numbers.

2. Compute $S(T)(\epsilon)$ and $S(T)(-\epsilon)$ for each sample ϵ by formula above.
3. Compute the average $(f(S(T)(\epsilon)) + (f(S(T)(-\epsilon)))/2$.
4. Average the result.

Notice that for given n we need to perform slightly less than twice as many operations as for the plain Monte-Carlo method. We need the variance to be reduced by factor at least 2 to gain something.

Exercise:

Implement the Antithetic method for European call options. Check how much the variance is reduced.

2 Importance sampling

This method exploits the idea that one would get a better estimate for $E(X)$ when one samples in particular in the parts of the space where X is large.

Suppose we want to evaluate integral

$$I = \int_0^1 f(t)\, dt.$$

Then

$$I = E(f),$$

where the expected value is taken with respect to the uniform distribution.

Let g be any positive probability density function defined on $[0, 1]$. Then

$$I = \int_0^1 \frac{f(t)}{g(t)}\, g(t)dt = E_g(\frac{f}{g}).$$

Thus we have another estimator of I:

$$\bar{I} = \frac{1}{n}\sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)},$$

where X_i are i.i.d. with density function g.

The variance is

$$\mathrm{var}(\bar{I}) = \frac{1}{n} \operatorname*{var}_g(\frac{f}{g}).$$

It is clear that if f is positive and g is proportional to f, then $\frac{f}{g}$ is a constant and $\mathrm{var}_g(\frac{f}{g}) = 0$. In this case optimal g is

$$g(t) = \frac{f(t)}{\int_0^1 f(z)\, dz}.$$

Obviously, the denominator here is I, which is what we require to find.

However, if the shape of g mimics the shape of f and it is easy to produce samples with density g, this method can give a significant variance reduction and be numerically efficient.

2.1 Importance sampling method for European calls

The payoff function for the European call options is zero when $S \leq X$, where X is the strike price. This can be used to reduce the variance of the Monte-Carlo method using the Importance sampling method.

The European call option price is given by

$$P_c = \int_{-\infty}^{\infty} f(S(T)(\epsilon))\phi(\epsilon)\, d\epsilon.$$

We have used here the same notation as above. Function ϕ denotes the density of the normal $N(0,1)$ distribution.

This integral is equal to

$$P_c = \int_{\epsilon_0}^{\infty} f(S(T)(\epsilon))\phi(\epsilon)\, d\epsilon,$$

where ϵ_0 is defined from the equation

$$S(T)(\epsilon_0) = X.$$

Let $\hat{\epsilon}$ be a random variable which distribution function is equal to zero if $\hat{\epsilon} <= \epsilon_0$ and is proportional to ϕ otherwise. Then

$$P_c = E_{\hat{\epsilon}}(\hat{I}\, f(S(T)(\hat{\epsilon}))),$$

where

$$\hat{I} = \int_{\epsilon_0}^{\infty} \phi(\epsilon)\, d\epsilon$$

is the normalising factor.

How cat we generate pseudo random numbers distributed as $\hat{\epsilon}$? First, notice that if u is a random variable uniformly distributed on $[0,1]$, then $\Phi^{-1}(u)$, where Φ denotes the $N(0,1)$ cumulative distribution function, is distributed $N(0,1)$ normally.

It is easy to check that if $\hat{u}$ is uniformly distributed on $[\Phi(\epsilon_0), 1]$, then $\Phi^{-1}(\hat{u})$ is distributed as $\hat{\epsilon}$.

Notice that

$$\hat{I} = 1 - \Phi(\epsilon_0).$$

Exercise:

Implement the Importance sampling method for European call options. Check how much the variance is reduced.

3 Stratified sampling

The idea of this is to reduce the variance of the Monte Carlo estimator by stratifying the sampling. For example, instead of sampling 10000 numbers uniformly in $[0,1]$, one samples 100 points uniformly in each of the 100 intervals $(i/100,(i+1)/100)$, $i=0,1,\dots,99$. What is the point of this?

More abstractly, suppose we want to compute

$$I=\int_\Omega f\,dt$$

where for example $\Omega=[0,1]^k$. Then take some partition $\Omega_1,\dots,\Omega_M$ of Ω and let a_j be the volume of Ω_j (so $a_1+\cdots+a_M=1$). Let X be the unstratified sampler with sample size n:

$$X=\frac{1}{n}\sum_{i=0}^{n}f(u_i)$$

with u_i i.i.d. uniform on Ω.

Let $n=n_1+\cdots+n_M$ with $n_j>0$,

$$Y_j=\frac{1}{n_j}\sum_{i=0}^{n_j}f(u_i^j)$$

with u_i^j uniform in Ω_j and let $Y=a_1Y_1+\cdots+a_MY_M$ be the stratified sampler. Then one can show that $E(X)=E(Y)=I$, $E(Y_j)=E(f(t)|t\in\Omega_j)=I_j$ where $I_j=\frac{1}{a_j}\int_{\Omega_j}f\,dt$.

Y_j are independent, $f(u_i^j)$ are independent as well, so

$$\mathrm{var}(Y)=\sum_{j=1}^{M}a_j^2\,\mathrm{var}(Y_j)=\sum_{j=1}^{M}\frac{a_j^2}{n_j}\,\mathrm{var}(f(u^j)).$$

A common scheme is *proportional stratified sampling* where $n_i=na_i$. In this case the previous formula becomes

$$\mathrm{var}(Y)=\frac{1}{n}\sum_{j=1}^{M}a_j\,\mathrm{var}(f(u^j)).$$

Now let us compare stratified variance with the variance of X.

$$\begin{aligned}
\mathrm{var}(X) &= \frac{E(f(u)^2) - I^2}{n} \\
&= \frac{1}{n}\left[\sum_{j=1}^{M} a_j E(f(u)^2 | u \in \Omega_j) - I^2\right] \\
&= \frac{1}{n}\left[\sum_{j=1}^{M} a_j (\mathrm{var}(f(u^j)) + I_j^2) - I^2\right] \\
&= \frac{1}{n}\sum_{j=1}^{M} a_j \,\mathrm{var}(f(u^j)) + \frac{1}{n}\sum_{j=1}^{M} a_j (I_j - I)^2
\end{aligned}$$

Thus we see that the variance of the stratified method is smaller than the variance of the naive method.

The choice $n_j = na_j$ is not optimal. Using Lagrange multipliers method one can compute that the optimal choice is

$$n_j = \frac{na_j \,\mathrm{var}(f(u^j))}{\sum_{j=1}^{M} a_j \,\mathrm{var}(f(u^j))}.$$

3.1 Stratified sampling for European calls

To apply the Stratified sampling technique to price the European call options we can use ideas similar to ones used for the Importance sampling method. In fact, we can combine both these methods.

Recall that

$$P_c = E_{\hat{\epsilon}}(\hat{I}\, f(S(T)(\hat{\epsilon}))).$$

Using random variable $\hat{u}$ we can rewrite this expression as

$$P_c = E_{\hat{u}}(\hat{I}\, f(S(T)(\Phi^{-1}(\hat{u})))\hat{I}).$$

The application of the Stratified sampling method to this formula is straightforward.

Exercise:

Implement the Stratified sampling method for European call options.

3.2 Post stratification

This method is very similar to the stratification method. The only difference is that the number n_j of samples in a stratum is not fixed and counted after a naive sampling was performed.

4 Control variates

Suppose that we know that $E(Y) = \nu$. The idea now is that instead of estimating $E(X)$, we may as well estimate

$$E(X_{contr})$$

where $X_{contr} = X + c \cdot (Y - \nu)$.

Clearly, $E(X_{contr}) = EX$, and

$$\mathrm{var}(X_{contr}) = \mathrm{var}(X) + c^2 \,\mathrm{var}(Y) + 2c \,\mathrm{cov}(X, Y).$$

Whether this makes sense, depends on whether we can find c reliably so that

$$\mathrm{var}(X_{contr}) < \mathrm{var}(X).$$

The optimal choice for c is

$$c = \frac{-\,\mathrm{cov}(X, Y)}{\mathrm{var}(Y)}.$$

The resulting variance is

$$\begin{aligned} \mathrm{var}(X_{contr}) &= \mathrm{var}(X) - \mathrm{cov}(X, Y)^2 / \mathrm{var}(Y) \\ &= \mathrm{var}(X)(1 - \mathrm{corr}(X, Y)^2). \end{aligned}$$

The challenge is to choose a good Y which is well correlated with X. The covariance, and hence the optimal coefficient c, can be estimated from the data.

4.1 Control variates method for European calls

We should try to find a random variable highly correlated to the payoff function and whose expectation we can directly compute.

The simplest choice is to start with $S(T) - X$ itself. From the formulas above it immediately follows that the results off the application of the Control variates method does not change if one shifts the control variate by a constant, so we can simply take $S(T)$.

In the case of the geometrical Brownian motion model we know that

$$E(S(T)) = S(0)e^{rT}.$$

The variance of $S(0)$ which we need to compute optimal c is also known and given by the formula:

$$\mathrm{var}(S(T)) = S(0)^2 e^{2rT}(e^{\sigma^2 T} - 1).$$

The covariance of $f(S(T))$ and $S(T)$ we can estimate empirically.

Exercise:

Implement the Control variates method for European call options. Check how much the variance is reduced.