1 Introduction to reducing variance in Monte Carlo simulations

1.1 Review of confidence intervals for estimating a mean

In statistics, we estimate an unknown mean $\mu = E(X)$ of a distribution by collecting n iid samples from the distribution, $X_1, \ldots, X_n$ and using the sample mean

$$\overline{X}(n) = \frac{1}{n}\sum_{j=1}^{n} X_j. \tag{1}$$

Letting $\sigma^2 = Var(X)$ denote the variance of the distribution, we conclude that

$$Var(\overline{X}(n)) = \frac{\sigma^2}{n}. \tag{2}$$

The *central limit theorem* asserts that as $n \to \infty$, the distribution of
$Z_n \stackrel{\text{def}}{=} \frac{\sqrt{n}}{\sigma}(\overline{X}(n) - \mu)$ tends to $N(0,1)$, the unit normal distribution. Letting Z denote a $N(0,1)$ rv, we conclude that for n sufficiently large,
$Z_n \approx Z$ in distribution. From here we obtain for any $z \geq 0$,

$$P(|\overline{X}(n) - \mu| > z\frac{\sigma}{\sqrt{n}}) \approx P(|Z| > z) = 2P(Z > z).$$

(We can obtain any value of $P(Z > z)$ by referring to tables, etc.)

For any $\alpha > 0$ no matter how small (such as $\alpha = 0.05$), letting $z_{\alpha/2}$ be such that $P(Z > z_{\alpha/2}) = \alpha/2$, we thus have

$$P(|\overline{X}(n) - \mu| > z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) \approx \alpha,$$

which implies that the unknown mean μ lies within the interval $\overline{X}(n) \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ with (approximately) probability $1 - \alpha$.

This allows us to construct *confidence intervals* for our estimate:

> *we say that the interval* $\overline{X}(n) \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ *is a* $100(1-\alpha)\%$ *confidence interval for the mean* μ.

Typically, we would use (say) $\alpha = 0.05$ in which case $z_{\alpha/2} = z_{0.025} = 1.96$, and we thus obtain a 95% confidence interval $\overline{X}(n) \pm (1.96)\frac{\sigma}{\sqrt{n}}$.

The length of the confidence interval is $2(1.96)\frac{\sigma}{\sqrt{n}}$ which of course tends to 0 as the sample size n gets larger.

In practice we would not actually know the value of σ^2; it would be unknown (just as μ is). But this is not really a problem: we instead use an estimate for it, the *sample variance* $s^2(n)$ defined by

$$s^2(n) = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2.$$

It can be shown that $s^2(n) \to \sigma^2$, with probability 1, as $n \to \infty$ and that $E(s^2(n)) = \sigma^2,\ n \geq 2$.

So, in practice we would use $s(n)$ is place of σ when constructing our confidence intervals. For example, a 95% confidence interval is given by $\overline{X}(n) \pm (1.96)\frac{s(n)}{\sqrt{n}}$.

The following recursions can be derived; they are useful when implementing a simulation requiring a confidence interval:

$$\overline{X}_{n+1} = \overline{X}_n + \frac{X_{n+1} - \overline{X}_n}{n+1},$$

$$S^2_{n+1} = \Big(1 - \frac{1}{n}\Big) S^2_n + (n+1)(\overline{X}_{n+1} - \overline{X}_n)^2.$$

1.2 Application to Monte Carlo simulation

In Monte Carlo simulation, instead of "collecting" the iid data $X_1, \ldots, X_n$, we simulate it. Moreover, we can choose n as large as we want; $n = 10,000$ for example, so the central limit theorem justification for constructing confidence intervals can safely be used. Thus we can immediately obtain confidence intervals for Monte Carlo estimates.

But simulation also allows us to be clever: We can purposely try to induce negative correlation among the variables $X_1, \ldots, X_n$, or generate copies that while having the same mean, have a smaller variance, so that the variance of the estimator in (1) becomes smaller than $\frac{\sigma^2}{n}$ resulting in a smaller confidence interval. The idea is to try to get even better estimates by reducing the uncertainty in our estimate. In the next sections, we explore ways of doing this.

1.3 Antithetic variates method

Let X_i denote our copies of X (each has the same distribution hence the same mean μ and variance σ^2) but let us not assume that they are independent. Let $n = 2m$, for some $m \geq 1$, that is, n is even. Note that

$$\overline{X}(n) = \frac{1}{2m} \sum_{j=1}^{2m} X_j = \frac{1}{m} \sum_{j=1}^{m} Y_j = \overline{Y}(m), \tag{3}$$

where

$$\begin{aligned} Y_1 &= \frac{X_1 + X_2}{2} \\ Y_2 &= \frac{X_3 + X_4}{2} \\ &\vdots \\ Y_m &= \frac{X_{n-1} + X_n}{2}, \end{aligned}$$

and we conclude that

The two estimators $\overline{Y}(m)$ and $\overline{X}(n)$ for $E(X)$ in (3) are identical.

Because they are identical, we can and will use $\overline{Y}(m)$ in what follows. Moreover, $E(Y_i) = E(X) = \mu$ (remember we are assuming that the X_i all have the same distribution hence the same mean). This means that for purposes of argument here we can view each Y_i as the end

"copy" that we wish to simulate from (instead of the X_i). We let $Y = \frac{X_1+X_2}{2}$ denote a generic Y_i. The problem of estimation can be re-cast as "we are trying to estimate $\mu = E(Y)$".

Computing variances,

$$\begin{aligned} Var(Y) &= (1/4)(\sigma^2 + \sigma^2 + 2Cov(X_1, X_2)) \\ &= (1/2)(\sigma^2 + Cov(X_1, X_2)). \end{aligned}$$

In the case when the X_i are iid, $Cov(X_1, X_2) = 0$ and thus $Var(Y) = \sigma^2/2$ yielding (as we already know, recall (2)) $Var(\overline{Y}(m)) = \frac{\sigma^2}{n}$.

But if $Cov(X_1, X_2) < 0$, then $Var(Y) < (1/2)\sigma^2$ yielding $Var(\overline{Y}(m)) < \frac{\sigma^2}{n}$; variance is reduced. So it is in our interest to somehow create some negative correlation within each pair (X_1, X_2), $(X_3, X_4), \ldots$, but keep the pairs iid so that the Y_i are iid (and thus the CLT still applies); for then $Var(\overline{Y}(m))$ will be lowered from what it would be if we simply used iid copies of the X_i.

To motivate how we might create the desired negative correlation, recall that we can generate an exponentially distributed rv $X_1 = -(1/\lambda)\ln(U)$ with U uniformly distributed on $(0, 1)$. Now instead of using a new independent uniform to generate a second such copy, use $1 - U$ which we well know is also uniformly distributed on $(0, 1)$; that is, define $X_2 = -(1/\lambda)\ln(1 - U)$. Clearly X_1 and X_2 are negatively correlated since if U increases, then $1 - U$ decreases and the function $\ln(y)$ is an increasing function of y: X_1 increases iff U increases iff $1 - U$ decreases iff X_2 decreases. More generally, for any distribution $F(x) = P(X \le x)$ with inverse $F^{-1}(y)$ we could generate a negatively correlated pair via $X_1 = F^{-1}(U)$, $X_2 = F^{-1}(1 - U)$ since $F^{-1}(y)$ is a monotone increasing function of y. The random variables U and $1 - U$ have a correlation coefficient $\rho = -1$, they are negatively correlated (to the largest extent), thus the monotonicity preserves the property of negative correlation; $\rho_{X_1,X_2} < 0$ (not necessarily -1 though).

In a general Monte Carlo simulation our X is of the form $X = h(U_1, \ldots, U_k)$, for some (perhaps very complicated) function h, and some k (perhaps large), that is, we need k iid U_i to generate each copy of X. For example, if we are considering

$$X = C_2 = (\frac{1}{2}\sum_{i=1}^{2} S_i - K)^+,$$

the payoff at time $T = 2$ of an Asian call option under the binomial lattice model, then re-writing

$$\frac{1}{2}\sum_{i=1}^{2} S_i = (1/2)S_0 Y_1[1 + Y_2],$$

where the Y_i are the iid up-down rvs, we have

$$h(U_1, U_2) = \Big((1/2)S_0(uI\{U_1 \le p\} + dI\{U_1 > p\})[1 + (uI\{U_2 \le p\} + dI\{U_2 > p\})] - K\Big)^+.$$

This function is monotone decreasing in U_1 and U_2 : as either variable increases, they will exceed the value p and hence the indicators will tend towards the lower value d as opposed to the higher value $u > d$. Because the vectors (U_1, U_2) and $(1 - U_1, 1 - U_2)$ are identically distributed, so are the rvs $X_1 = h(U_1, U_2)$ and $X_2 = h(1 - U_1, 1 - U_2)$; in particular they have the same mean $E(X)$. But the monotonicity of h results in negative correlation between them, $Cov(X_1, X_2) < 0$.

In general, as long as the function h is monotone (either increasing or decreasing) in each variable, then it can be shown that $X_1 = h(U_1, \ldots, U_k)$ and $X_2 = h(1 - U_1, \ldots, 1 - U_k)$ are

indeed negatively correlated, and are referred to as *antithetic variates*. Again, because the vectors $(U_1, U_2, \ldots U_k)$ and $(1-U_1, 1-U_2, \ \ldots 1-U_k)$ have the same distribution, so do X_1 and X_2; in particular they have the same mean $E(X)$. But because of the induced negative correlation (when h is monotone) the two are themselves negatively correlated copies:

Proposition 1.1 *If the function h for generating $X = h(U_1, \ldots, U_k)$ is monotone in each variable, then $X_1 = h(U_1, \ldots, U_k)$ and $X_2 = h(1-U_1, \ldots, 1-U_k)$ with the U_i iid uniform on $(0,1)$ are in fact negatively correlated; $Cov(X_1, X_2) < 0$.*
(Equivalently $E(X_1X_2) < E(X_1)E(X_2) = E^2(X)$.)

Algorithm for using antithetic variates to estimate $\mu = E(X)$, when $X = h(U_1, \ldots, U_k)$ is monotone in the U_i:

The method of simulating our pairs is straightforward:

1. Generate $U_1, \ldots U_k$. Construct a first pair: Set $X_1 = h(U_1, \ldots, U_k)$ and $X_2 = h(1 - U_1, \ldots, 1-U_k)$.
2. Now independently generate k new iid uniforms to construct another pair X_3, X_4 and so on pair by pair until reaching m (large) desired pairs.
3. Use the estimate
$$\overline{Y}(m) = \sum_{j=1}^{m} Y_j,$$
where
$$\begin{aligned} Y_1 &= \frac{X_1 + X_2}{2} \\ Y_2 &= \frac{X_3 + X_4}{2} \\ &\vdots \\ Y_m &= \frac{X_{2m-1} + X_{2m}}{2}. \end{aligned}$$

To construct our (new and better) confidence interval:

Define the sample variance as
$$s^2(m) = \frac{1}{m-1} \sum_{j=1}^{m} (Y_j - \overline{Y}_m)^2.$$

Then the interval $\overline{Y}(m) \pm z_{\alpha/2} \frac{s(m)}{\sqrt{n}}$ is a $100(1-\alpha)\%$ confidence interval for the mean μ.

As a very simple example recall that we can estimate π by observing that $\pi/4 = E(\sqrt{1-U^2})$. Since $h(y) = \sqrt{1-y^2}$ is monotone decreasing in y, we can use antithetic variates. Thus we would use $X_1 = \sqrt{1-U_1^2}$, $X_2 = \sqrt{1-(1-U_1)^2}$ for our first pair, $X_3 = \sqrt{1-U_2^2}$, $X_4 = \sqrt{1-(1-U_2)^2}$ and so on.

Remark 1.1 In a real simulation application, computing exactly $Cov(X_1, X_2)$ when X_1 and X_2 are antithetic is never possible in general; after all, we do not even know (in general) either $E(X)$ or $Var(X)$. But this is not important since our objective was only to reduce the variance, and we accomplished that.

1.4 Antithetic normal rvs

In many finance applications, the fundamental rvs needed to construct a desired output copy X are unit normals, $Z_1, Z_2, \ldots$. For example, when using geometric Brownian motion for asset pricing, our payoffs typically can be written in the form $X = h(Z_1, \ldots, Z_k)$. Noting that $-Z$ is also a unit normal if Z is, and that the correlation coefficient between them is $\rho = -1$, the following is the Gaussian analogue to Proposition 1.1

Proposition 1.2 *If the function h for generating $X = h(Z_1, \ldots, Z_k)$ is monotone in each variable, then $X_1 = h(Z_1, \ldots, Z_k)$ and $X_2 = h(-Z_1, \ldots, -Z_k)$ with the Z_i iid $N(0,1)$ are in fact negatively correlated; $Cov(X_1, X_2) < 0$.*

Example with an Asian call option:

Suppose for example you wish to estimate the expected payoff of an Asian call option (termination date T) averaged over the time points $0 = t_0 < t_1 < t_2 < \cdots < t_k = T$. The payoff is then

$$X = C_T = (\frac{1}{k}\sum_{i=1}^{k} S(t_k) - K)^+.$$

We next show how to construct the antithetic pairs.

1. Generate k iid $N(0,1)$ rvs, $Z_1, \ldots, Z_k$. Set

$$L_i = e^{\sigma\sqrt{t_i - t_{i-1}}Z_i + \mu(t_i - t_{i-1})},\ i \in \{1, 2, \ldots, k\}. \tag{4}$$

2. Recursively set

$$\begin{aligned} S(t_1) &= S(0)L_1 \\ S(t_2) &= S(t_1)L_2 = S_0 L_1 \times L_2 \\ &\vdots \\ S(t_k) &= S(t_{k-1})L_k = S_0 L_1 \times L_2 \times \cdots \times L_k. \end{aligned}$$

Set

$$X_1 = (\frac{1}{k}\sum_{i=1}^{k} S(t_k) - K)^+.$$

3. Now reset the L_i in (5) by using $-Z_1, \ldots, -Z_k$ in place of $Z_1, \ldots, Z_k$, that is, set

$$L_i = e^{-\sigma\sqrt{t_i - t_{i-1}}Z_i + \mu(t_i - t_{i-1})},\ i \in \{1, 2, \ldots, k\}. \tag{5}$$

4. Recursively set

$$\begin{aligned} S(t_1) &= S(0)L_1 \\ S(t_2) &= S(t_1)L_2 = S_0L_1 \times L_2 \\ &\vdots \\ S(t_k) &= S(t_{k-1})L_k = S_0L_1 \times L_2 \times \cdots \times L_k. \end{aligned}$$

Set

$$X_2 = (\frac{1}{k}\sum_{i=1}^{k} S(t_k) - K)^+.$$

5. Set

$$Y = \frac{X_1 + X_2}{2}.$$

Denoting the above copy by $Y_1 = Y$, we can generate a second independent such copy by starting again at (1) with a new (independent) $Z_1, \ldots, Z_k$ and setting$Y_2 = Y$. Repeating this procedure m times yields our desired m iid copies of Y.

1.5 Control variates

Suppose we wish to estimate $\mu = E(X)$ using Monte Carlo simulation (e.g., using $\overline{X}(n)$ with iid copies of X). Letting C be any other rv, with mean $E(C)$, and b a constant, note that the rv Y given by

$$Y = X - b(C - E(C)), \tag{6}$$

has the same desired mean: $E(Y) = E(X)$. Thus, if we could simulate iid copies of Y, $Y_1, \ldots Y_n$, then we could use as our estimate $\overline{Y}(n)$ instead of $\overline{X}(n)$.

Noting further that

$$\sigma_Y^2 = Var(Y) = \sigma_X^2 + b^2\sigma_C^2 - 2b\sigma_{X,C}, \tag{7}$$

we see that by choosing C and b wisely, it might be possible to reduce variance, that is, to have $\sigma_Y^2 < \sigma_X^2$ thus resulting in the lower variance estimator $\overline{Y}(n)$ than the usual $\overline{X}(n)$. If X is non-negative, then this would amount to choosing $b > 0$ and selecting C and X to have high positive correlation, but in general many possibilities might come into play. Before we investigate this further, note that it would be very helpful if C was already part of the simulation of X in the sense that whenever we simulated a copy of X, a copy of C necessarily came out for free along the way. Also we want C to be such that we exactly know the value $E(C)$. The idea being that we do not want to have to increase our work. An example would be $X = ([S(t_1)+S(t_2)]/2-K)^+$ and $C = S(t_2)$ for an Asian call option.

This method of introducing such a C for purpose of reducing variance is the *control variates method*, and $C - E(C)$ is called the *control variate* for estimating $E(X)$.

For a given C, we can view (7) as a function of b, $g(b) = \sigma_Y^2(b)$, and then using elementary calculus, set $g'(b) = 2b\sigma_C^2 - 2\sigma_{x,C} = 0$ and solve for the minimum b^*. This yields:

$$b^* = \frac{\sigma_{X,C}}{\sigma_C^2}, \tag{8}$$

$$\sigma_Y^2(b^*) = \sigma_X^2(1 - \rho_{X,C}^2), \tag{9}$$

where $\rho_{X,C} = \sigma_{X,C}/(\sigma_X \sigma_C)$ denotes the correlation coefficient. Thus by choosing any C for which $\sigma_{X,C} \neq 0$ we can always reduce variance, and it is desirable to choose a C that is strongly correlated with X.

In practice we would not be able to compute the value of b^* exactly since it is unlikely that we would know $\sigma_{X,C}$ and maybe not even σ_C^2. But we could estimate it in advance by simulation: Choose n large and use

$$b^* \approx b^*(n) = \frac{\sum_{i=1}^n (X_i - \overline{X}(n))(C_i - E(C))}{\sum_{i=1}^n (C_i - E(C))^2}. \tag{10}$$

In other words we would first (just once) run a simulation (large n) to obtain the estimate $b^*(n)$, and then use that fixed value throughout our desired Monte Carlo simulation.

Examples with GBM

1. *Asian call option:* With $0 < t_1 < \cdots < t_k = T$, the payoff at time T is

$$X = (\frac{1}{k} \sum_{i=1}^k S(t_i) - K)^+.$$

   A natural choice for C is the stock itself at the terminal value, $C = S(T) = S(t_k)$. We certainly can compute $E(C)$ it is part of the simulation anyhow and is clearly positively correlated with X.

   Another choice would be the payoff of a European call $C = (S(T) - K)^+$, since in this case $E(C)$ is known exactly from the Black-Scholes options pricing formula. An even better choice (more correlation) would be to use $C = (\Pi_{i=1}^k S(t_i) - K)^+$, the payoff of a geometrically averaged Asian option. Here to, it turns out that the expected payoff $E(C)$ is exactly known (a formula exists), and because it incorporates all k values of the GBM, it yields a higher correlation with X.

   It should not be surprising that the choice of C might also depend on the strike price K. For example, if K is very small compared to $S(0)$, one would, with high probability, obtain a positive payoff of $\frac{1}{k}\sum_{i=1}^k S(t_i) - K$, yielding a high correlation with (say) any of the choices of C mentioned above, whereas if K is very large compared to $S(0)$ then one would, with high probability, obtain no payoff at all, thus yielding a low correlation with a choice of $C = S(T)$ (but still a high one with the geometrically averaged payoff above).

   There is no obvious best choice that works with all payoffs; one must take into consideration the specific structure of a payoff, and its parameters.

2. In payoffs with multiple assets such as a spread option, one can use a control that uses all (or some) of the assets. For example, for $X = (|S_1(T) - S_2(T)| - K)^+$ one might try $C = S_1(T) - S_2(T)$.