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

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

This is justified by the *strong law of large numbers (SLLN)*, which asserts that this estimate converges with probability one (wp1) to the desired $\mu = E(X)$, as $n \to \infty$. But the SLLN does not tell us how good the approximation is; we consider this next.

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:

$$\overline{X}(n+1) = \overline{X}(n) + \frac{X_{n+1} - \overline{X}(n)}{n+1},$$

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

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 (e.g., in statistics "out in the field" applications, one might only have $n = 25$ or $n = 50$ samples!). 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. 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 $E(Y_i) = E(X) = \mu$. This means that for purpose of estimating $\mu = E(X)$, 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,

$$
\begin{aligned}
(4)\quad Var(Y) &= (1/4)(\sigma^2+\sigma^2+2Cov(X_1,X_2)) \\
(5)\quad &= (1/2)(\sigma^2+Cov(X_1,X_2)), \text{ hence} \\
(6)\quad Var(\overline{Y}(m)) &= \frac{1}{n}(\sigma^2+Cov(X_1,X_2)) \\
(7)\quad &= \sigma^2/n+Cov(X_1,X_2))/n \\
(8)\quad &= Var(\overline{X}(n))+Cov(X_1,X_2))/n.
\end{aligned}
$$

In the case when the X_i are iid, $Cov(X_1,X_2)=0$ and thus $Var(\overline{Y}(m))=\frac{\sigma^2}{n}=Var(\overline{X}(n))$, and we get back to where we started in (2).

But if $Cov(X_1,X_2)<0$, then $Var(\overline{Y}(m))<\frac{\sigma^2}{n}$; 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 $X=D_2$ for the FIFO GI/GI/1 queue, by using the inverse transform method for the service (G^{-1}) and interarrival times (A^{-1}), then we need $k=4$ because we need to generate 2 service times and 2 interarrival times; assuming $D_0=0$ we can write this as $D_1=(S_0-T_0)^+=(G^{-1}(U_1)-A^{-1}(U_2))^+$ and then $D_2=(D_1+S_1-T_1)^+$ or

$$D_2=\Big[(G^{-1}(U_1)-A^{-1}(U_2))^+ + G^{-1}(U_3)-A^{-1}(U_4)\Big]^+.$$

Thus $D_2=h(U_1,U_2,U_3,U_4)$ where

$$h(y_1,y_2,y_3,y_4)=\Big[(G^{-1}(y_1)-A^{-1}(y_2))^+ + G^{-1}(y_3)-A^{-1}(y_4)\Big]^+.$$

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 $X_1=h(1-U_1,\ldots,1-U_k)$ are indeed negatively correlated, and are referred to as *antithetic variates*.

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. In the above example for D_2, this is easily established since each inverse function is monotone; $h(y_1, y_2, y_3, y_4)$ increases in y_1 and y_3 and decreases in y_2 and y_4.

As another 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, $S_n = S_0 Y_1 \cdots Y_n$, 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

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

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, $Cov(X_1, X_2) < 0$.

We summarize (without proof):

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 $X_2 = h(1 - U_1, \ldots, 1 - U_k)$. Define $Y_1 = [X_1 + X_2]/2$ and note that $E(Y_1) = E(X) = \mu$.
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, and m iid random variables $Y_j = [X_{2j-1} + X_{2j}]/2,\ 1 \leq j \leq m$. These Y_j have the same mean $E(X) = \mu$, but have a smaller variance because of (8).
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:

The sample variance for these Y_j is given by

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

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.

Examples

1. *Estimating π:* As a very simple example, note that we can estimate π by observing that π = the area of a disk of radius 1 ($\{(x, y) : x^2 + y^2 \leq 1\}$); $\pi/4 = \int_0^1 \sqrt{1 - x^2}dx = E(\sqrt{1 - U^2})$. So Monte Carlo can be used to estimate π by generating copies of $X = \sqrt{1 - U^2}$ and averaging. Since $h(x) = \sqrt{1 - x^2}$ is monotone decreasing in x, 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.

2. *Customer delay in a FIFO single-server queue:* As another example, consider the delay recursion for a FIFO GI/GI/1 queue:

$$D_{n+1} = (D_n + S_n - T_n)^+, \ n \geq 0,$$

where $\{T_n : n \geq 0\}$ are iid customer interarrival times distributed as $A(x) = P(T \leq x)$, $x \geq 0$ and independently $\{S_n : n \geq 0\}$ are iid customer service times distributed as $G(x) = P(S \leq x)$, $x \geq 0$. We assume here that both $A^{-1}(y)$ and $G^{-1}(y)$ are explicitly known so that the inverse transform method can be applied. Then $D_n = h(U_1, \ldots, U_n)$ can be written as a monotone in each variable function. For example, $D_1 = h(U_1, U_2) = (D_0 + G^{-1}(U_1) - A^{-1}(U_2))^+$, which is monotone increasing in U_1 and monotone decreasing in U_2. We can then write

$$D_2 = (D_1 + G^{-1}(U_3) - A^{-1}(U_4))^+,$$

which is thus monotone increasing in both U_1 and U_3, and monotone decreasing in U_2 and U_4. This same idea extends to D_n for any n. Thus if we wanted to estimate (say) $E(D_{10})$, the expected delay of the 10^{th} customer (when (say) $D_0 = 0$), we could do so as follows: Generate $(U_1, \ldots U_{10})$ and $(V_1, \ldots V_{10})$ as iid uniforms over $(0, 1)$. Construct a copy $X_1 = D_{10}$ via using the following recursion for $0 \leq n \leq 9$:

$$D_{n+1} = (D_n + G^{-1}(U_{n+1}) - A^{-1}(V_{n+1}))^+.$$

Now repeat the construction using $(1-U_1, \ldots 1-U_{10})$ and $(1-V_1, \ldots 1-V_{10})$ to get the second (antithetic) copy $X_2 = D_{10}$:

$$D_{n+1} = (D_n + G^{-1}(1-U_{n+1}) - A^{-1}(1-V_{n+1}))^+.$$

X_1 and X_2 are thus negatively correlated copies of D_{10} as desired.

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, $Z_1, Z_2, \ldots$. (As opposed to uniforms.) 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; $Cov(X_1, X_2) < 0$.*