Variance Reduction Techniques

Anna Battauz

Bocconi University

1 Variance Reduction Techniques

Aim: reduce the variance of the estimation

to gain efficiency

Radius of the confidence interval:

$$z_{\frac{\delta}{2}} \frac{\sigma}{\sqrt{n}}$$

where

- $z_{\frac{\delta}{2}}$ is the $\left(1 - \frac{\delta}{2}\right) -$quantile of the standard normal distribution
- σ is the standard deviation of one simulation
- n is the number of simulations

2 Different methods to achieve Variance Reduction:

2.1 Control Variate technique

2.2 Antithetic Variate

2.3 Matching underlying assets (Moment matching and Weighted Monte Carlo)

2.4 Importance sampling

3 Control variate technique: basic idea (G.4.1)

1. Monte Carlo estimator of known and unknown quantities.
2. Use the error in the estimate of *known quantities* to reduce the error in the estimator of the *unknown quantities*

To fix notations:
Suppose you want to estimate $\mathbb{E}[Y]$ and you know how to compute $\mathbb{E}[X]$

$$Y \text{ has unknown } \mathbb{E}[Y]$$

$$X \text{ has known } \mathbb{E}[X]$$

$$\text{Monte Carlo estimator for } \mathbb{E}[Y] = \widehat{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \text{ where } Y_i \sim Y \text{ i.i.d.}$$

4 Control variate technique

Use a smart combination of Y_i and X_i to reduce the variance of the estimator

$$Y_i(b) := Y_i - b\left(X_i - \underbrace{\mathbb{E}[X]}_{\text{known}}\right)$$

where $Y_i \sim Y$ and $X_i \sim X$ i.i.d and

$$\widehat{Y}_n(b) := \frac{1}{n}\sum_{i=1}^{n} Y_i(b)$$

$\widehat{Y}_n(b)$ is consistent and unbiased estimator for $\mathbb{E}[Y]$ for any $b \in \Re$

In fact:

$$\lim_{n\to\infty} \widehat{Y}_n(b) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - b\left(X_i - \mathbb{E}[X]\right)\right] = \lim_{n\to\infty}\left[\widehat{Y}_n - b\left(\widehat{X}_n - \mathbb{E}[X]\right)\right] = \mathbb{E}[Y] - 0$$

$$\mathbb{E}\left[\widehat{Y}_n(b)\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left[Y_i - b\left(X_i - \mathbb{E}\left[X\right]\right)\right]\right] = \mathbb{E}\left[Y\right] - b\left(\mathbb{E}\left[X\right] - \mathbb{E}\left[X\right]\right) = \mathbb{E}\left[Y\right] - 0$$

The parameter b has to be chosen to minimize the variance of each replication of $Y_i(b)$, so that the variance of the estimator

$$Var\left(\widehat{Y}_n(b)\right) = \frac{Var\left(Y_i(b)\right)}{n}$$

will be minimized as well. We show that the optimal parameter is

$$b^* = \frac{Cov(X,Y)}{Var(X)},$$

the regression coefficient of the *unknown* Y over the *known* X.
In fact:

$$\begin{aligned} Var\left(Y_i(b)\right) &= Var\left(Y_i - b\left(X_i - \mathbb{E}\left[X\right]\right)\right) = \\ &= Var\left(Y_i\right) + b^2 Var\left(\left(X_i - \mathbb{E}\left[X\right]\right)\right) - 2bCov\left(Y_i, \left(X_i - \mathbb{E}\left[X\right]\right)\right) \\ &= b^2\sigma_X^2 - 2b\rho_{XY}\sigma_X\sigma_Y + \sigma_Y^2 \end{aligned}$$

is a *parabola* with respect to b.
The parabola reaches the minimum in the vertex

$$b^* = \frac{2\rho_{XY}\sigma_X\sigma_Y}{2\sigma_X^2} = \frac{Cov(X,Y)}{Var(X)}$$

5 Does b^* reduce the variance?

The control variate estimator $\widehat{Y}_n(b^*)$ has variance

$$Var\left(\widehat{Y}_n(b^*)\right) = \frac{Var\left(Y_i(b^*)\right)}{n}.$$

Hence the control variate estimator is better than the usual Monte Carlo $\widehat{Y}_n$ whose variance is

$$Var\left(\widehat{Y}_n\right) = \frac{Var\left(Y_i\right)}{n} = \frac{\sigma_Y^2}{n}$$

if and only if

$$Var\left(Y_i(b^*)\right) < Var\left(Y_i\right)$$

We have

$$Var\left(Y_i\left(b^*\right)\right)=\left(\underbrace{\frac{\rho_{XY}\sigma_Y}{\sigma_X}}_{b^*}\right)^2\sigma_X^2-2\left(\underbrace{\frac{\rho_{XY}\sigma_Y}{\sigma_X}}_{b^*}\right)\rho_{XY}\sigma_X\sigma_Y+\sigma_Y^2=\sigma_Y^2\left(1-\rho_{XY}^2\right)$$

Hence the control variate reduces the variance if and only if

$$Var\left(Y_i\left(b^*\right)\right)=\sigma_Y^2\left(1-\rho_{XY}^2\right)<Var\left(Y_i\right)=\sigma_Y^2$$

that is if and only if

$$1-\rho_{XY}^2<1$$

$$\rho_{XY}^2>0$$

$$\left|\rho_{XY}\right|>0$$

The bigger the $\left|\rho_{XY}\right|$ the bigger the variance reduction:
Size of correlation ρ_{XY} matters, sign does not.

6 How to compute b^*?

To compute

$$b^* = \frac{2\rho_{XY}\sigma_X\sigma_Y}{2\sigma_X^2} = \frac{Cov(X,Y)}{Var(X)}$$

we need

$$Cov(X,Y)$$

If unknown, use the estimator

$$\widehat{b}_n = \frac{\sum_{i=1}^n \left(X_i - \widehat{X}_n\right)\left(Y_i - \widehat{Y}_n\right)}{\sum_{i=1}^n \left(X_i - \widehat{X}_n\right)^2} \underset{n\to\infty}{\longrightarrow} b^* = \frac{Cov(X,Y)}{Var(X)}$$

the slope of the least-square regression line through the points (X_i, Y_i) and through $\left(\widehat{X}_n, \widehat{Y}_n\right)$.

Example (G4.1.1.)

In the Black-Scholes model, the Monte Carlo estimator of a call option is

$$\widehat{Y}_n := \frac{1}{n}\sum_{i=1}^{n}\left(S_i\left(T\right)-K\right)^{+}e^{-rT}$$

where

$$S_i\left(T\right)\sim S_i\left(0\right)e^{\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}Z_i}$$

Z_i are i.i.d standard normal random variables.
Use as a control

$$X = S\left(T\right) \text{ with } \textit{known } \mathbb{E}\left[X\right] = S\left(0\right)e^{rT}$$

for the *unknown*

$$Y = \left(S\left(T\right)-K\right)^{+}e^{-rT}$$

The Control-Variate estimator

$$\begin{aligned}\widehat{Y}_n(b) \;:\; &= \frac{1}{n}\sum_{i=1}^{n} Y_i(b) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - b\left(X_i - \mathbb{E}[X]\right)\right]\\ &= \frac{1}{n}\sum_{i=1}^{n}\left[\left(S_i(T) - K\right)^{+} e^{-rT} - b\left(S_i(T) - S(0)\,e^{rT}\right)\right]\end{aligned}$$

for

$$b = \widehat{b}_n$$

significantly reduces the variance of the estimator and also the error (in this ad-hoc example $\mathbb{E}[Y]$ is known, so that the errors both for the control variate

$$\left|\widehat{Y}_n\left(\widehat{b}_n\right) - \mathbb{E}[Y]\right|$$

and for the pure Monte Carlo

$$\left|\widehat{Y}_n - \mathbb{E}[Y]\right|$$

can be computed explicitly.

7 Output analysis

7.1 Radius of the confidence interval for the control variate estimator $\widehat{Y}_n\left(\widehat{b}_n\right)$

$$b^* = \frac{Cov(X,Y)}{Var(X)},$$

$$\widehat{b}_n = \frac{\widehat{Cov}(X,Y)}{\widehat{Var}(X)} = \frac{\sum_{i=1}^n \left(X_i - \widehat{X}_n\right)\left(Y_i - \widehat{Y}_n\right)}{\sum_{i=1}^n \left(X_i - \widehat{X}_n\right)^2} \underset{n \to \infty}{\longrightarrow} b^* = \frac{Cov(X,Y)}{Var(X)}$$

almost surely

Hence

$$\sqrt{n}\left(\widehat{Y}_n\left(\widehat{b}_n\right)-\widehat{Y}_n\left(b^*\right)\right) = \sqrt{n}\left(\underbrace{\widehat{Y}_n-\widehat{b}_n\left(\widehat{X}_n-\mathbb{E}\left[X\right]\right)}_{\widehat{Y}_n\left(\widehat{b}_n\right)}-\underbrace{\left(\widehat{Y}_n-b^*\left(\widehat{X}_n-\mathbb{E}\left[X\right]\right)\right)}_{\widehat{Y}_n\left(b^*\right)}\right)$$

$$= \left(b^*-\widehat{b}_n\right)\sqrt{n}\left(\widehat{X}_n-\mathbb{E}\left[X\right]\right) \underset{n\to\infty}{\Rightarrow} N(0;1)$$

because

$$b^*-\widehat{b}_n \underset{n\to\infty}{\longrightarrow} 0 \text{ almost surely}$$

and

$$\sqrt{n}\left(\widehat{X}_n-\mathbb{E}\left[X\right]\right) \underset{n\to\infty}{\Rightarrow} N(0;\sigma_X^2)$$

Hence

$$\widehat{Y}_n\left(\widehat{b}_n\right)\approx\widehat{Y}_n\left(b^*\right)$$

and

$$\frac{\widehat{Y}_n\left(\widehat{b}_n\right)-\mathbb{E}\left[Y\right]}{\frac{\sigma\left(\widehat{b}_n\right)}{\sqrt{n}}}\underset{n\to\infty}{\sim}\frac{\widehat{Y}_n\left(b^*\right)-\mathbb{E}\left[Y\right]}{\frac{\sigma\left(b^*\right)}{\sqrt{n}}}\underset{n\to\infty}{\Rightarrow}N(0;1)$$

Therefore the radius of the confidence interval for the control variate estimator is

$$\frac{1}{\sqrt{n}}\cdot z_{\frac{\delta}{2}}\cdot\underbrace{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i\left(\widehat{b}_n\right)-\widehat{Y}_n\left(\widehat{b}_n\right)\right)^2}}_{\text{sample standard deviation of a replication } Y_i\left(\widehat{b}_n\right)}$$

7.2 Bias because of $\widehat{b}_n$ instead of b^*

$$\begin{aligned}\text{Bias}\left(\widehat{Y}_n\left(\widehat{b}_n\right)\right) &= \mathbb{E}\left[\widehat{Y}_n\left(\widehat{b}_n\right)\right]-\mathbb{E}\left[Y\right]=\mathbb{E}\left[\widehat{Y}_n-\widehat{b}_n\left(\widehat{X}_n-\mathbb{E}\left[X\right]\right)\right]-\mathbb{E}\left[Y\right]\\ &= -\mathbb{E}\left[\widehat{b}_n\left(\widehat{X}_n-\mathbb{E}\left[X\right]\right)\right]\neq-\mathbb{E}\left[\widehat{b}_n\right]\cdot\mathbb{E}\left[\widehat{X}_n-\mathbb{E}\left[X\right]\right]=0\end{aligned}$$

because $\widehat{b}_n$ and $\widehat{X}_n$ are not independent.
But to remove the bias there is a simple solution:

- First run n_1 i.i.d simulations of Y_i and X_i to estimator

$$\widehat{b}_{n_1} = \frac{\sum_{i=1}^{n_1} \left(X_i - \widehat{X}_{n_1}\right)\left(Y_i - \widehat{Y}_{n_1}\right)}{\sum_{i=1}^{n_1} \left(X_i - \widehat{X}_{n_1}\right)^2}$$

- Then run n_2 independent i.i.d simulations of Y_i and X_i to compute

$$Y_i\left(\widehat{b}_{n_1}\right) = Y_i - \widehat{b}_{n_1}\left(X_i - \mathbb{E}\left[X\right]\right)$$

and

$$\widehat{Y}_{n_2}\left(\widehat{b}_{n_1}\right) = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i\left(\widehat{b}_{n_1}\right)$$

- Now

$$\text{Bias}\left(\widehat{Y}_{n_2}\left(\widehat{b}_{n_1}\right)\right) = -\mathbb{E}\left[\widehat{b}_{n_1}\right]\cdot\mathbb{E}\left[\widehat{X}_{n_2} - \mathbb{E}\left[X\right]\right] = 0$$

because of the independent simulations.

If the procedure is too costly, in general remember that Bias$\sim O\left(\frac{1}{n}\right)$ whereas the standard error $\sim O\left(\frac{1}{\sqrt{n}}\right)$

8 Antithetic Variates $[$G.4.2$]$

The key idea of the antithetic variates technique is to exploit a negative dependence between pair of replications.
More precisely:

$$Z_i \sim N(0;1) \text{ and } Y_i = f(Z_i)$$

Compute also

$$\widetilde{Y}_i = f(-Z_i)$$

so that

$$Z_i \to \left(Y_i, \widetilde{Y}_i\right)$$

is a pair of antithetic variables (since you reversed the sign of Z_i).
Since $-Z_i \sim N(0;1)$, the variables Y_i and $\widetilde{Y}_i$ have the same distribution,
but they are not independent!

The Antithetic Monte Carlo estimator for Y is

$$\widehat{Y}_n^{AV} := \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i+\widetilde{Y}_i}{2} = \text{sample mean of } \frac{Y_i+\widetilde{Y}_i}{2}$$

The Antithetic Monte Carlo $\widehat{Y}_n^{AV}$ is consistent and unbiased

$$\lim_{n\to\infty}\widehat{Y}_n^{AV} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i+\widetilde{Y}_i}{2} = \lim_{n\to\infty}\frac{1}{2}\sum_{i=1}^{n}\frac{Y_i+\widetilde{Y}_i}{n} = \frac{1}{2}\left(\mathbb{E}\left[Y\right]+\mathbb{E}\left[\widetilde{Y}\right]\right) = \mathbb{E}\left[Y\right]$$

$$\mathbb{E}\left[\widehat{Y}_n^{AV}\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i+\widetilde{Y}_i}{2}\right] = \frac{1}{2}\mathbb{E}\left[\sum_{i=1}^{n}\frac{Y_i+\widetilde{Y}_i}{n}\right] = \frac{1}{2}\left(\mathbb{E}\left[Y\right]+\mathbb{E}\left[\widetilde{Y}\right]\right) = \mathbb{E}\left[Y\right]$$

Hence denoting with

$$\sigma_{AV}^2 = Var\left(\frac{Y_i+\widetilde{Y}_i}{2}\right),$$

the central limit theorem ensures that

$$\frac{\widehat{Y}_n^{AV} - \mathbb{E}[Y]}{\frac{\sigma_{AV}}{\sqrt{n}}} \underset{n\to\infty}{\Rightarrow} N(0;1)$$

But when is $\widehat{Y}_n^{AV}$ better than the usual Monte Carlo estimator?
We first compare $\widehat{Y}_n^{AV}$ with $\widehat{Y}_{2n}$: *tough comparison* for our antithetic variate estimator, ***fair*** if the cost of computing n replications of $(f(Z_i), f(-Z_i))$ is *almost the same* of computing $2n$ replications of $f(Z_i)$.
Then

$$Var\left[\widehat{Y}_n^{AV}\right] < Var\left[\widehat{Y}_{2n}\right]$$

if and only if

$$\frac{Var\left(\frac{Y_i+\widetilde{Y}_i}{2}\right)}{n} < \frac{Var(Y_i)}{2n}$$

that is

$$\frac{1}{4}Var\left(Y_i+\widetilde{Y}_i\right)<\frac{Var\left(Y_i\right)}{2}$$

or

$$Var\left(Y_i+\widetilde{Y}_i\right)<2Var\left(Y_i\right)$$

Simple computations show that

$$Var\left(Y_i+\widetilde{Y}_i\right)=Var\left(Y_i\right)+Var\left(\widetilde{Y}_i\right)+2Cov\left(Y_i;\widetilde{Y}_i\right)<2Var\left(Y_i\right)$$

if and only if

$$Cov\left(Y_i;\widetilde{Y}_i\right)<0$$

That is the pairs

$$\left(Y_i,\widetilde{Y}_i\right)=\left(f(Z_i),f(-Z_i)\right)$$

must be *negatively correlated.*
This is true if the simulation map $f:Z_i\longrightarrow Y_i$ and $f:-Z_i\longrightarrow\widetilde{Y}_i$ is monotone

(intuitive, but difficult theorem).
If we compare $\widehat{Y}_n^{AV}$ with $\widehat{Y}_n$, assuming that the cost of computing n replications of $(f(Z_i), f(-Z_i))$ is *almost the same* of computing only n replications of $f(Z_i)$

$$Var\left[\widehat{Y}_n^{AV}\right] < Var\left[\widehat{Y}_n\right]$$

if

$$\frac{Var\left(\frac{Y_i+\widetilde{Y}_i}{2}\right)}{n} < \frac{Var(Y_i)}{n}$$

that is

$$\frac{1}{4}Var\left(Y_i+\widetilde{Y}_i\right) < Var(Y_i)$$

or

$$Var\left(Y_i+\widetilde{Y}_i\right) < 4Var(Y_i)$$

$$Var\left(Y_i+\widetilde{Y}_i\right) = 2Var(Y_i) + 2Cov\left(Y_i;\widetilde{Y}_i\right) < 4Var(Y_i)$$

if and only if

$$Cov\left(Y_i;\widetilde{Y}_i\right) < Var\left(Y_i\right)$$

which is true *if*

$$Cov\left(Y_i;\widetilde{Y}_i\right) < 0 < Var\left(Y_i\right)$$

that is true if the antithetic replications $\left(Y_i;\widetilde{Y}_i\right)$ are *negatively correlated.*

9 Matching underlying assets

Key idea: matching the moments of the underlying reduces the risk of mispricing derivatives.
Two techniques:

1. Moment matching [G.4.5.1] : Transform the simulated paths.
2. Weighted Monte Carlo [G.4.5.2]: weight paths so that their sample moments match the true ones.

9.1 Moment matching

To match the first moment (the mean) of the Geometric Brownian Motion

$$S(T) \sim S(0)\, e^{\left(r-\frac{\sigma^2}{2}\right)T+\sigma\sqrt{T}Z}$$

we can use

- a multiplicative correction

$$\widetilde{S}_i(T) = S_i(T)\,\frac{\mathbb{E}[S(T)]}{\widehat{S}_n(T)} \text{ for } i = 1, ..., n$$

PROS:

– the transformation preserves positivity:

$$\widetilde{S}_i(T) > 0$$

– the sample mean matches the first moment of $S(T)$

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{S}_i(T) = \mathbb{E}[S(T)]$$

CONS:

– the higher moments are difficultly matched

– there is a bias, because $\widetilde{S}_i(T)$ is not distributed as $S_i(T)$

- an additive correction

$$\widetilde{S}_i(T) = S_i(T) + \mathbb{E}[S(T)] - \widehat{S}_n(T) \text{ for } i = 1, ..., n$$

PROS:

– the sample mean matches the first moment of $S(T)$

$$\frac{1}{n}\sum_{i=1}^{n} \widetilde{S}_i(T) = \mathbb{E}[S(T)]$$

– the transformation preserves martingality

CONS:

– the transformation does not preserve positivity

– there is a bias, because $\widetilde{S}_i(T)$ is not distributed as $S_i(T)$

Remark: Some times it is convenient to apply the moment matching approach on the underlying Brownian motion.

10 Weighted Monte Carlo

1. Weigh the paths $S_i(T)$ for $i = 1, ..., n$ with weights ω_i for $i = 1, ..., n$ such that the moments of S are matched.
2. Then use these weights ω_i for $i = 1, ..., n$ to estimate expected payoffs

For instance:
Select ω_i for $i = 1, ..., n$ such that

$$\sum_{i=1}^{n} \omega_i = 1$$

$$\sum_{i=1}^{n} \omega_i S_i(T) = S(0)\, e^{rT} \text{ match the 1}^{st} \text{ moment of } S(T)$$

then use the ω_i you found to estimate a call option

$$\widehat{Y}_n^{WMC} := \sum_{i=1}^{n} \omega_i \left(S_i(T) - K\right)^+ e^{-rT}$$

instead of the standard Monte Carlo estimator

$$\widehat{Y}_n := \frac{1}{n} \sum_{i=1}^{n} \left(S_i\left(T\right) - K\right)^+ e^{-rT}$$

where weights are uniform, i.e.

$$\omega_i = \frac{1}{n}$$

For discretely-monitored path-dependent options it can be effective to select ω_i such that

$$\sum_{i=1}^{n} \omega_i = 1$$

and

$$\sum_{i=1}^{n} \omega_i S_i\left(t_m\right) = S\left(0\right) e^{rt_m}$$ match the 1st moment at different monitoring dates t_m

In general: n parameters to be determined and few constraints.

You can choose the weights to maximize *the (negative entropy) distance* from the uniform distribution

The *maximum (negative) entropy weights* are defined as follows. The negative entropy distance is the function:

$$H(\omega_1, ..., \omega_n) = \sum_{i=1}^{n} \omega_i \ln \omega_i$$

subject to

$$\begin{cases} \sum_{i=1}^{n} \omega_i = 1 \\ \sum_{i=1}^{n} \omega_i X_i = \mu_X \end{cases}$$

for $X, \mu_X \in \Re^d$.

Write the Lagrangian

$$\mathcal{L}(\omega_1, ..., \omega_n, \nu, \lambda_1, ..., \lambda_d) = \sum_{i=1}^{n} \omega_i \ln \omega_i - \nu \left(\sum_{i=1}^{n} \omega_i - 1 \right) - \lambda \cdot \left(\sum_{i=1}^{n} \omega_i X_i - \mu_X \right)$$

The FOC:

$$\frac{\partial}{\partial \omega_i}\mathcal{L}\left(\omega_1, ..., \omega_n, \nu, \lambda_1, ..., \lambda_d\right) = \ln \omega_i + \omega_i \frac{1}{\omega_i} - \nu - \lambda \cdot X_i = 0$$

delivers

$$\ln \omega_i = -1 + \nu + \lambda \cdot X_i$$

that is

$$\omega_i = \exp\left(-1 + \nu + \lambda \cdot X_i\right)$$

Since the ω_j's sum up to 1 we have that

$$1 = \sum_{j=1}^{n} \exp\left(-1 + \nu + \lambda \cdot X_j\right)$$

that is

$$-1 + \nu = \ln \frac{1}{\sum_{j=1}^{n} e^{\lambda \cdot X_j}} = -\ln \sum_{j=1}^{n} e^{\lambda \cdot Xj}$$

Hence

$$\begin{aligned}\omega_i &= e^{-1+\nu+\lambda\cdot X_i} \\ &= \frac{1}{\sum_{j=1}^{n} e^{\lambda\cdot X_j}} e^{\lambda\cdot X_i}\end{aligned}$$

and $\lambda_1, ..., \lambda_d$ can be found numerically by solving

$$\sum_{i=1}^{n} \omega_i X_i = \mu_X$$

that is

$$\sum_{i=1}^{n} \frac{1}{\sum_{j=1}^{n} e^{\lambda\cdot X_j}} e^{\lambda\cdot X_i} X_i = \mu_X$$