Financial Mathematics 32000

Lecture 6

Roger Lee

2023 April 26

Variance reduction: Control variates

Variance reduction: Importance sampling

Variance reduction: Conditional MC

Control variates

A control variate Y^* is a random variable, related to Y, such that $C^* := \mathbb{E}Y^*$ is known (e.g. via explicit formula). Examples:

- Let $\mathrm{d}S_t = \sigma_t S_t \mathrm{d}W_t$, where σ_t is stochastic. If Y is the discounted payoff of a call on S, we can let Y^* be the discounted payoff of a call on S^* where $\mathrm{d}S_t^* = \sigma S_t^* \mathrm{d}W_t$. So C^* is known (B-S formula). Here σ is some constant that approximates the σ_t process.
- If Y is discounted payoff of Asian-style call on GBM:

$$Y = e^{-rT}\left(\frac{1}{T}\int_0^T S_t \mathrm{d}t - K\right)^+$$

Can let Y^* be the discounted *European* call payoff on the GBM. Or let Y^* be the discounted payoff of a *geometric Asian* call: $Y^* = e^{-rT}\left(e^{\frac{1}{T}\int_0^T \log S_t \mathrm{d}t} - K\right)^+$. In either case, C^* has a formula.

Control variates

The control variate technique simulates the **difference** $Y - \beta Y^*$, where β is a constant. The difference has expectation

$$\mathbb{E}(Y - \beta Y^*) = C - \beta C^*$$

and variance $\mathrm{Var}(Y - \beta Y^*)$ which is, we hope, small. Therefore

$$C = \mathbb{E}Y + \beta(C^* - \mathbb{E}Y^*).$$

Replacing expectations by sample averages, we define the control variate Monte Carlo estimate of C to be

$$\hat{C}_M^{\mathrm{cv},\beta} := \bar{Y}_M + \beta(C^* - \bar{Y}_M^*) \quad \text{where} \quad \begin{aligned} \bar{Y}_M &:= \frac{1}{M}(Y_1 + \cdots + Y_M) \\ \bar{Y}_M^* &:= \frac{1}{M}(Y_1^* + \cdots + Y_M^*) \end{aligned}$$

and we use the *same* pseudo-random numbers generate Y_m and Y_m^*.

Control variates: variance analysis

Expectation:

$$\mathbb{E}\hat{C}_M^{\mathrm{cv},\beta} := \mathbb{E}\hat{C}_M + \beta(C^* - C^*) = C$$

so $\hat{C}_M^{\mathrm{cv},\beta}$ is an unbiased estimate of C.

Variance:

$$\begin{aligned}\mathrm{Var}(\hat{C}_M^{\mathrm{cv},\beta}) &= \mathrm{Var}(\bar{Y}_M - \beta\bar{Y}_M^*) = \frac{1}{M}\mathrm{Var}(Y - \beta Y^*) \\ &= \frac{1}{M}\left[\mathrm{Var}(Y) - 2\beta\mathrm{Cov}(Y, Y^*) + \beta^2\mathrm{Var}(Y^*)\right]\end{aligned}$$

Now choose β to minimize this. Obtain

$$\beta_{\mathrm{optimal}} = \mathrm{Cov}(Y, Y^*)/\mathrm{Var}(Y^*)$$

and $\mathrm{Var}(\hat{C}_M^{\mathrm{cv},\beta_{\mathrm{optimal}}}) = \mathrm{Var}(\hat{C}_M) \times (1 - \mathrm{Corr}^2(Y, Y^*)) \leq \mathrm{Var}(\hat{C}_M)$

Control variates: optimizing the coefficient

Unfortunately the Cov and Var are usually not both known.

But we can use the *sample* covariance and variance. Let

$$\hat{\beta} := \frac{\sum_m (Y_m - \bar{Y}_M)(Y_m^* - \bar{Y}_M^*)}{\sum_m (Y_m^* - \bar{Y}_M^*)^2}$$

And define

$$\hat{C}_M^{\mathrm{cv},\hat{\beta}} := \bar{Y}_M + \hat{\beta}(C^* - \bar{Y}_M^*).$$

Note $\mathbb{E}(C^* - \bar{Y}_M^*) = 0$ does not imply $\mathbb{E}[\hat{\beta}(C^* - \bar{Y}_M^*)] = 0$, so this creates bias in the estimate:

$$\mathbb{E}\hat{C}_M^{\mathrm{cv},\hat{\beta}} = C + \mathbb{E}[\hat{\beta}(C^* - \bar{Y}_M^*)].$$

As $M \to \infty$ the bias $\to 0$.

Control variates: confidence intervals

To compute confidence intervals for fixed β, note that

$$\hat{C}_M^{\mathrm{cv},\beta} = \frac{1}{M}\sum_m Y_m^{\mathrm{cv},\beta}$$

where the independent simulations

$$Y_m^{\mathrm{cv},\beta} := Y_m + \beta(C^* - Y_m^*)$$

have an estimated variance

$$\frac{1}{M-1}\sum_m (Y_m^{\mathrm{cv},\beta} - \hat{C}_M^{\mathrm{cv},\beta})^2.$$

Let $\hat{\sigma}_M^{\mathrm{cv},\beta}$ be the square root of this estimated variance.

Control variates: confidence intervals

Note that

$\hat{\sigma}_M^{\mathrm{cv},\beta}$ = estimated standard deviation of *one* of the $Y_m^{\mathrm{cv},\beta}$

$\hat{\sigma}_M^{\mathrm{cv},\beta}/\sqrt{M}$ = estimated st. dev. of $\hat{C}_M^{\mathrm{cv},\beta}$, the *average* of the $Y_m^{\mathrm{cv},\beta}$

Then an asymptotic $100(1-p)\%$ confidence interval has endpoints

$$\hat{C}_M^{\mathrm{cv},\beta} \pm \Phi^{-1}\left(1-\frac{p}{2}\right)\frac{\hat{\sigma}_M^{\mathrm{cv},\beta}}{\sqrt{M}}.$$

If instead of β we use $\hat{\beta}$ estimated from those M simulations, the variance estimate is biased for finite M, but as $M \to \infty$ its bias and variance $\to 0$, so the same confidence interval calculation (with $\hat{\beta}$ in place of β) is asymptotically valid.

Variance reduction: Control variates

Variance reduction: Importance sampling

Variance reduction: Conditional MC

Motivating example

Using Monte Carlo, a rare event with a big payoff is harder to price than a more frequent event with a smaller payoff.

- Consider an asset that pays 100 dollars with probability 2% (under some pricing measure), and pays zero otherwise.
  Its price is 2 and the payoff variance is $10000 \times 0.02 - 2^2 = 196$
- Now consider an asset that pays 10 dollars with probability 20%, zero otherwise.
  Its price is 2. But its payoff variance is only $100 \times 0.20 - 2^2 = 16$

So maybe if we *sample from a different distribution* and *modify the payoff* we can preserve the expectation, while decreasing the variance.

Importance sampling

Let X be a random vector in $\mathbb{R}^N$, with density function $f : \mathbb{R}^N \to \mathbb{R}$.

Let $h : \mathbb{R}^N \to \mathbb{R}$. Think of h as a discounted payoff function.

We want to estimate

$$C := \mathbb{E}h(X) = \int h(x)f(x)\mathrm{d}x$$

Ordinary Monte Carlo:

$$\hat{C} := \frac{1}{M}\sum_{m=1}^{M} h(X_{[m]})$$

where $X_{[1]}, \ldots, X_{[M]}$ are IID random draws from density f.

Can apply to $X =$ some financial variable (such as an asset price), or can apply to $X =$ as the randomness that drives the dynamics (such as a vector of Brownian increments).

Change of measure

Let g be any density on $\mathbb{R}^N$ that vanishes where fh does, meaning $f(x)h(x) = 0$ if $g(x) = 0$. Let $A := \{x : f(x)h(x) \neq 0\}$. Then

$$C = \int_A h(x)f(x)\mathrm{d}x = \int_A h(x)\frac{f(x)}{g(x)}g(x)\mathrm{d}x = \mathbb{E}^*\left[h(X)\frac{f(X)}{g(X)}\right]$$

where $\mathbb{E}^*$ denotes expectation with respect to a new measure $\mathbb{P}^*$ under which X has density g.

The *importance sampling* Monte Carlo estimate of C is

$$\hat{C}^{\text{is}} := \frac{1}{M}\sum_{m=1}^{M} h(X_{[m]})\frac{f(X_{[m]})}{g(X_{[m]})}$$

where $X_{[1]}, \dots, X_{[M]}$ are IID random draws from density g, which we call the *importance sampling* density.

Change of measure

So we have made two changes to the ordinary MC estimate

- Change the distribution from which X is sampled. Draw from density g instead of f.
- Change the payoff. Instead of $h(X)$, it becomes $h(X)$ times the *Radon-Nikodym derivative* or *likelihood ratio* $f(X)/g(X)$.

What about bias and variance?

- By construction, $\hat{C}^{\text{is}}$ is an unbiased estimator of C because

$$\mathbb{E}^* \hat{C}^{\text{is}} = \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}^* \left[h(X_{[m]}) \frac{f(X_{[m]})}{g(X_{[m]})} \right] = C$$

- To minimize variance, try to choose g such that hf/g is constant on the set where $g \neq 0$. Try to "flatten the curve" (the payout h).

Variance

The importance-sampling MC estimate $\hat{C}^{\text{is}}$ has (wrt $\mathbb{P}^*$) variance

$$\frac{1}{M}\mathrm{Var}^*\frac{h(X)f(X)}{g(X)} = \mathbb{E}^*\frac{h(X)^2 f(X)^2}{g(X)^2} - C^2 = \mathbb{E}\left[h(X)^2\frac{f(X)}{g(X)}\right] - C^2$$

which can be bigger or smaller than the ordinary MC variance:

$$\frac{1}{M}\mathrm{Var}\ h(X) = \mathbb{E}\big[h(X)^2\big] - C^2.$$

How to choose g to minimize $\mathrm{Var}^*\hat{C}^{\text{is}}$? Answer: if $h \geq 0$ and

$$g(x) = \text{constant} \times h(x)f(x)$$

then

$$\mathrm{Var}^*\hat{C}^{\text{is}} = \frac{1}{M}\mathrm{Var}^*\frac{h(X)f(X)}{g(X)} = 0.$$

Variance reduction

- However, note that the unique constant such that

$$g(x) = \text{constant} \times h(x)f(x)$$

  defines a legitimate density is $1/\int h(x)f(x)dx = 1/C$.

  If we knew the value of C, we wouldn't need Monte Carlo. So in practice, this recipe for choosing g cannot be followed exactly.

- What we *can* do is choose g to be a known density *approximately* proportional to the original density f times the payoff h.

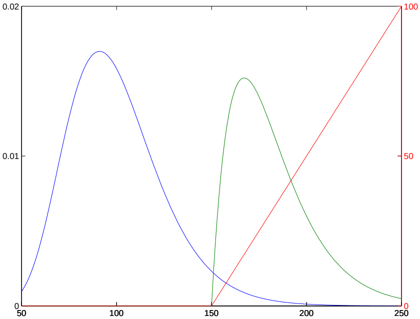
  Irrespective of whether this approximation is bad or good, we still have an unbiased estimate $\hat{C}^{\text{is}}$.

  A good approximation can make the variance of $\hat{C}^{\text{is}}$ small.

Variance reduction

If $X = (X_1, \ldots, X_N)$ are Brownian increments driving stock $S(X)$, where $S : \mathbb{R}^N \to \mathbb{R}$, then can apply this idea to $S(X)$ instead of X.

In other words, if $h(x) = H(S(x))$, then variance is zero if $S(X)$ has IS-density = constant $\times$ H $\times$ original density of $S(X)$.



Likelihood ratio computations

To compute $f(X)/g(X)$, usually you won't be specifying g. Rather you'd specify, for instance, a desired mean or drift, then compute g. Compute g by starting in X space.

Suppose, in particular, that the components of $X \in \mathbb{R}^N$ under the original measure are independent with densities $f_1, f_2, \ldots, f_N$, and that under importance sampling measure (not necessarily optimal), they are independent with densities $g_1, g_2, \ldots, g_N$. Then

$$\frac{f(x)}{g(x)} = \frac{f(x_1, \ldots, x_N)}{g(x_1, \ldots, x_N)} = \prod_{n=1}^{N} \frac{f_n(x_n)}{g_n(x_n)}$$

(Actually, independence is unnecessary, if you let f_n be the *conditional* density of X_n given $X_1, \ldots, X_{n-1}$, and similarly for g_n.)

Example: Changing of drift of BM

Let $X_1, \ldots, X_N$ be successive increments of BM W sampled at times $0, t_1, \ldots, t_N = T$. Then density f_n is Normal$(0, \Delta t)$, where $\Delta t = T/N$.

$$f_n(x_n) = \frac{1}{\sqrt{2\pi\Delta t}} e^{-x_n^2/(2\Delta t)}$$

The g_n can be the density of any other distribution supported on $\mathbb{R}$.

In particular, if we take g_n to be the Normal$(\lambda\Delta t, \Delta t)$ density, then

$$g_n(x_n) = \frac{1}{\sqrt{2\pi\Delta t}} e^{-(x_n - \lambda\Delta t)^2/(2\Delta t)}$$

So for all $x = (x_1, \ldots, x_n)$,

$$\frac{f_n(x_n)}{g_n(x_n)} = e^{\frac{1}{2\Delta t}(-x_n^2 + x_n^2 - 2x_n\lambda\Delta t + \lambda^2(\Delta t)^2)} = e^{-\lambda x_n + \frac{1}{2}\lambda^2\Delta t}$$

$$\frac{f(x)}{g(x)} = \prod_{n=1}^{N} \frac{f_n(x_n)}{g_n(x_n)} \Longrightarrow \frac{f(X)}{g(X)} = e^{-\lambda \sum X_n + \frac{1}{2}\lambda^2 T} = e^{-\lambda W_T + \frac{1}{2}\lambda^2 T}$$

Example: Changing the drift of BM

Implementation

- If you want to change the drift of W to λ, then work under $\mathbb{P}^*$ such that $W_t^* := W_t - \lambda t$ is $\mathbb{P}^*$-Brownian Motion. Then

$$f(X)/g(X) = e^{-\lambda \sum X_n + \frac{1}{2}\lambda^2 T} = e^{-\lambda W_T + \frac{1}{2}\lambda^2 T} = e^{-\lambda W_T^* - \frac{1}{2}\lambda^2 T}$$

  Multiply the payoff $h(X)$ by this.
- When simulating under the measure $\mathbb{P}^*$, the zero-mean normals that you generate are simulations of W^*, not W.
- The $f(X)/g(X)$ is also known as the *Radon-Nikodym derivative* of $\mathbb{P}$ with respect to $\mathbb{P}^*$.

Example: Changing the drift of BM

Choice of λ

- How to choose λ? Maybe too difficult to make S_T's distribution agree entirely with the optimal distribution. So let's just make S_T's *mean* $\approx$ the optimal distribution's *mean*. If S is GBM,

$$dS_t = rS_t dt + \sigma S_t dW_t = (r + \sigma\lambda)S_t dt + \sigma S_t dW_t^*$$

  So $\mathbb{E}^* S_T = S_0 e^{(r+\sigma\lambda)T}$. You could, for instance, try to choose λ such that $S_0 e^{(r+\sigma\lambda)T} \approx$ the optimal distribution's mean.

- Even if λ is not chosen optimally, the importance sampling MC estimate is still unbiased.

Variance reduction: Control variates

Variance reduction: Importance sampling

Variance reduction: Conditional MC

Conditional Monte Carlo

For any random variable X, by the law of iterated expectations,

$$C = \mathbb{E}Y = \mathbb{E}[\mathbb{E}(Y|X)] = \mathbb{E}f(X)$$

Assuming that the function $f(X) := \mathbb{E}(Y|X)$ can be computed analytically, Conditional Monte Carlo does the following:

- Simulate not Y but rather X, generating $X_1, \ldots, X_M$ IID as X.
- Estimate $\mathbb{E}[\mathbb{E}(Y|X)]$ by taking the average

$$\hat{C}_M^{\text{cmc}} := \frac{1}{M} \sum_m f(X_m).$$

The standard error is $\hat{\sigma}_M^{\text{cmc}}/\sqrt{M}$ where

$$\hat{\sigma}_M^{\text{cmc}} := \sqrt{\frac{1}{M-1} \sum_m (f(X_m) - \hat{C}_M^{\text{cmc}})^2}$$

Conditional Monte Carlo: variance analysis

Expectation: The CMC estimate is unbiased, because

$$\mathbb{E}\hat{C}_M^{\text{cmc}} = \mathbb{E}f(X) = C.$$

Variance: CMC has smaller variance than ordinary MC because

$$\text{Var}(\hat{C}_M^{\text{cmc}}) = \frac{1}{M}\text{Var}f(X) = \frac{1}{M}\text{Var}[\mathbb{E}(Y|X)] \leq \frac{1}{M}\text{Var}(Y) = \text{Var}(\hat{C})$$

where the last step is because $\text{Var}(Y) = \text{Var}[\mathbb{E}(Y|X)] + \mathbb{E}\text{Var}(Y|X)$.

Intuition: suppose $Y = F(\mathbf{X})$ where $\mathbf{X} \in \mathbb{R}^D$. Then ordinary MC computes a D-dimensional integral $\int F(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x}$ by drawing random $\mathbf{X}_m \in \mathbb{R}^D$ with density p. But if, along d of those dimensions, the integration can be performed *analytically*, then we only need to draw random vectors $X_m \in \mathbb{R}^{D-d}$. This reduces the noise in the estimate.

Conditional Monte Carlo: example 1

Let $0 < t_1 < t_2 < t_3$. Consider at time 0 an Asian-style call on a GBM S, with expiry t_3 and discounted payoff

$$Y := e^{-rt_3}\left(\frac{S_{t_1} + S_{t_2} + S_{t_3}}{3} - k\right)^+$$

Let $X = (S_{t_1}, S_{t_2})$. Then $\mathbb{E}(Y|X)$ is analytically known, because

$$Y = \frac{1}{3}e^{-rt_3}\Big(S_{t_3} - (3k - S_{t_1} - S_{t_2})\Big)^+$$

So $\mathbb{E}(Y|X)$ can be analytically evaluated using the B-S formula.

CMC: Generate M realizations of (S_{t_1}, S_{t_2}) and take the average of

$$\mathbb{E}(Y|X) = \frac{1}{3}e^{-rt_2}C^{BS}(S = S_{t_2},\ t = t_2,\ K = 3k - (S_{t_1} + S_{t_2}),\ T = t_3).$$

This random variable has *smaller variance* than Y.

Conditional Monte Carlo: example 2a

Consider a call with discounted payoff $Y = e^{-rT}(S_T - K)^+$ where

$$\mathrm{d}S_t = rS_t\mathrm{d}t + \sqrt{V_t}S_t\mathrm{d}W_{1t}$$

$$\mathrm{d}V_t = \alpha(V_t)\mathrm{d}t + \beta(V_t)\mathrm{d}W_{2t}$$

and W_1 and W_2 are (in this particular example) *independent* BM.

Note that *conditional on the entire path of* V_t, the dynamics of S are $\mathrm{d}S_t = rS_t\mathrm{d}t + \sqrt{V_t}S_t\mathrm{d}W_{1t}$ where the V_t path is conditionally *fixed*, and W_1 is still conditionally a BM. So Y has conditional expectation

$$\mathbb{E}(Y|X) = C^{BS}(\bar{\sigma}) \qquad \text{where } \bar{\sigma}^2 := \frac{1}{T}\int_0^T V_t\mathrm{d}t$$

and so Y has unconditional expectation given by the "mixing formula"

$$C = \mathbb{E}C^{BS}(\bar{\sigma}).$$

Conditional Monte Carlo: example 2a

Two approaches:

- Ordinary MC:

  Generate M realizations of the two-dimensional path of (S, V).

  On each path calculate the discounted call payoff. Average.

- Conditional MC:

  Generate M realizations of the one-dimensional path of V.

  On each path calculate $C^{BS}(\bar{\sigma})$. Average.

CMC reduces variance by removing from the simulations the noise due to the randomness of S.

Conditional Monte Carlo: example 2b

Stochastic volatility model with nonzero correlation: Let

$$W_t = \sqrt{1-\rho^2}\mathrm{d}W_{1t} + \rho\mathrm{d}W_{2t}$$

where W_1 and W_2 are independent BMs, so W and W_2 have correlation ρ. Let

$$\begin{aligned}\mathrm{d}S_t &= rS_t\mathrm{d}t + \sigma_t S_t \mathrm{d}W_t\\ \sigma_t &= \sqrt{V_t}\\ \mathrm{d}V_t &= \alpha(V_t)\mathrm{d}t + \beta(V_t)\mathrm{d}W_{2t},\end{aligned}$$

Conditional Monte Carlo: example 2b

Then $L := \log S$ satisfies

$$\begin{aligned} \mathrm{d}L_t &= r\mathrm{d}t - \frac{1}{2}\sigma_t^2\mathrm{d}t + \sigma_t\sqrt{1-\rho^2}\mathrm{d}W_{1t} + \sigma_t\rho\mathrm{d}W_{2t} \\ &= r\mathrm{d}t - \frac{1-\rho^2}{2}\sigma_t^2\mathrm{d}t + \sigma_t\sqrt{1-\rho^2}\mathrm{d}W_{1t} - \frac{\rho^2}{2}\sigma_t^2\mathrm{d}t + \sigma_t\rho\mathrm{d}W_{2t} \end{aligned}$$

So conditional on $\mathcal{F}_T^{W_2}$,

$$L_T \sim \mathrm{Normal}\left(L_0 + rT + \log M_T(\rho) - \frac{1-\rho^2}{2}\bar{\sigma}^2T,\ (1-\rho^2)\bar{\sigma}^2T\right)$$

where

$$M_T(\rho) := \exp\left(-\frac{\rho^2}{2}\int_0^T \sigma_t^2\mathrm{d}t + \rho\int_0^T \sigma_t\mathrm{d}W_{2t}\right)$$

So even with correlation, S is still W_2-conditionally lognormal.

Conditional Monte Carlo: example 2b

"Mixing formula" for price of option paying $C_T = C(S_T)$

$$C_0 = \mathbb{E}(e^{-rT}C_T) = \mathbb{E}(\mathbb{E}(e^{-rT}C_T|\mathcal{F}_T^{W_2})) = \mathbb{E}C^{BS}(S_0 M_T(\rho),\ \bar{\sigma}\sqrt{1-\rho^2})$$

where

$$M_T(\rho) := \exp\left(-\frac{\rho^2}{2}\int_0^T \sigma_t^2 \mathrm{d}t + \rho\int_0^T \sigma_t \mathrm{d}W_{2t}\right)$$

Option price is expectation of the B-S formula for that option,

evaluated at random volatility *and* a randomized spot.

Therefore:

Simulate W_2. Given W_2, evaluate $C^{BS}(S_0 M_T, \bar{\sigma}\sqrt{1-\rho^2})$.

Average across all simulations of W_2 . No need to simulate W_1.