Control Variates for Monte Carlo Integration

Independent Study Exercises

Monte Carlo Rendering & Inverse Rendering

Contents

1 Introduction to Control Variates

Background

In Monte Carlo rendering, we frequently encounter integrals of the form

$$I = \int_\Omega f(x)\, d\mu(x) = \mathbb{E}[f(X)]$$

where X is a random variable with distribution μ over domain Ω. The standard Monte Carlo estimator uses n independent samples $X_1, \ldots, X_n$ drawn from μ:

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^{n} f(X_i).$$

This estimator is unbiased with variance $\mathrm{Var}(\hat{I}_n) = \sigma_f^2/n$, where $\sigma_f^2 = \mathrm{Var}(f(X))$.
The **control variate** method is a variance reduction technique that exploits correlation with a function whose integral is known analytically.

Definition 1 (Control Variate)**.** Let $g : \Omega \to \mathbb{R}$ be a function with **known** expected value $\mu_g = \mathbb{E}[g(X)]$. We call g a *control variate* for f if g is correlated with f. The *control variate estimator* is defined as:

$$\hat{I}_n^{CV} = \frac{1}{n} \sum_{i=1}^{n} \left[f(X_i) - c\,(g(X_i) - \mu_g)\right]$$

where $c \in \mathbb{R}$ is a constant coefficient. Equivalently, using sample means:

$$\hat{I}_n^{CV} = \bar{f}_n - c(\bar{g}_n - \mu_g)$$

where $\bar{f}_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$ and $\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$.

Remark 1 (Intuition)**.** The key insight is that since we know $\mathbb{E}[g(X)] = \mu_g$, the quantity $\bar{g}_n - \mu_g$ represents the "error" in our Monte Carlo estimate of μ_g. If f and g are correlated, this error is informative about the error in $\bar{f}_n$. By subtracting a scaled version of this known error, we can reduce the overall variance.

2 Basic Properties of Control Variates

Exercise 1 (Unbiasedness of the Control Variate Estimator)**.** Prove that the control variate estimator $\hat{I}_n^{CV}$ is an unbiased estimator of $I = \mathbb{E}[f(X)]$ for any choice of the coefficient $c \in \mathbb{R}$.

Exercise 2 (Variance of the Control Variate Estimator)**.** Derive the variance of the control variate estimator $\hat{I}_n^{CV}$. Express your answer in terms of:

- $\sigma_f^2 = \mathrm{Var}(f(X))$
- $\sigma_g^2 = \mathrm{Var}(g(X))$
- $\sigma_{fg} = \mathrm{Cov}(f(X), g(X))$

Exercise 3 (Optimal Coefficient)**.** Find the value of c that minimizes the variance of the control variate estimator. Then show that the minimum variance is:

$$\mathrm{Var}^*\left(\hat{I}_n^{CV}\right) = \frac{\sigma_f^2}{n}\left(1-\rho^2\right)$$

where $\rho = \mathrm{Corr}(f(X), g(X)) = \sigma_{fg}/(\sigma_f \sigma_g)$ is the Pearson correlation coefficient.

3 Degenerate Cases and Practical Considerations

Background

In the previous section, we derived the optimal coefficient $c^* = \sigma_{fg}/\sigma_g^2$. However, **in practice, this optimal value cannot be computed exactly** because it depends on:

- $\sigma_{fg} = \mathrm{Cov}(f(X), g(X))$ — the covariance between f and g
- $\sigma_g^2 = \mathrm{Var}(g(X))$ — the variance of the control variate

These population quantities are *unknown* and must be estimated from samples. Common approaches include:

1. Using a small pilot sample to estimate c^*, then applying control variates with this estimate
2. Jointly estimating c^* from the same samples used for the final estimate
3. Using domain knowledge or prior experiments to choose a reasonable c

Each approach introduces additional uncertainty, and the coefficient actually used may differ significantly from the true optimal. The following exercises explore the consequences of this, examining when control variates help, when they don't, and when they can actually *hurt* performance.

Exercise 4 (Zero Variance)**.** Under what conditions is the variance of the optimal control variate estimator exactly zero? Is this achievable in practice?

Exercise 5 (Variance Inflation)**.** Since the optimal coefficient $c^* = \sigma_{fg}/\sigma_g^2$ cannot be computed exactly in practice, we must use some estimate or heuristic value c.

(a) For what values of c does the control variate estimator have *higher* variance than the standard Monte Carlo estimator (i.e., when does it hurt rather than help)?

(b) Provide a concrete numerical example demonstrating variance inflation.

4 Random Control Variates

Background

In many rendering applications, we have access to a function g that is:

- Strongly correlated with our target function f
- **Cheaper** to evaluate than f
- Has **unknown** integral $\mu_g = \mathbb{E}[g(X)]$

Examples in rendering include:

- Using a neural network approximation of the rendering equation as a control variate for path tracing
- Using single-bounce illumination as a control variate for multi-bounce global illumination

Since μ_g is unknown, we must *estimate* it with Monte Carlo. This leads to the concept of **random control variates** (also called *estimated control variates*).

Definition 2 (Random Control Variate Estimator)**.** Let f be the target function and g be a cheaper function (control variate) with unknown mean μ_g. We draw:

- n **paired samples** $(f(X_i), g(X_i))$ for $i = 1, \dots, n$
- m **additional samples** $g(Y_j)$ for $j = 1, \dots, m$ (independent of the X_i)

Define the sample means:

$$\bar{f}_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i), \quad \bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i), \quad \tilde{g}_m = \frac{1}{m}\sum_{j=1}^{m} g(Y_j).$$

The **random control variate estimator** is:

$$\hat{I}^{RCV}_{n,m} = \bar{f}_n - c\,(\bar{g}_n - \tilde{g}_m)$$

where $\tilde{g}_m$ serves as our estimate of μ_g.

Remark 2. The key idea is that since g is cheaper to evaluate than f, we can afford many more samples of g (i.e., $m \gg n$). This makes $\tilde{g}_m$ a good estimate of μ_g, allowing us to still benefit from the variance reduction of control variates despite not knowing μ_g analytically.

5 Analysis of Random Control Variates

Exercise 6 (Unbiasedness of Random Control Variates)**.** Prove that the random control variate estimator $\hat{I}^{RCV}_{n,m}$ is an unbiased estimator of $I = \mathbb{E}[f(X)]$.

Exercise 7 (Variance of Random Control Variates)**.** Derive the variance of the random control variate estimator $\hat{I}^{RCV}_{n,m}$.

Exercise 8 (Optimal Coefficient for Random Control Variates)**.** Find the optimal coefficient c^* that minimizes the variance of $\hat{I}_{n,m}^{RCV}$.

Exercise 9 (Optimal Variance for Random Control Variates)**.** Substitute c^* from Exercise 8 into the variance expression to find the minimum variance.

6 Optimal Sample Allocation

In practice, we have a fixed computational budget. If f costs C_f units to evaluate and g costs C_g units, our total budget B constrains our choice of n and m:

$$nC_f + nC_g + mC_g = B$$

The term nC_g accounts for evaluating g on the paired samples, and mC_g for the additional control variate samples.

Defining the cost ratio $r = C_f/C_g > 1$, we can write the budget constraint as:

$$n(r+1) + m = T$$

where $T = B/C_g$ is the total budget in units of g-evaluations.

Derivation of Optimal Allocation

From the constraint, $m = T - n(r+1)$, which gives $m + n = T - nr$. Substituting into the optimal variance formula from Exercise 9:

$$V^*(n) = \frac{\sigma_f^2}{n}\left(1 - \frac{T - n(r+1)}{T - nr}\rho^2\right)$$

Simplifying the fraction $\frac{T-n(r+1)}{T-nr} = \frac{(T-nr)-n}{T-nr} = 1 - \frac{n}{T-nr}$, we can rewrite:

$$V^*(n) = \frac{\sigma_f^2}{n}\left(1 - \rho^2 + \frac{n\rho^2}{T-nr}\right) = \frac{\sigma_f^2(1-\rho^2)}{n} + \frac{\sigma_f^2\rho^2}{T-nr}$$

To minimize, we differentiate with respect to n and set to zero:

$$\frac{dV^*}{dn} = -\frac{\sigma_f^2(1-\rho^2)}{n^2} + \frac{\sigma_f^2\rho^2 r}{(T-nr)^2} = 0$$

Rearranging: $(1-\rho^2)(T-nr)^2 = \rho^2 r n^2$. Taking square roots (both sides positive):

$$\sqrt{1-\rho^2}(T - nr) = \rho\sqrt{r} \cdot n$$

Expanding and solving for n:

$$T\sqrt{1-\rho^2} = n\left(r\sqrt{1-\rho^2} + \rho\sqrt{r}\right)$$

$$\boxed{n^* = \frac{T\sqrt{1-\rho^2}}{r\sqrt{1-\rho^2} + \rho\sqrt{r}}}$$

Optimal Sample Ratio

From the budget constraint, $m^* = T - n^*(r+1)$, so:

$$\frac{m^*}{n^*} = \frac{T}{n^*} - (r+1) = \frac{r\sqrt{1-\rho^2} + \rho\sqrt{r}}{\sqrt{1-\rho^2}} - (r+1)$$

$$= r + \frac{\rho\sqrt{r}}{\sqrt{1-\rho^2}} - r - 1 = \frac{\rho\sqrt{r}}{\sqrt{1-\rho^2}} - 1$$

For large r, the -1 term becomes negligible:

$$\frac{m^*}{n^*} \approx \frac{\rho}{\sqrt{1-\rho^2}} \cdot \sqrt{r}$$

The factor $\frac{\rho}{\sqrt{1-\rho^2}} = 1$ when $\rho = 1/\sqrt{2} \approx 0.707$. Thus, for moderate correlation ($\rho \approx 0.7$) and large r:

$$\boxed{\frac{m^*}{n^*} \approx \sqrt{r}}$$

Practical Guidance

- **Heuristic:** Set $m/n \approx \sqrt{r} = \sqrt{C_f/C_g}$. This is a good approximation for moderate correlation.
- **When $r \gg 1$:** Use $m \gg n$. Cheap g evaluations allow accurate estimation of μ_g.
- **High ρ:** The ratio m^*/n^* increases; more control variate samples are beneficial.
- **Low ρ:** Control variates may not be worthwhile—the cost of g may exceed the benefit.

7 Additional Exercises

Exercise 10 (Adaptive Coefficient Estimation)**.** In practice, we must estimate the optimal coefficient $c^* = \sigma_{fg}/\sigma_g^2$ from samples. Two natural approaches are:

Method 1 (Single-sample): Use all n samples both to estimate $\hat{c}$ and to compute the final control variate estimator.

Method 2 (Split-sample): Split the n samples into two disjoint sets: use n_1 samples to estimate $\hat{c}$, and the remaining $n_2 = n - n_1$ samples to compute the final estimate with this fixed $\hat{c}$.

(a) For each method, write down the estimator and determine whether it is biased or unbiased.

(b) Discuss the bias-variance tradeoffs between the two methods.

Exercise 11 (Application to Path Tracing)**.** Consider estimating the radiance at a point using path tracing. Let $f(X)$ be the contribution of a random path X, and let $g(X)$ be a neural network approximation of the radiance that can be evaluated without tracing additional rays.

(a) Why might g not have a known integral μ_g?

(b) Propose a strategy using random control variates for this scenario. Use the results from Section 6 to determine an appropriate budget allocation between paired samples and control variate samples.

(c) What properties should the neural network have to be an effective control variate?

8 Summary

Estimator	**Optimal Coefficient**	**Optimal Variance**
Standard MC	—	$\dfrac{\sigma_f^2}{n}$
Deterministic CV	$c^* = \dfrac{\sigma_{fg}}{\sigma_g^2}$	$\dfrac{\sigma_f^2}{n}(1-\rho^2)$
Random CV	$c^* = \dfrac{\sigma_{fg}}{\sigma_g^2} \cdot \dfrac{m}{m+n}$	$\dfrac{\sigma_f^2}{n}\left(1 - \dfrac{m}{m+n}\rho^2\right)$

Key Takeaways:

- Control variates reduce variance by exploiting correlation with functions of known (or estimable) integrals.
- The variance reduction factor is $(1-\rho^2)$ for perfect control variates, so high correlation is essential.
- Random control variates trade some variance reduction for practicality when exact integrals are unknown.
- Careful allocation of samples between target and control variate is important for efficiency.
- Poorly chosen coefficients can *increase* variance—estimation and validation are crucial.