

The Control Variates Method

Reducing Monte Carlo variance with a known expectation

Junqi Tang

Computational Statistics
University of Birmingham

The expectation we want to estimate

For $X \sim p$, our quantity of interest is

$$\mathcal{I} = \mathbb{E}_p[f(X)] = \int f(x)p(x) dx.$$

With $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} p$,

$$\hat{\mathcal{I}}_{\text{MC}} = \frac{1}{n} \sum_{i=1}^n f(X_i), \quad \text{Var}(\hat{\mathcal{I}}_{\text{MC}}) = \frac{\text{Var}_p(f(X))}{n}.$$

- Assume $\mathbb{E}_p[f(X)^2] < \infty$ for the variance calculation.
- For fixed f and p , increasing n reduces the variance.

Can we reduce the variance at the same sample size?

Importance sampling changes where we sample

Generate $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} q$ and use $w = p/q$:

$$\hat{\mathcal{I}}_{\text{IS}} = \frac{1}{n} \sum_{i=1}^n f(Y_i) w(Y_i).$$

Under the appropriate support and moment conditions,

$$\mathbb{E}[\hat{\mathcal{I}}_{\text{IS}}] = \mathcal{I}, \quad \text{Var}(\hat{\mathcal{I}}_{\text{IS}}) = \frac{\text{Var}_q(f(Y)w(Y))}{n}.$$

- The choice of q changes the variance.
- The weights preserve the target expectation.
- A suitable proposal can improve accuracy.

Control variates change the function we average

Keep the sampling distribution p .

Construct a new function f_{CV} such that

$$\mathbb{E}_p[f_{\text{CV}}(X)] = \mathbb{E}_p[f(X)] = \mathcal{I}.$$

Our goal is

$$\text{Var}_p(f_{\text{CV}}(X)) < \text{Var}_p(f(X)).$$

Then apply ordinary Monte Carlo to f_{CV} .

Preserve the mean while reducing the fluctuations.

A control function has a known expectation

Choose a function g for which we know

$$g_0 = \mathbb{E}_p[g(X)].$$

- The expectation is under the same distribution p .
- For example, a linear g is usable when the required mean of X is known.
- To analyse variance, assume $f(X)$ and $g(X)$ have finite second moments.

A known expectation is essential. A useful correlation comes next.

Adding a centred control

For a fixed coefficient $c \in \mathbb{R}$, define

$$f_{\text{CV}}(x) = f(x) + c(g(x) - g_0).$$

The control is centred:

$$\mathbb{E}_p[g(X) - g_0] = g_0 - g_0 = 0.$$

- c may be positive, negative, or zero.
- $c = 0$ recovers the original function f .

We use the lecture's plus-sign convention throughout.

The target expectation is unchanged

By linearity of expectation,

$$\begin{aligned}\mathbb{E}_p[f_{\text{CV}}(X)] &= \mathbb{E}_p[f(X)] + c\mathbb{E}_p[g(X) - g_0] \\ &= \mathbb{E}_p[f(X)] + c(\mathbb{E}_p[g(X)] - g_0) \\ &= \mathbb{E}_p[f(X)] \\ &= \boxed{\mathcal{I}}.\end{aligned}$$

Every fixed c preserves the mean when g_0 is exact.

The control variates estimator

Use $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} p$:

$$\hat{\mathcal{I}}_{\text{CV}} = \frac{1}{n} \sum_{i=1}^n [f(X_i) + c(g(X_i) - g_0)].$$

Equivalently, with $\bar{f} = n^{-1} \sum_i f(X_i)$ and $\bar{g} = n^{-1} \sum_i g(X_i)$,

$$\boxed{\hat{\mathcal{I}}_{\text{CV}} = \bar{f} + c(\bar{g} - g_0).}$$

Evaluate f and g at the **same sample points**.

The estimator is unbiased for fixed coefficients

For fixed c and known g_0 ,

$$\mathbb{E}[\hat{\mathcal{I}}_{\text{CV}}] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_p[f_{\text{CV}}(X_i)] = \mathcal{I}.$$

Hence its mean squared error is

$$\text{MSE}(\hat{\mathcal{I}}_{\text{CV}}) = \mathbb{E}[(\hat{\mathcal{I}}_{\text{CV}} - \mathcal{I})^2] = \text{Var}(\hat{\mathcal{I}}_{\text{CV}}).$$

At this stage, accuracy is determined by the variance.

Two choices determine the variance

Independence across samples gives

$$\text{Var}(\hat{\mathcal{I}}_{\text{CV}}) = \frac{\text{Var}_p(f_{\text{CV}}(X))}{n}.$$

The numerator depends on

- the control function g ,
- the coefficient c .

A poor choice can increase variance. We first fix g and optimise c .

A constant shift does not affect variance

For this calculation, write $f = f(X)$ and $g = g(X)$. All moments are under p .

$$\begin{aligned}\mathrm{Var}(f_{\mathrm{CV}}) &= \mathrm{Var}(f + c(g - g_0)) \\ &= \mathrm{Var}(f + cg - cg_0) \\ &= \mathrm{Var}(f + cg).\end{aligned}$$

This uses $\mathrm{Var}(Z + a) = \mathrm{Var}(Z)$ for any constant a .

The covariance term makes reduction possible

For random variables with finite second moments,

$$\text{Var}(U + V) = \text{Var}(U) + \text{Var}(V) + 2 \text{Cov}(U, V).$$

Therefore

$$\begin{aligned} \text{Var}(f_{\text{CV}}) &= \text{Var}(f) + \text{Var}(cg) + 2 \text{Cov}(f, cg) \\ &= \boxed{\text{Var}(f) + c^2 \text{Var}(g) + 2c \text{Cov}(f, g)}. \end{aligned}$$

The factor of two is essential. Independence is not assumed.

The variance is a quadratic in the coefficient

With g fixed, define

$$V(c) = \text{Var}(f) + c^2 \text{Var}(g) + 2c \text{Cov}(f, g).$$

- Only c varies in this optimisation.
- $\text{Var}(f)$, $\text{Var}(g)$, and $\text{Cov}(f, g)$ are fixed numbers.
- If $\text{Var}(g) > 0$, the quadratic is strictly convex.

$$c^* = \arg \min_{c \in \mathbb{R}} V(c).$$

The optimal coefficient

Assume $\text{Var}(g) > 0$. Differentiate:

$$V'(c) = 2c \text{Var}(g) + 2 \text{Cov}(f, g).$$

Setting $V'(c) = 0$ gives

$$c^{\star} = -\frac{\text{Cov}(f, g)}{\text{Var}(g)}.$$

Since $V''(c) = 2 \text{Var}(g) > 0$, this is the unique global minimum.

Positive covariance gives a negative optimal coefficient.

Substituting the optimal coefficient

Insert c^* into $V(c)$:

$$\begin{aligned} V(c^*) &= \text{Var}(f) + \left(-\frac{\text{Cov}(f, g)}{\text{Var}(g)}\right)^2 \text{Var}(g) \\ &\quad + 2 \left(-\frac{\text{Cov}(f, g)}{\text{Var}(g)}\right) \text{Cov}(f, g) \\ &= \text{Var}(f) + \frac{\text{Cov}(f, g)^2}{\text{Var}(g)} - 2 \frac{\text{Cov}(f, g)^2}{\text{Var}(g)} \\ &= \boxed{\text{Var}(f) - \frac{\text{Cov}(f, g)^2}{\text{Var}(g)}}. \end{aligned}$$

The minimum variance involves squared correlation

When both variances are positive, define

$$\rho = \text{Corr}_p(f(X), g(X)) = \frac{\text{Cov}(f, g)}{\sqrt{\text{Var}(f) \text{Var}(g)}}.$$

Then

$$\rho^2 = \frac{\text{Cov}(f, g)^2}{\text{Var}(f) \text{Var}(g)}.$$

Factoring out $\text{Var}(f)$ yields

$$\boxed{V(c^*) = \text{Var}(f)(1 - \rho^2).}$$

The variance of the optimally adjusted estimator

Restore the factor $1/n$:

$$\text{Var}(\hat{\mathcal{I}}_{\text{CV}}) = \frac{\text{Var}_p(f(X))}{n}(1 - \rho^2), \quad c = c^\star.$$

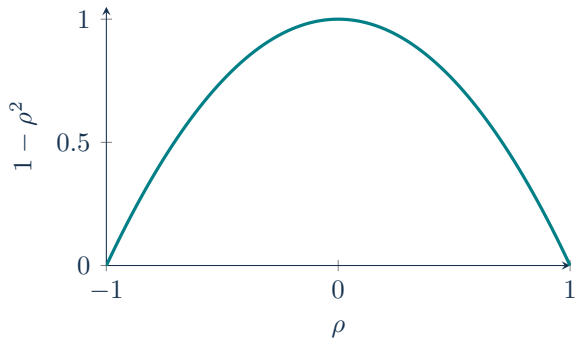
Thus

$$\frac{\text{Var}(\hat{\mathcal{I}}_{\text{CV}})}{\text{Var}(\hat{\mathcal{I}}_{\text{MC}})} = 1 - \rho^2.$$

- $0 \leq \rho^2 \leq 1$, so the optimal variance cannot increase.
- Nonzero correlation gives a strict reduction.

The guarantee applies to $c^\star$, not to an arbitrary coefficient.

The magnitude of correlation determines the gain



At the optimal coefficient:

- $\rho = 0$: no reduction.
- $|\rho|$ near one: large reduction.
- $|\rho| = 1$: zero variance.

Zero correlation gives no linear improvement

If $\rho = 0$, then $\text{Cov}(f, g) = 0$ and $c^* = 0$.

$$\text{Var}(\hat{\mathcal{I}}_{\text{CV}}) = \frac{\text{Var}_p(f(X))}{n}.$$

Zero correlation means **uncorrelated**.

It does **not** imply independence in general.

Independence implies zero covariance when the relevant moments exist.

Perfect correlation gives the ideal limiting case

If $\rho^2 = 1$, then

$$V(c^\star) = \text{Var}(f)(1 - \rho^2) = 0.$$

For nonconstant $f(X)$ and $g(X)$, this means an almost-sure affine relation:

$$f(X) = a + b g(X), \quad b \neq 0.$$

Taking $c^\star = -b$ makes

$$f_{\text{CV}}(X) = a + b g_0 = \mathcal{I} \quad \text{almost surely.}$$

Choosing the integrand itself is circular

The obvious candidate $g = f$ has $\rho = 1$ and $c^\star = -1$.

$$f_{\text{CV}}(X) = f(X) - (f(X) - g_0) = g_0.$$

But the required known expectation is

$$g_0 = \mathbb{E}_p[g(X)] = \mathbb{E}_p[f(X)] = \mathcal{I}.$$

This choice requires the very answer we wanted to compute.

A useful control combines tractability and correlation

We want a function g such that

- $g_0 = \mathbb{E}_p[g(X)]$ is known,
- $|\text{Corr}_p(f(X), g(X))|$ is large.

The quantity we want to make small is

$$1 - \rho^2.$$

Strong negative correlation is as useful as strong positive correlation.

The control variates construction

For a control with known mean g_0 and positive variance,

$$\widehat{\mathcal{I}}_{\text{CV}} = \overline{f} + c(\overline{g} - g_0), \quad \mathbb{E}[\widehat{\mathcal{I}}_{\text{CV}}] = \mathcal{I} \quad \text{for fixed } c.$$

The best population coefficient is

$$c^{\star} = -\frac{\text{Cov}_p(f(X), g(X))}{\text{Var}_p(g(X))}.$$

With positive variances,

$$\text{Var}(\widehat{\mathcal{I}}_{\text{CV}}) = \text{Var}(\widehat{\mathcal{I}}_{\text{MC}})(1 - \rho^2), \quad c = c^{\star}.$$