

Multiple Control Variates and Antithetic Variates

Using auxiliary information and dependent pairs

Junqi Tang

Computational Statistics
University of Birmingham

From one control to dependent pairs

Our target remains

$$\mathcal{I} = \mathbb{E}_p[f(X)], \quad X \sim p.$$

1. Review how correlation controls the variance reduction.
2. Combine several controls and solve for their coefficients.
3. Find controls using cheaper approximations.
4. Construct an estimator from antithetic pairs.

A centred control preserves the expectation

Suppose $g_0 = \mathbb{E}_p[g(X)]$ is known exactly. For fixed c ,

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

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

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

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

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

The expectation is unchanged for every fixed coefficient.

The optimal coefficient determines the gain

For $\text{Var}(g(X)) > 0$,

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

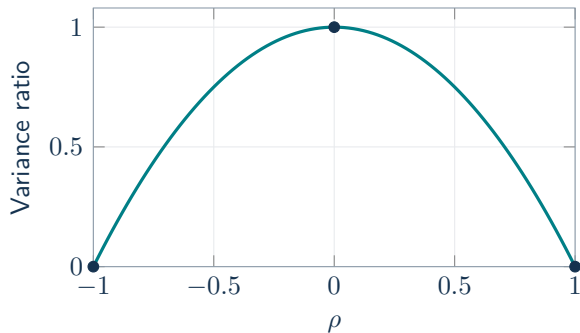
Let $\rho = \text{Corr}(f(X), g(X))$, with both variances positive.

$$\text{Var}(f(X) + c^*(g(X) - g_0)) = \text{Var}(f(X))(1 - \rho^2).$$

For the same number n of independent samples,

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

Strong correlation in either direction helps



$$0 \leq 1 - \rho^2 \leq 1.$$

- $\rho = 0$: no reduction.
- $|\rho| = 0.9$: retain 19% of the variance.
- $|\rho| = 1$: zero variance.

What the extreme cases mean

Perfect correlation in absolute value

$$|\rho| = 1 \iff f(X) = a + b g(X) \text{ almost surely,}$$

for nonconstant variables with finite variances and some $b \neq 0$.

- If a, b and g_0 are known, then $\mathcal{I} = a + b g_0$ directly.
- A perfect control is possible, but generally unavailable.

Zero correlation

$$\rho = 0 \implies c^* = 0.$$

The variables may still be dependent. This single linear correction gives no gain.

Choosing a useful control

The basic method needs an exactly known expectation:

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

For a substantial gain, we also want

$$|\text{Corr}(f(X), g(X))| \text{ close to } 1.$$

- An easy integral can still provide a weak control.
- A close approximation may have a difficult expectation.
- The cost of evaluating g affects the overall efficiency.

Control design depends on the information available in the problem.

Several controls can contribute together

Suppose we have L controls and know their means:

$$g_1, \dots, g_L, \quad g_{0,\ell} = \mathbb{E}_p[g_\ell(X)].$$

Use one coefficient for each centred control:

$$f_{\text{CV}}(x) = f(x) + \sum_{\ell=1}^L c_\ell (g_\ell(x) - g_{0,\ell}).$$

- The controls can be correlated with one another.
- Choose the coefficients jointly to account for this dependence.

Each added term has mean zero

For fixed $c_1, \dots, c_L$,

$$\begin{aligned}\mathbb{E}_p[f_{\text{CV}}(X)] &= \mathbb{E}_p[f(X)] + \sum_{\ell=1}^L c_\ell (\mathbb{E}_p[g_\ell(X)] - g_{0,\ell}) \\ &= \mathcal{I}.\end{aligned}$$

Therefore, for independent samples from p ,

$$\hat{\mathcal{I}}_{\text{CV}} = \frac{1}{n} \sum_{i=1}^n \left[f(X_i) + \sum_{\ell=1}^L c_\ell (g_\ell(X_i) - g_{0,\ell}) \right]$$

is unbiased.

Expanding the variance

Write $F = f(X)$ and $G_\ell = g_\ell(X)$. Constants do not affect variance.

$$\begin{aligned} V(\mathbf{c}) &= \text{Var} \left(F + \sum_{\ell=1}^L c_\ell G_\ell \right) \\ &= \text{Var}(F) + 2 \sum_{\ell=1}^L c_\ell \text{Cov}(F, G_\ell) \\ &\quad + \sum_{\ell=1}^L \sum_{k=1}^L c_\ell c_k \text{Cov}(G_\ell, G_k). \end{aligned}$$

The last term includes covariances between different controls.

Differentiation gives a linear system

For $j = 1, \dots, L$,

$$\frac{\partial V}{\partial c_j} = 2 \operatorname{Cov}(F, G_j) + 2 \sum_{k=1}^L c_k \operatorname{Cov}(G_j, G_k).$$

Setting each derivative to zero gives

$$\sum_{k=1}^L \operatorname{Cov}(G_j, G_k) c_k^* = -\operatorname{Cov}(F, G_j), \quad j = 1, \dots, L.$$

There are L linear equations for L coefficients.

The covariance matrix collects the equations

Define

$$\Sigma_{jk} = \text{Cov}(G_j, G_k), \quad b_j = \text{Cov}(G_j, F), \quad \mathbf{c} = (c_1, \dots, c_L)^\top.$$

Then

$$V(\mathbf{c}) = \text{Var}(F) + 2\mathbf{c}^\top \mathbf{b} + \mathbf{c}^\top \Sigma \mathbf{c}.$$

The optimality equations are

$$\boxed{\Sigma \mathbf{c}^* = -\mathbf{b}.}$$

If Σ is invertible,

$$\mathbf{c}^* = -\Sigma^{-1} \mathbf{b}.$$

When is the solution unique?

A covariance matrix satisfies $\Sigma \succeq 0$, so

$$\nabla^2 V(\mathbf{c}) = 2\Sigma \succeq 0.$$

- $\Sigma \succ 0$: the quadratic is strictly convex and the solution is unique.
- Linearly dependent centred controls make Σ singular.
- Remove redundant controls or use a pseudoinverse.

A stationary solution minimises the population variance.

A cheaper approximation can provide a control

Suppose evaluating $f(x)$ dominates the simulation cost. Construct a simpler approximation $g(x)$.

- It should be inexpensive to evaluate.
- Its fluctuations should track those of $f(X)$ under p .
- Ideally, its expectation can be calculated analytically.

$$f_{CV}(x) = f(x) + c(g(x) - \mathbb{E}_p[g(X)]).$$

Cheap evaluations can estimate the control mean

If one evaluation of g costs 5% of one evaluation of f ,

$$\frac{\text{cost of one } f \text{ evaluation}}{\text{cost of one } g \text{ evaluation}} = 20.$$

Ignoring overhead, the same evaluation budget buys twenty times as many values of g .

- A large cheap simulation may estimate g_0 accurately.
- The full budget must cover both the main calculation and the control mean.
- An estimated mean introduces additional uncertainty.

An accurate estimate of g_0 still has an error to account for.

Antithetic variates use negatively dependent pairs

Consider two random variables with

$$U \stackrel{d}{=} V, \quad \text{Cov}(U, V) < 0.$$

- Their marginal distributions are the same.
- Their joint distribution makes high and low values tend to occur together.
- Averaging the pair can cancel part of the fluctuation.

The useful dependence is between the quantities being averaged.

A normal variable and its reflection

If $Z \sim N(0, 1)$, then

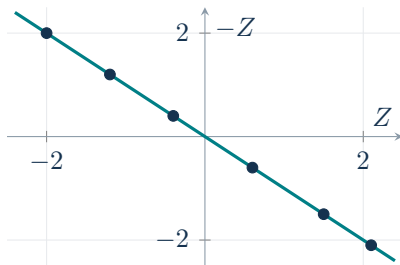
$$-Z \sim N(0, 1).$$

Their dependence is exact:

$$\text{Corr}(Z, -Z) = -1.$$

Their pair average is

$$\frac{Z + (-Z)}{2} = 0.$$



The partner must work for the function values

We want $\mathcal{I} = \mathbb{E}_p[f(X)]$. Suppose we can construct h such that

$$f(X) \stackrel{d}{=} h(X), \quad \text{Cov}(f(X), h(X)) < 0.$$

Equality in distribution gives

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

- The partner must have the correct marginal distribution.
- Check the covariance of the outputs $f(X)$ and $h(X)$.

The factor of two preserves the target

The pair sum has twice the target expectation:

$$\mathbb{E}_p[f(X) + h(X)] = 2\mathcal{I}.$$

Define the pair average

$$f_{\text{AV}}(x) = \frac{f(x) + h(x)}{2}.$$

Then

$$\mathbb{E}_p[f_{\text{AV}}(X)] = \frac{\mathcal{I} + \mathcal{I}}{2} = \mathcal{I}.$$

We can apply standard Monte Carlo to the pair average.

The antithetic Monte Carlo estimator

Generate n independent inputs $X_i \sim p$ and form n pairs:

$$\hat{\mathcal{I}}_{\text{AV}} = \frac{1}{2n} \sum_{i=1}^n [f(X_i) + h(X_i)].$$

- The two values within a pair are dependent.
- Different pairs are independent.
- n counts pairs, so the estimator uses $2n$ function values.

$$\mathbb{E}[\hat{\mathcal{I}}_{\text{AV}}] = \mathcal{I}.$$

What this lecture establishes

Multiple control variates

$$f_{\text{CV}} = f + \sum_{\ell=1}^L c_{\ell}(g_{\ell} - g_{0,\ell}), \quad \Sigma \mathbf{c}^* = -\mathbf{b}.$$

Known expectations centre the controls. Joint covariance determines the coefficients.

Antithetic variates

$$\hat{\mathcal{I}}_{\text{AV}} = \frac{1}{2n} \sum_{i=1}^n [f(X_i) + h(X_i)].$$

Equal marginals preserve the mean. Negative output covariance motivates the pairing.

Next lecture: the variance of the antithetic estimator.