CSCI 145 Problem 2: The Optimal Control Variate

October 27, 2026

Submission Instructions

Please upload your work to Gradescope by **11:59pm on Wednesday, September 9**.

- Turn in a single PDF of *handwritten* solutions. Typed (including LaTeX) solutions will not be accepted.
- A complete, good-faith attempt earns submission credit. The grader may select any recent assigned problem for a brief conversation, so keep your work and be prepared to explain each solution and answer a related question. Scheduling details will be announced in class.

Problem 2: The Optimal Control Variate

Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector with distribution P, and let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent and identically distributed samples from P. A scientific simulation computes an expensive output $F(\mathbf{X}_i)$ and a surrogate $G(\mathbf{X}_i)$ from each sample, where $F, G : \mathbb{R}^d \to \mathbb{R}$. The quantity of interest is $\mu = \mathbb{E}[F(\mathbf{X})]$, while $\mathbb{E}[G(\mathbf{X})]$ is known exactly from the model. Assume $F(\mathbf{X})$ and $G(\mathbf{X})$ have finite positive variances. For a chosen constant $c \in \mathbb{R}$, the control-variate estimator is

$$\widehat{\mu}_{n,c} = \frac{1}{n} \sum_{i=1}^{n} \left[F(\mathbf{X}_i) - c \left(G(\mathbf{X}_i) - \mathbb{E}[G(\mathbf{X})] \right) \right].$$

Additional simulation runs are costly, so we want to choose the member of this family that extracts as much information as possible from G.

1. What is $\mathrm{Var}(\widehat{\mu}_{n,c})$, and which value of c globally minimizes it?
2. What multiple of the ordinary Monte Carlo variance $\mathrm{Var}(F(\mathbf{X}))/n$ is attained at this optimum?

Hint: The estimator above is unbiased for every c because $c(G(\mathbf{X}_i) - \mathbb{E}[G(\mathbf{X})])$ has mean zero. For independent, identically distributed variables with finite variance, the variance of their average is the one-sample variance divided by n. You may also use $\mathrm{Var}(U + aV) = \mathrm{Var}(U) + 2a\,\mathrm{Cov}(U, V) + a^2\,\mathrm{Var}(V)$.

Takeaway: *A second measurement with a known mean can make an estimate more precise without running more simulations. It helps most when it is strongly correlated with the quantity we want to estimate.*