Some variance reduction methods

Felipe Uribe

Computational Engineering
School of Engineering Sciences
Lappeenranta-Lahti University of Technology (LUT)

Special Course on Inverse Problems
Lappeenranta, FI — January-February, 2024

Why variance reduction?

- We have seen that standard MC typically has an error variance of the form σ^2/n. We get a better answer with larger n, but the computing time grows with n.

- Sometimes we can find a way to reduce σ instead. We construct a new Monte Carlo problem with the same answer as our original one but with a lower $\sigma \implies$ variance reduction techniques.

Why variance reduction?

- We have seen that standard MC typically has an error variance of the form σ^2/n. We get a better answer with larger n, but the computing time grows with n.

- Sometimes we can find a way to reduce σ instead. We construct a new Monte Carlo problem with the same answer as our original one but with a lower $\sigma \implies$ variance reduction techniques.

- We can group the methods in the following categories:
  - Type-1: antithetic sampling, stratification, and common random numbers.
  - Type-2: conditioning and control variates.
  - Type-3: importance sampling and its variants (we will skip this one in the interest of time).

- These methods are also used in combination with MCMC.

This lecture...

- The lecture is based on multiple references. However, we mostly follow Chapters 8 and 9 of the book by **Art Owen**[1], which is freely available online.

[1] A. B. Owen. *Monte Carlo theory, methods and examples*. `artowen.su.domains/mc/`, 2018.

Variance reduction: type-1 methods (“using clever samples”)

Antithetic sampling: intro

- Random variables X, Y on the same probability space are antithetic, if they have the same distribution and their covariance is negative.
- When we are using Monte Carlo averages of quantities $f(\boldsymbol{x}_i)$ then the randomness in the algorithm leads to some error cancellation. In antithetic sampling, we try to get even more cancellation.
- An antithetic sample $\widetilde{\boldsymbol{x}}$ is one that gives the opposite value of $f(\boldsymbol{x})$, i.e., being low when $f(\boldsymbol{x})$ is high and vice versa. Ordinarily, we get an opposite f by sampling at a point $\widetilde{\boldsymbol{x}}$ that is *somehow* opposite to $\boldsymbol{x}$.
- Let $\mu = \mathbb{E}[\boldsymbol{X}]$ for $\boldsymbol{X} \sim \pi$, where π is a symmetric density on $\mathbb{R}^d$. Here, symmetry is with respect to reflection through the *center point* $\boldsymbol{c}$ of $\mathbb{R}^d$.

Antithetic sampling: estimator

- If we reflect $\boldsymbol{x}$ through $\boldsymbol{c}$, we have $\widetilde{\boldsymbol{x}} - \boldsymbol{c} = -(\boldsymbol{x} - \boldsymbol{c})$, and we get the point $\widetilde{\boldsymbol{x}} = 2\boldsymbol{c} - \boldsymbol{x}$. For basic examples, when $\pi = \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ then $\widetilde{\boldsymbol{x}} = -\boldsymbol{x}$. When $\pi = \mathcal{U}(0,1)^d$, we have $\widetilde{\boldsymbol{x}} = 1 - \boldsymbol{x}$ (componentwise).

- The antithetic sampling estimate of μ is:

$$\mu \approx \hat{\mu}_{\text{anti}} = \frac{1}{n} \sum_{i=1}^{n/2} f(\boldsymbol{x}_i) + f(\widetilde{\boldsymbol{x}}_i), \tag{1}$$

  where $\boldsymbol{x}_i \overset{\text{idd}}{\sim} \pi$ and n is an even number. This estimator is also **unbiased**.

- The rationale for antithetic sampling is that each value of $\boldsymbol{x}$ is *balanced* by its opposite $\widetilde{\boldsymbol{x}}$, satisfying $(\boldsymbol{x} + \widetilde{\boldsymbol{x}})/2 = \boldsymbol{c}$.

Antithetic sampling

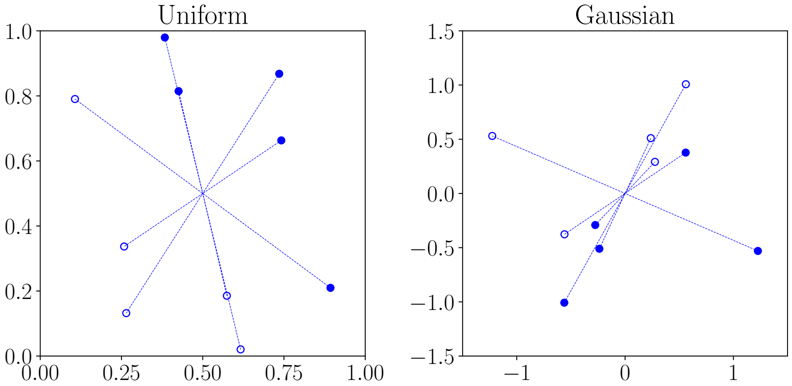

Figure: Five points and their antithetics. Left: from a standard uniform. Right: from a standard Gaussian.

Antithetic sampling: variance

- Whether the balance is helpful or not depends on f. If f is nearly linear, we could obtain a large improvement.
- The variance of antithetic sampling is:

$$\mathbb{V}[\hat{\mu}_{\mathsf{anti}}] = \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n/2} f(\boldsymbol{x}_i) + f(\widetilde{\boldsymbol{x}}_i)\right] = \frac{n/2}{n^2}\mathbb{V}\left[f(\boldsymbol{X}) + f(\widetilde{\boldsymbol{X}})\right] \tag{2}$$

$$= \frac{1}{2n}\left(\mathbb{V}[f(\boldsymbol{X})] + \mathbb{V}\left[f(\widetilde{\boldsymbol{X}})\right] + 2\mathbb{C}\mathrm{ov}\left[f(\boldsymbol{X}), f(\widetilde{\boldsymbol{X}})\right]\right) = \frac{\sigma^2}{n}(1+\rho) \tag{3}$$

- Since $-1 \leq \rho \leq 1$, we obtain $0 \leq \sigma^2(1+\rho) \leq 2\sigma^2$. In the best case, antithetic sampling gives the exact answer from just one pair of function evaluations. In the worst case, it doubles the variance.

Antithetic sampling: when it works?

- Hence, the variance of standard MC and antithetics can be written as:

$$\begin{bmatrix} \mathbb{V}[\hat{\mu}] \\ \mathbb{V}[\hat{\mu}_{\text{anti}}] \end{bmatrix} = \frac{1}{n} \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} \sigma_e^2 \\ \sigma_o^2 \end{bmatrix}; \tag{4}$$

  antithetic sampling eliminates the variance of f_o but doubles the contribution from f_e.

- **Tip**: antithetic sampling reduces the variance if $\rho < 0$ (e.g., monotone function), or equivalently if $\sigma_o^2 > \sigma_e^2$. This analysis is appropriate when the most of the computation is in evaluating f.

- Because antithetic samples have dependent values within pairs. We can define $y_i = f_e(\boldsymbol{x}_i) = (f(\boldsymbol{x}_i) + f(\widetilde{\boldsymbol{x}}_i))/2$, for $i = 1, \ldots, m = n/2$, then

$$\hat{\mu}_{\text{anti}} = \frac{1}{m}\sum_{i=1}^{m} y_i, \qquad \sigma^2_{\text{anti}} = \frac{1}{m-1}\sum_{i=1}^{m}(y_i - \hat{\mu}_{\text{anti}})^2. \tag{5}$$

Antithetic sampling: example (I)

Consider the expected logarithmic return of a portfolio:

- There are K stocks and the portfolio has proportion $\lambda_k \geq 0$ in stock k, with $\sum_{k=1}^{K} \lambda_k = 1$.
- The expected logarithmic return is defined as

$$\mu(\lambda) = \mathbb{E}\left[\log\left(\sum_{k=1}^{K} \lambda_k \, \exp(X_k)\right)\right], \tag{6}$$

  where $\boldsymbol{X} \in \mathbb{R}^K$ is the vector of returns.

- If one keeps reinvesting/rebalancing the portfolio at N regular time intervals then, by the LLN, our fortune grows as $\exp(N\mu + \mathcal{O}(N))$, assuming of course that the $\boldsymbol{X}$ for each time period are idd.

Antithetic sampling: example (II)

- The log-optimal choice λ is the allocation that maximizes μ. Finding a model for the distribution of $\boldsymbol{X}$ and then choosing λ are challenging problems. We focus on the problem of evaluating $\mu(\lambda)$ for a given λ.
- We take $\lambda_k = 1/K$ with $K = 500$. We also suppose that each marginal distribution is $X_k \sim \mathcal{N}(\delta, \sigma^2)$ but that $\boldsymbol{X}$ has the $t(0, \nu, \boldsymbol{\Sigma})$ copula. Here $\delta = 0.001$ and $\sigma = 0.03$ ($\approx$ one week time frame). And $\nu = 4$ with covariance is $\boldsymbol{\Sigma} = \rho \mathbf{1}_K \mathbf{1}_K^{\mathsf{T}} + (1-\rho)\mathbf{I}_k^{\mathsf{T}}$ for $\rho = 0.3$.
- Letting $f(\boldsymbol{X}) = \log\left(\sum_{k=1}^{K} \exp(X_k)/K\right)$, the MC estimate is $\hat{\mu} = 1/n \sum_{i=1}^{n} f(\boldsymbol{X}_i)$.
- The antithetic to $\boldsymbol{X}_i$ has components $\widetilde{X}_{ik} = 2\delta - X_{ik}$.
- Continue on code...

Variance reduction: type-2 methods ("using things we know")

Control variates: intro

- Control variates provide a way to exploit closed form results. With control variates we use some other problem, quite similar to our given one, but for which an exact answer is known.

- Suppose first that we want to find $\mu = \mathbb{E}[f(\boldsymbol{X})]$ and that we know the value $\theta = \mathbb{E}[h(\boldsymbol{X})]$, where $h(\boldsymbol{X}) \approx f(\boldsymbol{X})$. Using the MC estimators for each of these quantities:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{X}_i) \qquad \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} h(\boldsymbol{X}_i) \tag{7}$$

  we can estimate μ, using the (unbiased) **difference estimator**:

$$\hat{\mu}_{\text{diff}} = \frac{1}{n}\sum_{i=1}^{n}\left(f(\boldsymbol{X}_i) - h(\boldsymbol{X}_i)\right) + \theta = \hat{\mu} - \hat{\theta} + \theta. \tag{8}$$

Control variates: estimators

- The variance of the difference estimator is

$$\mathbb{V}[\hat{\mu}_{\text{diff}}] = \frac{1}{n}\mathbb{V}[f(\boldsymbol{X}) - h(\boldsymbol{X})]\,. \tag{9}$$

- If h is similar to f in the sense that the difference $f(\boldsymbol{X}) - h(\boldsymbol{X})$ has smaller variance than $f(\boldsymbol{X})$, we will reduce the variance. In this setting, $h(\boldsymbol{X})$ is called the control variate.

- The difference estimator is not the only way to use a control variate. The ratio and product estimators are also used:

$$\hat{\mu}_{\text{ratio}} = \frac{\hat{\mu}}{\hat{\theta}}\theta \qquad \hat{\mu}_{\text{prod}} = \frac{\hat{\mu}\hat{\theta}}{\theta}; \tag{10}$$

  however, the ratio and product estimators are usually biased.

Control variates: regression estimator (I)

- By far the most common way of using a control variate is through the regression. For a value $\beta \in \mathbb{R}$, the (unbiased) regression estimator of μ is:

$$\hat{\mu}_\beta = \frac{1}{n}\sum_{i=1}^{n}(f(\boldsymbol{X}_i) - \beta h(\boldsymbol{X}_i)) + \beta\theta = \hat{\mu} - \beta(\hat{\theta} - \theta); \quad (11)$$

  note that $\beta = 0$ gives standard MC and $\beta = 1$ yields the difference estimator.

- The variance of this estimator is:

$$\mathbb{V}[\hat{\mu}_\beta] = \frac{1}{n}\left(\mathbb{V}[f(\boldsymbol{X})] - 2\beta\mathbb{C}\mathrm{ov}\left[f(\boldsymbol{X}), h(\boldsymbol{X})\right] + \beta^2\mathbb{V}[h(\boldsymbol{X})]\right). \quad (12)$$

- **Intuition**: control variates create a new random vector $\boldsymbol{Z} = f(\boldsymbol{X}) + \beta(h(\boldsymbol{X}) - \theta)$, that allows us to leverage θ in order to compute $\mathbb{E}[f(\boldsymbol{X})]$ in an easier way.

Control variates: regression estimator (II)

- We can find the optimal value of β as:

$$\beta_{\mathsf{opt}} = \arg\min_{\beta} \mathbb{V}[\hat{\mu}_\beta] = \frac{\mathbb{C}\mathrm{ov}\left[f(\boldsymbol{X}), h(\boldsymbol{X})\right]}{\mathbb{V}[h(\boldsymbol{X})]} \quad \text{and} \quad \mathbb{V}\left[\hat{\mu}_{\beta_{\mathsf{opt}}}\right] = \frac{\sigma^2}{n}(1-\rho^2); \tag{13}$$

  note that in the regression estimator, any control variate that correlates with f is helpful, even one that correlates negatively.

- Since we do not know β_{opt} in practice, it can be estimated as

$$\beta_{\mathsf{opt}} \approx \hat{\beta} = \frac{\sum_{i=1}^{n}(f(\boldsymbol{X}_i) - \hat{\mu})(h(\boldsymbol{X}_i) - \hat{\theta})}{\sum_{i=1}^{n}(h(\boldsymbol{X}_i) - \hat{\theta})^2}; \tag{14}$$

  note that the estimator $\hat{\mu}_{\hat{\beta}}$ is no longer unbiased. But the bias is very small !

Control variates: regression estimator (III)

- The estimated variance of $\hat{\mu}_{\hat{\beta}}$ is

$$\hat{\sigma}^2_{\hat{\beta}} = \mathbb{V}\left[\hat{\mu}_{\hat{\beta}}\right] = \frac{1}{n^2}\sum_{i=1}^{n}\left(f(\boldsymbol{X}_i) - \hat{\mu}_{\hat{\beta}} - \hat{\beta}(h(\boldsymbol{X}_i) - \hat{\theta})\right)^2. \qquad (15)$$

  and a 99% confidence interval is $\hat{\mu}_{\hat{\beta}} \pm 2.58\,\hat{\sigma}_{\hat{\beta}}$.

- The variance with a control variate is never worse than the MC one. Whether the control variate is helpful ultimately depends on **how much it costs to use it**.

- A significant advantage of the regression estimator is that it generalizes easily to handle multiple control variates. The potential value is greatest when f is expensive but is approximately equal to a linear combination of inexpensive control variates.

Variance reduction: final comments

- Variance reduction is an ongoing field of research in UQ, for both forward and inverse problems.
- Many of the methods exposed here can be extended to the case of inverse problems within the Bayesian framework.
- We will see that practical UQ for inverse problems requires a solid foundation on stochastic simulation (i.e., the methods discussed in the past lectures).

References

[1] H. Kahn et al. "Methods of reducing sample size in Monte Carlo computations". In: *Journal of the Operations Research Society of America* 1.5 (1953), pp. 263–278.

[2] A. B. Owen. *Monte Carlo theory, methods and examples*. `artowen.su.domains/mc/`, 2018.

Disclaimer: all figures are either generated by the Author or under Creative Commons licenses