Optimization and Simulation

Variance reduction

Michel Bierlaire

Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne

Simulation runs and random inputs

Random inputs

A simulation run uses random numbers, typically

- uniform draws: $U_{r,1}, \ldots, U_{r,m} \sim U(0,1)$,
- normal draws: $\xi_{r,1}, \ldots, \xi_{r,k} \sim N(0,1)$,

all independent.

Simulation output

The output of run r is obtained by a deterministic function

$$X_r = h(U_{r,1}, \ldots, U_{r,m}, \xi_{r,1}, \ldots, \xi_{r,k}).$$

Repeating the simulation produces i.i.d. draws

$$X_1, \ldots, X_R \quad \text{with} \quad \mathbb{E}[X_r] = \theta.$$

What is variance reduction?

Baseline simulation

Using the function h, repeated runs produce i.i.d. outputs

$$X_1, \ldots, X_R, \qquad \mathbb{E}[X_r] = \theta,$$

Variance reduction

Variance reduction modifies the construction of the output by using another function $\tilde{h}$, applied to the same types of random inputs:

$$\tilde{X}_r = \tilde{h}(U_{r,1}, \ldots, U_{r,m}, \xi_{r,1}, \ldots, \xi_{r,k}).$$

What is variance reduction?

Variance reduction

The resulting outputs

$$\tilde{X}_1, \ldots, \tilde{X}_R$$

are still i.i.d., satisfy

$$\mathbb{E}[\tilde{X}_r] = \theta, \qquad \mathrm{Var}(\tilde{X}_r) < \mathrm{Var}(X_r).$$

Estimating the mean by simulation

Estimator

We estimate

$$\theta = \mathbb{E}[X_r]$$

using the sample mean

$$\hat{\theta}_R = \frac{1}{R}\sum_{r=1}^{R} X_r,$$

or

$$\tilde{\theta}_R = \frac{1}{R}\sum_{r=1}^{R} \tilde{X}_r.$$

Estimating the mean by simulation

Precision (bootstrap lecture)

If $X_1, \ldots, X_R$ are i.i.d., then

$$\mathbb{E}\left[(\hat{\theta}_R - \theta)^2\right] = \frac{\mathrm{Var}(X_r)}{R}.$$

The typical estimation error is therefore of order

$$\sqrt{\frac{\mathrm{Var}(X_r)}{R}}.$$

Why variance reduction is useful

Baseline vs variance-reduced simulation

$$\mathbb{E}\left[(\hat{\theta}_R - \theta)^2\right] = \frac{\mathsf{Var}(X_r)}{R}; \;\; \mathbb{E}\big[(\tilde{\theta}_R - \theta)^2\big] = \frac{\mathsf{Var}(\tilde{X}_r)}{R}.$$

Key consequence

Since

$$\mathsf{Var}(\tilde{X}_r) < \mathsf{Var}(X_r),$$

the variance-reduced estimator has a smaller mean square error for the same number of simulation runs R.

Outline

Anthitetic draws

Control variates

Other techniques

Beyond the mean

Antithetic draws

Intuition

Instead of simulating two independent scenarios, we simulate two opposite scenarios so that their errors tend to cancel.

Example

Use simulation to compute

$$I = \int_0^1 e^x \, dx$$

We know the solution: $e - 1 = 1.7183$

Simulation: consider draws two by two

- Let $r_1, \ldots, r_R$ be independent draws from $U(0,1)$.
- Let $s_1, \ldots, s_R$ be independent draws from $U(0,1)$.

$$I \approx \frac{1}{2R}\left(\sum_{i=1}^{R} e^{r_i} + \sum_{i=1}^{R} e^{s_i}\right) = \frac{1}{R}\sum_{i=1}^{R} \frac{e^{r_i} + e^{s_i}}{2}$$

Example

Simulation: consider draws two by two

- Use $R = 10'000$ (that is, a total of 20'000 draws)
- Mean over R draws from $(e^{r_i} + e^{s_i})/2$: 1.720,variance: 0.123.

Example

Now, use half the number of draws

- Idea: if $X \sim U(0,1)$, then $(1-X) \sim U(0,1)$
- Let $r_1, \ldots, r_R$ be independent draws from $U(0,1)$.

$$I \approx \frac{1}{R} \sum_{i=1}^{R} \frac{e^{r_i} + e^{1-r_i}}{2}$$

- Use $R = 10'000$
- Mean over R draws of $(e^{r_i} + e^{1-r_i})/2$: 1.7183,variance: 0.00388.
- Compared to: mean of $(e^{r_i} + e^{s_i})/2$: 1.720,variance: 0.123.

Antithetic draws: h and $\tilde{h}$

Baseline simulation

For one simulation run r, generate two independent uniforms

$$U_{r,1}, U_{r,2} \sim U(0,1),$$

and define

$$X_r = h(U_{r,1}, U_{r,2}) = \frac{e^{U_{r,1}} + e^{U_{r,2}}}{2}.$$

The outputs $X_1, \ldots, X_R$ are i.i.d.

Antithetic draws: h and $\tilde{h}$

Antithetic construction

For one simulation run r, generate a single uniform

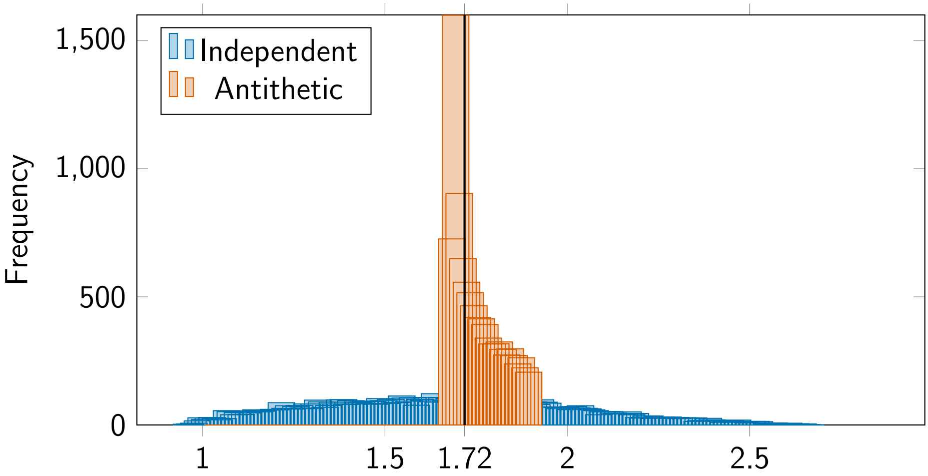
$$U_r \sim U(0,1),$$

and define

$$\tilde{X}_r = \tilde{h}(U_r) = \frac{e^{U_r} + e^{1-U_r}}{2}.$$

The outputs $\tilde{X}_1, \ldots, \tilde{X}_R$ are also i.i.d.

Example



Independent
Antithetic
Frequency
1,500
1,000
500
0
1
1.5
1.72
2
2.5

Antithetic draws

- Let X_1 and X_2 i.i.d. r.v. with mean θ.
- Then

$$\mathrm{Var}\left(\frac{X_1+X_2}{2}\right)=\frac{1}{4}\left(\mathrm{Var}(X_1)+\mathrm{Var}(X_2)+2\,\mathrm{Cov}(X_1,X_2)\right).$$

- If X_1 and X_2 are independent, then $\mathrm{Cov}(X_1,X_2)=0$.
- If X_1 and X_2 are negatively correlated, then $\mathrm{Cov}(X_1,X_2)<0$, and the variance is reduced.

Back to the example

Independent draws

- $X_1 = e^U$, $X_2 = e^U$

$$
\begin{aligned}
\text{Var}(X_1) = \text{Var}(X_2) &= \quad E[e^{2U}] \quad - \quad E[e^U]^2 \\
&= \int_0^1 e^{2x}\,dx \quad - \quad (e-1)^2 \\
&= \quad \tfrac{e^2-1}{2} \quad - \quad (e-1)^2 \\
&= \quad 0.2420
\end{aligned}
$$

$$\text{Cov}(X_1, X_2) = 0$$

$$\text{Var}\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4}\,(0.2420 + 0.2420)) = 0.1210$$

Back to the example

Antithetic draws

- $X_1 = e^U$, $X_2 = e^{1-U}$

$$\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = 0.2420$$

$$\begin{aligned} \mathrm{Cov}(X_1, X_2) &= \mathrm{E}[e^U e^{1-U}] - E[e^U]E[e^{1-U}] \\ &= e - (e-1)(e-1) \\ &= -0.2342 \end{aligned}$$

$$\mathrm{Var}\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4}\left(0.2420 + 0.2420 - 2\ 0.2342)\right) = 0.0039$$

Antithetic draws: generalization

- Suppose that
$$X_1 = h(U_1, \ldots, U_m),$$
where $U_1, \ldots U_m$ are i.i.d. $U(0, 1)$.
- Define
$$X_2 = h(1 - U_1, \ldots, 1 - U_m).$$
- X_2 has the same distribution as X_1
- If h is monotonic in each of its coordinates, then X_1 and X_2 are negatively correlated.
- If h is not monotonic, there is no guarantee that the variance will be reduced.

Another example

$$I = \int_0^1 \left(x - \frac{1}{2}\right)^2 dx$$

- Antithetic draws:

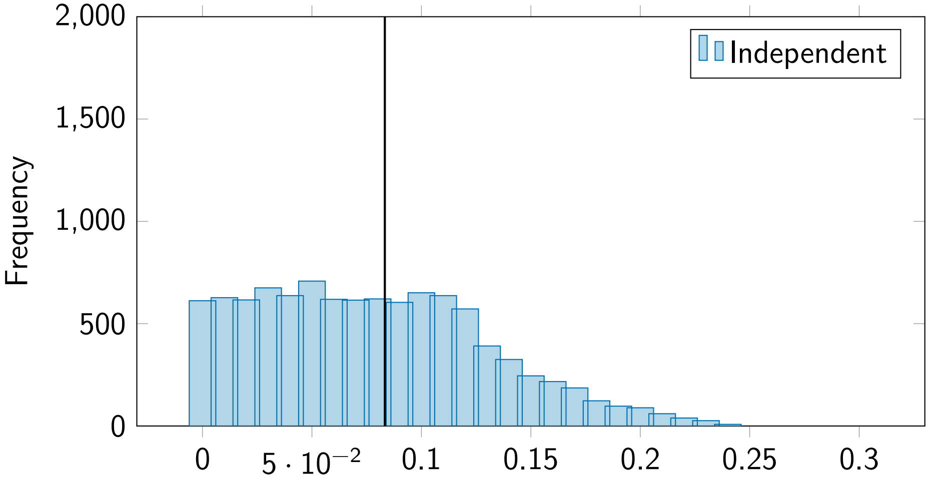
$$X_1 = \left(U - \frac{1}{2}\right)^2, \; X_2 = \left((1 - U) - \frac{1}{2}\right)^2$$

- The covariance is positive:

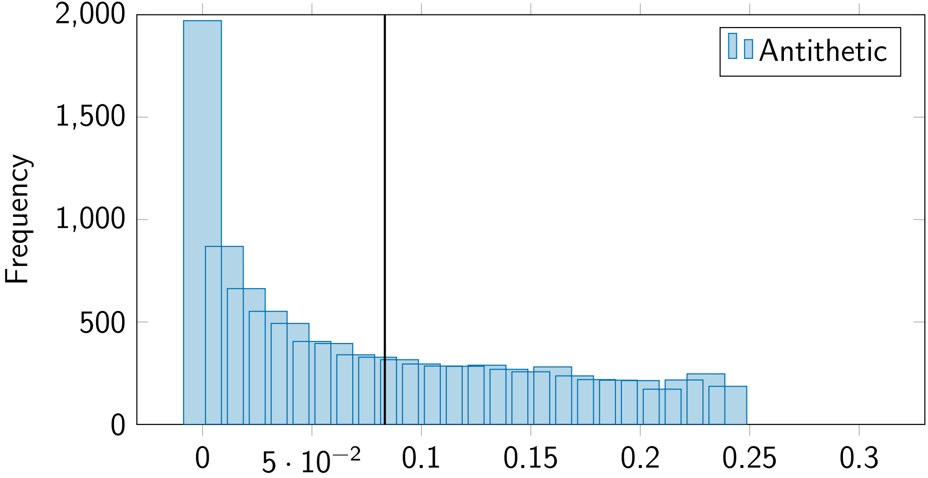
$$\mathrm{Cov}(X_1, X_2) = \frac{1}{180} > 0.$$

- The variance will therefore be (slightly) increased!

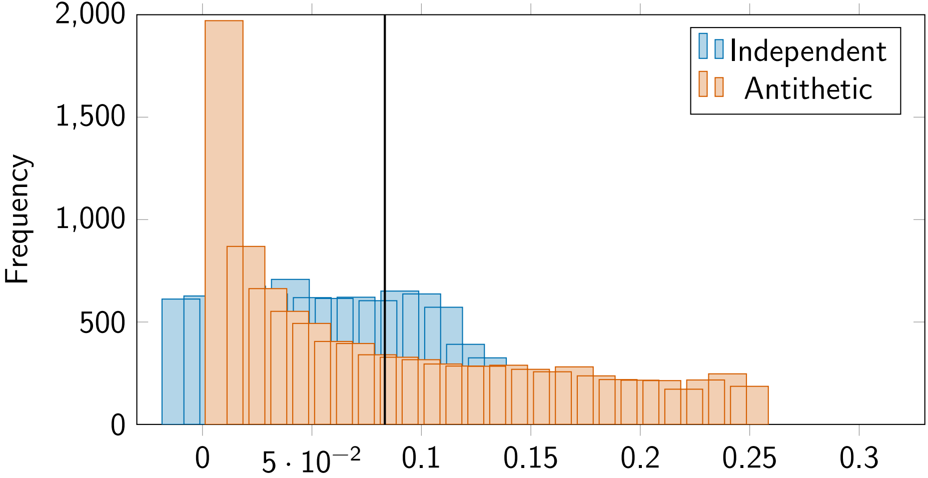
Another example



Another example



Another example



Antithetic draws: practical rule

Core idea

Applying a simple symmetry to the underlying random inputs.

Most common cases

- **Uniform draws** If $U \sim U(0,1)$, the antithetic draw is

$$U^{\text{anti}} = 1 - U.$$

- **Normal draws** If $\xi \sim N(0,1)$, the antithetic draw is

$$\xi^{\text{anti}} = -\xi.$$

Key property

In both cases, the antithetic draw has the same marginal distribution as the original draw, but is perfectly negatively correlated with it.

Outline

Control variates

Motivation

Suppose the simulator produces an output X whose mean we want, but also another output Y:

- strongly correlated with X,
- whose expectation $\mathbb{E}[Y] = \mu$ is known exactly.

Can we exploit this information to reduce variance?

Control variates: h and $\tilde{h}$

Simulation output

For one simulation run r, the simulator produces

$$(X_r, Y_r) = h(U_{r,1}, \ldots, U_{r,m}, \xi_{r,1}, \ldots, \xi_{r,k}),$$

where

$$\mathbb{E}[X_r] = \theta, \qquad \mathbb{E}[Y_r] = \mu \ \text{(known)}.$$

Control variates: h and $\tilde{h}$

Variance-reduced output

Using the same run, define

$$\tilde{X}_r = \tilde{h}(X_r, Y_r) = X_r + c\,(Y_r - \mu).$$

For any constant c,

$$\mathbb{E}[\tilde{X}_r] = \theta.$$

Control variates

- We use simulation to estimate $\theta = \mathrm{E}[X]$, where X is an output of the simulation.
- Let Y be another output of the simulation, such that we know $\mathrm{E}[Y] = \mu$.
- We consider the quantity:

$$Z = X + c(Y - \mu).$$

- By construction, $\mathrm{E}[Z] = \mathrm{E}[X]$.
- Its variance is

$$\mathrm{Var}(Z) = \mathrm{Var}(X + cY) = \mathrm{Var}(X) + c^2\,\mathrm{Var}(Y) + 2c\,\mathrm{Cov}(X, Y).$$

- Find c such that $\mathrm{Var}(Z)$ is minimum.

Control variates

- First derivative:
$$2c\,\mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y).$$
- Zero if
$$c^* = -\frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(Y)}.$$
- Second derivative:
$$2\,\mathrm{Var}(Y) > 0.$$
- We use
$$Z^* = X - \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(Y)}(Y - \mu).$$
- Its variance
$$\mathrm{Var}(Z^*) = \mathrm{Var}(X) - \frac{\mathrm{Cov}(X, Y)^2}{\mathrm{Var}(Y)} \leqslant \mathrm{Var}(X).$$

Control variates

In practice...

- $\text{Cov}(X, Y)$ and $\text{Var}(Y)$ are usually not known.
- We can use their sample estimates:

$$\widehat{\text{Cov}}(X, Y) = \frac{1}{n-1}\sum_{r=1}^{R}(X_r - \bar{X})(Y_r - \bar{Y}),$$

and

$$\widehat{\text{Var}}(Y) = \frac{1}{n-1}\sum_{r=1}^{R}(Y_r - \bar{Y})^2.$$

Control variates

In practice...

- Alternatively, use linear regression

$$X = aY + b + \varepsilon$$

  where $\varepsilon \sim N(0, \sigma^2)$.

- The least square estimators of a and b are

$$\hat{a} = \frac{\widehat{\text{Cov}}(X, Y)}{\widehat{\text{Var}}(Y)} = \frac{\sum_{r=1}^{R}(X_r - \bar{X})(Y_r - \bar{Y})}{\sum_{r=1}^{R}(Y_r - \bar{Y})^2}$$

$$\hat{b} = \bar{X} - \hat{a}\bar{Y}.$$

- Therefore

$$c^* = -\hat{a}.$$

Control variates

- Moreover,

$$\begin{aligned} \hat{b} + \hat{a}\mu &= \bar{X} - \hat{a}\bar{Y} + \hat{a}\mu \\ &= \bar{X} - \hat{a}(\bar{Y} - \mu) \\ &= \bar{X} + c^*(\bar{Y} - \mu) \\ &= \widehat{\theta}. \end{aligned}$$

- Therefore, the control variate estimate $\widehat{\theta}$ of θ is obtained by the estimated linear model, evaluated at μ.

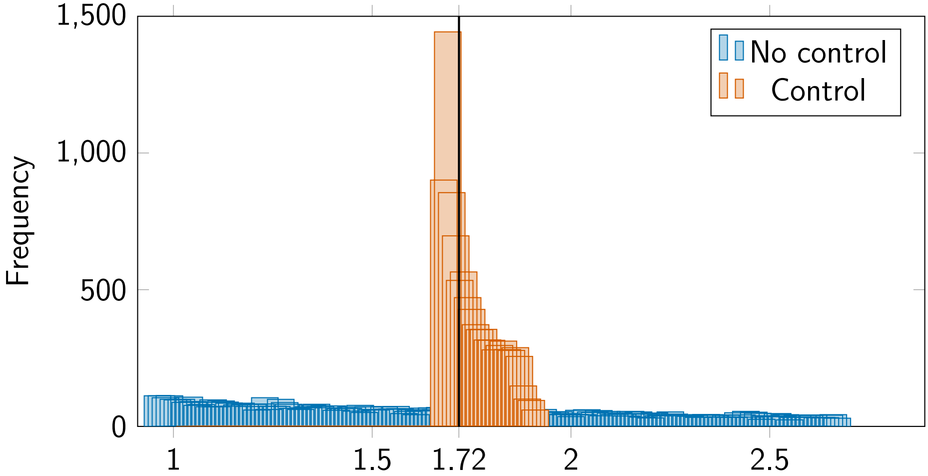
Back to the example

- Use simulation to compute $I = \int_0^1 e^x \, dx$.
- $X = e^U$.
- $Y = U$, $E[Y] = 1/2$, $Var(Y) = 1/12$.
- $Cov(X, Y) = (3 - e)/2 \approx 0.14$.
- Therefore, the best c is

$$c^* = -\frac{Cov(X, Y)}{Var(Y)} = -6(3 - e) \approx -1.69.$$

- Test with $R = 10'000$.
- Result of the regression: $\hat{a} = 1.6893$, $\hat{b} = 0.8734$.
- Estimate: $\hat{b} + \hat{a}/2 = 1.7180$, Variance: 0.003847 (compared to 0.24).

Back to the example

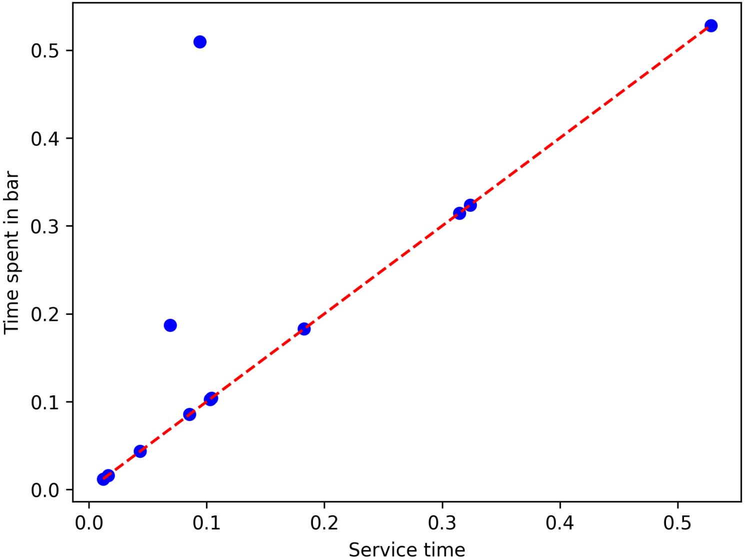


Satellite simulation

Variables

- X: average time spent by the customers in the bar.
- Y: average service time.

Satellite simulation: one run



Satellite simulation

True value of $\mathrm{E}[Y]$

- The average service time $\mu = 0.2$ is known.
- Therefore,

$$\mathrm{E}[Y] = \mu = 0.2.$$

Important

Do not use simulated values to calculate this quantity.

Satellite simulation:

Scenario: closure: 100,inter-arrival time: 1

	R	Service time	E[X]	E[Z]	Var[X]	Var[Z]
0	1000	0.1	0.1115	0.1111	0.0001676	3.129e-05
1	10000	0.1	0.1107	0.1111	0.0001857	3.153e-05
2	100000	0.1	0.1110	0.1110	0.0001827	3.111e-05
3	1000	1	7.665	7.771	21.91	12.74
4	10000	1	7.820	7.800	22.23	13.66
5	100000	1	7.780	7.773	22.04	13.69
6	1000	3	102.3	102.2	509.1	275.5
7	10000	3	102.9	102.9	532.5	302.4
8	100000	3	103.0	102.9	526.2	303.2

Comments

- The true value μ of the mean of the control variable Y must be available.
- Using the sample mean does **not** work.
- The higher the correlation between X and Y, the better.

Outline

Anthitetic draws

Control variates

Other techniques

Beyond the mean

Variance reductions techniques

Other variance reduction techniques

- Conditioning — analytically remove randomness
- Stratified sampling — enforce balanced exploration
- Importance sampling — focus on rare but important events
- Draw recycling — reuse randomness across scenarios

Takeaway

All these methods exploit structure or correlation to reduce variance more efficiently than brute-force simulation.

In general

Correlation helps!

Outline

Scope of variance reduction

What variance reduction guarantees

Variance reduction methods are designed to preserve $\mathbb{E}[X]$ and to reduce the variance of estimators of the form

$$\hat{\theta}_R = \frac{1}{R} \sum_{r=1}^{R} X_r.$$

Implicit limitation

These guarantees rely on:

- linearity of expectation,
- estimators based on sample averages.

Other indicators than the mean

Examples

In practice, we often want to estimate:

- quantiles (e.g. median, 95% percentile),
- probabilities (e.g. $P(X > c)$),
- extrema (maxima, minima),
- risk measures.

Key difference

These indicators are nonlinear functions of the distribution of X.

Pitfall 1: bias can be introduced

Key fact

Variance reduction preserves $\mathbb{E}[X]$, but does not preserve

$$\mathbb{E}[T(\hat{F}_R)]$$

for a nonlinear statistic T.

Reason

In general,

$$\mathbb{E}[T(\hat{F}_R)] \neq T(\mathbb{E}[\hat{F}_R]).$$

Expectation does not commute with nonlinear transformations.

Consequence

A variance-reduced estimator of a quantile or probability may be **biased**.

Pitfall 2: variance may not decrease

Mean vs nonlinear indicators

Reducing $\mathrm{Var}(X_r)$ does not imply that the variance of a nonlinear estimator (e.g. a quantile) is reduced.

Why

- Variance reduction reshapes the empirical distribution.
- Nonlinear indicators depend on order, tails, or ranks.
- Improved balance in one region may worsen accuracy elsewhere.

Conclusion

There is **no universal variance reduction guarantee** for indicators other than the mean.

What remains valid

Important clarification

Variance reduction methods:

- do **not** invalidate simulation,
- do **not** break independence across runs,
- do **not** change the target distribution.

What changes

They modify the finite-sample behavior of estimators built from the simulated outputs.

Practical recommendations (1)

Always ask first

- What is the indicator of interest?
- Is it linear (mean) or nonlinear (quantile, probability)?

Rule of thumb

- Mean $\Rightarrow$ variance reduction is theoretically safe.
- Other indicators $\Rightarrow$ proceed with caution.

Practical recommendations (2)

No closed-form guarantee

For nonlinear indicators, analytical variance or bias formulas are rarely available.

Recommended approach

- Use variance reduction as a candidate improvement.
- Assess its effect empirically:
  - variance,
  - bias,
  - mean square error.

Tool

Bootstrap resampling provides a natural assessment framework.

Final takeaway

Big picture

- Variance reduction is a powerful tool for estimating means.
- For other indicators, it is neither universally good nor bad.
- Its impact must be evaluated, not assumed.

Good practice

$$\text{Variance reduction} + \text{bootstrap assessment}$$