

Unit 1: Monte Carlo Foundations

Lecture 3

The Monte Carlo estimator and its error

The target is a fixed expectation

$$X \sim p, \quad f : \mathbb{R}^d \rightarrow \mathbb{R}, \quad I = \mathbb{E}_p[f(X)].$$

Define $Y = f(X)$. Then $I = \mathbb{E}[Y]$.

Known ingredients

The distribution p and the deterministic function f .

The simulated sample, and hence the estimator, will be random.

Computational task

Evaluate I without an analytical integral or an observed dataset.

Sampling the output through the input

Assume we can draw $X_1, \dots, X_N \stackrel{\text{iid}}{\sim} p$. Set

$$Y_i = f(X_i), \quad i = 1, \dots, N.$$

- Each Y_i has the same distribution as $Y = f(X)$.
- Applying a fixed function separately preserves independence.
- We can simulate Y without an explicit density for Y .

Note

Sampling from p is an assumption for now. Later lectures explain how to construct suitable samplers.

Example: squaring uniform draws

For $U \sim U(0, 1)$, let $Y = U^2$.

Draw i	1	2	3	4
Illustrative u_i	0.20	0.70	0.40	0.90
Output $y_i = u_i^2$	0.04	0.49	0.16	0.81

$$\bar{y} = 0.375, \quad (\bar{u})^2 = 0.3025.$$

Pause and discuss

Which calculation estimates $\mathbb{E}[U^2]$?

The basic Monte Carlo algorithm

- 1 Draw $X_1, \dots, X_N \stackrel{\text{iid}}{\sim} p$.
- 2 Evaluate $Y_i = f(X_i)$ for each draw.
- 3 Return

$$\hat{I}_N = \bar{Y}_N = \frac{1}{N} \sum_{i=1}^N f(X_i).$$

Note

Monte Carlo integration applies sample-mean estimation to simulated output.

Bias and variance account for MSE

$$\text{MSE}(\hat{I}_N) = \mathbb{E}[(\hat{I}_N - I)^2] = \text{Var}(\hat{I}_N) + \text{Bias}(\hat{I}_N)^2.$$

$$\text{Bias}(\hat{I}_N) = \mathbb{E}[\hat{I}_N] - I.$$

- I is fixed throughout repeated simulations.
- The expectation averages over possible simulated samples.
- We will calculate the bias first, then the variance.

Unbiasedness: the full calculation

Assume $\mathbb{E}|f(X)| < \infty$. By linearity,

$$\begin{aligned}\mathbb{E}[\hat{I}_N] &= \mathbb{E}\left[\frac{1}{N} \sum_{i=1}^N f(X_i)\right] \\ &= \frac{1}{N} \sum_{i=1}^N \mathbb{E}[f(X_i)] \\ &= \frac{1}{N} \sum_{i=1}^N I = I.\end{aligned}$$

Hence $\text{Bias}(\hat{I}_N) = 0$.

What unbiasedness tells us

- The average estimate over repeated simulations equals I .
- Independence was not used in the expectation calculation.
- A particular estimate can still be far from I .

With a finite second moment,

$$\text{MSE}(\hat{I}_N) = \text{Var}(\hat{I}_N).$$

Pause and discuss

Could an unbiased estimator have a very large MSE?

Variance: the constant must be squared

Write $Y_i = f(X_i)$. Then

$$\begin{aligned}\text{Var}(\hat{I}_N) &= \text{Var}\left(\frac{1}{N} \sum_{i=1}^N Y_i\right) \\ &= \frac{1}{N^2} \text{Var}\left(\sum_{i=1}^N Y_i\right).\end{aligned}$$

Scaling rule

$\text{Var}(aZ) = a^2 \text{Var}(Z)$, whereas $\mathbb{E}[aZ] = a\mathbb{E}[Z]$.

Variance: where independence enters

For finite second moments,

$$\text{Var} \left(\sum_{i=1}^N Y_i \right) = \sum_{i=1}^N \text{Var}(Y_i) + 2 \sum_{i < j} \text{Cov}(Y_i, Y_j).$$

Independent inputs give independent outputs, so

$$\text{Cov}(Y_i, Y_j) = 0 \quad (i \neq j).$$

Note

Variance adds here because the covariance terms vanish. Pairwise uncorrelatedness would also suffice.

Variance: identical distributions finish the proof

Let $\sigma_f^2 = \text{Var}(f(X)) < \infty$. Each Y_i has variance σ_f^2 .

$$\begin{aligned}\text{Var}(\hat{I}_N) &= \frac{1}{N^2} \sum_{i=1}^N \text{Var}(Y_i) \\ &= \frac{1}{N^2} N \sigma_f^2 \\ &= \boxed{\frac{\sigma_f^2}{N}}.\end{aligned}$$

The numerator is the variance of $f(X)$, which can differ from $\text{Var}(X)$.

Worked example: the square integral

Let $U \sim U(0, 1)$ and $Y = U^2$. Then

$$I = \int_0^1 u^2 \, du = \frac{1}{3}, \quad \mathbb{E}[Y^2] = \mathbb{E}[U^4] = \frac{1}{5}.$$

$$\sigma_f^2 = \frac{1}{5} - \left(\frac{1}{3}\right)^2 = \frac{4}{45}.$$

Therefore

$$\text{MSE}(\hat{I}_N) = \frac{4}{45N}, \quad \text{RMSE}(\hat{I}_N) = \sqrt{\frac{4}{45N}}.$$

Exercise: interpreting an error formula

Exercise

For the square integral, $\sigma_f^2 = 4/45$.

- 1 Compute the RMSE for $N = 1000$.
- 2 Find the RMSE when $N = 4000$.
- 3 Does the second run necessarily have a smaller absolute error?
- 4 Is the output SD also divided by two?

Work for two minutes, then compare with a neighbour.

Solution: repeated-sampling accuracy

$$\text{RMSE}_{1000} = \sqrt{\frac{4}{45000}} \approx 0.00943, \quad \text{RMSE}_{4000} \approx 0.00471.$$

- Four times the samples halves the RMSE.
- A particular run can still have a larger absolute error.
- The output SD remains $\sqrt{4/45} \approx 0.2981$.

Note

Increasing N stabilises the average without changing the output distribution.

The square-root law is an exact RMSE formula

For iid Monte Carlo with finite variance,

$$\text{RMSE}(\hat{I}_N) = \frac{\sigma_f}{\sqrt{N}}.$$

Desired reduction in RMSE	Required multiplier of N
2	4
5	25
10	100

This identity does not require a normal approximation. The CLT explains the approximate shape of the error distribution.

Choosing N for an RMSE target

To obtain $\text{RMSE} \leq \varepsilon$, we need

$$N \geq \frac{\sigma_f^2}{\varepsilon^2}.$$

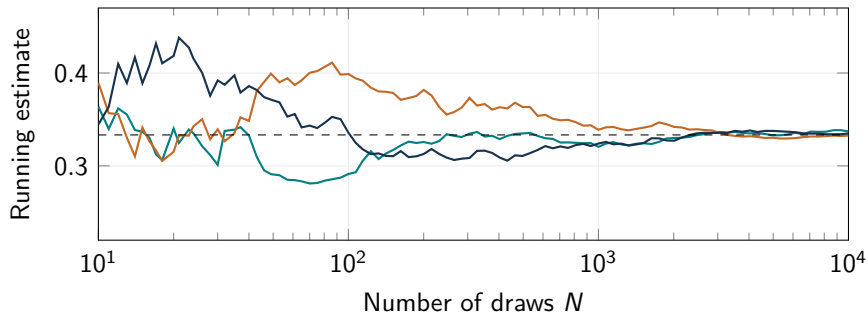
For $\sigma_f^2 = 4/45$ and $\varepsilon = 10^{-3}$,

$$N \geq 88888.\bar{8}, \quad N_{\min} = 88889.$$

RMSE and coverage are different targets

An RMSE of 10^{-3} is not a guarantee that a particular absolute error is below 10^{-3} , nor a 95% error bound of that width.

A running estimate can move away from the target



Three independent runs for $\mathbb{E}[U^2]$, $U \sim U(0, 1)$. Dashed line: $1/3$.

Note

Convergence does not mean that the absolute error decreases at every step.

Student question: only means?

Many targets are expectations or functions of expectations:

$$\mathbb{P}(X \in A) = \mathbb{E}[\mathbf{1}_{\{X \in A\}}], \quad m_k = \mathbb{E}[X^k].$$

For finite second moment,

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

An extra estimation step can matter

If the mean is unknown, substituting its estimate can introduce bias. Quantiles generally require estimating and inverting a distribution function.

Introducing a distribution for an integral

For $a < b$, take $U \sim U(a, b)$, whose density is $1/(b - a)$.

$$J = \int_a^b g(x) \, dx = (b - a) \mathbb{E}[g(U)].$$

The Monte Carlo estimator is

$$\hat{J}_N = \frac{b - a}{N} \sum_{i=1}^N g(U_i).$$

Pause and discuss

Where does the interval-length factor come from?

Exercise: an interval with length two

Exercise

Estimate $J = \int_0^2 x^2 dx$ using $U_i \stackrel{\text{iid}}{\sim} U(0, 2)$.

- 1 Write an unbiased Monte Carlo estimator.
- 2 Find J analytically.
- 3 For $Y = 2U^2$, calculate $\text{Var}(Y)$.
- 4 State the RMSE after N draws.

Hint: $\mathbb{E}[U^k] = 2^k/(k+1)$.

Solution: the scaling also affects variance

$$\hat{J}_N = \frac{2}{N} \sum_i U_i^2, \quad J = \frac{8}{3}.$$

$$\text{Var}(U^2) = \frac{16}{5} - \left(\frac{4}{3}\right)^2 = \frac{64}{45}.$$

Since $Y = 2U^2$,

$$\text{Var}(Y) = 4 \text{Var}(U^2) = \frac{256}{45}, \quad \text{RMSE}(\hat{J}_N) = \frac{16}{\sqrt{45N}}.$$

Both the estimator and its uncertainty must include the interval scaling.

More accuracy requires more computation

Suppose drawing X and evaluating $f(X)$ costs c time units.

$$T \approx Nc, \quad \text{RMSE} = \frac{\sigma_f}{\sqrt{N}}.$$

RMSE reduction	Sample multiplier	Time if current run takes 1 hour
2	4	4 hours
10	100	100 hours

A single evaluation of a complex simulation model may itself be expensive.

The motivation for variance reduction

For the current iid estimator,

$$\text{MSE} = \frac{\sigma_f^2}{N}.$$

- Increasing N costs more sampling and function evaluations.
- A different unbiased estimator can have smaller variance.
- The target I must remain the same.
- Compare methods at a similar total computational budget.

Note

Can we gain accuracy by changing the estimator rather than only increasing the sample size?

Checkpoint

- ① Equal distributions are enough to prove unbiasedness. True or false?
- ② Variance always distributes over a sum. True or false?
- ③ The $N^{-1/2}$ RMSE formula requires normal output. True or false?
- ④ An unbiased estimate has zero error. True or false?

Discussion

State the missing qualification or give a counterexample.

Lecture 3 summary

$$\hat{I}_N = \frac{1}{N} \sum_i f(X_i), \quad \mathbb{E}[\hat{I}_N] = I.$$

For iid draws and finite output variance,

$$\text{MSE}(\hat{I}_N) = \frac{\sigma_f^2}{N}, \quad \text{RMSE}(\hat{I}_N) = \frac{\sigma_f}{\sqrt{N}}.$$

Next lecture

Estimate the unknown variance, report simulation uncertainty, and implement a reproducible calculation.