Unit 1: Monte Carlo Foundations

Lecture 3

The Monte Carlo estimator and its error

Simulation, unbiasedness and the square-root law

Learning outcomes

- Generate samples of $Y = f(X)$ by transforming simulated inputs.
- Derive the bias, variance and MSE of the iid sample mean.
- Distinguish the N^{-1} MSE rate from the $N^{-1/2}$ RMSE rate.
- Express integrals and moments as expectations.
- Explain the computational motivation for variance reduction.

The target is a fixed expectation

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

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

Known ingredients

The distribution p and the deterministic function f.

Computational task

Evaluate I without an analytical integral or an observed dataset.

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

Sampling the output through the input

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

$$Y_i = f(X_i), \qquad i = 1, \ldots, N.$$

- 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.

Key point

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

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

Pause and discuss

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

The basic Monte Carlo algorithm

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

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

Key point

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

Assumptions behind the calculations

Result	Conditions used here
Well-defined integrable output	f measurable and $\mathbb{E}\|f(X)\| < \infty$
Unbiased sample mean	Correct marginal distribution for each X_i
Variance σ_f^2/N	Iid draws and $\sigma_f^2 = \mathrm{Var}(f(X)) < \infty$
Normal approximation later	Iid draws and $0 < \sigma_f^2 < \infty$

The variance calculation needs stronger assumptions than unbiasedness.

Bias and variance account for MSE

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

- 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,

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

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

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,

$$\boxed{\mathrm{MSE}(\widehat{I}_N) = \mathrm{Var}(\widehat{I}_N).}$$

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

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

Scaling rule

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

Variance: where independence enters

For finite second moments,

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

Independent inputs give independent outputs, so

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

Key point

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

Variance: identical distributions finish the proof

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

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

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

MSE and RMSE have different rates

Quantity	Exact formula	Dependence on N
Bias	0	Zero
Variance and MSE	σ_f^2/N	N^{-1}
SD and RMSE	$\sigma_f/\sqrt{N}$	$N^{-1/2}$

Key point

The square-root law follows directly from the exact variance calculation. It does not require a central limit theorem.

Worked example: the square integral

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

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

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

Therefore

$$\mathrm{MSE}(\widehat{I}_N) = \frac{4}{45N}, \qquad \mathrm{RMSE}(\widehat{I}_N) = \sqrt{\frac{4}{45N}}.$$

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

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

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

Key point

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,

$$\mathrm{RMSE}(\widehat{I}_N) = \frac{\sigma_f}{\sqrt{N}}.$$

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 RMSE $\leq \varepsilon$, we need

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

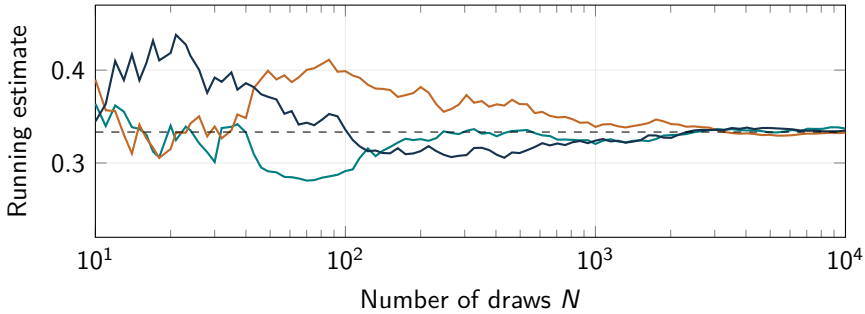
For $\sigma_f^2 = 4/45$ and $\varepsilon = 10^{-3}$,

$$N \geq 88888.\overline{8}, \qquad N_{\min} = 88889.$$

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: $I = 1/3$.

Key point

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

Student question: only means?

Many targets are expectations or functions of expectations:

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

For finite second moment,

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

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 $1/(b-a)$.

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

The Monte Carlo estimator is

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

Pause and discuss

Where does the interval-length factor come from?

Exercise: an interval with length two

Exercise

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

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

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

Solution: the scaling also affects variance

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

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

Since $Y = 2U^2$,

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

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.

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

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,

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

- 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.

Key point

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

Checkpoint: which assumptions matter?

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

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

For iid draws and finite output variance,

$$\mathrm{MSE}(\widehat{I}_N) = \frac{\sigma_f^2}{N}, \qquad \mathrm{RMSE}(\widehat{I}_N) = \frac{\sigma_f}{\sqrt{N}}.$$

Next lecture

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

Optional extensions: convergence and dimension

- Historical context for the name Monte Carlo.
- Laws of large numbers and finite-sample probability bounds.
- The role of dimension and heavy tails.
- Further practice with accuracy and computational cost.

These slides are outside the core 50-minute plan.

Historical context

- The modern computer-based method developed at Los Alamos around 1946–1947.
- Ulam, von Neumann, Metropolis and collaborators developed the approach for problems including neutron transport.
- Metropolis proposed the name Monte Carlo, with its association with gambling.
- Statistical sampling ideas predate these developments.

N. Metropolis (1987), *The Beginning of the Monte Carlo Method*, Los Alamos Science, pp. 125–130.

The law of large numbers

If $Y_1, Y_2, \ldots$ are iid and $\mathbb{E}[|Y_1|] < \infty$, then

$$\widehat{I}_N = \bar{Y}_N \longrightarrow I \quad \text{almost surely as } N \to \infty.$$

In particular, for every $\varepsilon > 0$,

$$\mathbb{P}(|\widehat{I}_N - I| > \varepsilon) \longrightarrow 0.$$

This latter property is called **consistency**.

Key point

The law of large numbers establishes eventual convergence. It does not specify a useful finite sample size.

Chebyshev's inequality gives a finite-sample bound

For an unbiased estimator with variance σ_f^2/N,

$$\mathbb{P}(|\widehat{I}_N - I| \geq \varepsilon) \leq \frac{\sigma_f^2}{N\varepsilon^2}.$$

To make this upper bound at most α, it is sufficient to choose

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

- This is a guarantee under the stated assumptions.
- It can be conservative because it uses only a variance.

Worked example: a conservative sample size

For $Y = X^2$ with $X \sim U(0,1)$, $\sigma_f^2 = 4/45$. We want

$$\mathbb{P}(|\widehat{I}_N - 1/3| \geq 0.01) \leq 0.05.$$

Chebyshev gives the sufficient requirement

$$N \geq \frac{4/45}{0.05(0.01)^2} = 17777.\overline{7}.$$

Thus $N = 17778$ suffices for this bound.

Pause and discuss

Does the calculation prove that every smaller N fails?

Grid cost and dimension

A tensor grid with m points per coordinate contains m^d points.

Dimension d	Grid points for $m = 10$
1	10
3	10^3
6	10^6
10	10^{10}

Monte Carlo retains the formal RMSE dependence $N^{-1/2}$. In low dimensions, smooth integrands can favour deterministic quadrature.

Dimension still affects the variance constant

Let $X_1, \ldots, X_d$ be independent $U(0,1)$ variables.

$$f_d(X) = \sum_{j=1}^{d} X_j, \qquad \mathrm{Var}(f_d(X)) = \frac{d}{12},$$

$$g_d(X) = \frac{1}{d}\sum_{j=1}^{d} X_j, \qquad \mathrm{Var}(g_d(X)) = \frac{1}{12d}.$$

For N independent vectors, the respective RMSEs are

$$\sqrt{\frac{d}{12N}} \quad \text{and} \quad \sqrt{\frac{1}{12dN}}.$$

Key point

The dependence on dimension depends on the target. Sampling and evaluation costs can also grow with d.

A dimension-dependent rare event

For independent $X_j \sim U(0,1)$, consider

$$q_d = \mathbb{P}(X_1 \le 1/2, \ldots, X_d \le 1/2) = 2^{-d}.$$

The crude Monte Carlo relative SD is

$$\sqrt{\frac{1-q_d}{Nq_d}} \approx \sqrt{\frac{2^d}{N}}.$$

To maintain fixed relative accuracy, the required N grows roughly like 2^d.

A limit of the dimension-free rate statement

The $N^{-1/2}$ exponent does not imply dimension-independent difficulty.

Finite mean with infinite variance

Let $U \sim U(0,1)$ and $Y = U^{-3/4}$.

$$\mathbb{E}[Y] = \int_0^1 u^{-3/4}\,\mathrm{d}u = 4, \qquad \mathbb{E}[Y^2] = \int_0^1 u^{-3/2}\,\mathrm{d}u = \infty.$$

- The iid sample mean is unbiased and the law of large numbers applies.
- Its MSE is infinite for every finite N.
- The usual finite-variance CLT and $\sigma_f/\sqrt{N}$ formula do not apply.

Occasional draws very near zero can dominate the average.

Independent practice: moments and cost

Exercise

1. For $Y = U^{-a}$, $U \sim U(0,1)$ and $a > 0$, find when its mean and variance are finite.
2. An estimator has RMSE 0.02 after 5 seconds. Under the same iid cost model, how long is needed for RMSE 0.005?
3. Method B halves the variance per draw but costs three times as much as A. Compare them at equal large budgets.

Independent practice: answers

1. $\mathbb{E}[Y] = 1/(1-a)$ for $0 < a < 1$. The second moment is $1/(1-2a)$ for $0 < a < 1/2$. Thus the mean is finite for $a < 1$ and variance is finite for $a < 1/2$.
2. Reducing RMSE by four needs sixteen times the work: 80 seconds.
3. B has variance–cost product $0.5 \times 3 = 1.5$ times that of A. At equal budget, B has approximately 50% larger estimator variance.

Key point

More accurate individual samples can still yield a less efficient overall estimator.