

Computational Statistics

Section 3.1

Introduction to importance sampling

Recap: direct Monte Carlo

The target is a fixed expectation

$$I = \mathbb{E}_p[f(X)] = \int f(x)p(x) \, dx.$$

With $X_1, \dots, X_N \stackrel{\text{iid}}{\sim} p$,

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

For finite output variance,

$$\text{MSE}(\hat{I}_{\text{MC}}) = \frac{\text{Var}_p(f(X))}{N}.$$

The cost of increasing the sample size

For the same estimator and output distribution,

Accuracy target	Sample multiplier	Cost multiplier*
Divide variance by 10	10	10
Divide RMSE by 10	100	100

*Assuming constant cost per draw and function evaluation.

Pause and discuss

Can a different estimator reduce MSE while using the same number of expensive evaluations of f ?

The target density and the proposal density

Object	Meaning	Role
$p(x)$	Target density	Part of the specified problem
$f(x)$	Function of interest	Defines the desired expectation
$q(x)$	Proposal density	Chosen for the simulation

We keep $I = \int fp$ fixed while changing how we sample.

Note

A proposal must be a valid density that we can sample from and use to evaluate the required weights.

The proposal must cover the contributing region

A standard sufficient support condition is

$$p(x) > 0 \implies q(x) > 0 \text{ almost everywhere.}$$

For a particular f , the weaker condition

$$f(x)p(x) = 0 \text{ almost everywhere on } \{q(x) = 0\}$$

is sufficient for the expectation identity.

Why support matters

No weight can recover a contribution from a region that the proposal never samples.

Rewriting the same integral

On the region where $q(x) > 0$,

$$\begin{aligned} I &= \int f(x)p(x) \, dx \\ &= \int_{q>0} f(x) \frac{p(x)}{q(x)} q(x) \, dx. \end{aligned}$$

The support condition ensures that the omitted region contributes zero.

Pause and discuss

Which factor will play the role of the sampling density?

The importance sampling identity

Define the importance weight and weighted output

$$w(x) = \frac{p(x)}{q(x)}, \quad h(x) = f(x)w(x).$$

If $Z \sim q$, then

$$I = \mathbb{E}_p[f(X)] = \mathbb{E}_q[h(Z)] = \mathbb{E}_q[f(Z)w(Z)].$$

Note

The expectation has the same value even though both the sampling law and the output function change.

The importance sampling algorithm

- 1 Choose a valid proposal q and draw $Z_1, \dots, Z_N \stackrel{\text{iid}}{\sim} q$.
- 2 Evaluate $H_i = f(Z_i)p(Z_i)/q(Z_i)$.
- 3 Return

$$\hat{h}_{\text{IS}} = \frac{1}{N} \sum_{i=1}^N H_i.$$

Ordinary importance sampling

The estimator averages weighted outputs. It does not divide by the observed sum of weights.

Direct sampling and importance sampling

	Direct Monte Carlo	Importance sampling
Draws	$X_i \sim p$	$Z_i \sim q$
Output averaged	$f(X_i)$	$f(Z_i)p(Z_i)/q(Z_i)$
Expectation	I	I
Choice to make	Sample size	Proposal and sample size

Importance sampling is itself a Monte Carlo method.

Unbiasedness: averaging the weighted outputs

By linearity of expectation,

$$\begin{aligned}\mathbb{E}[\hat{I}_{\text{IS}}] &= \frac{1}{N} \sum_{i=1}^N \mathbb{E}_q[H_i] \\ &= \frac{1}{N} \sum_{i=1}^N I = I.\end{aligned}$$

Therefore $\text{Bias}(\hat{I}_{\text{IS}}) = 0$.

Note

Both methods are unbiased. When their variances are finite, their MSE comparison is a variance comparison.

Variance: scaling and independence

Assume $H_1, \dots, H_N$ are iid with finite variance.

$$\begin{aligned}\text{Var}(\hat{l}_{\text{IS}}) &= \text{Var}\left(\frac{1}{N} \sum_i H_i\right) \\ &= \frac{1}{N^2} \text{Var}\left(\sum_i H_i\right) \\ &= \frac{1}{N^2} \sum_i \text{Var}(H_i).\end{aligned}$$

The factor $1/N$ is squared. Independence removes covariance terms.

Variance: identical distributions

Define

$$\sigma_q^2 = \text{Var}_q \left(f(Z) \frac{p(Z)}{q(Z)} \right).$$

Then

$$\text{Var}(\hat{l}_{\text{IS}}) = \frac{N\sigma_q^2}{N^2} = \boxed{\frac{\sigma_q^2}{N}}.$$

$$\text{MSE}(\hat{l}_{\text{IS}}) = \frac{\sigma_q^2}{N}, \quad \text{RMSE}(\hat{l}_{\text{IS}}) = \frac{\sigma_q}{\sqrt{N}}.$$

Proposal choice changes the constant in the square-root law.

The second-moment expression in the notes

Since $\mathbb{E}_q[f w] = I$,

$$\begin{aligned}\sigma_q^2 &= \mathbb{E}_q[f(Z)^2 w(Z)^2] - I^2 \\ &= \int_{q>0} \frac{f(x)^2 p(x)^2}{q(x)} dx - I^2 \\ &= \mathbb{E}_p[f(X)^2 w(X)] - I^2.\end{aligned}$$

Thus the estimator variance is σ_q^2/N .

Finite-variance requirement

$$\int_{q>0} f(x)^2 p(x)^2 / q(x) dx < \infty.$$

No automatic variance ordering

At the same N , compare

$$\frac{\text{Var}_p(f(X))}{N} \quad \text{and} \quad \frac{\text{Var}_q(f(Z)w(Z))}{N}.$$

- A suitable proposal can reduce the numerator.
- Another proposal can leave it unchanged or increase it.
- A poor proposal can make the weighted second moment infinite.

Pause and discuss

Why does the extra factor w alone not determine which variance is smaller?

The special case $q = p$

If $q(x) = p(x)$ on the target support, then

$$w(x) = 1, \quad H_i = f(Z_i), \quad Z_i \sim p.$$

Hence importance sampling recovers direct Monte Carlo:

$$\hat{I}_{\text{IS}} = \frac{1}{N} \sum_i f(Z_i).$$

Note

Changing the name of the estimator does not improve accuracy. The proposal must provide the improvement.

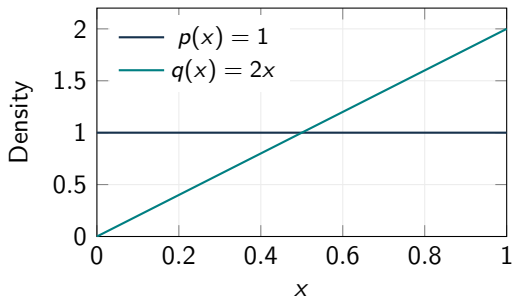
Worked example: the square integral

$$I = \int_0^1 x^2 dx = \frac{1}{3}.$$

Target: $p(x) = 1$ on $(0, 1)$.

Proposal: $q(x) = 2x$ on $(0, 1)$.

Both integrate to one. The proposal draws larger inputs more often.



Sampling and weighting in the example

The proposal CDF is $Q(z) = z^2$ for $0 \leq z \leq 1$. If $U \sim U(0, 1)$, then $Z = \sqrt{U}$ has density $q(z) = 2z$.

$$w(z) = \frac{1}{2z}, \quad H = f(Z)w(Z) = \frac{Z^2}{2Z} = \frac{Z}{2}.$$

Therefore

$$\hat{l}_{\text{IS}} = \frac{1}{2N} \sum_i Z_i.$$

The simulation requires one uniform draw per output.

Exercise: does this proposal reduce variance?

For $Z \sim q(z) = 2z$ on $(0, 1)$, the output is $H = Z/2$.

Exercise

- 1 Calculate $\mathbb{E}_q[Z]$ and $\mathbb{E}_q[Z^2]$.
- 2 Verify that $\mathbb{E}_q[H] = 1/3$.
- 3 Calculate $\text{Var}_q(H)$.
- 4 Compare it with $\text{Var}_p(X^2) = 4/45$ for $X \sim U(0, 1)$.

Hint: $\mathbb{E}_q[Z^k] = \int_0^1 z^k (2z) dz$. Allow three minutes.

Solution: the proposal reduces the output variance

$$\mathbb{E}_q[Z] = \frac{2}{3}, \quad \mathbb{E}_q[Z^2] = \frac{1}{2}.$$
$$\mathbb{E}_q[H] = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}, \quad \text{Var}_q(H) = \frac{1}{4} \left(\frac{1}{2} - \frac{4}{9} \right) = \frac{1}{72}.$$

Thus

$$\text{Var}(\hat{l}_{\text{IS}}) = \frac{1}{72N} < \frac{4}{45N} = \text{Var}(\hat{l}_{\text{MC}}).$$

The target expectation remains unchanged.

How large is the improvement?

Method	Variance per output	RMSE at $N = 1000$
Direct Monte Carlo	4/45	0.00943
Importance sampling	1/72	0.00373

$$\frac{\text{Var}(\hat{l}_{\text{MC}})}{\text{Var}(\hat{l}_{\text{IS}})} = 6.4, \quad \frac{\text{RMSE}_{\text{MC}}}{\text{RMSE}_{\text{IS}}} = \sqrt{6.4} \approx 2.53.$$

Direct Monte Carlo needs 6.4 times as many draws for the same variance.

Failure through missing support

Keep $p(x) = 1$ and $f(x) = x^2$ on $(0, 1)$, but choose

$$q(x) = 2\mathbf{1}_{(1/2, 1)}(x).$$

The weighted expectation is only

$$\mathbb{E}_q[f(Z)w(Z)] = \int_{1/2}^1 x^2 dx = \frac{7}{24}.$$

The target is $1/3 = 8/24$, so the estimator's bias is $-1/24$.

The problem persists as N grows

Increasing the sample size cannot recover the missing interval.

Failure through infinite variance

Now choose $q(x) = 6x^5$ on $(0, 1)$, which covers the support.

$$H = \frac{Z^2}{6Z^5} = \frac{1}{6Z^3}, \quad \mathbb{E}_q[H] = \frac{1}{3}.$$

But

$$\mathbb{E}_q[H^2] = \int_0^1 \frac{x^4}{6x^5} dx = \frac{1}{6} \int_0^1 \frac{dx}{x} = \infty.$$

Note

The estimator is unbiased, yet has infinite MSE for every finite N . Rare small draws can dominate the average.

Efficiency at equal computational cost

If each weighted output costs c_q and the budget is T , then

$$N \approx T/c_q, \quad \text{Var}(\hat{I}_{\text{IS}}) \approx \frac{\sigma_q^2 c_q}{T}.$$

For the square-integral example, importance sampling wins at equal large budget when

$$\frac{c_q}{72} < \frac{4c_p}{45} \iff \frac{c_q}{c_p} < 6.4.$$

Include sampling, density evaluation and function evaluation costs.

Names for the proposal distribution

The same q may be called an

- importance sampling distribution or importance distribution;
- proposal distribution;
- biasing distribution;
- alternative distribution.

Biasing the sampling law

We deliberately change the distribution of the draws. Correct weighting still gives an unbiased estimator of I .

Checkpoint: diagnosing an importance sampler

- ① We draw from q and average $f(Z_i)$. What do we estimate?
- ② We draw from q but use weights $q(Z_i)/p(Z_i)$. Is the identity valid?
- ③ We choose $q = p$. What accuracy improvement follows?
- ④ The proposal covers the support. Is finite variance guaranteed?

Pause and discuss

Explain each answer using the expectation or second-moment integral.

Summary

$$I = \mathbb{E}_p[f(X)] = \mathbb{E}_q\left[f(Z)\frac{p(Z)}{q(Z)}\right].$$

With valid support, integrability and iid proposal draws,

$$\mathbb{E}[\hat{I}_{\text{IS}}] = I, \quad \text{Var}(\hat{I}_{\text{IS}}) = \frac{\text{Var}_q(f(Z)w(Z))}{N}$$

when the weighted second moment is finite.

Next question

How should we choose a proposal that places samples where they are most useful?