

Computational Statistics

"Optimal" importance sampling

Recap: the importance sampling estimator

$$I = \int f(x)p(x) \, dx = \mathbb{E}_q[f(Z)w(Z)], \quad w(x) = \frac{p(x)}{q(x)}.$$

With $Z_1, \dots, Z_N \stackrel{\text{iid}}{\sim} q$,

$$\hat{l}_q = \frac{1}{N} \sum_i f(Z_i)w(Z_i), \quad \mathbb{E}[\hat{l}_q] = I.$$

The proposal covers the contributing support.

Note

We change the sampling law while keeping the target I fixed.

What does optimal mean?

For finite weighted second moment,

$$\text{Var}(\hat{I}_q) = \frac{\sigma_q^2}{N}, \quad \sigma_q^2 = \text{Var}_q(f(Z)w(Z)).$$

At fixed N , the ideal design problem is

$$\min_{q \text{ admissible}} \sigma_q^2.$$

- Admissible proposals are densities covering the contributing region.
- A restricted family may not contain the unrestricted optimum.
- Variance optimality at fixed N does not include computational cost.

The variance identity keeps two squares distinct

For any square-integrable random variable V ,

$$\text{Var}(V) = \underbrace{\mathbb{E}[V^2]}_{\text{expectation of the square}} - \underbrace{(\mathbb{E}[V])^2}_{\text{square of the expectation}}.$$

Taking $H = f(Z)w(Z)$ with $Z \sim q$,

$$\sigma_q^2 = \mathbb{E}_q[f(Z)^2 w(Z)^2] - I^2.$$

The second term is fixed because $\mathbb{E}_q[H] = I$ for every valid proposal.

The second moment as an integral

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

The equivalent expectation under p

Since $w = p/q$,

$$\int_{q>0} \frac{f^2 p^2}{q} dx = \int_{q>0} f^2 w p dx = \mathbb{E}_p[f(X)^2 w(X)].$$

Hence equation (3.1) in the notes gives

$$\text{Var}(\hat{l}_q) = \frac{1}{N} \left(\mathbb{E}_p[f(X)^2 w(X)] - l^2 \right).$$

The estimator still uses samples from q . This is an algebraic rewriting of its variance.

The nonnegative case

For the zero-variance construction, assume

$$f(x) \geq 0, \quad 0 < I = \int f(x)p(x) \, dx < \infty.$$

- Nonnegativity makes fp a possible density shape.
- Finite positive I allows us to normalise that shape.
- If $I = 0$, then $f(X) = 0$ almost surely under p .

The sign assumption matters

A function with both positive and negative contributions generally has a positive minimum variance.

The candidate optimal proposal

Choose a density proportional to the nonnegative integrand:

$$q^*(x) \propto f(x)p(x).$$

Its normalised form is

$$q^*(x) = \frac{f(x)p(x)}{\int f(u)p(u) \, du} = \frac{f(x)p(x)}{I}.$$

Pause and discuss

Why is the normalising constant I , rather than $\int f$ or $\int p$?

Verification that q^* is a density

Both density requirements hold:

$$q^*(x) \geq 0,$$
$$\int q^*(x) dx = \frac{1}{I} \int f(x)p(x) dx = \frac{I}{I} = 1.$$

Note

The numerator determines the shape. The denominator fixes the total probability mass.

Support and zeros of the integrand

If $f(x) = 0$, the ideal density can satisfy $q^*(x) = 0$ even when $p(x) > 0$.

The condition needed for this expectation is

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

An event indicator

If $f = \mathbf{1}_A$, the ideal proposal places all its mass inside A . Points outside A contribute zero to the target.

The weighted output becomes constant

Where $q^*(x) > 0$,

$$w^*(x) = \frac{p(x)}{q^*(x)} = \frac{I}{f(x)}.$$

Therefore, for $Z \sim q^*$,

$$f(Z)w^*(Z) = I \quad \text{almost surely.}$$

Note

The input remains random, but the weighted output is the same for every draw.

The full variance substitution

$$\begin{aligned}\sigma_{q^*}^2 &= \int_{q^* > 0} \frac{f(x)^2 p(x)^2}{q^*(x)} dx - I^2 \\ &= \int_{q^* > 0} \frac{f(x)^2 p(x)^2}{f(x)p(x)/I} dx - I^2 \\ &= I \int_{q^* > 0} f(x)p(x) dx - I^2 \\ &= I^2 - I^2 = 0.\end{aligned}$$

The region where $q^* = 0$ contributes zero to $\int fp$.

Optimality and the one-sample result

Variance is nonnegative, so $\sigma_{q^*}^2 = 0$ is optimal. For any positive integer N ,

$$\text{Var}(\hat{I}_{q^*}) = \frac{0}{N} = 0, \quad \text{MSE}(\hat{I}_{q^*}) = 0.$$

Thus $\hat{I}_{q^*} = I$ almost surely, even with $N = 1$.

An ideal mathematical statement

This assumes exact sampling and evaluation. If direct Monte Carlo already has zero variance, the optimum gives no strict improvement.

Exercise: constructing a zero-variance proposal

Let $p(x) = 1$ for $0 < x < 1$ and $f(x) = \sqrt{x}$.

Exercise

- 1 Calculate $I = \int_0^1 \sqrt{x} \, dx$.
- 2 Write $q^*(x)$ and check that it integrates to one.
- 3 Find $w^*(x)$ and the weighted output.
- 4 Compare the variance with direct Monte Carlo.

Work for three minutes. For direct Monte Carlo, use $\mathbb{E}_p[f(X)^2] = 1/2$.

Solution: the normalising constant cancels the randomness

$$I = \frac{2}{3}, \quad q^*(x) = \frac{3}{2}\sqrt{x}, \quad \int_0^1 q^*(x) dx = 1.$$

$$w^*(x) = \frac{2}{3\sqrt{x}}, \quad f(x)w^*(x) = \frac{2}{3}.$$

$$\text{Var}(\hat{I}_{\text{MC}}) = \frac{1}{N} \left(\frac{1}{2} - \frac{4}{9} \right) = \frac{1}{18N}, \quad \text{Var}(\hat{I}_{q^*}) = 0.$$

Why the ideal proposal is usually unavailable

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

- The normalising constant is the unknown expectation I .
- The exact ordinary importance weight therefore contains the answer.
- Knowing only the unnormalised shape fp does not supply that weight.

Note

The ideal density characterises the best possible variance, but generally does not directly compute an unknown integral.

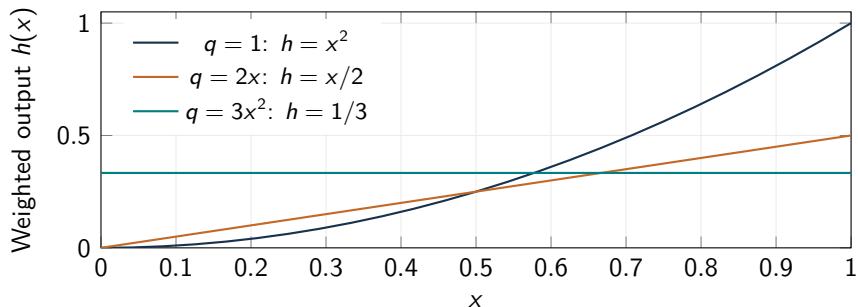
A familiar benchmark: three choices of proposal

For $p(x) = 1$, $f(x) = x^2$ on $(0, 1)$, $I = 1/3$.

Proposal	$q(x)$	Weighted output $h(x)$	Output variance
Direct sampling	1	x^2	4/45
Useful alternative	$2x$	$x/2$	1/72
Ideal	$3x^2$	1/3	0

The ideal proposal makes fp/q constant. A practical proposal aims to reduce its variation.

How the proposal changes the weighted output



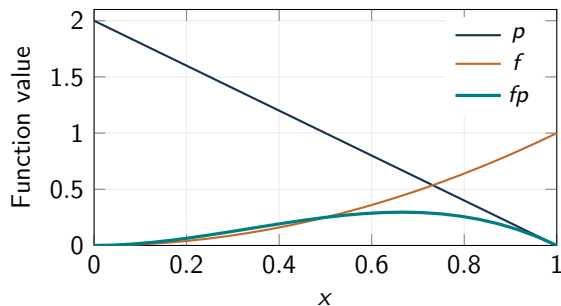
Each curve has mean $1/3$ under its own sampling density. The flat curve has zero variance.

The shape of the product matters

Consider $0 < x < 1$:

$$f(x) = x^2, \quad p(x) = 2(1 - x).$$

f increases and p decreases. Their product concentrates its contribution in the interior.



The curves f and fp are not normalised densities.

Exercise: normalising the product

For $f(x) = x^2$ and $p(x) = 2(1 - x)$ on $(0, 1)$:

Exercise

- 1 Calculate $I = \int_0^1 2x^2(1 - x) dx$.
- 2 Find the ideal density $q^*(x)$.
- 3 Find where q^* attains its maximum.
- 4 Show that the weighted output is constant.

Work for three minutes. Compare the maximiser with those of f and p .

Solution: an interior concentration of probability

$$I = 2 \left(\frac{1}{3} - \frac{1}{4} \right) = \frac{1}{6}, \quad q^*(x) = 12x^2(1 - x).$$

$$\frac{d}{dx} q^*(x) = 12x(2 - 3x), \quad x_{\text{mode}} = \frac{2}{3}.$$

$$w^*(x) = \frac{2(1 - x)}{12x^2(1 - x)} = \frac{1}{6x^2}, \quad f(x)w^*(x) = \frac{1}{6}.$$

The ideal proposal reflects both the function and the target density.

Two approaches to practical proposal design

Problem-specific design	Systematic fitting
Use knowledge of the model and contributing regions.	Choose a family and fit its parameters using an objective.
Exploit a suitable change of location, scale or shape.	Estimate the objective and solve a numerical fitting problem.
Depends on understanding the problem.	Still requires assumptions, a representation and computational choices.

What a practical proposal must achieve

- Cover every region where fp contributes to the target.
- Permit sampling and evaluation of the required density ratio.
- Keep $\int f^2 p^2 / q$ finite and reasonably small.
- Justify its sampling, weighting and construction costs.

Guidance from the ideal

For $f \geq 0$, approximate the shape of fp . A visual match alone does not guarantee a small weighted second moment.

Checkpoint: interpreting the ideal proposal

- ① Does zero variance mean the input draw is deterministic?
- ② Can q^* be zero somewhere that p is positive?
- ③ If we can sample using the unnormalised shape fp , do we automatically know the exact weight?
- ④ Does every problem admit a strict improvement over direct sampling?

Discuss for two minutes before checking the answers.

Checkpoint answers

- ① No. The input can be random while $f(Z)w^*(Z) = I$ is constant.
- ② Yes, where $fp = 0$. Such a region contributes nothing to I .
- ③ No. The normalising constant may still be unknown, and the ordinary weight requires it.
- ④ No. Direct sampling already has zero variance when f is constant under p .

Note

The ideal proposal is a statement about the weighted output and its variance.

Summary

For $f \geq 0$ and $0 < I < \infty$,

$$q^* = \frac{fp}{I}, \quad fw^* = I, \quad \text{Var}(\hat{I}_{q^*}) = 0.$$

- The unknown normalising constant usually prevents direct use.
- The shape of fp guides practical proposals.
- Signed functions require the $|f|_p$ qualification.