Monte Carlo Estimation and the Limits of Importance Sampling

Maria Alba Fernández Bonet
Facultat de Matemàtiques i Estadística
Universitat Politècnica de Catalunya
Barcelona, Spain
maria.alba.fernandez@estudiantat.upc.edu

Sebastián Serra Peña
Facultat de Matemàtiques i Estadística
Universitat Politècnica de Catalunya
Barcelona, Spain
sserrape@estudiantat.upc.edu

***Abstract*—This mini-project explores Monte Carlo estimation methods through a sequence of experiments. Starting with direct Monte Carlo using i.i.d. samples from the standard normal distribution, we verify convergence properties (bias, variance, MSE) for estimators of means and tail probabilities. We then introduce Importance Sampling (IS) to improve efficiency when the target distribution is a bimodal mixture, evaluating both unnormalized and self-normalized IS estimators. We analyze the sensitivity of IS performance to the proposal's parameters and demonstrate how poor overlap leads to high estimator variance. Finally, we show that as the target distribution's dimension increases, IS weights become severely imbalanced (degenerate), the effective sample size collapses, and estimator MSE grows rapidly. These findings illustrate the fundamental challenges of IS in high-dimensional spaces and highlight the necessity of careful proposal design or adaptive alternatives.**

***Index Terms*—Monte Carlo Methods, Importance Sampling, Mean Squared Error, Effective Sample Size, Curse of Dimensionality**

I. Raw Monte Carlo Estimation and Characterization

We begin our project of Monte Carlo methods with the most fundamental approach: direct (or "raw") Monte Carlo estimation, using i.i.d. samples drawn from the target distribution. In this first experiment, we draw $N = 20$ samples from the standard normal distribution $\mathcal{N}(0, 1)$ and focus on estimating two functionals of interest:

- The expectation $\mathbb{E}[X]$ (i.e., $f(x) = x$).
- The tail probability $\Pr(X > \gamma)$ with $\gamma = 3$ (i.e., $f(x) = \mathbb{I}_{|x|>\gamma}(x)$).

To characterize the performance of the estimators, we replicate the Monte Carlo procedure independently 10^4 times and compute the empirical bias, variance, and mean squared error (MSE) as functions of the sample size $n \in 1, \ldots, 500$:

$$\text{Bias}_n = \mathbb{E}[\hat{I}_n] - I,\ \text{Var}_n \quad = \mathbb{V}[\hat{I}_n],\ \text{MSE}_n = \text{Bias}_n^2 + \text{Var}_n.$$

Figure 1 presents the evolution of these metrics for both quantities of interest.

As expected, the Monte Carlo estimators exhibit convergence towards the true values of the functionals:

- The empirical bias decreases and stabilizes around zero and 0.0013 respectively as n grows, in agreement with the unbiasedness of the sample mean estimator for $\mathbb{E}[X]$ and the indicator-based estimator for $\Pr(X > \gamma)$.
- The empirical variance decreases at the canonical $\mathcal{O}(1/n)$ rate, confirming the standard convergence property of Monte Carlo estimators.
- The MSE, as the sum of squared bias and variance, is dominated by variance and thus also decays with $1/n$.

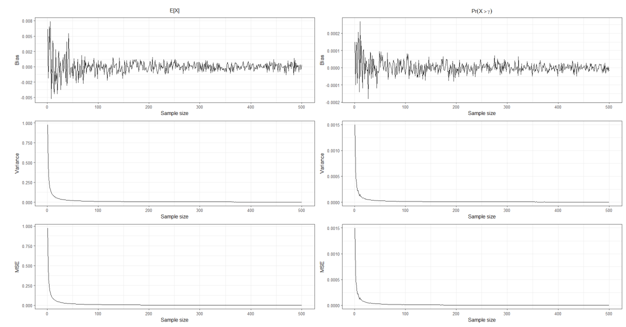

Fig. 1. Empirical bias (top), variance (middle), and mean squared error (bottom) of Monte Carlo estimators for $\mathbb{E}[X]$ (left) and $\Pr(X > 3)$ (right) as functions of the sample size. Results based on 10^4 independent replications.

This experiment illustrates the fundamental properties of Monte Carlo estimation: consistency, unbiasedness (under suitable conditions), and the $\mathcal{O}(1/\sqrt{n})$ rate of convergence. It also reveals practical limitations when estimating rare event probabilities (e.g., $\Pr(X > 3) \approx 0.00135$), where the estimator remains highly variable for small n.

These insights motivate the need for variance reduction techniques when estimating low-probability or high-dimensional events—a problem that raw Monte Carlo struggles with. In the following section, we introduce *importance sampling*, a powerful strategy to tackle such inefficiencies by changing the sampling distribution while re-weighting appropriately.

II. Importance Sampling for a Bimodal Target Distribution

To highlight the potential of importance sampling (IS), we consider a more challenging scenario where the target distribution is bimodal:

$$p(x) = \tfrac{1}{2}\mathcal{N}(-3, 1) + \tfrac{1}{2}\mathcal{N}(5, 4).$$

This mixture exhibits multiple modes and heavy tails, posing significant difficulties for naive sampling strategies. We aim to estimate the following expectations under the target distribution:

- $\mathbb{E}_p[X]$ (mean),
- $\mathbb{E}_p[X^2]$ (second moment),
- $\Pr_p(X > \gamma)$ with $\gamma = 2$.

We employ importance sampling using two distinct proposals $q(x)$:

1) **Good proposal:** $\mathcal{N}(0, 25)$ — a wide Gaussian encompassing both modes.
2) **Bad proposal:** $\mathcal{N}(-3, 25)$ — a wide Gaussian centered near one mode.

Figure 2 illustrates the overlap between $p(x)$ and each proposal, as well as the importance weights $w(x) = p(x)/q(x)$ for samples $x \sim q$.

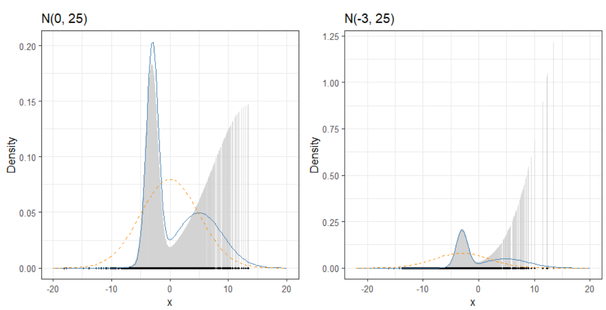

Fig. 2. Target distribution $p(x)$ (blue), proposal $q(x)$ (orange dashed), and scaled importance weights (gray vertical lines) for two proposals: $\mathcal{N}(0, 25)$ (left) and $\mathcal{N}(-3, 25)$ (right).

The first proposal better covers both modes of the target, while the second leads to high-variance importance weights due to insufficient overlap in the right tail.

We study two versions of the IS estimator:

- The *unnormalized IS estimator (UIS)*:

$$\hat{I}^{\text{UIS}} = \frac{1}{n}\sum_{i=1}^{n} f(x_i)\frac{p(x_i)}{q(x_i)}$$

- The *self-normalized IS estimator (SNIS)*:

$$\hat{I}^{\text{SNIS}} = \frac{\sum_{i=1}^{n} f(x_i) w_i}{\sum_{i=1}^{n} w_i}, \quad \text{with } w_i = \frac{p(x_i)}{q(x_i)}$$

For each estimator and each quantity ($\mathbb{E}[X]$, $\mathbb{E}[X^2]$, $\Pr(X > \gamma)$), we simulate 10^4 independent replications for each sample size $n = 1, 2, \ldots, 500$, with $N = 20$ samples per estimation. We then compute the empirical bias, variance, and MSE across replications with the good proposal distribution.

From Figure 3, we observe:

- **Bias:** Both UIS and SNIS estimators are asymptotically unbiased. SNIS tends to exhibit slightly larger bias at small n due to the normalization denominator introducing nonlinearity.
- **Variance:** UIS can suffer from large variance if the importance weights are highly variable, especially in cases where $q(x)$ fails to cover high-weight regions. SNIS reduces this variance but at the cost of introducing bias.
- **MSE:** SNIS achieves consistently lower MSE in small to moderate sample sizes, especially when estimating $\Pr(X > 2)$—a rare event located in the tail of the distribution.

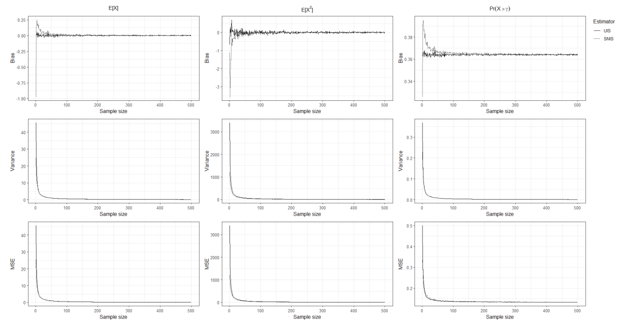

Fig. 3. Empirical bias (top), variance (middle), and MSE (bottom) of UIS and SNIS estimators for $\mathbb{E}[X]$ (left), $\mathbb{E}[X^2]$ (center), and $\Pr(X > 2)$ (right) across sample sizes.

This comparison demonstrates the delicate balance between bias and variance in IS estimators. To better understand this sensitivity, we next explore the effect of small perturbations in the proposal's mean and variance. Even slight misalignment between the proposal $q(x)$ and the target $p(x)$ can substantially increase the variance of the weights $w(x) = p(x)/q(x)$, thereby degrading estimator efficiency. This motivates further study of robustness and proposal tuning strategies before we escalate to the even more challenging high-dimensional case.

Effect of Proposal Perturbations on IS Performance

To complement our evaluation of UIS and SNIS, we now investigate how changes in the *proposal distribution's mean and variance* affect the quality of the SNIS estimator. This sensitivity analysis is essential, as IS efficiency strongly depends on the overlap between the proposal $q(x)$ and the target $p(x)$.

For this experiment, we fix the number of samples at $n = 20$ and explore a grid of proposals:

- Means: $\mu_q \in \{-3, 1, 4\}$,
- Standard deviations: $\sigma_q \in \{2, 5, 8\}$.

For each configuration, we estimate $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, and $\Pr(X > 2)$ using SNIS. Each estimator is empirically characterized through 10^4 repetitions, and the performance is assessed via bias, variance, and mean squared error (MSE).

From Table I, several key patterns emerge:

- The highlighted case ($\mu_q = 1, \sigma_q = 5$) yields the lowest MSE for $\mathbb{E}[X]$, demonstrating the optimality of aligning the proposal's center and spread with those of the target. This confirms theoretical expectations: good coverage of both modes and tails stabilizes the importance weights.

Performance Metrics for all estimators (SNIS, n = 20)

Estimator	Proposal Mean	Proposal SD	Bias	Variance	MSE
E(X)	-3	2	-3.200	1.223	11.465
		5	-0.400	4.873	5.033
		8	0.127	2.630	2.646
	1	2	-0.868	4.448	5.202
		5	0.148	2.240	2.262
		8	0.271	2.380	2.454
	4	2	3.113	3.944	13.634
		5	0.497	3.280	3.527
		8	0.399	2.841	3.000
E(X^2)	-3	2	-16.598	2.649	278.141
		5	-3.594	342.434	355.354
		8	0.472	98.608	98.831
	1	2	-13.360	71.466	249.943
		5	0.330	104.918	105.027
		8	1.102	69.403	70.618
	4	2	1.783	259.455	262.633
		5	1.831	77.272	80.626
		8	1.790	69.950	73.155
P(X > γ)	-3	2	0.038	0.031	0.033
		5	0.335	0.035	0.147
		8	0.376	0.025	0.167
	1	2	0.309	0.047	0.143
		5	0.379	0.019	0.163
		8	0.389	0.023	0.175
	4	2	0.710	0.048	0.553
		5	0.411	0.030	0.199
		8	0.401	0.027	0.188

TABLE I

- For $\mathbb{E}[X^2]$, although the same configuration is still competitive, the overall error remains much higher across all proposals, especially in variance. This indicates higher-order moments are more sensitive to weight variability.
- Estimation of $\Pr(X > 2)$ is generally more robust. Even moderate misalignment in $q(x)$ does not drastically degrade MSE, likely due to the event being less extreme under this bimodal target compared to previous normal cases.
- Notably, proposals with small standard deviation (e.g., $\sigma_q = 2$) lead to dramatically inflated variance and MSE, especially when the proposal mean is far from the target's effective support. This reflects the high kurtosis and poor coverage of $q(x)$ in these settings.

In conclusion, the SNIS estimator is highly sensitive to the proposal distribution's parameters. Ensuring good coverage of the target—particularly capturing both modes and enough tail mass—is essential for controlling variance and achieving low MSE. This motivates not only heuristic tuning of $q(x)$ but also adaptive strategies that adjust the proposal based on sample feedback.

In the next section, we explore how these challenges amplify in higher dimensions, where poor overlap becomes even more likely, and adaptive methods such as population Monte Carlo and effective sample size monitoring become crucial.

III. IS in Higher Dimensions: Curse of Dimensionality

In this final section, we investigate how Importance Sampling (IS) deteriorates as the dimension of the target distribution increases. While IS can perform well in low dimensions when the proposal distribution is reasonably aligned with the target, this alignment becomes increasingly difficult to maintain in high-dimensional spaces.

We consider a factorized target distribution $\pi(x)$ given by a product of d independent components, where each marginal follows a bimodal distribution:

$$\pi(x) = \prod_{j=1}^{d} \left[\tfrac{1}{2}\mathcal{N}(-3,1) + \tfrac{1}{2}\mathcal{N}(5,4)\right].$$

The proposal distribution $q(x)$ is chosen to be a diagonal Gaussian with independent marginals $q_j(x_j) = \mathcal{N}(1, \sqrt{24.5})$ for all $j = 1, \ldots, d$. This symmetric and over-dispersed form reasonably overlaps the target in low dimensions, but fails to remain effective as d grows.

To estimate:

- $\mathbb{E}[X_1]$
- $\mathbb{E}[X_1^2]$
- $\Pr(X_1 > \gamma)$ with $\gamma = 2$

we generate $n = 20$ samples per run from $q(x)$, simulating dimension-by-dimension (possible due to the diagonal structure). For each sample $x^{(i)}$, the IS weight is computed as:

$$w^{(i)} = \frac{\pi(x^{(i)})}{q(x^{(i)})} = \prod_{j=1}^{d} \frac{\pi_j(x_j^{(i)})}{q_j(x^{(i)}j)},$$

and normalized as:

$$\tilde{w}^{(i)} = \frac{w^{(i)}}{\sum j = 1^n w^{(j)}}.$$

Weight Degeneracy and Effective Sample Size

Figure 4 shows the sorted normalized IS weights for a single simulation at $d = 5$. Ideally, the weights should be close to uniform: $\tilde{w}^{(i)} \approx 1/n = 0.05$. However, the plot reveals:

- The largest weight is over 0.33.
- Most weights are nearly zero.

This indicates that only a few samples dominate the estimation — a phenomenon known as *weight degeneracy*. This occurs when the proposal $q(x)$ fails to place sufficient mass in regions where $\pi(x)$ is large. In high dimensions, the product of many slightly misaligned marginal densities leads to exponentially small overlap.

The Effective Sample Size (ESS) can be used to quantify this phenomenon:

$$\widehat{ESS} = \frac{1}{\sum_{i=1}^{n} (\tilde{w}^{(i)})^2},$$

which evaluates to n when all weights are equal, and tends toward 1 when one sample dominates. In our example, ESS is approximately 2–3, implying that although we use 20 samples, only a handful contribute meaningfully.

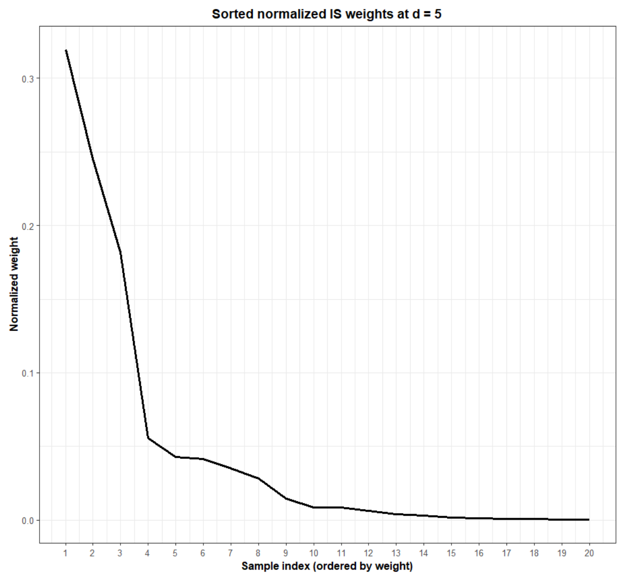

Fig. 4. Sorted normalized IS weights at dimension $d = 5$.

Mean Squared Error as a Function of Dimension

To assess performance, we run 10^4 simulations at increasing dimension $d \in \{1, 2, 3, 4, 5\}$ and compute the empirical Mean Squared Error (MSE) for all three estimators. Results are displayed in Figure 5.

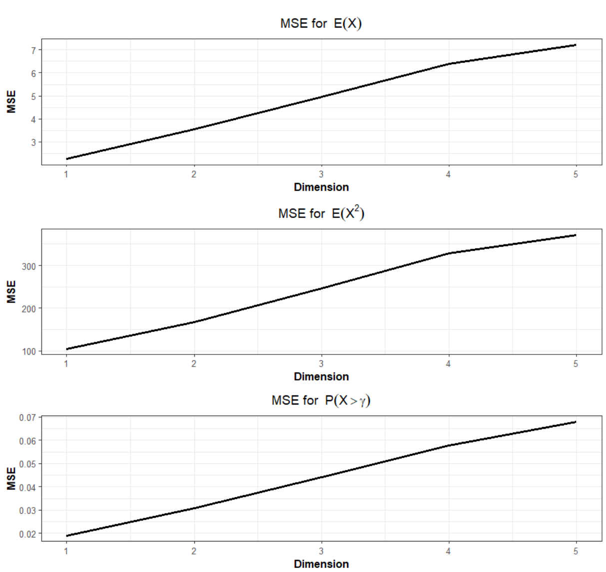

Fig. 5. Empirical MSE of IS estimators for $\mathbb{E}[X_1]$, $\mathbb{E}[X_1^2]$, and $\Pr(X_1 > 2)$ as a function of target dimension d.

We observe that:

- All MSEs increase monotonically with dimension.
- The growth is especially steep for $\mathbb{E}[X_1^2]$, where variance accumulates multiplicatively.
- The MSE for $\Pr(X_1 > 2)$ grows more moderately but still significantly, showing that even tail event estimation becomes unreliable.

This confirms the **curse of dimensionality** in IS:

> As d increases, the probability that any of the proposal samples fall into regions of high target density decays rapidly. IS weights become highly imbalanced, leading to unstable, inefficient, and biased estimates.

In practical terms, this means that classical IS with naive proposals becomes computationally infeasible for even moderate dimensions. Adaptive techniques, variance reduction methods, or alternative samplers must be used to mitigate this collapse in efficiency.

The code implementation can be found in the GitHub repository [1], and some parts are inspired by prior material from [2] and the lecture slides in [3].

IV. Conclusions

This mini-project has explored Monte Carlo estimation methods across a range of progressively difficult settings. We began with raw Monte Carlo, which, while consistent and unbiased, becomes inefficient for estimating rare events or when direct sampling from the target is difficult. Importance Sampling (IS) improves estimation by shifting the sampling distribution but is extremely sensitive to the choice of proposal.

Self-normalized IS (SNIS) partially alleviates the instability caused by large weight variance and can outperform unnormalized IS in small-sample regimes. However, even minor deviations in the proposal's mean or variance can severely degrade performance. Our sensitivity analysis confirmed that optimal proposals must align closely with the target's support and tail behavior.

In high-dimensional settings, IS deteriorates rapidly. The effective sample size collapses as importance weights become concentrated on a few samples, leading to unstable and inefficient estimators. This phenomenon illustrates the curse of dimensionality: in such regimes, traditional IS is no longer viable without adaptation or more advanced sampling strategies.

Overall, while IS can significantly reduce variance when well-configured, its practical success hinges on careful proposal tuning, especially in complex or high-dimensional problems.

References

[1] M. A. Fernández Bonet and S. Serra Peña, "Monte-carlo-methods," GitHub repository, 2025, accessed on July 2025. [Online]. Available: https://github.com/SSerraPe/Monte-Carlo-Methods.git

[2] A. Garbuno, "simulacion," GitHub repository, 2023, accessed on July 2025. [Online]. Available: https://github.com/agarbuno/simulacion.git

[3] V. Elvira, "Monte carlo methods for bayesian inference and beyond (l2)," 2025, phD course on Bayesian filtering and Monte Carlo methods.