Week #4: Variance

Stochastic Methods Course

Notes by: Francisco Richter

05.03.2025

Monte Carlo Methods and Variance

Monte Carlo methods are a powerful class of numerical techniques that approximate solutions to complex mathematical problems by leveraging random sampling. They are especially useful when deterministic or closed-form methods either do not exist or are prohibitively difficult to apply.

Estimating Integrals

A canonical example is estimating the integral of a real-valued function $f(x)$ over $[a, b]$:

$$I = \int_a^b f(x)\,\mathrm{d}x.$$

A fundamental Monte Carlo estimator proceeds by:

1. Drawing independent samples $x_1, x_2, \ldots, x_N$ from the uniform distribution on $[a, b]$.
2. Taking the sample mean of f at these points.

Since $X \sim \text{Uniform}(a, b)$ has density

$$p(x) = \frac{1}{b-a}, \quad x \in [a, b],$$

we observe that

$$\mathbb{E}[f(X)] = \frac{1}{b-a}\int_a^b f(x)\,\mathrm{d}x = \frac{I}{b-a}.$$

Hence, multiplying by $(b-a)$ yields:

$$I = (b-a)\,\mathbb{E}[f(X)].$$

An unbiased estimator for I is then:

$$\hat{I} = (b-a)\,\frac{1}{N}\sum_{i=1}^{N} f(x_i).$$

By the *Law of Large Numbers*, $\hat{I} \to I$ almost surely as $N \to \infty$.

Definition. Variance

For a random variable X with mean $\mu = \mathbb{E}[X]$, the *variance* is given by

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

Since $\hat{I}$ is an average of i.i.d. random variables, we may also invoke the *Central Limit Theorem* (CLT). The CLT tells us that for large N,

$$\hat{I} \;\approx\; \mathcal{N}\Big(I,\, \frac{(b-a)^2\,\mathrm{Var}\big(f(X)\big)}{N}\Big),$$

where $X \sim \mathrm{Uniform}(a,b)$. In other words, the standard error of $\hat{I}$ behaves like $\mathcal{O}(N^{-1/2})$. This confirms that Monte Carlo estimators converge relatively slowly compared to, say, deterministic methods for low-dimensional integrals. However, they remain feasible in high-dimensional settings where grid-based or quadrature approaches fail due to the *curse of dimensionality*.

Importance Sampling

In some applications, especially high-dimensional ones, uniform sampling might be extremely inefficient if $f(x)$ is sharply peaked in a small region of $[a,b]$. *Importance sampling* addresses this by drawing points from a *proposal* or *importance* density $p(x)$ that closely resembles the shape of $f(x)$. We then rewrite:

$$I \;=\; \int f(x)\,\mathrm{d}x \;=\; \int \frac{f(x)}{p(x)}\,p(x)\,\mathrm{d}x \;=\; \mathbb{E}_p\Big[\frac{f(X)}{p(X)}\Big].$$

So an alternative unbiased estimator is:

$$\hat{I}_{\mathrm{IS}} \;=\; \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}, \quad x_i \overset{\mathrm{iid}}{\sim} p(x).$$

The variance of $\hat{I}_{\mathrm{IS}}$ is

$$\mathrm{Var}\big(\hat{I}_{\mathrm{IS}}\big) \;=\; \frac{1}{N}\,\mathrm{Var}_p\Big(\frac{f(X)}{p(X)}\Big).$$

A well-chosen $p(x)$ can drastically reduce this variance relative to uniform sampling. In an extreme but instructive case, the *optimal* importance density is $p^*(x) = \frac{|f(x)|}{\int |f(x')|\mathrm{d}x'}$, which, if feasible to sample from, can reduce the integral variance to zero (for an integrable f). In practice, one picks a tractable p approximating f.

1 Rejection Sampling

Sometimes we want to sample from a distribution with density $f(x)$ (normalized or unnormalized) but cannot directly draw from it. *Rejection sampling* overcomes this by comparing f to a *proposal* q for which sampling is straightforward. Suppose we know a constant $c \geq 1$ such that

$$f(x) \;\leq\; c\,q(x) \quad \text{for all } x.$$

The method proceeds:

1. Draw $X \sim q$ and $U \sim \mathrm{Uniform}(0,1)$ independently.
2. Accept X if $U \leq \frac{f(X)}{c\,q(X)}$; otherwise, reject and repeat.

Theorem 1.1 (Correctness of Rejection Sampling)**.** *Let $\{(X,U)\}$ be as above. The conditional distribution of X given that X is accepted coincides with the target distribution f. Specifically,*

$$\mathbb{P}\big(X \in \mathrm{d}x \,\big|\, \textit{accepted}\big) \;=\; \frac{f(x)\,\mathrm{d}x}{\int f(y)\,\mathrm{d}y}.$$

Proof. The event "accept" occurs if

$$U \;\le\; \frac{f(X)}{c\,q(X)}.$$

Because U is uniform(0,1), for each fixed x,

$$\mathbb{P}\big(\text{accept} \mid X = x\big) \;=\; \frac{f(x)}{c\,q(x)}.$$

Hence, the joint density of (X, accept) is

$$q(x)\,\frac{f(x)}{c\,q(x)} \;=\; \frac{f(x)}{c}.$$

Integrating over x gives the acceptance probability

$$\mathbb{P}(\text{accept}) \;=\; \frac{1}{c}\int f(x)\,\mathrm{d}x \;=\; \frac{1}{c}.$$

Thus, conditioned on acceptance, the density is

$$\frac{\frac{f(x)}{c}}{\frac{1}{c}} \;=\; f(x)\Big/\int f(y)\,\mathrm{d}y,$$

as required. □

The efficiency of rejection sampling depends on how tight the bound $f(x) \le c\,q(x)$ is. The probability of acceptance is $1/c$. If c is large, many samples get rejected.

2 Dependence and Independence

Concepts of dependence and independence among random variables are foundational in both probability theory and Monte Carlo methods. Two real-valued random variables X and Y are *independent* if

$$f_{X,Y}(x,y) \;=\; f_X(x)\,f_Y(y),$$

which equivalently means

$$\mathbb{P}\big(X \in A, Y \in B\big) \;=\; \mathbb{P}\big(X \in A\big)\,\mathbb{P}\big(Y \in B\big) \quad \text{for all measurable sets } A, B.$$

Example 2.1. *If X and Y are the results of rolling two fair dice, each taking values in $\{1,\ldots,6\}$, then*

$$p_{X,Y}(i,j) \;=\; \frac{1}{6}\,\frac{1}{6} \;=\; \frac{1}{36},$$

indicating independence. In contrast, a single die roll's value and that value squared would not be independent.

Conditional Probability and Bayes' Theorem

Definition. Conditional Probability

For random variables X and Y with joint density $f_{X,Y}$ and marginal f_Y (assuming $f_Y(y) \neq 0$), the conditional probability density is

$$f_{X|Y}(x \mid y) \;=\; \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

The multiplication rule for densities states

$$f_{X,Y}(x,y) \;=\; f_{X|Y}(x \mid y)\,f_Y(y).$$

Theorem 2.1 (Bayes' Theorem)**.** *For random variables X and Y with $f_Y(y) \neq 0$,*

$$f_{X|Y}(x \mid y) \;=\; \frac{f_{Y|X}(y \mid x)\, f_X(x)}{f_Y(y)}.$$

Proof. From the multiplication rule,

$$f_{X,Y}(x,y) \;=\; f_{X|Y}(x \mid y)\, f_Y(y) \;=\; f_{Y|X}(y \mid x)\, f_X(x).$$

Rearrange to solve for $f_{X|Y}(x \mid y)$:

$$f_{X|Y}(x \mid y) \;=\; \frac{f_{Y|X}(y \mid x)\, f_X(x)}{f_Y(y)}.$$

□

Bayesian inference and many advanced Monte Carlo techniques (like Markov chain Monte Carlo) rely heavily on iterated conditioning and Bayes' theorem. Understanding conditional distributions and independence lays the groundwork for nearly all stochastic simulation methods.

Questions

- **Exercise 1.** Propose an explicit algorithm for simulating a random variable whose density on $(0, 1)$ is:

  $$f(x) \;=\; 30\,(x^2 \;-\; 2x^3 \;+\; x^4) \;=\; 30\,x^2\,(1-x)^2.$$

  (Hint: Factor and consider whether a suitable Beta distribution might match this form.)

- **Exercise 2.** Let $0 \leq X \leq a$ for some constant $a > 0$. Show that:

  (a) $\mathbb{E}[X^2] \;\leq\; a\,\mathbb{E}[X]$.

  (b) $\mathrm{Var}(X) \;\leq\; \mathbb{E}[X]\,\big(a - \mathbb{E}[X]\big)$.

  (c) $\mathrm{Var}(X) \;\leq\; \frac{a^2}{4}$.

- **Exercise 3.** Suppose

  $$\sigma^2 \;=\; \mathbb{E}[X^2] \;=\; \frac{\int_{-\infty}^{\infty} x^2 \exp\{-|x|^3/3\}\,\mathrm{d}x}{\int_{-\infty}^{\infty} \exp\{-|x|^3/3\}\,\mathrm{d}x},$$

  where X has density $q(x) \propto \exp\{-|x|^3/3\}$.

  (a) Estimate σ^2 using *importance sampling* with a suitable proposal distribution and (standardized) weights.

  (b) Repeat the estimation using *rejection sampling*. Analyze the acceptance rate in terms of your chosen proposal.