Week #1: Introduction to Stochastic Methods

Stochastic Methods Course

Notes by: Francisco Richter

18.02.2025

1 Introduction

Probability theory is the mathematical study of uncertainty. It provides the tools for modeling random phenomena, making predictions, and understanding systems influenced by chance.

2 Random Numbers and Their Generators

A **random number** is an unpredictable value generated independently of other numbers, lacking any discernible pattern.

Definition. Random Number Generator

A *Random Number Generator* (RNG) is an algorithm that produces a sequence of numbers that appears random. Formally, an RNG is a function

$$R : S \to T,$$

where:

- S is the *seed space* (a finite set of initial states). A seed $s \in S$ is used to initialize the RNG.
- T is the *target space* (typically the interval $[0, 1)$ or a set of integers).
- The function R maps each seed s to a value $u \in T$ in a manner that appears random.

A high-quality RNG should satisfy the following:

- **Unpredictability:** Future values cannot be deduced without knowing the seed and algorithm.
- **Reproducibility:** The same seed produces the same sequence.
- **True Randomness Representation:** All outcomes have an equitable chance.
- **Long Period:** The sequence takes a long time to repeat.
- **Efficiency:** Numbers are generated quickly.

Example 2.1 (Linear Congruential Generator)**.** *The Linear Congruential Generator (LCG) produces a sequence via:*

$$X_{n+1} = (aX_n + c) \mod m,$$

where X_n is the current number, a is the multiplier, c is the increment, m is the modulus and $X_0 = S$ is the initial seed.

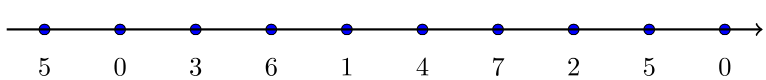

Figure 1: A sequence of values from a Linear Congruential Generator (LCG).

Example 2.2 (PCG64 RNG)**.** *PCG64 is the default RNG in recent versions of NumPy. It combines a 128-bit LCG with an output permutation to yield high-quality 64-bit pseudorandom numbers. The algorithm proceeds in two steps:*

1. ***State Update:*** *The 128-bit state s_n is updated by*

$$s_{n+1} = (s_n \cdot a + c) \mod 2^{128},$$

where a and c are carefully chosen constants.

2. ***Output Permutation:*** *A permutation is applied to the updated state to produce the final output.*

This method ensures uniformity, independence, and an extremely long period.

3 Random Variables and Distributions

Definition. ***Random Variable***

*A **random variable** is a function $X : S \to \mathbb{R}$ that assigns a real number to each outcome in a sample space S. For any subset $C \subset \mathbb{R}$, the probability that X falls in C is*

$$P\{X \in C\} = P(\{s \in S : X(s) \in C\}).$$

Random variables are classified as follows:

1. ***Discrete Random Variables:*** *Take values in a countable set and are characterized by a* Probability Mass Function (PMF)*.*
2. ***Continuous Random Variables:*** *Take values in an uncountable set and are described by a* Probability Density Function (PDF)*.*

Definition. ***Probability Mass Function (PMF)***

For a discrete random variable X taking values in a set $A \subset \mathbb{R}$, the PMF is defined by

$$f(x) = P(X = x), \quad x \in A.$$

For any subset $C \subset A$,

$$P\{X \in C\} = \sum_{x \in C} f(x).$$

Definition. ***Probability Density Function (PDF)***

For a continuous random variable X, a function $f(x)$ is its PDF if, for any interval $[a, b]$,

$$P\{a \leq X \leq b\} = \int_a^b f(x)\, dx,$$

with the normalization condition

$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$

Definition. ***Cumulative Distribution Function (CDF)***

*The **cumulative distribution function** of a random variable X is*

$$F(x) = P\{X \leq x\}.$$

For discrete variables,

$$F(x) = \sum_{t \le x} f(t),$$

and for continuous variables,

$$F(x) = \int_{-\infty}^{x} f(t)\,dt.$$

A common method for generating samples from any given distribution is the Inverse Transform Sampling.

Theorem 3.1 (Inverse Transform Sampling)**.** *Let U be uniformly distributed on $[0,1]$ and let $F_X(x)$ be the CDF of a random variable X with an invertible inverse $F_X^{-1}(u)$. Then, the variable*

$$X = F_X^{-1}(U)$$

has CDF $F_X(x)$.

Proof. For any $x \in \mathbb{R}$,

$$P(X \le x) = P\big(F_X^{-1}(U) \le x\big) = P\big(U \le F_X(x)\big) = F_X(x),$$

since U is uniformly distributed on $[0,1]$ and F_X is strictly increasing. □

4 Exercises

Exercise 1: Random Quadratic.

Let b be uniformly distributed on $(0,1)$. Consider the quadratic equation

$$x^2 + b\,x + 2 = 0.$$

Find the probability that this quadratic has real roots.

Exercise 2: Number of Hearts in a 5-Card Hand.

From a standard 52-card deck, draw 5 cards without replacement. Let

$$X = \textit{number of hearts in the hand.}$$

(a) Determine the probability mass function $p_X(k) = P(X = k)$ for $k = 0, 1, \ldots, 5$.

(b) Determine the cumulative distribution function $F_X(k) = P(X \le k)$.

Exercise 3: Coin Toss Difference.

Flip 4 fair coins. Define

$$X = (\textit{number of heads}) - (\textit{number of tails}).$$

(a) Find the probability mass function $p_X(x)$.

(b) Find the cumulative distribution function $F_X(x)$.

Exercise 4: Counting Uniform$(0,1)$ Observations Below 0.4.

Let $U_1, U_2, \ldots, U_{10}$ be independent random variables uniformly distributed on $(0,1)$. Define

$$X = \textit{number of } U_i \textit{ with } U_i < 0.4.$$

(a) Determine $p_X(k) = P(X = k)$ for $k = 0, 1, \ldots, 10$.

(b) *Determine* $F_X(k) = P(X \leq k)$.

Exercise 5: A Discrete Random Variable.
Suppose X *takes values* $k = 0, 1, 2, \ldots$ *with*

$$P(X = k) = e^{-3}\,\frac{3^k}{k!}.$$

(a) *Show that* $\sum_{k=0}^{\infty} P(X = k) = 1$.

(b) *Determine the cumulative distribution function* $F_X(k) = P(X \leq k)$.

Exercise 6: Distance Between Two Uniform$(0,1)$ Points.
Let Y_1 *and* Y_2 *be independent random variables uniformly distributed on* $(0, 1)$. *Define*

$$X = |\,Y_1 - Y_2\,|.$$

(a) *Derive the probability density function (PDF) and cumulative distribution function (CDF) of* X.

(b) *Verify that* X *takes values in* $[0, 1]$ *and discuss the shape of its distribution.*