Week #1: Randomness

March 10, 2025

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

1 Random Numbers and Their Generators

A *random number* is an unpredictable value, generated independently from preceding or succeeding numbers. It lacks any discernible pattern or regularity, making it impossible to deduce without understanding the underlying random generation process.

Moving from individual random numbers, it is essential to understand how we can generate a series of such numbers, leading us to the concept of a random number generator.

Definition. Random number generator

A *Random Number Generator* (RNG) is an algorithm that produces a sequence of numbers that lacks any pattern, i.e., appears random. More formally, an RNG is defined as a function:

$$R : S \to T$$

where:

- S is the *seed space*, a finite set of initial states. An RNG is typically initialized with a value in S, known as the *seed*.
- T is the *target space*, typically the set of real numbers in the interval $[0, 1)$ or a set of integer values.
- The function R maps each seed $s \in S$ to a target $u \in T$ in a manner that appears random.

A high-quality random number generator should have the following properties:

- **Unpredictability:** Without knowing the algorithm and seed, it should be impossible to predict future numbers.
- **Reproducibility:** Given the same seed, the RNG should produce the same sequence of numbers.
- **Representation of True Randomness:** The RNG should accurately represent true randomness, ensuring an equitable chance for all potential outcomes.

- **Long period:** The sequence of numbers should be long before repeating.
- **Efficiency:** The RNG should generate numbers quickly.

When these properties are fulfilled, the target space can be conceived as an *aleatory universe*.

Example. Linear Congruential Generator

The *linear congruential generator* generates a sequence of random numbers via the following linear recurrence relation:

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

where:

- X_{n+1} is the next number in the sequence.
- X_n is the current number.
- a, c, and m are constants, known as the *multiplier*, *increment*, and *modulus*, respectively.
- $\mod$ denotes the modulus operation.
- The initial or seed value $X_0 = S$ is required to start the sequence.

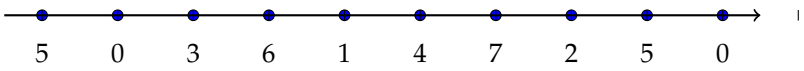

Figure 1: A sequence of realizations from a linear congruential generator (LCG).

Example. PCG64 RNG

PCG64 is the default generator in newer versions of NumPy. It combines a 128-bit linear congruential generator (LCG) with an output permutation to produce high-quality 64-bit pseudorandom numbers. The algorithm proceeds in two main stages:

(1) State Update: PCG64 maintains a 128-bit state s_n. The state is updated via the LCG:

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

where:

- a is a carefully chosen 128-bit multiplier,
- c is a 128-bit odd increment (ensuring a full period).

(2) Output Permutation: Instead of outputting the raw state, PCG64 applies a permutation.

This combination yields high-quality 64-bit pseudorandom numbers that are uniform, independent, and have an extremely long period.

2 Random Variables and Distributions

Definition. Random Variable

A **random variable** is a measurable function that assigns a real number to each outcome in a sample space S:

$$X : S \to \mathbb{R}.$$

For any subset $C \subset \mathbb{R}$, the probability that X takes a value in C is defined by

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

In probability theory, we distinguish between two types of random variables based on how probabilities are assigned to outcomes:

1. Discrete Random Variables: These take values in a countable set. Their distribution is characterized by a *probability mass function (PMF)*, which assigns a probability to each individual value.

Definition. Probability Mass Function (PMF)

Let X be a discrete random variable that takes values in a countable set $A \subset \mathbb{R}$. The **probability mass function** is defined by

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

For any subset $C \subset A$, the probability that X takes a value in C is given by

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

2. Continuous Random Variables: These have uncountably many possible values, and their distribution is described by a *probability density function (PDF)*.

Definition. Probability Density Function (PDF)

Let X be a continuous random variable. If there exists a function $f(x)$ such that for any interval $[a, b]$,

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

then $f(x)$ is called the **probability density function** of X. It must satisfy

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

Regardless of whether X is discrete or continuous, its distribution can be fully described by its cumulative distribution function:

Definition. Cumulative Distribution Function (CDF)

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

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

For a discrete random variable, this can be written as

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

and for a continuous random variable with PDF $f(x)$, it is given by

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

In both cases, $F(x)$ represents the total probability that X does not exceed x.

In particular, if a target distribution has a continuously invertible CDF $F_X(x)$, one can obtain a random variable X with that distribution by setting

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

where U is uniform on $[0, 1]$.

Theorem 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 defined by

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

has CDF $F_X(x)$.

Proof. We show that for any $x \in \mathbb{R}$,

$$\begin{aligned} P\Big(X \leq x\Big) &= P\Big(F_X^{-1}(U) \leq x\Big) \\ &= P\Big(U \leq F_X(x)\Big) \quad \text{(since } F_X \text{ is strictly increasing)} \\ &= F_X(x) \quad \text{(because } U \text{ is uniformly distributed on } [0, 1]). \end{aligned}$$

Thus, $X = F_X^{-1}(U)$ indeed has CDF $F_X(x)$. □

Because every number in $[0, 1]$ is equally likely, the uniform distribution is ideal for generating samples from other distributions via inverse transform sampling.

3 Exercises

Exercise 1: Random Quadratic.

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

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

Find the probability that this quadratic equation has real roots.

Exercise 2: Number of Hearts in a 5-Card Hand.
From a standard 52-card deck, draw 5 cards at random (without replacement). Let

$$X \;=\; \text{the number of hearts in those 5 cards.}$$

(a) Find the probability mass function $p_X(k) = P(X = k)$ for $k = 0, 1, 2, 3, 4, 5$.

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

Exercise 3: Coin Toss Difference.
Flip 4 fair coins. Define

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

(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 ten independent random variables, each uniformly distributed on $(0, 1)$. Define

$$X \;=\; \text{the number of these } U_i \text{ that are less than } 0.4.$$

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

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

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

$$P(X = k) \;=\; e^{-3}\,\frac{3^k}{k!} \quad (k = 0, 1, 2, \ldots).$$

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

(b) Find 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, each uniformly distributed on $(0, 1)$. Define

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

(a) Find the probability distribution of X (both its PDF and its CDF).

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