Week #2: Random Variables

March 10, 2025

Often, we require random numbers in the continuous interval $[0, 1)$. A common technique to achieve this is to normalize the LCG's output. If the LCG produces integers in $\{0, 1, \ldots, m-1\}$, then we define

$$U_{n+1} = \frac{X_{n+1}}{m}.$$

This operation maps each integer to a real number in $[0, 1)$.

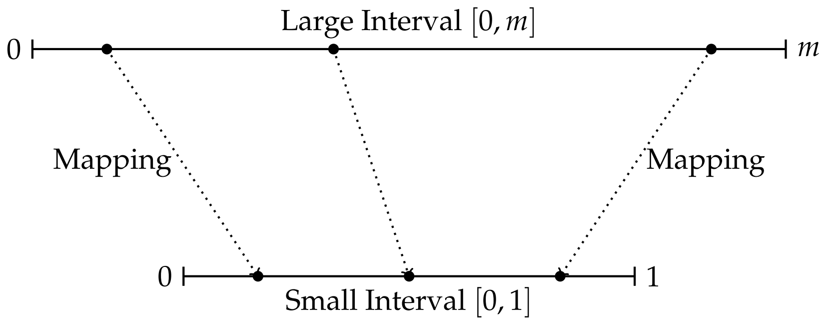

Figure 1: Mapping of random numbers from a large interval to a small interval.

Many experiments have outcomes defined on different sets. For instance, a six-sided die yields

$$U_d = \{1, 2, 3, 4, 5, 6\},$$

and a coin toss yields

$$U_c = \{H, T\}.$$

By running an RNG and then mapping its output (for example, via a modulus operation or parity check), we can sample from these universes. In essence, we first generate a number in $[0, 1)$ and then use a suitable transformation to obtain the desired outcome.

Example. Probability and Frequency

Consider the LCG defined by

$$X_{n+1} = (1664525 \times X_n + 1013904223) \mod 2^{32}.$$

Since the LCG produces numbers uniformly in its range, half of the values lie below 2^{31} and half above. We map values below 2^{31} to `Head` (H) and those above to `Tail` (T). Figure 2 shows this mapping for the first few iterations.

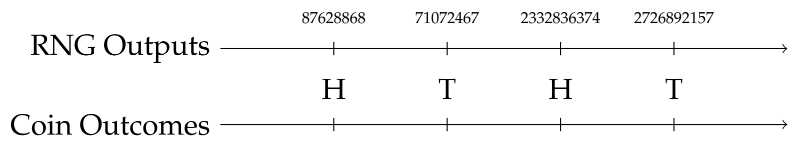

Figure 2: Mapping LCG outputs to coin toss outcomes.

Define the random variable X_n as

$$X_n = \begin{cases} 0, & \text{if the outcome is } H, \\ 1, & \text{if the outcome is } T. \end{cases}$$

Then, as the number of iterations n approaches 2^{32}, we have

$$P(X_n = 0) = P(X_n = 1) = 0.5.$$

Figure 3 provides empirical evidence: as the sample size increases, the frequency of H and T converges to 0.5.

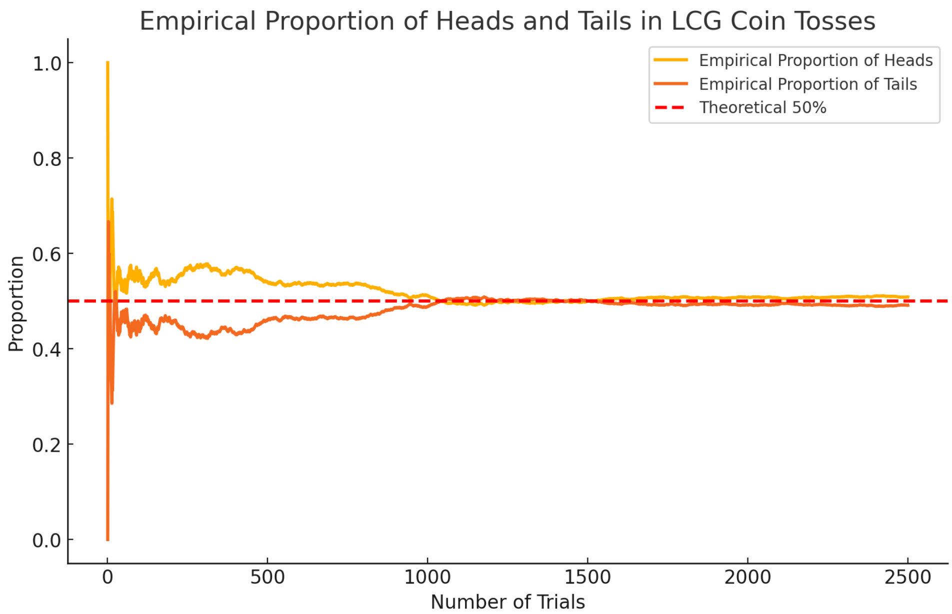

Figure 3: Empirical validation of the head-tail experiment.

To provide a unified framework for probability, we define a standard measure on the interval $[0, 1]$. For any subinterval $[a, b] \subseteq [0, 1]$ with $0 \leq a < b \leq 1$, we assign

$$\mu([a, b]) = b - a.$$

This measure represents the length of the interval and, when normalized so that $\mu([0, 1]) = 1$, it serves as the canonical probability measure on $[0, 1]$. In practice, every set that can be constructed from intervals by countable unions, intersections, and complements (the so-called Borel sets) is measurable with respect to μ.

This construction is universal. Many experiments yield outcomes defined on various sets. For example, a fair die gives

$$U_d = \{1, 2, 3, 4, 5, 6\},$$

and a coin toss gives

$$U_c = \{H, T\}.$$

Often we generate a random number in $[0, 1)$ using an RNG—say, by normalizing an LCG's output—and then map that number to the desired set. More generally, if there is a measurable mapping

$$\phi : U \to [0, 1],$$

we can define a probability measure P_U on U by *pulling back* μ:

$$P_U(B) = \mu\big(\phi(B)\big),$$

for any measurable subset $B \subset U$. This approach transfers the well-understood properties of the Lebesgue measure on $[0, 1]$ to any outcome space U, ensuring that probabilities are assigned consistently.

Definition. Borel Sets on $[0, 1$

] The Borel σ-algebra on $[0, 1]$, denoted by $\mathcal{B}([0, 1])$, is the collection of all sets that can be formed from open intervals in $[0, 1]$ by applying countable unions, countable intersections, and complementation.

Example.

Consider the set of rational numbers in $[0, 1]$. Since the rationals are countable, we can enumerate them as

$$\mathbb{Q} \cap [0, 1] = \bigcup_{n=1}^{\infty} \{q_n\},$$

where each q_n is a rational number in $[0, 1]$. Each singleton $\{q_n\}$ is a closed set in $\mathbb{R}$ and thus a Borel set. Because the Borel σ-algebra is closed under countable unions, the set $\mathbb{Q} \cap [0, 1]$ is a Borel set.

1 Random Variables

With a probability measure P established on our sample space U, we now introduce random variables as functions that assign numerical values to outcomes.

Definition. Random Variable

Let $(U, \mathcal{F}, P)$ be a probability space, where U is the sample space, $\mathcal{F}$ is a collection of measurable subsets of U (for instance, the Borel sets when $U \subset \mathbb{R}$), and P is the probability measure. A *random variable* is a measurable function

$$X : U \to \mathbb{R},$$

meaning that for every Borel set $B \subset \mathbb{R}$, the preimage $X^{-1}(B)$ is in $\mathcal{F}$. The set

$$R(X) = \{X(u) : u \in U\} \subset \mathbb{R}$$

is called the *range* or *image* of X. If $R(X)$ is countable, X is said to be *discrete*.

For discrete random variables, the distribution is fully characterized by the function f_X defined on $R(X)$.

Definition. Probability Mass Function

Let X be a discrete random variable with range $R(X)$. The *probability mass function* (pmf) of X is the function

$$f_X : R(X) \to [0, 1],$$

defined by

$$f_X(x) = P\big(\{u \in U : X(u) = x\}\big), \quad \text{for each } x \in R(X).$$

The function f_X must satisfy:

1. $0 \le f_X(x) \le 1$ for all $x \in R(X)$,
2. $\sum_{x \in R(X)} f_X(x) = 1$.

Example. Discrete Uniform Distribution

Suppose X takes values in the finite set $\{a, a+1, \ldots, b\}$ with equal likelihood. Then, for every $x \in \{a, a+1, \ldots, b\}$,

$$f_X(x) = \frac{1}{b - a + 1}.$$

Example. Bernoulli Distribution

Let X represent the outcome of a binary experiment, defined by

$$X(u) = \begin{cases} 1, & \text{if the outcome is a success,} \\ 0, & \text{if the outcome is a failure.} \end{cases}$$

If $P(X = 1) = p$ and $P(X = 0) = 1 - p$, then the pmf of X is

$$f_X(x) = \begin{cases} p, & \text{if } x = 1, \\ 1 - p, & \text{if } x = 0. \end{cases}$$

Example. Binomial Distribution

If we conduct n independent Bernoulli trials with success probability p and let X denote the number of successes, then X follows a binomial distribution with pmf

$$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{for } k = 0, 1, \ldots, n.$$

Example. Poisson Distribution

A random variable X is said to have a Poisson distribution with parameter $\lambda > 0$ if

$$f_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad \text{for } k = 0, 1, 2, \ldots.$$

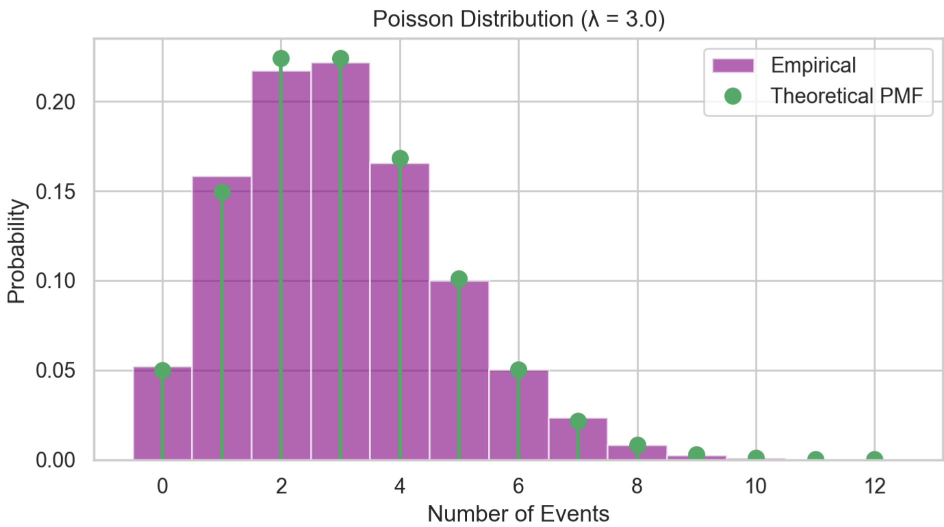

Figure 4: Plot of the Poisson Distribution.

Now imagine partitioning the interval $[0, 1]$ into k equal subintervals, each of width $\Delta x = \frac{1}{k}$. Assign a probability $f_X(x_i)\Delta x$ to the ith subinterval, where x_i is a representative point in that interval. Then, the total probability that X falls within an interval $[a, b]$ is approximated by the Riemann sum

$$\sum_{i=1}^{n} f_X(x_i)\Delta x.$$

As the number of subintervals increases (i.e., as $\Delta x \to 0$), this sum converges to the integral

$$\int_a^b f_X(x)\, dx.$$

Definition. Probability Density Function

For a continuous random variable X with range $R(X) \subseteq \mathbb{R}$, the distribution is described by the *probability density function* (pdf) f_X. This function is nonnegative and satisfies

$$P(a \leq X \leq b) = \int_a^b f_X(x)\,dx,$$

for any interval $[a, b] \subseteq R(X)$, with the normalization

$$\int_{R(X)} f_X(x)\,dx = 1.$$

Example. Exponential Distribution

Let X be a continuous random variable that follows an exponential distribution with rate parameter $\lambda > 0$. Its probability density function (pdf) is given by

$$f_X(x) = \lambda e^{-\lambda x}, \quad \text{for } x \geq 0,$$

and $f_X(x) = 0$ for $x < 0$. The exponential distribution is often used to model waiting times between independent events. Figure 5 shows a plot of the exponential pdf with parameter $\lambda - 0.1$.

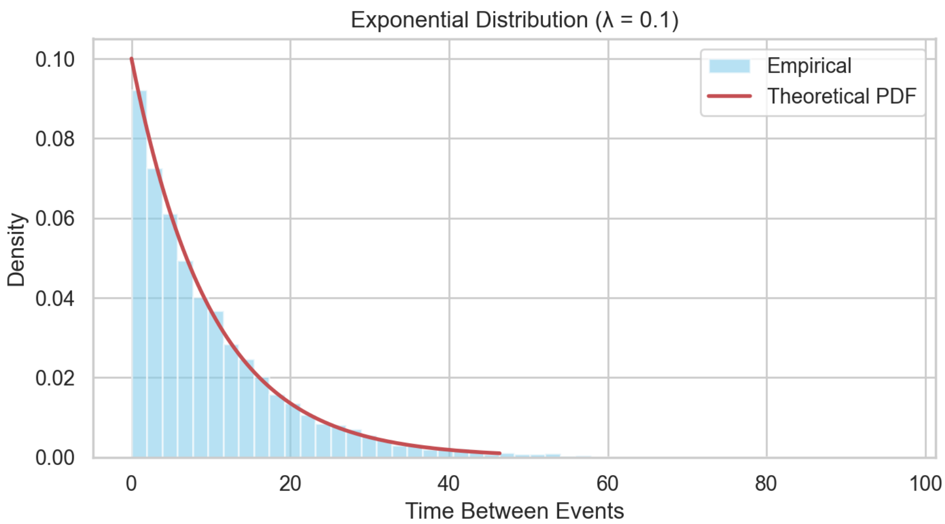

Figure 5: Plot of the exponential probability density function for a given λ.

Example. Normal Distribution

Let X be a continuous random variable that follows a normal distribution with mean μ and variance σ^2. Its probability density function (pdf) is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \text{for } x \in \mathbb{R}.$$

The normal distribution is symmetric about μ and is widely used in statistics to model natural variations and measurement errors. The parameters μ and σ^2 represent the center and spread of the distribution, respectively. Figure 6 illustrates a typical normal pdf.

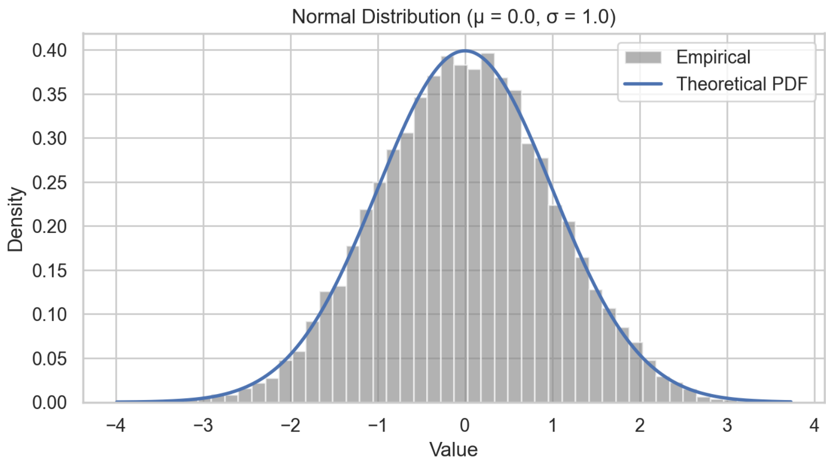

Figure 6: Plot of the normal probability density function with parameters μ and σ^2.

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 $[0, 1]$, as seen in the previous class.

2 Exercises

Exercise 1: Maximum of Five Uniform$(0, 1)$ Variables.

Let $Z_1, Z_2, \ldots, Z_5$ be independent random variables, each uniformly distributed on $(0, 1)$. Define

$$M = \max\{Z_1, Z_2, Z_3, Z_4, Z_5\}.$$

(a) Find $P(M \leq x)$ for $0 \leq x \leq 1$.

(b) Differentiate to find the PDF of M, and discuss its interpretation.

Exercise 2: Sum of Two Uniform$(0,1)$ Variables.

Let W_1 and W_2 be independent random variables, each uniformly distributed on $(0,1)$. Define

$$S = W_1 + W_2.$$

(a) Find the probability density function of S.

(b) Find the cumulative distribution function of S.

Exercise 3: Repeated Uniform$(0,1)$ Picks Until Sum > 1.

Independently pick random numbers $V_1, V_2, \ldots$ each uniformly distributed on $(0,1)$. Stop as soon as

$$V_1 + V_2 + \cdots + V_N > 1.$$

Let $X = N$, the number of picks required.

(a) Find $p_X(n) = P(X = n)$.

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

Exercise 4: From Binomial to Poisson.

Let $X_n \sim \text{Bin}(n, p)$ be a binomial random variable with parameters n and p. Define $\lambda = np$ and consider the limit as $n \to \infty$ while $p \to 0$ in such a way that λ remains constant. Show that the probability mass function of X_n converges to the Poisson distribution with mean λ.

Exercise 5: From Poisson to Exponential.

Consider a Poisson process with rate λ, where $N(t) \sim \text{Poisson}(\lambda t)$ represents the number of events occurring in time t. Let T be the waiting time until the first event, defined as

$$T = \inf\{t > 0 \mid N(t) \geq 1\}.$$

Show that the probability of no events occurring in $[0, t]$ is given by $P(T > t) = e^{-\lambda t}$, and conclude that $T \sim \text{Exp}(\lambda)$.

Exercise 6: Memoryless Property of the Exponential Distribution.

Let $X \sim \text{Exp}(\lambda)$ be an exponentially distributed random variable with rate λ. Show that for any $s, t \geq 0$,

$$P(X > s + t \mid X > s) = P(X > t).$$