Lecture 07

CONTINUOUS RANDOM VARIABLES

Probability & Statistics

Francisco Richter, Martina Boschi and Ernst Wit

Consider the probability as a *measure* of chance. For a discrete random variable X taking values $x_1, x_2, \ldots, x_k$, the probability mass function assigns a probability to each possible outcome x_i. The sum of these probabilities over all possible values of X equals one:

$$\sum_{i=1}^{k} p_X(x_i) = 1.$$

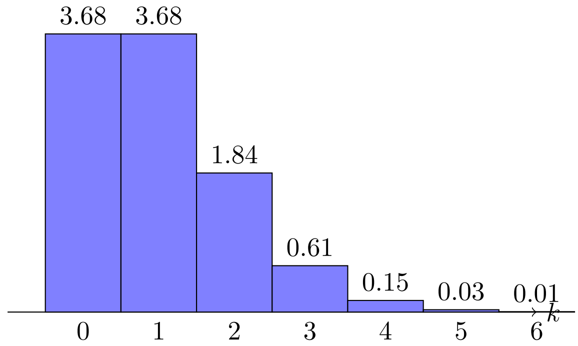


Figure 1: Poisson Distribution PMF for $\lambda = 1$

If you consider the figure above, each probability mass $p_X(x_i)$ represents the area of a rectangle with a width of $\Delta x = 1$, which corresponds to the discrete nature of the outcomes. Thus, the probability of each outcome is visualized as the height of each bar in a histogram, where the area of each bar is simply $p_X(x_i) \cdot 1 = p_X(x_i)$. When we sum these areas over all possible values of X, we obtain the total probability, which equals 1. For the Poisson distribution with $\lambda = 1$, this can be expressed as:

$$\sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} \frac{e^{-1} \cdot 1^k}{k!} = e^{-1} \sum_{k=0}^{\infty} \frac{1}{k!} = e^{-1} \cdot e = 1$$

In probability theory, the transition from a discrete to a continuous probability distribution can be understood by considering a sequence of increasingly finer partitions of the universe.

In the discrete case, we approximate the interval $[a, b]$ by dividing it into n small subintervals $[x_i, x_{i+1}]$ of equal width $\Delta x = \frac{b-a}{n}$. The probability of X falling within the i-th interval can be approximated as:

$$p_X(x_i \leq x < x_{i+1}) \approx f(x_i) \cdot \Delta x$$

where $f(x_i)$ represents the probability density at point x_i.

Consider the example of the Linear Congruential Generator, introduced in the previous chapters. Recall that LCGs that satisfy particular properties generate a sequence of random numbers between 0 and $m - 1$. To obtain a random number between 0 and 1, we simply divide the output of the LCG by m. This operation maps the range from 0 to $m - 1$ to the range from 0 to 1 (excluding 1).

Mathematically, this can be represented as:

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

where X_{n+1} is the output of the LCG and U_{n+1} is the corresponding random number between 0 and 1.

Imagine mapping each outcome X_i to an interval on the real line $[0, 1]$. Specifically, assign each outcome to a subinterval of width $\Delta x = \frac{1}{k}$:

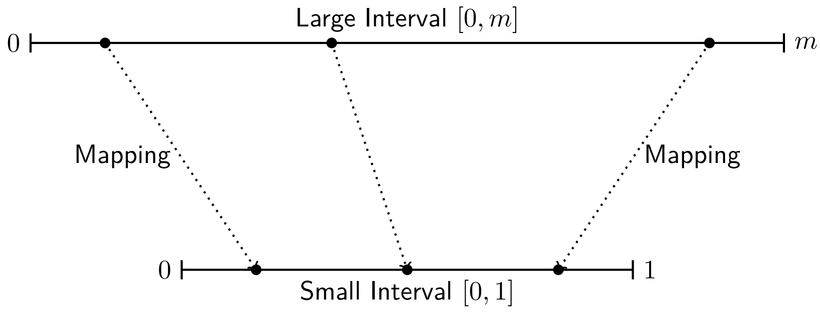


Figure 2: Mapping of random numbers from a large interval to a small interval.

$$X_i \leftrightarrow \left[\frac{i-1}{k}, \frac{i}{k}\right), \quad \text{for } i = 1, 2, \ldots, k.$$

The total probability of X falling within $[a, b]$ can be approximated by summing the probabilities for each subinterval:

$$\sum_{i=1}^{n} f(x_i) \cdot \Delta x$$

As we increase the number of intervals n (and hence make each interval width Δx smaller), this sum becomes a better approximation of the total probability. In the limit as $n \to \infty$, the width $\Delta x \to 0$, and the sum converges to the integral of $f(x)$ over $[a, b]$:

$$\lim_{n \to \infty} \sum_{i=1}^{n} f(x_i) \cdot \Delta x = \int_a^b f(x)\, dx$$

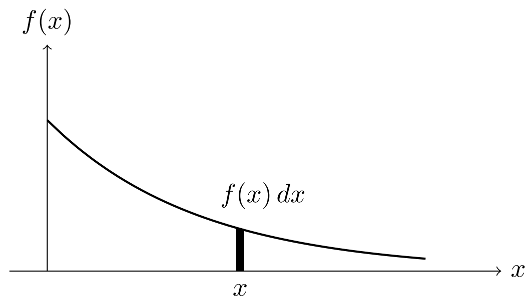
This limit process shows that the sum of probabilities in the discrete case can be viewed as an approximation to the integral in the continuous case. The integral $\int_a^b f(x)\, dx$ gives the exact probability of the continuous random variable X taking a value within the interval $[a, b]$.

Definition 1. *A function $f(x)$ is called a **probability density function (pdf)** of a continuous random variable X if it satisfies the following properties:*

- $f(x) \geq 0$ *for all* x,
- $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

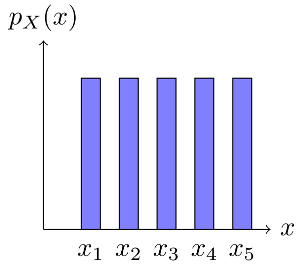
The probability that X falls within any interval $[a, b] \subseteq \mathbb{R}$ is then given by:

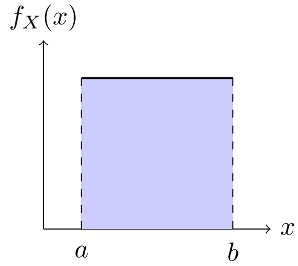
$$p_X(a \leq x \leq b) = \int_a^b f(x)\, dx.$$



Definition:
A continuous uniform distribution on the interval $[0, 1]$ is defined by its probability density function (pdf):

$$f_X(x) = \begin{cases} 1 & \text{for } 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$





In the continuous limit:

- Each discrete outcome's probability tends to zero, reflecting the uncountably infinite outcomes in the continuous case.
- The probability density $f(x)$ remains constant at 1, ensuring that the integral over $[0, 1]$ is 1:

$$\int_0^1 f(x)\,dx = \int_0^1 1\,dx = 1$$

A continuous random variable X follows a uniform distribution over the interval $[a, b]$ if it has the same probability density for all values within this interval. The pdf of a uniformly distributed random variable X is given by:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

This means that for any two sub-intervals of equal length within the interval $[a, b]$, the probability that X falls into either sub-interval is the same.

Definition 2 (Cumulative Distribution Function for Continuous Random Variables)**.** *Let X be a continuous random variable with pdf $f_X(x)$. The cumulative distribution function of X is defined for every number x by the following integral:*

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt$$

where $F_X(x)$ represents the probability that the random variable X takes on a value less than or equal to x.

Example (Cumulative Distribution Function of Continuous Uniform Distribution)**.** Consider a continuous random variable X that follows a uniform distribution over the interval $[a, b]$. We previously defined its pdf as:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

The cdf of X, denoted $F_X(x)$, is defined as the integral of the pdf from the lower bound of the distribution up to x:

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

After calculating, we obtain:

$$F_X(x) = \begin{cases} 0 & \text{for } x < a, \\ \frac{x-a}{b-a} & \text{for } a \leq x \leq b, \\ 1 & \text{for } x > b. \end{cases}$$

This cdf indicates the probability that the random variable X will take a value less than or equal to x. Note that $F_X(x)$ increases linearly from 0 to 1 as x ranges from a to b, consistent with our intuition of the uniform distribution. $\square$

Theorem 3. *Let T be a continuous random variable with cumulative distribution function (CDF) $F_T(t)$. Then the probability density function (PDF) $f_T(t)$ of T is given by the derivative of the CDF with respect to t:*

$$f_T(t) = \frac{d}{dt} F_T(t).$$

Proof. By definition, the cumulative distribution function $F_T(t)$ of a continuous random variable T is given by

$$F_T(t) = P(T \leq t) = \int_{-\infty}^{t} f_T(x)\,dx.$$

To find the probability density function $f_T(t)$, we differentiate $F_T(t)$ with respect to t. Applying the Fundamental Theorem of Calculus, we get

$$\frac{d}{dt} F_T(t) = \frac{d}{dt} \int_{-\infty}^{t} f_T(x)\,dx = f_T(t).$$

Thus, we have shown that

$$f_T(t) = \frac{d}{dt} F_T(t),$$

as required. $\square$

Continuous processes and discrete processes are both interconnected parts of the real world. Let T represent the **waiting time** until the first Poisson event occurs. Since we are dealing with time, T is a continuous random variable. Given a Poisson process with rate λ, we want to find the probability that the first event happens after a certain time t. This probability can be represented as:

$$p_T(T > t) = p(\text{no events occur in the interval } [0, t]).$$

Since the Poisson process assumes independent events occurring at a rate λ, the probability of observing zero events in an interval of length t is given by the Poisson distribution:

$$P(\text{no events in } [0, t]) = p_X(x = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}.$$

Thus, we have:

$$p_T(T > t) = e^{-\lambda t}.$$

This represents the probability that the waiting time T exceeds t. To obtain the probability density function (pdf) of T, we consider:

$$p_T(t) = p_T(T \leq t) = 1 - P(T > t) = 1 - e^{-\lambda t}.$$

Now, differentiating $F_T(t)$ with respect to t gives us the pdf of T, which we denote as $f_T(t)$:

$$f_T(t) = \frac{d}{dt} F_T(t) = \frac{d}{dt}\left(1 - e^{-\lambda t}\right) = \lambda e^{-\lambda t}.$$

This pdf, $f_T(t) = \lambda e^{-\lambda t}$, defines the exponential distribution with rate parameter λ.

Definition 4 (Exponential Distribution (1))**.** *A continuous random variable X follows an Exponential distribution with a rate parameter $\lambda > 0$ if its pdf is given by:*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases}$$

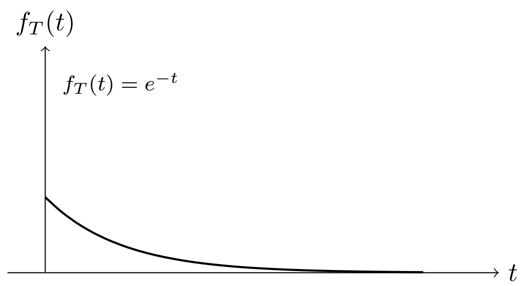


Figure 3: Probability Density Function of the Exponential Distribution with $\lambda = 1$

First, let's verify that $f_X(x)$ is a valid pdf. For it to be a pdf, the integral over its entire domain must be equal to 1.

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{\infty} \lambda e^{-\lambda x}\,dx$$

After integrating, we find that:

$$\int_0^{\infty} \lambda e^{-\lambda x}\, dx = -e^{-\lambda x}\Big|_0^{\infty} = 0 - (-1) = 1$$

Now, let's find the cdf of X, denoted $F_X(x)$. The cdf is defined as the integral of the PDF from the lower bound of the distribution up to x:

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

After calculating, we obtain:

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0, \\ 1 - e^{-\lambda x} & \text{for } x \geq 0. \end{cases}$$

This CDF gives us the probability that the random variable X will take a value less than or equal to x. As x approaches infinity, $F_X(x)$ approaches 1, consistent with our intuition of the Exponential distribution.

Generating Continuous Random Variables

We now need to understand how to generate continuous random variables using with specific probability density functions. We'll focus on inverse transform sampling method, which we previously introduced in a basic form for discrete random variables

Theorem 5 (Inverse Transform Sampling)**.** *Let U be a uniform random variable in the interval $[0,1]$, and let $F_X^{-1}(u)$ be the inverse function of $F_X(x)$, the cdf of X. Then $X = F_X^{-1}(U)$ will have the PDF $f_X(x)$.*

Proof. We need to prove that if $U \sim U(0,1)$ and $F(\cdot)$ is a valid invertible cdf, then $X = F^{-1}(U)$:

$$\begin{aligned} P(X \leq x) &= P(F_X^{-1}(U) \leq x) \\ &= P(U \leq F_X(x)) \\ &= F_U(F_X(x)) \\ &= F_X(x). \end{aligned}$$

□

This theorem provides us a convenient way to generate random variables that follow any given distribution, as long as we can compute its inverse cdf $F_X^{-1}(u)$.

Example (Exponential Distribution)**.** The exponential distribution is commonly used to model waiting times between events in a Poisson process. Its pdf is given by:

$$f_X(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

The cdf is then:

$$F_X(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

To generate a random variable X that follows an exponential distribution with rate λ, we can first generate a uniform random variable U in $[0,1]$, and then use the inverse of the CDF:

$$X = F_X^{-1}(U) = -\frac{1}{\lambda} \ln(1 - U)$$

□

By using the inverse transform sampling method, we can simulate any continuous random variable, provided we can find the inverse of its cdf. This theorem provides a straightforward way to generate numbers from an arbitrary probability distribution by simulating uniformly distributed numbers and calculating the proper transformation.

Moments

.

The expectation of a random variable X, often denoted $E[X]$, intuitively represents the weighted average of all possible outcomes, where the weight corresponds to the probability of each outcome.

For a discrete random variable, the mathematical expression for expectation is:

$$E[X] = \sum_x x \cdot p_X(x)$$

For a continuous random variable, the integral replaces the sum:

$$E[X] = \int x \cdot f_X(x)\, dx$$

The expectation, or expected value, is a measure of central tendency, indicating the average or "center" of a distribution. It is also known as the first moment about the origin and serves as the foundation for defining higher-order moments, which describe other characteristics of the distribution.

Definition 6. *The k-th moment of a random variable X is defined as*

$$E[X^k] = \int_{-\infty}^{\infty} x^k \cdot f_X(x)\, dx,$$

where $f_X(x)$ is the probability density function of X. Moments provide information about the distribution's shape, with the first moment giving the mean, the second moment relating to variance, and higher moments describing aspects like skewness and kurtosis.

Example. Let X follow an exponential distribution with rate parameter $\lambda > 0$, so that $X \sim \text{Exp}(\lambda)$ and $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. The k-th moment of X is given by

$$E[X^k] = \int_0^{\infty} x^k \cdot \lambda e^{-\lambda x}\, dx.$$

To compute $E[X^k]$, we use integration by parts or recognize this as a known integral involving the gamma function. The result is

$$E[X^k] = \frac{k!}{\lambda^k}.$$

For example, the first moment (mean) of X is $E[X] = \frac{1}{\lambda}$, and the second moment is $E[X^2] = \frac{2}{\lambda^2}$. □

The variance, denoted by $\mathsf{V}[X]$ or $V[X]$, measures how much a distribution deviates from its mean. Variance is defined as the expected value of the squared deviation of X from its mean $E[X]$. For a discrete random variable, the variance is given by

$$\mathsf{V}[X] = \sum_x (x - E[X])^2 \cdot p_X(x),$$

while for a continuous random variable, the integral replaces the sum:

$$\mathsf{V}[X] = \int_{-\infty}^{\infty} (x - E[X])^2 \cdot f_X(x)\, dx.$$

We can also express variance in an alternative form by expanding the squared term $(x - E[X])^2$:

$$\begin{aligned}
\mathsf{V}[X] &= E[(X - E[X])^2] \\
&= E[X^2 - 2X \cdot E[X] + (E[X])^2] \\
&= E[X^2] - 2E[X] \cdot E[X] + (E[X])^2 \\
&= E[X^2] - (E[X])^2.
\end{aligned}$$

Thus, the variance can be written as:

$$\mathsf{V}[X] = E[X^2] - (E[X])^2.$$

This form of the variance formula is often useful for calculations, as it expresses the variance in terms of the first two moments of X.

Example. For an exponential random variable $X \sim \text{Exp}(\lambda)$, we know from the previous example that $E[X] = \frac{1}{\lambda}$ and $E[X^2] = \frac{2}{\lambda^2}$. Using the variance formula, we can calculate $\mathsf{V}[X]$ as follows:

$$\mathsf{V}[X] = E[X^2] - (E[X])^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Therefore, the variance of an exponential random variable with rate λ is $\mathsf{V}[X] = \frac{1}{\lambda^2}$. □

After understanding the expectation and variance of a random variable, it's interesting to explore another crucial aspect of its distribution: the **skewness**. Skewness provides insight into the asymmetry of the probability distribution of a real-valued random variable about its mean.

Definition 7. *Skewness The* ***skewness*** *of a random variable X measures the degree of asymmetry of its distribution around its mean. It is defined as the third standardized moment:*

$$\gamma_1 = \frac{E\left[(X-\mu)^3\right]}{\sigma^3}$$

where:

- $\mu = E[X]$ *is the mean of X,*
- $\sigma^2 = V(X)$ *is the variance of X,*
- $E\left[(X-\mu)^3\right]$ *is the third central moment of X.*
- **Positive Skewness ($\gamma_1 > 0$)**: The right tail (higher values) of the distribution is longer or fatter than the left tail. The bulk of the data is concentrated on the left.
- **Negative Skewness ($\gamma_1 < 0$)**: The left tail (lower values) of the distribution is longer or fatter than the right tail. The bulk of the data is concentrated on the right.
- **Zero Skewness ($\gamma_1 = 0$)**: The distribution is perfectly symmetric around the mean. While not all symmetric distributions have zero skewness, all perfectly symmetric distributions do.

Example. Continuing with the Exponential distribution example, we will compute its skewness.
Recall that for an Exponential distribution with rate parameter $\lambda > 0$:

$$f_X(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

We have already established that:

$$E[X] = \frac{1}{\lambda}, \quad \mathsf{V}(X) = \frac{1}{\lambda^2}, \quad \text{and} \quad E[X^3] = \frac{6}{\lambda^3}.$$

Using the formula for the third central moment, we find:

$$E\left[(X-\mu)^3\right] = E[X^3] - 3\mu E[X^2] + 2\mu^3.$$

Substituting the values:

$$\mu = \frac{1}{\lambda}, \quad E[X^2] = \frac{2}{\lambda^2}, \quad E[X^3] = \frac{6}{\lambda^3},$$

we get:

$$\begin{aligned} E\left[(X-\mu)^3\right] &= \frac{6}{\lambda^3} - 3 \cdot \frac{1}{\lambda} \cdot \frac{2}{\lambda^2} + 2 \cdot \left(\frac{1}{\lambda}\right)^3 \\ &= \frac{6}{\lambda^3} - \frac{6}{\lambda^3} + \frac{2}{\lambda^3} \\ &= \frac{2}{\lambda^3}. \end{aligned}$$

then,

$$\gamma_1 = \frac{E\left[(X-\mu)^3\right]}{\sigma^3} = \frac{\frac{2}{\lambda^3}}{\left(\frac{1}{\lambda^2}\right)^{3/2}} = \frac{\frac{2}{\lambda^3}}{\frac{1}{\lambda^3}} = 2.$$

The Exponential distribution has a skewness of 2, indicating a moderate right skew. This asymmetry is characteristic of the Exponential distribution, where the tail on the right side is longer or fatter than the left side.

□

Skewness helps in understanding the shape of the distribution beyond its central tendency and variability. While the expectation provides the "center" and variance measures the "spread," skewness reveals whether the distribution leans towards higher or lower values relative to the mean.

Example (Uniform Distribution Skewness)**.** Consider a continuous random variable X that follows a uniform distribution over the interval $[a, b]$. We have already defined its probability density function as:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

Let's compute its skewness.

For a uniform distribution:

$$\mu = \frac{a+b}{2}, \quad \sigma^2 = \frac{(b-a)^2}{12}$$

The third central moment for a uniform distribution is:

$$E\left[(X-\mu)^3\right] = 0$$

then,

$$\gamma_1 = \frac{E\left[(X-\mu)^3\right]}{\sigma^3} = \frac{0}{\left(\frac{(b-a)^2}{12}\right)^{3/2}} = 0$$

The uniform distribution is perfectly symmetric around its mean μ. Due to this symmetry, the third central moment, which measures asymmetry, is zero.

The uniform distribution has a skewness of 0, indicating perfect symmetry around its mean. This aligns with our understanding that the uniform distribution is symmetric, with equal probability spread evenly across the interval $[a, b]$. □

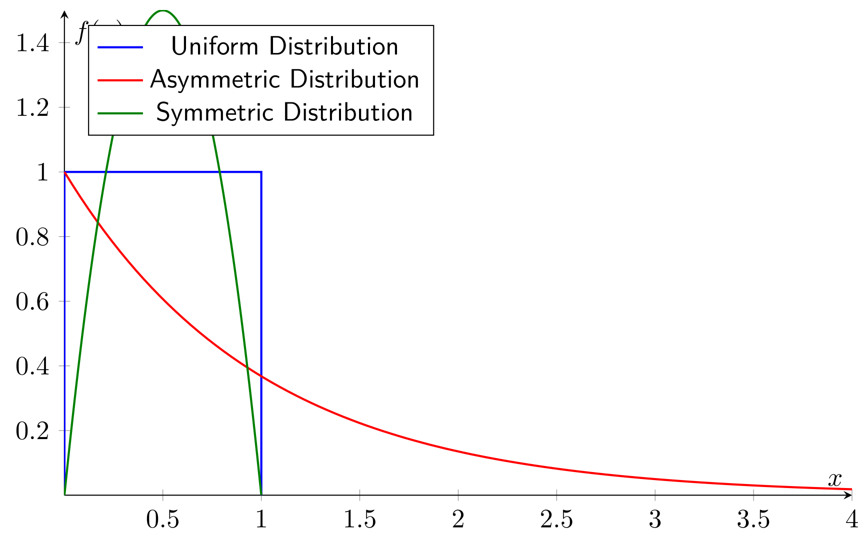


Figure 4: Comparison of Skewness in Different Distributions. The Uniform distribution is symmetric ($\gamma_1 = 0$), the Exponential distribution is positively skewed ($\gamma_1 = 2$), and the Beta distribution shown here is symmetric.

Central Limit Theorem

The Central Limit Theorem (CLT) is a fundamental result in probability theory and statistics. It states that the sum of a large number of independent and identically distributed (i.i.d.) random variables, each with finite mean and variance, tends to be normally distributed, regardless of the original distribution of the variables.

Definition 8. *A continuous random variable X is said to follow a* ***normal distribution*** *with mean μ and variance σ^2 if its probability density function (PDF) is given by*

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

where $-\infty < x < \infty$, $-\infty < \mu < \infty$, and $\sigma > 0$. This is denoted by $X \sim \mathcal{N}(\mu, \sigma^2)$.

The normal distribution is one of the most important distributions in probability and statistics due to the Central Limit Theorem, which states that the sum of many independent, identically distributed random variables tends to follow a normal distribution regardless of the original distribution.

Lemma 9. *Let $Z \sim \mathcal{N}(0,1)$. Then the characteristic function $\varphi_Z(t)$ of Z is given by*

$$\varphi_Z(t) = E[e^{itZ}] = \exp\left(-\frac{t^2}{2}\right).$$

Proof. By definition, the characteristic function $\varphi_Z(t)$ is

$$\varphi_Z(t) = E[e^{itZ}] = \int_{-\infty}^{\infty} e^{itz} \cdot f_Z(z)\, dz.$$

Since $Z \sim \mathcal{N}(0,1)$, we substitute the PDF of Z:

$$\varphi_Z(t) = \int_{-\infty}^{\infty} e^{itz} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz.$$

To simplify this integral, we complete the square in the exponent. Note that

$$itz - \frac{z^2}{2} = -\frac{1}{2}\left(z^2 - 2itz\right) = -\frac{1}{2}\left(z - it\right)^2 + \frac{t^2}{2}.$$

Thus,

$$\varphi_Z(t) = \exp\left(-\frac{t^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(z - it\right)^2\right) dz.$$

The remaining integral is the integral of a normal density with mean it and variance 1, which integrates to 1. Therefore,

$$\varphi_Z(t) = \exp\left(-\frac{t^2}{2}\right).$$

This completes the proof. □

With this, we are ready to present one of main results in probability theory and statistics

Theorem 10 (Central Limit Theorem)**.** *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with mean μ and variance σ^2, both finite. Define the standardized sum:*

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$$

Then, as $n \to \infty$,

$$Z_n \xrightarrow{d} \mathcal{N}(0,1)$$

where $\mathcal{N}(0,1)$ denotes the standard normal distribution.

To understand this theorem, let's analyze the characteristics of this result.

Aggregation of Random Effects

When summing many independent random variables, each contributing its own randomness, individual irregularities tend to "average out," leading to predictable overall behavior.

Consider n i.i.d. random variables $X_1, X_2, \ldots, X_n$, each with mean μ and variance σ^2.

$$\begin{aligned} S_n &= X_1 + X_2 + \cdots + X_n \\ E[S_n] &= n\mu \\ \mathsf{V}(S_n) &= n\sigma^2 \end{aligned}$$

To analyze the behavior as n grows, we standardize the sum:

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

As n increases, the standardized sum Z_n becomes more stable. The "aggregation" of individual random effects leads to a reduction in relative fluctuations, making Z_n less influenced by the variability of any single X_i.

Symmetry Through Averaging

As more variables are added, the influence of any single variable diminishes. This averaging effect induces symmetry in the distribution of the sum.

Define the standardized individual variables:

$$Y_i = \frac{X_i - \mu}{\sigma}$$

Thus, the standardized sum can be expressed as:

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$$

Each Y_i has:

$$E[Y_i] = 0, \quad \mathsf{V}(Y_i) = 1$$

The Law of Large Numbers ensures that:

$$\frac{1}{n} \sum_{i=1}^{n} Y_i \xrightarrow{a.s.} 0 \quad \text{as} \quad n \to \infty$$

If the original distribution of Y_i is skewed, the sum $\sum Y_i$ tends to balance out the skewness as positive and negative deviations cancel each other out. The skewness of Z_n diminishes as n increases:

$$\text{Skewness}(Z_n) = \frac{E[Z_n^3]}{(\mathsf{V}(Z_n))^{3/2}} = \frac{\gamma}{\sqrt{n}} \to 0 \quad \text{as} \quad n \to \infty$$

Where γ is the third central moment of Y_i.

The distribution of Z_n becomes increasingly symmetric around zero as n grows, regardless of the original distribution's symmetry. This emerging symmetry is a crucial step toward the normal distribution's characteristic bell shape.

Emergence of the Bell Curve

The normal distribution (bell curve) is inherently symmetric and arises naturally when multiple independent random factors contribute to a single outcome.

The characteristic function of Z_n is given by:

$$\phi_{Z_n}(t) = E\left[e^{itZ_n}\right] = E\left[e^{it\frac{S_n - n\mu}{\sigma\sqrt{n}}}\right]$$

Expanding S_n:

$$\phi_{Z_n}(t) = e^{-it\frac{n\mu}{\sigma\sqrt{n}}} \prod_{i=1}^{n} \phi_{Y_i}\left(\frac{t}{\sqrt{n}}\right)$$

Assuming Y_i has finite moments, expand $\phi_{Y_i}(t)$ around $t = 0$:

$$\phi_{Y_i}\left(\frac{t}{\sqrt{n}}\right) \approx 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)$$

Taking the product:

$$\prod_{i=1}^{n}\left(1 - \frac{t^2}{2n}\right) \approx \left(1 - \frac{t^2}{2n}\right)^n \approx e^{-t^2/2} \quad \text{as} \quad n \to \infty$$

Then,

$$\phi_{Z_n}(t) \approx e^{-t^2/2}$$

The characteristic function $e^{-t^2/2}$ uniquely corresponds to the standard normal distribution $\mathcal{N}(0, 1)$.

As n increases, the characteristic function of Z_n converges to that of the normal distribution, indicating that Z_n approaches $\mathcal{N}(0, 1)$.

Universal Attraction

Regardless of the original distribution of the individual variables (provided they have finite mean and variance), their sum tends to exhibit normal behavior as the number of variables grows.

As demonstrated earlier, the characteristic function of Z_n converges to that of $\mathcal{N}(0, 1)$, regardless of the form of $\phi_{Y_i}(t)$, provided that Y_i has finite mean and variance.

A more generalized condition ensuring CLT holds, even for non-identically distributed variables, is the Lindeberg-Feller condition. It states that no single variable dominates the sum, ensuring the CLT's applicability beyond identical distributions.

Role of Independence:

Independence ensures that individual random fluctuations cancel out when aggregated, preventing any single X_i from skewing the overall distribution.

The universal attraction of the normal distribution means that the CLT is remarkably powerful and widely applicable. It explains the ubiquity of the normal distribution in natural phenomena, finance, engineering, and more, as it arises from the collective behavior of numerous small, independent random effects.

Experiments

Imagine a gambler tossing dice at a casino. The outcomes of individual dice rolls are random, but what about the average outcome over a series of rolls? Interestingly, this average tends to behave in a very predictable manner as the number of dice increases, thanks to a fundamental result in probability theory known as the Central Limit Theorem (CLT).
In previous discussions, we delved into how the average outcome of multiple dice rolls doesn't follow a uniform distribution. For example, when you roll 10 dice, the sum of 35 is much more likely to occur than a sum of 10. We'll now extend this experiment to understand how these averages behave as the number of dice rolls increases.

Experiment 1 (Distribution of Dice Roll Averages)**.** ***Objective:*** *To investigate how the distribution of the averages of dice rolls changes as the number of dice rolls increases.*
Procedure:

- *Set* $N = 10,000$ *trials for the experiment.*
- *Let* $M = \{1, 2, 5, 10, 100, 1000\}$ *represent different dice counts.*
- *For each* $m \in M$*:*
  1. *Initialize an array for storing averages.*
  2. *For* $i = 1, \ldots, N$*:*
     - *Roll* m *dice.*
     - *Compute and store the average.*
- *Plot the distributions for each* m*.*

Observation: *The distribution narrows and assumes a symmetric shape as* m *increases.*

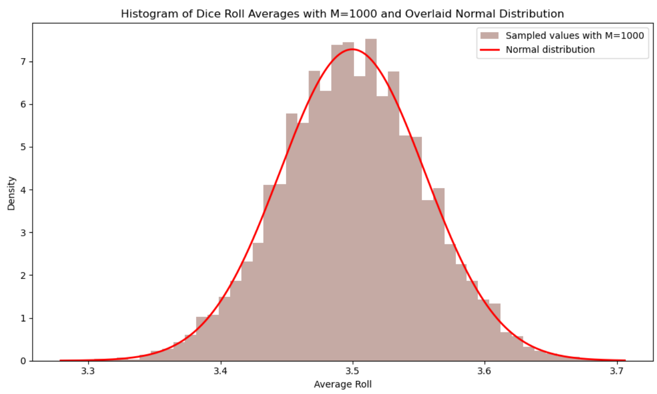


Figure 5: Histogram of Dice Roll Averages with $M = 1$ and Overlaid Normal Distribution

The CLT mathematically formalizes the phenomenon we observed in the dice experiment.

Experiment 2 (Conformance to Normal Distribution)**.** ***Objective:*** *To verify that the average of a large number of dice rolls approximates a normal distribution, in line with the Central Limit Theorem.*
Procedure:

1. *Use the averages obtained from the previous experiment where* $M = 1000$*.*
2. *Calculate* μ *and* σ *of these averages.*
3. *Plot a histogram of the averages and overlay it with a normal distribution curve parameterized by* μ *and* σ*.*

Insight: *The histogram closely matches the overlaid normal distribution, substantiating the Central Limit Theorem and reinforcing its applicability to gambling scenarios.*
Given that $\mu = 3.5$ *and* $\sigma^2 = 2.92$ *for this dice game, one can transform the sample means into Z-scores using*

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Plotting these Z-scores should result in a distribution that approximates a standard normal distribution $P(Z = z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$*.*