Week #3: Expectation

Stochastic Methods Course

Notes by: Francisco Richter

May 15, 2025

The concept of *expectation* (or expected value) is central in probability theory as it quantifies the ”average” outcome of a random experiment. It is rigorously defined via integration (for continuous variables) or summation (for discrete variables) with respect to a probability measure.

Definition. Expected Value

Let X be a random variable on a probability space $(\Omega, \mathcal{F}, P)$.

- If X is continuous with probability density function $f_X(x)$, then

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

- If X is discrete with countable range $R(X) = \{x_1, x_2, \dots\}$ and probability mass function $f_X(x)$, then

$$E[X] = \sum_{x \in R(X)} x\, f_X(x).$$

Investor Example

Consider an investor who commits a fixed investment every month in a new technology. The monthly revenue multiplier is modeled by the random variable M:

$$M = \begin{cases} 15, & \text{with probability } 0.10, \\ 1, & \text{with probability } 0.70, \\ 0.1, & \text{with probability } 0.20. \end{cases}$$

Since a multiplier of 1 represents no gain or loss, the net monthly change is $M - 1$. The expected monthly multiplier is

$$E[M] = (15)(0.10) + (1)(0.70) + (0.1)(0.20) = 1.5 + 0.7 + 0.02 = 2.22,$$

so the expected net gain per month is

$$E[M] - 1 = 2.22 - 1 = 1.22.$$

Assuming monthly outcomes are independent, after 12 months the total expected net gain is

$$12 \times (E[M] - 1) = 12 \times 1.22 = 14.64.$$

Linearity of Expectation

One of the most useful properties of expectation is its linearity. This property holds regardless of whether the random variables are independent.

Theorem 0.1 (Linearity of Expectation). *For any random variables $X_1, X_2, \ldots, X_n$ and any constants $a_1, a_2, \ldots, a_n$,*

$$E\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i E[X_i].$$

Proof. The proof follows directly from the definition of expectation and the linearity of integrals (or sums). Since the constants a_i are not random, they can be factored out of the expectation:

$$E[a_i X_i] = a_i E[X_i],$$

and summing over all i yields the result. □

Example 0.1 (Bernoulli Distribution). *Let X be a Bernoulli random variable with success probability p. Then,*

$$E[X] = 1 \cdot p + 0 \cdot (1-p) = p.$$

Example 0.2 (Poisson Distribution). *Let X be a Poisson random variable with parameter λ. Its expected value is given by:*

$$E[X] = \sum_{k=0}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} = \lambda.$$

1 Law of Large Numbers

The Law of Large Numbers (LLN) guarantees that as the number of independent observations increases, the sample average converges to the expected value.

Theorem 1.1 (Law of Large Numbers). *Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables with finite expectation $E[X_i] = \mu$. Then, as $n \to \infty$,*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu,$$

i.e., the sample average converges in probability to μ.

Example 1.1 (LLN in the Investment Scenario). *If the investor repeats the monthly investment over a large number of months, the average net gain per month will converge to 1.22 as predicted by the LLN.*

2 Central Limit Theorem

The Central Limit Theorem (CLT) is one of the cornerstones of probability theory. It states that the sum (or average) of a large number of i.i.d. random variables, when properly normalized, converges in distribution to a normal distribution.

Theorem 2.1 (Central Limit Theorem). *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ and variance σ^2. 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 $\mathcal{N}(0,1)$ is the standard normal distribution.

Discussion: Regardless of the original distribution of the X_i's (provided they have finite mean and variance), the distribution of the normalized sum tends to be normal. This explains why the normal distribution is prevalent in statistical applications.

3 Monte Carlo Simulation

Monte Carlo methods use random sampling to approximate numerical results, such as integrals or areas. A central application is Monte Carlo integration.

Monte Carlo Integration

Suppose we wish to evaluate an integral

$$I = \int_D f(x)\,dx,$$

where D is a domain with finite measure $|D|$. If $x_1, x_2, \dots, x_N$ are independent random samples uniformly drawn from D, then the estimator

$$\hat{I}_N = |D| \cdot \frac{1}{N} \sum_{i=1}^{N} f(x_i)$$

approximates I. By the Law of Large Numbers, as $N \to \infty$, $\hat{I}_N$ converges to I.

Example: Estimating π

Consider estimating π using Monte Carlo simulation. Let the domain be the unit square $D = [0,1] \times [0,1]$. Define the indicator function

$$I(x,y) = \begin{cases} 1, & \text{if } x^2 + y^2 \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

The probability that a random point in D falls inside the quarter circle (of radius 1) is

$$P((x,y) \in Q) = \frac{\pi/4}{1} = \frac{\pi}{4}.$$

Thus, by the relation

$$\pi = 4\,E[I(x,y)],$$

if we generate N points uniformly in D and compute the fraction $\hat{p}$ that lie inside the quarter circle, then

$$\hat{\pi} = 4\,\hat{p}$$

is an estimator for π.

4 Exercises

Exercise 1: Maximum of Five Uniform $(0,1)$ Variables.
Let $Z_1, Z_2, \dots, Z_5$ be independent random variables uniformly distributed on $(0,1)$. Define

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

(a) Derive an expression for $P(M \le x)$ for $0 \le x \le 1$.

(b) Differentiate $P(M \le x)$ to obtain the pdf of M and interpret the result.

Exercise 2: Sum of Two Uniform $(0,1)$ Variables.
Let W_1 and W_2 be independent random variables uniformly distributed on $(0,1)$. Define

$$S = W_1 + W_2.$$

(a) Derive the probability density function of S.

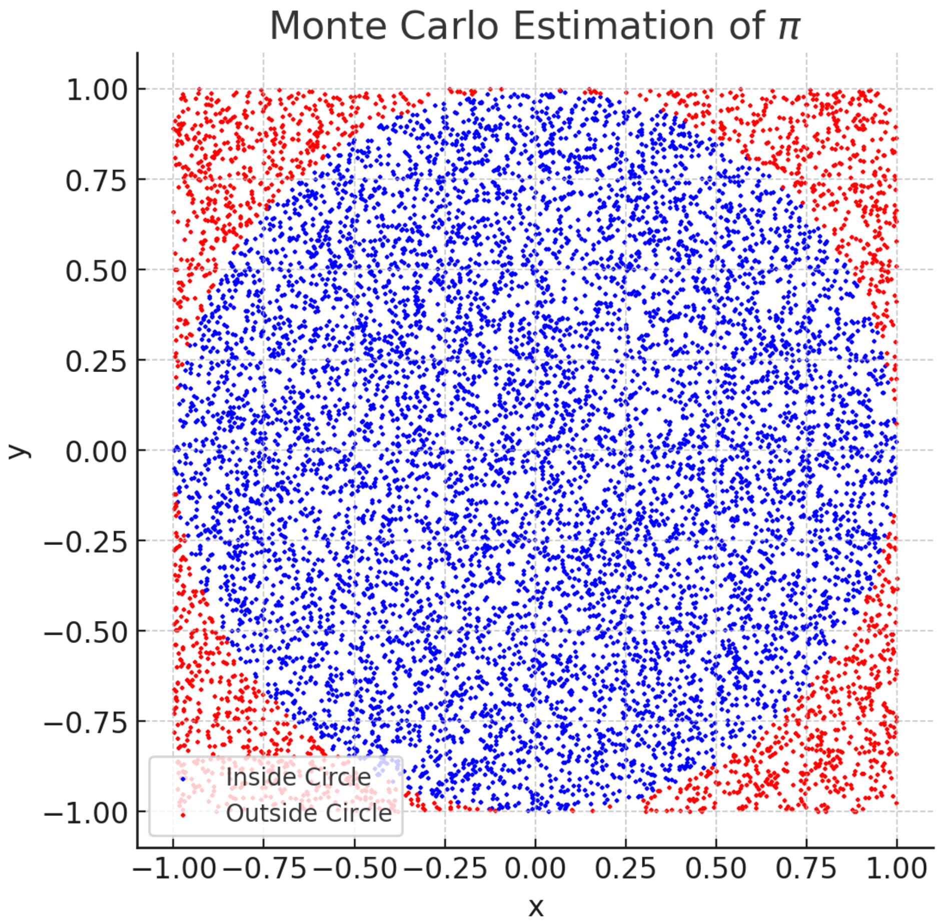

Figure 1: Monte Carlo estimation of π: random points in a unit square.

(b) Derive the cumulative distribution function of S.

Exercise 3: Repeated Uniform $(0, 1)$ Picks Until Sum > 1.
Independently generate random numbers $V_1, V_2, \ldots$ uniformly distributed on $(0, 1)$ until

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

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

(a) Determine the pmf $p_X(n) = P(X = n)$.
(b) Determine the cdf $F_X(n) = P(X \le n)$.

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 $\lambda = np$. Show that as $n \to \infty$ and $p \to 0$ with λ constant, the pmf of X_n converges to that of a Poisson random variable with mean λ.

Exercise 5: From Poisson to Exponential.
Consider a Poisson process with rate λ, where $N(t) \sim \text{Poisson}(\lambda t)$ denotes the number of events in time t. Let

$$T = \inf\{t > 0 : N(t) \ge 1\}$$

be the waiting time until the first event. Prove that

$$P(T > t) = e^{-\lambda t},$$

and conclude that $T \sim \text{Exp}(\lambda)$.

Exercise 6: Memoryless Property of the Exponential Distribution.
Let $X \sim \text{Exp}(\lambda)$. Prove that for any $s, t \geq 0$,

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