Week #3: Expectation

March 10, 2025

1 Expectation

Consider an investor who, every month, invests a fixed amount in a new technology. In each month, the technology produces one of three outcomes: with a 10% chance it yields a 15-fold return, with a 70% chance it leaves the investment unchanged (1× return), and with a 20% chance it reduces the investment to 0.1× its value. In other words, if we let the random variable M denote the revenue multiplier in one month, then

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

If the investor's baseline is recovering the original investment (a multiplier of 1), then the net gain or loss in a month is measured by $M - 1$. Naturally, if he makes this investment every month, we want to know his expected overall gain or loss after one year (i.e., 12 months).

The concept underlying this analysis is the *expected value*.

Definition. Expected Value

if X is a random variable defined on a probability space $(U, \mathcal{F}, P)$, its expected value is defined by the integral

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

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

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

For a discrete random variable—where the probability measure P is concentrated on a countable set $R(X) = \{x_1, x_2, \dots\}$—the integral reduces to the sum

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

since the measure of a singleton $\{x\}$ is given by $P(\{x\}) = f_X(x)$.

Returning to our investor's example, let M be the monthly revenue multiplier as defined above. Then the expected monthly multiplier is

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

This means that, on average, the investment multiplies the money by 2.22 in a month. Since a multiplier of 1 corresponds to no gain or loss, the expected net gain per month is

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

If the investor makes this investment each month for one year (12 months), his total expected net gain is

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

Thus, for every unit of currency invested monthly, he can expect an overall gain of 14.64 units over one year.

[Linearity of Expectation] Let $X_1, X_2, \ldots, X_n$ be random variables (not necessarily independent). For any constants $a_1, a_2, \ldots, a_n$, the expectation of their linear combination is given by:

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

Proof. Let $X_1, X_2, \ldots, X_n$ be random variables and $a_1, a_2, \ldots, a_n$ be constants. We want to prove that:

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

By the definition of the expectation of a sum of random variables:

$$E\left[\sum_{i=1}^{n} a_i X_i\right] = E[a_1 X_1 + a_2 X_2 + \cdots + a_n X_n]$$

One of the fundamental properties of expectation is its linearity, which means that the expectation of a sum is the sum of the expectations:

$$E[a_1 X_1 + a_2 X_2 + \cdots + a_n X_n] = E[a_1 X_1] + E[a_2 X_2] + \cdots + E[a_n X_n]$$

Since the constants a_i are not random, they can be factored out of the expectation:

$$E[a_1 X_1] = a_1 E[X_1], \quad E[a_2 X_2] = a_2 E[X_2], \quad \ldots, \quad E[a_n X_n] = a_n E[X_n]$$

Thus, we have:

$$E[a_1 X_1 + a_2 X_2 + \cdots + a_n X_n] = a_1 E[X_1] + a_2 E[X_2] + \cdots + a_n E[X_n]$$

Finally, we combine these terms into a summation:

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

$\square$

Example. Expectation of the Bernoulli Distribution

Let X be a random variable following a Bernoulli distribution with success probability p. Using the expectation formula for the Bernoulli distribution, we have:

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

Hence, the expected value of X is p.

Example. Expectation of the Poisson Distribution

Let X be a random variable following a Poisson distribution with parameter λ. To find the expectation $E[X]$, we compute:

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

This can be split into two parts: the $k = 0$ term and the sum from $k = 1$ to infinity. The $k = 0$ term gives a contribution of 0 to the sum.

$$E[X] = \sum_{k=1}^{\infty} k \cdot \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-1)!}$$

Multiplying and dividing by λ in our expression for $E[X]$, we notice that:

$$E[X] = \lambda \cdot e^{-\lambda} \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

By replacing $h = k - 1$ and recognizing the Taylor series expansion for e^{λ}, which is $e^{\lambda} = \sum_{h=0}^{\infty} \frac{\lambda^h}{h!}$, $E[X]$ becomes:

$$E[X] = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda$$

Hence, the expected value of X coincides with its rate parameter λ.

The Law of Large Numbers provides an answer, guaranteeing that, under specific conditions, the sample average converges to the theoretical expectation as the number of samples increases. This law underpins the intuitive notion that as we collect more independent observations, their average tends towards the true mean of the distribution.

Theorem 1. Law of Large Numbers

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with a finite expectation denoted by $\mathbb{E}[X_i]$. Then, as n approaches infinity, the sample average converges to the expected value:

$$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{} \mathbb{E}[X_i].$$

In practical terms, the average outcome from a large number of trials will approximate the expected value, and this approximation improves with more trials.

Example. Law of Large Numbers in an Investment Scenario

Consider an investor who, each month, invests a fixed amount in a new technology that yields one of three outcomes:

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

Here, M represents the monthly revenue multiplier, and the net gain per month is measured by $M - 1$. The theoretical 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

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

In this example, we simulate the investor's monthly outcomes over 500 months and generate 100 independent simulation paths. Figure 1 shows the cumulative average net gain for each simulation. Notice that during the first 12 months there is considerable variability; however, as the number of months increases, the cumulative averages converge and stabilize around the theoretical expected net gain of 1.22. This behavior exemplifies the Law of Large Numbers.

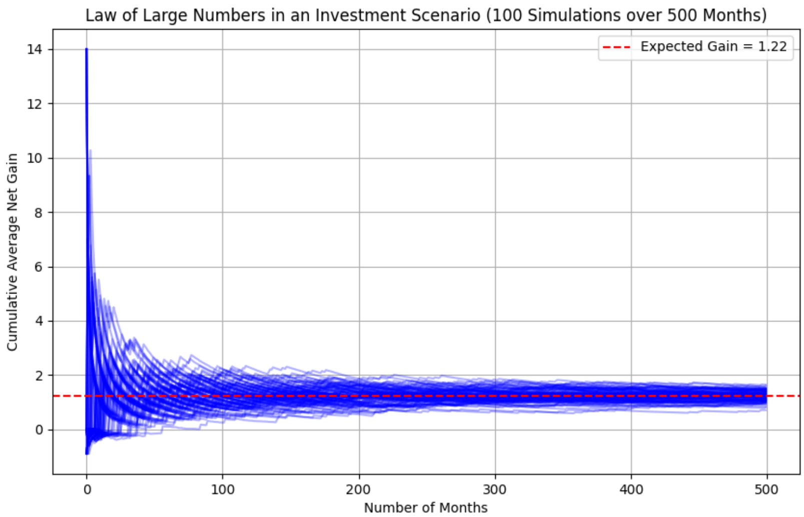

Figure 1: Cumulative average net gain over 500 months for 100 simulation paths. Early on (first 12 months) there is high variability, but with more months the average stabilizes around 1.22, as predicted by the Law of Large Numbers.

2 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.

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

Theorem 2. 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.

- **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.

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

- **Universal attraction**

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

  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.

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.

To illustrate the CLT visually, consider the following 2×2 grid of histograms. Each panel shows the distribution of normalized sums (or averages) from large samples of independent random variables drawn from one of these base distributions:

- Uniform (0,1)
- Exponential
- Poisson
- Geometric

As the number of summands increases, these normalized sums converge to the standard normal distribution.

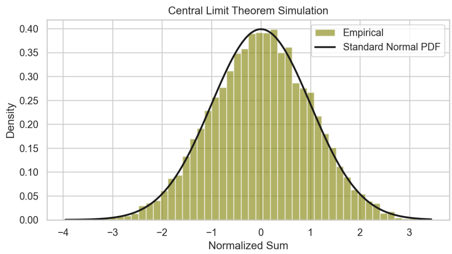

Figure 2: *
Uniform (0,1)

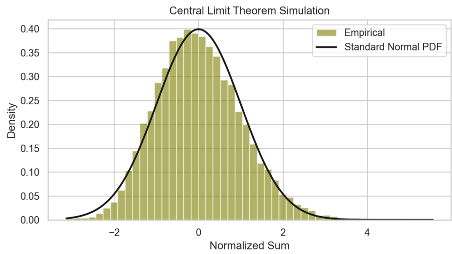

Figure 3: *
Exponential

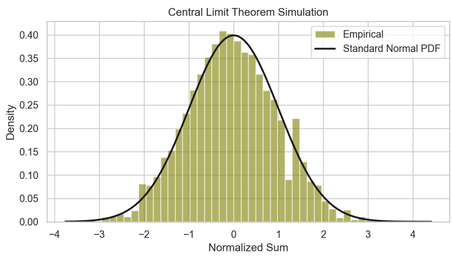

Figure 4: *
Poisson

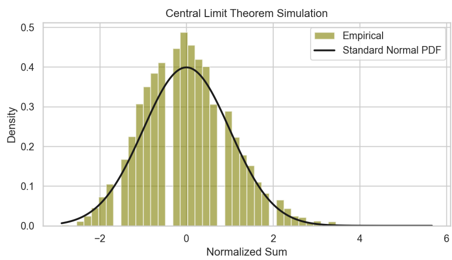

Figure 5: *
Geometric

Figure 6: Normalized averages from four base distributions, illustrating the Central Limit Theorem.

To generate data and the plots of this lecture note, please visit here.

3 Monte Carlo Simulation

Monte Carlo methods use random sampling to approximate numerical results. A central application is Monte Carlo integration, which estimates the area of a region—or more generally, the value of an integral—by interpreting the problem in terms of expected values.

For example, consider the classical problem of estimating π using a unit square. Define the domain

$$D = [0,1] \times [0,1],$$

and let the indicator function $I(x,y)$ be

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

A point (x, y) drawn uniformly from D falls inside the quarter circle (of radius 1) with probability

$$\mathbb{E}[I(x,y)] = \frac{\text{Area of quarter circle}}{\text{Area of } D} = \frac{\pi/4}{1} = \frac{\pi}{4}.$$

This gives the key relationship

$$\pi = 4\,\mathbb{E}[I(x,y)].$$

In practice, we approximate the expectation by drawing N independent random points $(x_1, y_1), \ldots, (x_N, y_N)$ in D and computing

$$\hat{p} = \frac{1}{N}\sum_{i=1}^{N} I(x_i, y_i).$$

Then an estimator for π is

$$\hat{\pi} = 4\,\hat{p} = \frac{4}{N}\sum_{i=1}^{N} I(x_i, y_i).$$

Figure 7 illustrates such a simulation, where blue dots represent points inside the circle and red dots those outside.

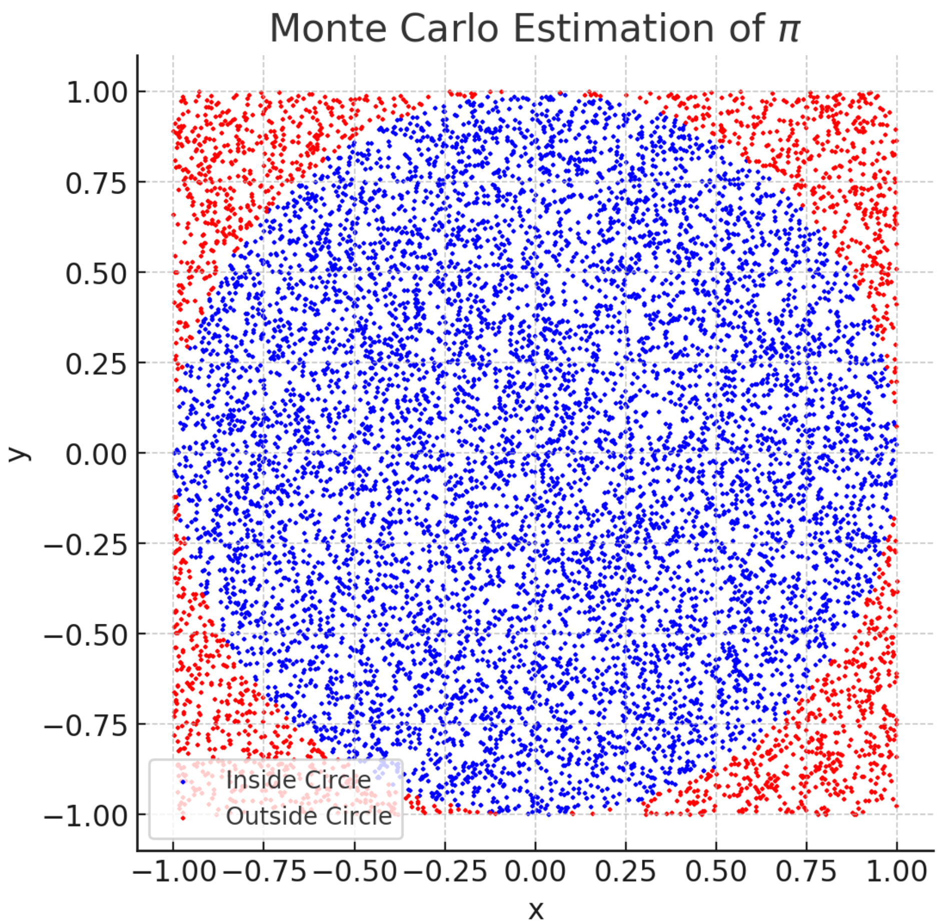

Figure 7: Monte Carlo estimation of π: random points in a unit square.

More generally, suppose we wish to estimate the area A_R of an arbitrary region R that is enclosed within a known rectangle. One common strategy is as follows:

---

1. **Enclose the Region:** Choose a rectangle that completely contains R, with area
$$A_{\text{rect}} = (b - a) \times M,$$
where, for example, $R = \{(x, y) \mid a \le x \le b,\ 0 \le y \le f(x)\}$ and $M \ge \max_{x \in [a,b]} f(x)$.

2. **Random Sampling:** Generate a large number N_{total} of random points uniformly distributed within the rectangle.

3. **Counting Points:** Let N_R be the number of points that fall inside R.

4. **Area Estimation:** Then the area is estimated by
$$A_R \approx \frac{N_R}{N_{\text{total}}} A_{\text{rect}}.$$

This method relies on the Law of Large Numbers: as N_{total} increases, the ratio $\frac{N_R}{N_{\text{total}}}$ converges to the true probability that a point lies in R.

Monte Carlo integration also provides a general method for estimating integrals by interpreting them as expected values. Suppose $f(x)$ is a real-valued function defined on a domain D with measure $|D|$ (for example, the area of D). If x is drawn uniformly from D, then the expected value of $f(x)$ is
$$\mathbb{E}[f(x)] = \frac{1}{|D|} \int_D f(x)\, dx,$$
or equivalently,
$$\int_D f(x)\, dx = |D|\, \mathbb{E}[f(x)].$$
A Monte Carlo estimator for the integral is
$$I_N = |D| \frac{1}{N} \sum_{i=1}^{N} f(x_i),$$
where $x_1, x_2, \ldots, x_N$ are independent samples drawn uniformly from D.

Theorem 3. Fundamental Theorem of Monte Carlo Integration

Let $f(x)$ be an integrable function on a domain D with measure $|D|$, and let $x_1, x_2, \ldots, x_N$ be independent random samples drawn uniformly from D. Then the Monte Carlo estimator
$$I_N = |D| \frac{1}{N} \sum_{i=1}^{N} f(x_i)$$
satisfies
$$\lim_{N \to \infty} I_N = \int_D f(x)\, dx.$$
That is, as $N \to \infty$, the estimator converges to the true value of the integral, in accordance with the Law of Large Numbers.

Monte Carlo integration can also be expressed using indicator random variables, which is especially useful when D or R is irregularly shaped. For example, define
$$I(x, y) = \begin{cases} 1, & \text{if } (x, y) \in R, \\ 0, & \text{otherwise.} \end{cases}$$

Then, after N independent samples,

$$\hat{p} = \frac{1}{N} \sum_{i=1}^{N} I(x_i, y_i)$$

estimates the probability $P((x, y) \in R)$, and the area of R is estimated by

$$\hat{A}_R = \hat{p}\, A_{\text{rect}}.$$

Example.

Consider the function $f(x) = x^2$ on the interval $[0, 1]$. The exact value of the integral is

$$\int_0^1 x^2\, dx = \frac{1}{3} \approx 0.3333.$$

To estimate this integral via Monte Carlo integration:

1. Enclose the area under $f(x)$ in the rectangle $[0, 1] \times [0, 1]$ (with area 1).
2. Generate N random points uniformly in the rectangle.
3. Count the number of points N_f that fall below the curve $y = x^2$.
4. Then, the integral is approximated by

$$I \approx \frac{N_f}{N}.$$

Figure 8 shows a typical simulation, where green dots lie below the curve and red dots lie above it. In this simulation, the estimated value of the integral was approximately 0.3349, very close to 1/3.

Exercises

Exercise 1. Let $f(x) = \sin(x)$ on the interval $[0, \pi]$. Consider the rectangle

$$D = [0, \pi] \times [0, 1],$$

and define the indicator function

$$\mathbf{1}_R(x, y) = \begin{cases} 1, & \text{if } 0 \le y \le \sin(x), \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show that the expected value of $\mathbf{1}_R$ when sampling uniformly from D is

$$E[\mathbf{1}_R(x, y)] = \frac{\text{Area under } \sin(x)}{|D|}.$$

(b) Explain how you would use Monte Carlo simulation to estimate the area under the curve $\sin(x)$.

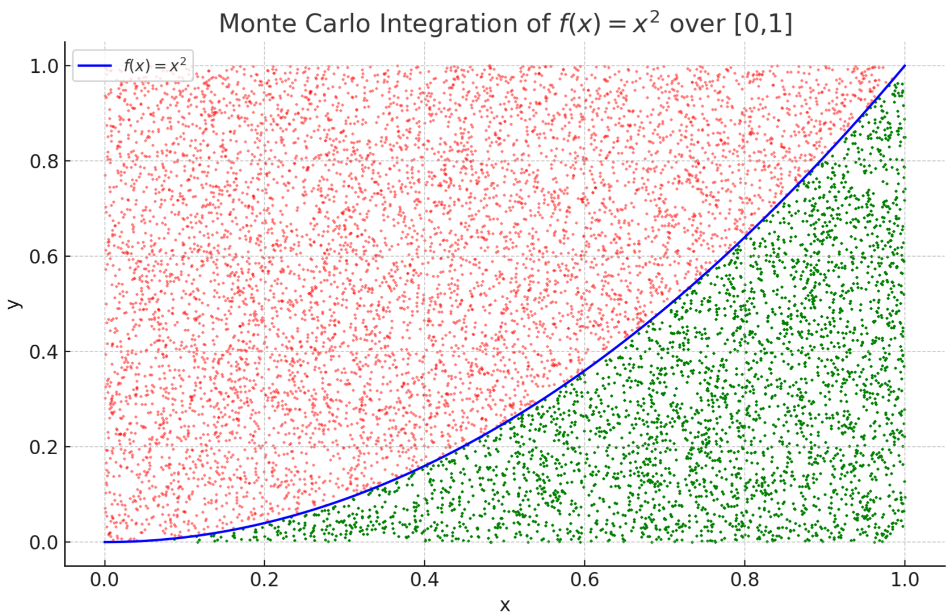

Figure 8: Monte Carlo Integration of $f(x) = x^2$ over $[0, 1]$. Green dots indicate points below the curve; red dots are above.

Exercise 2. Let X be an exponential random variable with rate parameter $\lambda = 2$. Define the indicator function for the event $\{X > 1\}$ as

$$\mathbf{1}_{\{X>1\}}(x) = \begin{cases} 1, & x > 1, \\ 0, & x \leq 1. \end{cases}$$

(a) Show that

$$E[\mathbf{1}_{\{X>1\}}] = P(X > 1)$$

and compute $P(X > 1)$ analytically.

(b) Describe a simulation strategy to estimate $P(X > 1)$ by generating N samples from the exponential distribution and computing the sample average of $\mathbf{1}_{\{X>1\}}$.

Exercise 3. Consider a random point (x, y) chosen uniformly from the unit square $[0, 1] \times [0, 1]$. Let C be the circle of radius 0.5 centered at $(0.5, 0.5)$. Define the indicator function

$$\mathbf{1}_C(x, y) = \begin{cases} 1, & \text{if } (x - 0.5)^2 + (y - 0.5)^2 \leq 0.25, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Derive an expression for $E[\mathbf{1}_C(x, y)]$ in terms of the area of the circle.

(b) Calculate the theoretical probability that a point lies inside the circle.

(c) Explain how Monte Carlo simulation can be used to estimate this probability.

Exercise 4. Let X be uniformly distributed on $[0, 1]$. Define the sets or events:

$$A = \{x | x < 0.3\} \quad \text{and} \quad B = \{x | x > 0.6\},$$

with corresponding indicator functions

$$\mathbf{1}_A(x) = \begin{cases} 1, & x < 0.3, \\ 0, & x \geq 0.3, \end{cases} \quad \text{and} \quad \mathbf{1}_B(x) = \begin{cases} 1, & x > 0.6, \\ 0, & x \leq 0.6. \end{cases}$$

(a) Prove that $E[\mathbf{1}_A] = P(A)$ and $E[\mathbf{1}_B] = P(B)$.

(b) Define the indicator function for the union $A \cup B$ and show how its expected value relates to $P(A \cup B)$.

(c) Using linearity of expectation, explain the relationship between

$$E[\mathbf{1}_A + \mathbf{1}_B] \quad \text{and} \quad P(A \cup B),$$

especially noting what happens when the events are disjoint.