Exercise Walkthrough: Alternative Expectation Formulas

Justin Lanfermann

25. June of 2025

Abstract

This document provides a detailed, step-by-step walkthrough for an exercise on computing the expectation of random variables. The exercise demonstrates how the expectation can be expressed as a sum or integral of the tail probabilities of a distribution. We will prove this for the discrete case and then apply the result to the Geometric and Exponential distributions, referencing definitions and concepts from the "Discrete Probability Theory" script by Niki Kilbertus.

1 The Exercise Statement

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

(i) Let $X : \Omega \to \mathbb{N}_0$ be a discrete random variable with $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. Show that

$$E[X] = \sum_{n=1}^{\infty} P(X \geq n)$$

and describe in words what this means.

(ii) Assume now that $X : \Omega \to \mathbb{N}_{>0}$ is geometrically distributed with parameter $p \in (0, 1)$, i.e., $X \sim \text{Geo}(p)$. Use (i) to calculate the expectation $E[X]$.

(iii) If $X : \Omega \to \mathbb{R}_{\geq 0}$ is a non-negative, continuous RVRV with cdf F_X, then it holds that

$$E[X] = \int_0^{\infty} (1 - F_X(x)) \, dx$$

Using this, calculate the expectation of an exponentially distributed random variable $X \sim \text{Exp}(\lambda)$, where $\lambda > 0$.

2 Solution Walkthrough

2.1 Part (i): Expectation as a Sum of Tail Probabilities

Overview Our goal is to prove that the expectation of a non-negative integer-valued random variable is the sum of its "tail probabilities" $P(X \geq n)$. This is a fantastic trick that reframes the calculation from "value times probability" to just summing probabilities.

Step-by-Step Proof Let's start from the definition of expectation [2] for a discrete random variable [1] X that takes values in $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. The formula is given in Definition 2.1 of the script (p. 42):

$$E[X] = \sum_{k=0}^{\infty} k \cdot P(X = k)$$

Since the $k = 0$ term is $0 \cdot P(X = 0) = 0$, we can start the sum from $k = 1$:

$$E[X] = \sum_{k=1}^{\infty} k \cdot P(X = k)$$

The key insight is to represent the integer k as a sum of ones: $k = \underbrace{1 + 1 + \cdots + 1}_{k \text{ times}}$. We can write this as $k = \sum_{n=1}^{k} 1$. Substituting this into our expectation formula gives:

$$E[X] = \sum_{k=1}^{\infty} \left(\sum_{n=1}^{k} 1 \right) P(X = k)$$

We now have a double summation. The term $P(X = k)$ is summed for every pair (n, k) where $k \geq 1$ and $1 \leq n \leq k$. We can visualize these pairs as points on a grid and realize that we can swap the order of summation [4]. Instead of summing over k first, we sum over n first.

$$\text{Original order: } \sum_{k=1}^{\infty} \sum_{n=1}^{k} \quad \Longrightarrow \quad \text{Swapped order: } \sum_{n=1}^{\infty} \sum_{k=n}^{\infty}$$

Applying this swap, we get:

$$E[X] = \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P(X = k)$$

Now, let's look at the inner sum, $\sum_{k=n}^{\infty} P(X = k)$. This is the sum of probabilities of all outcomes where X is equal to n, or $n + 1$, or $n + 2$, and so on. This is precisely the definition of the event $\{X \geq n\}$. Therefore:

$$\sum_{k=n}^{\infty} P(X = k) = P(X \geq n)$$

Substituting this back gives us our final result:

$$\boxed{E[X] = \sum_{n=1}^{\infty} P(X \geq n)}$$

Intuitive Explanation In a single sentence: The expected value can be thought of as the sum of the probabilities of surviving to each level n, summed over all possible levels.

Think of building the expectation as stacking blocks. The standard formula $\sum k \cdot P(X = k)$ calculates the total volume by summing up columns of height k and width $P(X = k)$. The formula we just proved calculates the same volume by summing up horizontal slices. The slice at height n has a width of $P(X \geq n)$, which is the total probability of all columns that are at least n blocks high. Summing up the areas of these horizontal slices gives the total volume.

A Quick Question The formula you were given in the exercise sheet was $\sum_{n=0}^{\infty} P(X \geq n)$, but we proved it for a sum starting at $n = 1$. What would be the difference if X can take the value 0? *Hint: What is $P(X \geq 0)$ for a random variable that only takes non-negative values?* The identity is often also written as $E[X] = \sum_{n=0}^{\infty} P(X > n)$, which is equivalent to our version.

2.2 Part (ii): Expectation of the Geometric Distribution

Overview Now we'll use our shiny new formula to calculate the expectation of a Geometrically distributed random variable, $X \sim \text{Geo}(p)$. This distribution models the number of trials needed to get the first success.

Step-by-Step Calculation

1. **Recall the Geometric Distribution [5]:** From Example 1.36 (vi) in the script (p. 20), the geometric distribution $X \sim \text{Geo}(p)$ is defined on the sample space $\Omega = \mathbb{N}_{>0} = \{1, 2, 3, \dots\}$. Its probability mass function (pmf) is:

$$P(X = k) = (1-p)^{k-1}p \quad \text{for } k \in \{1, 2, 3, \dots\}$$

   This represents the probability of having $k-1$ failures (each with probability $1-p$) followed by one success (with probability p).

2. **Calculate the Tail Probability** $P(X \geq n)$**:** Since our expectation formula from (i) requires $P(X \geq n)$, we need to compute this for our geometric RV. It's easier to first compute the complement, $P(X < n)$, which is $P(X \leq n - 1)$.

$$P(X \leq n - 1) = \sum_{k=1}^{n-1} P(X = k) = \sum_{k=1}^{n-1} (1-p)^{k-1}p$$

   This is a finite geometric series [3]. Let's pull out the constant p and re-index the sum by letting $j = k - 1$:

$$P(X \leq n - 1) = p \sum_{j=0}^{n-2} (1-p)^j$$

   Using the formula for a finite geometric series, $\sum_{j=0}^{N} q^j = \frac{1-q^{N+1}}{1-q}$:

$$P(X \leq n - 1) = p \left(\frac{1 - (1-p)^{(n-2)+1}}{1 - (1-p)} \right) = p \left(\frac{1 - (1-p)^{n-1}}{p} \right) = 1 - (1-p)^{n-1}$$

   Now we can easily find our tail probability for $n \geq 1$:

$$P(X \geq n) = 1 - P(X < n) = 1 - P(X \leq n - 1) = 1 - (1 - (1-p)^{n-1}) = (1-p)^{n-1}$$

   This is a surprisingly simple and elegant result! The probability of needing *at least* n trials is the same as the probability of having *at least* $n - 1$ failures.

3. **Sum the Tail Probabilities:** Now we plug this into our expectation formula. Since X is defined on $\mathbb{N}_{>0}$, the formula from (i) applies directly.

$$E[X] = \sum_{n=1}^{\infty} P(X \geq n) = \sum_{n=1}^{\infty} (1-p)^{n-1}$$

This is an infinite geometric series. Let's re-index with $m = n - 1$. As n goes from 1 to ∞, m goes from 0 to ∞:

$$E[X] = \sum_{m=0}^{\infty} (1-p)^m$$

Using the formula for an infinite geometric series $\sum_{m=0}^{\infty} q^m = \frac{1}{1-q}$ (which converges since $0 < p < 1 \implies 0 < 1 - p < 1$):

$$E[X] = \frac{1}{1-(1-p)} = \frac{1}{p}$$

So, the expected number of trials until the first success is $\boxed{E[X] = 1/p}$.

Check Your Intuition If you are flipping a fair coin ($p = 0.5$), how many flips would you expect to make on average to get the first heads? Our formula says $1/0.5 = 2$. This seems very reasonable! If an event is very rare, say $p = 0.01$, you'd expect to wait a long time: $1/0.01 = 100$ trials on average. The formula makes intuitive sense.

2.3   Part (iii): Expectation of the Exponential Distribution

Overview This part is the continuous analogue of part (i). We'll use the given integral formula to find the expectation of an exponentially distributed random variable, which is the continuous counterpart to the geometric distribution.

Step-by-Step Calculation

1. **Recall the Exponential Distribution [5]:** From Example 1.56 (iv) in the script (p. 29), an exponential random variable $X \sim \text{Exp}(\lambda)$ is defined for $\lambda > 0$ on $\mathbb{R}_{\geq 0}$. Its probability density function (pdf) is $p_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. The cumulative distribution function (cdf) $F_X(x) = P(X \leq x)$ is the integral of the pdf:

$$F_X(x) = \int_0^x \lambda e^{-\lambda t}\, dt = \left[-e^{-\lambda t}\right]_0^x = -e^{-\lambda x} - (-e^0) = 1 - e^{-\lambda x} \quad \text{for } x \geq 0$$

2. **Calculate $1 - F_X(x)$:** This is the "survival function," the continuous equivalent of the tail probability $P(X \geq n)$.

$$1 - F_X(x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x} \quad \text{for } x \geq 0$$

3. **Integrate the Survival Function:** Now we plug this into the given formula:

$$E[X] = \int_0^\infty (1 - F_X(x)) \, dx = \int_0^\infty e^{-\lambda x} \, dx$$

This is a standard integral.

$$\int_0^\infty e^{-\lambda x} \, dx = \left[-\frac{1}{\lambda} e^{-\lambda x} \right]_0^\infty$$

Evaluating at the limits:

$$E[X] = \lim_{b \to \infty} \left(-\frac{1}{\lambda} e^{-\lambda b} \right) - \left(-\frac{1}{\lambda} e^{-\lambda \cdot 0} \right) = (0) - \left(-\frac{1}{\lambda} \cdot 1 \right) = \frac{1}{\lambda}$$

Thus, for $X \sim \text{Exp}(\lambda)$, the expectation is $\boxed{E[X] = 1/\lambda}$.

3   Summary and Takeaways

This exercise showcased a powerful and elegant method for calculating expectations for non-negative random variables.

- **For discrete RVs on** $\mathbb{N}_0$, the expectation is the sum of the tail probabilities: $E[X] = \sum_{n=1}^\infty P(X \geq n)$.

- **For continuous RVs on** $\mathbb{R}_{\geq 0}$, the expectation is the integral of the survival function: $E[X] = \int_0^\infty (1 - F_X(x)) \, dx$.

- We applied these formulas to find that $E[\text{Geo}(p)] = 1/p$ and $E[\text{Exp}(\lambda)] = 1/\lambda$, reinforcing the deep connection between these two "memoryless" distributions.

This technique is not just a computational shortcut; it provides a different, and often more intuitive, perspective on what the expected value represents.

In-depth Explanations

[1] Random Variable (RV)   A random variable is not a variable in the traditional algebraic sense. It's a function that maps outcomes from a sample space Ω (the set of all possible outcomes of an experiment) to a set of numerical values (commonly the real numbers $\mathbb{R}$). For example, if our experiment is flipping a coin, $\Omega = \{$Heads, Tails$\}$. A random variable X could map this to $X(\text{Heads}) = 1$ and $X(\text{Tails}) = 0$. This allows us to use the tools of arithmetic and calculus to analyze the outcomes of random phenomena. (See script Def 1.45, p. 24).

[2] Expectation (Expected Value)   The expectation of a random variable is its long-run average value, weighted by the probability of each outcome. For a discrete RV X, it's the sum of each possible value multiplied by its probability: $E[X] = \sum_k k \cdot P(X = k)$. For a continuous RV, it's the integral of each value multiplied by its probability density: $E[X] = \int x \cdot p_X(x) \, dx$. It's a measure of the "center" or "central tendency" of a distribution. (See script Def 2.1, p. 42).

[3] Geometric Series   A geometric series is a sum where each term is found by multiplying the previous one by a fixed, non-zero number called the common ratio q.

- **Finite Series:** $\sum_{k=0}^{N} q^k = \frac{1-q^{N+1}}{1-q}$.

- **Infinite Series:** If $|q| < 1$, the series converges to a finite value: $\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$.

This formula is fundamental for working with the Geometric distribution.

[4] Changing the Order of Summation/Integration   The step where we swapped the order of summation is a discrete version of a powerful result in analysis called Fubini's Theorem. The theorem states that for "well-behaved" functions (specifically, non-negative functions in our case), you can compute a double integral (or sum) by integrating with respect to the variables in either order. The mental model is to imagine a grid of values you want to sum. You can either sum up all the values in each column first and then add up the column totals, or you can sum up all the values in each row first and then add up the row totals. Both methods must yield the same grand total.

[5] Geometric and Exponential Distributions   These two distributions are intrinsically linked.

- The **Geometric distribution** is discrete and models the number of independent Bernoulli trials needed to get one success. It is the only discrete distribution with the *memoryless property*.

- The **Exponential distribution** is continuous and models the waiting time until an event occurs in a Poisson process (where events happen at a constant average rate). It is the only continuous distribution with the *memoryless property*.

Their expectations, $1/p$ and $1/\lambda$, reflect this relationship: if success happens with probability p per trial, you expect $1/p$ trials. If events happen at rate λ per unit time, you expect to wait $1/\lambda$ units of time. (See script p. 20 for Geo, p. 29 for Exp).