Further Applications of Expectation

Math Circle Competition Team (sourced from brilliant.org)

0 Introduction

In probability theory, an **expected value**, or expectation, is the theoretical mean value of a numerical experiment over many repetitions of the experiment. Expected value is a measure of **central tendency**; a value for which the results will tend to.

Any given random variable contains a wealth of information. It can have many (or infinite) possible outcomes, and each outcome could have different likelihood. The expected value is a way to summarize all this information in a single numerical value. This concept is incredibly important in probability theory, statistics, and combinatorics.

1 Review of Definitions

When the sample space of a probability experiment can be encoded with numerical outcomes, then a **random variable** is the variable that represents the outcomes. For example, the result of rolling a fair six-sided die is a random variable that takes each of the values from 1 to 6 with probability $\frac{1}{6}$. This is an example of a **discrete random variable**.

For a discrete random variable, the expected value can be calculated by multiplying each numerical outcome by the probability of that outcome, and then summing those products together.

Definition 1. Let X be a discrete random variable. Then the expected value of X, denoted as $\mathrm{E}[X]$ or μ_X, is

$$\mathrm{E}[X] = \mu_X = \sum_{\text{all } x} xP(X = x).$$

Example 1. A stack of cards contains one card labeled with 1, two cards labeled with 2, three cards labeled with 3, and four cards labeled with 4. If the stack is shuffled and a card is drawn, what is the expected value of the card drawn?

Solution. Let X be a random variable that represents the value of a drawn card. We have $P(X = 1) = \frac{1}{10}$, $P(X = 2) = \frac{2}{10}$, $\ldots$, $P(X = 4) = \frac{4}{10}$. Thus, $\mathrm{E}[X] = \sum_{x=1}^{4} xP(X = x) = \boxed{3}$.

Recall that many random variables have an underlying **probability distribution**. To review, a probability distribution is a mathematical function that provides the probabilities of occurrence of different possible outcomes in an experiment. For instance, if the random variable X is used to denote the outcome of a fair coin toss, then the probability distribution of X would take the value 0.5 for $X =$ heads, and 0.5 for $X =$ tails. Below are the expectations of common distributions:

	Uniform	Hypergeometric	Bernoulli	Binomial	Geometric	Negative Binomial
$\mathrm{E}[X]$	$\frac{a+b}{2}$	$n\frac{K}{N}$	p	np	$\frac{1}{p}$ or $\frac{1}{p} - 1$	$\frac{n}{p}$ or $\frac{n}{p} - n$

To review these distributions, refer to the combinatorics handouts. Note that the expectation of the uniform distribution applies in both the discrete and continuous cases.

2 Properties

Theorem 1. *For any random variable X and constant c,*

$$\mathrm{E}[X+c] = \mathrm{E}[X] + c.$$

Proof. By the definition of expectation,

$$\begin{aligned}
\mathrm{E}[X+c] &= \sum_{\text{all } x} (x+c)P(X+c = x+c) \\
&= \sum_{\text{all } x} xP(X=x) + \sum_{\text{all } x} cP(X=x) \\
&= \mathrm{E}[X] + c\sum_{\text{all } x} P(X=x) = \mathrm{E}[X] + c.
\end{aligned}$$

The last line follows from the usual probability properties. Note $\mathrm{E}[c] = c$ for all constants. □

Theorem 2. *For any random variable X and constant c,*

$$\mathrm{E}[cX] = c\,\mathrm{E}[X].$$

Proof. By the definition of expectation,

$$\begin{aligned}
\mathrm{E}[cX] &= \sum_{\text{all } x} (cx)P(cX = cx) \\
&= c\sum_{\text{all } x} xP(X=x) = c\,\mathrm{E}[X].
\end{aligned}$$

□

The first theorem shows that translating all variables by a constant also translates the expected value by the same constant. This is intuitive, since if all variables are translated by a constant, the central or mean value should also be translated by the constant. The second theorem shows that scaling the values of a random variable by a constant c also scales the expected value by c.

Example 2. What is the expected dice roll of a fair six-sided die labeled on each face with the numbers 5 through 10? What if it is labeled on each face with the first six positive multiples of 5?

Solution. We know the expected value of a regular six-sided die is $\mathrm{E}[X] = 3.5$. The first part asks for $\mathrm{E}[X+4] = 3.5 + 4 = \boxed{7.5}$. The second part asks for $\mathrm{E}[5X] = 5 \cdot 3.5 = \boxed{17.5}$.

Theorem 3 (Linearity of Expectation)**.** *Let X and Y be random variables. We have*

$$\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y].$$

More generally, let $X_1, \dots, X_n$ be random variables and $c_1, \dots, c_n$ be constants. We have

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

Linearity of expectation holds even when the random variables being summed are not independent. It is an extremely useful result for calculating expectation of recursive processes.

Example 3. A fair six-sided die is rolled repeatedly until three sixes are rolled consecutively. What is the expected number of rolls?

Solution. Let X_n be a random variable that represents the the number of rolls required to get n consecutive sixes. In order to get n consecutive sixes, there must first be $n-1$ consecutive sixes. Then, the n^{th} consecutive six will occur on the next roll with probability $\frac{1}{6}$, or the process will start over on the next roll with probability $\frac{5}{6}$.

$$\mathrm{E}[X_n] = \mathrm{E}\left[\frac{1}{6}(X_{n-1}+1) + \frac{5}{6}(X_{n-1}+1+X_n)\right] = \mathrm{E}\left[X_{n-1}+1+\frac{5}{6}X_n\right].$$

By linearity of expectation, $\mathrm{E}[X_n] = \mathrm{E}[X_{n-1}] + 1 + \frac{5}{6}\mathrm{E}[X_n]$. Solving yields $\mathrm{E}[X_n] = 6\,\mathrm{E}[X_{n-1}] + 6$. Note that $\mathrm{E}[X_1] = 6$ from the expectation of the geometric distribution, so $\mathrm{E}[X_2] = 6 \cdot 6 + 6 = 42$ and $\mathrm{E}[X_3] = 6 \cdot 42 + 6 = \boxed{258}$. This recurrence can also be solved explicitly.

3 Conditional Expectation

Recall that that many probabilities change depending on prior information. This can be encoded with **conditional probability**, and can be calculated utilizing the definition of conditional probability and **Bayes' theorem**. The properties of conditional probability also apply to **conditional distributions** and **conditional expectation**.

Definition 2. Let X and Y be discrete random variables. The expected value of X given that the event $Y = y$ has occurred is defined as

$$\mathrm{E}[X|Y=y] = \sum_{\text{all } x} xP(X=x|Y=y).$$

Example 4. What is the expected number of heads in 5 flips of a fair coin given that the number of heads flipped is greater than 2?

Solution. Let X represent the number of heads in 5 flips. Note that it is binomially distributed. We are conditioning on the event $X > 2$. Recall that conditional probability is defined as $P(X=x|Y=y) = \dfrac{P(X=x \cap Y=y)}{P(Y=y)}$. Thus, our desired sum is

$$\mathrm{E}[X|X>2] = 3 \cdot \frac{P(X=3)}{P(X>2)} + 4 \cdot \frac{P(X=4)}{P(X>2)} + 5 \cdot \frac{P(X=5)}{P(X>2)}.$$

Note we ignore the $X = 1$ and $X = 2$ cases because their probabilities are 0 in this conditional distribution. Using the binomial distribution, we have $P(X=3) = \frac{5}{16}$, $P(X=4) = \frac{5}{32}$, and $P(X=5) = \frac{1}{32}$. Thus, $P(X>2) = \frac{1}{2}$. Plugging this in yields $\mathrm{E}[X|X>2] = \boxed{\frac{55}{16}}$.

3.1 Problems

1. A fair coin is tossed repeatedly until 5 consecutive heads occur. What is the expected number of coin tosses?

2. The casino *Magnicifecto* was having difficulties attracting its hotel guests down to the casino floor. The empty casino prompted management to take drastic measures, and they decided to forgo the house cut. They decided to offer an "even value" game. Whatever bet size the player places (say $\$A$), there is a 50% chance that he will get $+\$A$ and a 50% chance he will get $-\$A$. They felt that since the expected value of every game is 0, they should not be making or losing money in the long run.

   Scrooge, who was on vacation, decided to exploit this even value game. He has an infinite bankroll (money) and decides to play the first round (entire series) in the following manner:

   - First make a bet of \$10.
   - If he wins, keep the earnings and leave.
   - Each time that he loses, he doubles the size of his previous bet and plays again.

   What is the expected value of Scrooge's (total) winnings?

3. Call the vertices of a regular tetrahedron A, B, C, and D. A bug starts on vertex A. Every second it randomly walks along one edge to another vertex. What is the expected value of the number of seconds it will take for it to reach the vertex D?

4. Brooke has two boxes. The left box contains 1 red ball and 2 blue balls while the right box contains 3 red balls and 2 blue balls. In a game, Brooke has to randomly and simultaneously pick up one ball from each box using both hands. If Brooke gets a blue ball from the left box, she'll be rewarded \$4, but she'll lose \$5 if she gets a red. On the other hand, a blue ball drawn from the right box will reward \$8 but a red ball will lose \$7 dollars. What is the expected value of a single round of this game?

5. A bug starts on one vertex A of a regular dodecahedron. A second vertex adjacent to the one it starts on is called B. Every second, it randomly walks along one edge to another vertex. What is the expected number of seconds it will take for the bug to reach the vertex B?

   (*Hint*: Every second it chooses randomly among the three edges available to it, including the one it might have just walked along. Recall a regular dodecahedron has 12 faces that are regular pentagons, 30 edges, and 20 vertices. Each vertex is connected to 3 others.)

6. A fair coin is flipped several times. Find the expected number of flips needed to make in order to see the pattern TXT, where T is tails, and X is either heads or tails.

   (*Hint*: We want the pattern TTT or THT. That is, one will keep flipping until one of these two patterns is seen, then the number of flips made is counted.)

7. A bug starts on one vertex of a regular icosahedron. Every second it randomly walks along one edge to another vertex. What is the expected value of the number of seconds it will take for it to reach the vertex opposite to the original vertex he was on?

   (*Hint*: Every second it chooses randomly among the five edges available to him, including the one it might have just walked along. Recall a regular icosahedron has 20 faces that are equilateral triangles, 30 edges, and 12 vertices. Each vertex is connected to 5 others.)

4 Fun with Cycles and Permutations

In group theory, a **cyclic permutation**, or **cycle**, is a permutation of the elements of some set X which maps the elements of some subset $S \subseteq X$ to each other in a cyclic fashion, while **fixing** (that is, mapping to themselves) all other elements of X. If S has k elements, the cycle is called a **k-cycle**.

For example, given $X = \{1, 2, 3, 4\}$, the permutation that sends $1 \to 3$, $3 \to 2$, $2 \to 4$ and $4 \to 1$ (so $S = X$) is a 4-cycle.

The permutation that sends $1 \to 3$, $3 \to 2$, $2 \to 1$ and $4 \to 4$ (so $S = \{1, 2, 3\}$ and 4 is a fixed element) is a 3-cycle.

On the other hand, the permutation that sends $1 \to 3$, $3 \to 1$, $2 \to 4$ and $4 \to 2$ is not a cyclic permutation because it separately permutes the pairs $\{1, 3\}$ and $\{2, 4\}$.

The set S is called the **orbit** of the cycle. Every permutation on finitely many elements can be decomposed into cycles on disjoint orbits. The cyclic parts of a permutation are cycles, thus the second example is composed of a 3-cycle and a 1-cycle (or fixed point) and the third is composed of two 2-cycles.

Definition 3. A permutation is called a cyclic permutation if and only if it has a single nontrivial cycle (a cycle of length > 1).

Example 5. Let a random permutation of $X = \{1, 2, \ldots, n\}$. What is the expected number of cycles in any such permutation?

Solution. We first find the expected number of fixed points for a random permutation. Let p_i be the number of permutations that fix position i. There are $(n-1)!$ of these, as the remaining $n-1$ non-fixed elements are free to permute. Thus, all permutations have a total of $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} (n-1)! = n(n-1)!$ fixed points. Since there are $n!$ permutations, the average permutation has $\frac{n(n-1)!}{n!} = 1$ fixed point. Note that a fixed point can be considered a cycle of length 1. Counting the 2-cycles is a similar process. One has to fix i in any of the n positions and j in a remaining $n-1$ positions. Then there are $(n-2)!$ permutations of the remaining elements, but we must divide by the cyclical orderings of our 2-cycle. This yields $\sum_{j=1}^{n-1} \sum_{i=1}^{n} p_{i,j} = \sum_{j=1}^{n-1} \sum_{i=1}^{n} \frac{(n-2)!}{2} = \frac{n(n-1)(n-2)!}{2}$ with an average of $\frac{n(n-1)(n-2)!}{2n!} = \frac{1}{2}$ 2-cycles. A similar calculation gives $\frac{1}{3}$ 3-cycles and so on.
We can generalize and say that the expected number of cycles of length k is $\frac{1}{k}$. Thus, the expected number of cycles of any length is

$$\sum_{k=1}^{n} \frac{1}{k} = \boxed{H_n}.$$

In Example 5, H_n is the general notation for the sum of the reciprocals of the first n natural numbers. It is known as the n^{th} **harmonic number**. For example, the expected number of cycles of a random permutation of $X = \{1, 2, 3\}$ is $1 + \frac{1}{2} + \frac{1}{3} = \frac{11}{6} = H_3$. This can be verified directly by counting. There are 2 permutations of type (abc) with one cycle, 3 permutations of type $(a)(bc)$ with two cycles, and 1 permutation of type $(a)(b)(c)$ with three cycles, for a total of $2 \cdot 1 + 3 \cdot 2 + 1 \cdot 3 = 11$ cycles over all $3! = 6$ permutations.

Note. When counting permutations according to their number of cycles, the **Stirling numbers of the first kind** often arise. These numbers are closely related to the coefficients 2, 3, and 1 in the above calculation. Stirling and harmonic numbers both appear frequently in mathematics.

5 Additional Worked Examples

Example 6 (Coupon Collector's Problem)**.** There are n distinct coupons, and a coupon collector wants to collect at least one of each. His only method of doing so is to buy random coupons one at a time. All n coupons are equally likely to be purchased at any given time. What is the expected number of purchases before the collector has at least one of each?

Solution. One can consider this process as sampling from the discrete uniform distribution with n elements with replacement (where duplicates are allowed). We wish to calculate the average number of samples before every element has been "seen." Let X_i be a random variable that represents the number of purchases until the i^{th} unique coupon is bought. Note that X_i is geometrically distributed. With the geometric distribution, we can first see that $P(X_1 = 1) = 1 = \frac{n}{n}$ and $\mathrm{E}[X_1] = \frac{1}{P(X_1=1)} = 1 = \frac{n}{n}$, as the collector is guaranteed to get a unique coupon on his first purchase. After this first coupon is collected, we can see that $P(X_2 = 1) = \frac{n-1}{n}$, as the probability of buying a unique coupon decreases now that one has been "seen." Thus, $\mathrm{E}[X_2] = \frac{n}{n-1}$. We can continue in this fashion and obtain:

$$P(X_i = 1) = \frac{n-i+1}{n}, \quad \mathrm{E}[X_i] = \frac{n}{n-i+1}.$$

We wish to calculate $\mathrm{E}[X_1 + X_2 + \ldots + X_n]$. By linearity of expectation, this becomes

$$\begin{aligned}\mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{E}[X_i] &= \frac{n}{n} + \frac{n}{n-1} + \ldots + \frac{n}{1} \\ &= n\left(\frac{1}{n} + \frac{1}{n-1} + \ldots + 1\right) = \boxed{nH_n}.\end{aligned}$$

Example 7 (Extended Birthday Paradox)**.** What is the average number of people required to find a pair with the same birthday? Assume all 365 days are equally likely to be born on and ignore leap years.

Solution. If we consider the 365 days as a discrete uniform distribution, then we wish to calculate the expected number of samples uniformly, with replacement, and from a population of days with size 365 required for the first repeated sampling of some day. This can be calculated as $1 + Q(n)$, where

$$\begin{aligned}Q(n) = \sum_{i=1}^{n} \frac{n!}{(n-1)!n^i} &= 1 + \frac{n-1}{n} + \frac{(n-1)(n-2)}{n^2} + \ldots + \frac{(n-1)(n-2)\ldots 1}{n^{n-1}} \\ &\approx \sqrt{\frac{\pi n}{2}} - \frac{1}{3} + \frac{1}{12}\sqrt{\frac{\pi}{2n}} - \frac{4}{135n} + \ldots\end{aligned}$$

When $n = 365$, we calculate $1 + Q(365) \approx \boxed{24.62}$.

Note. The usual birthday paradox asks for the median number of days. The asymptotic expansion on the last line is due to **Ramanujan**. Example 7 is relevant to several hashing algorithms analyzed by **Donald Knuth** in his book *The Art of Computer Programming*.

Example 8. Random real numbers are repeatedly observed from the Uniform(0,1) distribution. At every iteration, the observation is added to the total sum of all previous samples. Continue observing and adding until the total sum of all the numbers exceeds 1. Find the expected number of observations.

Solution. Let X_i be a random variable representing the i^{th} observation from the Uniform(0,1) distribution. We denote by N the number of X_i we need to add for the sum to exceed 1. We first find the distribution (probability mass function) of N. The easiest way to do this is by computing $P(N > n)$ for $n = 1, 2, \ldots$. Once we know this, we can compute $P(N = n) = P(N > n-1) - P(N > n)$. The event that $N > n$ corresponds to the sum of the first n uniform random variables not exceeding 1. The probability of this event is therefore $P(N > n) = P(X_1 + \ldots + X_n < 1)$. To calculate this probability, envision n orthogonal axes from 0 to 1 in n-dimensional space. We wish to calculate a geometric probability that corresponds to the n-volume of this n-dimensional **simplex**. That is, we wish to find the n-volume of the set

$$\left\{(X_1, \ldots, X_n) : \sum_{i=1}^{n} X_i \leq 1,\ X_1, \ldots, X_n > 0\right\}.$$

To visualize this, when $n = 1$ in one dimension, the set is simply a line segment with length 1. When $n = 2$ in two dimensions, the set is a right isosceles triangle with sides of length 1 and area $\frac{1}{2}$. When $n = 3$, the set is a right triangular pyramid (tetrahedron) with side lengths 1 and volume $\frac{1}{6}$. In general, the n-volume of the n-simplex is $\frac{1}{n!}$. The probability is simply the ratio of these n-volumes to the n-volume of the unit hypercube, which is 1. Thus, $P(N > n) = \frac{1}{n!}$. We conclude that $P(N = n) = P(N > n-1) - P(N > n) = \frac{1}{(n-1)!} - \frac{1}{n!} = \frac{n-1}{n!}$. This gives us the probability mass function. We finally calculate

$$\mathrm{E}[N] = \sum_{n=1}^{\infty} nP(N = n) = \sum_{n=1}^{\infty} \frac{n(n-1)}{n!} = \sum_{n=0}^{\infty} \frac{1}{n!} = e.$$

Thus, we expect e trials before exceeding 1 and $\boxed{1 + e}$ trials to exceed 1.

5.1 Competition Problems

1. [AMC 12A 2002] Tina randomly selects two distinct numbers from the set $\{1, 2, 3, 4, 5\}$, and Sergio randomly selects a number from the set $\{1, 2, \ldots, 10\}$. What is the probability that Sergio's number is larger than the sum of the two numbers chosen by Tina?

2. [HMMT 2010] Indecisive Andy starts out at the midpoint of the 1-unit-long segment HT. He flips 2010 coins. On each flip, if the coin is heads, he moves halfway towards endpoint H, and if the coin is tails, he moves halfway towards endpoint T. After his 2010 moves, what is the expected distance between Andy and the midpoint of HT?

3. [HMMT 2010] Let $\triangle ABC$ be a triangle with $AB = 8$, $BC = 15$, and $AC = 17$. Point X is chosen at random on line segment AB. Point Y is chosen at random on line segment BC. Point Z is chosen at random on line segment CA. What is the expected area of triangle $\triangle XYZ$?

4. [HMMT 2015] Consider an 8×8 grid of squares. A rook is placed in the lower left corner, and every minute it moves to a square in the same row or column with equal probability (the rook must move; i.e. it cannot stay in the same square). What is the expected number of minutes until the rook reaches the upper right corner?

5. [HMMT 2015] Consider a 10×10 grid of squares. One day, Daniel drops a burrito in the top left square, where a wingless pigeon happens to be looking for food. Every minute, if the pigeon and the burrito are in the same square, the pigeon will eat 10% of the burritos original size and accidentally throw it into a random square (possibly the one it is already in). Otherwise, the pigeon will move to an adjacent square, decreasing the distance between it and the burrito. What is the expected number of minutes before the pigeon has eaten the entire burrito?

6. [HMMT 2018] A permutation of $\{1, 2, ..., 7\}$ is chosen uniformly at random. A partition of the permutation into contiguous blocks is correct if, when each block is sorted independently, the entire permutation becomes sorted. For example, the permutation $(3, 4, 2, 1, 6, 5, 7)$ can be partitioned correctly into the blocks $[3, 4, 2, 1]$ and $[6, 5, 7]$, since when these blocks are sorted, the permutation becomes $(1, 2, 3, 4, 5, 6, 7)$. Find the expected value of the maximum number of blocks into which the permutation can be partitioned correctly.

7. [SMT 2018] Morris plays a game using a fair coin. He starts with \$2 and proceeds using the following rules:

   - If Morris flips a heads, he gains \$2.
   - If Morris flips a tails, he loses half his money.
   - If Morris flips two tails in a row, the game ends (but he doesn't lose any more money)

   For instance, if Morris flips the sequence $THTT$, he will end up with \$1.50. What is the expected amount of money in dollars Morris will have after the game ends?

5.2 Real-World Interview Problems

1. [Citadel Investment Group] What is the expected number of draws from a standard deck of 52 cards until an ace is drawn?

2. [Jump Trading] Aaron makes an observation from the Uniform(0,1) distribution. Then Brooke repeatedly observes from the same distribution until she obtains a number higher than Aaron's. How many observations is she expected to make?

3. [Tower Trading Group] Find the expected number of cycles of length greater than $\frac{n}{2}$ in a random permutation of $\{1, \ldots, n\}$.

4. [Five Rings Capital] Brooke is betting on a fair coin flip. She can bet any percentage of the \$100. If she wins, she gains 1.2 times her bet (and her bet back), but if she loses, she loses her bet. What is her optimal bet size (proportion of her wealth) to maximize long-run expected earnings?