

CS30 (Discrete Math in CS), Summer 2021 : Lecture 17

Topic: Probability: Expectation

Disclaimer: These notes have not gone through scrutiny and in all probability contain errors.

Please discuss in Piazza/email errors to deeparnab@dartmouth.edu

- **Linearity of Expectation.** This is one of the most powerful equations in all of probability. Literally. It states the following. It literally has a four line proof.

Theorem 1. For any two random variables X and Y , let $Z := X + Y$. Then,

$$\mathbf{Exp}[Z] = \mathbf{Exp}[X] + \mathbf{Exp}[Y]$$

Proof.

$$\begin{aligned}\mathbf{Exp}[Z] &= \sum_{\omega \in \Omega} Z(\omega) \mathbf{Pr}[\omega] && \text{Definition of Expectation} \\ &= \sum_{\omega \in \Omega} (X(\omega) + Y(\omega)) \mathbf{Pr}[\omega] && \text{Definition of } Z \\ &= \sum_{\omega \in \Omega} X(\omega) \mathbf{Pr}[\omega] + \sum_{\omega \in \Omega} Y(\omega) \cdot \mathbf{Pr}[\omega] && \text{Distributivity} \\ &= \mathbf{Exp}[X] + \mathbf{Exp}[Y] && \text{Definition of Expectation}\end{aligned}$$

□

As a corollary, by applying the above again and again $k - 1$ times, we get:

Theorem 2. For any k random variables X_1, X_2, \dots, X_k ,

$$\mathbf{Exp}\left[\sum_{i=1}^k X_i\right] = \sum_{i=1}^k \mathbf{Exp}[X_i]$$

Examples of applications.

1. We roll two fair dice. X takes the value of the sum. In this case, $X = Y + Z$ where Y, Z are random variables of the kind from the previous bullet point.
Tailor-made application. $\mathbf{Exp}[Y] = \mathbf{Exp}[Z] = 3.5$, the expected value of a single roll of a die. Thus, $\mathbf{Exp}[X] = \mathbf{Exp}[Y + Z] = 7$ by linearity of expectation.
2. We have a biased coin which lands heads with probability p . We toss it 100 times. Let Z be the number of heads we see. What is $\mathbf{Exp}[Z]$? Note that earlier we had the question for $p = 0.5$.

Remark: Try doing this the “first-principle” way. That is, for each $0 \leq k \leq 100$, figure out the probability $\mathbf{Pr}[X = k]$ (that is, the probability we get exactly k heads), and then sum $\sum_{k=0}^{100} k \cdot \mathbf{Pr}[X = k]$. Please try it; feel the sweat needed to do this. It will make you appreciate the next three lines more!

Define new random variables; define X_i to take the value 1 if the i th toss is heads, and 0 otherwise. Note, $X = X_1 + X_2 + \dots + X_{100}$. Note, $\mathbf{Exp}[X_i] = p$ (it is a Bernoulli random variable). Thus, linearity of expectation gives $\mathbf{Exp}[X] = 100p$.

3. n people checked in their hats, but on their way out, were handed back hats randomly. What is the expected number of people who get their correct hats?

Define X_i to be 1 if the i th person gets his or her back correctly. What is $\mathbf{Exp}[X_i]$? It is $1/n$; it is the probability that $\sigma(i) = i$ for a random ordering σ . This question was there in the UGP. Let $Z = \sum_{i=1}^n X_i$. Note, Z is the number of people who get their correct hats. By linearity of expectation, $\mathbf{Exp}[Z] = 1$.

4. In a party of n people there are some pairs of people who are friends, and some pairs who are not. In all there are m pairs of friends. The host randomly divides the party by taking each person and sending them left or right at the toss of a fair coin. How many friends, in expectation, are sundered apart?

Remark: In terms of graphs (which we will see soon) the question is: a graph with m edges is randomly partitioned. How many edges, in expectation, have endpoints in different parts?

For each pair of friends (u, v) , define X_{uv} which takes the value 1 if u and v are split, and takes the value 0 if u and v are not split. The probability u and v are split is $1/2$ (either u is sent left, v is sent right, or vice-versa – do you see this?). Thus, $\mathbf{Exp}[X_{uv}] = 1/2$. Define $Z = \sum_{(u,v): \text{ friends}} X_{uv}$; Z is the number of friends sent apart. $\mathbf{Exp}[Z] = \sum_{(u,v): \text{ friends}} \mathbf{Exp}[X_{uv}] = m/2$. In expectation, half the friendships are sundered apart.

5. In an ordering σ of $(1, 2, \dots, n)$, an inversion is a pair $i < j$ such that $\sigma(i) > \sigma(j)$. How many inversions, in expectation, are there in a random permutation?

Let σ be a random permutation. Define the indicator random variable X_{ij} for $i < j$, which takes the value 1 if $\sigma(i) > \sigma(j)$, and 0 otherwise. Note that $\Pr[X_{ij} = 1] = \frac{1}{2}$; there are equally many orderings with $\sigma(i) > \sigma(j)$ as $\sigma(i) < \sigma(j)$. Now note that $Z = \sum_{i=1}^n \sum_{j>i} X_{ij}$ is the number of inversions in σ . Thus, $\mathbf{Exp}[Z] = \sum_{i=1}^n \sum_{j>n} \mathbf{Exp}[X_{ij}] = \frac{1}{2} \cdot \frac{n(n-1)}{2}$.

- **Independent Random Variables.** Two random variables X and Y are independent, if for any $x \in \text{range}(X)$ and any $y \in \text{range}(Y)$,

$$\Pr[X = x, Y = y] = \Pr[X = x] \cdot \Pr[Y = y]$$

Examples:

- If we roll two dice, and X_1 and X_2 indicate the value of the rolls, then X_1 and X_2 are independent.
- If we have two independent events \mathcal{A} and \mathcal{B} , then their indicator random variables $\mathbf{1}_{\mathcal{A}}$ and $\mathbf{1}_{\mathcal{B}}$ are independent.
- Consider a random variable X taking value +1 if a toss of a coin is head, and –1 if its tails. Such random variables are called *Rademacher random variables*. Suppose we toss the coin twice and X_1 and X_2 are the corresponding random variables. Then X_1 and X_2 are independent.

A set of k random variables X_1, \dots, X_k are *mutually independent* if for any x_1, x_2, \dots, x_k with $x_i \in \text{range}(X_i)$, we have

$$\Pr[X_1 = x_1, X_2 = x_2, \dots, X_k = x_k] = \prod_{i=1}^k \Pr[X_i = x_i]$$

Theorem 3. If X and Y are two independent random variables, then

$$\mathbf{Exp}[XY] = \mathbf{Exp}[X] \cdot \mathbf{Exp}[Y]$$

Proof.

$$\begin{aligned} \mathbf{Exp}[XY] &= \sum_{x \in \text{range}(X), y \in \text{range}(Y)} (xy) \cdot \Pr[X = x, Y = y] && \text{Definition of Expectation} \\ &= \sum_{x \in \text{range}(X), y \in \text{range}(Y)} (xy) \cdot \Pr[X = x] \cdot \Pr[Y = y] && \text{Independence} \\ &= \left(\sum_{x \in \text{range}(X)} x \cdot \Pr[X = x] \right) \cdot \left(\sum_{y \in \text{range}(Y)} y \cdot \Pr[Y = y] \right) && \text{Algebra} \\ &= \mathbf{Exp}[X] \cdot \mathbf{Exp}[Y] && \text{Definition of Expectation} \end{aligned}$$

□

Of course, there is no need to stick to two random variables. The theorem easily generalizes (do you see how?) to mutually independent random variables as follows.

Theorem 4. If X_1, X_2, \dots, X_k are mutually independent random variables, then

$$\mathbf{Exp} \left[\prod_{i=1}^k X_i \right] = \prod_{i=1}^k \mathbf{Exp}[X_i]$$

Examples.

- Let X_i and X_j be two independent Rademacher random variables. Recall, X_i takes +1 with probability 1/2 and -1 with probability 1/2. Then note (a) $\mathbf{Exp}[X_i] = \mathbf{Exp}[X_j] = 0$, (b) $\mathbf{Exp}[X_i \cdot X_i] = \mathbf{Exp}[X_j \cdot X_j] = 1$, and (c) $\mathbf{Exp}[X_i X_j] = \mathbf{Exp}[X_i] \cdot \mathbf{Exp}[X_j] = 0$. This is a very useful fact.
- Consider rolling a die n times, independently. Let Z be the random variable indicating the *product* of all the numbers seen. What is $\mathbf{Exp}[Z]$? To solve this, let X_i be the roll of the i th die. We know that $\mathbf{Exp}[X_i] = 3.5$ for all i . We also know X_1, X_2, \dots, X_n are all independent random variables. Thus, $\mathbf{Exp}[Z] = (3.5)^n$.

• **Some Famous Random Variables.**

1. **Bernoulli Random Variable.** A random variable $X \sim \text{Ber}(p)$ is a Bernoulli Random Variable with parameter p if $X = 1$ with probability p and 0 with probability $1 - p$. Basically, a coin toss where the probability of “heads” (with $X(\text{heads}) = 1$) is p instead of being $1/2$.

Recall, given any event \mathcal{E} , the indicator random variable $\mathbf{1}_{\mathcal{E}}$ takes the value 1 if the event occurs and 0 otherwise (more precisely, $\mathbf{1}_{\mathcal{E}}(\omega) = 1$ for $\omega \in \mathcal{E}$ and 0 otherwise); indicator random variables are Bernoulli random variables with parameter $p = \mathbf{Pr}[\mathcal{E}]$.

The expectation $\mathbf{Exp}[X]$ of $X \in \text{Ber}(p)$ is precisely p .

2. **Binomial Random Variable.** A random variable $X \sim \text{Bin}(n, p)$ is the number of “heads” seen when one tosses n coins independently, where each coin comes heads with probability p . In other words, X is a sum of n independent Bernoulli random variables with probability p . The “shape” of this random variable is precisely given by

$$\text{For integer } 0 \leq k \leq n, \quad \mathbf{Pr}[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}$$

The expectation, however, is easy to calculate using linearity of expectation : $\mathbf{Exp}[X] = np$ since it’s a sum of n Bernoulli’s and each Bernoulli has expectation p .

3. **Geometric Random Variable.** A random variable $X \sim \text{Geom}(p)$ with parameter p , is the number of *times* a coin whose probability of heads is p needs to be tossed before we see the first heads. This is an interesting random variable whose range is *unbounded* (unlike the previous two); that is, given any integer k , there is a finite probability that $X > k$. Indeed, here is the “shape” (see you understand this)

$$\text{For integer } k \geq 1, \quad \mathbf{Pr}[X = k] = \underbrace{(1 - p)^{k-1}}_{\text{the first } (k-1) \text{ tosses are tails}} \cdot \underbrace{p}_{\text{and the } k\text{th toss is heads}}$$

In the last lecture, we saw a random variable which counted the number of while-loops in a snippet of code. Check that it was an example of a geometric random variable. What was its parameter?

- **Expectation of a Geometric Random Variable.** One could calculate the expectation of $X \sim \text{Geom}(p)$ “directly” via a calculation of the form

$$\mathbf{Exp}[X] = \sum_{k=1}^{\infty} k \cdot (1 - p)^{k-1} \cdot p$$

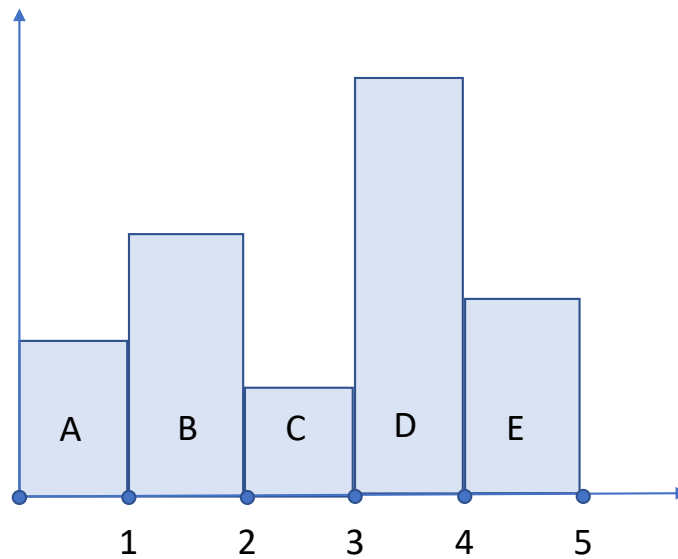
This “smells” of the sum of a geometric series. To remind every one, for any number $0 < a < 1$, the sum $\sum_{i=0}^{\infty} a^i = \frac{1}{1-a}$. If the above summation didn’t have the extra “ k ” multiplying the thing inside the summation, we could apply the above. However, one can still figure out the summations like above analytically. Instead, I want to take this opportunity to show another interpretation of expectation which is also useful to keep in mind.

Theorem 5. For any positive integer valued random variable X , one has

$$\mathbf{Exp}[X] = \sum_{k=1}^{\infty} \mathbf{Pr}[X \geq k]$$

Remark: One has the same for any random variable except if the range is not discrete the summation must be replaced by an integration.

Proof. The proof is best illustrated by a picture. In the picture below, we are plotting the histogram/“shape” of the random variable. The x -axis has the positive integers. The y -axis has the probabilities. Now note that the area of the rectangle whose bottom right corner is at the integer k is *precisely* $\Pr[X = k]$. This is because the gap between k and $k - 1$ is 1. In particular, the picture below shows only 5 rectangles presumably because the range of X is $\{1, 2, 3, 4, 5\}$ and the probability $\Pr[X = 1]$ is the area A , the probability $\Pr[X = 2]$ is the area B , and so on.



Note that the sum of the area $A + B + C + D + E$ is precisely 1. This is because the sum of the probabilities is 1. In particular, the area under this histogram is 1.

What is the expectation? Well, it is a weighted sum : $\mathbf{Exp} = A + 2B + 3C + 4D + 5E$. The key is to note that this sum can be thought of as taking all 5 rectangles, then taking all but the first, then taking all but the first two, and so on. More precisely,

$$\mathbf{Exp} = (A + B + C + D + E) + (B + C + D + E) + (C + D + E) + (D + E) + E$$

And now note the 5 different sums precisely correspond to $\Pr[X \geq k]$; for instance, $C + D + E = \Pr[X \geq 3]$. And this proves the theorem, at least pictorially. Now that we “feel” the proof, let’s go and write it precisely.

$$\begin{aligned}
\mathbf{Exp}[X] &= \sum_{k=1}^{\infty} k \cdot \mathbf{Pr}[X = k] && \text{Definition of Expectation} \\
&= \sum_{k=1}^{\infty} k \cdot (\mathbf{Pr}[X \geq k] - \mathbf{Pr}[X \geq k+1]) && \text{An algebraic manipulation} \\
&= \sum_{k=1}^{\infty} \mathbf{Pr}[X \geq k] \cdot (k - (k-1)) && \text{Rearranging} \\
&= \sum_{k=1}^{\infty} \mathbf{Pr}[X \geq k] && (1)
\end{aligned}$$

The algebraic manipulation uses the fact that the random variable is integer valued. □

Now, we can use this to evaluate the expectation of a Geometric random variable.

Theorem 6. Let $X \sim \text{Geom}(p)$ be a geometric random variable with parameter p . Then, $\mathbf{Exp}[X] = \frac{1}{p}$.

Proof. First let us observe that $\mathbf{Pr}[X \geq k] = (1-p)^{k-1}$. Why is this? Well the event $\{X \geq k\}$ is **equivalent** to the first $(k-1)$ tosses coming in tails. Make sure you see this. Therefore,

$$\mathbf{Exp}[X] = \sum_{k \geq 1} (1-p)^{k-1} = \frac{1}{p}$$

where we directly used the sum of geometric series. □