EE5137 : Stochastic Processes (Spring 2022)
Examples of Using the Linearity of Expectation

Vincent Y. F. Tan

January 21, 2022

1 Linearity of Expectation

In class, we saw that for two arbitrary random variables X and Y defined on the same probability space, the expectation of the sum is the sum of the expectations, i.e.,

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] \tag{1}$$

We emphasize that X and Y need not be independent. This statement admits a simple proof that you should know. Clearly, this statement can be extended to n random variables, i.e.,

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i]. \tag{2}$$

You should of course know how to prove (1) and (2). But the applications below are completely **optional reading**. They are just for "fun" (at least to me). The examples below illustrate the so-called "Probabilistic Method" [AS08], popularized by the great Hungarian mathematician Paul Erdős (Fig. 1). Here's a quote on Paul Erdős:

> Working with Paul Erdős was like taking a walk in the hills. Every time when I thought that we had achieved our goal and deserved a rest, Paul pointed to the top of another hill and off we would go.
> – Professor Fan Chung

2 Application 1: Coloring k-Uniform Hypergraphs

A *graph* $G = (V, E)$ consists of a set of vertices (or nodes) $V = \{1, 2, \ldots, n\}$ and a set of pairs of nodes or edges E. Thus, E is a subset of the set of distinct pairs of nodes in V; this is often denoted as $E \subset \binom{V}{2}$.

A *hypergraph* $H = (V, E)$ consists of a set of vertices (or nodes) $V = \{1, 2, \ldots, n\}$ and a set of tuples of nodes E. Each $e \in E$ is known as an *edge* or *hyperedge*. For example if $n = 10$, $E = \{\{1, 2, 3\}, \{3, 4, 5, 6\}, \{5, 6, 7, 8, 9\}, \{1, 10\}\}$. A hypergraph is called *k-regular* if each edge $e \in E$ contains precisely k nodes. For example, $H = (V, E)$ with $E = \{\{1, 2, 3\}, \{3, 4, 5\}, \{6, 7, 8\}, \{1, 2, 10\}\}$ is 3-regular.

An edge is *monochromatic* if all the nodes of that edge have the same color. We say that H is *2-colorable* if we can color each of the nodes of V using either red or blue such that no edge is monochromatic. Clearly, the more edges one has in a k-regular graph H, the more constraints one has and the less likely H is 2-colorable. The pertinent question is thus: What is the maximum number of edges of a k-uniform hypergraph such that it is 2-colorable?

The following theorem is due to Erdős (1963) [Erd63].

Theorem 1. *Every k-uniform hypergraph with less than 2^{k-1} edges is 2-colorable.*

Figure 1: Counter-clockwise from left: Erdős, Fan Chung, and her husband Ronald Graham, Japan 1986

The proof involves an appropriate construction of a probability space, a collection of random variables, and the linearity of expectation. As the reader can see, the linearity of expectation, though simple, is immensely powerful in establishing this non-trivial result.

Even though the proof is relatively simple, the result in Theorem 1 is somewhat tight. Erdős (1964) [Erd64] showed that there exists a k-uniform hypergraph with $O(2^k \cdot k^2)$ hyperedges which is not 2-colorable. So the base 2 is sharp.

Proof. Let $H = (V, E)$ be a k-uniform hypergraph with $< 2^{k-1}$ edges. Color each node in V red with probability 1/2 and blue with probability 1/2. This defines a probability space/model. For each edge e, define the random variable

$$X_e = \begin{cases} 1 & e \text{ is monochromatic} \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

Consider the random variable

$$X = \sum_{e \in E} X_e. \tag{4}$$

X represents the total number of monochromatic edges in the randomly colored k-uniform hypergraph H. By the linearity of expectation,

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{e \in E} X_e\right] = \sum_{e \in E} \mathbb{E}[X_e]. \tag{5}$$

Now, we need to calculate $\mathbb{E}[X_e]$. By the definition of the expectation

$$\mathbb{E}[X_e] = 1 \times \Pr(e \text{ is monochromatic}) + 0 \times \Pr(e \text{ is not monochromatic}) = \Pr(e \text{ is monochromatic}). \tag{6}$$

But the probability of an edge being monochromatic is exactly $2/2^k = 2^{-k+1}$ (note that being monochromatic means that e can be all red or all blue). Hence,

$$\mathbb{E}[X] = (\text{number of edges in } H) \times 2^{-k+1} < 2^{k-1} \cdot 2^{-k+1} = 1. \tag{7}$$

Note that X can only take on values $0, 1, 2, \ldots$. So the expectation is

$$\mathbb{E}[X] = p_X(0) \times 0 + p_X(1) \times 1 + p_X(2) \times 2 + \ldots \tag{8}$$

Since $\mathbb{E}[X] < 1$, necessarily $p_X(0) > 0$; otherwise if $p_X(0) = 0$, the above equation means that $\mathbb{E}[X] \geq 1$. This means that the probability that X has 0 monochromatic edges is strictly positive. Hence, there must be a coloring of the hypergraph H with no monochromatic edges, meaning that H is 2-colorable. $\square$

3 Application 2: Unbalancing Lights

Theorem 2. *Let $a_{ij} = \pm 1$ for $1 \le i, j \le n$. Then there exists $x_i, y_j = \pm 1$ for $1 \le i, j \le n$ such that*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i y_j \ge \left(\sqrt{\frac{2}{\pi}} + o(1) \right) n^{3/2} \tag{9}$$

The interpretation of this result is as follows. Let an $n \times n$ array of lights be given, each either on ($a_{ij} = +1$) or off ($a_{ij} = -1$). Suppose that for each row and column there is a switch so that the switch is pulled ($x_i = -1$ for row i and $y_j = -1$ for column j) all of the lights in that line are "switched": on to off and off to on. Then for any initial configuration, it is possible to perform switches so that the number of lights on minus the number of lights off is at least $(\sqrt{2/\pi} + o(1))n^{3/2}$.

Proof. Surprisingly, linearity of expectation is sufficient to prove this result. First forget the x's. Let $y_1, \ldots, y_n = \pm 1$ be selected independently and uniformly and set

$$R_i := \sum_j a_{ij} y_j, \quad \text{and} \quad R := \sum_i |R_i|. \tag{10}$$

Fix i. Regardless of a_{ij}, $a_{ij} y_j$ is ± 1 with probability $1/2$ and their values (over j) are independent; that is, whatever the i-th row is initially after random switching it becomes a uniformly distributed row, all 2^n possibilities equally likely. Thus R_i has distribution S_n—the distribution of the sum of n independent uniform $\{-1, +1\}$ random variables Z_k—and so

$$\mathbb{E}[|R_i|] = \mathbb{E}[|S_n|] = \mathbb{E}\left[\left| \sum_{k=1}^{n} Z_k \right| \right] = \left(\sqrt{\frac{2}{\pi}} + o(1) \right) \sqrt{n} \tag{11}$$

These asymptotics may be found by estimating S_n by $\sqrt{n} Z$, where Z is standard normal and using elementary calculus. In fact, we know that $\frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k$ converges in distribution to a Gaussian Z with zero mean and variance 1. Thus, $\mathbb{E}[|\frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k|] \to \mathbb{E}[|Z|] = \sqrt{2/\pi}$ (this follows from Fatou's lemma and uniform integrability). Applying linearity of expectation to R

$$\mathbb{E}[R] = \sum_i \mathbb{E}[|R_i|] = \left(\sqrt{\frac{2}{\pi}} + o(1) \right) n^{3/2}. \tag{12}$$

Thus by the probabilistic method, there exists $y_1, \ldots, y_n = \pm 1$ with R at least this value. Finally, pick x_i with the same sign as R_i so that

$$\sum_i x_i \sum_j a_{ij} y_j = \sum_i x_i R_i = \sum_i |R_i| = R \ge \left(\sqrt{\frac{2}{\pi}} + o(1) \right) n^{3/2}, \tag{13}$$

completing the proof. □

4 Application 3: Buffon's Needle

Suppose a needle of length 1 is dropped onto a floor with strips of wood 1 unit apart (see Fig. 2). What is the probability that the needle will land across two strips of wood? The traditional solution to this problem uses calculus, but we're going to show how to solve it with linearity of expectation instead!

First, we're going to argue that the expected number of times a needle crosses the strips of woods is directly proportional to the length l of the needle. Suppose that the needle is made up of n linear pieces

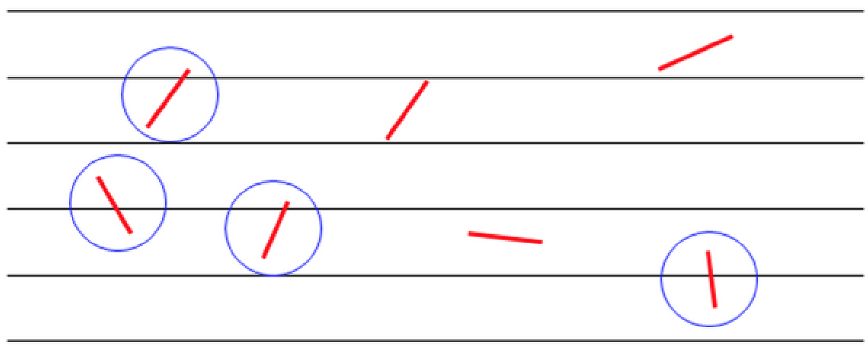

Figure 2: Buffon's needle

of equal length, and let X_i be an indicator variable on the i^{th} piece crossing two strips of wood. Then, the expected number of total crosses, by linearity of expectation, is

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n\mathbb{E}[X_1]. \tag{14}$$

Holding the length of these small pieces constant, we see that the expected number of times is proportional to the length of the needle.

Thus, as a function of the length l, the expected number of crossings is cl for some constant c. However, consider a circle with diameter 1 (so circumference π); with probability 1, this circle will intersect exactly 2 of the wood-crossings. Hence, $c\pi = 2$ so $c = 2/\pi$. If you're uncomfortable with the idea of a circle, consider approximating the circle by combining a bunch of very, very small linear segments.

For any needle (such as ours) which can intersect at most one wood-crossing, $\sum_{i=1}^{n} X_i$ is in fact an indicator variable on the event that the needle lands across two strips of wood, so the expected value is precisely the probability of this occurring. Thus, the probability of a crossing with a needle of length 1 is simply $2/\pi \approx 64\%$.

References

[AS08] N. Alon and J. H. Spencer. *The Probabilistic Method*. Wiley Interscience Series in Discrete Mathematics and Optimization, 3rd edition, 2008.

[Erd63] P. Erdős. On a combinatorial problem. *Nordisk Mat. Tidsskr*, 11:5–10, 1963.

[Erd64] P. Erdős. On a combinatorial problem II. *Acta Mathematica Academiae Scientiarum Hungaricae*, 15(3–4):455–447, 1964.