

## Expected Value I

The *expectation* or *expected value* of a random variable is a single number that tells you a lot about the behavior of the variable. Roughly, the expectation is the average value, where each value is weighted according to the probability that it comes up. Formally, the expected value (also known as the average or mean) of a random variable  $R$  defined on a sample space  $S$  is:

$$\text{Ex}(R) = \sum_{w \in S} R(w) \Pr\{w\}$$

To appreciate its significance, suppose  $S$  is the set of students in a class, and we select a student uniformly at random. Let  $R$  be the selected student's exam score. Then  $\text{Ex}(R)$  is just the class average—the first thing everyone wants to know after getting their test back! In the same way, expectation is usually the first thing one wants to determine about any random variable.

Let's work through an example. Let  $R$  be the number that comes up on a fair, six-sided die. Then the expected value of  $R$  is:

$$\begin{aligned} \text{Ex}(R) &= \sum_{k=1}^6 k \left(\frac{1}{6}\right) \\ &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\ &= \frac{7}{2} \end{aligned}$$

This calculation shows that the name “expected value” is a little misleading; the random variable might *never* actually take on that value. You can't roll a  $3\frac{1}{2}$  on an ordinary die!

Also note that the mean of a random variable is not the same as the median. The median is the midpoint of a distribution, i.e., the point  $x$  for which the random variable is at most  $x$  a half of the time, and greater than  $x$  the other half of the time.

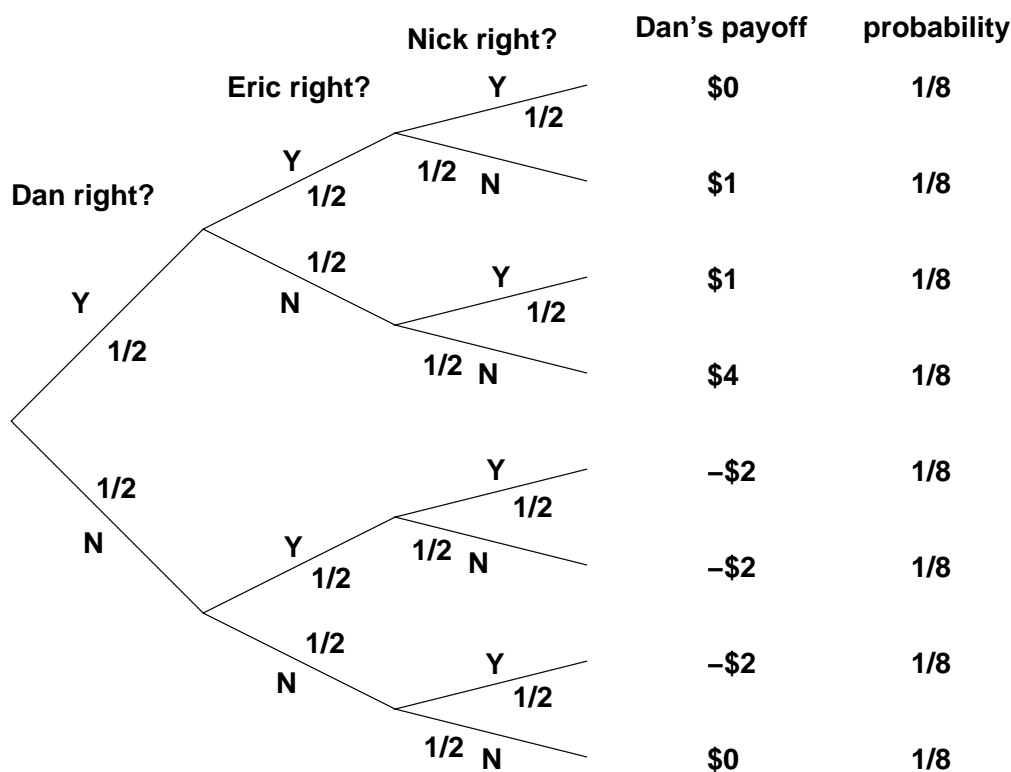
**Definition 1.** *The median of  $R$  is  $x \in \text{Range}(R)$  such that  $\Pr(R \leq x) \leq \frac{1}{2}$  and  $\Pr(R > x) < \frac{1}{2}$ .*

We note that sometimes the median of  $R$  is defined as the point for which  $\Pr(R \leq x) < \frac{1}{2}$  and  $\Pr(R > x) \leq \frac{1}{2}$ . For example, for a single roll of a 6-sided die, the median is 4. In this class we will not focus much attention on the median, but rather we will focus on the expected value, which is much more interesting and useful.

# 1 Betting on Coins

Dan, Eric, and Nick decide to play a fun game. Each player puts \$2 on the table and secretly writes down either “heads” or “tails”. Then one of them tosses a fair coin. The \$6 on the table is divided evenly among the players who correctly predicted the outcome of the coin toss. If everyone guessed incorrectly, then everyone takes their money back. After many repetitions of this game, Dan has lost a lot of money— more than can be explained by bad luck. What’s going on?

A tree diagram for this problem is worked out below, under the assumptions that everyone guesses correctly with probability  $1/2$  and everyone is correct independently.



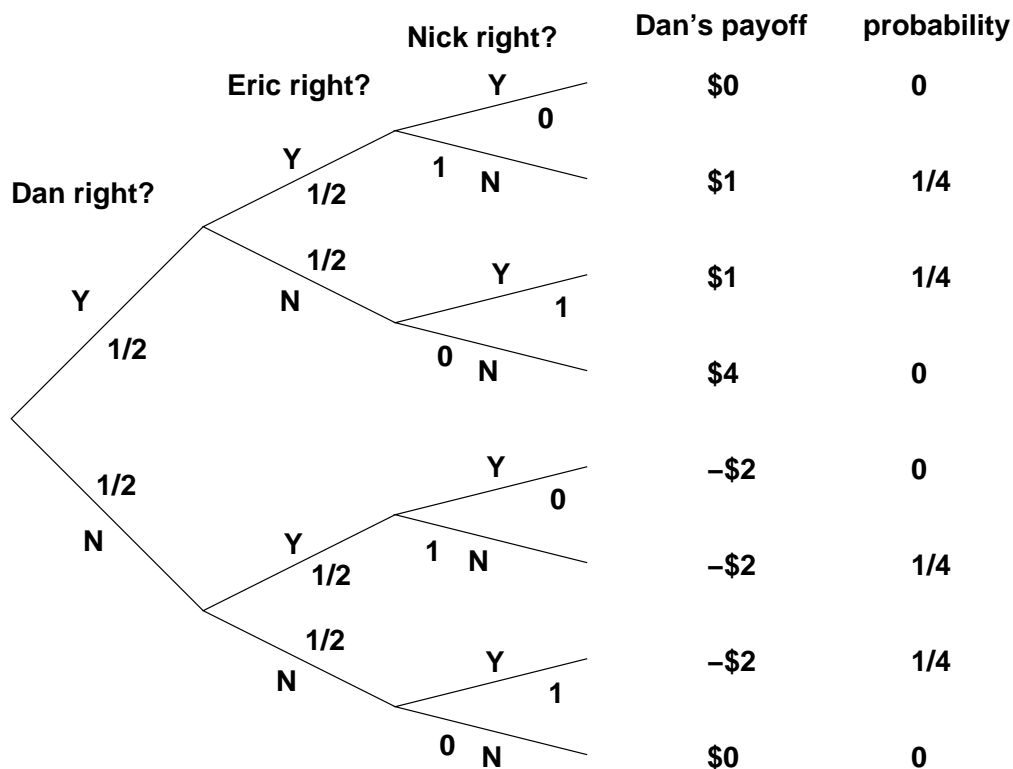
In the “payoff” column, we’re accounting for the fact that Dan has to put in \$2 just to play. So, for example, if he guesses correctly and Eric and Nick are wrong, then he takes all \$6 on the table, but his net profit is only \$4. Working from the tree diagram, Dan’s expected payoff is:

$$\begin{aligned} \text{Ex}(\text{payoff}) &= 0 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} + (-2) \cdot \frac{1}{8} + (-2) \cdot \frac{1}{8} + (-2) \cdot \frac{1}{8} + 0 \cdot \frac{1}{8} \\ &= 0 \end{aligned}$$

So the game perfectly fair! Over time, he should neither win nor lose money.

The trick is that Eric and Nick are collaborating; in particular, they always make *opposite* guesses. So our assumption that everyone is correct independently is wrong; actually the

events that Dan is correct and Eric is correct are mutually exclusive! As a result, Dan can never win all the money on the table. When he guesses correctly, he always has to split her winnings with someone else. This lowers his overall expectation, as the corrected tree diagram below shows:



From this revised tree diagram, we can work out Dan's actual expected payoff:

$$\begin{aligned}
 \text{Ex (payoff)} &= 0 \cdot 0 + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 4 \cdot 0 + (-2) \cdot 0 + (-2) \cdot \frac{1}{4} + (-2) \cdot \frac{1}{4} + 0 \cdot 0 \\
 &= -\frac{1}{2}
 \end{aligned}$$

So he loses an average of a half-dollar per game.

Similar opportunities for subtle cheating come up in many betting games. For example, a group of friends once organized a football pool where each participant would guess the outcome of every game each week relative to the spread. This may mean nothing to you, but the upshot is that everyone was effectively betting on the outcomes of 12 or 13 coin tosses each week. The person who correctly predicts the most coin tosses won a lot of money. The organizer, thinking in terms of the first tree diagram, swore up and down that there was no way to get an unfair "edge". But actually the number of participants was small enough that just two players betting oppositely could gain a substantial advantage!

Another example involves a former MIT professor of statistics, Herman Chernoff. State lotteries are the worst gambling games around because the state pays out only a fraction of

the money it takes in. But Chernoff figured out a way to win! Here are rules for a typical lottery:

- All players pay \$1 to play and select 4 numbers from 1 to 36.
- The state draws 4 numbers from 1 to 36 uniformly at random.
- The state divides 1/2 the money collected among the people who guessed correctly and spends the other half repairing the Big Dig.

This is a lot like our betting game, except that there are more players and more choices. Chernoff discovered that a small set of numbers was selected by a large fraction of the population— apparently many people think the same way (not on purpose as in the previous game with Angelina and Arvind), e.g., based on Manny’s batting average, or today’s date. It was as if the players were collaborating to lose! If any one of them guessed correctly, then they’d have to split the pot with many other players. By selecting numbers uniformly at random, Chernoff was unlikely to get one of these favored sequences. So if he won, he’d likely get the whole pot! Thus, in this case people unknowingly collaborate to lose! By analyzing actual state lottery data, he determined that he could win an average of 7 cents on the dollar (based on a profile of bets) this way!

There is another variation of the accidental collusion with a betting pool. For example, how many people ever participated in a Super Bowl betting pool where the goal was to get closest to the total number of points scored in the game?

Suppose the average Super Bowl has a total of 30 points scored and everyone knows this, and 50 people are in the pool. Then most people will guess around 30 points. Where should you guess? Well, you should guess just outside of this range because you get to cover a lot more ground and you don’t share the pot if you win. Of course, if you are in a pool with 6.042 students and they all know this strategy, then maybe you should guess 30 points after all.

## 2 Equivalent Definitions of Expectation

There are some other ways of writing the definition of expectation. Sometimes using one of these other formulations can make computing an expectation a lot easier. One option is to group together all outcomes on which the random variable takes on the same value.

**Theorem 1.**

$$\text{Ex}(R) = \sum_{x \in \text{range}(R)} x \cdot \Pr\{R = x\}$$

*Proof.* We'll transform the left side into the right. Let  $[R = x]$  be the event that  $R = x$ .

$$\begin{aligned}
 \text{Ex}(R) &= \sum_{w \in S} R(w) \Pr\{w\} \\
 &= \sum_{x \in \text{range}(R)} \sum_{w \in [R=x]} R(w) \Pr\{w\} \\
 &= \sum_{x \in \text{range}(R)} \sum_{w \in [R=x]} x \Pr\{w\} \\
 &= \sum_{x \in \text{range}(R)} \left( x \cdot \sum_{w \in [R=x]} \Pr\{w\} \right) \\
 &= \sum_{x \in \text{range}(R)} x \cdot \Pr\{R = x\}
 \end{aligned}$$

On the second line, we break the single sum into two. The outer sum runs over all possible values  $x$  that the random variable takes on, and the inner sum runs over all outcomes taking on that value. Thus, we're still summing over every outcome in the sample space exactly once. On the last line, we use the definition of the probability of the event  $[R = x]$ .  $\square$

**Corollary 2.** *If  $R$  is a natural-valued random variable, then:*

$$\text{Ex}(R) = \sum_{i=0}^{\infty} i \cdot \Pr\{R = i\}$$

There is another way to write the expected value of a random variable that takes on values only in the natural numbers,  $\mathbb{N} = \{0, 1, 2, \dots\}$ .

**Theorem 3.** *If  $R$  is a natural-valued random variable, then:*

$$\text{Ex}(R) = \sum_{i=0}^{\infty} \Pr\{R > i\}$$

*Proof.* Consider the sum:

$$\begin{array}{ccccccc}
 \Pr\{R = 1\} & + & \Pr\{R = 2\} & + & \Pr\{R = 3\} & + & \dots \\
 & & + & \Pr\{R = 2\} & + & \Pr\{R = 3\} & + \dots \\
 & & & & + & \Pr\{R = 3\} & + \dots \\
 & & & & & & + \dots
 \end{array}$$

The columns sum to  $1 \cdot \Pr\{R = 1\}$ ,  $2 \cdot \Pr\{R = 2\}$ ,  $3 \cdot \Pr\{R = 3\}$ , etc. Thus, the whole sum is equal to:

$$\sum_{i=0}^{\infty} i \cdot \Pr\{R = i\} = \text{Ex}(R)$$

Here, we're using Corollary 2. On the other hand, the rows sum to  $\Pr\{R > 0\}$ ,  $\Pr\{R > 1\}$ ,  $\Pr\{R > 2\}$ , etc. Thus, the whole sum is also equal to:

$$\sum_{i=0}^{\infty} \Pr\{R > i\}$$

These two expressions for the whole sum must be equal, which proves the theorem.  $\square$

Sometimes it is easier to calculate  $\sum_{i=0}^{\infty} \Pr\{R > i\}$  rather than  $\sum_{i=0}^{\infty} i \Pr\{R = i\}$ , so this theorem can be useful in this case. A good example is when you want to compute the mean time to failure in a system.

## 2.1 Mean Time to Failure

Let's look at a problem where one of these alternative definitions of expected value is particularly helpful. A computer program crashes at the end of each hour of use with probability  $p$ , if it has not crashed already. What is the expected time until the program crashes?

If we let  $R$  be the number of hours until the crash, then the answer to our problem is  $\text{Ex}(R)$ . This is a natural-valued variable, so we can use the formula:

$$\text{Ex}(R) = \sum_{i=0}^{\infty} \Pr\{R > i\}$$

We have  $R > i$  only if the system remains stable after  $i$  opportunities to crash, which happens with probability  $(1 - p)^i$ . Plugging this into the formula above gives:

$$\begin{aligned} \text{Ex}(R) &= \sum_{i=0}^{\infty} (1 - p)^i \\ &= \frac{1}{1 - (1 - p)} \\ &= \frac{1}{p} \end{aligned}$$

The closed form on the second line comes from the formula for the sum of an infinite geometric series where the ratio of consecutive terms is  $1 - p$ .

So, for example, if there is a 1% chance that the program crashes at the end of each hour, then the expected time until the program crashes is  $1/0.01 = 100$  hours. The general principle here is well-worth remembering: if a system fails at each time step with probability  $p$ , then the expected number of steps up to the first failure is  $1/p$ .

## 2.2 Making a Baby Girl

A couple really wants to have a baby girl. There is a 50% chance that each child they have is a girl, and the genders of their children are mutually independent. If the couple insists on having children until they get a girl, then how many baby boys should they expect first?

This is really a variant of the previous problem. The question, “How many hours until the program crashes?” is mathematically the same as the question, “How many children must the couple have until they get a girl?” In this case, a crash corresponds to having a girl, so we should set  $p = \frac{1}{2}$ . By the preceding analysis, the couple should expect a baby girl after having  $1/p = 2$  children. Since the last of these will be the girl, they should expect just 1 boy.

## 3 An Expectation Paradox

Next we’re going to look at a nasty example that you see a lot in experimental work. Suppose you are trying to estimate the average delay across a communications channel. So you set up an experiment to measure how long it takes to send a test packet from one end to the other and you run the experiment 100 times and record the latency each time. You assume there is some probability distribution function for the probability that a packet has a certain amount of delay.

Let  $D$  be the delay of a packet on the channel, and let  $f(x) = \Pr(D = x)$  be the probability density function for  $D$ . You expect  $f(x)$  to be a decreasing function since fewer packets are observed to have longer delays. Suppose from your 100 trials that you decide to assume that  $\Pr(D \geq i) = 1/i$  for  $i \geq 1$ . Suppose the unit of measurement here is milliseconds.

The problem is to find the expected delay. There are two ways to proceed. Often what you see people do in practice is to compute the average of the delay on 100 test packets. For this distribution there is a very good chance the average of 100 packet delays would be under 10 millisecond. If they are careful, they might even do another 100 test packets and compare the results.

A better way to approach this problem is to use the formula for the expectation.

**Claim 4.**  $\text{Ex}(D) = \infty$ .

*Proof.*

$$\begin{aligned} \text{Ex}(T) &= \sum_{i=1}^{\infty} \Pr\{D \geq i\} \\ &= \sum_{i=1}^{\infty} \frac{1}{i} \\ &= \infty \end{aligned}$$

The first  $n$  terms of this sum total to  $H_n$ , the  $n$ -th harmonic number, which is at least  $\ln n$ . Since  $\ln n$  goes to infinity as  $n$  goes to infinity, the expectation is infinite.  $\square$

What went wrong?

The approach of sampling the distribution function and taking the average of samples doesn't get you the expected value if the expected value is infinite! It is much better to figure out the probability density function and then evaluate the formula for the expectation. This mistake is made all the time in practice. We'll see later that this experimental approach does work well for finite distributions, but it is still better to use the formula.

## 4 Linearity of Expectation

Expected values obey a wonderful rule called *linearity of expectation*. This says that the expectation of a sum is the sum of the expectations.

**Theorem 5** (Linearity of Expectation). *For every pair of random variables  $R_1$  and  $R_2$ :*

$$\text{Ex}(R_1 + R_2) = \text{Ex}(R_1) + \text{Ex}(R_2)$$

*Proof.* Let  $S$  be the sample space.

$$\begin{aligned} \text{Ex}(R_1 + R_2) &= \sum_{w \in S} (R_1(w) + R_2(w)) \cdot \Pr\{w\} \\ &= \sum_{w \in S} R_1(w) \cdot \Pr\{w\} + \sum_{w \in S} R_2(w) \cdot \Pr\{w\} \\ &= \text{Ex}(R_1) + \text{Ex}(R_2) \end{aligned}$$

$\square$

Linearity of expectation generalizes to any finite collection of random variables by induction:

**Corollary 6.** *For any random variables  $R_1, R_2, \dots, R_k$ ,*

$$\text{Ex}(R_1 + R_2 + \dots + R_k) = \text{Ex}(R_1) + \text{Ex}(R_2) + \dots + \text{Ex}(R_k)$$

The reason linearity of expectation is so wonderful is that, unlike many other probability rules, *the random variables are not required to be independent*. This probably sounds like a “yeah, whatever” technicality right now. But when you give an analysis using linearity of expectation, someone will almost invariably say, “No, you're wrong. There are all sorts of complicated dependencies here that you're ignoring.” But that's the magic of linearity of expectation: you *can* ignore such dependencies!

## 4.1 Expected Value of Two Dice

What is the expected value of the sum of two fair dice?

Let the random variable  $R_1$  be the number on the first die, and let  $R_2$  be the number on the second die. At the start of these Notes, we showed that the expected value of one die is  $3\frac{1}{2}$ . We can find the expected value of the sum using linearity of expectation:

$$\begin{aligned}\text{Ex}(R_1 + R_2) &= \text{Ex}(R_1) + \text{Ex}(R_2) \\ &= 3\frac{1}{2} + 3\frac{1}{2} \\ &= 7\end{aligned}$$

Notice that we did *not* have to assume that the two dice were independent. The expected sum of two dice is 7, even if they are glued together! (This is provided that gluing somehow does not change weights to make the individual dice unfair.)

Proving that the expected sum is 7 with a tree diagram would be hard; there are 36 cases. And if we did not assume that the dice were independent, the job would be a nightmare!

## 4.2 The Hat-Check Problem

There is a dinner party where  $n$  men check their hats. The hats are mixed up during dinner, so that afterward each man receives a random hat. In particular, each man gets his own hat with probability  $1/n$ . What is the expected number of men who get their own hat?

Without linearity of expectation, this would be a very difficult question to answer. We might try the following. Let the random variable  $R$  be the number of men that get their own hat. We want to compute  $\text{Ex}(R)$ . By the definition of expectation, we have:

$$\text{Ex}(R) = \sum_{k=0}^{\infty} k \cdot \Pr(R = k)$$

Now we're in trouble, because evaluating  $\Pr(R = k)$  is a mess and we then need to substitute this mess into a summation. Furthermore, to have any hope, we would need to fix the probability of each permutation of the hats. For example, we might assume that all permutations of hats are equally likely.

Now let's try to use linearity of expectation. As before, let the random variable  $R$  be the number of men that get their own hat. The trick is to express  $R$  as a sum of indicator variables. In particular, let  $R_i$  be an indicator for the event that the  $i$ th man gets his own hat. That is,  $R_i = 1$  is the event that he gets his own hat, and  $R_i = 0$  is the event that he gets the wrong hat. The number of men that get their own hat is the sum of these indicators:

$$R = R_1 + R_2 + \cdots + R_n$$

These indicator variables are *not* mutually independent. For example, if  $n - 1$  men all get their own hats, then the last man is certain to receive his own hat. But, since we plan to use linearity of expectation, we don't have worry about independence!

Let's take the expected value of both sides of the equation above and apply linearity of expectation:

$$\begin{aligned}\text{Ex}(R) &= \text{Ex}(R_1 + R_2 + \cdots + R_n) \\ &= \text{Ex}(R_1) + \text{Ex}(R_2) + \cdots + \text{Ex}(R_n)\end{aligned}$$

All that remains is to compute the expected value of the indicator variables  $R_i$ . We'll use an elementary fact that is worth remembering in its own right:

**Fact 1.** *The expected value of an indicator random variable is the probability that the indicator is 1. In symbols:*

$$\text{Ex}(I) = \Pr\{I = 1\}$$

*Proof.*

$$\begin{aligned}\text{Ex}(I) &= 1 \cdot \Pr\{I = 1\} + 0 \cdot \Pr\{I = 0\} \\ &= \Pr\{I = 1\}\end{aligned}$$

□

So now we need only compute  $\Pr(R_i = 1)$ , which is the probability that the  $i$ th man gets his own hat. Since every man is as likely to get one hat as another, this is just  $1/n$ . Putting all this together, we have:

$$\begin{aligned}\text{Ex}(R) &= \text{Ex}(R_1) + \text{Ex}(R_2) + \cdots + \text{Ex}(R_n) \\ &= \Pr\{R_1 = 1\} + \Pr\{R_2 = 1\} + \cdots + \Pr\{R_n = 1\} \\ &= n \cdot \frac{1}{n} = 1.\end{aligned}$$

So we should expect 1 man to get his own hat back on average!

Notice that we did not assume that all permutations of hats are equally likely or even that all permutations are possible. We only needed to know that each man received his own hat with probability  $1/n$ . This makes our solution very general, as the next example shows.

### 4.3 The Chinese Appetizer Problem

There are  $n$  people at a circular table in a Chinese restaurant. On the table, there are  $n$  different appetizers arranged on a big Lazy Susan. Each person starts munching on the appetizer directly in front of him or her. Then someone spins the Lazy Susan so that everyone is faced with a random appetizer. What is the expected number of people that end up with the appetizer that they had originally?

This is just a special case of the hat-check problem, with appetizers in place of hats. In the hat-check problem, we assumed only that each man received his own hat with probability  $1/n$ . Beyond that, we made no assumptions about how the hats could be permuted. This problem is a special case because we happen to know that appetizers are cyclically shifted relative to their initial position. This means that either everyone gets their original appetizer back, or no one does. But our previous analysis still holds: the *expected* number of people that get their own appetizer back is 1.

The nice thing about solving problems with linearity of expectations is that you don't need to know very much about the underlying distribution to compute interesting facts about it. Because you don't use much information about the distribution, the calculations are also usually much easier using this method.

In fact, linearity of expectations provides a very general method for computing the expected number of events that will happen.

**Theorem 7.** *Given any collection of  $N$  events  $A_1, \dots, A_N \subseteq S$ , the expected number of events that will occur is  $\sum_{1 \leq i \leq N} \Pr(A_i)$ .*

For example,  $A_i$  could be the event that the  $i$ th man gets the right hat back. But in general, it could be any subset of the sample space, and we are asking for the expected number of events that will contain a random sample point.

*Proof.* Define  $R_i$  to be the indicator variable for  $A_i$ , where  $R_i(w) = 1$  if  $w \in A_i$ , and  $R_i(w) = 0$  if  $w \notin A_i$ . Let  $R = R_1 + R_2 + \dots + R_N$ . Then

$$\begin{aligned} \text{Ex}(R) &= \sum_i \text{Ex}(R_i) \\ &= \sum_i \Pr(R_i = 1) \\ &= \sum_i \sum_{w \in A_i} \Pr(w) \\ &= \sum_i \Pr(A_i). \end{aligned}$$

□

So whenever you are asked for the expected number of events that occur, all you have to do is sum the probabilities that each event occurs. Independence is not needed.

As a final example, suppose you flip  $N$  fair coins. Let  $R$  be the number of heads, and  $R_i$  the event that the  $i$ th coin is a head. Then  $\text{Ex}(R) = \sum_i \text{Ex}(R_i) = \frac{1}{2} + \frac{1}{2} + \dots + \frac{1}{2} = \frac{N}{2}$ .

This is the easy way. You could also solve this the hard way, assuming the coins are

independent. In this case,

$$\begin{aligned}\text{Ex}(R) &= \sum_{i=0}^N i \Pr(R = i) \\ &= \sum_{i=0}^N i \binom{N}{i} 2^{-N}\end{aligned}$$

It is not so obvious that this is  $\frac{N}{2}$ . In fact, our first approach using linearity of expectations gives a proof of the combinatorial identity  $\sum_{i=0}^N i \binom{N}{i} = N2^{N-1}$ . This method is much harder, and required independence of the random variables. Thus, on an exam, the first method, i.e., linearity of expectations, is a much more useful and handy way of solving these kinds of problems.