Lecture Notes

1 Expected Value

This lecture describes a very important characteristic of a random variable called the *expected value.*

Definition The expected value of a random variable R is denoted $\mathrm{Ex}(R)$ and defined:

$$\mathrm{Ex}(R) = \sum^{x \in \mathrm{range}(R)} x \cdot \Pr(R = x)$$

The expected value of a random variable is also sometimes called the *mean* or *average.* Intuitively, the expected value is the average of all possible values of a random variable, where a value is weighted according to the probability that it is attained.

If the image of a random variable R is not countable, then the summation above becomes an integral. We will not deal with this in 6.042.

Note: In the previous lecture, a random variable was defined as a function $R : S \mapsto \mathcal{R}$. That is, the range of a random variable was defined to be the real numbers. In this lecture, we relax this definition and consider random variables over a more limited range. In particular, we often consider random variables $R : S \mapsto \mathcal{N}$.

1.1 Median is Not the Mean

Expected value, average, and mean are the same thing, but median is entirely different. The median is defined below, but only to make the distinction clear.

Definition The *median* of a random variable R is the value x in the range of R such that:

$$\begin{aligned} \Pr(R < x) &\leq \frac{1}{2} \\ \text{and} \quad \Pr(R > x) &< \frac{1}{2} \end{aligned}$$

Sometimes the $\leq$ and $<$ are swapped in the two conditions above. We will not discuss the median further in this lecture.

1.2 Expected Value of One Die

Suppose we roll a fair, six-sided die. Let the random variable R be the number that comes up. We can compute the expected value of R directly from the definition of expected value.

$$\begin{aligned}
\mathrm{Ex}(R) &= \sum_{i=1}^{6} i \cdot \Pr(R = i) \\
&= \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} \\
&= 3\frac{1}{2}
\end{aligned}$$

The average value thrown on a fair die is $3\frac{1}{2}$.

1.3 Modified Carnival Dice

Let's look at a modified version of Carnival Dice. The player chooses a number from 1 to 6. He then throws three fair and mutually independent dice. He wins one dollar for *each* die that matches his number, and he loses one dollar if no die matches.

This is better than the original game where the player received one dollar if any die matched, and lost a dollar otherwise. At first glance the new game appears to be fair; after all, the player is now "justly compensated" if he rolls his number on more than one die. In fact, there is still another variant of Carnival Dice in which the payoff is \$2.75 instead of \$3 if all three dice match. In this case, the game appears fair except for the lost quarter in the rare case that all three dice match. This looks like a tiny, tolerable edge for the house.

Let's check our intuition by computing the expected profit of the player in one round of the \$3 variant of Carnival Dice. Let the random variable R be the amount of money won or lost by the player in a round. We can compute the expected value of R as follows:

$$\begin{aligned}
\mathrm{Ex}(R) &= -1 \cdot \Pr(\text{no dice match}) + 1 \cdot \Pr(\text{one die matches}) + \\
&\quad\; 2 \cdot \Pr(\text{two dice match}) + 3 \cdot \Pr(\text{three dice match}) \\
&= -1 \cdot \left(\frac{5}{6}\right)^3 + 1 \cdot 3 \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2 + 2 \cdot 3 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right) + 3 \cdot \left(\frac{1}{6}\right)^3 \\
&= \frac{-125 + 75 + 30 + 3}{216} \\
&= \frac{-17}{216}
\end{aligned}$$

Our intuition was wrong! Even with a \$3 payoff for three matching dice, the player can expect to lose $\frac{17}{216}$ of a dollar, or about 8 cents, in every round. This is still a horrible game for the player!

The \$2.75 variant is deceptive. One is tempted to believe that a player is shortchanged only a quarter in the rare case that all three dice match. This is a tiny amount. In fact, though, the player loses this tiny amount *in addition* to the comparatively huge 8 cents per game!

1.4 Another Way to Find Expected Values

There is a nice, alternative way to compute the expected value of a random variable with range $\mathcal{N}$.

Theorem 1.1 If R is a random variable with range $\mathcal{N}$, then

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

Proof. We will begin with the right-hand expression and transform it into $\operatorname{Ex}(R)$.

$$\begin{aligned}
\sum_{i=0}^{\infty} \Pr(R > i) &= \Pr(R > 0) + \Pr(R > 1) + \Pr(R > 2) + \ldots \\
&= \underbrace{\Pr(R = 1) + \Pr(R = 2) + \Pr(R = 3) + \ldots}_{=\Pr(R>0)} \\
&\quad + \underbrace{\Pr(R = 2) + \Pr(R = 3) + \ldots}_{=\Pr(R>1)} \\
&\quad + \underbrace{\Pr(R = 3) + \ldots}_{=\Pr(R>2)} \\
&= \Pr(R = 1) + 2 \cdot \Pr(R = 2) + 3 \cdot \Pr(R = 3) + \ldots \\
&= \sum_{i=0}^{\infty} i \cdot \Pr(R = i) \\
&= \operatorname{Ex}(R)
\end{aligned}$$

In the first step, the summation is rewritten with the $\ldots$ notation. In the second step, we rewrite each term $\Pr(R > i)$ as a series; if R is greater than i, then it must be one of $i+1$, $i+2$, etc. (This is where we use the fact that the range of R is $\mathcal{N}$.) Next, we sum all the resulting series, and then revert to Σ notation. The last summation is the definition of $\operatorname{Ex}(R)$. ∎

We can now compute the expected value of a random variable R by summing terms of the form $\Pr(R > i)$ instead of terms of the form $\Pr(R = i)$. Sometimes this makes a problem easier, as we will see in the next section. Remember, though, that the theorem only holds if the range of R is $\mathcal{N}$!

1.5 Mean Time to Failure

American astronaut David Wolf has just arrived on the Mir space station. Suppose that Mir's main computer has probability p of failing every hour, and assume that failures occur independently. (A failure is not catastrophic; the Mir computer is constantly on the blink.) How long can David expect to wait until the main computer fails?

Let the random variable R be the number of hours until the first failure. We want to compute the expected value of R. Since the range of R is $\mathcal{N}$, we can apply Theorem 1.1:

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

All that remains is to compute $\Pr(R > i)$, the probability that the first failure occurs sometime after hour i. We can compute this with the usual four-step method, though there is a twist!

Step 1: Find the Sample Space. We can regard the sample space as a set of infinite strings such as $WWFWFFW\ldots$. A W in the i-th position means that the main computer is working during hour i. An F in the i-th position means that the computer is down during hour i.

Step 2: Define Events of Interest. We are concerned with the event that $R > i$. This event consists of all outcomes with no F in the first i positions.

Step 3: Compute Outcome Probabilities. We want to compute the probability of a particular outcome, say the probability of the sequence $WWFWFFW\ldots$. There is a problem! The number of outcomes in this experiment is infinite and, in fact, uncountably infinite! We must skip computing the probability of an individual outcome and hope to compute the probability of the event $R > i$ directly.

Step 4: Compute Event Probabilities. We want to compute $\Pr(R > i)$. There is no F in the first position of an outcome string with probability $1 - p$, no F in the second position with probability $1 - p$, etc. Since failures occur independently, we can multiply probabilities; the probability that there is no F in the first i positions is $(1 - p)^i$. Therefore, $\Pr(R > i) = (1 - p)^i$.

Substituting this result into the formula for $\text{Ex}(R)$ given above, we can find the mean time until the first failure of the main computer.

$$\begin{aligned}
\text{Ex}(R) &= \sum_{i=0}^{\infty} (1-p)^i \\
&= 1 + (1-p) + (1-p)^2 + (1-p)^3 + \ldots \\
&= \frac{1}{1-(1-p)} \\
&= \frac{1}{p}
\end{aligned}$$

In the first step, we substitute our expression for $\Pr(R > i)$. Next, we rewrite the summation in the $\ldots$ notation. This makes clear that we have the sum of a geometric series. The third step is an application of our old formula for the sum of a geometric series, and the last step is simplification.

The expected time until the main computer fails is $\frac{1}{p}$. For example, if the computer has a 1% chance of failing every hour, then we would expect a failure in 100 hours, or about four days. On the bright side, this means that David Wolf can expect 99 comfortable hours *without* a computer failure.

1.6 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 main computer fails?" is mathematically the same as the question, "How many children must the couple have until they get a girl?" In this case, a computer failure 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 $\frac{1}{p} = 2$ children. Since the last of these will be the girl, they should expect just 1 baby boy.

For homework, you will determine the expected number of children needed to get *both* a girl and a boy.

2 An Expectation Paradox

Here is a game that reveals a strange property of expectations.

First, you think of a probability distribution function on the natural numbers. This distribution can be absolutely anything you like. For example, you might choose a uniform distribution on $1, 2, \ldots, 6$, giving something like a fair die. Or you might choose a binomial distribution on $0, 1, \ldots, n$. You can even give every natural number a non-zero probability, provided, of course, that the sum of all probabilities is 1.

Next, I pick a random number z according to whatever distribution you invent. Finally, you pick a random number y according to the same distribution. If your number is bigger than mine ($y > z$), then the game ends. Otherwise, if our numbers are equal or mine is bigger ($y \leq z$), then you pick again, and keep picking until you get a value that is bigger than z. What is the expected number of picks that you must make?

Certainly, you always need at least one pick, so the expected number is greater than one. An answer like 2 or 3 sounds reasonable, though one might suspect that the answer depends on the distribution. The real answer is amazing: the expected number of picks that you need is always *infinite*, regardless of the distribution you choose!

This makes sense if you choose, say, the uniform distribution on $1, 2, \ldots, 6$. After all, there is a $\frac{1}{6}$ chance that I will pick 6. In this case, you must pick forever— you can never beat me!

In general, what is the probability that you need more than one pick? There are two cases to consider. If our numbers are different, then by symmetry there is a $\frac{1}{2}$ chance that mine is the larger, and you have to pick again. Otherwise, if our numbers are the same, then you always have to pick again. In either case, you need more than one pick with probability at least $\frac{1}{2}$.

What is the probability that you need more than two picks? Here is an erroneous argument. On the first pick, you beat me with probability about $\frac{1}{2}$. On the second pick, you beat me with probability about $\frac{1}{2}$. The probability that you fail to beat me on both picks is only $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Therefore, the probability that you need more than two picks is around $\frac{1}{4}$. The problem is that beating me on your first pick is not independent of beating me on your second pick; multiplying the probabilities of these two events is therefore invalid.

Here is a correct argument for the probability that you need more than two picks. Suppose I pick z and then you pick y_1 and y_2. There are two cases. If there is a unique largest number among these three, then there is a $\frac{1}{3}$ chance that my number z is it, and you must pick again. After all, the largest number is equally likely to be chosen first, second, or third, regardless of the distribution. Otherwise, two or three of the numbers are tied for largest. My number is as likely to be among the largest as either of yours, so there is a better than $\frac{1}{3}$ chance that my number is as large as all of yours, and you must pick again. In both cases, you need more than two picks with probability at least $\frac{1}{3}$.

By the same argument, the probability that you need more than i picks is at least $\frac{1}{i+1}$. Suppose I pick z and you pick $y_1, y_2, \ldots, y_i$. Again, there are two cases. If there is a unique largest number among our picks, then my number is as likely to be it as any one of yours; with probability $\frac{1}{i+1}$ you must pick again. Otherwise, there are several numbers tied for largest. My number is as likely to be one of these as any of your numbers, so with probability greater than $\frac{1}{i+1}$ you must pick again. In both cases, with probability at least $\frac{1}{i+1}$, you need more than i picks to beat me.

These arguments suggest that you should choose a distribution such that ties are very rare. For example, you might choose the uniform distribution on $[1, 10^{100}]$. In this case, the probability that you need more than i picks to beat me is very close to $\frac{1}{i+1}$ for reasonable i. For example, the probability that you need more than 99 picks is almost exactly 1%. This sounds very promising for you; intuitively, you might expect to win within a reasonable number of picks on average!

Unfortunately for intuition, there is a simple proof that the expected number of picks that you need in order to beat me is infinite, regardless of the distribution.

Theorem 2.1 If the random variable T is the number picks you need to beat me, then $\mathrm{Ex}(T) = \infty$.

Proof.

$$\begin{aligned} \mathrm{Ex}(T) &= \sum_{i=0}^{\infty} \Pr(T > i) \\ &\geq \sum_{i=0}^{\infty} \frac{1}{i+1} \\ &= \infty \end{aligned}$$

In the first step, we express the expectation using Theorem 1.1. This is valid, since the range of T is $\mathcal{N}$. In the second step, we use the observation from above that you need more than i picks with probability at least $\frac{1}{i+1}$. This gives an unbounded sum! ∎

This phenomenon can cause all sorts of confusion. For example, suppose you have a communication network. Assume that a packet has a $1/i$ chance of being delayed by i or more steps. This sounds good; there is only a 1% chance of being delayed by 100 or more steps. But, by the argument above, the expected delay for a packet is actually infinite!

3 Linearity of Expectation

Expected values obey a really nice rule called *linearity of expectation*. This rule says that the expected value of a sum of random variables is the sum of the expected values of the variables.

Theorem 3.1 (Linearity of Expectation) For any random variables R_1 and R_2,

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

In other words, expectation is a linear function. The same rule holds for more than two random variables:

Corollary 3.2 For any random variables $R_1, R_2, \ldots, R_k$,

$$\mathrm{Ex}(R_1 + R_2 + \ldots + R_k) = \mathrm{Ex}(R_1) + \mathrm{Ex}(R_2) + \ldots + \mathrm{Ex}(R_k)$$

Given the theorem, we can prove the corollary by induction on k. To prove the theorem, we need a result from tutorial two weeks ago called the Theorem of Total Probability. The result is restated below; the proof is in the tutorial notes.

Theorem 3.3 (Theorem of Total Probability) Let $B_1, B_2, \ldots$ be disjoint events such that the union is the entire sample space S. Then for all events $A \subseteq S$,

$$\Pr(A) = \sum_i \Pr(A \cap B_i)$$

Now we are ready to prove linearity of expectation.

Proof.

We will transform $\mathrm{Ex}(R_1 + R_2)$ into $\mathrm{Ex}(R_1) + \mathrm{Ex}(R_2)$ with a sequence of equalities. We start by writing out $\mathrm{Ex}(R_1 + R_2)$ using the definition of expectation. Then we split the sum over all values of $R_1 + R_2$ into two sums, one over values of R_1 and the other over values of R_2.

$$\begin{aligned}
\mathrm{Ex}(R_1 + R_2) &= \sum^{x \in \mathrm{range}(R_1 + R_2)} x \cdot \Pr(R_1 + R_2 = x) \\
&= \sum^{x_1 \in \mathrm{range}(R_1)} \sum^{x_2 \in \mathrm{range}(R_2)} (x_1 + x_2) \Pr((R_1 = x_1) \cap (R_2 = x_2))
\end{aligned}$$

Next, we split the summation into two parts. We then swap the the second pair of summation symbols, and continue by pulling out x_1 in the first pair of summations and pulling out x_2 in the second pair.

$$\begin{aligned}
&= \sum^{x_1}\sum^{x_2} x_1 \Pr((R_1 = x_1) \cap (R_2 = x_2)) + \sum^{x_1}\sum^{x_2} x_2 \Pr((R_1 = x_1) \cap (R_2 = x_2)) \\
&= \sum^{x_1}\sum^{x_2} x_1 \Pr((R_1 = x_1) \cap (R_2 = x_2)) + \sum^{x_2}\sum^{x_1} x_2 \Pr((R_1 = x_1) \cap (R_2 = x_2)) \\
&= \sum^{x_1} x_1 \sum^{x_2} \Pr((R_1 = x_1) \cap (R_2 = x_2)) + \sum^{x_2} x_2 \sum^{x_1} \Pr((R_1 = x_1) \cap (R_2 = x_2))
\end{aligned}$$

Now we apply the Theorem of Total probability to each pair of summations. For the first pair, event A in the theorem is the event $R_1 = x_1$. The B_i events in the theorem correspond to events of the form $R_2 = x_2$. The theorem is similarly applied to the second pair of summations.

$$\begin{aligned}
&= \sum^{x_1} x_1 \Pr(R_1 = x_1) + \sum^{x_2} x_2 \Pr(R_2 = x_2) \\
&= \operatorname{Ex}(R_1) + \operatorname{Ex}(R_2)
\end{aligned}$$

In the final step, we substitute $\operatorname{Ex}(R_1)$ and $\operatorname{Ex}(R_2)$ in place of their definitions. ■

The best thing about linearity of expectation is that no independence is required. This is great, because dealing with independence is pain! On one hand, we often need to work with random variables that are not independent. And even if we believe that some random variables are independent, we know from the last lecture that proving independence requires a lot of work.

3.1 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. Earlier in this lecture, 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:

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

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 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!

3.2 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 $\frac{1}{N}$. 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 are 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:

$$R = R_1 + R_2 + \ldots + R_N$$

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. That is, if $R_1 = R_2 = \ldots = R_{N-1} = 1$, then we know that $R_N = 1$. Therefore, R_N is not independent of the other indicator variables. But, since we plan to use linearity of expectation, we do not *care* whether the indicator variables are independent!

We can take the expected value of both sides of the equation above and apply linearity of expectation without worrying about independence:

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

All that remains is to compute the expected value of an indicator variable R_i. Applying the definition of expectation, we find that the expected value of an indicator variable R_i is just the probability that the indicator is 1.

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

(In general, the expected value of an indicator variable is always the probability that the indicator is 1.)

The quantity $\Pr(R_i = 1)$ is the probability that the i-th man gets his own hat, which is just $\frac{1}{N}$. We can now compute the expected number of men that get their own hat:

$$\begin{aligned} \text{Ex}(R) &= \text{Ex}(R_1) + \text{Ex}(R_2) + \ldots + \text{Ex}(R_N) \\ &= \frac{1}{N} + \frac{1}{N} + \ldots + \frac{1}{N} \\ &= 1 \end{aligned}$$

We should expect exactly 1 man to get the right hat!

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 $\frac{1}{N}$. This makes our solution very general, as the next example shows.

3.3 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 them. 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 $\frac{1}{N}$; 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. Our previous analysis still holds; the expected number of people that get their original appetizer is 1.

(Of course, the event that exactly one person gets his original appetizer never happens. Either everyone does or no one does. The name "expected value" can be misleading, since that value may never occur!)

3.4 Expected Number of Events that Occur

We can generalize the hat-check problem even further. Suppose that we have a collection of events in a sample space. What is the expected number of events that occur? For example, A_i might be the event that the i-th man receives his own hat. The number of events that occur is then the number of men that receive their own hat. Linearity of expectation gives a general solution to this problem:

Theorem 3.4 Given any collection of events $A_1, A_2, \ldots, A_N \subseteq S$, the expected number of these events that occur is

$$\sum_{i=1}^{N} \Pr(A_i).$$

The theorem says that the expected number of events that occur is the sum the probabilities of the events. For example, in the hat-check problem the probability of the event that the i-th man receives his hat is $\frac{1}{N}$. Since there are N such events, the theorem says that the expected number of men that receive their hat is $N \cdot \frac{1}{N} = 1$. This matches our earlier result.

Proof. Let the random variable R be the number of events that occur. Let R_i be an indicator variable for the event A_i; that is, $R_i(w) = 1$ if event A_i occurs in outcome w (i.e., if $w \in A_i$), and $R_i(w) = 0$ otherwise. Then $R_i = 1$ and A_i are the same event.

The number of events that occur is the sum of the indicator variables:

$$R = R_1 + R_2 + \ldots + R_N$$

Taking the expected value of both sides, we find:

$$\begin{aligned}
\mathrm{Ex}(R) &= \mathrm{Ex}(\sum_{i=1}^{n} R_i) \\
&= \sum_{i=1}^{n} \mathrm{Ex}(R_i) \\
&= \sum_{i=1}^{n} \Pr(R_i = 1) \\
&= \sum_{i=1}^{n} \Pr(A_i)
\end{aligned}$$

The second equation follows by linearity of expectation. In the third step, we use the fact that the expectation of an indicator variable is the probability that the indicator is 1. The final step follows because $R_i = 1$ and A_i are the same event. ∎

3.5 Flipping Fair Coins

Suppose that we flip N fair coins. What is the expected number that come up heads?

Consider the event that a particular coin comes up heads. Since the coin is fair, the probability of the event is $\frac{1}{2}$. Since there are N coins in all, there are N such events. By the theorem in the last section, the expected number of events that occur (the number of coins that come up heads) is $N \cdot \frac{1}{2} = \frac{N}{2}$.

Let's try to solve the same problem the hard way. In this case, assume that the coins are fair. Let the random variable R be the number of heads. We want to compute the expected value of R.

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

The first equation follows from the definition of expectation. In the second step, we evaluate $\Pr(R = i)$. An outcome of tossing the N coins can be represented by a length N sequence of H's and T's. An H in position i indicates that the i-th coin is heads, and a T indicates that the i-th coin is tails. The sample space consists of all 2^N such sequences. The outcomes are equiprobable, and so each has probability 2^{-N}. The number of outcomes with exactly i heads is the number of length N sequences with i H's, which is $\binom{N}{i}$. Therefore, $\Pr(R = i) = \binom{N}{i} 2^{-N}$.

The answer from linearity expectation and from the hard way must be the same, so we can equate the two results to obtain a neat identity.

$$\begin{aligned}
\sum_{i=0}^{N} i \binom{N}{i} 2^{-N} &= \frac{N}{2} \\
\sum_{i=0}^{N} i \binom{N}{i} &= N 2^{N-1}
\end{aligned}$$

In fact, we proved this identity by another method earlier in the term. Note that linearity of expectation solves the problem more easily and that the result is more general, since we do not need to assume that the coins are independent. The expected number of heads is $\frac{N}{2}$, even if some coins are glued together.