Expectations

Definition of a random variable

A random variable X is a function from the sample space S of an experiment to the real numbers. X(S) denotes the range of the function X.

The random variable is basically some number that we want to associate with outcomes the experiment. For instance, roll two dice - red and blue and define the random number D to be the sum of the two dice. In more mathematical terms let x be the value on the blue die and y be the value on the red die, then $D(x, y) = x + y$.

If X is a random variable defined on the sample space S of an experiment and r is a real number, then X = r is an event. The event X = r consists of all outcomes s in the sample space such that X(s) = r. p(X = r) is the sum of p(s) for all s such that X(s) = r.

Expressed in more verbose language, the probability that the random variable X takes on the value r, is the sum of the probabilities for all outcomes s for which the value of the random variable in r.

Distribution of a random variable

The distribution of a random variable is the set of all pairs (r, p(X = r)) such that $r \in X(S)$.

A histogram (a visualization that most of you have seen), is nothing more than a plot of these points.

Expectation

The expected value of a random variable is what we in common language just refer to as the average. Formally in math, if you have a random variable X, the expected value of the random variable is

$$E[X] = \sum_{s \in S} X(s)p(s)$$

where p(s) is the probability of an outcome.

Example - Lottery (taken from Zybook)

A lottery is run in which every ticket has six numbers. Each of the six numbers is in the range from 1 through 50. The person purchasing a lottery ticket can select the numbers on their ticket. On each Friday of the week, the state lottery commission holds a drawing in which six random numbers are generated. Each number in the range from 1 through 50

is equally likely. In order to win the jackpot, the ticket must match each number selected in order. If a person has a winning ticket, they win $100,000,000. What is the expected winning from a single ticket?

If we think of the random variable X as representative of the winnings from a lottery drawing, then X takes on only two values. 0, which happens most of the time, corresponding to not winning the lottery and then $100,000,000 which corresponds to winning the lottery. To compute the expected value we need to find the probabilities associated with the event of winning and the event of losing.

$$E[X] = (100,000,000) \cdot P(winning) + 0 \cdot (1 - P(winning))$$

What is the probability of winning? Our sample space consists of all possible arrangements of 6 numbers, where each number ranges from 1 to 50. The first number has 50 options, the second number has 50 options and so on. And of course, numbers could be repeated. So the sample space has the size 50^6.

How do you win? Only 1 way. When all the numbers match! So $P(winning) = \frac{1}{50^6}$.

Therefore

$$E[X] = (100,000,000) \cdot \frac{1}{50^6} + 0 \cdot (1 - \frac{1}{50^6}) = 0.0064$$

So the expected winnings in this lottery for a single ticket is 0.64 cents.

Example - Roulette

A roulette wheel has 38 slots. These slots consist of the number 1 through 36. Half of them are read and half of them are black. You also have 0 and 00.

Let us say you bet $1 on Black. If a black number comes up, you receive your dollar back plus $1; otherwise you lose your dollar. Let X be your net winnings in one game. Then X can take on the values +1 and -1. What are your expected winnings.

$P[X = 1] = 18/38$ and $P[X = (-1)] = 20/38$.

Thus

$$E[X] = 1 \cdot \frac{18}{38} + -1 \cdot \frac{20}{38} = \frac{18}{38} - \frac{20}{38} = -\frac{1}{19}$$

Again, you are expected to lose some money on each turn (about a nickel).

Linearity of Expectation

Quite possibly, one of the nicest results when it comes to computing Expectations is what is called Linearity of Expectations. We will state it formally and then see how a lot of calculations become really simple once you use this idea.

If X and Y are random variables defined on the same sample space S and a and b are two real numbers, then the following result is true

$$E[aX + bY] = aE[X] + bE[Y]$$

The applications of this are immediate. For instance going back to the roulette question, if instead of playing roulette just once, you decide to spin 100 times, how much do you stand to gain or lose? Let X_1 be the random variable corresponding to the earning the first time around. X_2 be the random variable corresponding to the earnings the second time around and so on.

If X is the total amount earned in this process, then $X = X_1 + X_2 + \ldots + X_{100}$.

Now thanks to linearity of expectations, you get the $E[X] = 100 \cdot \frac{-1}{19} = 5.26$. So the more you play, the more you lose.

Example of coin flips

Suppose you flip a coin n times and the random variable X is used to denote the number of heads. What the expected value of this random variable?

Intuition suggests that about half the time the number will be heads and as we will see over here, intuition will be right. But there are two ways to approach this problem and one is a lot easier.

Method 1

Let X_i denote the number of heads we see on the ith flip. Obviously X_i can take values 0 or 1 only.

then $X = X_1 + X_2 + \ldots + X_n$. This immediately means

$$E(X) = E(X_1 + X_2 + \ldots X_n) = 0.5n$$

Method 2

Suppose instead of breaking the problem up into each individual coin flip and then using linearity of expectations, you decide to do it directly. X being the number of heads can take on any value from 0 to n. What is the probability that X takes on a specific value, let's call it i.

This basically amounts to answering the question, in how many ways can we get i coins to be heads. Each coin flip is independent, so the chance of getting heads at any point is 0.5. To get i heads we just need to choose i to the coin flips to land as heads. The other $(n - i)$ times, we get tails and this event has a probability of 0.5 as well. Therefore $P(X = i) = \binom{n}{i}(0.5)^i(0.5)^{(n-i)}$.

Now that we know the probability of getting (X=i), we can compute the expectation as follows

$$E[X] = \sum_{i=0}^{n} i \cdot P(X = i) = \sum_{i=0}^{n} i \cdot \binom{n}{i} (0.5)^i (0.5)^{(n-i)}$$
$$= (0.5)^n \sum_{i=0}^{n} i \cdot \binom{n}{i}$$

The summation is not particularly easy to compute until you realize that you can use the binomial theorem and some basic calculus to show that it is in fact $n2^{n-1}$. Putting everything together you get the same result as before that the expected value will be 0.5n.

This example shows that while expectation can be calculated using the direct formula, it is usually better to try and see if there is any possible way that linearity of expectations can be used. It leads to tremendous simplifications.

Indicator random variables

Indicator random variables are random variables that take the value 1 if the event happens and 0 if the event does not happen. While seemingly trivial, these random variables have a nice property that makes them invaluable in expectation calculations. $E(X_i) = P(X_i = 1) =$ probability of the event happening.

Recall the derangements (hat check problem) that involved people going into a bar, getting drunk and then not getting their hats back. We had previously computed the probability of no one getting their hats back. Now let us compute the expected number of people who get their hats back?

Let X_i be the indicator random variable for the ith person getting their hat back. So X_i is 1 if person i gets their hat back and 0 otherwise. Given this setup, the total number of people who get their hats back will be $X_1 + X_2 + \ldots X_n$.

$E(X_i)$, since it is an indicator variable, is the same thing as the probability that the ith person gets their hat. That is a $\frac{1}{n}$ chance. Obviously this expectation is independent of i, so every single indicator variable has the expected value of $\frac{1}{n}$.

The expected number of people getting their hats will be $E(X_1) + E(X_2) + \ldots + E(X_n)$. That means the expectation is $nE(X_1)$ if we use linearity of expectation and the fact that each of the expectations is basically the same.

This means the expected number of people who get their hats back is $n\frac{1}{n}$ and that gives us the somewhat surprising result that on average only person gets their hat back. Rather surprisingly, the number does not depend upon n at all.

Usage of indicator random variables

One area in algorithms that has gained a lot of popularity is the concept of randomized algorithms. While we will visit this topic in more detail either at the end of this course or

at the beginning of 596, the methods of analyzing a randomized algorithm are all rooted in probability and expectation.

So here is an example of using this with some actual code.

```
def findMin(arr):
    minimum = arr[0]
    for i in range(0,len(arr)):
        if arr[i] < minimum :
            minimum = arr[i]
    return minimum
```

We are interested in knowing the expected number of times the variable minimum gets reassigned (number of updates).

When asked to analyze an algorithm like this, the assumption really is that you are dealing with a random input array. Because anything else and you are biasing towards some answer.

So let's say we have a random array arr, with elements arr[0] through arr[n].

Let us define indicator variables X_i to correspond to the event that arr[i] becomes the minimum.

What is $E(X_i) = P(arr[i]$ becomes the minimum while iterating through this loop)

That is the same as P(arr[i] is the minimum in the set $\{arr[0], arr[1], arr[2], ...arr[i-1]\}$)

What is the probability of that?

One way of thinking of it is that there is a $\frac{1}{i}$ chance that when you randomly take i elements, the lowest one is at the end.

What is the expected number of updates. Since we want to use linearity of expectation, if we just say let $X = \sum_{i=0}^{n-1} X_i$ then clearly the value of X represents the number of times the variables minimum gets updated.

So

$$E[X] = \sum_{i=0}^{n-1} \frac{1}{i+1}$$

Unfortunately there is no closed form expression for this (a formula). But there is an approximation which comes from the theory of Harmonic numbers that basically approximates this to be close to the natural logarithm $ln(n)$. Intuition might suggest that you actually have a linearly increasing number of updates that take place as the size of your array goes up. Applying the theory of expectation, we see that actually the increase is a lot slower.

Bernoulli Trial

An experiment which has a single outcome of success or of failure. Generally, the probability of success is represented with p.

A coin-flip is the most traditional variant of a Bernoulli trial.

The generalized version of computing the probability of a certain number of heads in n coin flips is this theorem.

Theorem 1. *The probability of exactly k successes in a sequence of n independent Bernoulli trials, with probability of success p and probability of failure $q = 1 - p$ is*

$$\binom{n}{k} p^k q^{n-k}$$