CIS 2033, Spring 2017,
Important Examples of Discrete Random Variables

Instructor: David Dobor

February 18, 2017

Last time we introduced the notion of a random variable. It is now time to see some examples of random variables.

Bernoulli and indicator random variables

We start with the simplest conceivable random variable – a random variable that takes the values of 0 or 1 with certain given probabilities. Such a random variable is called a Bernoulli random variable. And the distribution of this random variable is determined by parameter p, which is a given number that lies in the interval between 0 and 1.

The simplest random variable: Bernoulli with parameter $p \in [0, 1]$

$$X = \begin{cases} 1, & \text{w.p. } p \\ 0, & \text{w.p. } 1 - p \end{cases}$$

So this random variable takes on the value of 0 with probability $1 - p$ and it takes on the value of 1 with probability p. If you wish to plot the PMF of this random variable, the plot is rather simple. It consists of two bars, one at 0 and one at 1 and is shown in figure 1.

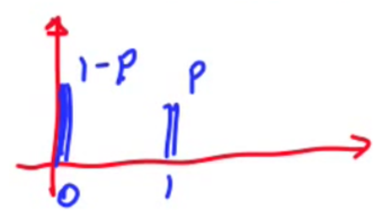


Figure 1: **Bernoulli PMF.** This random variable takes on the value of 1 with probability p and it takes on the value 0 with probability $1 - p$.

Bernoulli random variables show up whenever you're trying to model a situation where you run a trial that can result in two alternative outcomes, either success or failure, or heads versus tails, and so on.

Another situation where Bernoulli random variables show up is when we're making a connection between events and random variables. Here's how this connection is made.

We have our sample space Ω. Within that sample space we have a certain event A; outside of the event A, of course, we have A^c. Our random variable is defined so that it takes a value of 1, whenever the outcome of the experiment lies in A; it takes a value of 0 whenever the outcome of the experiment lies in A^c. This random variable is called the *indicator random variable of the event* A and is denoted by I_A. Thus I_A is equal to 1 if and only if event A occurs. This is shown in figure 2.

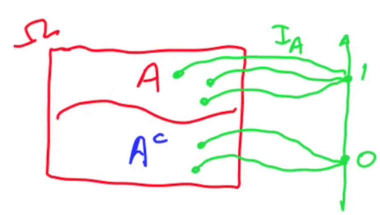


Figure 2: An Indicatior Random Variable, denoted by I_A here.

The PMF of that random variable can be found as follows:

$$p_{I_A}(1) = P(I_A = 1) = P(A).$$

And so what we have is that the indicator random variable is a Bernoulli random variable with a parameter p equal to the probability of the event of interest.

Indicator random variables are very useful because they allow us to translate a manipulation of events to a manipulation of random variables. And sometimes the algebra of working with random variable is easier than working with events, as we will see in some later examples.

Check Your Understanding: Indicator variables

Let A and B be two events (subsets of the same sample space Ω), with nonempty intersection. Let I_A and I_B be the associated indicator random variables.

For each of the two cases below, select one statement that is true.

(A) $I_A + I_B$:

(a) is an indicator variable of $A \cup B$.

(b) is an indicator variable of $A \cap B$.

(c) is not an indicator variable of any event.

(B) $I_A \cdot I_B$:

(a) is an indicator variable of $A \cup B$.

(b) is an indicator variable of $A \cap B$.

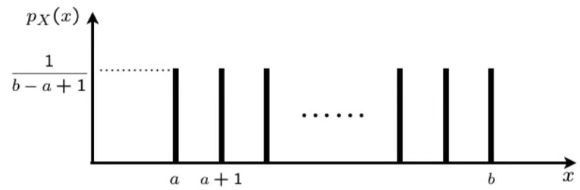
(c) is not an indicator variable of any event.

Uniform random variables

In this segment and the next two, we will introduce a few useful random variables that show up in many applications – discrete uniform random variables, binomial random variables, and geometric random variables.

WE'LL START WITH a discrete uniform. A discrete uniform random variable is one that has a PMF that is shown in the figure on the next page.

So the discrete uniform random variable takes values in a certain range – here the range is from a to b, both ends inclusive – and each one of the values in that range – of which there are $b - a + 1$ – has the



same probability, which of course has to be $1/(b-a+1)$ so that the probabilities sum up to one.

A discrete uniform is completely determined by two parameters that are two integers, a and b, which are the beginning and the end of the range of that random variable. We're thinking of an experiment where we're going to pick an integer at random among the values that are between a and b with the end points a and b included, where all of these values are equally likely.

Note that the outcome of the experiment just described is already a number. And the numerical value of the random variable is just the number that we happen to pick in the range from a to b. So in this context, there isn't really a distinction between the outcome of the experiment and the numerical value of the random variable. They are one and the same.

In summary,

Discrete uniform random variable; parameters a, b

- **Parameters:** integers a, b; $a \leq b$
- **Experiment:** Pick one of $a, a+1, \ldots, b$ at random; all equally likely
- **Sample space:** $\{a, a+1, \ldots, b\}$
- **Random variable X:** $X(\omega) = \omega$

What does this random variable model in the real world? It models a case where we have a range of possible values, and we have complete ignorance, no reason to believe that one value is more likely than the other.

As an example, suppose that you look at your digital clock, and the time that it tells you is 11:52:26 seconds. Suppose that you just look at the seconds. The seconds reading is something that takes values in the set from 0 to 59. So there are 60 possible values. If you choose to look at your clock at a completely random time, there's no reason to expect that any one reading would be any more or less likely than any other – all readings should be equally likely, and each reading should have a probability of 1/60.

One final comment – let us look at the special case where the beginning and the endpoint of the range of possible values is the same, which means that our random variable can only take one value, namely that particular number a.

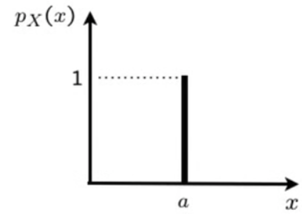


Figure 3: **Special Case.** $a = b$ constant/deterministic r.v.

In that case, the random variable that we're dealing with is really a constant. It doesn't have any randomness. It is a deterministic random variable that takes a particular value of a with probability equal to 1. It is not random in the common sense of the world, but mathematically we can still consider it a random variable that just happens to be the same no matter what the outcome of the experiment is.

Binomial random variables

The next random variable that we will discuss is the binomial random variable. It is one that is already familiar to us in most respects. It is associated with the experiment of taking a coin and tossing it n times independently. At each toss, there is a probability, p, of obtaining heads. So the experiment is completely specified in terms of two parameters – n, the number of tosses, and p, the probability of heads at each one of the tosses.

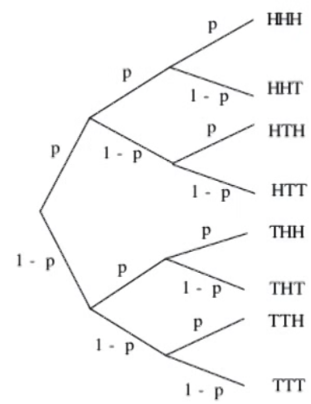


Figure 4: The leaves in this tree are the outcomes of the experiment of flipping a coin three times. At each flip, the coin lands heads with probability p, independently of other flips.

Binomial random variable; **parameters:** positive integer n; $p \in [0, 1]$

- **Experiment:** n independent tosses of a coin with $\mathbf{P}(\text{Heads}) = p$
- **Sample space:** Set of sequences of H and T, of length n
- **Random variable** X**:** number of Heads observed
- **Model of:** number of successes in a given number of independent trials

We can represent this experiment by the usual sequential tree diagram. The leaves of the tree are the elements of the sample space – the possible outcomes of the experiment. A typical outcome is a particular sequence of heads and tails that has length n. In figure 4 we took $n = 3$.

We can now define a random variable associated with this experiment. We denote our random variable by X. X is the number of heads that are observed. For example, if the outcome happens to be *tails, heads, heads*, that is, we observed 2 heads, then the numerical value of our random variable X is equal to 2. We write this *event* as $X = 2$.

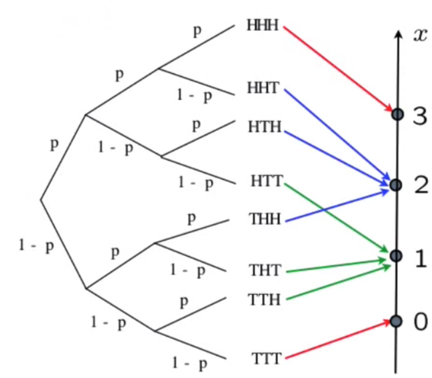


Figure 5: The binomial random variable X measures the number of heads obtained in three independent flips of a coin.

In general, a binomial random variable can be used to model any situation in which we have a fixed number of independent and identical trials, where each trial can result in success or failure, and where the probability of success is equal to some given number p. The number of successes obtained in these trials is, of course, random and it is modeled by a binomial random variable.

We can now proceed and calculate the PMF of this random variable. Instead of calculating the whole PMF, let us look at just one typical entry of the PMF. Let's look at the probability of the event that $X = 2$, which means that we've observed 2 heads in three tosses of this coin. In our notation, we want to compute

$$p_X(2) = P(X = 2)$$

Now, this random variable taking the numerical value of 2, is an event that can happen in three possible ways that we can identify in the sample space. We can have 2 heads followed by a tail. We can have heads, tails, heads. Or we can have tails, heads, heads. That is,

$$\begin{aligned} p_X(2) &= P(X = 2) \\ &= P(HHT) + P(HTH) + P(THH) \\ &= 3p^2(1-p) \end{aligned}$$

Now the last term in this equality can also be written this way:

$$p_X(2) = \binom{3}{2} p^2(1-p)$$

since 3 is the same as $\binom{3}{2}$. It's the number of ways that you can choose 2 heads, where they will be placed in a sequence of 3 slots or 3 trials.

More generally, we have the familiar binomial formula:

$$\boxed{p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{for } k = 0, 1, \ldots, n}$$

This is a formula that you have already seen. It's the probability of obtaining k successes in a sequence of n independent trials. The only thing that is new is that instead of using the traditional probability notation, now we're using our new PMF notation.

To get a feel for the binomial PMF, it's instructive to look at some plots. Suppose that we toss the coin three times and that the coin tosses are fair, so that the probability of heads is equal to 1/2. In

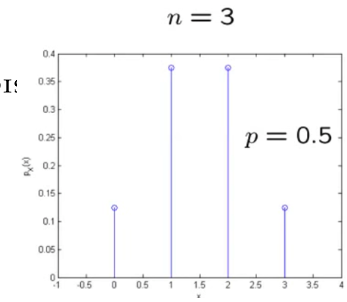
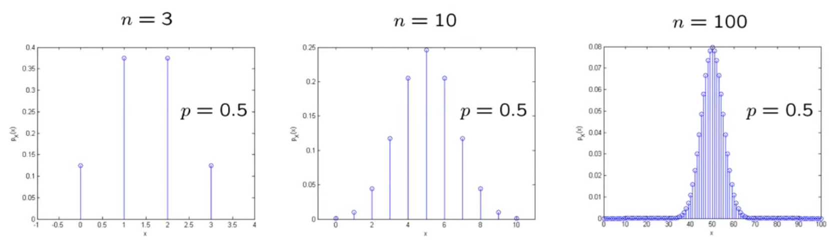


Figure 6: The binomial random variable X measures the number of successes (heads) in 3 flips of a fair coin. This PMF plot shows the probabilities with which each of the events $X = 0, X = 1, X = 2, X = 3$ occur.

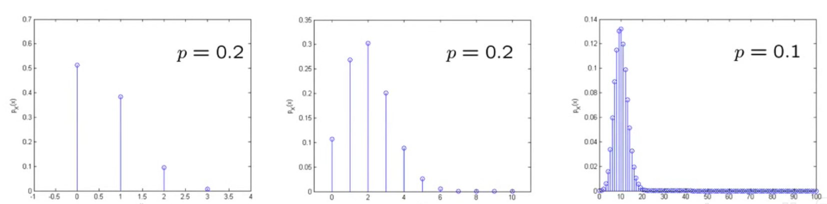
figure 6 we see that 1 head or 2 heads are equally likely, and they are more likely than the outcome of 0 or 3 heads.

Now, if we change the number of tosses and toss the coin 10 times, then we see that the most likely result is to have 5 heads. The probability of the number of heads being greater than 5 or smaller than 5 becomes smaller and smaller. Now, if we toss the coin many times, let's say 100 times, the coin is still fair, then we see that the number of heads that we're going to get is most likely to be somewhere in the range, let's say, 35 and 65. These are values of the random variable that have some noticeable or high probabilities. But anything below 30 or anything above 70 is extremely unlikely to occur.



We can generate similar plots for unfair coins. Suppose now that our coin is biased and the probability of heads is quite low, equal to 0.2. In that case, the most likely result is that we're going to see 0 heads. There's smaller and smaller probability of obtaining more and more heads. On the other hand, if we toss the coin 10 times, we expect to see a few heads, not a very large number, but some number of heads between, let's say, 0 and 4.

Finally, if we toss the coin 100 times and we take the coin to be an extremely unfair one, what do we expect to see? You will be plotting the PMFs of binomial random variables for various values of parameters n and p at the recitation.



Check Your Understanding: The binomial PMF

You roll a fair six-sided die (all 6 of the possible results of a die roll are equally likely) 5 times, independently. Let X be the number of times that the roll results in 2 or 3. Find the numerical values of the following.

(a) $p_X(2.5)$

(b) $p_X(1)$

Geometric random variables

The last discrete random variable that we will discuss (for now) is the so-called geometric random variable. It shows up in the context of the following experiment. We have a coin that we toss *infinitely* many times and independently. At each coin toss we have a fixed probability of heads, which is some given number, p. This is a parameter that specifies the experiment.

(When we say that the infinitely many tosses are independent, what we mean is that any finite subset of those tosses are independent of each other. I'm only making this comment because we introduced a definition of independence of finitely many events, but had never defined the notion of independence or infinitely many events.)

Figure 7: A typical outcome of our experiment of tossing a coin infinitely many times. The dots indicate that the tosses continue forever.

The sample space for this experiment is the set of *infinite* sequences of heads and tails. A typical outcome of this experiment is shown in figure 7. It's a sequence of heads and tails in some arbitrary order. Of course, it's an infinite sequence, so it continues forever, but we are only showing here the beginning of that sequence.

We're interested in the random variable X which is *the number of tosses until the first heads*. So if our sequence was like the one shown in figure 7 or 8, our random variable would be taking a value of 5.

Figure 8: The random variable X is the number of tosses until the first heads. It took us 5 tosses to get the first heads, so $X = 5$.

A random variable of this kind appears in many applications and many real world contexts. In general, it models situations where we're waiting for something to happen. Suppose that we keep performing trials and that each time we do the trial, it can result either in success or failure. We're counting the *number of trials it takes until a success is observed for the first time*.

Now, these trials could be experiments of some kind, could be processes of some kind, or they could be whether a customer shows up in a store in a particular second or not. So there are many diverse interpretations of the words trial and of the word success that would allow us to apply this particular model to a given situation.

In summary,

Geometric random variable; parameter p: $0 < p \leq 1$

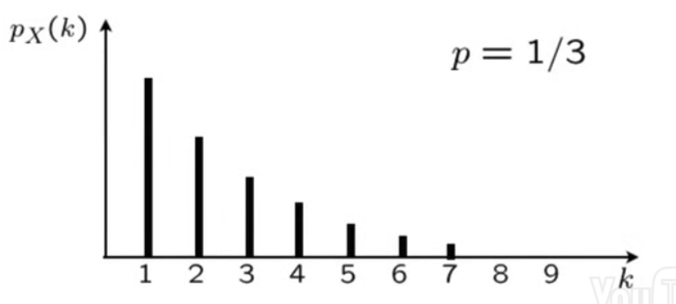
- **Experiment:** infinitely many independent tosses of a coin; $\mathbf{P}(\text{Heads}) = p$
- **Sample space:** Set of infinite sequences of H and T
- **Random variable** X**:** number of tosses until the first Heads
- **Model of:** waiting times; number of trials until a success

Now, let us move to the calculation of the PMF of this random variable. By definition, what we need to calculate is the probability that the random variable takes on a particular numerical value. What does it mean for X to be equal to k? What it means is that the first heads was observed in the k-th trial, which means that the first $k-1$ trials were tails, and then were followed by heads in the k-th trial.

$$p_X(k) = P(X = k) = P(\underbrace{T \dots T}_{k-1 \text{ times}} H) = (1-p)^{k-1}p$$

Because the time of the first head can only be a positive integer, this formula applies for $k = 1, 2, \dots$. So our random variable takes values in a discrete but infinite set.

Let's plot this geometric PMF for some particular value of p, say $p = 1/3$.



The probability that the first head shows up in the first trial is equal to p, so the height of the leftmost bar here is p, the probability of heads. The probability that the first head appears in the second trial is the probability that we had heads following a tail. So we have the probability of a tail times the probability of a head. This means that the height of the second bar from the left is $(1-p)p$. How about the height of the third bar? Yes, that's $(1-p)^2 p$ because the probability of obtaining the first head in the third trial can only happen if the first two trials resulted in tails and the third trial resulted in a head. And this pattern continues forever: $p_X(k) = (1-p)^{k-1}p$, for any integer $k \geq 1$.

FINALLY, one little technical remark. There's a possible and rather annoying outcome of this experiment, which would be that we observe a sequence of tails forever, no heads ever showing up. In that case, our random variable is not well-defined, because there is no first heads to consider. You might say that in this case our random variable takes a value of infinity, but we would rather not have to deal with random variables that could be infinite.

Fortunately, it turns out that this particular event has 0 probability of occurring, which I will now try to show.

Let us compare the event that we always see tails to the event where we see tails in the first k trials. How do these two events relate?

- Event 1: no Heads ever
- Event 2: tails in the first k trials: $\underbrace{T \ldots T}_{k \text{ times}} \ldots$

If we have always tails, then we will have tails in the first k trials. So event 1 implies event 2 – event 1 is smaller than event 2. Therefore the probability of event 1 is less than or equal to the probability of event 2. Moreover, the probability of event 2 is $(1-p)^k$:

$$P(\text{no Heads ever}) \leq P(\underbrace{T \ldots T}_{k \text{ times}}) = (1-p)^k$$

Now, this is true no matter what k we choose. And by taking k arbitrarily large, $P(\text{no Heads ever})$ becomes arbitrarily small. Why does it become arbitrarily small? Well, we're assuming that p is positive, so $1-p$ is a number less than 1. And when we multiply a number strictly less than 1 by itself over and over, we get arbitrarily small numbers: $(1-p)^k \to 0$ as k grows large. So the probability of never seeing a head is less than or equal to an arbitrarily small positive number, and this means that we can ignore the outcome of "no Heads ever".

As a side consequence of this, the sum of the probabilities of the different possible values of k is going to be equal to 1, because we're certain that the random variable is going to take a finite value. And so when we sum probabilities of all the possible finite values, that sum will have to be equal to 1. Indeed, you can use the formula for the geometric series to verify that, and that would be a nice exercise to do.

Check Your Understanding: Geometric random variables

Let X be a geometric random variable with parameter p. Find the probability that $X \geq 10$.

The following topics will be discussed over the next few classes.

Expectation

Check Your Understanding: Expectation calculation

The PMF of the random variable Y satisfies $p_Y(-1) = 1/6, p_Y(2) = 2/6, p_Y(5) = 3/6$, and $p_Y(y) = 0$ for all other values y. The expected value of Y is:

$$E[Y] = \dots$$

$$E[Y] = (-1) \cdot \frac{1}{6} + 2 \cdot \frac{2}{6} + 5 \cdot \frac{3}{6} = \frac{18}{6} = 3$$

Elementary properties of expectation

Check Your Understanding: Random variables with bounded range

Suppose a random variable X can take any value in the interval $[-1, 2]$ and a random variable Y can take any value in the interval $[-2, 3]$.

(A) The random variable $X - Y$ can take any value in an interval $[a, b]$. Find the values of a and b:

- (a) $a = \dots$
- (b) $b = \dots$

(B) Can the expected value of $X + Y$ be equal to 6?

- (a) Yes, why not?
- (b) No, no way!

(A) The smallest possible value of $X - Y$ is obtained if X takes its smallest value, -1, and Y takes its largest value, 3, resulting in $X - Y = -1?3 = -4$. Similarly, the largest possible value of $X - Y$ is obtained if X takes its largest value, 2, and Y takes its smallest value, -2, resulting in $X - Y = 2 - (-2) = 4$.

(B) No, no way! No matter what the outcome of the experiment is, the value of $X + Y$ will be at most 5, and so the expected value can be at most 5.

The expected value rule

Check Your Understanding: The expected value rule

Let X be a uniform random variable on the range $\{-1, 0, 1, 2\}$. Let $Y = X^4$. Use the expected value rule to calculate $E[Y]$.

$$E[Y] = \dots$$

We are dealing with $Y = g(X)$, where g is the function defined by $g(x) = x^4$. Thus,

$$\begin{aligned} E[Y] &= E[X^4] = \sum_x x^4 p_X(x) = (-1)^4 \cdot \frac{1}{4} + 0^4 \cdot \frac{1}{4} + 1^4 \cdot \frac{1}{4} + 2^4 \cdot \frac{1}{4} \\ &= \frac{1}{4} + \frac{1}{4} + \frac{16}{4} \\ &= 4.5 \end{aligned}$$

Linearity of expectations

Check Your Understanding: Linearity of expectations

The random variable X is known to satisfy $E[X] = 2$ and $E[X^2] = 7$. Find the expected value of $8 - X$ and of $(X-3)(X+3)$.

(a) $E[8 - X] = \dots$

(b) $E[(X-3)(X+3)]$

(a) The random variable $8 - X$ is of the form $aX + b$, with $a = -1$ and $b = 8$. By linearity, $E[8 - X] = -E[X] + 8 = -2 + 8 = 6$.

(b) The random variable $(X-3)(X+3)$ is equal to $X^2 - 9$ and therefore its expected value is $E[X^2] - 9 = 7 - 9 = -2$.