Chapter 3. Discrete Random Variables

3.4: Zoo of Discrete RVs Part I

(From "Probability & Statistics with Applications to Computing" by Alex Tsun)

In this section, we'll define formally what it means for random variables to be independent. Then, for the rest of the chapter (3.4, 3.5, 3.6), we'll discuss commonly appearing random variables for which we can just cite its properties like its PMF, mean, and variance without doing any work! These situations are so common that we name them, and can refer to them and related quantities easily!

3.4.1 Independence of Random Variables

Definition 3.4.1: Independence

Random variables X and Y are independent, denoted $X \perp Y$, if for all $x \in \Omega_X$ and all $y \in \Omega_Y$, any of the following three equivalent properties holds:

1. $\mathbb{P}(X = x \mid Y = y) = \mathbb{P}(X = x)$
2. $\mathbb{P}(Y = y \mid X = x) = \mathbb{P}(Y = y)$
3. $\mathbb{P}(X = x \cap Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)$

Note, that this is the same as the event definition of independence, but it must hold for all events $\{X = x\}$ and $\{Y = y\}$.

Theorem 3.4.1: Variance Adds for Independent RVs

If $X \perp Y$, then

$$\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y)$$

This will be proved a bit later, but we can start using this fact now! It is important to remember that you *cannot* use this formula if the random variables are not independent (unlike linearity).

A common misconception is that $\mathsf{Var}(X - Y) = \mathsf{Var}(X) - \mathsf{Var}(Y)$, but this actually isn't true, otherwise we could get a negative number. In fact, if $X \perp Y$, then

$$\mathsf{Var}(X - Y) = \mathsf{Var}(X + (-Y)) = \mathsf{Var}(X) + \mathsf{Var}(-Y) = \mathsf{Var}(X) + (-1)^2\mathsf{Var}(Y) = \mathsf{Var}(X) + \mathsf{Var}(Y)$$

3.4.2 The Bernoulli Process and Bernoulli Random Variable

There are several random variables that occur naturally and frequently! It is often useful to be able to recognize these random variables by their characterization, so we can take advantage of relevant properties such as probability mass functions, expected values, and variance. In the rest of this section and chapter 3,

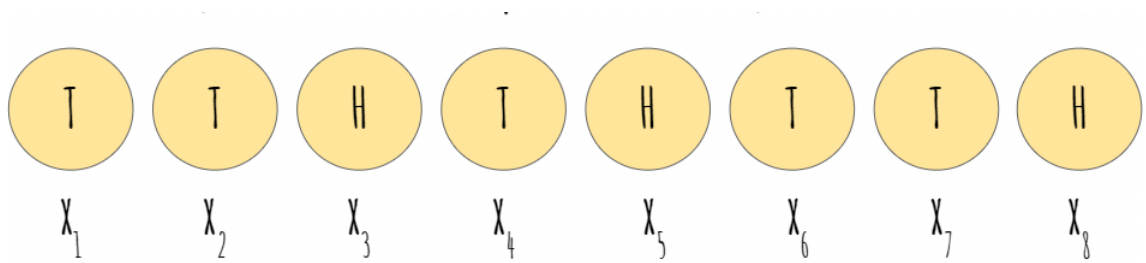
we will explore some fundamental discrete random variables, while finding those aforementioned properties, so that we can just cite them instead of doing all the work again!

Before diving into the random variables themselves, let's look at a situation that arises often...

Definition 3.4.2: Bernoulli Process

A Bernoulli process with parameter p is a sequence of independent coin flips $X_1, X_2, X_3, \ldots$ where $\mathbb{P}(\text{head}) = p$. If flip i is heads, then we encode $X_i = 1$; otherwise, $X_i = 0$. From this process we can measure many interesting things.

Let's illustrate how this might be useful with an example. Suppose we independently flip 8 coins that land heads with probability p, and get the following sequence of coin flips



This series of flips is a Bernoulli process. We call each of these coin flips a Bernoulli random variable (or indicator rv).

Definition 3.4.3: Bernoulli/Indicator Random Variable

A random variable X is Bernoulli (or indicator), denoted $X \sim \text{Ber}(p)$, if and only if X has the following PMF:

$$p_X(k) = \begin{cases} p, & k = 1 \\ 1 - p, & k = 0 \end{cases}$$

Each X_i in the Bernoulli process with parameter p is Bernoulli/indicator random variable with parameter p. It simply represents a binary outcome, like a coin flip.

Additionally,

$$\mathbb{E}[X] = p \text{ and } \mathsf{Var}(X) = p(1-p)$$

Proof of Expectation and Variance of Bernoulli.
Suppose $X \sim \text{Ber}(p)$. Then

$$\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1-p) = p$$

Now for the variance, we compute $\mathbb{E}\left[X^2\right]$ first by LOTUS:

$$\mathbb{E}\left[X^2\right] = 1^2 \cdot p + 0^2 \cdot (1-p) = p$$

$$\mathsf{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = p - p^2 = p(1-p)$$

□

Notice how we found a situation whose general form comes up quite often, and derived a random variable that models that situation well. Now, anytime we need a Bernoulli/indicator random variable we can denote it as follows: $X \sim \text{Ber}(p)$.

3.4.3 The Binomial Random Variable

If you recall, one of the main reasons that indicator random variables are useful is because they can compose more complicated random variables. For example, in the sequence of coin flips above, we might be interested in modeling the probability of getting a certain number of heads. In order to do this, it would be useful to use a random variable that is equal to the sum of the Bernoulli's that each represent a single flip. These types of random variables also come up frequently, so we have a special name for them, binomial random variables.

$$X \sim \text{Bin}(n,p) = \sum_{i=1}^{n} X_i, \text{ where } X_i\text{'s are } \textit{independent} \text{ Bernoulli random variables}$$

That is, we write $X \sim \text{Bin}(n,p)$ to be the number of heads in n *independent* flips of a coin with $\mathbb{P}(\text{head}) = p$. Why is it true that if X is the number of heads in n flips, that $X = \sum_{i=1}^{n} X_i$ (recall the mermaids in 3.3)?

Let's try to derive the PMF of a binomial rv. Its range is $\Omega_X = \{0, 1, \ldots, n\}$ since we can get anywhere from 0 heads to n heads. Let's consider the case of $n = 5$ flips, and figure out the probability we get exactly 4 heads, $\mathbb{P}(X = 4)$.

Here's one sample sequence of heads and tails with exactly four heads, HTHHH, and its probability is (by independence):

$$\mathbb{P}(HTHHH) = p \cdot (1-p) \cdot p \cdot p \cdot p = p^4(1-p)^{5-4}$$

But this is not the only sequence of flips which gives exactly 4 heads! How many such sequences are there? There are $\binom{5}{4}$ since we choose 4 out of the 5 positions to put the heads, and the remaining must be tails. Haha, so our counting knowledge is finally being applied! So we must sum these $\binom{5}{4}$ disjoint cases, and we get

$$\mathbb{P}(X = 4) = \binom{5}{4} p^4 (1-p)^{5-1}$$

We can generalize this as follows to get the PMF of a binomial random variable:

$$p_X(k) = \mathbb{P}(X = k) = \mathbb{P}(\text{exactly k heads in n Bernoulli trials}) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \Omega_X$$

This hopefully sheds some light on why $\binom{n}{k}$ is called a binomial coefficient and X a binomial random variable. Before computing its expectation, let's make sure we didn't make a mistake, and check that our probabilities sum to 1. This will use the binomial *theorem* we learned in chapter 1 finally: $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$.

$$\begin{aligned}
\sum_{k=0}^{n} p_X(k) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} && \text{[PMF of Binomial RV]} \\
&= (p + (1-p))^n && \text{[binomial theorem]} \\
&= 1^n = 1
\end{aligned}$$

Definition 3.4.4: Binomial Random Variable

A random variable X has a Binomial distribution, denoted $X \sim \text{Bin}(n, p)$, if and only if X has the following PMF for $k \in \Omega_X = \{0, 1, 2, \ldots, n\}$:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

X is the sum of n independent $\text{Ber}(p)$ random variables, and represents the number of heads in n independent coin flips where $\mathbb{P}(\text{head}) = p$.

Additionally,

$$\mathbb{E}[X] = np \text{ and } \mathsf{Var}(X) = np(1-p)$$

Proof of Expectation and Variance of Binomial. We can use linearity of expectation to compute the expected value of a particular binomial variable (i.e. the expected number of successes in n Bernoulli trials). Let $X \sim \text{Bin}(n, p)$.

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i] && \text{[linearity of expectation]} \\
&= \sum_{i=1}^{n} p && \text{[expectation of Bernoulli]} \\
&= np
\end{aligned}$$

This makes sense! If $X \sim \text{Bin}(100, 0.5)$ (number of heads in 100 independent flips of a fair coin), you expect 50 heads, which is just $np = 100 \cdot 0.5 = 50$. Variance can be found in a similar manner

$$\begin{aligned}
\mathsf{Var}(X) &= \mathsf{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \sum_{i=1}^{n} \mathsf{Var}(X_i) && \text{[variance adds if independent rvs]} \\
&= \sum_{i=1}^{n} p(1-p) && \text{[variance of Bernoulli]} \\
&= np(1-p)
\end{aligned}$$

Like Bernoulli rvs, Binomial random variables have a special place in our zoo. Argually, Binomial rvs are probably the **most important** discrete random variable, so make sure to understand everything above and be ready to use it!

It is important to note for the hat check example in 3.3 that we had the sum of n Bernoulli/indicator rvs BUT that they were NOT independent. This is because if we know one person gets their hat back, someone else is more likely to (since there are $n-1$ possibilities instead of n). However, linearity of expectation works **regardless of independence**, so we were able to still add their expectations like so

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n \cdot \frac{1}{n} = 1$$

It would be **incorrect** to say that $X \sim \text{Bin}\left(n, \frac{1}{n}\right)$ because the indicator rvs were NOT independent. □

Example(s)

A factory produces 100 cars per day, but a car is defective with probability 0.02. What's the probability that the factory produces 2 or more defective cars on a given day?

Solution Let X be the number of defective cars that the factory produces. $X \sim \text{Bin}(100, 0.02)$, so

$$\begin{aligned}
\mathbb{P}(X \geq 2) &= 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) && \text{[complement]} \\
&= 1 - \binom{100}{0}(0.02)^0(1-0.02)^{100} - \binom{100}{1}(0.02)^1(1-0.02)^{99} && \text{[plug in binomial PMF]} \\
&\approx 0.5967
\end{aligned}$$

So, there is about a 60% chance that 2 or more cars produced on a given day will be defective. □

3.4.4 Exercises

1. An elementary school wants to keep track of how many of their 200 students have acceptable attendance. Each student shows up to school on a particular day with probability 0.85, independently of other days and students.

   (a) A student has acceptable attendance if they show up to class at least 4 out of 5 times in a school week. What is the probability a student has acceptable attendance?

   (b) What is the probability that at least 170 out of the 200 students have acceptable attendance? Assume students' attendance are independent since they live separately.

   (c) What is the expected number of students with acceptable attendance?

   Solution: Actually, this is a great question because it has nested binomials!

   (a) Let X be the number of school days a student shows up in a school week. Then, $X \sim \text{Bin}(n = 5, p = 0.85)$ since a students' attendance on different days is independent as mentioned earlier. We want $X \geq 4$,

   $$\mathbb{P}(X \geq 4) = \mathbb{P}(X = 4) + \mathbb{P}(X = 5) = \binom{5}{4}0.85^4 0.15^1 + \binom{5}{5}0.85^5 0.15^0 = 0.83521$$

   .

   (b) Let Y be the number of students who have acceptable attendance. Then, $Y \sim \text{Bin}(n = 200, p = 0.83521)$ since each students' attendance is independent of the rest. So,

   $$\mathbb{P}(Y \geq 170) = \sum_{k=170}^{200} \binom{200}{k} 0.83521^k (1 - 0.83521)^{200-k} \approx 0.3258$$

   (c) We have $\mathbb{E}[Y] = np = 200 \cdot 0.83521 = 167.04$ as the expected number of students! We can just cite it now that we've identified Y as being Binomial!

2. [From Stanford CS109] When sending binary data to satellites (or really over any noisy channel) the bits can be flipped with high probabilities. In 1947 Richard Hamming developed a system to more

reliably send data. By using Error Correcting Hamming Codes, you can send a stream of 4 bits with 3 (additional) redundant bits. If zero or one of the seven bits are corrupted, using error correcting codes, a receiver can identify the original 4 bits. Let's consider the case of sending a signal to a satellite where each bit is independently flipped with probability $p = 0.1$. (Hamming codes are super interesting. It's worth looking up if you haven't seen them be- fore! All these problems could be approached using a binomial distribution (or from first principles).)

(a) If you send 4 bits, what is the probability that the correct message was received (i.e. none of the bits are flipped).

(b) If you send 4 bits, with 3 (additional) Hamming error correcting bits, what is the probability that a correctable message was received?

(c) Instead of using Hamming codes, you decide to send 100 copies of each of the four bits. If for every single bit, more than 50 of the copies are not flipped, the signal will be correctable. What is the probability that a correctable message was received?

Solution:

(a) We have $X \sim \text{Bin}(n = 4, p = 0.9)$ to be the number of correct (unflipped) bits. So the binomial PMF says:

$$\mathbb{P}(X = 4) = \binom{4}{4} 0.9^4 (0.1)^{4-0} = 0.9^4 = 0.656$$

Note we could have also approached this by letting $Y \sim \text{Bin}(4, 0.1)$ be the number of corrupted (flipped) bits, and computing $\mathbb{P}(Y = 0)$. This is the same result!

(b) Let Z be the number of corrupted bits, then $Z \sim \text{Bin}(n = 7, p = 0.1)$, so we can use its PMF. A message is correctable if $Z = 0$ or $Z = 1$ (mentioned above), so

$$\mathbb{P}(Z = 0) + \mathbb{P}(Z = 1) = \binom{7}{0} 0.1^7 0.9^0 + \binom{7}{1} 0.1^6 0.9^1 = 0.850$$

This is a 30% (relative) improvement compared to above by just using 3 extra bits!

(c) For $i = 1, \ldots, 4$, let $X_i \sim \text{Bin}(n = 100, p = 0.9)$. We need $X_1 > 50$, $X_2 > 50$, $X_3 > 50$, and $X_4 > 50$ for us to get a correctable message. For $X_i > 50$, we just sum the binomial PMF from 51 to 100:

$$\mathbb{P}(X_i > 50) = \sum_{k=51}^{100} \binom{100}{k} 0.9^k (1-p)^{100-k}$$

Then, since we need all 4 to work, by independence, we get

$$\begin{aligned}
\mathbb{P}(X_1 > 50, X_2 > 50, X_3 > 50, X_4 > 50) &= \mathbb{P}(X_1 > 50)\,\mathbb{P}(X_2 > 50)\,\mathbb{P}(X_3 > 50)\,\mathbb{P}(X_4 > 50) \\
&= \left(\sum_{k=51}^{100} \binom{100}{k} 0.9^k (1-p)^{100-k} \right)^4 \\
&> 0.999
\end{aligned}$$

But this required 400 bits instead of just the 7 required by Hamming codes! This is well worth the tradeoff.

3. Suppose A and B are random, independent (possibly empty) subsets of $\{1, 2, ..., n\}$, where each subset is equally likely to be chosen. Consider $A \cap B$, i.e., the set containing elements that are in both A and B. Let X be the random variable that is the size of $A \cap B$. What is $\mathbb{E}[X]$?

Solution: Then, $X \sim Bin\left(n, \frac{1}{4}\right)$, so $\mathbb{E}[X] = \frac{n}{4}$ (since we know the expected value of $Bin(n,p)$ is np). How did we do that??

Choosing a random subset of $\{1, \ldots, n\}$ can be thought of as follows: for each element $i = 1, \ldots, n$, with probability 1/2 take the element (and with probability 1/2 don't take it), independently of other elements. This is a crucial observation.

For each element $i = 1, \ldots, n$, the element is either in $A \cap B$ or not. So let X_i be the indicator/Bernoulli rv of whether $i \in A \cap B$ or not. Then, $\mathbb{P}(X_i = 1) = \mathbb{P}(i \in A, i \in B) = \mathbb{P}(i \in A)\,\mathbb{P}(i \in B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ because A, B are chosen independently, and each element is in A or B with probability 1/2. Note that these X_i's are independent because one element being in the set does not affect another element being in the set. Hence, $X = \sum_{i=1}^{n} X_i$ the number of elements in our intersection, so $X \sim \text{Bin}\left(n, \frac{1}{4}\right)$ and $\mathbb{E}[X] = np = \frac{n}{4}$.

Note that it was not necessary that these variables were independent; we could have still applied linearity of expectation anyway to get $\frac{n}{4}$. We just wouldn't have been able to say $X \sim \text{Bin}\left(n, \frac{1}{4}\right)$.