Lisa Yan and Jerry Cain
CS109

Lecture Notes #7
September 28, 2020

Bernoulli and Binomial Random Variables

Based on a chapter by Chris Piech and Lisa Yan

Pre-recorded lecture: Sections 1 and 2 (definitions only).
In-lecture: Section 3.

There are some classic random variable abstractions that show up in many problems. At this point in the class you will learn about several of the most significant discrete distributions. When solving problems, if you are able to recognize that a random variable fits one of these formats, then you can use its precalculated probability mass function (PMF), expectation, variance, and other properties.

Random variables of this sort are called "parametric" random variables. If you can argue that a random variable falls under one of the studied parametric types, you simply need to provide parameters. A good analogy is a *class* in programming. Creating a parametric random variable is very similar to calling a constructor with input parameters.

1 Bernoulli Random Variable

A **Bernoulli random variable** is the simplest kind of random variable. It can take on two values, 1 and 0. It takes on a 1 if an experiment with probability p resulted in success and a 0 otherwise. Some example uses include a coin flip, a random binary digit, whether a disk drive crashed, and whether someone likes a Netflix movie.

If X is a Bernoulli random variable, denoted $X \sim \text{Ber}(p)$:

Probability mass function:	$P(X = 1) = p$ $P(X = 0) = (1 - p)$
Expectation:	$E[X] = p$
Variance:	$\text{Var}(X) = p(1 - p)$

Bernoulli random variables and **indicator variables** are two aspects of the same concept. As a review, a random variable I is called an indicator variable for an event A if $I = 1$ when A occurs and $I = 0$ if A does not occur. $P(I = 1) = P(A)$ and $E[I] = P(A)$. Indicator random variables *are* Bernoulli random variables, with $p = P(A)$.

2 Binomial Random Variable

A **Binomial random variable** is random variable that represents the number of successes in n successive independent trials of a Bernoulli experiment. Some example uses include the number of heads in n coin flips, the number of disk drives that crashed in a cluster of 1000 computers, and the number of advertisements that are clicked when 40,000 are served.

If X is a Binomial random variable, we denote this $X \sim \text{Bin}(n, p)$, where p is the probability of success in a given trial. A binomial random variable has the following properties:

$$
\begin{aligned}
&P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{if } k \in \mathbb{N}, 0 \le k \le n \text{ (0 otherwise)} \\
&E[X] = np \\
&\text{Var}(X) = np(1-p)
\end{aligned}
$$

There is a strong relationship between the Binomial random variable and the Bernoulli random variable—in that a Binomial RV is the sum of n independent Bernoulli RVs. Next week we'll talk more about what independence in the context of random variables means, but for now:

Let $X_i \sim \text{Ber}(p)$, for $i = 1, \ldots, n$. Let $Y = \sum_{i=1}^{n} X_i$. If the X_i are independent, then $Y \sim \text{Bin}(n, p)$. Another way to think about this is that for the Binomial RV Y to take on value k, it must be true that there are exactly k of the X_i's take on value 1, and all other X_i's must take on value 0. There are $\binom{n}{k}$ ways to pick which X_i's will have value 1. (after which we set the rest to 0). When we sum up these k ones and $n - k$ zeros, we get $Y = k$.

Example: Coin flips

Let X = number of heads after a coin is flipped three times. $X \sim \text{Bin}(3, 0.5)$. What is the probability of each of the different values of X?

$$
\begin{aligned}
P(X = 0) &= \binom{3}{0} p^0 (1-p)^3 = \frac{1}{8} \\
P(X = 1) &= \binom{3}{1} p^1 (1-p)^2 = \frac{3}{8} \\
P(X = 2) &= \binom{3}{2} p^2 (1-p)^1 = \frac{3}{8} \\
P(X = 3) &= \binom{3}{3} p^3 (1-p)^0 = \frac{1}{8}
\end{aligned}
$$

3 Exercises: Binomial random variables

Example: The NBA finals

It's 2019, and the Bay Area basketball team the Golden State Warriors are going to play the Toronto Raptors in a 7-game series during the NBA finals. A team wins the series if they win at least 4 games (suppose that in this world, teams must play all 7 games). The Warriors have a probability of 58% of winning each game, independently. What is the probability of the Warriors winning the series?

Solution: Let X be a random variable representing the number of games the Warriors win. $X \sim \text{Bin}(7, 0.58)$. $P(X \geq 4) = \sum_{k=4}^{7} P(X = k) = \sum_{k=4}^{7} \binom{7}{k}(0.58)^k(0.42)^{7-k}$.

Example: Generic inheritance

Suppose that there is a specific pair of genes that determines eye color. A child receives 1 gene from each parent (where a parent is equally likely to give either of their own two genes to their child). Further suppose that brown is a "dominant" eye color—where a child will have brown eyes if either (or both) genes are brown—and blue is a "recessive" eye color, where a child will have blue eyes only if both genes are blue. The eye colors of children are independent given their parents' genes.

Suppose two parents—who each have 1 brown and 1 blue gene—have 4 biological children. What is the probability that exactly 3 of the 4 children have brown eyes?

Solution: Let X be the number of brown-eyed children in the family. $X \sim \text{Bin}(4, p)$, where there are 4 children ("independent trials") and p is the probability that a child has brown eyes. A child has brown eyes if they do not have blue eyes (which happens with probability $(0.5)^2$), and therefore $p = 1 - 0.25 = 0.75$. Therefore $P(X = 3) = \binom{4}{3}(0.75)^3(0.25)^1$.

Example: Hamming code

When sending messages over a network, there is a chance that the bits will be corrupted. A Hamming code allows for a 4 bit code to be encoded as 7 bits, with the advantage that if 0 or 1 bit(s) are corrupted, then the message can be perfectly reconstructed. You are working on the Voyager space mission and the probability of any bit being lost in space is 0.1. How does reliability change when using a Hamming code?

Image we use error correcting codes. Let $X \sim \text{Bin}(7, 0.1)$.

$$\begin{aligned} P(X = 0) &= \binom{7}{0}(0.1)^0(0.9)^7 \approx 0.468 \\ P(X = 1) &= \binom{7}{1}(0.1)^1(0.9)^6 = 0.372 \\ P(X = 0) + P(X = 1) &= 0.850 \end{aligned}$$

What if we didn't use error correcting codes? Let $X \sim \text{Bin}(4, 0.1)$.

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

Using Hamming Codes improves reliability by about 30%!