



Slide Deck C2:

Bernoulli and Binomial Distributions

The section in which we learn about two named discrete distributions. When a distribution is encountered frequently, we name it and try to calculate formulas for its parameters. The Binomial distribution is one of the most useful distributions when dealing with categorical random variables. When a random variable follows a Binomial distribution, it is the number of successes out of the number of trials.

Start of Lecture Material
Bernoulli Distributions
Binomial Distributions
Five or So Examples
End of Lecture Material

Today's Objectives

Today's Objectives

By the end of this slidedeck, you should

- ➊ determine what random variable follow a Bernoulli distribution using its definition
- ➋ calculate the expected value, variance, median, and probabilities associated with a **Bernoulli** random variable
- ➌ determine what random variable follow a Binomial distribution using its definition
- ➍ calculate the mean and variance of a **Binomial** distribution
- ➎ determine if the Bernoulli and Binomial distributions are skewed
- ➏ calculate probabilities from a Binomial distribution

Definition of Bernoulli Experiment

Definition

A random variable with only two possible outcomes follows a **Bernoulli distribution** with success probability parameter, p .

Examples

- head on a single flip of a coin
- getting a 6 on one roll of a die
- correct choice on a True/False question
- using a credit card for a transaction
- voting in an election
- stopping at Starbucks in the morning
- war breaking out between two countries in a year

Bernoulli Probability Mass Function

Recall that the probability mass function (pmf) provides the probability of each element of the sample space. For a Bernoulli random variable, there are only two outcomes: failure and success — 0 and 1. From this, the probability mass function for a Bernoulli random variable is always

$$\mathbb{P}[X = x] = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & \text{Otherwise} \end{cases}$$

This can also be written as

$$\mathbb{P}[X = x] = p^x (1 - p)^{1-x}$$

Why should we use this form? It helps create a closely-related distribution.

Bernoulli Expected Value

The expected value of a Bernoulli random variable is

$$\begin{aligned}\mathbb{E}[X] &:= \sum_{x \in S} x \mathbb{P}[X = x] \\ &= 0 \mathbb{P}[X = 0] + 1 \mathbb{P}[X = 1] \\ &= 0(1 - p) + 1(p) \\ &= p\end{aligned}$$

Question: Does this result seem reasonable?

Bernoulli Variance

The variance of a Bernoulli random variable is

$$\begin{aligned}\mathbb{V}[X] &:= \sum_{x \in S} (x - \mu)^2 \mathbb{P}[X = x] \\ &= (0 - p)^2(1 - p) + (1 - p)^2(p) \\ &= p p (1 - p) + (1 - p) (1 - p) p \\ &= p(1 - p) (p + (1 - p)) \\ &= p(1 - p)\end{aligned}$$

Question: For what value of p are we most uncertain of a Bernoulli outcome?

Bernoulli Cumulative Distribution Function (CDF)

Recall the probability mass function (pmf):

$$\mathbb{P}[X = x] = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & \text{Otherwise} \end{cases}$$

The cumulative distribution function (CDF) is the function defined as $F(x) = \mathbb{P}[X \leq x]$. The CDF for a Bernoulli is

$$\mathbb{P}[X \leq x] = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$

Note: The CDF will become very important when we deal with continuous distributions.

Bernoulli Median

The median depends on the Bernoulli parameter, p . The rule to calculate the median is to start low and calculate the *cumulative* probability until that sum is at least 0.500. This is most easily done with a CDF.

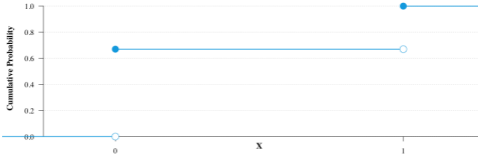
$$\mathbb{P}[X \leq x] = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$

The median of a Bernoulli random variable is

$$\tilde{X} = \begin{cases} 0 & p < 0.500 \\ [0, 1] & p = 0.500 \\ 1 & p > 0.500 \end{cases}$$

Bernoulli Median

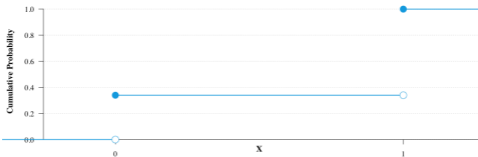
Why is that the formula for the median?



Think: As I increase x , when does the cumulative probability attain 0.500, when $p < 0.500$ (as here)? That will be the median, $\tilde{X} = 0$.

Bernoulli Median

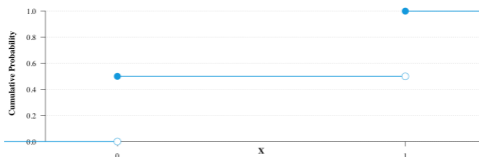
Why is that the formula for the median?



Think: As I increase x , when does the cumulative probability attain 0.500, when $p > 0.500$ (as here)? That will be the median, $\tilde{X} = 1$.

Bernoulli Median

Why is that the formula for the median?



Think: As I increase x , when does the cumulative probability attain 0.500, when $p = 0.500$ (as here)? That will be the median, $0 \leq \tilde{X} < 1$ (or $\tilde{X} = \frac{1}{2}$).

Definition of Binomial Random Variable

Definition

A **Binomial random variable** is the sum of n independent and identically distributed Bernoulli random variables.

This definition leads to the **five requirements of a Binomial random variable**:

- ➊ The number of trials, n , is known
- ➋ each trial has two possible outcomes (is Bernoulli)
- ➌ the success probability, p , is constant from trial to trial
- ➍ the trials are independent
- ➎ the random variable is the number of successes

Note that this means the sample space is $S = \{0, 1, \dots, n\}$.

Examples of Binomial Distribution

Examples of Binomial random variables are

- times getting heads on n flips of a coin
- times getting a 6 on n rolls of a die
- times getting the correct choice on n True/False questions
- times using a credit card for n transactions
- times voting in n elections
- times stopping at Starbucks in n mornings
- times war breaks out between two countries in n years

Notice the difference between Binomial rvs and Bernoulli rvs: the Bernoulli is the number of successes in *one* trial; the Binomial is the number of successes in n trials.

Binomial Probability Mass Function

Recall that a Binomial random variable is just the sum of n Bernoulli random variables. From this, we will be able to determine the pmf of the Binomial:

$$\mathbb{P}[X = x] = \underbrace{\binom{n}{x}}_{\text{combinations}} \underbrace{p^x (1-p)^{n-x}}_{\text{Bernoulli probabilities}}$$

This can be further broken down:

$$\mathbb{P}[X = x] = \underbrace{\binom{n}{x}}_{\text{combinations}} \underbrace{p^x}_{\text{successes}} \underbrace{(1-p)^{n-x}}_{\text{failures}}$$

Binomial Parameters

The expected value and variance of a Binomial are easily calculated from its statistical definition:

Definition (Binomial Random Variable)

Let $X_i \stackrel{\text{ind}}{\sim} \mathcal{Bern}(p)$. Define $Y = \sum_{i=1}^n X_i$. Then $Y \sim \mathcal{Bin}(n, p)$.

In words: Let us have a sample of n independent and identically distributed Bernoulli random variables. Define Y as the sum of those random variables. By definition, Y follows a Binomial distribution with parameters n and p .

In fact, the proof seems *too* easy, as you will see.

Binomial Parameters (Expected Value)

Using the statistical definition, the expected value is easily calculated using a property of the expected value:

$$\begin{aligned}\mathbb{E}[Y] &= \mathbb{E}\left[\sum_{i=1}^n X_i\right] \\ &= \sum_{i=1}^n \mathbb{E}[X_i] \\ &= \sum_{i=1}^n p \\ &= n p\end{aligned}$$

Of course: We should stop and make sure that this result makes sense.

Binomial Parameters (Variance)

Similarly, the formula for the variance is also easily calculated:

$$\begin{aligned}\mathbb{V}[Y] &= \mathbb{V}\left[\sum_{i=1}^n X_i\right] \\ &= \sum_{i=1}^n \mathbb{V}[X_i] \\ &= \sum_{i=1}^n p(1-p) \\ &= n p(1-p)\end{aligned}$$

Aside: The step from 1 to 2 requires that the Bernoulli experiments are independent.

Binomial Example 1: Fair Coins

Example

Let Y be the number of heads flipped in $n = 10$ flips of a coin. If the coin is fair, then it is clear that $Y \sim \text{Bin}(n = 10, p = 0.500)$.

- ➊ What is the probability of getting three heads?
- ➋ What is the probability of getting three tails?
- ➌ What is the probability of getting at most three heads or at least seven heads?

Binomial Example 1: Fair Coins

Example

Let Y be the number of heads flipped in $n = 10$ flips of a coin. If the coin is fair, $Y \sim \text{Bin}(n = 10, p = 0.500)$.

- What is the probability of getting three heads?

We are asked to calculate $\mathbb{P}[Y = 3]$. This is a simple application of the pmf:

$$\begin{aligned}\mathbb{P}[Y = y] &= \binom{n}{y} p^y (1-p)^{n-y} \\ \mathbb{P}[Y = 3] &= \binom{10}{3} 0.500^3 (1 - 0.500)^{10-3} \\ &= (120)(0.125)(0.0078125) \\ &= 0.1171875\end{aligned}$$

Binomial Example 1: Fair Coins

In R, this is

```
dbinom(3, size=10, prob=0.500)
```

Binomial Example 1: Fair Coins

Example

Let Y be the number of heads flipped in $n = 10$ flips of a coin. If the coin is fair, $Y \sim \text{Bin}(n = 10, p = 0.500)$.

- What is the probability of getting three tails?

We are asked to calculate $\mathbb{P}[Y = 7]$. This is a simple application of the pmf:

$$\begin{aligned}\mathbb{P}[Y = y] &= \binom{n}{y} p^y (1-p)^{n-y} \\ \mathbb{P}[Y = 7] &= \binom{10}{7} 0.500^7 (1 - 0.500)^{10-7} \\ &= (120)(0.0078125)(0.125) \\ &= 0.1171875\end{aligned}$$

Binomial Example 1: Fair Coins

In R, this is

```
dbinom(7, size=10, prob=0.500)
```

Binomial Example 1: Fair Coins

Example

Let Y be the number of heads flipped in $n = 10$ flips of a coin. If the coin is fair, $Y \sim \text{Bin}(n = 10, p = 0.500)$.

- What is the probability of getting at most three heads or at least seven heads?

We are asked to calculate $\mathbb{P}[Y \leq 3 \cup Y \geq 7]$:

$$\begin{aligned}\mathbb{P}[Y \leq 3 \cup Y \geq 7] &= \mathbb{P}[Y \leq 3] + \mathbb{P}[Y \geq 7] \\ &= \mathbb{P}[Y = 0] + \mathbb{P}[Y = 1] + \mathbb{P}[Y = 2] + \mathbb{P}[Y = 3] + \\ &\quad \mathbb{P}[Y = 7] + \mathbb{P}[Y = 8] + \mathbb{P}[Y = 9] + \mathbb{P}[Y = 10]\end{aligned}$$

Binomial Example 1: Fair Coins

Example

Let Y be the number of heads flipped in $n = 10$ flips of a coin. If the coin is fair, $Y \sim \text{Bin}(n = 10, p = 0.500)$.

- What is the probability of getting at most three heads or at least seven heads?

To do this in [R](#) (or most other statistical packages like Excel), the problem needs to be rewritten using cumulative probabilities, $\mathbb{P}[Y \leq y]$. This is easily done using complements from Chapter 4:

$$\mathbb{P}[Y \geq 7] = 1 - \mathbb{P}[Y \leq 6]$$

Thus, $\mathbb{P}[Y \leq 3 \cup Y \geq 7]$ is

$$\mathbb{P}[Y \leq 3] + (1 - \mathbb{P}[Y \leq 6])$$

Binomial Example 1: Fair Coins

Example

Let Y be the number of heads flipped in $n = 10$ flips of a coin. If the coin is fair, $Y \sim \text{Bin}(n = 10, p = 0.500)$.

- What is the probability of getting at most three heads or at least seven heads?

In R,

$$\mathbb{P}[Y \leq 3] + (1 - \mathbb{P}[Y \leq 6])$$

is

```
= pbinom(3, size=10, prob=0.500) +  
  (1-pbinom(6, size=10, prob=0.500))  
  
= 0.34375
```

Binomial Example 2: Interstate War

Example

Given two randomly-selected countries, the probability of war between them in a given year is one in ten-thousand $p = 0.0001$. Calculate the probability of *at least one* war in a given year.

Binomial Example 2: Interstate War

Solution:

There are currently 194 countries in the world (members of the United Nations). This means there are $\binom{194}{2} = 18,721$ pairs of countries. If W is the number of wars in a given year, we are asked to calculate $P[W \geq 1]$.

This is just

$$P[W \geq 1] = \sum_{w=1}^{18,721} \binom{18,721}{w} 0.0001^w 0.9999^{18,721-w}$$

Note that this is a rather extensive summation (18,721 terms to add). It would be easier if we could calculate this with fewer steps.

Binomial Example 2: Interstate War

Instead of summing 18,721 terms, we can use **complements** to reduce the number of terms needed in the summation.

Complementary Events: Note that *all* values of W can be placed in one of these two sets

$$\{W < 1\} \qquad \{W \geq 1\}$$

Because of this, we know

$$1 = P[W < 1] + P[W \geq 1]$$

or, more usefully,

$$P[W \geq 1] = 1 - P[W < 1]$$

Binomial Example 2: Interstate War

Thus, using complements, the probability of war in a given year is

$$\begin{aligned}\mathbb{P}[W \geq 1] &= 1 - \mathbb{P}[W < 1] \\ &= 1 - \mathbb{P}[W \leq 0] \\ &= 1 - \binom{18721}{0} 0.0001^0 0.9999^{18721-0} \\ &= 1 - 0.1537859 \\ &= 0.8462141\end{aligned}$$

Binomial Example 2: Interstate War

In R, this is

```
1 - pbinom(0, size=18721, prob=0.0001)
```

Thus, even though war is relatively rare between two randomly-selected countries, it is not rare between *every* two countries.

Binomial Example 3: Puppy Cancer

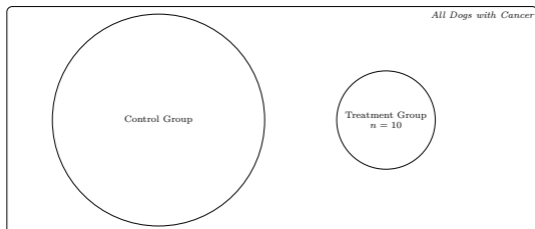
Example

While at Knox, Natalie (a former student of mine) studied an unfortunately not-too-rare type of canine cancer (*hemangiosarcoma*) and a medicine that is supposed to increase canine longevity. Without the medicine, the median survival time for the dog is 90 days from detection.

She wanted to determine if the medicine helped dogs live longer. To do this, she gave the treatment to 10 dogs and measured how long they lived. Of the 10 dogs given the new treatment, eight lived longer than 90 days.

What is the probability of observing this (or something even *less* likely) if the medicine does *not* work?

Binomial Example 4: Time Use



Binomial Example 3: Puppy Cancer

Solution:

If D is the number of dogs surviving beyond 90 days, then we are asked to calculate $\mathbb{P}[D \geq 8]$, given $D \sim \text{Bin}(10, 0.500)$. This is just

$$\begin{aligned}\mathbb{P}[D \geq 8] &= \sum_{d=8}^{10} \binom{10}{d} 0.500^d (1 - 0.500)^{10-d} \\ &= 1 - \sum_{d=0}^7 \binom{10}{d} 0.500^d (1 - 0.500)^{10-d} \\ &= 1 - \mathbb{P}[D \leq 7]\end{aligned}$$

Have fun doing this calculation by hand.

Binomial Example 3: Puppy Cancer

Here it is in R:

```
1 - pbinom(7, size=10, prob=0.500) = 0.0546875.
```

Conclusion:

Thus, if the medicine did not help, we would experience this particular reality about one time in 19 (5.5%). Since this seems to be a relatively rare event, we can conclude that this medicine is worthy of further research.

Binomial Example 4: Time Use

Example

Yesol, another former student of mine, developed a calendar app to increase productivity. As a part of her Statistics minor research project, she wanted to determine if using her calendar app would help people better manage their time.

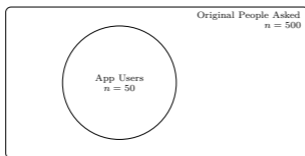
To test this, she collected asked 500 people whether they felt they managed their time well. Of those people, 80% stated they managed their time “well.” Then, she provided the app to 50 of those people, asking them to use it.

After a couple of months, she surveyed those 50 to see how many now claimed they managed their time “well.” Of those 50, a total of 48 said they now manage their time “well.”

Does this provide sufficient evidence that her app helped people manage their time?

Binomial Example 4: Time Use

All People



Binomial Example 4: Time Use

Solution:

If X is the number of people managing their time “well,” then we are asked to calculate $\mathbb{P}[X \geq 48]$, given $X \sim \text{Bin}(50, 0.800)$. This is just

$$\mathbb{P}[X \geq 48] = \sum_{x=48}^{50} \binom{50}{x} 0.800^x (1 - 0.800)^{10-x}$$

Using complements, we can write this in terms of the CDF:

$$\begin{aligned} &= 1 - \mathbb{P}[X < 48] \\ &= 1 - \mathbb{P}[X \leq 47] \end{aligned}$$

Again, calculating this by hand is time-prohibitive.

Binomial Example 4: Time Use

Here it is in R:

```
1 - pbinom(47, size=50, prob=0.800) = 0.001285415.
```

Or, if you wanted to perform the original calculations:

```
dbinom(48, size=50, prob=0.800) + dbinom(49, size=50, prob=0.800) +  
dbinom(50, size=50, prob=0.800)
```

This gives the same result.

Conclusion:

Because this probability is so low ($p\text{-value} = 0.0013$), there is significant evidence that those people using her app **are** more likely to (think that they) manage their time well.

Binomial Example 5: Warm Fuzzies

Example

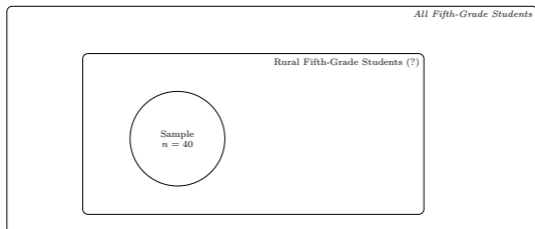
Back at Oklahoma State University, a Ph.D. candidate in Education approached me asking for help with the statistical analysis of her work. She devised a Teaching Unit aimed at improving feelings towards science for fifth-grade students.

The data she provided to me consisted of scores on two tests: a pre-test (before teaching the Unit) and a post-test (taken after she taught the Unit). The teaching took place for $n = 40$ students in a small, rural school in North Dakota.

The pre-test suggested that only 25% of fifth-grade students felt warmth towards science. Of the 40 students who took the post-test, 12 reported positive (warm) feelings towards science.

Does this provide evidence that her Teaching Unit help students feel warmer towards science?

Binomial Example 5: Warm Fuzzies



Binomial Example 5: Warm Fuzzies

Solution:

If S is the number of students feeling warm towards science, then we are asked to calculate $\mathbb{P}[S \geq 12]$, given $S \sim \text{Bin}(40, 0.25)$. This is just

$$\begin{aligned}\mathbb{P}[S \geq 12] &= \sum_{s=12}^{40} \binom{40}{s} 0.25^s (1 - 0.25)^{40-s} \\ &= 1 - \mathbb{P}[S \leq 11]\end{aligned}$$

Have fun doing this calculation by hand. Here it is in R:

```
1 - pbinom(11, size=40, prob=0.25) = 0.2848556.
```

Because this probability is not low, there is *little evidence* that students experiencing her Teaching Unit are more likely to feel warm to science.

Today's Objectives

Now that we have concluded this lecture, you should be able to

- ➊ determine what random variable follow a Bernoulli distribution using its definition
- ➋ calculate the expected value, variance, median, and probabilities associated with a **Bernoulli** random variable
- ➌ determine what random variable follow a Binomial distribution using its definition
- ➍ calculate the mean and variance of a **Binomial** distribution
- ➎ determine if the Bernoulli and Binomial distributions are skewed
- ➏ calculate probabilities from a Binomial distribution

Today's R Functions

In this slide deck, we covered the following R functions:

- `dbinom(x, size, prob)` is the pmf, p for: $\mathbb{P}[X = x] = p$
- `pbinom(x, size, prob)` is the CDF, p for: $\mathbb{P}[X \leq x] = p$
- `qbinom(p, size, prob)` is the quantile function, x for: $\mathbb{P}[X \leq x] = p$
- `rbinom(n, size, prob)` generates n random values from: $\text{Bin}(\text{size}, \text{prob})$

Please do not forget to access the `allProbabilities` document that provides all of the probability functions in R.

Supplemental Activities

The following may be of interest to you in terms of today's topics:

- SCA 5a is for some discrete distributions

Note that you can access all Statistical Computing Activities here:

<https://www.kvasaheim.com/courses/stat200/sca/>

In addition to the SCA, **Laboratory Activity B** is helpful for learning how to handle discrete distributions (including the Binomial distribution). The lab actually shows the connection between sampling and discrete distributions. It uses three named distributions.

<https://www.kvasaheim.com/courses/stat200/labs/>

Supplemental Readings

The following are some readings that may be of interest to you in terms of understanding discrete distributions:

- Hawkes Learning: Section 5.2
- Intro to Modern Statistics: None
- [R](#) for Starters: Appendix A.2
Appendix A.3
- Wikipedia: Bernoulli Distribution
Binomial Distribution