CSSS/SOC/STAT 321
Case-Based Statistics I

Random Variables & Probability Distributions I: Discrete Distributions

Christopher Adolph

Department of Political Science

and

Center for Statistics and the Social Sciences

University of Washington, Seattle

Two examples

Flipping coins In the play *Rosenscratz and Guildenstern are dead*, R & G flip a coin repeatedly, and get heads 89 times. How likely is this outcome? In general, how likely is it get X *or more* heads out of N flips? Can you think of real-world situations like this one?

Two examples

Flipping coins In the play *Rosenscratz and Guildenstern are dead*, R & G flip a coin repeatedly, and get heads 89 times. How likely is this outcome? In general, how likely is it get X *or more* heads out of N flips? Can you think of real-world situations like this one?

Anticipating a landslide A Democrat-controlled state legislature redistricts a 55% Democratic state so that they have a 55% majority in every district. They say they expect to win all 20 Congressional seats in the state. “Majority rules, after all.” Are they right?

Sample spaces & sets

One way to understand sample spaces is to list every outcome

These lists are *sets*, or collections of elements, which could be numbers

$$\mathcal{A} = \{23, 5.3, 1000, 4\}$$

Sample spaces & sets

One way to understand sample spaces is to list every outcome

These lists are *sets*, or collections of elements, which could be numbers

$$\mathcal{A} = \{23, 5.3, 1000, 4\}$$

But they need not be quantitative at all,

$$\mathcal{A} = \{\text{Democrat}, \text{Republican}, \text{Green}\}$$

Sample spaces & sets

One way to understand sample spaces is to list every outcome

These lists are *sets*, or collections of elements, which could be numbers

$$\mathcal{A} = \{23, 5.3, 1000, 4\}$$

But they need not be quantitative at all,

$$\mathcal{A} = \{\text{Democrat}, \text{Republican}, \text{Green}\}$$

And we will for now leave them as mathematical objects

$$\mathcal{A} = \{a_1, a_2, a_3\}$$

Sample spaces & sets

One way to understand sample spaces is to list every outcome

These lists are *sets*, or collections of elements, which could be numbers

$$\mathcal{A} = \{23, 5.3, 1000, 4\}$$

But they need not be quantitative at all,

$$\mathcal{A} = \{\text{Democrat}, \text{Republican}, \text{Green}\}$$

And we will for now leave them as mathematical objects

$$\mathcal{A} = \{a_1, a_2, a_3\}$$

a_1 is an element of $\mathcal{A}$, which we write $a_1 \in \mathcal{A}$

Sample spaces & sets

One way to understand sample spaces is to list every outcome

These lists are *sets*, or collections of elements, which could be numbers

$$\mathcal{A} = \{23, 5.3, 1000, 4\}$$

But they need not be quantitative at all,

$$\mathcal{A} = \{\text{Democrat}, \text{Republican}, \text{Green}\}$$

And we will for now leave them as mathematical objects

$$\mathcal{A} = \{a_1, a_2, a_3\}$$

a_1 is an element of $\mathcal{A}$, which we write $a_1 \in \mathcal{A}$
A set may also be empty, e.g., $\mathcal{B} = \emptyset = \{\}$

Sample space for the coin flip example

Suppose we toss a coin twice and record the results.

We can use a set to record this complex event.

For example, we might see a head and a tail, or H, T.

Sample space for the coin flip example

Suppose we toss a coin twice and record the results.

We can use a set to record this complex event.

For example, we might see a head and a tail, or H,T.

The *sample space*, or universe of possible results, is in this case a set of sets:

$$\Omega = \{\{H,H\},\{H,T\},\{T,H\},\{T,T\}\}$$

Note that our sample space has separate entries for every *ordering* of heads or tails we could see.

Sample space for two dice

Now suppose that we roll two dice. The sample space is:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Sample space for two dice

Now suppose that we roll two dice. The sample space is:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

The sum of the dice rolls for each event:

2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

Note that many sums repeat

Sample spaces and complex events

sum of the dice rolls for each event

2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

What are the odds?

Outcome	2	3	4	5	6	7
Frequency						
Probability						

Outcome	8	9	10	11	12
Frequency					
Probability					

Sample spaces and complex events

sum of the dice rolls for each event

2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

What are the odds?

Outcome	2	3	4	5	6	7
Frequency	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$
Probability						

Outcome	8	9	10	11	12
Frequency	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
Probability					

Sample spaces and complex events

sum of the dice rolls for each event

2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

Each event (a, b) is equally likely. But each sum, $a + b$, is not.

Outcome	2	3	4	5	6	7
Frequency	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$
Probability	0.028	0.056	0.115	0.111	0.139	0.167

Outcome	8	9	10	11	12
Frequency	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
Probability	0.139	0.111	0.083	0.056	0.028

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result
military casualties	each soldier's status

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result
military casualties	each soldier's status
education outcomes	each student's passing status

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result
military casualties	each soldier's status
education outcomes	each student's passing status
presidential popularity	each citizen's opinion
...	...

How can we reduce the space to something manageable?

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result
military casualties	each soldier's status
education outcomes	each student's passing status
presidential popularity	each citizen's opinion
...	...

How can we reduce the space to something manageable?

$\rightarrow$ map the sample space Ω to one or more *random variables*:
Ω for coin flips $\quad \rightarrow \quad X = \#$ of heads

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result
military casualties	each soldier's status
education outcomes	each student's passing status
presidential popularity	each citizen's opinion
...	...

How can we reduce the space to something manageable?

$\rightarrow$ map the sample space Ω to one or more *random variables*:

Ω for coin flips $\rightarrow$ $X = \text{\# of heads}$

Ω for military casualties $\rightarrow$ $D = \text{\# of deaths}$

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result
military casualties	each soldier's status
education outcomes	each student's passing status
presidential popularity	each citizen's opinion
...	...

How can we reduce the space to something manageable?

$\rightarrow$ map the sample space Ω to one or more *random variables*:

Ω for coin flips $\rightarrow$ $X = \#$ of heads
Ω for military casualties $\rightarrow$ $D = \#$ of deaths
Ω for education outcomes $\rightarrow$ $Y = \#$ of students passing
Ω for presidential popularity $\rightarrow$ $S = \#$ support pres

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result
military casualties	each soldier's status
education outcomes	each student's passing status
presidential popularity	each citizen's opinion
...	...

How can we reduce the space to something manageable?

$\rightarrow$ map the sample space Ω to one or more *random variables*:

Ω for coin flips	$\rightarrow$	$X = \#$ of heads
Ω for military casualties	$\rightarrow$	$D = \#$ of deaths
Ω for education outcomes	$\rightarrow$	$Y = \#$ of students passing
Ω for presidential popularity	$\rightarrow$	$S = \#$ support pres

This mapping can produce discrete or continous variables,
and each will have a different distribution of probabilities

Probability for random variables

Consider the random variable $X = \#$ of heads in M coin flips

Five things we'd like to know about the theoretical distribution of X:

$\Pr(X) = f(X)$ How do we summarize the random distribution of X?

Probability for random variables

Consider the random variable $X = \#$ of heads in M coin flips

Five things we'd like to know about the theoretical distribution of X:

$\Pr(X) = f(X)$ How do we summarize the random distribution of X?

$\Pr(X = x)$ What is the probability X is some specific value like $x = 1$?

Probability for random variables

Consider the random variable $X = \#$ of heads in M coin flips

Five things we'd like to know about the theoretical distribution of X:

$\Pr(X) = f(X)$ How do we summarize the random distribution of X?

$\Pr(X = x)$ What is the probability X is some specific value like $x = 1$?

$\Pr(X \leq x)$ What is the probability that X is *at least* equal to some specific value, like $x = 1$?

Probability for random variables

Consider the random variable $X = \#\text{ of heads in } M \text{ coin flips}$

Five things we'd like to know about the theoretical distribution of X:

$\Pr(X) = f(X)$ How do we summarize the random distribution of X?

$\Pr(X = x)$ What is the probability X is some specific value like $x = 1$?

$\Pr(X \leq x)$ What is the probability that X is *at least* equal to some specific value, like $x = 1$?

$\mathrm{E}(X)$ What is the expected number of heads we will see on average?

Probability for random variables

Consider the random variable $X = \#\text{ of heads in } M \text{ coin flips}$

Five things we'd like to know about the theoretical distribution of X:

$\Pr(X) = f(X)$ How do we summarize the random distribution of X?

$\Pr(X = x)$ What is the probability X is some specific value like $x = 1$?

$\Pr(X \leq x)$ What is the probability that X is *at least* equal to some specific value, like $x = 1$?

$\mathrm{E}(X)$ What is the expected number of heads we will see on average?

$\mathrm{sd}(X)$ On average, how much do we expect a given random outcome to differ from the expected result?

The probability distribution of a random variable

We want to describe the behavior of a sum of coin fips

Let's start smaller: a single coin flip, $X = 1$ if heads, and $X = 0$ if tails

The probability distribution of a random variable

We want to describe the behavior of a sum of coin fips

Let's start smaller: a single coin flip, $X = 1$ if heads, and $X = 0$ if tails

We start by defining a *probability distribution*, or mathematical function which the random variable should follow in theory

Let

$$\Pr(X = x) = f(X = x)$$

Note that we haven't yet spelled out what $f(\cdot)$ looks like.

The probability distribution of a random variable

$$\Pr(X = x) = f(X = x)$$

The probability distribution of a random variable

$$\Pr(X = x) = f(X = x)$$

Sometimes we will have a lot of choice in specifying $f(\cdot)$;
other times, based on our knowledge of X, we will have very little

The probability distribution of a random variable

$$\Pr(X = x) = f(X = x)$$

Sometimes we will have a lot of choice in specifying $f(\cdot)$;
other times, based on our knowledge of X, we will have very little

Usually, there are some *parameters* of this function which help determine $\Pr(X)$:

$$\Pr(X|\theta) = f(X, \theta)$$

Note that the probability of X is now conditional on our parameters θ

The probability distribution of a random variable

$$\Pr(X = x) = f(X = x)$$

Sometimes we will have a lot of choice in specifying $f(\cdot)$;
other times, based on our knowledge of X, we will have very little

Usually, there are some *parameters* of this function which help determine $\Pr(X)$:

$$\Pr(X|\theta) = f(X, \theta)$$

Note that the probability of X is now conditional on our parameters θ

For coin flips, the obvious parameter is the probability a single flipped coin comes out as heads, or π

$$\Pr(X|\pi) = f(X, \pi)$$

The Bernoulli distribution

The probability of a coin flip is described by a simple probability function known as the Bernoulli distribution:

$$f_{\text{Bern}}(X|\pi) = \begin{cases} 1-\pi & \text{if } X = 0 \\ \pi & \text{if } X = 1 \end{cases}$$

The Bernoulli distribution

The probability of a coin flip is described by a simple probability function known as the Bernoulli distribution:

$$f_{\text{Bern}}(X|\pi) = \begin{cases} 1-\pi & \text{if } X = 0 \\ \pi & \text{if } X = 1 \end{cases}$$

If we are clever, we can write it much more conveniently:

$$f_{\text{Bern}}(X|\pi) = \pi^X(1-\pi)^{1-X}$$

Another name for a probability distribution function is *pdf*,
so this is also called the Bernoulli pdf

The Bernoulli distribution

$\Pr(X) = f(X)$ How do we summarize the random distribution of X?

The Bernoulli distribution

$\Pr(X) = f(X)$ How do we summarize the random distribution of X?
The pdf above.

The Bernoulli distribution

$\Pr(X) = f(X)$ How do we summarize the random distribution of X?
The pdf above.

$\Pr(X = x)$ What is the probability X is some specific value like $x = 1$?

The Bernoulli distribution

$\Pr(X) = f(X)$ How do we summarize the random distribution of X?
The pdf above.

$\Pr(X = x)$ What is the probability X is some specific value like $x = 1$?
Plug it in to the pdf.

The Bernoulli distribution

$\Pr(X) = f(X)$ How do we summarize the random distribution of X?
The pdf above.

$\Pr(X = x)$ What is the probability X is some specific value like $x = 1$?
Plug it in to the pdf.

$\Pr(X \leq x)$ What is the probability that X is *at least* equal to some specific value, like $x = 1$? Sum up the pdf over all $X \leq x$.

The Bernoulli distribution

$\Pr(X) = f(X)$ How do we summarize the random distribution of X?
The pdf above.

$\Pr(X = x)$ What is the probability X is some specific value like $x = 1$?
Plug it in to the pdf.

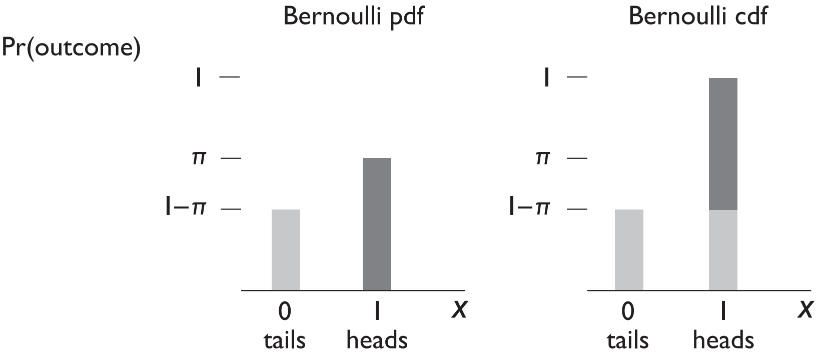
$\Pr(X \leq x)$ What is the probability that X is *at least* equal to some specific value, like $x = 1$? Sum up the pdf over all $X \leq x$.

The sum of a pdf over a range of values is the *cumulative density*

We can write a cumulative density function (cdf) for any discrete distribution as $F(X)$:

$$F(X = x) = \sum_{\forall X \leq x} f(x)$$

The Bernoulli distribution



For the Bernoulli, the pdf and cdf are simple functions

Note that both have *support* on the values $0, 1$ only –
the probability of X is defined as 0 for any other X beside 0 and 1.

The Bernoulli distribution

A distribution has an expected value and variance, just like a sample does, though they are now *theoretical*, pre-existing any data:

$$\mathrm{E}(X) \quad = \quad \sum_{\forall i} X_i f_{\mathrm{Bern}}(X_i|\pi)$$

The Bernoulli distribution

A distribution has an expected value and variance, just like a sample does, though they are now *theoretical*, pre-existing any data:

$$\begin{aligned} \mathrm{E}(X) &= \sum_{\forall i} X_i f_{\text{Bern}}(X_i|\pi) \\ &= 0 \times f_{\text{Bern}}(0|\pi) + 1 \times f_{\text{Bern}}(1|\pi) \end{aligned}$$

The Bernoulli distribution

A distribution has an expected value and variance, just like a sample does, though they are now *theoretical*, pre-existing any data:

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{\forall i} X_i f_{\mathrm{Bern}}(X_i|\pi) \\
&= 0 \times f_{\mathrm{Bern}}(0|\pi) + 1 \times f_{\mathrm{Bern}}(1|\pi) \\
&= 0 + \pi = \pi
\end{aligned}$$

The Bernoulli distribution

A distribution has an expected value and variance, just like a sample does, though they are now *theoretical*, pre-existing any data:

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{\forall i} X_i f_{\mathrm{Bern}}(X_i|\pi) \\
&= 0 \times f_{\mathrm{Bern}}(0|\pi) + 1 \times f_{\mathrm{Bern}}(1|\pi) \\
&= 0 + \pi = \pi
\end{aligned}$$

$$\mathrm{Var}(X) = \mathrm{E}\left[(X - \mathrm{E}(X))^2\right]$$

The Bernoulli distribution

A distribution has an expected value and variance, just like a sample does, though they are now *theoretical*, pre-existing any data:

$$\begin{aligned} \mathrm{E}(X) &= \sum_{\forall i} X_i f_{\mathrm{Bern}}(X_i|\pi) \\ &= 0 \times f_{\mathrm{Bern}}(0|\pi) + 1 \times f_{\mathrm{Bern}}(1|\pi) \\ &= 0 + \pi = \pi \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{E}\left[(X - \mathrm{E}(X))^2\right] \\ &= \mathrm{E}\left[(X - \pi)^2\right] \end{aligned}$$

The Bernoulli distribution

A distribution has an expected value and variance, just like a sample does, though they are now *theoretical*, pre-existing any data:

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{\forall i} X_i f_{\text{Bern}}(X_i|\pi) \\
&= 0 \times f_{\text{Bern}}(0|\pi) + 1 \times f_{\text{Bern}}(1|\pi) \\
&= 0 + \pi = \pi
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}\left[(X - \mathrm{E}(X))^2\right] \\
&= \mathrm{E}\left[(X - \pi)^2\right] \\
&= \sum_{\forall i} (X_i - \pi)^2 f_{\text{Bern}}(X_i|\pi)
\end{aligned}$$

The Bernoulli distribution

A distribution has an expected value and variance, just like a sample does, though they are now *theoretical*, pre-existing any data:

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{\forall i} X_i f_{\mathrm{Bern}}(X_i|\pi) \\
&= 0 \times f_{\mathrm{Bern}}(0|\pi) + 1 \times f_{\mathrm{Bern}}(1|\pi) \\
&= 0 + \pi = \pi
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}\left[(X - \mathrm{E}(X))^2\right] \\
&= \mathrm{E}\left[(X - \pi)^2\right] \\
&= \sum_{\forall i} (X_i - \pi)^2 f_{\mathrm{Bern}}(X_i|\pi) \\
&= (0 - \pi)^2 \times f_{\mathrm{Bern}}(0|\pi) + (1 - \pi)^2 \times f_{\mathrm{Bern}}(1|\pi)
\end{aligned}$$

The Bernoulli distribution

A distribution has an expected value and variance, just like a sample does, though they are now *theoretical*, pre-existing any data:

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{\forall i} X_i f_{\mathrm{Bern}}(X_i|\pi) \\
&= 0 \times f_{\mathrm{Bern}}(0|\pi) + 1 \times f_{\mathrm{Bern}}(1|\pi) \\
&= 0 + \pi = \pi
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}\left[(X - \mathrm{E}(X))^2\right] \\
&= \mathrm{E}\left[(X - \pi)^2\right] \\
&= \sum_{\forall i} (X_i - \pi)^2 f_{\mathrm{Bern}}(X_i|\pi) \\
&= (0 - \pi)^2 \times f_{\mathrm{Bern}}(0|\pi) + (1 - \pi)^2 \times f_{\mathrm{Bern}}(1|\pi) \\
&= \pi^2(1 - \pi) + (1 - \pi)^2\pi
\end{aligned}$$

The Bernoulli distribution

A distribution has an expected value and variance, just like a sample does, though they are now *theoretical*, pre-existing any data:

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{\forall i} X_i f_{\mathrm{Bern}}(X_i|\pi) \\
&= 0 \times f_{\mathrm{Bern}}(0|\pi) + 1 \times f_{\mathrm{Bern}}(1|\pi) \\
&= 0 + \pi = \pi
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}\left[(X - \mathrm{E}(X))^2\right] \\
&= \mathrm{E}\left[(X - \pi)^2\right] \\
&= \sum_{\forall i} (X_i - \pi)^2 f_{\mathrm{Bern}}(X_i|\pi) \\
&= (0 - \pi)^2 \times f_{\mathrm{Bern}}(0|\pi) + (1 - \pi)^2 \times f_{\mathrm{Bern}}(1|\pi) \\
&= \pi^2(1 - \pi) + (1 - \pi)^2\pi \\
&= \pi(1 - \pi)
\end{aligned}$$

The Bernoulli distribution

$\mathrm{E}(X)$ What is the expected number of heads we will see on average?

$\mathrm{E}(X) = \pi$

The Bernoulli distribution

$\mathrm{E}(X)$ What is the expected number of heads we will see on average?
$\mathrm{E}(X) = \pi$

$\mathrm{sd}(X)$ On average, how much do we expect a given random outcome to differ from the expected result?

We know $\mathrm{Var}(X) = \pi(1 - \pi)$.

The Bernoulli distribution

$\mathrm{E}(X)$ What is the expected number of heads we will see on average?
$\mathrm{E}(X) = \pi$

$\mathrm{sd}(X)$ On average, how much do we expect a given random outcome to differ from the expected result?

We know $\mathrm{Var}(X) = \pi(1 - \pi)$.

So $\mathrm{sd}(X) = \sqrt{\pi(1 - \pi)}$.

$\mathrm{E}(X)$ What is the expected number of heads we will see on average?
$\mathrm{E}(X) = \pi$

$\mathrm{sd}(X)$ On average, how much do we expect a given random outcome to differ from the expected result?

We know $\mathrm{Var}(X) = \pi(1 - \pi)$.

So $\mathrm{sd}(X) = \sqrt{\pi(1 - \pi)}$.

Note that this is maximized at $\pi = 0.5$. Why?

The binomial distribution

Now let's go back to our original question:
what is the distribution of the sum of M coin flips?

Real world applications are many:

- rainy days in a month
- students our of a class who pass
- votes from a known population of voters (elections, Congressional votes)

The binomial distribution

Now let's go back to our original question:
what is the distribution of the sum of M coin flips?

Real world applications are many:

- rainy days in a month
- students our of a class who pass
- votes from a known population of voters (elections, Congressional votes)

Suppose we observe several Bernoulli RVs and count the successes
(might imagine that the underlying 1s and 0s are lost)

The binomial distribution

Now let's go back to our original question:
what is the distribution of the sum of M coin flips?

Real world applications are many:

- rainy days in a month
- students our of a class who pass
- votes from a known population of voters (elections, Congressional votes)

Suppose we observe several Bernoulli RVs and count the successes
(might imagine that the underlying 1s and 0s are lost)

Key assumption:
each trial is identically and independently distributed (iid) Bernoulli
For the moment, take this to mean that each trial has the same π of success

The binomial distribution

How do we come up with a pdf for these assumptions?
Back to sample spaces for a moment

Suppose we wanted to calculate the probability that e get two heads out of a total of three flips: $\Pr(X = 2|\pi, M = 3)$

There are three events in the sample space that meet these criteria:

$$\Pr(X = 2|\pi, M = 3) = \Pr(H, H, T)$$

The binomial distribution

How do we come up with a pdf for these assumptions?
Back to sample spaces for a moment

Suppose we wanted to calculate the probability that e get two heads out of a total of three flips: $\Pr(X = 2|\pi, M = 3)$

There are three events in the sample space that meet these criteria:

$$\Pr(X = 2|\pi, M = 3) = \Pr(H, H, T) + \Pr(H, T, H)$$

The binomial distribution

How do we come up with a pdf for these assumptions?
Back to sample spaces for a moment

Suppose we wanted to calculate the probability that e get two heads out of a total of three flips: $\Pr(X = 2|\pi, M = 3)$

There are three events in the sample space that meet these criteria:

$$\Pr(X = 2|\pi, M = 3) \quad = \quad \Pr(H, H, T) + \Pr(H, T, H) + \Pr(T, H, H)$$

The binomial distribution

How do we come up with a pdf for these assumptions?
Back to sample spaces for a moment

Suppose we wanted to calculate the probability that e get two heads out of a total of three flips: $\Pr(X = 2|\pi, M = 3)$

There are three events in the sample space that meet these criteria:

$$\begin{aligned}\Pr(X = 2|\pi, M = 3) &= \Pr(H, H, T) + \Pr(H, T, H) + \Pr(T, H, H) \\ &= \pi\pi(1 - \pi)\end{aligned}$$

The binomial distribution

How do we come up with a pdf for these assumptions?
Back to sample spaces for a moment

Suppose we wanted to calculate the probability that e get two heads out of a total of three flips: $\Pr(X = 2|\pi, M = 3)$

There are three events in the sample space that meet these criteria:

$$\begin{aligned} \Pr(X = 2|\pi, M = 3) &= \Pr(H, H, T) + \Pr(H, T, H) + \Pr(T, H, H) \\ &= \pi\pi(1 - \pi) + \pi(1 - \pi)\pi \end{aligned}$$

The binomial distribution

How do we come up with a pdf for these assumptions?
Back to sample spaces for a moment

Suppose we wanted to calculate the probability that e get two heads out of a total of three flips: $\Pr(X = 2|\pi, M = 3)$

There are three events in the sample space that meet these criteria:

$$\begin{aligned} \Pr(X = 2|\pi, M = 3) &= \Pr(H, H, T) + \Pr(H, T, H) + \Pr(T, H, H) \\ &= \pi\pi(1 - \pi) + \pi(1 - \pi)\pi + (1 - \pi)\pi\pi \end{aligned}$$

The binomial distribution

How do we come up with a pdf for these assumptions?
Back to sample spaces for a moment

Suppose we wanted to calculate the probability that e get two heads out of a total of three flips: $\Pr(X = 2|\pi, M = 3)$

There are three events in the sample space that meet these criteria:

$$\begin{aligned}
\Pr(X = 2|\pi, M = 3) &= \Pr(H, H, T) + \Pr(H, T, H) + \Pr(T, H, H) \\
&= \pi\pi(1-\pi) + \pi(1-\pi)\pi + (1-\pi)\pi\pi \\
&= 3\pi\pi(1-\pi)
\end{aligned}$$

The binomial distribution

How do we come up with a pdf for these assumptions?
Back to sample spaces for a moment

Suppose we wanted to calculate the probability that e get two heads out of a total of three flips: $\Pr(X = 2|\pi, M = 3)$

There are three events in the sample space that meet these criteria:

$$\begin{aligned}\Pr(X = 2|\pi, M = 3) &= \Pr(H, H, T) + \Pr(H, T, H) + \Pr(T, H, H) \\ &= \pi\pi(1 - \pi) + \pi(1 - \pi)\pi + (1 - \pi)\pi\pi \\ &= 3\pi\pi(1 - \pi) \\ &= 3\pi^2(1 - \pi)^{3-2}\end{aligned}$$

The binomial distribution

How do we come up with a pdf for these assumptions?
Back to sample spaces for a moment

Suppose we wanted to calculate the probability that e get two heads out of a total of three flips: $\Pr(X = 2|\pi, M = 3)$

There are three events in the sample space that meet these criteria:

$$\begin{aligned}
\Pr(X = 2|\pi, M = 3) &= \Pr(H, H, T) + \Pr(H, T, H) + \Pr(T, H, H) \\
&= \pi\pi(1-\pi) + \pi(1-\pi)\pi + (1-\pi)\pi\pi \\
&= 3\pi\pi(1-\pi) \\
&= 3\pi^2(1-\pi)^{3-2} \\
\Pr(X|\pi, M) &= \binom{M}{X}\pi^X(1-\pi)^{M-X}
\end{aligned}$$

The binomial distribution

How do we come up with a pdf for these assumptions?
Back to sample spaces for a moment

Suppose we wanted to calculate the probability that e get two heads out of a total of three flips: $\Pr(X = 2|\pi, M = 3)$

There are three events in the sample space that meet these criteria:

$$\begin{aligned}
\Pr(X = 2|\pi, M = 3) &= \Pr(H, H, T) + \Pr(H, T, H) + \Pr(T, H, H) \\
&= \pi\pi(1-\pi) + \pi(1-\pi)\pi + (1-\pi)\pi\pi \\
&= 3\pi\pi(1-\pi) \\
&= 3\pi^2(1-\pi)^{3-2} \\
\Pr(X|\pi, M) &= \binom{M}{X}\pi^X(1-\pi)^{M-X} \\
&= \frac{M!}{X!(M-X)!}\pi^X(1-\pi)^{M-X}
\end{aligned}$$

The binomial distribution

$$f_{\text{Bin}}(X|\pi, M) = \frac{M!}{X!(M-X)!}\pi^{X}(1-\pi)^{M-X}$$

Similarity to the Bernoulli evident, especially in the moments:

The binomial distribution

$$f_{\text{Bin}}(X|\pi, M) = \frac{M!}{X!(M-X)!}\pi^X(1-\pi)^{M-X}$$

Similarity to the Bernoulli evident, especially in the moments:

$\text{E}(x) = M\pi$

The binomial distribution

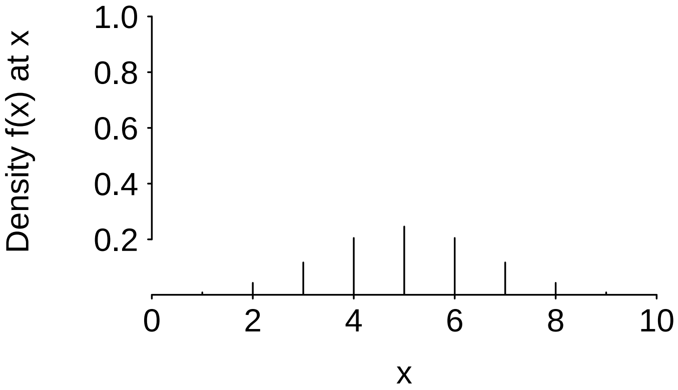
$$f_{\text{Bin}}(X|\pi, M) = \frac{M!}{X!(M-X)!}\pi^{X}(1-\pi)^{M-X}$$

Similarity to the Bernoulli evident, especially in the moments:

$$\mathrm{E}(x) = M\pi \qquad \mathrm{Var}(x) = M\pi(1-\pi)$$

The binomial distribution: PDF

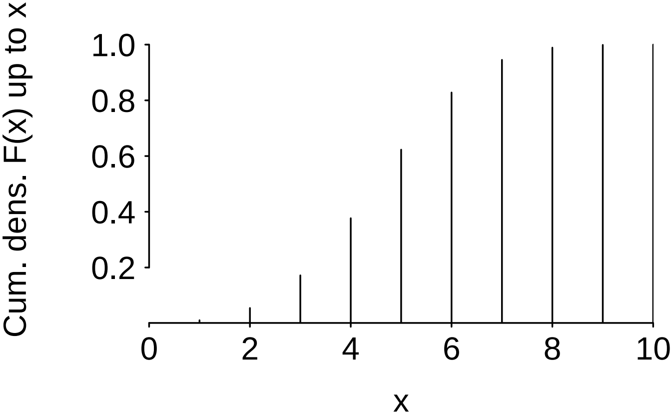
This graph shows the probability of each outcome X



The pdf of binomial sums over 10 trials,
with each trial having an 0.5 probability of success

The binomial distribution: CDF

This graph shows the probability of seeing an outcome *less than or equal to* X



The cdf of binomial sums over 10 trials,
with each trial having an 0.5 probability of success

Example: Rosencrantz and Guildenstern are still dead

Rosencrantz and Guildenstern are minor characters in Shakespeare's *Hamlet*

Claudius has killed his brother, the King, and married his sister-in-law, the Queen, to assume the throne

He dispossesses Prince Hamlet, without ever explicitly acknowledging the fact

Hamlet becomes depressed and sullen. Expecting Hamlet to eventually challenge him, Claudius summons two of Hamlet's childhood friends, R & G, to spy on the prince

Claudius tricks R & G into delivering Hamlet to England for execution. Hamlet tricks the (none-too-bright) R & G into being executed in his place.

Tom Stoppard's play *Rosencrantz and Guildenstern are Dead* retells this story through R & G's perspective, emphasizing their powerlessness and lack of real choice

Example: Rosencrantz and Guildenstern are still dead

At the start of Stoppard's play, Rosenstern and Guildencrantz have flipped a coin 89 times, getting heads every time

How unlikely is this event?

Example: Rosencrantz and Guildenstern are still dead

At the start of Stoppard's play, Rosenstern and Guildencrantz have flipped a coin 89 times, getting heads every time

How unlikely is this event?

Apply the binomial pdf:

$$\Pr(X|\pi, M) \quad = \quad f_{\text{Bin}}(X|\pi, M)$$

Example: Rosencrantz and Guildenstern are still dead

At the start of Stoppard's play, Rosenstern and Guildencrantz have flipped a coin 89 times, getting heads every time

How unlikely is this event?

Apply the binomial pdf:

$$\begin{aligned} \Pr(X|\pi, M) &= f_{\text{Bin}}(X|\pi, M) \\ &= \frac{M!}{X!(M-X)!}\pi^{X}(1-\pi)^{M-X} \end{aligned}$$

Example: Rosencrantz and Guildenstern are still dead

At the start of Stoppard's play, Rosenstern and Guildencrantz have flipped a coin 89 times, getting heads every time

How unlikely is this event?

Apply the binomial pdf:

$$\begin{aligned}
\Pr(X|\pi, M) &= f_{\text{Bin}}(X|\pi, M) \\
&= \frac{M!}{X!(M-X)!}\pi^{X}(1-\pi)^{M-X} \\
\Pr(X=89|\pi=0.5, M=89) &= \frac{89!}{89!(89-89)!}0.5^{89}(1-0.5)^{89-89}
\end{aligned}$$

Example: Rosencrantz and Guildenstern are still dead

At the start of Stoppard's play, Rosenstern and Guildencrantz have flipped a coin 89 times, getting heads every time

How unlikely is this event?

Apply the binomial pdf:

$$\begin{aligned} \Pr(X|\pi, M) &= f_{\text{Bin}}(X|\pi, M) \\ &= \frac{M!}{X!(M-X)!}\pi^X(1-\pi)^{M-X} \\ \Pr(X=89|\pi=0.5, M=89) &= \frac{89!}{89!(89-89)!}0.5^{89}(1-0.5)^{89-89} \end{aligned}$$

Working with large factorials might "break" your calculator.

$89!$ has 136 digits before the decimal, for example

Example: Rosencrantz and Guildenstern are still dead

$$\Pr(X|\pi, M) \quad = \quad f_{\text{Bin}}(X|\pi, M)$$

Example: Rosencrantz and Guildenstern are still dead

$$
\begin{aligned}
\Pr(X|\pi, M) &= f_{\text{Bin}}(X|\pi, M) \\
&= \frac{M!}{X!(M-X)!}\pi^X(1-\pi)^{M-X}
\end{aligned}
$$

Example: Rosencrantz and Guildenstern are still dead

$$
\begin{aligned}
\Pr(X|\pi, M) &= f_{\text{Bin}}(X|\pi, M) \\
&= \frac{M!}{X!(M-X)!}\pi^{X}(1-\pi)^{M-X} \\
\Pr(X=89|\pi=0.5, M=89) &= \frac{89!}{89!(89-89)!}0.5^{89}(1-0.5)^{89-89}
\end{aligned}
$$

In `R`, the above formula is implemented using the `dbinom()` command.
(d stands for density, as in pdf)

Example: Rosencrantz and Guildenstern are still dead

$$\begin{aligned}\Pr(X|\pi, M) &= f_{\text{Bin}}(X|\pi, M) \\ &= \frac{M!}{X!(M-X)!}\pi^X(1-\pi)^{M-X} \\ \Pr(X=89|\pi=0.5, M=89) &= \frac{89!}{89!(89-89)!}0.5^{89}(1-0.5)^{89-89}\end{aligned}$$

In `R`, the above formula is implemented using the `dbinom()` command.
(d stands for density, as in pdf)

`dbinom(89, size=89, prob=0.5)` gives us:

Example: Rosencrantz and Guildenstern are still dead

$$\begin{aligned}\Pr(X|\pi, M) &= f_{\text{Bin}}(X|\pi, M) \\ &= \frac{M!}{X!(M-X)!}\pi^X(1-\pi)^{M-X} \\ \Pr(X=89|\pi=0.5, M=89) &= \frac{89!}{89!(89-89)!}0.5^{89}(1-0.5)^{89-89}\end{aligned}$$

In `R`, the above formula is implemented using the `dbinom()` command.
(d stands for density, as in pdf)

`dbinom(89, size=89, prob=0.5)` gives us:

$$\Pr(89 \text{ heads out of } 89|\text{fair coin}) = 0.0000000000000000000000000162$$

Example: Rosencrantz and Guildenstern are still dead

$$\begin{aligned}
\Pr(X|\pi, M) &= f_{\text{Bin}}(X|\pi, M) \\
&= \frac{M!}{X!(M-X)!}\pi^X(1-\pi)^{M-X} \\
\Pr(X=89|\pi=0.5, M=89) &= \frac{89!}{89!(89-89)!}0.5^{89}(1-0.5)^{89-89}
\end{aligned}$$

In `R`, the above formula is implemented using the `dbinom()` command.
(d stands for density, as in pdf)

`dbinom(89, size=89, prob=0.5)` gives us:

$$\Pr(89 \text{ heads out of } 89|\text{fair coin}) = 0.00000000000000000000000000162$$

Or a 1 in 618,000,000,000,000,000,000,000,000 chance

Example: Rosencrantz and Guildenstern are still dead

$$
\begin{aligned}
\Pr(X|\pi, M) &= f_{\text{Bin}}(X|\pi, M) \\
&= \frac{M!}{X!(M-X)!}\pi^X(1-\pi)^{M-X} \\
\Pr(X=89|\pi=0.5, M=89) &= \frac{89!}{89!(89-89)!}0.5^{89}(1-0.5)^{89-89}
\end{aligned}
$$

In `R`, the above formula is implemented using the `dbinom()` command. (d stands for density, as in pdf)

`dbinom(89, size=89, prob=0.5)` gives us:

$\Pr(89 \text{ heads out of } 89|\text{fair coin}) = 0.00000000000000000000000000162$

Or a 1 in 618,000,000,000,000,000,000,000,000 chance

Or one chance in 618 million trillion trillion tries

Example: Rosencrantz and Guildenstern are still dead

How small is one chance in 618 million trillion trillion tries?

Example: Rosencrantz and Guildenstern are still dead

How small is one chance in 618 million trillion trillion tries?

For context, the universe began 432 thousand trillion seconds ago

Example: Rosencrantz and Guildenstern are still dead

How small is one chance in 618 million trillion trillion tries?

For context, the universe began 432 thousand trillion seconds ago

If a billion people each flippled a set of 89 coins once a second
– repeating this action *every* second since time began –
we would expect approximately *one* person to get a single set of all heads.

Example: Rosencrantz and Guildenstern are still dead

If you see such a freak occurrence in life you must ask yourself:

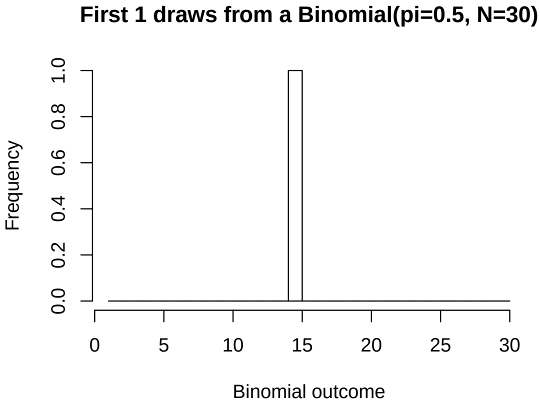
"Is it more likely that I saw 89 fair coins come up heads by chance, *or* that someone has changed the probability of a heads to something close to 1?"

Clearly, something is manipulating the coins,
just like Claudius & Hamlet are manipulating R & G

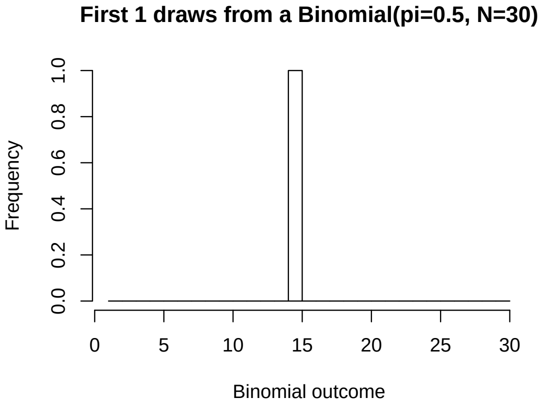
One way to use probability distributions:
How likely is an event to have been "mere chance"?

Random samples from the binomial

Another way to get a handle on pdfs and cdfs
is to just draw repeatedly from the distribution of interest

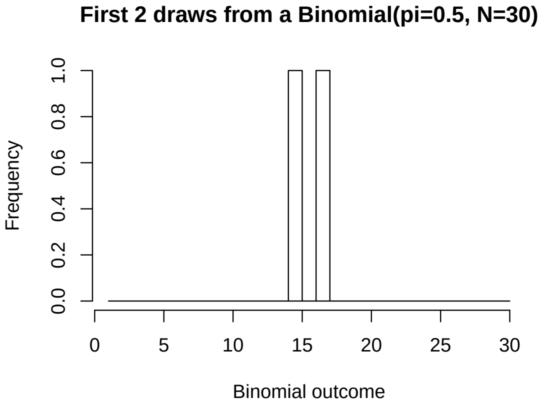


Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram



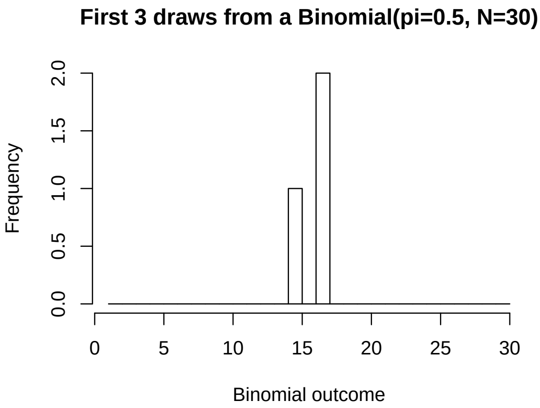
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



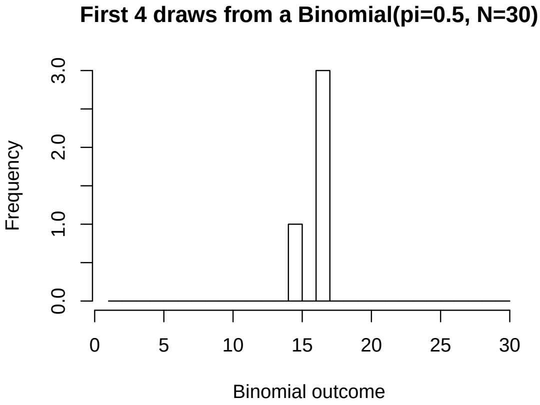
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



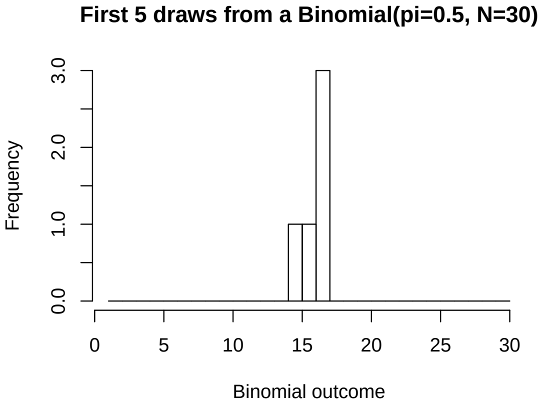
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



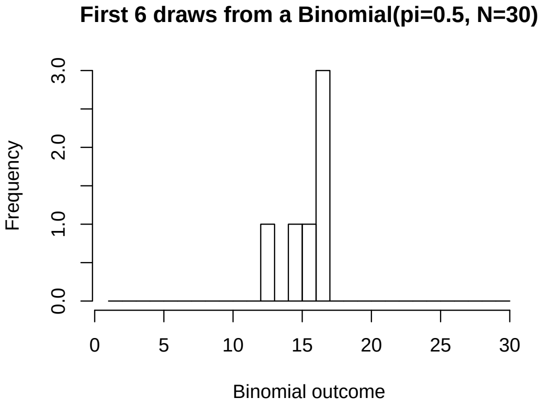
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



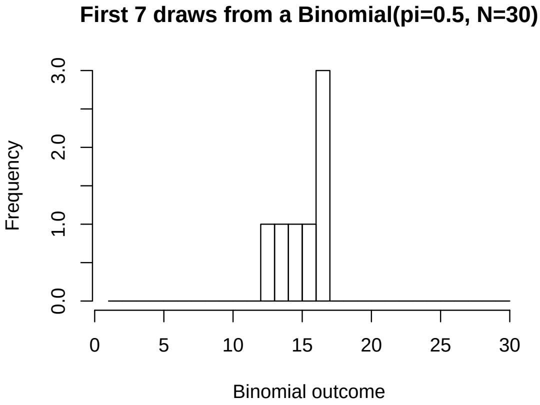
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



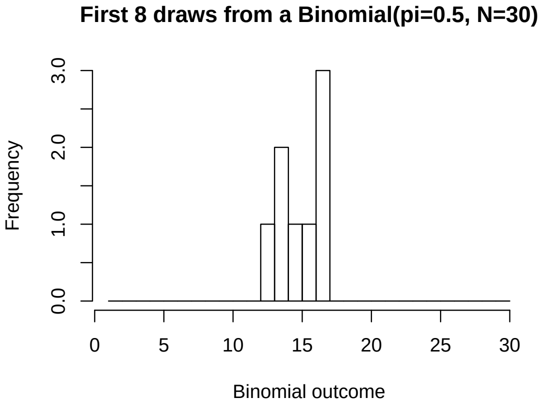
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



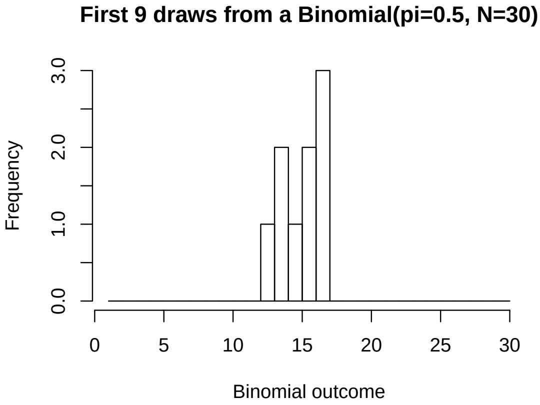
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



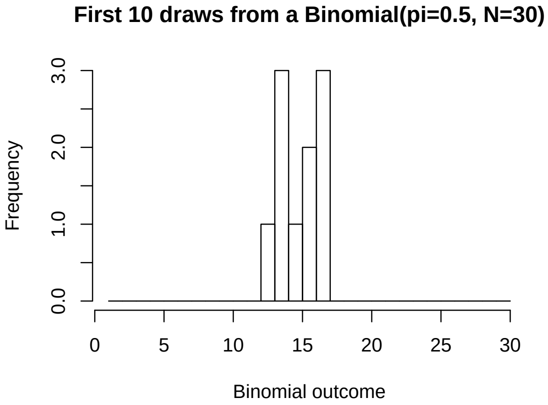
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



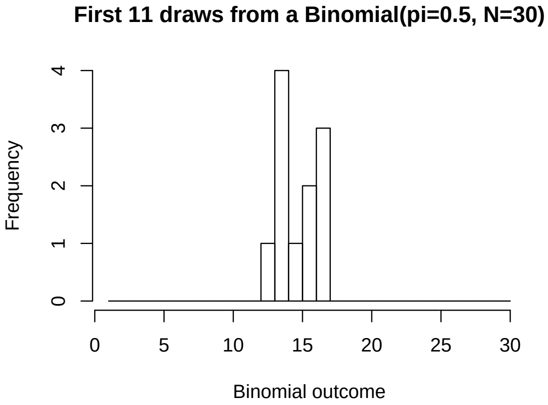
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



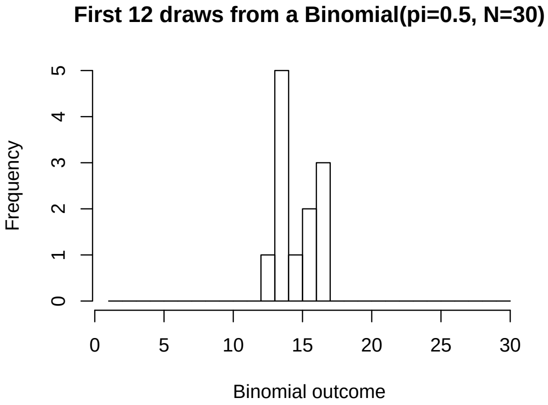
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



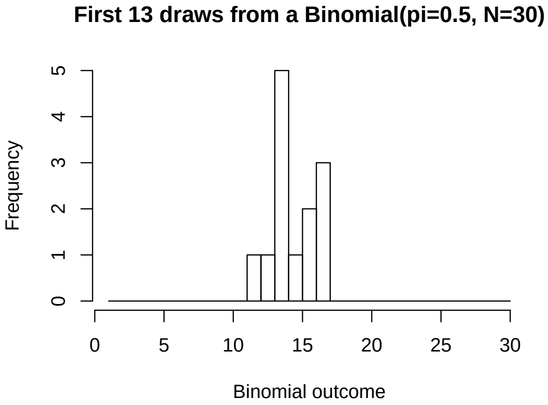
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



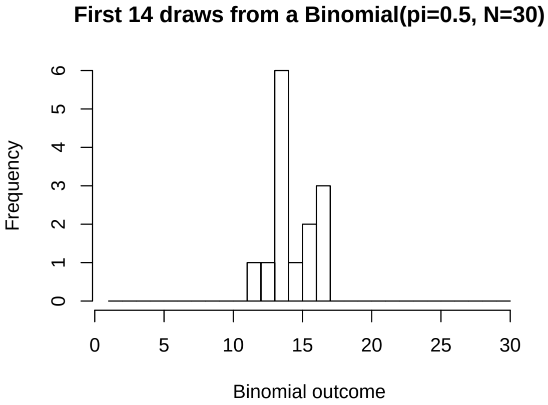
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



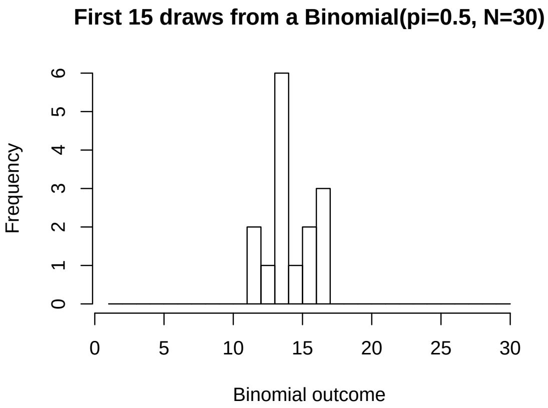
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



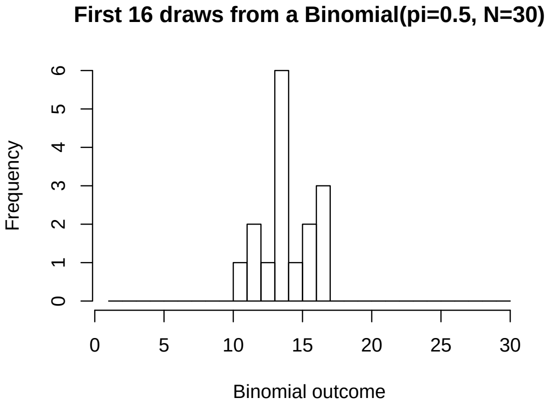
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



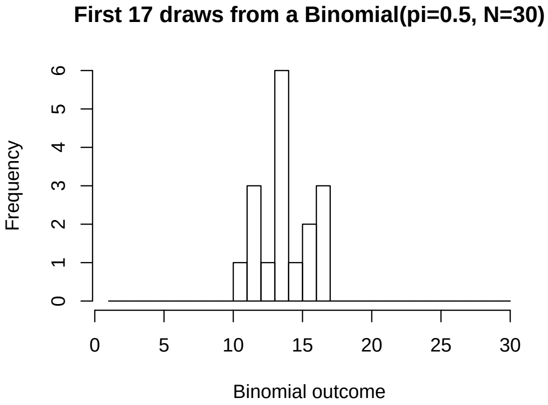
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



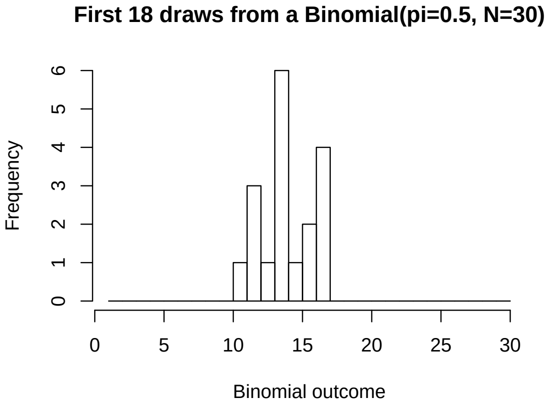
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



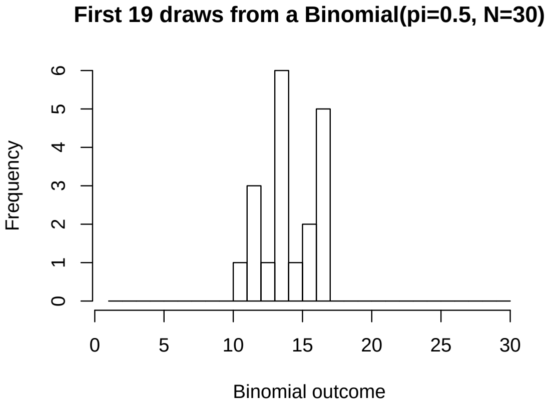
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



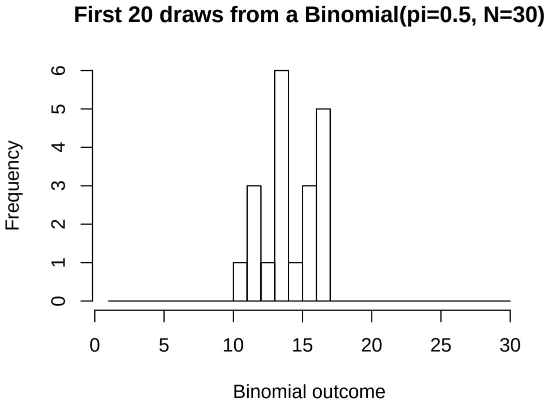
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



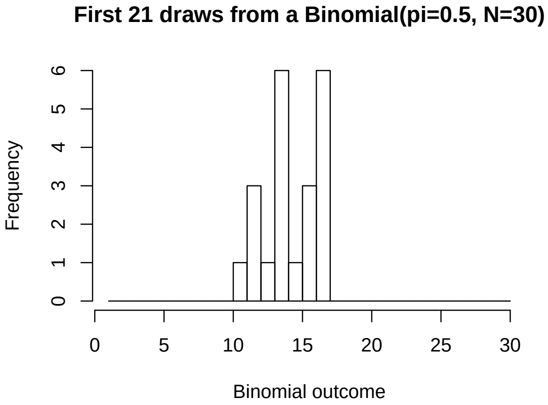
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



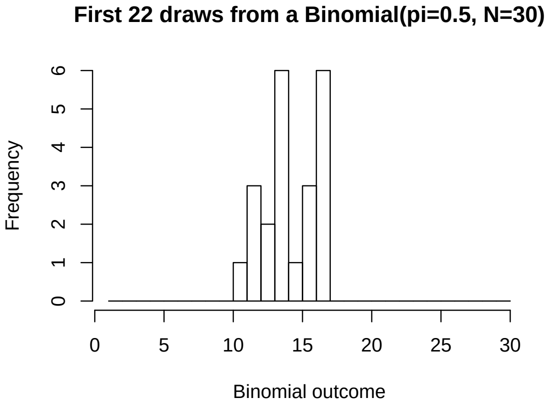
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



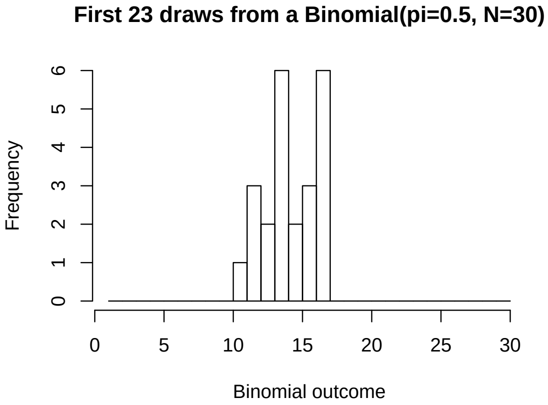
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



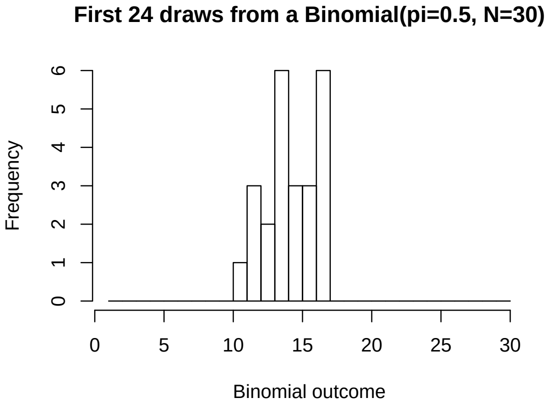
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



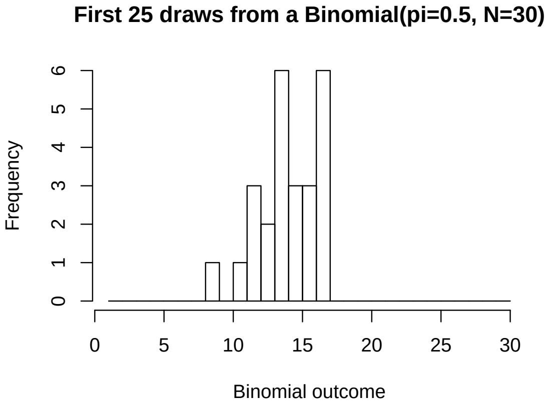
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



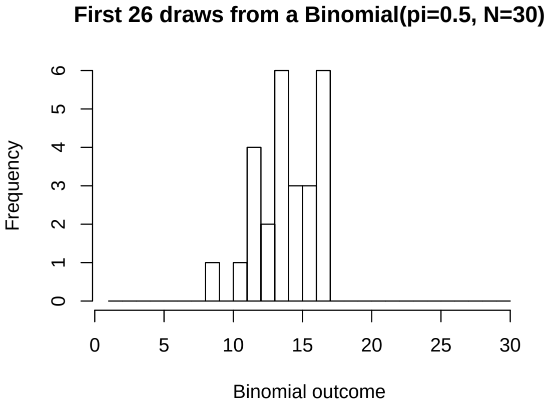
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



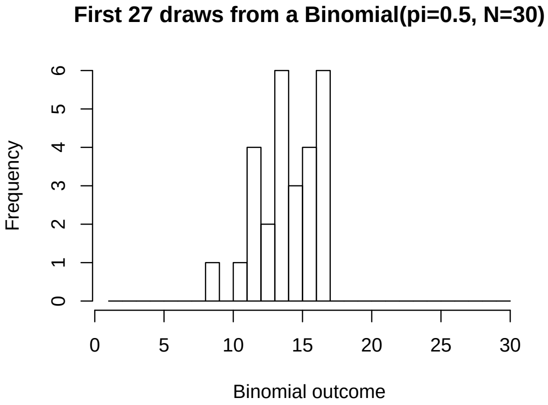
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



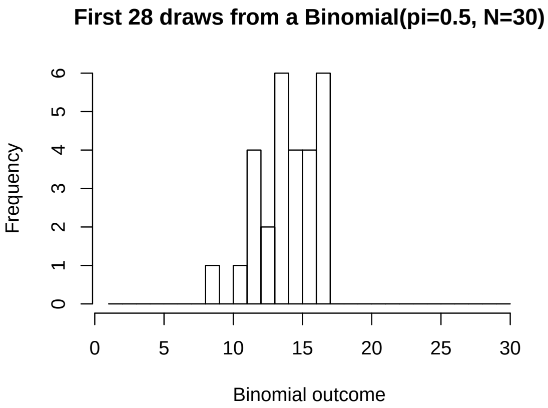
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



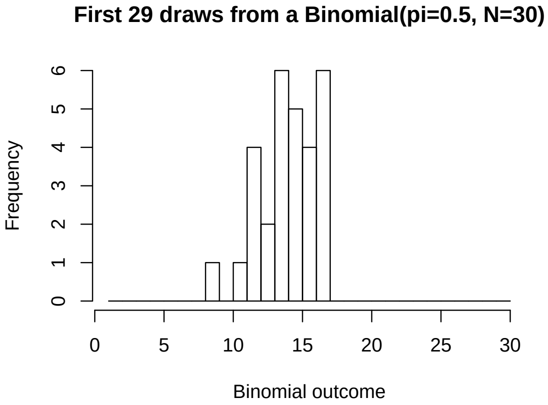
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



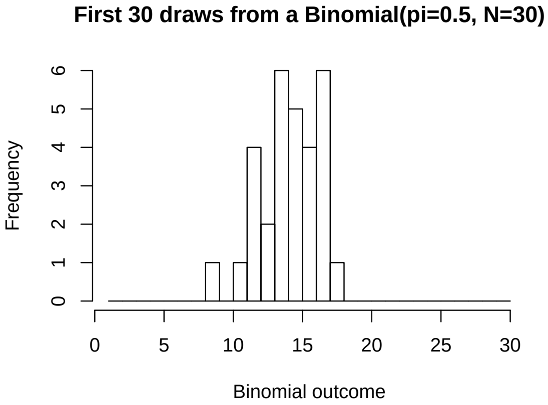
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



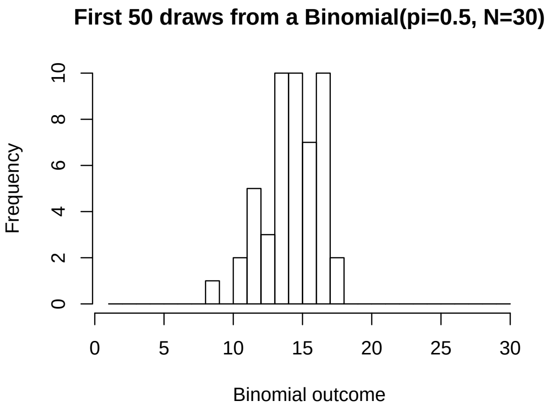
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



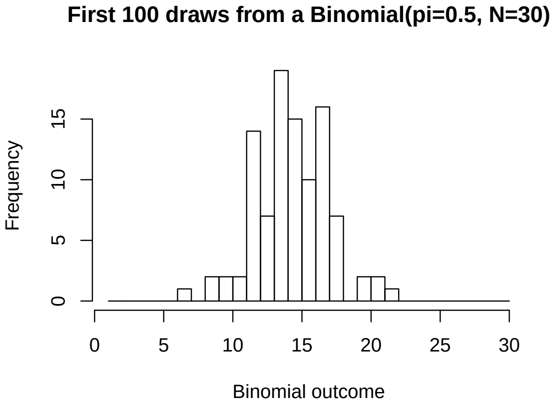
Let's draw repeatedly from the Binomial with 30 trials, each with probability 0.5 of success, and save our results in a histogram

Watch the scale of the vertical axis (frequency) closely



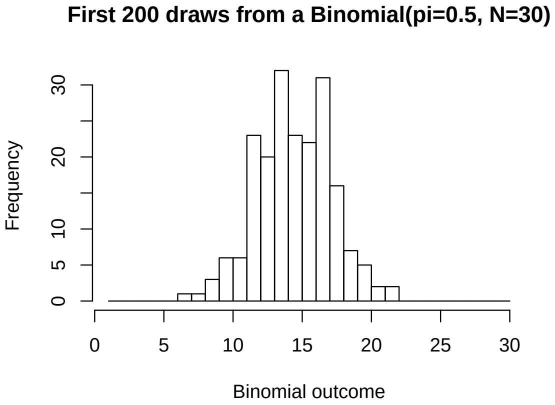
Now that we have the basic idea, let's add more than one draw to the histogram at a time

Now we're up to 50 draws



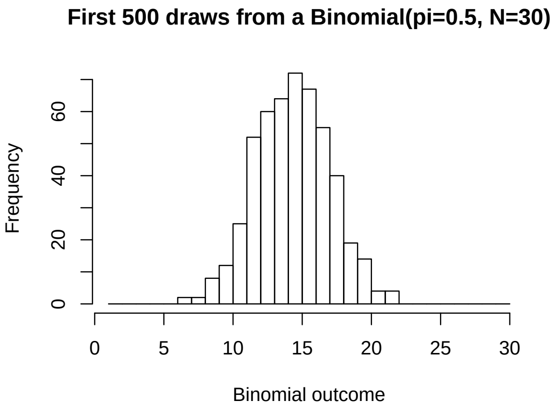
Now that we have the basic idea, let's add more than one draw to the histogram at a time

Now we're up to 100 draws



Now that we have the basic idea, let's add more than one draw to the histogram at a time

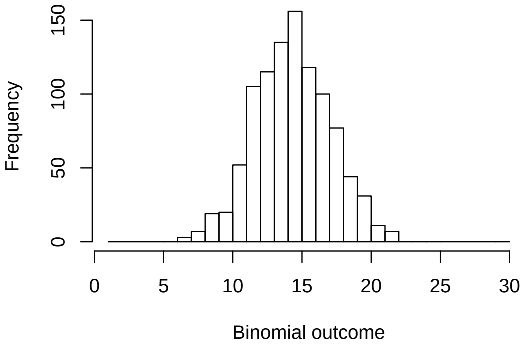
Now we're up to 200 draws



Now that we have the basic idea, let's add more than one draw to the histogram at a time

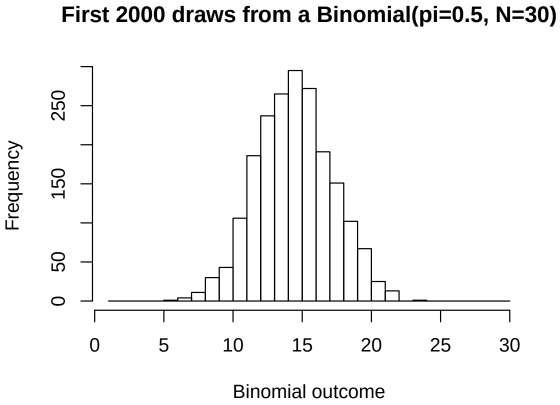
Now we're up to 500 draws

First 1000 draws from a Binomial(pi=0.5, N=30)



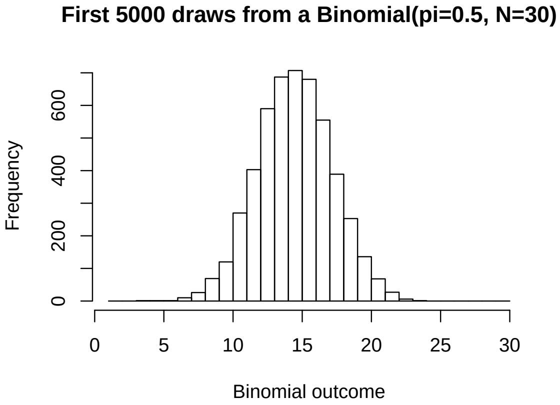
Now that we have the basic idea, let's add more than one draw to the histogram at a time

Now we're up to 1,000 draws



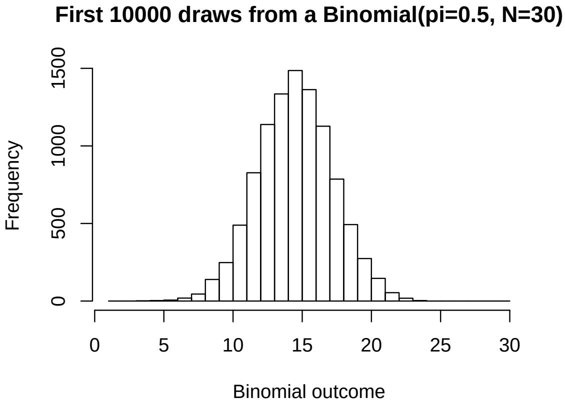
Now that we have the basic idea, let's add more than one draw to the histogram at a time

Now we're up to 2,000 draws



Now that we have the basic idea, let's add more than one draw to the histogram at a time

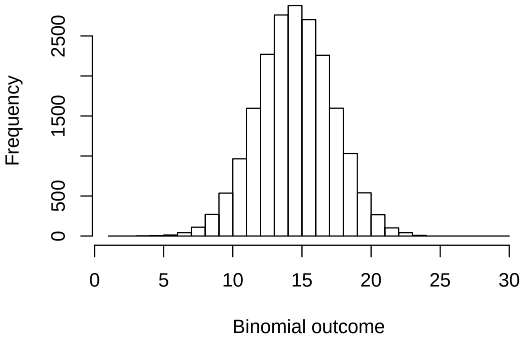
Now we're up to 5,000 draws



Now we're up to 10,000 draws

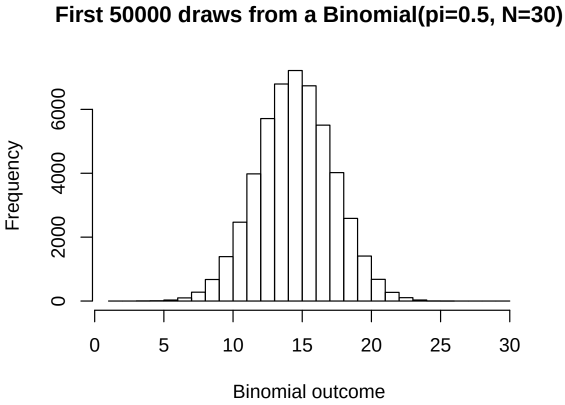
Does this distribution remind you of any other distribution you've read about?

First 20000 draws from a Binomial(pi=0.5, N=30)



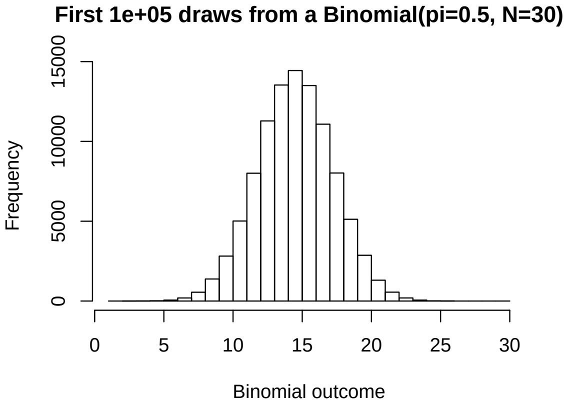
Now we're up to 20,000 draws

As the number of draws from the binomial gets large, it starts to look like a Normal, or bell-shaped distribution



Now we're up to 50,000 draws

Won't ever be exactly normal: it's still discrete (you can't get 10.5, for example), and it only has support on [0,30]



Now we're up to 100,000 draws

(Does this remind you of anything?)

Example: Anticipating a landslide

Every ten years, the states redraw the lines of Congressional districts

Example: Anticipating a landslide

Every ten years, the states redraw the lines of Congressional districts

Why we have a decennial census: how many districts should each state get, and how should they be drawn to have equal numbers of residents

Example: Anticipating a landslide

Every ten years, the states redraw the lines of Congressional districts

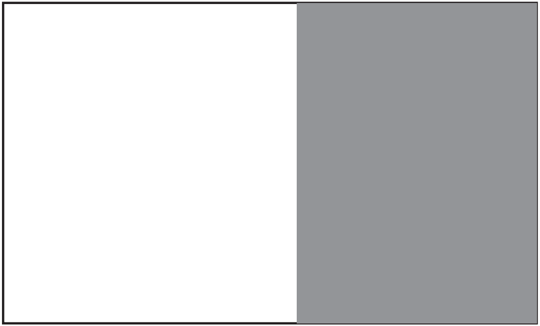
Why we have a decennial census: how many districts should each state get, and how should they be drawn to have equal numbers of residents

Political parties use redistricting to strategically draw district boundaries that achieve their goals

Key (conflicting!) goals: incumbent protection and maximizing total seats

Called political gerrymandering. Currently quite legal.

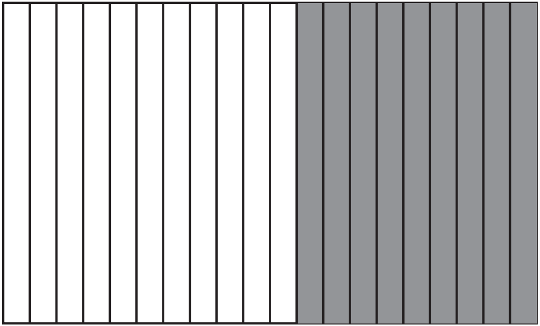
Example: Anticipating a landslide



Imagine a state of 10 million people where the population is 55% Democratic, and 45% Republican.

Suppose that every Democrat lives on the west side of the state, and every Republican on the east.

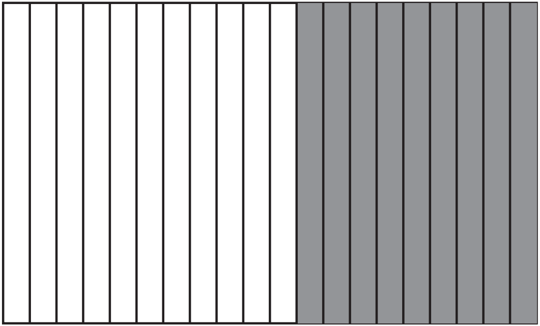
Example: Anticipating a landslide



Currently, there are 20 districts of 500,000 people each, draw in North-South strips.

Each district has homogenous D or R population

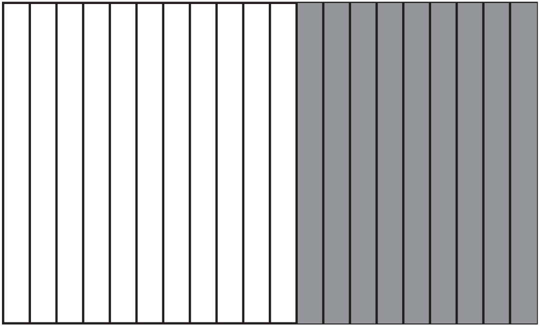
Example: Anticipating a landslide



Normally in an election, the side that can get the most voters to turnout wins (approximately half of the eligible voters actual go to the polls)

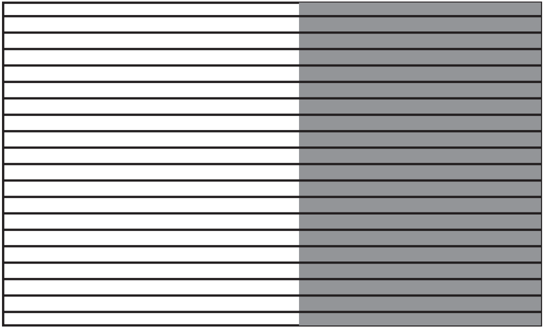
But if each district is homogenous, this doesn't matter: Every election, regardless of turnout, the D's win 11 seats, and R's win 9 seats

Example: Anticipating a landslide



The Democrats also control the state legislature, and have a clever idea.

Example: Anticipating a landslide

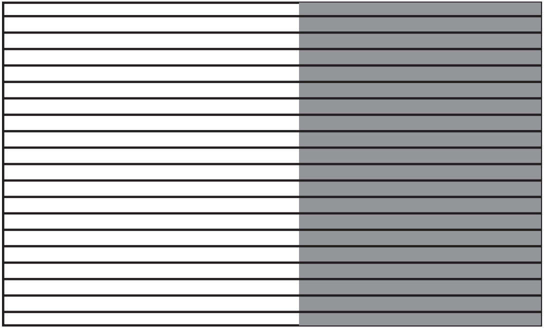


"Let's redraw the map this year as a series of East-West districts, each 55% Democratic.

We'll have a majority in every district, so we should expect to win 20 seats!"

Are these legislators correct? How many seats should they expect to win?

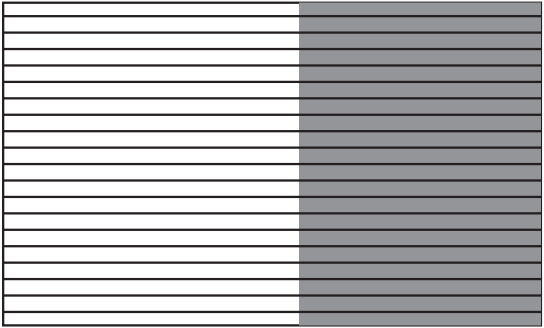
Example: Anticipating a landslide



The legislators have assumed the probability of victory is 100% in a district that is 55% Democratic and 45% Republican

Not a good assumption. Let's instead assume the probability of victory is a healthy 75%

Example: Anticipating a landslide



The Democrats will usually win, but 1 in 4 years, the Republicans will manage to get enough voters to the polls to overwhelm the D's registration advantage

Under the new map, what is the expected number of Democratic seats across a large number of elections?

Example: Anticipating a landslide

Under the new map, what is the expected number of Democratic seats across a large number of elections?

Example: Anticipating a landslide

Under the new map, what is the expected number of Democratic seats across a large number of elections?

$$\mathrm{E}(\text{seats}) = M\pi$$

Example: Anticipating a landslide

Under the new map, what is the expected number of Democratic seats across a large number of elections?

$$\begin{aligned} \mathrm{E}(\text{seats}) &= M\pi \\ &= 20 \times 0.75 \\ &= 15 \end{aligned}$$

The Demcrats should expect to add 4 seats, not 9 seats

Failure to understand binomial probability has lead to a massive overestimate of the effects of redistricting

A common mistake in redistricting and in news articles about it

Example: Anticipating a landslide

Gains from redistricting are not a sure thing.

How much should the Democrats expect their winning to vary on average from 15?

Example: Anticipating a landslide

Gains from redistricting are not a sure thing.

How much should the Democrats expect their winning to vary on average from 15?

$$\mathrm{sd}(x) = \sqrt{M\pi(1-\pi)}$$

Example: Anticipating a landslide

Gains from redistricting are not a sure thing.

How much should the Democrats expect their winning to vary on average from 15?

$$\begin{aligned}
\mathrm{sd}(x) &= \sqrt{M\pi(1-\pi)} \\
&= \sqrt{20 \times 0.75 \times 0.25} \\
&= \sqrt{3.75} \\
&= 1.94
\end{aligned}$$

How would you explain all this to a Democratic strategist?

Example: Anticipating a landslide

Gains from redistricting are not a sure thing.

How much should the Democrats expect their winning to vary on average from 15?

$$\begin{aligned} \text{sd}(x) &= \sqrt{M\pi(1-\pi)} \\ &= \sqrt{20 \times 0.75 \times 0.25} \\ &= \sqrt{3.75} \\ &= 1.94 \end{aligned}$$

How would you explain all this to a Democratic strategist?
If you redistrict for maximum gain, you may expect to win 15 seats out of 20 on average, but you'll typically overshoot or undershoot by 2 seats. So don't be surprised if you only pick up two new seats, for a total of 13!

Example: Anticipating a landslide

What is the probability the Democrats will actually get all 20 seats from their gerrymander?

Example: Anticipating a landslide

What is the probability the Democrats will actually get all 20 seats from their gerrymander? 11 seats exactly?

Example: Anticipating a landslide

What is the probability the Democrats will actually get all 20 seats from their gerrymander? 11 seats exactly? less than 11?

Example: Anticipating a landslide

What is the probability the Democrats will actually get all 20 seats from their gerrymander? 11 seats exactly? less than 11? more than 11?

Example: Anticipating a landslide

What is the probability the Democrats will actually get all 20 seats from their gerrymander? 11 seats exactly? less than 11? more than 11?

To calculate probabilities of specific outcomes, we need to use the binomial pdf, which in `R` is given by `dbinom()`

Example: Anticipating a landslide

What is the probability the Democrats will actually get all 20 seats from their gerrymander? 11 seats exactly? less than 11? more than 11?

To calculate probabilities of specific outcomes, we need to use the binomial pdf, which in `R` is given by `dbinom()`

To calculate probabilities of ranges of outcomes, we need to use the binomial cdf, which in `R` is given by `pbinom()`

Desired prob	`R` command	Result
$\Pr(20 \mid \pi = 0.75, M = 20)$	`dbinom(20,size=20,prob=0.75)`	0.003

Example: Anticipating a landslide

What is the probability the Democrats will actually get all 20 seats from their gerrymander? 11 seats exactly? less than 11? more than 11?

To calculate probabilities of specific outcomes, we need to use the binomial pdf, which in `R` is given by `dbinom()`

To calculate probabilities of ranges of outcomes, we need to use the binomial cdf, which in `R` is given by `pbinom()`

Desired prob	`R` command	Result
$\Pr(20 \mid \pi = 0.75, M = 20)$	`dbinom(20,size=20,prob=0.75)`	0.003
$\Pr(11 \mid \pi = 0.75, M = 20)$	`dbinom(11,size=20,prob=0.75)`	0.027

Example: Anticipating a landslide

What is the probability the Democrats will actually get all 20 seats from their gerrymander? 11 seats exactly? less than 11? more than 11?

To calculate probabilities of specific outcomes, we need to use the binomial pdf, which in `R` is given by `dbinom()`

To calculate probabilities of ranges of outcomes, we need to use the binomial cdf, which in `R` is given by `pbinom()`

Desired prob	`R` command	Result
$\Pr(20\|\pi = 0.75, M = 20)$	`dbinom(20,size=20,prob=0.75)`	0.003
$\Pr(11\|\pi = 0.75, M = 20)$	`dbinom(11,size=20,prob=0.75)`	0.027
$\Pr(< 11\|\pi = 0.75, M = 20)$	`pbinom(10,size=20,prob=0.75)`	0.014

Example: Anticipating a landslide

What is the probability the Democrats will actually get all 20 seats from their gerrymander? 11 seats exactly? less than 11? more than 11?

To calculate probabilities of specific outcomes, we need to use the binomial pdf, which in `R` is given by `dbinom()`

To calculate probabilities of ranges of outcomes, we need to use the binomial cdf, which in `R` is given by `pbinom()`

Desired prob	`R` command	Result
Pr$(20 \vert \pi = 0.75, M = 20)$	`dbinom(20,size=20,prob=0.75)`	0.003
Pr$(11 \vert \pi = 0.75, M = 20)$	`dbinom(11,size=20,prob=0.75)`	0.027
Pr$(< 11 \vert \pi = 0.75, M = 20)$	`pbinom(10,size=20,prob=0.75)`	0.014
Pr$(> 11 \vert \pi = 0.75, M = 20)$	`1-pbinom(11,size=20,prob=0.75)`	0.959