Binomial Distribution

The **binomial distribution** is the theoretical probability distribution appropriate when modeling the expected outcome, X, of N trials (or event sequences) that have the following characteristics:

- The outcome on every trial is binary
  - also called a **Bernoulli trial**
- The probability of the target outcome (usually called a “success”) is the same for all N trials
- The outcomes of the trials are *independent*
  - The probability of a success in any one trial must be the same from trial to trial
- The number of trials is fixed (you know how many times you're flipping a coin)

If these assumptions hold then X is a binomial random variable representing the *expected number of successes* over N trials, with expected success on each trial of θ .

A common and compact way of stating the same thing is:

$$X \sim B(N, \theta)$$

Note: Last lecture we used p to denote the probability of success. This time we'll use θ. θ is more correct for the population parameter, but you'll see p used a lot, too.

The probability distribution for X is defined by the following **probability mass function**:

$$P(X|\theta, N) = \frac{N!}{X!(N-X)!}\theta^X(1-\theta)^{N-X}$$

The probability mass function tells us what to expect for any particular value of X in the sample space.

The probability distribution for X is defined by the following **probability mass function**:

$$P(X = k|n, p) = \binom{n}{k} p^k (q)^{n-k}$$

This is the same exact thing as previous slide!

All theoretical distributions have a mass function (if discrete) or a density function (if continuous). These are the defining equations that tells us the generating process for the behavior of X.

$$P(X|\theta, N) = \frac{N!}{X!(N-X)!}\theta^X(1-\theta)^{N-X}$$

$\mathbf{P(X|\theta, N)}$ is a conditional probability: the probability of X **given** θ and N.

- X is the number of successful trials (in our other formula this is k -- as in r correct "choices")
- N is the number of trials; must be independent
- θ is the probability of success on any given trial (in our other formula, this is p)

θ and N are parameters of the binomial distribution.

The probability mass function tells us what to expect for any particular value of X in the sample space.

$$P(X|\theta, N) = \frac{N!}{X!(N-X)!}\theta^X(1-\theta)^{N-X}$$

$\boldsymbol{\theta^X(1-\theta)^{N-X}}$ is the probability of any particular instance of X.

- This is just a general form of the basic probability rule:

$$A \text{ and } B = P(A \cap B) = P(A)P(B)$$

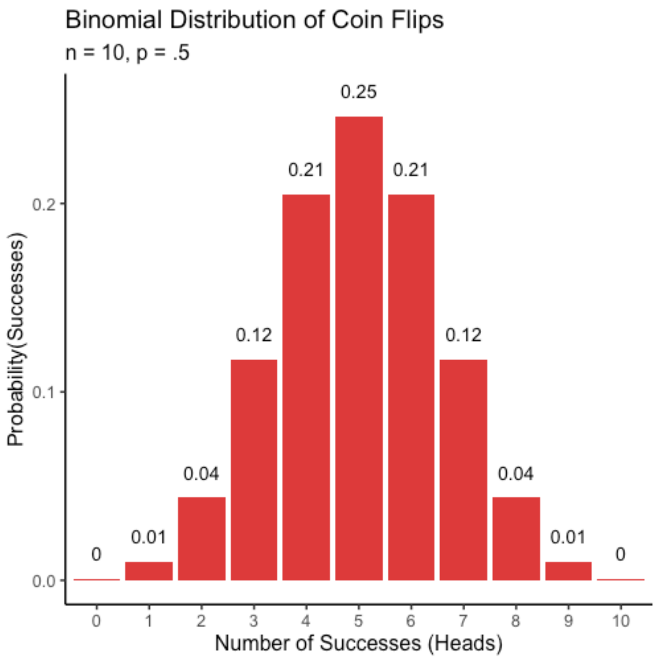
- AKA the intersection AKA $p(success) \times p(failure)$
- Note that this form of the rule assumes *independent events*.

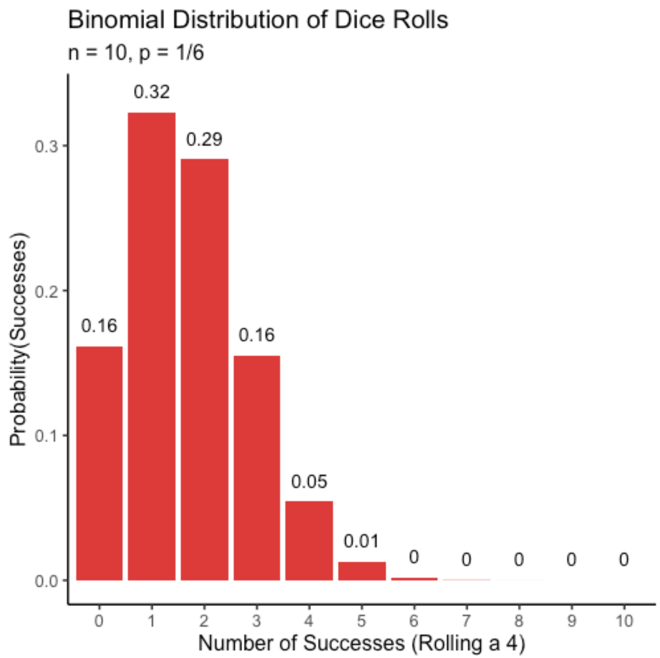
Binomial Distribution

$$p(X = k); n, p = \binom{n}{k} p^k (1-p)^{n-k}$$

- Parameters of this distribution are n and p
- If we change the parameters, we change the distribution
- **Family** of distributions

Family of Distributions





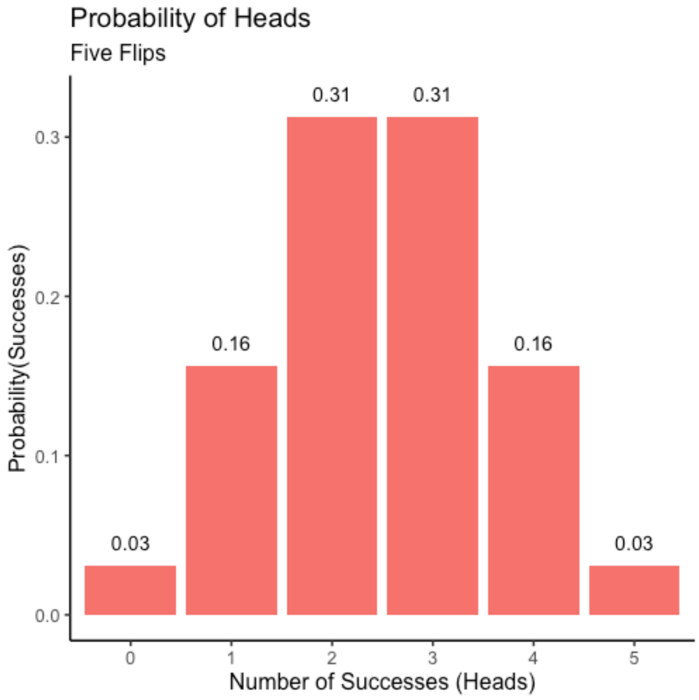
Parameters

Every probability distribution has an **expected value**.

- Expected Value is essentially the average value if you repeated the experiment...a lot
- Most likely result of the probability function
- The thing we would expect to happen if we have no other information than the parameters of the distribution
- The *long-run* average over an infinite amount of trials or samples

Averages vs. Expected Value

- What is the average number of successes out of 5 flips?
- In 5 coin flips, what is the average number of times you'd get "heads"?



Parameters

Every probability distribution has an **expected value**.

For the binomial distribution:

$$E(X) = N\theta$$

For 3 Heads out of 5 flips, $E(X) = 5 \times .5 = 2.5$

Each probability distribution also has a variance. For the binomial:

$$Var(X) = N\theta(1 - \theta)$$

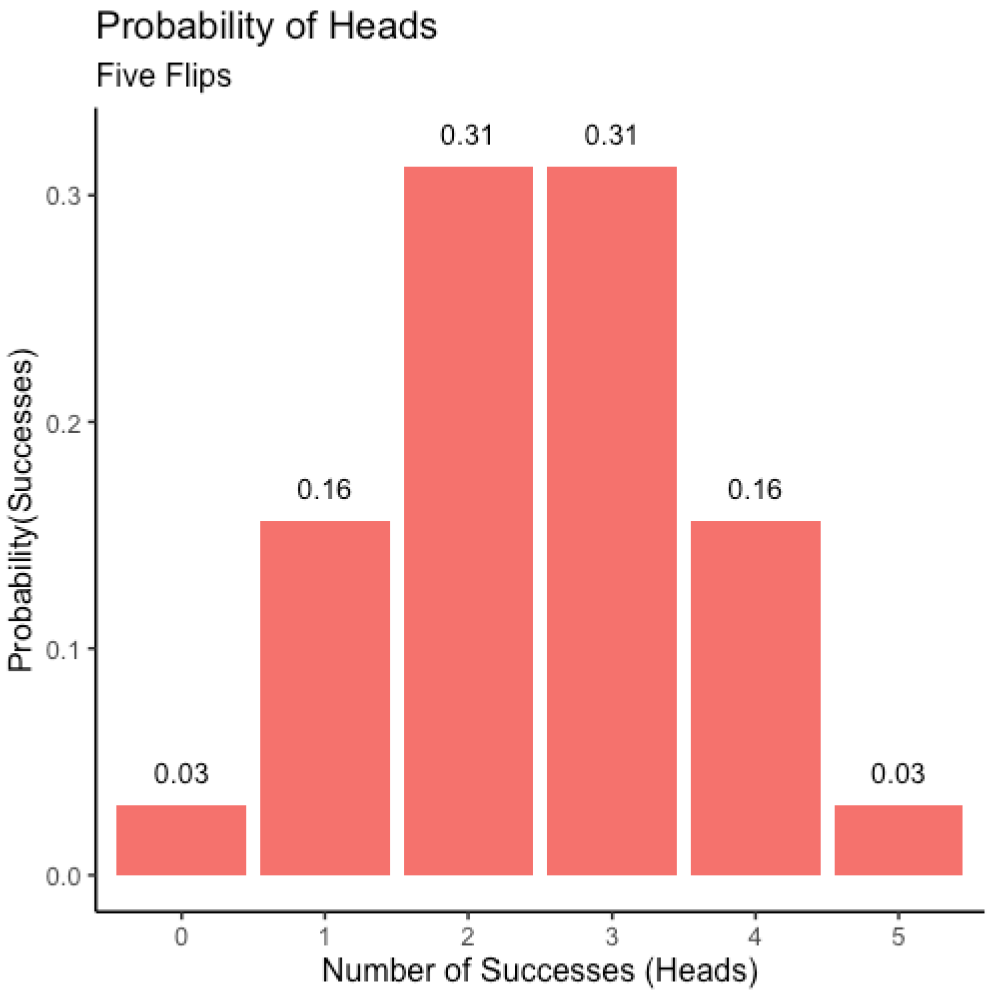
Importantly, this means our mean and variance are related in the binomial distribution, because they both depend on θ. How are they related?

Let's set our mean = variance

$$N\theta = N\theta(1-\theta)$$
$$Np = Np(1-p)$$
$$Np = Np(q)$$
$$\text{mean} = \text{mean} \times q$$

- If q = 1, then $E(X) = Var(X)$
- Else, $E(X) > Var(X)$

The mean, 2.5, does not exist in the sample space, and rounding up to 3 and claiming that to be the most typical outcome isn't right either.



- If you have a discrete distribution with a small N, these estimates may not have a sensible meaning

The **probability mass (density) function** allows us to answer other questions about the sample space that might be more important, or at least realistic.

- mass = discrete
- density = continuous

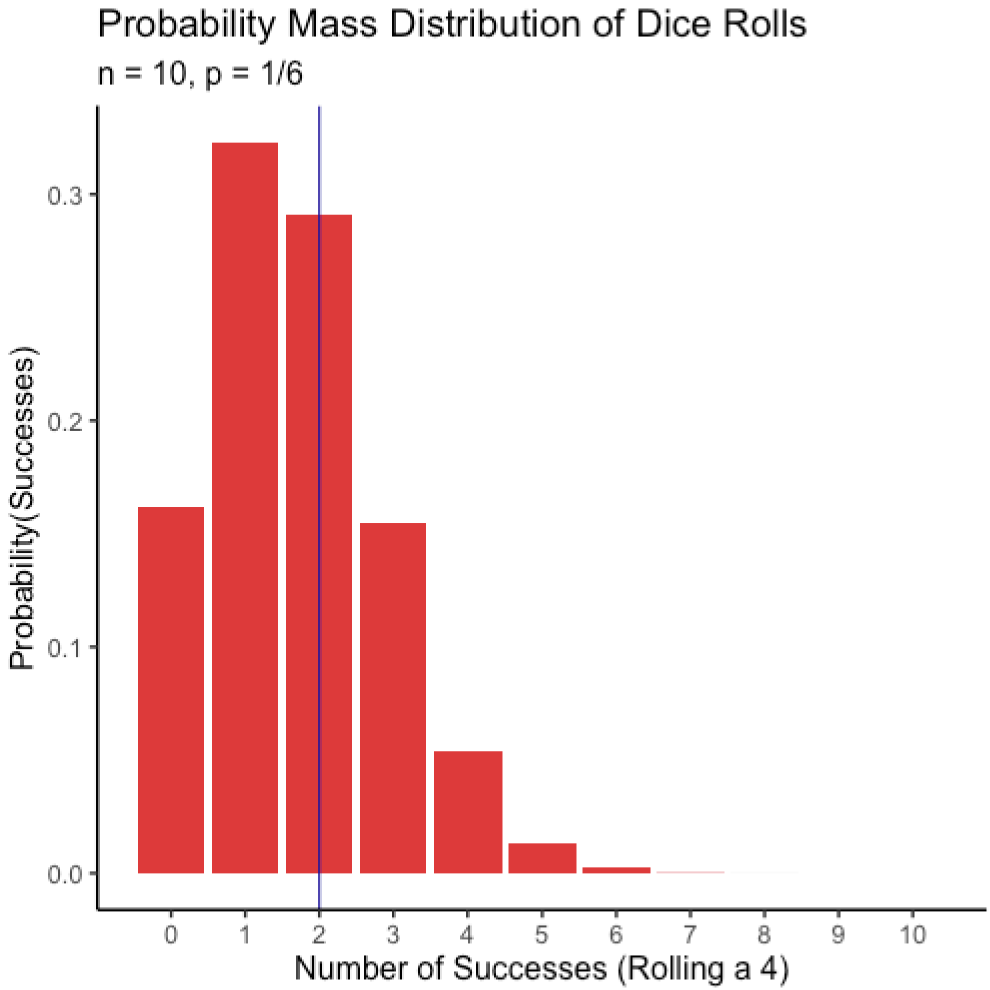
I might want to know the value in the sample space at or below which a certain proportion of outcomes fall.

> "At or below what outcome in the sample space do 75% of the outcomes fall?"

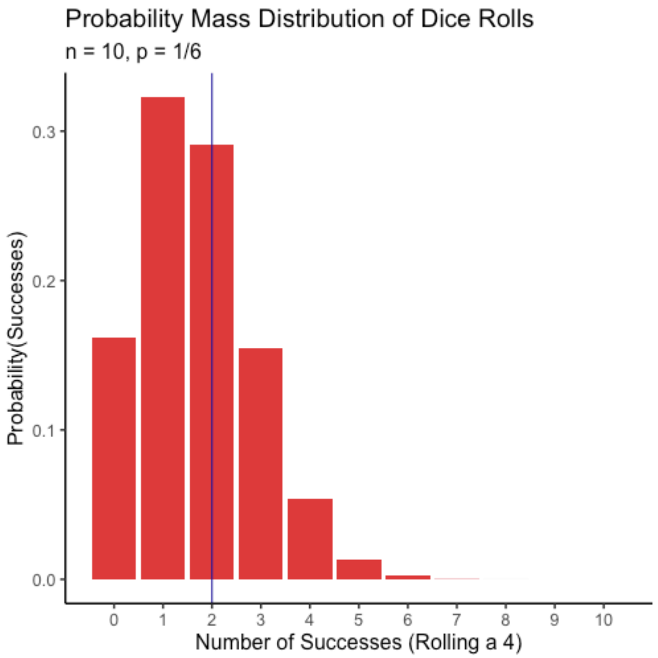
This is a **percentile or quantile** question.

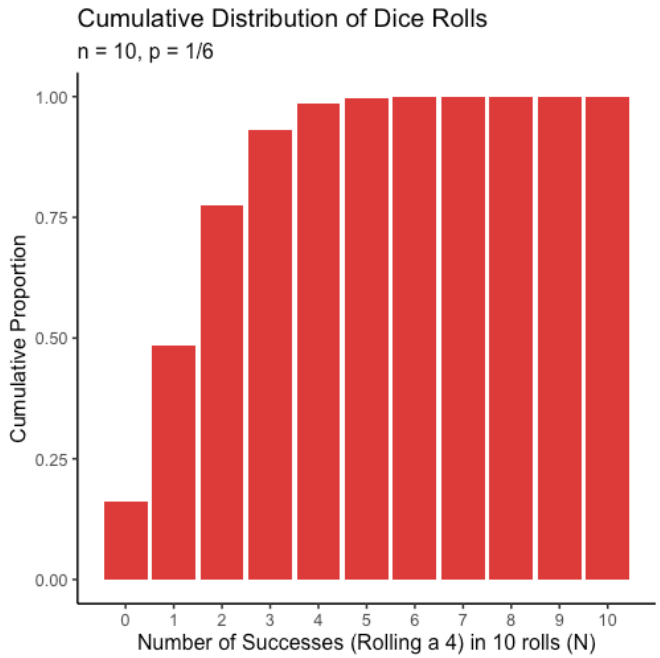
I might want to know the proportion of outcomes in the sample space that fall at or below a particular outcome. This is a **cumulative proportion** question.

At or below what outcome in the sample space do 75% of the outcomes fall? What is the outcome? Providing the 75%, trying to find the 2...



What proportion of outcomes in the sample space that fall at or below a given outcome? What percentage of outcomes fall at or below 2? Providing the 2, looking for the 75%



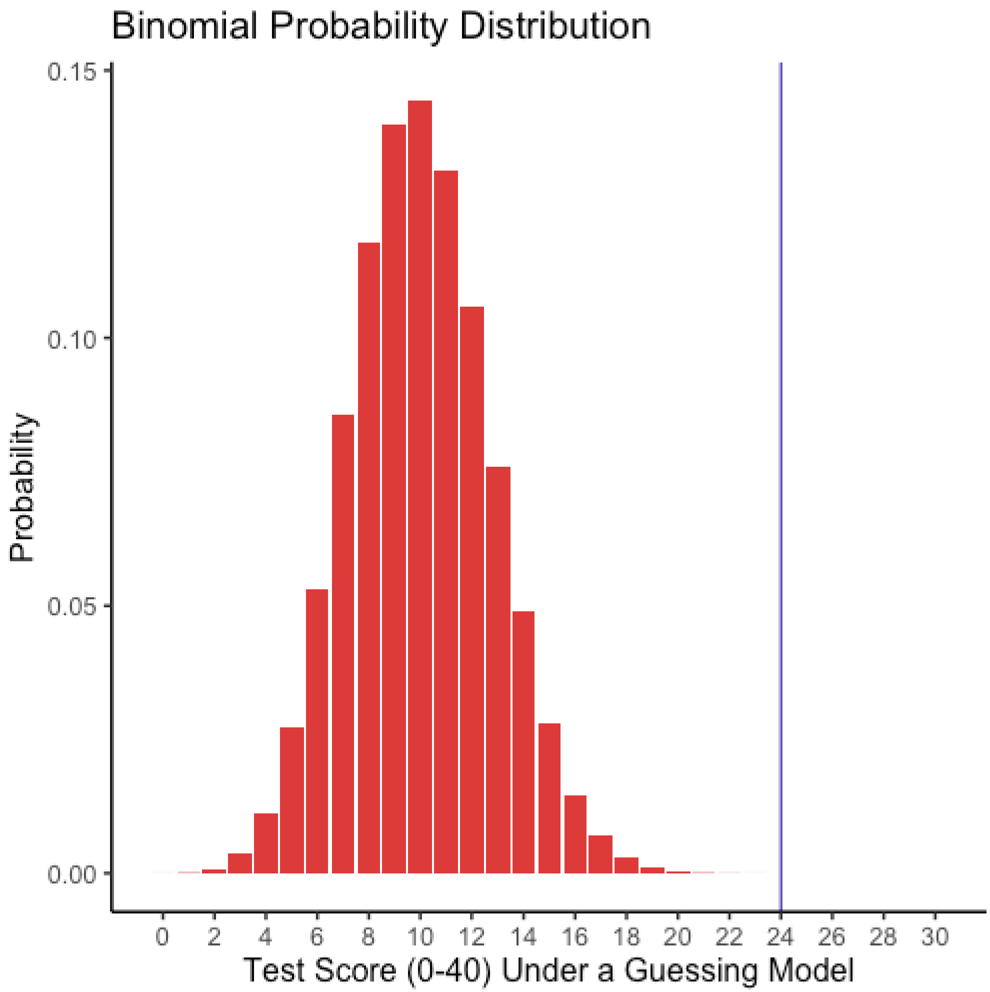


The binomial is of interest beyond describing the behavior of dice and coins.

Many practical outcomes might be best described by a binomial distribution.

For example, suppose I give a 40-item multiple choice test, with each question having 4 options.

- I am worried that students might do well by chance alone. I would not want to pass students in the class if they were just showing up for the exams and guessing for each question.
- What are the parameters in the binomial distribution that will help me address this question?
    - $N = 40$
    - $\theta = .25$


Binomial Probability Distribution
Probability
0.15
0.10
0.05
0.00
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Test Score (0-40) Under a Guessing Model

By hand

How likely is it that a guesser would score above the threshold (60%) necessary to pass the class by the most minimal standards?

$$P(24|.25,40)+$$
$$P(25|.25,40)+$$
$$P(26|.25,40)+$$
$$\dots$$
$$P(40|.25,40)$$

In R

How likely is it that a guesser would score above the threshold (60%) necessary to pass the class by the most minimal standards?

In R, we can calculate the cumulative probability (X or lower), using the `pbinom` function.

The probability of getting 23 questions or fewer correct:

```
pbinom(q = 23, size = 40, prob = .
```

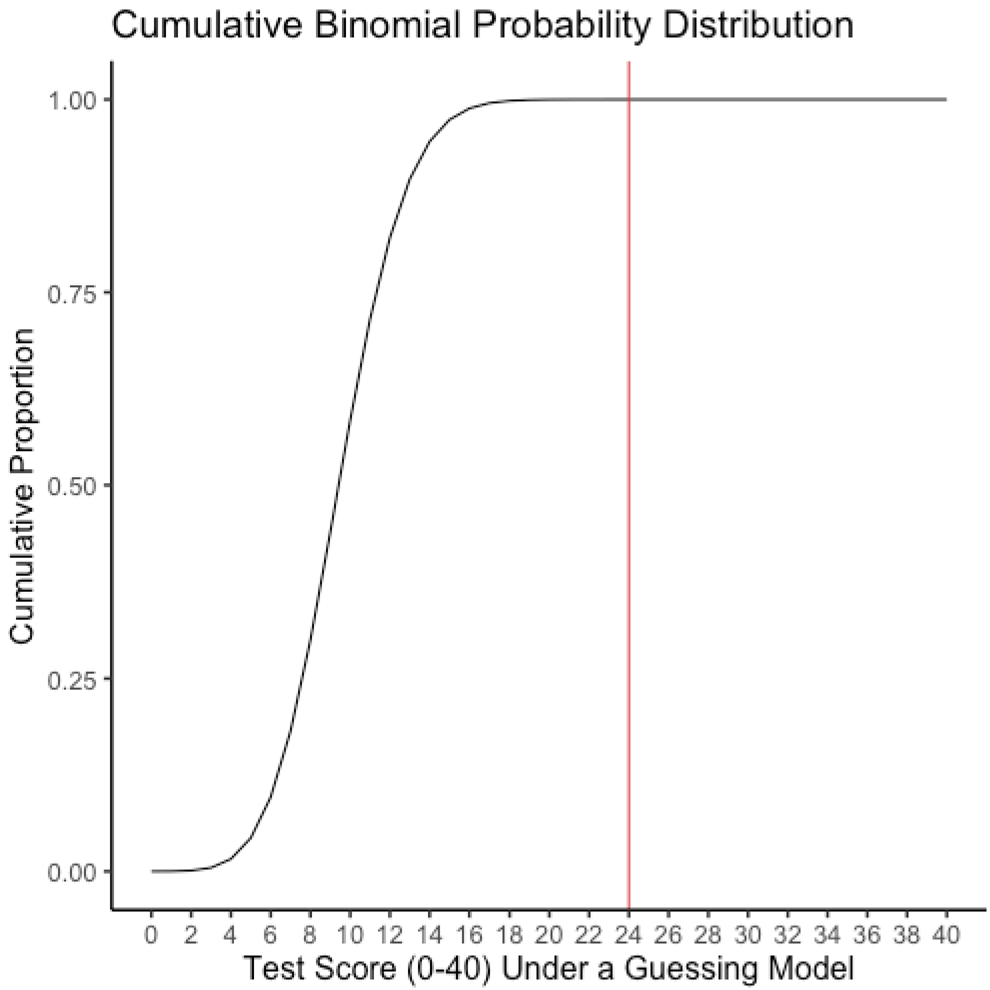
```
## [1] 0.9999972
```

The probability of getting 24 or more questions correct:

```
#Note the use of the Law of Total
1-pbinom(q = 23, size = 40, prob =
```

```
## [1] 2.825967e-06
```

Cumulatively, what proportion of guessers will fall below each score?



There's always a but

But, what assumptions are we making and what consequences will they have?

- The outcome on every trial is binary (also called a Bernoulli trial)
- The probability of the target outcome (usually called a "success") is the same for all N trials
- The trials are **independent** $P(A \cap B) = P(A|B)P(B) = P(A)P(B)$
- The number of trials is fixed

In probability and statistics, if the assumptions are wrong then inferences based on those assumptions could be wrong too, perhaps seriously so.

> All models are wrong, but some models are useful.
> (G.E.P. Box)

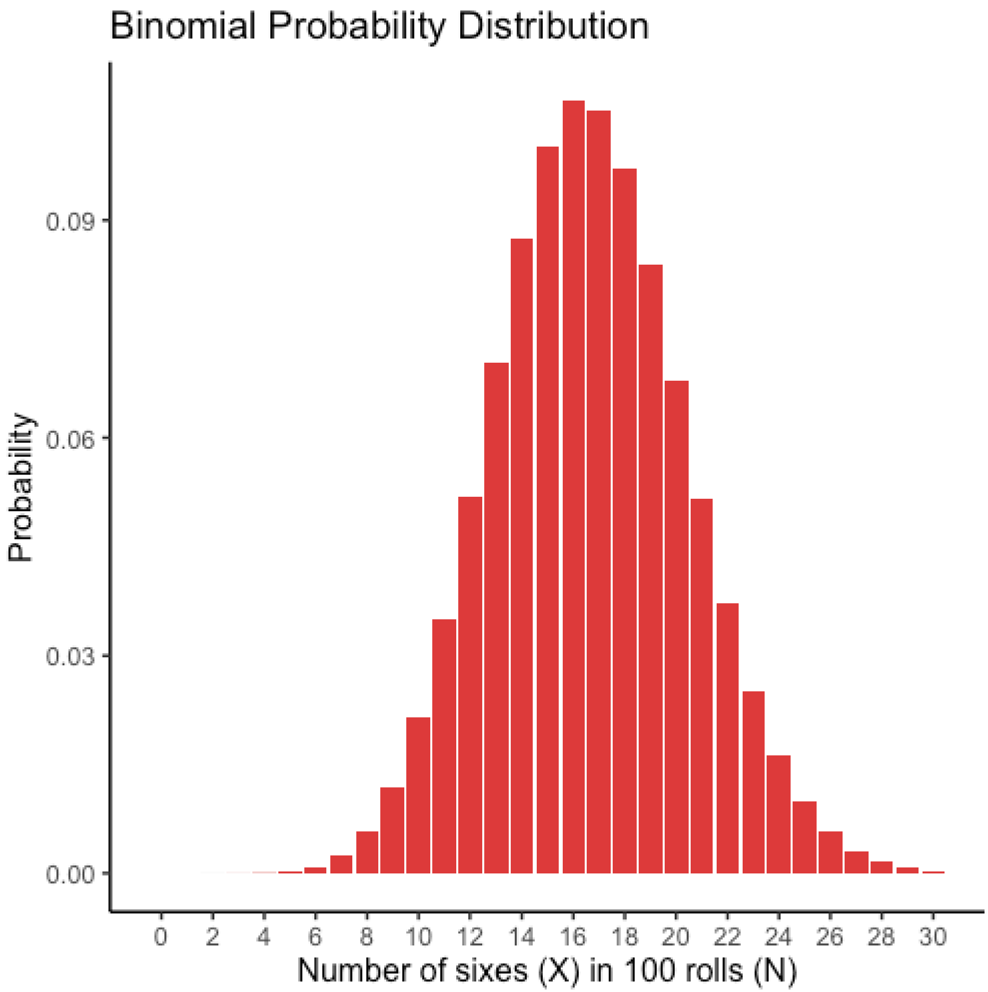
We might have viable alternative models:

- **Geometric distribution:** Used if we are interested in the number of trials required for one "success" to occur
  - "how many times do I start my computer before it fails to start at all?"

- **Negative binomial distribution:** Used if we are interested in the number of successes in a series of repeated trials until a specified number of failures are seen
  - "A child won't return from trick or treating until they get 5 full-size candy bars. What is the probability that they will have to visit 34 homes to get this?"

As N increases, the binomial becomes more normal in appearance.

Because of the difficulties in calculating large factorials, there is a large-sample normal approximation to the binomial. The normal distribution is useful for a lot of other reasons too.



Next time...

the normal distribution