Probability (part 2)

Ryan Miller

Babies Revisited

- On the first day of class we discussed a study involving babies choosing between a "helper" and "hinderer" toy
  - Recall that 14 of 16 infants chose the "helper" toy
  - We used simulation to determine that this result would be very unlikely to happen by random chance alone
- We're now ready to reach this conclusion more precisely using probability

Babies Revisited

- Let A_i denote the event that the i^{th} baby chooses the "helper" toy

Babies Revisited

- Let A_i denote the event that the i^{th} baby chooses the "helper" toy
- Of interest is the probability that *at least 14 babies* chose the helper toy under the *null model* that each choice was random (we called this probability the p-value)
  - What does the null model suggest regarding $P(A_i)$?

Babies Revisited

- Let A_i denote the event that the i^{th} baby chooses the "helper" toy
- Of interest is the probability that *at least 14 babies* chose the helper toy under the *null model* that each choice was random (we called this probability the p-value)
  - What does the null model suggest regarding $P(A_i)$?
  - If the null model were true, $P(A_i) = 0.5$

Babies Revisited

- Because each baby's choice is independent, the multiplication rule is a useful starting point
  - $P(A_1 \text{ and } A_2 \text{ and } \ldots) = P(A_1) * P(A_2) * \ldots$
- We might calculate the probability of seeing 14 "helper" and 2 "hinder" choices as $(.5)^{14}(.5)^{2}$, so is this the *p*-value?

Babies Revisited

- Because each baby's choice is independent, the multiplication rule is a useful starting point
  - $P(A_1 \text{ and } A_2 \text{ and } \ldots) = P(A_1) * P(A_2) * \ldots$
- We might calculate the probability of seeing 14 "helper" and 2 "hinder" choices as $(.5)^{14}(.5)^2$, so is this the p-value?
- Unfortunately the answer is "no", this calculation ignores two key things...

Combinations

- There are very many ways that 14 of 16 babies could choose the "helper" toy, but we only considered one of them

Combinations

- There are very many ways that 14 of 16 babies could choose the "helper" toy, but we only considered one of them
- As a simplified example, consider two babies and the result that 1 of 2 chose the "helper"
  - This result could happen in two ways: the first baby chose the "helper" and the second baby chose the "hinderer", or the first baby chose the "hinderer" and the second baby chose the "helper"

Combinations

- There are very many ways that 14 of 16 babies could choose the "helper" toy, but we only considered one of them
- As a simplified example, consider two babies and the result that 1 of 2 chose the "helper"
  - This result could happen in two ways: the first baby chose the "helper" and the second baby chose the "hinderer", or the first baby chose the "hinderer" and the second baby chose the "helper"
- Thus, we need to consider the number of different *combinations* that could result in 14 of 16 infants choosing the "helper" when calculating the *p*-value

Combinations

Generally speaking, the number of ways that k binary "successes" can occur in n trials is expressed by:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

This expression is read "n choose k", and the exclamation point denotes a factorial (ie: $4! = 4 * 3 * 2 * 1 = 24$)

Babies Revisited (attempt #2)

Each possible combination is equally likely (since the babies choose independently), so we can revise our calculation of the probability of seeing 14 "helper" and 2 "hinder" choices under the *null model*:

$$\binom{16}{14}(.5)^{14}(.5)^{2}$$

`R` can help us with the calculation:

```
choose(16,14)*(.5)^14*(.5)^2
```

```
## [1] 0.001831055
```

So is *this* the *p*-value?

Babies Revisited (attempt #2)

Each possible combination is equally likely (since the babies choose independently), so we can revise our calculation of the probability of seeing 14 "helper" and 2 "hinder" choices under the *null model*:

$$\binom{16}{14}(.5)^{14}(.5)^{2}$$

`R` can help us with the calculation:

```
choose(16,14)*(.5)^14*(.5)^2
```

```
## [1] 0.001831055
```

So is *this* the *p*-value? Unfortunately the answer is still no. . .

Definition of a p-value

- A **p-value** is the probability of seeing results *at least as extreme* as the observed outcome under a null model

Definition of a p-value

- A **p-value** is the probability of seeing results *at least as extreme* as the observed outcome under a null model
- Thus, for our babies example, the p-value is calculated (using the addition rule):

$$\binom{16}{14}(.5)^{14}(.5)^{2}+\binom{16}{15}(.5)^{15}(.5)^{1}+\binom{16}{16}(.5)^{16}(.5)^{0}=0.0021$$

Definition of a *p*-value

- A **p-value** is the probability of seeing results *at least as extreme* as the observed outcome under a null model
- Thus, for our babies example, the *p*-value is calculated (using the addition rule):

$$\binom{16}{14}(.5)^{14}(.5)^{2}+\binom{16}{15}(.5)^{15}(.5)^{1}+\binom{16}{16}(.5)^{16}(.5)^{0}=0.0021$$

- The past several slides have introduced one of the key applications of probability (calculating a *p*-value using a null model)
- We'll now go back and introduce some terminology to make the procedure more generalizable

Random Variables and their Distributions

- Statisticians use the term **random variable** to describe a random process with a numerical outcome
  - Random variables are denoted by a capital letter (usually X, Y, or Z)

Random Variables and their Distributions

- Statisticians use the term **random variable** to describe a random process with a numerical outcome
  - Random variables are denoted by a capital letter (usually X, Y, or Z)
- We often consider a set of random variables from independent repeats of the same random process
  - For example, $X_1, X_2, \ldots, X_n$ might describe each of the choice of one of the $n = 16$ infants (where "helper" is mapped to the numerical outcome of 1 and "hinderer" is mapped to 0)

Random Variables and their Distributions

- Statisticians use the term **random variable** to describe a random process with a numerical outcome
  - Random variables are denoted by a capital letter (usually X, Y, or Z)
- We often consider a set of random variables from independent repeats of the same random process
  - For example, $X_1, X_2, \ldots, X_n$ might describe each of the choice of one of the $n = 16$ infants (where "helper" is mapped to the numerical outcome of 1 and "hinderer" is mapped to 0)
- *Realized* or observed values of a random variable are denoted using lower case
  - For example, if the first infant selected the "helper", then $x_1 = 1$

Random Variables and their Distributions

- Statisticians use the term **random variable** to describe a random process with a numerical outcome
  - Random variables are denoted by a capital letter (usually X, Y, or Z)
- We often consider a set of random variables from independent repeats of the same random process
  - For example, $X_1, X_2, \ldots, X_n$ might describe each of the choice of one of the $n = 16$ infants (where "helper" is mapped to the numerical outcome of 1 and "hinderer" is mapped to 0)
- *Realized* or observed values of a random variable are denoted using lower case
  - For example, if the first infant selected the "helper", then $x_1 = 1$
- Random variables have *probability distributions*. In our example, the null model prompted the following distribution:

x	0	1
$P(X = x)$	.5	.5

The Bernoulli Distribution

- A random process with a binary outcome is called a **Bernoulli Trial**
- One of the outcomes is considered a "success" and denoted by a numeric value of 1:

x	0	1
$P(X = x)$	.5	.5

- Notice the *sample proportion*, $\hat{p}$, is actually just the sample mean of a bunch of Bernoulli random variables, for example:

$$\hat{p} = \frac{\text{number of successes}}{\text{number of trials}} = \frac{1+1+1+0+1+0+0+1+1+0}{10} = .6$$

- As you might expect, $\hat{p}$ can serve as an estimate of p, the true probability of a "success" (something we'll revisit later)

The Binomial Distribution

- Situations involving several independent Bernoulli random variables are common enough that statisticians use a more complex distribution to represent them

The Binomial Distribution

- Situations involving several independent Bernoulli random variables are common enough that statisticians use a more complex distribution to represent them
- Let X be a random variable representing the number of "successes" in k a Bernoulli trial repetitions
  - For example, X might denote the number of babies choosing the "helper" toy
- We've already seen how to find probability distribution for this random variable (under the null model where $p = 0.5$):

x	0	1	2	...
$P(X = x)$	$\binom{16}{0}(.5)^0(.5)^{16}$	$\binom{16}{1}(.5)^1(.5)^{15}$	$\binom{16}{2}(.5)^2(.5)^{14}$	...

The Binomial Distribution

- In this example you'll notice that $P(X)$ can be described by a particular function:

$$P(X = x) = \binom{n}{x}(.5)^x(.5)^{n-x}$$

- The **binomial distribution**, which characterized by the *probability function* above, describes the probability of observing exactly x "successes" in n independent Bernoulli Trials

The Binomial Distribution

- There was a time when using the binomial distribution to calculate p-values was very tedious
- For example, consider repeating the infant-choice study with 210 babies and observing 140 "helper" choices

The Binomial Distribution

- There was a time when using the binomial distribution to calculate *p*-values was very tedious
- For example, consider repeating the infant-choice study with 210 babies and observing 140 "helper" choices
- Calculating the *p*-value would require the summation of *70 different binomial terms*

$$\sum_{k=140}^{210} \binom{n}{k} (.5)^k (.5)^{n-k}$$

Binomial Distribution in R

The calculation is trivial for modern computers:

```
binom.test(140, 210, p = 0.5, alternative = "greater")
```

```
##
##  Exact binomial test
##
## data:  140 and 210
## number of successes = 140, number of trials = 210, p-value = 7.773e-07
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.6092413 1.0000000
## sample estimates:
## probability of success
##              0.6666667
```

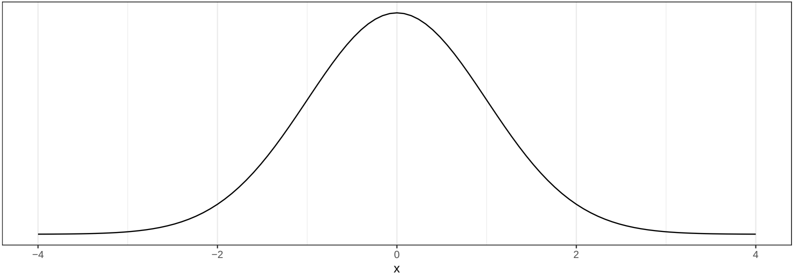
```
pbinom(139,210, prob = 0.5, lower.tail = FALSE)
```

```
## [1] 7.772811e-07
```

But before modern computing, statisticians needed to avoid such a tedious calculation. . .

The Normal Distribution

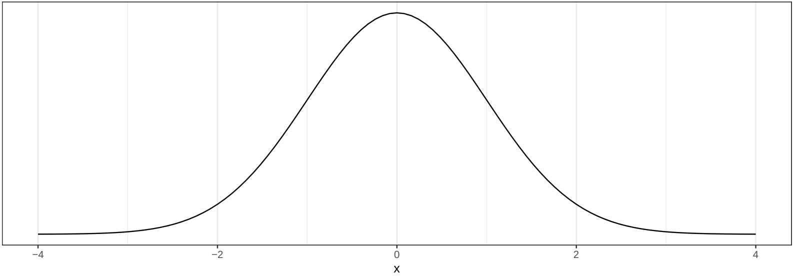
Among all of the different probability distributions used by statisticians, perhaps the most common is the **normal distribution**:



- The normal curve is a symmetric, bell-shaped distribution that depends upon two quantities: a center μ , and a standard deviation σ

The Normal Distribution

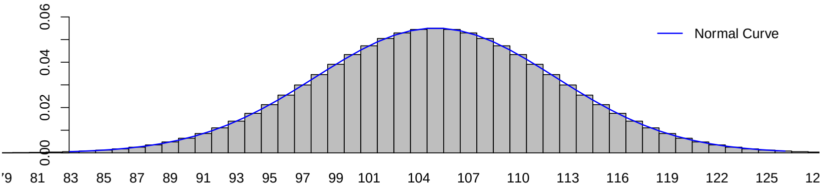
Among all of the different probability distributions used by statisticians, perhaps the most common is the **normal distribution**:



- The normal curve is a symmetric, bell-shaped distribution that depends upon two quantities: a center μ , and a standard deviation σ
- The **standard normal** distribution is depicted above, it's centered at 0 with a standard deviation of 1
  - We use often use the shorthand: $N(0, 1)$

Normal Approximation

- The binomial distribution for the scenario we were considering (observing 140 of 210 successes) can be approximated by the normal curve
  - Below the normal density overlaid on some of the values of the random variable (ie: 0 through 210) and their corresponding binomial distribution probabilities



- Given this normal curve, how might we use it to calculate an approximate *p*-value?

Normal Approximation

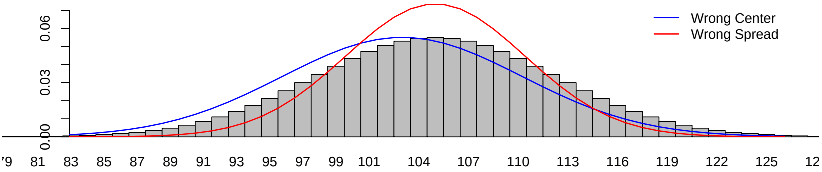
- We can use a normal approximation to calculate a *p*-value by finding the *area under the curve* in the regions of interest
- To do so, we'll need the *normal density function*:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- We'll first need to determine the proper values of μ and σ
  - Otherwise the center and spread of the curve won't match the binomial distribution in our application

Examples of Bad Normal Approximations

Below are some bad normal approximations (ie: their values of μ and σ are inappropriate for our scenario):



It's obvious to see how a bad approximation will yield an incorrect estimate of the *p*-value

Expected Values

- To properly center our normal curve, we'll use the **expected value**, or "average" outcome, of our random variable

Expected Values

- To properly center our normal curve, we'll use the **expected value**, or "average" outcome, of our random variable
- For a *discrete* random variable, the expected value is the *sum of each possible outcome value weighted by its probability*
  - Generally speaking, this can be expressed (using $i \in \{1, \ldots, k\}$ to index the different outcomes):

$$E(X) = \sum_{i=1}^{k} x_k P(X = x_k)$$

- In our latest example:

$$E(X) = 0 * P(X = 0) + 1 * P(X = 1) + \ldots + 210 * P(X = 210)$$

Expected Value (Binomial)

- Expected value calculations can be cumbersome, but if X is a *binomial random variable* the calculation yields:

$$E(X) = n * p$$

- Where n is the number of Bernoulli trials and p is the success probability
  - We won't cover the details, but general strategy is to consider X as the sum of n independent Bernoulli trials and sum their individual expected values (which are each p)

Expected Value (Binomial)

- Expected value calculations can be cumbersome, but if X is a *binomial random variable* the calculation yields:

$$E(X) = n * p$$

- Where n is the number of Bernoulli trials and p is the success probability
  - We won't cover the details, but general strategy is to consider X as the sum of n independent Bernoulli trials and sum their individual expected values (which are each p)
- So, we center our approximation of the probability distribution of X by a normal curve with $\mu = n * p$
  - But how do we find the proper value of σ?

Variance

- The **variance** of a random variable is defined:

$$Var(X) = E((X - E(X))^2)$$

- In words, variance is the *expected squared distance* of a random variable from its expected value
  - Standard deviation is just the square root of variance!

Variance

- The **variance** of a random variable is defined:

$$Var(X) = E((X - E(X))^2)$$

- In words, variance is the *expected squared distance* of a random variable from its expected value
  - Standard deviation is just the square root of variance!
- If X is a binomial random variable:

$$Var(X) = n * p * (1 - p)$$

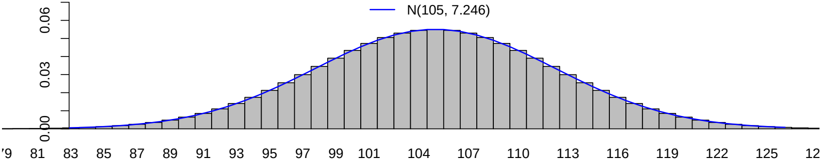
- Proof is omitted, but it's essentially the same strategy used to find the expected value of X
  - Note that a Bernoulli random variable has a variance of $p * (1 - p)$

Putting it all Together

- We've now determined that if X is a binomial random variable representing the outcome of n trials with success probability p
  - $E(X) = n * p$
  - $Var(X) = n * p * (1 - p)$ meaning Std Dev$(X) = \sqrt{n * (p) * (1 - p)}$

Putting it all Together

- We've now determined that if X is a binomial random variable representing the outcome of n trials with success probability p
  - $E(X) = n * p$
  - $Var(X) = n * p * (1 - p)$ meaning Std Dev$(X) = \sqrt{n * (p) * (1 - p)}$
- Thus, we can approximate the probability distribution of X by the normal curve: $N(np, \sqrt{np(1-p)})$
  - In our ongoing example, $210 * 0.5 = 105$ and $\sqrt{210 * 0.5 * 0.5} = 7.246$, leading to the following:



Conclusion

We haven't yet used to normal curve to find probabilities (like the p-value), but that's where we're headed next. For now, here is a review of the key terms and concepts from this lecture:

- **Random Variable** - a numerical value resulting from a random process
- **Probability Distribution** - a mapping of a random variable's values to probabilities
  - Examples so far include the **Bernoulli**, **binomial**, and **normal** distributions
- **Expected Value** - the average outcome of a random variable
- **Variance** - the average squared distance of a random variable from it's expected value
  - **Standard deviation** - the square root of variance