Quantitative Methods

Serena DeStefani – Lecture 9 – 7/16/2020

Announcements

- Review on Tuesday and Wednesday: please email questions/ topic for review
- Midterm on Thursday

Random Phenomena Vocabulary

Trial

- Each occasion which we observe a random phenomena (the “thing” that is happening)

Outcome

- The value of the trial for the random phenomena (what *could* happen in a trial, specifically)

Event

- The combination of the trial’s outcomes (what actually happened)

Sample Space

- The collection of all possible outcomes (what could ever possibly happen)

Flipping Two Coins

Trial

- The flipping of the two coins

Outcome

- Heads or tails for each flip

Event

- HT, for example

Sample Space

- **S** = {HH, HT, TH, TT}

The Law of Large Numbers

- If you flip a coin once, you will either get 100% heads or 0% heads.
- If you flip a coin 1000 times…
- …you will probably get close to 50% heads.

The **Law of Large Numbers** states that for many trials, the proportion of times an event occurs settles down to one number.

- This number is called the empirical probability.

The Nonexistent Law of Averages

Wrong

- If you flip a coin 6 times and get 6 heads, then you are due for a tail on the next flip.
- You put 10 quarters in the slot machine and lose each time. You are just a bad luck person, so you have a smaller chance of winning on the 11th try.

- There is no such thing as the Law of Averages for **short** runs.

Theoretical Probability

American Roulette

- 18 Red, 18 Black, 2 Green
- If you bet on Red, what is the probability of winning?

Theoretical Probability

- $P(\mathbf{A}) = \dfrac{\text{\# of outcomes in } \mathbf{A}}{\text{\# of possible outcomes}}$
- $P(\text{red}) = \dfrac{18}{38}$

Rules: 1, 2 and 3

Rule 1: $0 \leq P(\mathbf{A}) \leq 1$

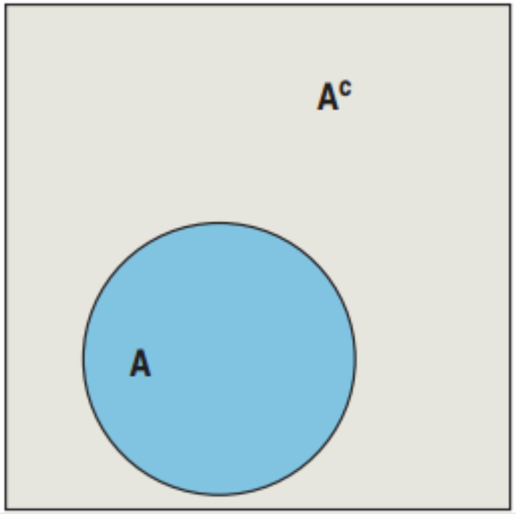
That's how we define probability

Rule 2: $P(\mathbf{S}) = 1$

The set of all possible outcomes has probability 1

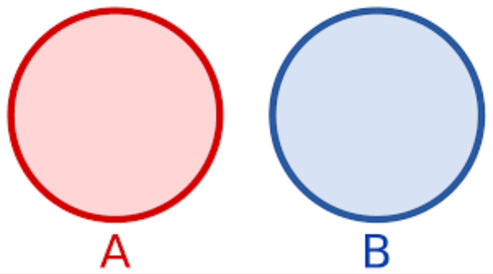
The Rule of Complements: $P(\mathbf{A}^C) = 1 - P(\mathbf{A})$

$\mathbf{A}^C$ is the event of "**A** not happening".

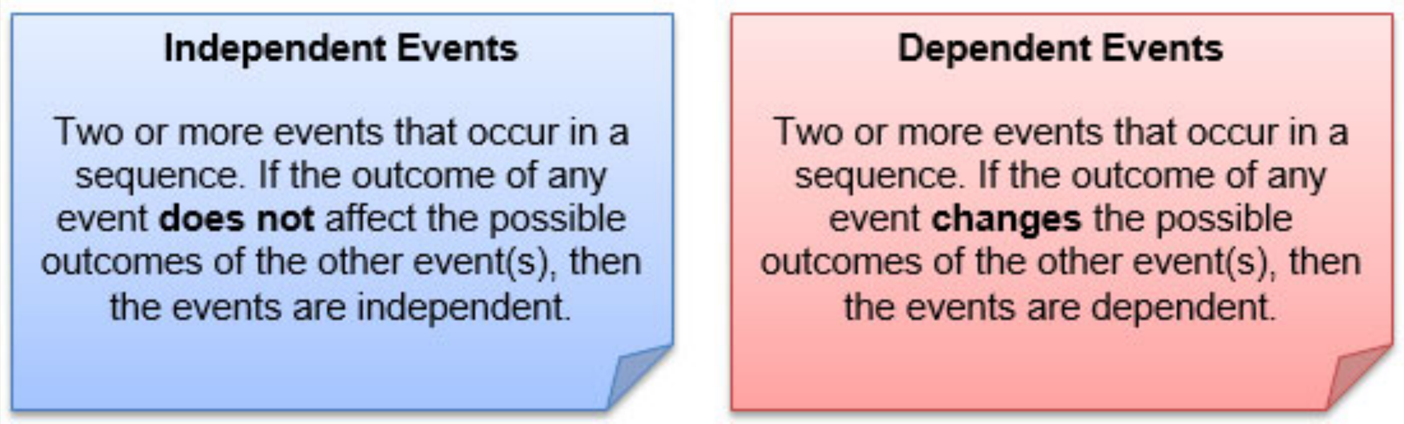


Events

Disjoint



Independent and Dependent

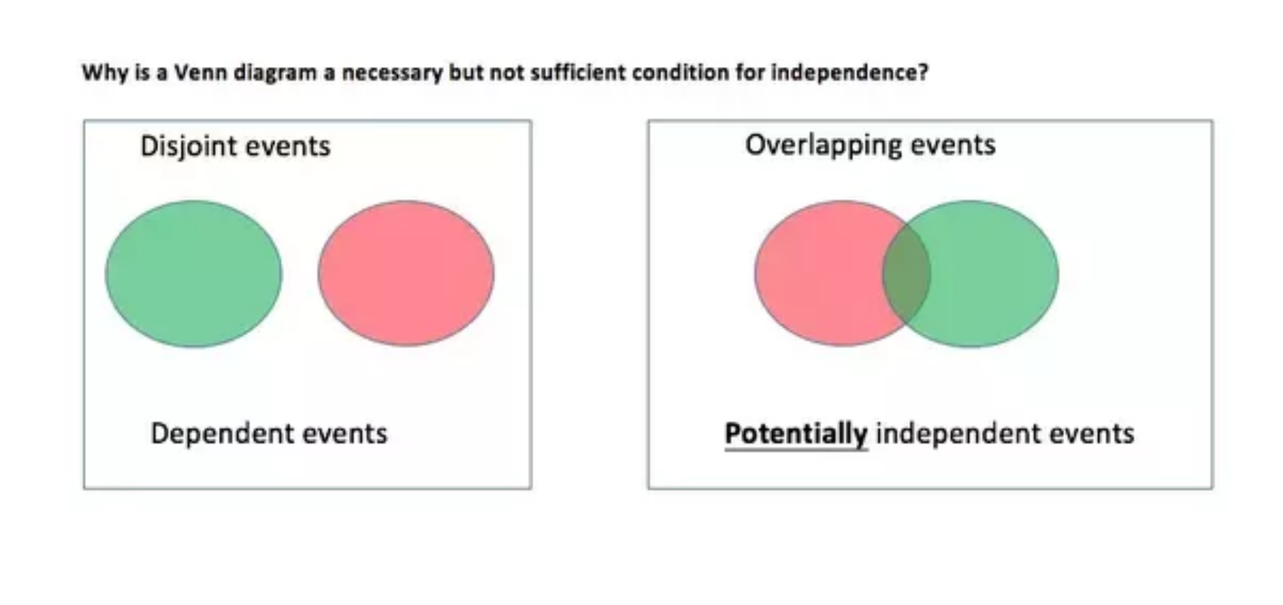


Independent Events

Two or more events that occur in a sequence. If the outcome of any event **does not** affect the possible outcomes of the other event(s), then the events are independent.

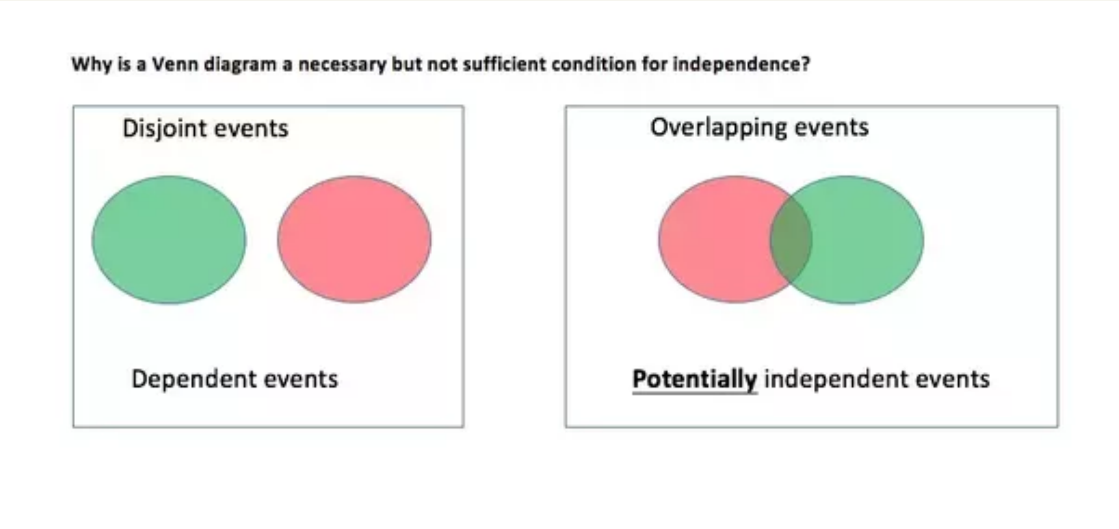
Dependent Events

Two or more events that occur in a sequence. If the outcome of any event **changes** the possible outcomes of the other event(s), then the events are dependent.



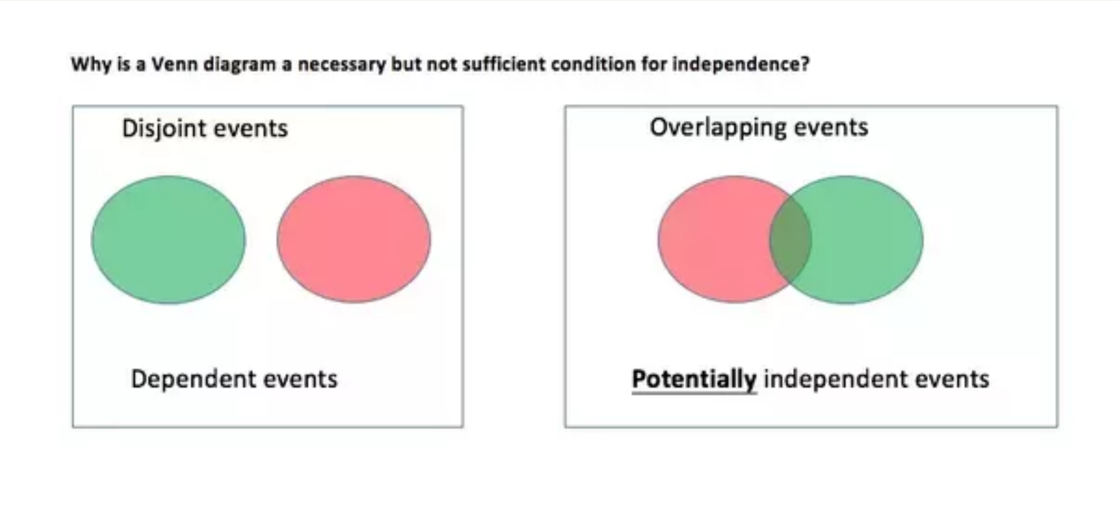
Why is a Venn diagram a necessary but not sufficient condition for independence?

Events


Why is a Venn diagram a necessary but not sufficient condition for independence?

Disjoint events:
OR: Addition
AND: makes no sense
But what if two events overlap?
They can be either independent or dependent
OR but not disjoint? General addition rule
AND?
If they are independent: Multiplication rule
If they are dependent, General Multiplication rule

Events



Why is a Venn diagram a necessary but not sufficient condition for independence?

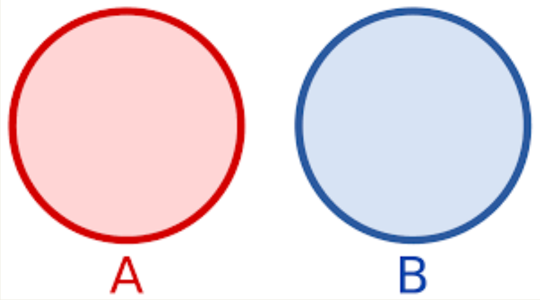
	Disjoint	Overlapping
Dependent	YES	YES
Independent	DOES NOT EXIST	YES

Rule 4: The Addition Rule

Suppose

P(sophomore) = 0.2 and *P*(junior) = 0.3

- Find *P*(sophomore OR junior)
- Solution: 0.2 + 0.3 = 0.5
- This works because sophomore and junior are **disjoint events**. They have no outcomes in common.



The Addition Rule

- If **A** and **B** are **disjoint** events, then

$$P(\mathbf{A} \text{ OR } \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B})$$

Rule 5: The Multiplication Rule

The probability that an Atlanta to Houston flight is on time is 0.85.

- If you have to fly every Monday, find the probability that your first two Monday flights will be on time.

Multiplication Rule: For **independent** events **A** and **B**:

$P(\mathbf{A} \text{ AND } \mathbf{B}) = P(\mathbf{A}) \times P(\mathbf{B})$

- $P(1^{st} \text{ on time AND } 2^{nd} \text{ on time})$

$$= P(\text{1st on time}) \times P(\text{2nd on time})$$
$$= 0.85 \times 0.85$$
$$= 0.7225$$

Red Light AND Green Light AND Yellow Light

Find the probability that the light will be red on Monday, green on Tuesday, and yellow on Wednesday.

- The multiplication rule works for more than 2 events.

- P(red Mon. AND green Tues. AND yellow Wed.)

  = P(red Mon.) × P(green Tues.) × P(yellow Wed.)

  = 0.61 × 0.35 × 0.04

  = 0.00854

Or But Not Disjoint

Your Wallet

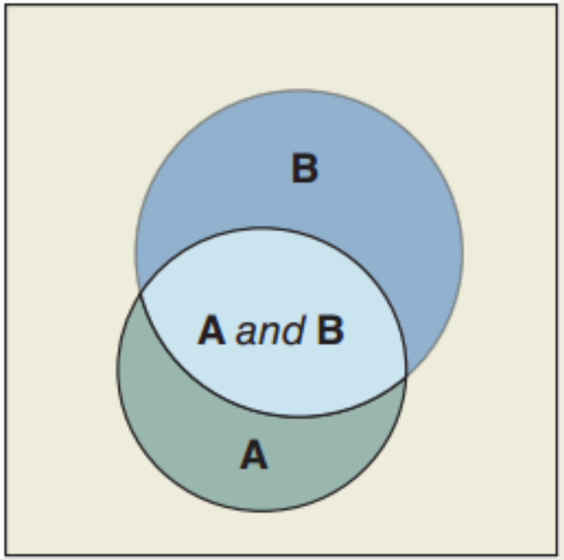
- **S** = {\$1, \$2, \$5, \$10, \$20, \$50, \$100}
- **A** = {odd numbered value} = {\$1, \$5}
- **B** = {bill with a building} = {\$5, \$10, \$20, \$50, \$100}

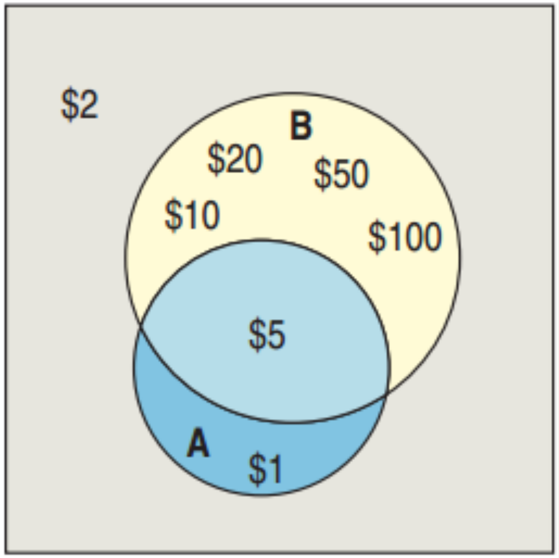
- Why is *P*(**A** *or* **B**) ≠ *P*(**A**) + *P*(**B**)?
- Answer: **A** and **B** are **not disjoint**.
- The intersection **A** *and* **B** = {\$5} is double counted.
- To find *P*(**A** *or* **B**), subtract *P*(**A** *and* **B**).

The General Addition Rule

P(**A** *or* **B**) = *P*(**A**) + *P*(**B**) – *P*(**A** *and* **B**)

- **The General Addition Rule in words:** Add the probabilities of the two events and then subtract the probability of their intersection.





P(odd amount *or* bill with a building)

= *P*(**A**) + *P*(**B**) – *P*(**A** and **B**}

= *P*({$1, $5}) + *P*({$5, $10, $20, $50, $100}) – *P*({$5})

33% Relationship, 25% Sport, 11% Both

Events

- **R** = {in a relationship}
- **S** = {involved in sports}

Calculations

- $P(\mathbf{R}\ or\ \mathbf{S}) = P(\mathbf{R}) + P(\mathbf{S}) - P(\mathbf{R}\ and\ \mathbf{S})$

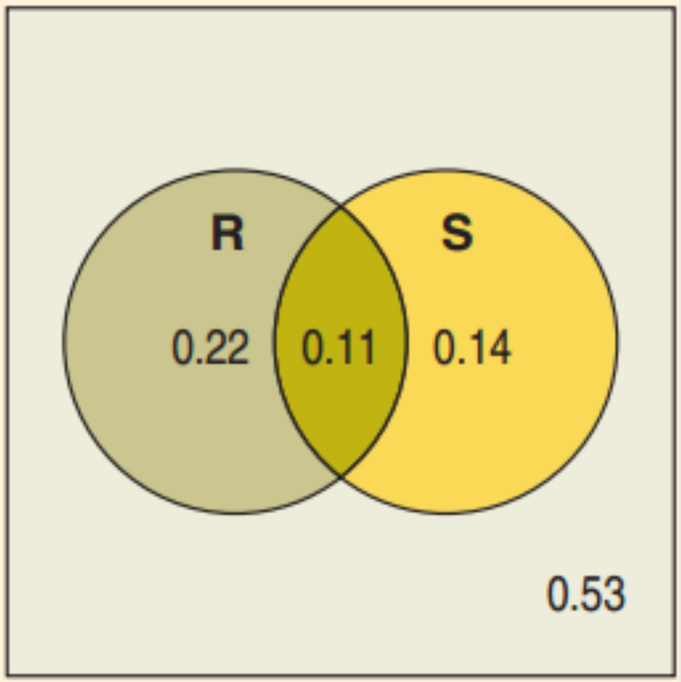
$$= 0.33 + 0.25 - 0.11$$

$$= 0.47$$

Conclusion

- There is a 47% chance that a randomly selected student is in a relationship or is involved sports.

Using Venn Diagrams



P(not in relationship *and* no sports)

- *P*(**R**^{C} *and* **S**^{C})
- This is the part outside of both circles: 0.53.

P(in a relationship *but* no sports)

- *P*(**R** *and* **S**^{C})
- This is the part in the circle **R** that is outside **S**: 0.22.

P(in a relationship *or* involved in sports *but* not both)

- *P*((**R** *and* **S**^{C}) *or* (**R**^{C} *and* **S**))
- This is the combination of the circles minus the intersection: 0.22 + 0.14 = 0.36

Relationships, Sports, and Independence

33% in a relationship, 25% involved in sports, 11% both

- Are being in a relationship and being involved in sports independent?
  - *P*(relationship) = 0.33
  - *P*(sports) = 0.25
  - *P*(relationship *and* sports) = 0.11
  - 0.33 × 0.25 = 0.0825 ≠ 0.11
  - No, they are **dependent**.
- Are they disjoint?
  - *P*(relationship *and* sports) = 0.11 ≠ 0
  - No, they are **not disjoint**.

Conditional probability

Probability of **B** *Given* **A**:

$$P(\mathbf{B} \mid \mathbf{A}) = \frac{P(\mathbf{A} \text{ and } \mathbf{B})}{P(\mathbf{A})}$$

Events

DEFINITIONS	OR ADDITION more general → P $P(\mathbf{A} \text{ or } \mathbf{B}) =$	AND MULTIPLICATION more restrictive → P $P(\mathbf{A} \text{ and } \mathbf{B}) =$
Disjoint dependent (mutually exclusive) $P(\mathbf{A} \text{ and } \mathbf{B}) = 0$	Addition rule $P(\mathbf{A}) + P(\mathbf{B})$ because $P(\mathbf{A} \text{ and } \mathbf{B}) = 0$	NA
Overlapping independent $P(\mathbf{A} \text{ and } \mathbf{B}) = P(\mathbf{A}) \times P(\mathbf{B})$ $P(\mathbf{B} \mid \mathbf{A}) = P(\mathbf{B})$ $P(\mathbf{A} \mid \mathbf{B}) = P(\mathbf{A})$	General addition rule $P(\mathbf{A}) + P(\mathbf{B}) - P(\mathbf{A} \text{ and } \mathbf{B})$ If specified as exclusionary: $P(\mathbf{A}) + P(\mathbf{B}) - 2P(\mathbf{A} \text{ and } \mathbf{B})$	Multiplication rule $P(\mathbf{A}) \times P(\mathbf{B})$ because $P(\mathbf{B} \mid \mathbf{A}) = P(\mathbf{B})$ $P(\mathbf{A} \mid \mathbf{B}) = P(\mathbf{A})$
Overlapping dependent $P(\mathbf{A} \text{ and } \mathbf{B}) \neq P(\mathbf{A}) \times P(\mathbf{B})$ $P(\mathbf{B} \mid \mathbf{A}) \neq P(\mathbf{B})$ $P(\mathbf{A} \mid \mathbf{B}) \neq P(\mathbf{A})$	General addition rule $P(\mathbf{A}) + P(\mathbf{B}) - P(\mathbf{A} \text{ and } \mathbf{B})$ If specified as exclusionary: $P(\mathbf{A}) + P(\mathbf{B}) - 2P(\mathbf{A} \text{ and } \mathbf{B})$	General multiplication rule $P(\mathbf{A}) \times P(\mathbf{B} \mid \mathbf{A})$ $P(\mathbf{B}) \times P(\mathbf{A} \mid \mathbf{B})$

Overview

- Till now we have mostly dealt with probabilities of single events
- Now: is an outcome (”Result A”) **typical** or **unusual**…
- … *compared* with similar outcomes
- We are about to start a long journey to try to answer this question!
- What we want: a probability model (or probability distribution) for our outcomes --> compare ”Result A” with other results
- How do we build it?
- We use the concept of random variable

Chapter 15

Random Variables

15.1

Center: The Expected Value

Random variables

A random variable, *X*, is a variable whose possible values are the numerical outcomes of a random phenomenon.

A random variable assumes a value based on the outcome of such random event.

- We use a capital letter, like *X*, to denote a random variable.
- A particular value of a random variable will be denoted with the corresponding lower case letter, in this case *x*.

Discrete and continuous random variables

There are two types of random variables:

A discrete random variable is one which may take on only a countable number of distinct values such as 0,1,2,3,4

- Example: Number of children in a family: 0,1,2,3,4,5,6,7…?

A continuous random variables can take any numeric value within a range of values.

- Example: Cost of books per term: Any number from $0 to $400?

Discrete random variables

Usually (but not necessarily) counts.

If a random variable can take only a finite number of distinct values, then it must be discrete.

Examples:

- the number of credit hours per term
- the Friday night attendance at a cinema
- the number of patients in a doctor's surgery
- the number of defective light bulbs in a box of ten

Discrete probability distribution

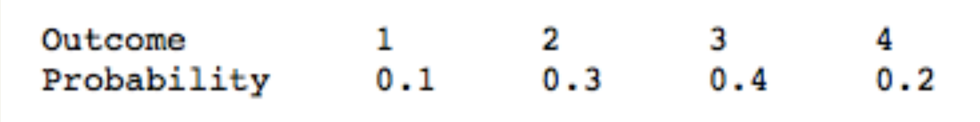
The probability distribution of a discrete random variable is a list of probabilities associated with each of its possible values.

It is also sometimes called the probability model or probability function.

Discrete probability distribution example

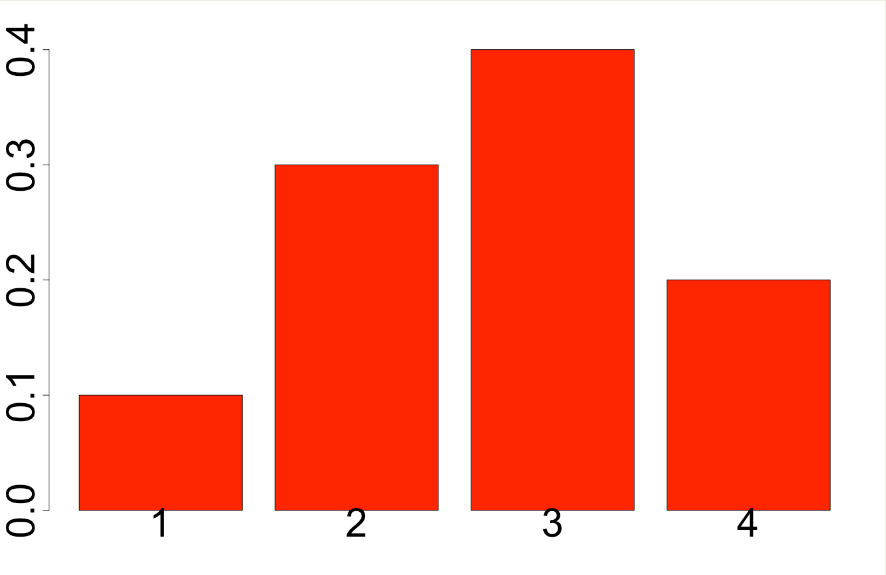
Suppose a discrete variable X can take the values 1, 2, 3, or 4.

The probabilities associated with each disjoint outcome are described by the following table:

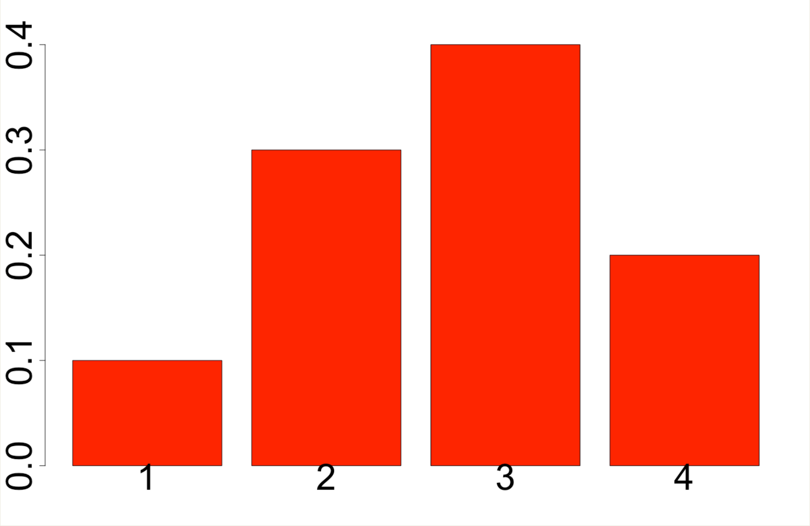


Outcome	1	2	3	4
Probability	0.1	0.3	0.4	0.2

```
data <- c(0.1,0.3,0.4,0.2)
barplot(data,
        names.arg=c("1", "2", "3", "4"),
        cex.axis = 4,
        cex.names = 4,
        col=("red"))
```



Example (cont.)



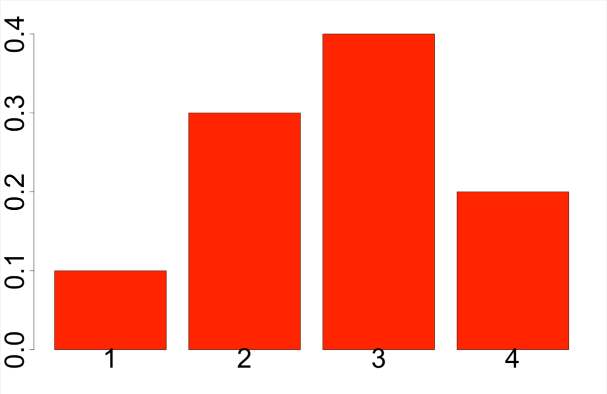
What's the probability that *X* is equal to 2 or 3?

It is the sum of the two probabilities:
P(X = 2 or X = 3) = *P(X = 2)* + *P(X = 3)* = 0.3 + 0.4 = 0.7

What's the probability that *X* is greater than 1?

It is equal to 1 - *P(X = 1)* = 1 - 0.1 = 0.9, by the complement rule.

Expected Value: Center



A probability model
for a random variable consists of:

- The collection of all possible values of a random variable AND
- the probabilities that the values occur.

Of particular interest is the value we **expect** a random variable to take on, notated *μ* (for **population mean**) or *E(X)* for expected value.

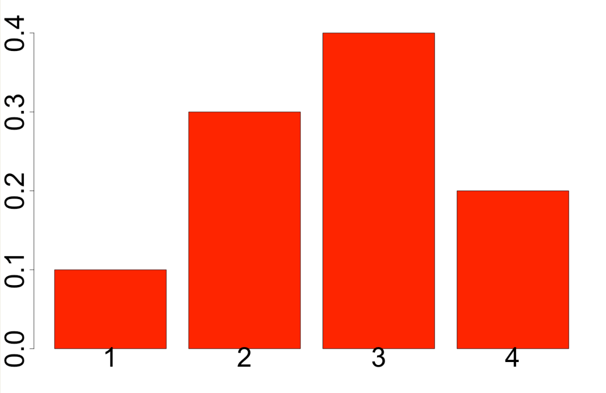
When we looked at a dataset, we calculated the **sample mean** ($\overline{X}$) as average.

Formulas for population mean and sample mean are different…

What is the Population Mean?

The expected value of a (discrete) random variable can be found by summing the products of each possible value by the probability that it occurs:

$$\mu = E(X) = \sum x \cdot P(x)$$



E(X) = (1*0.1)+(2*0.3)+(3*0.4)+(4*0.2)

We sum the values weighted by their probabilities (or relative frequencies)

The **sample mean** is an average. It varies from sample to sample.

The **population mean** is a sum of values weighted by their probabilities. It does not vary

Expected Value and Sample Mean

Imagine you are have a fair die. Knowing all the possible outcomes and their probabilities, we can calculate E(X):

E(X)= 1*1/6+2*1/6+...6*1/6 = 3.5
This number is constant.

Now imagine that you roll the die 10 times obtaining:
{5,6,4,5,3,2,1,2,4,6}
The sample mean, for this specific sample, is 3.8.

Imagine that you take another sample; this time the mean is 3.4. You keep taking sample and calculating the mean; ultimately the mean of the means will converge to 3.5.

15.2

Spread: The Standard Deviation

Review: measure of spread for data

For **data**, we calculated the **sample variance** by first computing the deviation from the mean, then squaring it and averaging it.

$$s^2 = \frac{\sum (y - \bar{y})^2}{n-1}$$

The **sample standard deviation** is the square root of the sample variance

→ The standard deviation expresses the average distance form the mean

$$s = \sqrt{\frac{\sum (y - \bar{y})^2}{n-1}}$$

First Center, Now Spread…

The population variance for a random variable is:

$$\sigma^2 = Var(X) = \sum (x - \mu)^2 \cdot P(x)$$

The standard deviation for a random variable is:

$$\sigma = SD(X) = \sqrt{Var(X)}$$

Formulas for population variance and sample variance are different.

15.3

Shifting and Combining Random Variables

More About Means and Variances

Adding or subtracting a constant from data shifts the mean but doesn’t change the variance or standard deviation:

$$E(X \pm c) = E(X) \pm c \quad Var(X \pm c) = Var(X)$$

- Example: Consider everyone in a company receiving a $5000 increase in salary.

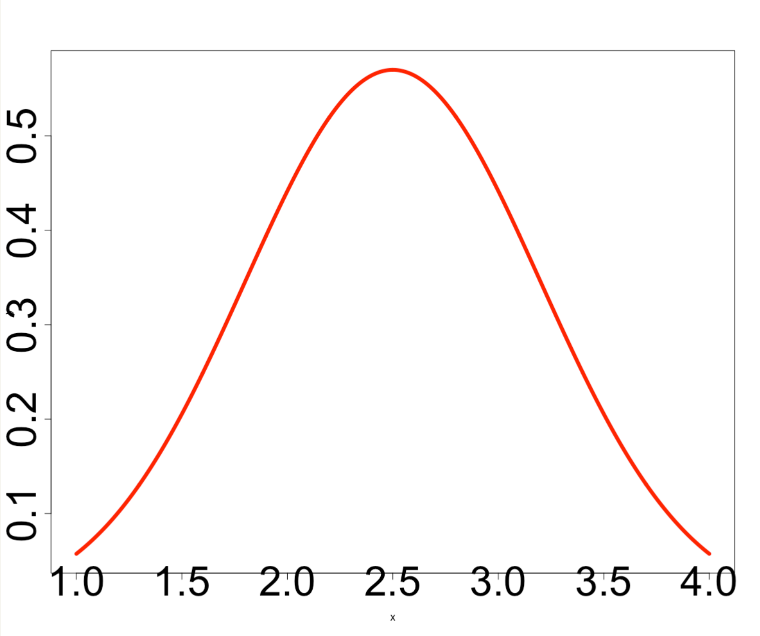
15.4

Continuous Random Variables

Continuous Random Variables

Random variables that can take on any value in a range of values are called continuous random variables.

```
x   <- seq(1,4,length=1000)
y   <- dnorm(x,mean=2.5, sd=0.7)
plot(x,y,
      type="l",
      lwd=6,
      cex.axis=4,
      col=("red"))
```

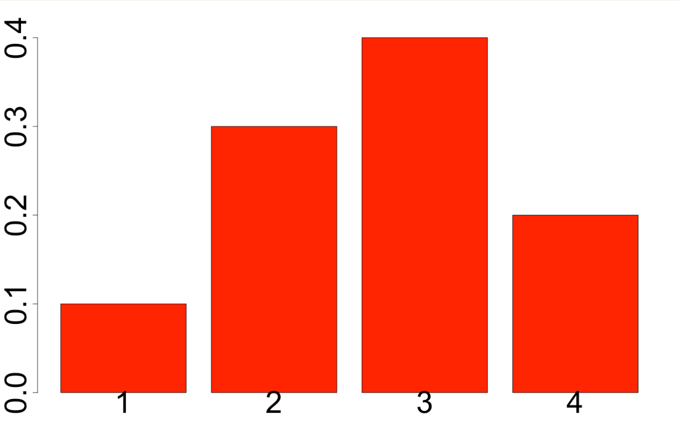


Probability as an area

Random variables that can take on **<u>any value</u>** in a range of values are called continuous random variables.

The fact that we are talking about <u>any value</u> has some <u>consequences</u>.

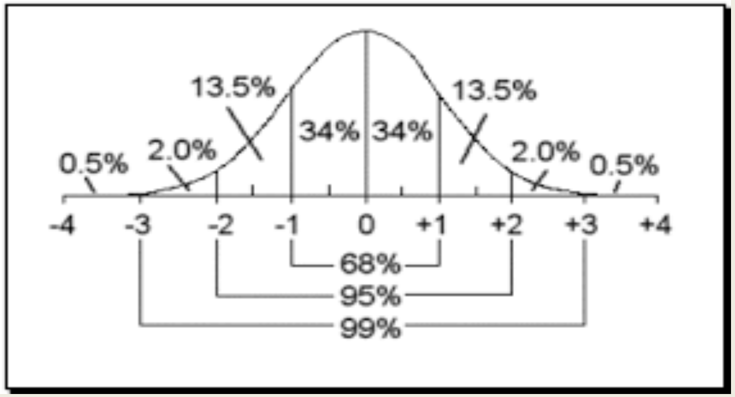
Let’s take a step back, looking at a discrete distribution again.



Which is bigger, *P(X = 1 or X = 2)* or *P(X = 3 or X = 4)* ?
We can think about probability as an area.

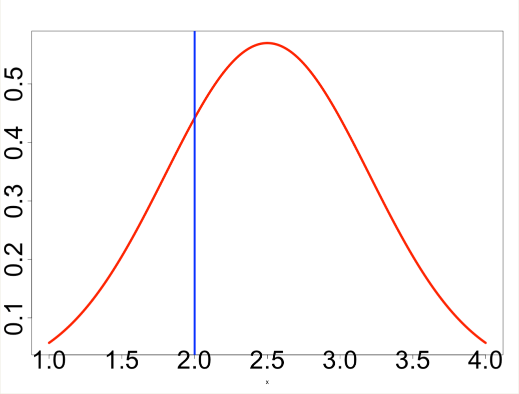
Continuous Random Variables

Random variables that can take on **<u>any value</u>** in a range of values are called continuous random variables.



We can think about probability as an area under a curve…

Any value → any *single value* won't have a probability!



But continuous random variables do have means (expected values) and variances.

What Can Go Wrong?

Don’t assume everything’s Normal.

- You must *Think* about whether the Normality Assumption is justified.

What Can Go Wrong?

Probability models are still just models.

- Models can be useful, but *they are not reality*.
- Question probabilities as you would data, and think about the assumptions behind your models.

If the model is wrong, so is everything else.

Chapter 16

Probability Models

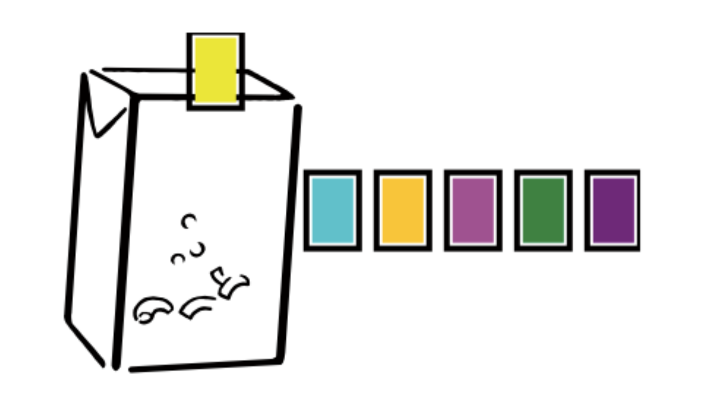
Overview

- Is a result typical or unusual?
- How do we build a probability model for our data?
- Random variable
- There are many different distribution we can define and use, discrete or continuous.
- Examples:
  - Discrete: geometric or binomial
  - Continuous: normal, uniform, exponential
- Focus on binomial and normal distributions
- Before we do that, we need to define a basic "experiment", a Bernoulli trial

16.1

Bernoulli Trials

Bernoulli Trials



You're collecting cards form cereal boxes. You don't care about completing the whole card collection, but you've just *got* to have the blue card.
How many boxes do you expect you'll have to open before you find the blue card?

We'll keep the assumption that cards are distributed at random and we'll trust the manufacturer's claim that 20% of the cards are blue.
So, when you open the box, the probability that you succeed in finding blue card is 0.20. Now we'll call the act of opening *each* box a trial, and note that:

Bernoulli Trials

- There are only two possible outcomes (called *success* and *failure*) on each trial. Either you get the blue card (success), or you don’t (failure).
- The probability of success, denoted *p*, is the same on every trial. Here *p* = 0.20.
- The trials are independent. Finding the blue card in the first box does not change what might happen when you reach for the next box.

Situations like this occur often and are called **Bernoulli trials**.

Bernoulli Trials

- Since the blue card is in 20% of the boxes, we expect we will need to open, on average, five boxes:

$$E(Y) = \frac{1}{0.2} = 5 \text{ boxes}$$

Not easy to prove!

Common Examples of Bernoulli Trials

- tossing a coin
- looking for defective products rolling off an assembly line
- shooting free throws in a basketball game

16.2

The Geometric Model

The Geometric Model: How long it takes?

- We want to model how long it will take to achieve the first success in a series of Bernoulli trials.

- The model that tells us this probability is called the **Geometric probability model.**

Geometric Probability Model for Bernoulli Trials: GEOM(*n*, *p*)

p = probability of success
q = 1 – *p* = probability of failure
X = number of trials until the first success occur

$$P(X=x)=q^{x-1}p$$

Expected value: $E(X)=\mu=\dfrac{1}{p}$

Standard deviation: $\sigma=\sqrt{\dfrac{q}{p^2}}$

Independence – The 10% Condition

One of the important requirements for Bernoulli trials is that the trials be independent.

The 10% Condition: Bernoulli trials must be independent. If that assumption is violated, it is still okay to proceed as long as the sample is smaller than 10% of the population.

Universal Blood Donors

6% of all people are O^-, **Universal Donors**.

1. If donors line up at random for a blood drive, how many do you expect to examine before you find someone who has O-negative blood?
2. What's the probability that the first O-negative donor found is one of the first four people in line?

Plan: Either success (O^-) or Failure (not O^-)
p = 0.06, independent, 10% condition

Variable: *X* = no. of donors until O^-
Model: *X* is Geom(0.06)

Universal Blood Donors

Mechanics:

Calculate probability that *X* = 1, 2, 3, 4

$$E\left(X\right)=\frac{1}{0.06}\approx 16.7$$

$$P\left(X\le 4\right)=P\left(X=1\right)+P\left(X=2\right)+P\left(X=3\right)+P\left(X=4\right)$$

$$=\left(0.06\right)+\left(0.94\right)\left(0.06\right)+\left(0.94\right)^2\left(0.06\right)+\left(0.94\right)^2\left(0.06\right)$$

$$\approx 21.93$$

Universal Blood Donors

Conclusion:

Blood drives such as this one expect to examine an average of 16.7 people to find a universal donor. About 22% of the time there will be one within the first 4 people in line.

16.3

The Binomial Model

The Binomial Model: How many successes?

- A **Binomial probability model** describes the number of successes in a specified number of trials.

- It takes two parameters to specify this model: the number of trials *n* and the probability of success *p*.

Binomial Probability Model for Bernoulli Trials: BINOM(*n*, *p*)

x = number of trials
p = probability of success
q = 1 – *p* = probability of failure
X = number of successes in *n* trials

$$P\left(X=x\right)={}_{n}C_{x}p^{x}q^{n-x}, \qquad {}_{n}C_{x}=\frac{n!}{x!\left(n-x\right)!}$$

$$\text{Mean: } \mu = np$$

$$\text{Standard deviation: } \sigma = \sqrt{npq}$$

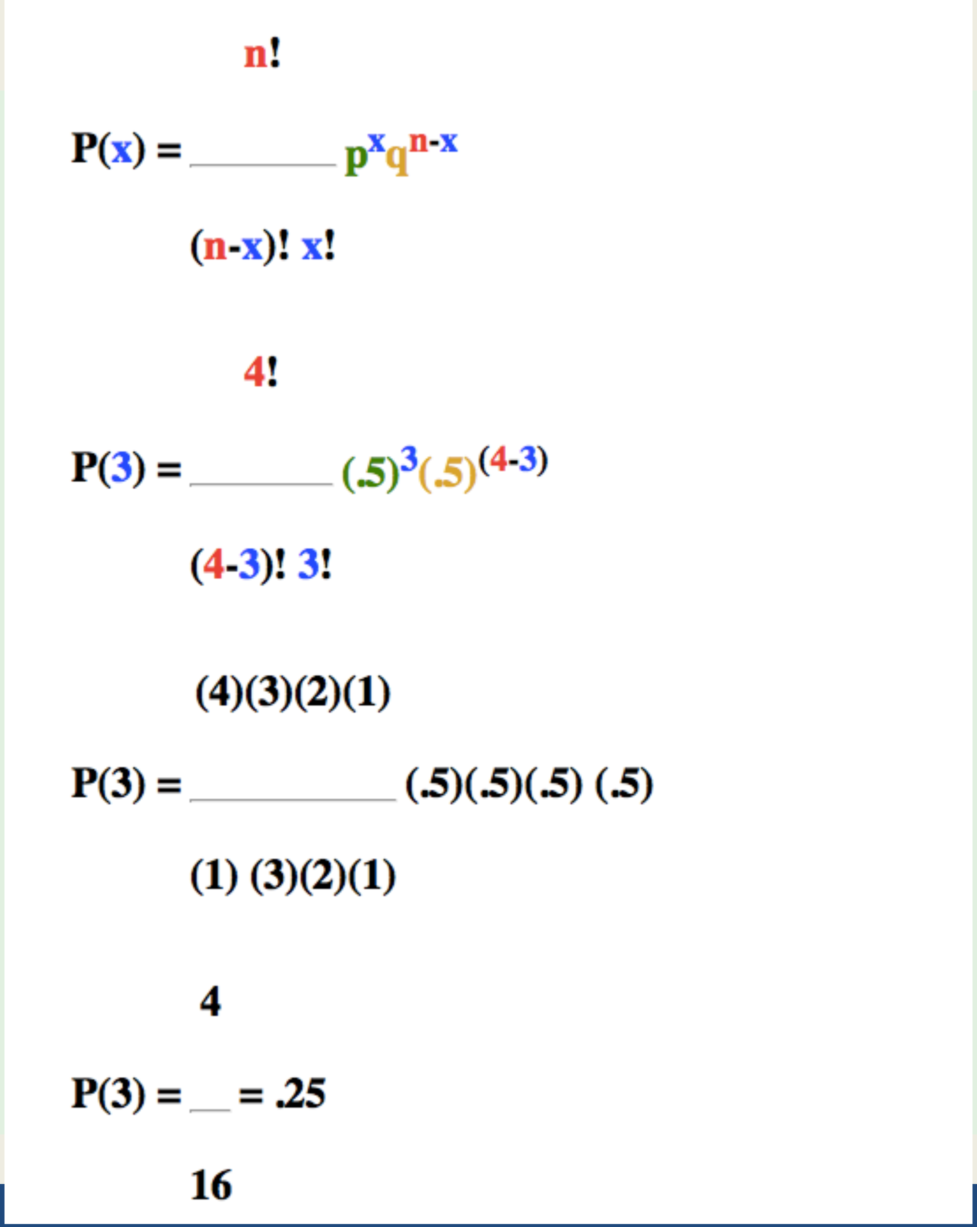
Binomial Probability Model: Example

$n = 4$

$x = 3$

$p = 0.5$

$q = 1 – p = 0.5$



$$P(x) = \frac{n!}{(n-x)!\ x!}\ p^{x}q^{n-x}$$

$$P(3) = \frac{4!}{(4-3)!\ 3!}\ (.5)^{3}(.5)^{(4-3)}$$

$$P(3) = \frac{(4)(3)(2)(1)}{(1)\ (3)(2)(1)}\ (.5)(.5)(.5)\ (.5)$$

$$P(3) = \frac{4}{16} = .25$$

Universal Blood Donors

6% of all people are O^-. If 20 donate at the blood drive, find the mean and sd of the number of universal donors and find the probability that 2 or 3 are O^-.

- Plan: Either success (O^-) or Failure (not O^-)
  p = 0.06, independent, 10% condition
  Therefore binomial

  Variable: X = no. of O^-, n = 20 people

  Model: X is Binom(20, 0.06)

Universal Blood Donors

6% of all people are O^-. If 20 donate at the blood drive, find the mean and sd of # of universal donors and find the probability that 2 or 3 are O^-.

- **Mechanics:**

$$\text{Mean} = np = (20)(0.06) = 1.2$$

$$SD = \sqrt{npq} = \sqrt{(20)(0.06)(0.94)} \approx 1.06$$

$$P(X = 2 \text{ or } 3) = P(X = 2) + P(X = 3)$$

$$= {}_{20}C_2\,(0.06)^2\,(0.94)^{18} + {}_{20}C_3(0.06)^3 + (0.94)^{17}$$

$$= 0.2246 + 0.0860$$

$$= 0.3106$$

Universal Blood Donors

6% of all people are O^-. If 20 donate at the blood drive, find the mean and sd of # of universal donors and find the probability that 2 or 3 are O^-.

- **Conclusion:**
  In Groups of 20 randomly selected blood donors, I expect to find an average of 1.2 universal blood donors with a standard deviation of 1.06. There is about a 31% chance that 2 or 3 of the 20 donors are O^-.

16.4

Approximating the Binomial with the Normal Model

The Trouble with Large Sample Sizes

The Tennessee Red Cross has 32,000 donors and needs at least 1850 that are O$^-$. Will they run out?

- The computations involve ridiculously large numbers.
- "At least" requires $P(X = 1850)$, $P(X = 1851)$, all the way up to $P(X = 32{,}000)$.
- Mean = np = 1920
  $SD = \sqrt{npq} \approx 42.48$
-

The Solution for Large Sample Sizes

The Tennessee Red Cross has 32,000 donors and needs at least 1850 that are O⁻. Will they run out (less than)?

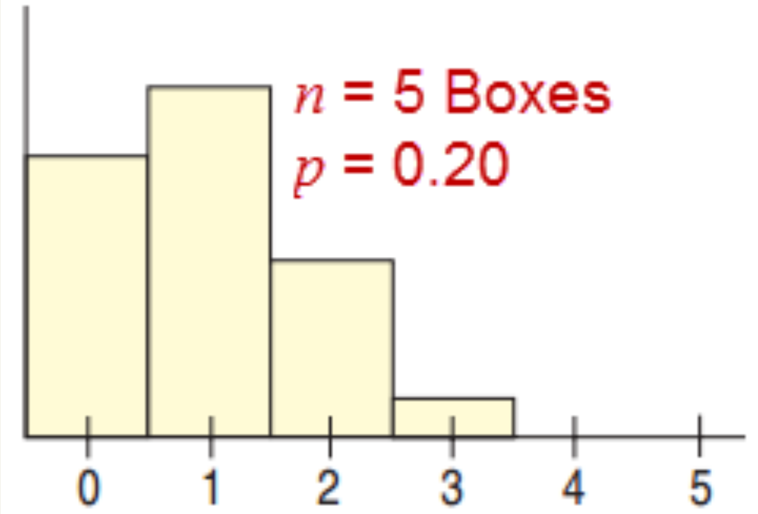
- Mean = np = 1920 $\quad SD = \sqrt{npq} \approx 42.48$
- The **normal model** with the same mean and standard deviation is a very good approximation.
- $P(X < 1850) \approx P\left(z < \dfrac{1850 - 1920}{42.48}\right) \approx \mathrm{P}(z < -1.65) \approx 0.05$
- There is about a 5% chance that they will run out.

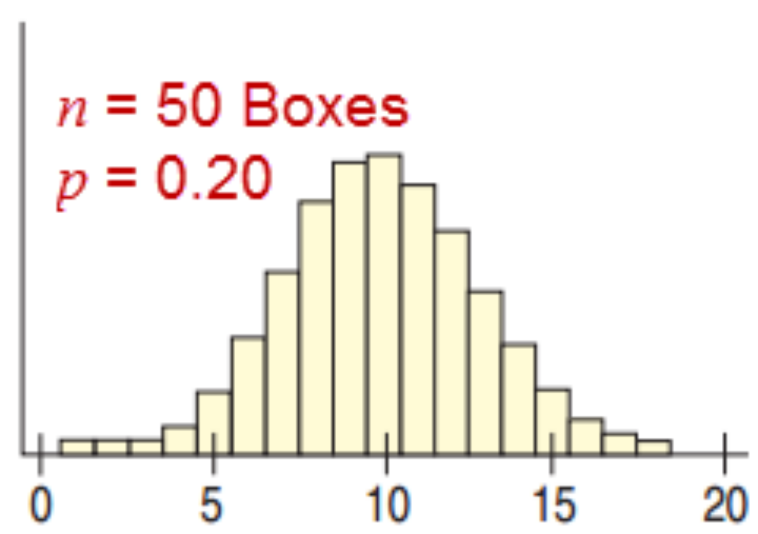
How Large is “Large Enough”

The Success/Failure Condition

- A Binomial is approximately Normal if we expect at least 10 successes and 10 failures.
  np ≥ 10, *nq* ≥ 10

- This comes from the binomial being skewed for a small number of successes or failures expected.





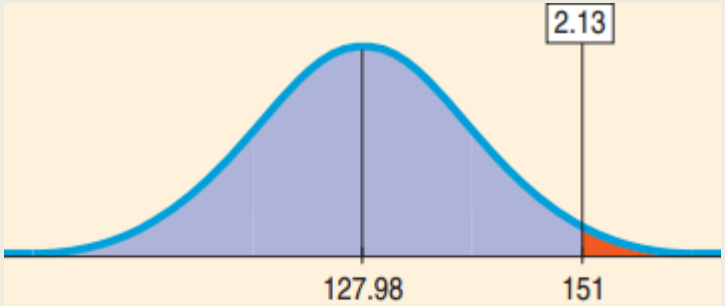
Example: Spam and the Normal Approximation to the Binomial

Only 151 of 1422 emails got through your spam filter. Might the filter be too aggressive?

- What is the **probability** that **no more than 151** of the emails are **real** messages, **if probability** of being a real message is **0.09**?
- Emails arrive randomly and independently
- These emails represent less than 10% of all emails.
- $np = (1422)(0.09) = 127.98 \geq 10$
- $nq = (1422)(0.91) = 1294.02 \geq 10$
- Yes, the **Normal model** is a good approximation.

Example: Spam and the Normal Approximation to the Binomial

- What is the probability that no more than 151 of the emails are real messages?
- $\mu = np = 127.98$
- $\sigma = \sqrt{npq} \approx 10.79$



- $P(X \leq 151) \approx P\left(z \leq \dfrac{151\text{-}127.98}{10.79}\right) \approx P(z \leq 2.13) \approx 0.9834$
- There is over a 98% chance that no more than 151 of them were real messages. The filter may be working.

What Can Go Wrong?

Be sure you have Bernoulli Trials.

- two possible outcome, constant probability of success, independence

Don’t confuse Geometric and Binomial models.

- Geometric: repeat until get first success
- Binomial: counting success in specified number of trials

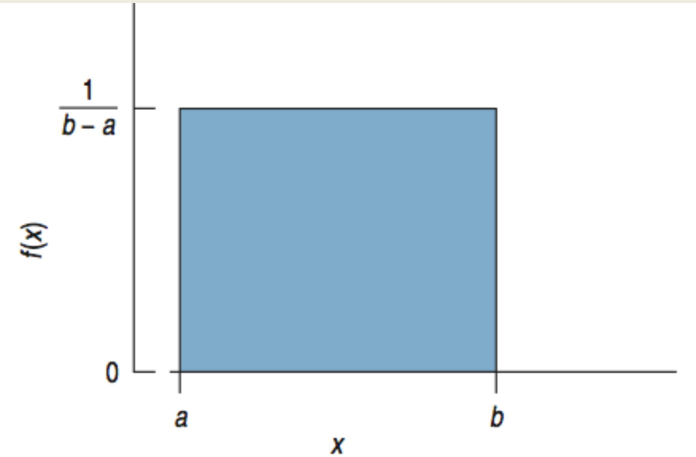
Don’t use the normal approximation for small *n*.

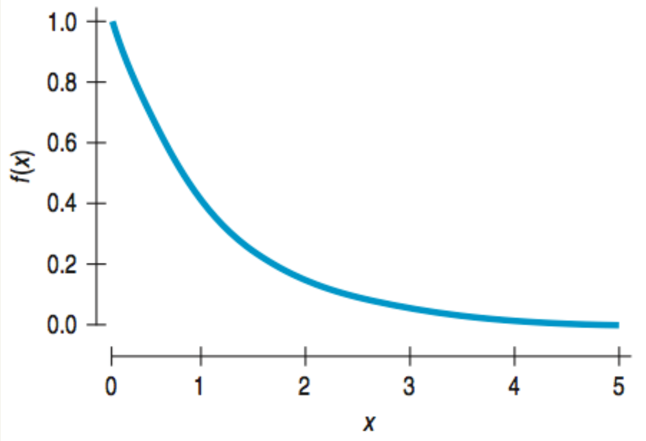
- $np > 10$, $nq > 10$

What Can Go Wrong?

Don’t assume everything is Normal. There are many other continuous probability models: Uniform, Exponential, Chi-Square etc.

- Check there the distribution is not skewed and that there is a bell shape.


$\frac{1}{b-a}$
f(x)
a
b
x



1.0
0.8
0.6
0.4
0.2
0.0
f(x)
0
1
2
3
4
5
x