# Math 32, Lecture #5

Prof. Suzanne S. Sindi
ssindi@ucmerced.edu
Prof. Lihong Zhao
lzhao33@ucmerced.edu
Thursday, September 10, 2020

# Today’s lecture

- More neat-o reasons why studying probability and statistics is awesome.
- Conditional Probability Continued
- Bayes’ Rule

# Why Study Probability?
# Um…have you seen the news…

Jeff Kao Follow
Data Scientist. Author of WaPo cited Net Neutrality study https://tinyurl.com/fcc-nn. Alum at Metis SF, Columbia Law and UWaterloo
Nov 23, 2017 · 10 min read

**More than a Million Pro-Repeal Net Neutrality Comments were Likely Faked**

I used natural language processing techniques to analyze net neutrality comments submitted to the FCC from April-October 2017, and the results were disturbing.

"In the matter of restoring Internet freedom. I'd like to recommend the commission to undo The Obama/Wheeler power grab to control Internet access. Americans, as opposed to Washington bureaucrats, deserve to enjoy the services they desire. The Obama/Wheeler power grab to control Internet access is a distortion of the open Internet. It ended a hands-off policy that worked exceptionally successfully

CNN
**How Covid-19 misinformation circulates on social media**
CNN's Don Lemon speaks to medical analyst Dr. Seema Yasmin and Dr. Craig Spencer about the rampant coronavirus misinformation ...
1 week ago

BBC News
**'Hundreds dead' because of Covid-19 misinformation**
A study says at least 800 people may have died globally because of coronavirus-related misinformation.
1 month ago


USA TODAY
**Your social feed is crowded with misinformation about coronavirus. Here's how to spot it.**
I took a peek at one of my childhood friends' Facebook feeds and spotted 16-times he's shared a debunked conspiracy theory, cheapfake video ...
3 weeks ago


We live in a highly connected digital era and understanding the flow of data will be an increasing part of our world. Better understanding will mean you are better engaged in the world and (if you are interested) can be your paycheck.

# Conditional Probability

Probability of B given A

$P(B \mid A) = P(A \cap B)/P(A)$

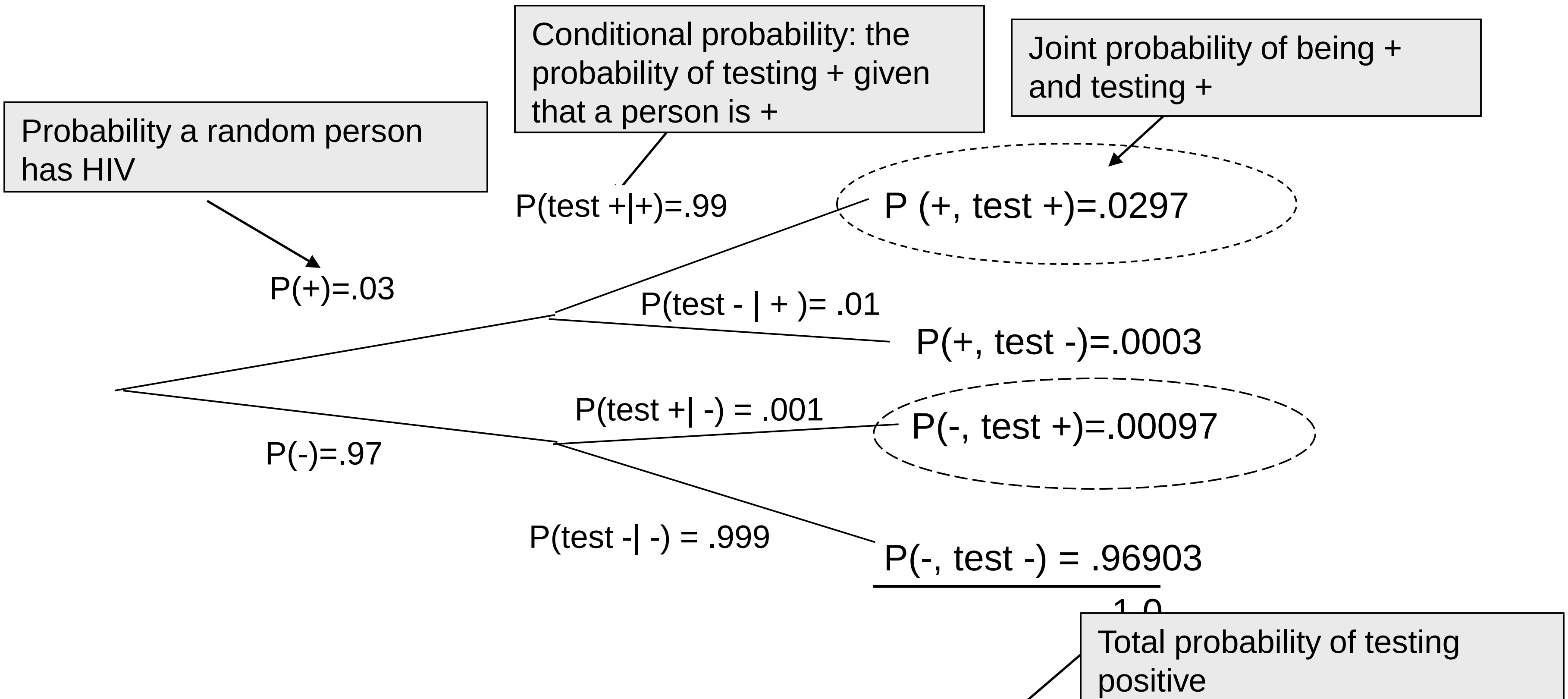
# A Conceptual Problem

If HIV has a **prevalence of 3%** in San Francisco, and a particular HIV test has a **false positive probability of .001** and a **false negative probability of .01**, what is the probability that a random person selected off the street will test positive?

# False Positive and False Negative

- **False Positive Probability**:
  P(Test Result Positive | No Disease)
- **False Negative Probability**:
  P(Test is Negative | Disease)
- We will use +/- for both test and disease result

# What is the probability that a random person selected off the street will test positive?




$$\therefore P(\text{test }+)=.0297+.00097=.03067$$

$P(+\text{ disease and test}+) \neq P(+\text{ disease}) \times P(\text{test}+)$

$.0297 \neq .03 * .03067\ (=.00092)$

$\therefore$ Dependent!

# Law of total probability

$$P(\text{test} +) = P(\text{test} + | \text{HIV+})P(\text{HIV+}) + P(\text{test} + | \text{HIV-})P(\text{HIV-})$$

One of these has to be true (mutually exclusive, collectively exhaustive). They sum to 1.0.

$$P(\text{test} +) = .99(.03) + .001(.97)$$

# Law of total probability

- Formal Rule: The probability for event A (also called the marginal probability)

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) \cdots + P(A|B_n)P(B_n)$$

- As long as:

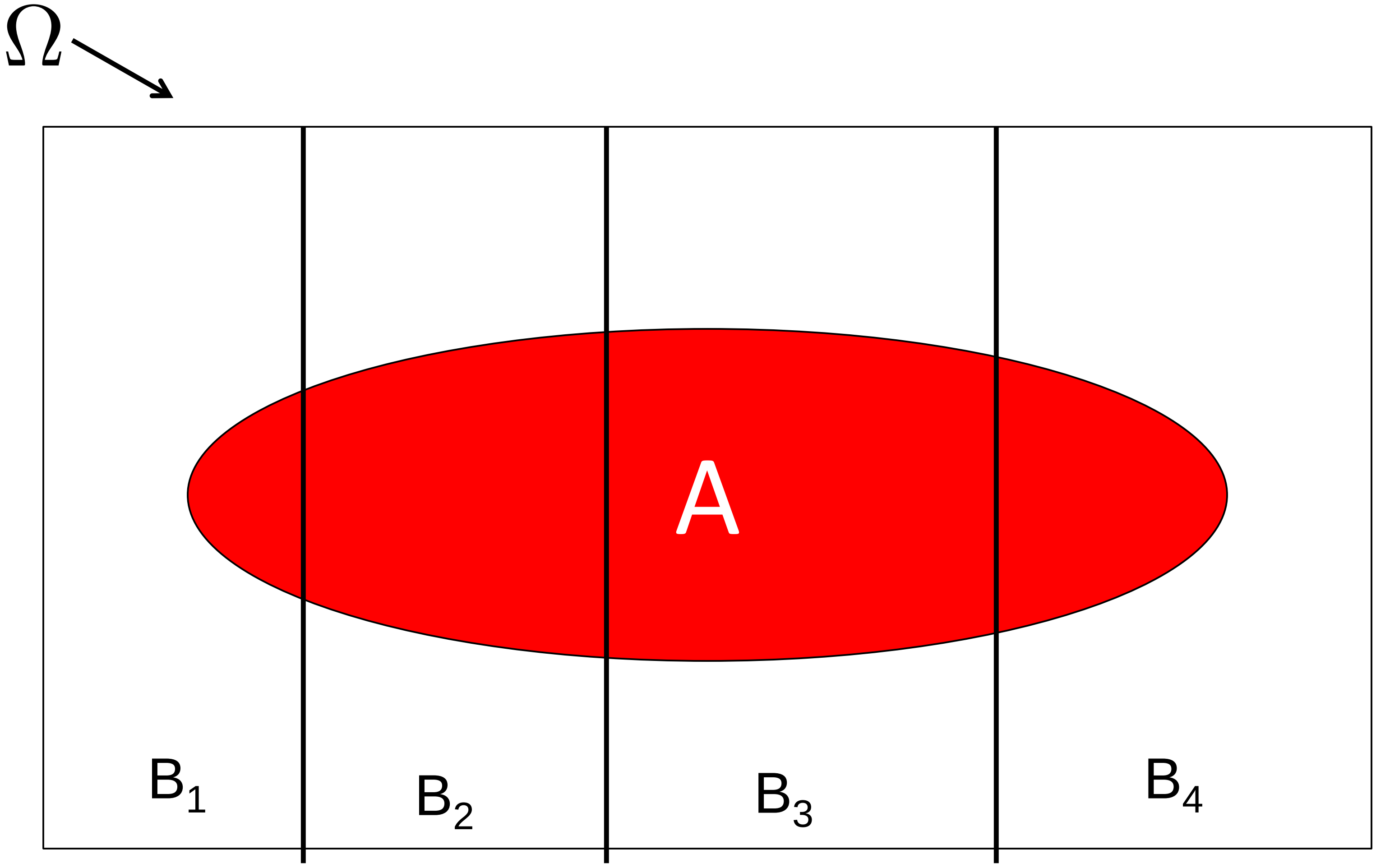
$$\sum_{i=1}^{n} P(B_i) = 1$$

Take up entire sample space.

$$B_i \cap B_j = \emptyset$$

Events are Disjoint

# Law of total probability




$$P(A) = P(A|B_1) \times P(B_1) + P(A|B_2) \times P(B_2) + P(A|B_3) \times P(B_3) + P(A|B_4) \times P(B_4)$$

# Insurance Problem

An insurance company believes that drivers can be divided into two classes—those that are of **high risk** and those that are of **low risk**.

Their statistics show that a high-risk driver will have an accident at some time within a year with probability .4, but this probability is only .1 for low risk drivers.

a) Assume 20% of the drivers are high-risk, what is the probability that a new policy holder will have an accident within a year of purchasing a policy?
b) If a new policy holder has an accident within a year of purchasing a policy, what is the probability that he is a high-risk type driver?

(a) Assuming that 20% of the drivers are of high-risk, what is the probability that a new policy holder will have an accident within a year of purchasing a policy?

**Answer:**

Use law of total probability:

P(accident)

= P(accident | high risk) x P(high risk) + P(accident | low risk) x P(low risk)

= .40(.20) + .10(.80) = .08 + .08 = .16

# Bayes Rule

- To answer (b), it's helpful to know Bayes Rule. (more details after this example)

$$P(A \cap B) = P(B|A)P(A)$$

$$P(A \cap B) = P(A|B)P(B)$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

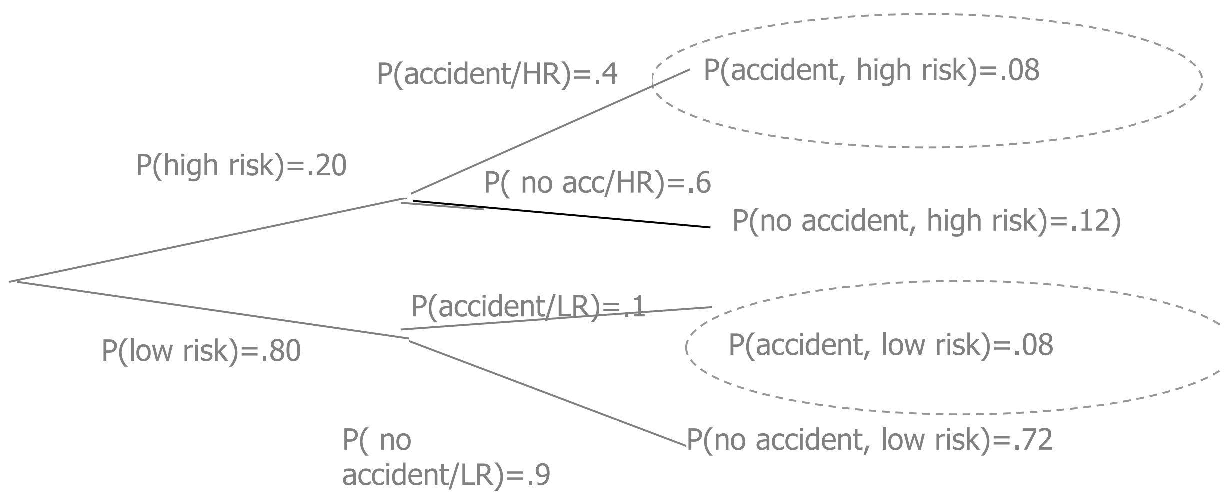
(b) If a new policy holder has an accident within a year of purchasing a policy, what is the probability that he is a high-risk type driver?

**Answer:**

$$\text{P(high-risk | accident)} = \frac{\text{P(accident | high risk) x P(high risk)}}{\text{P(accident)}}$$

$$= .40(.20)/.16 \ = 50\%$$

Or use tree:

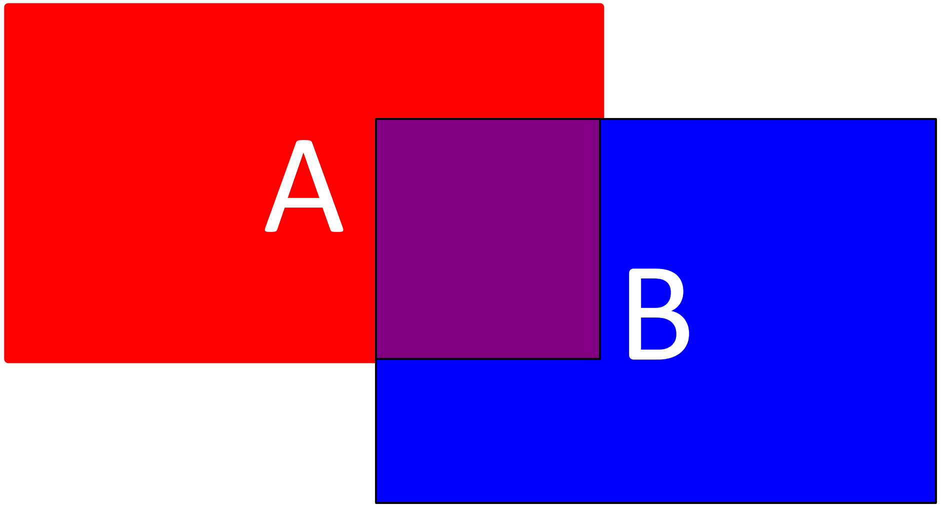



P(high risk | accident)=.08/(0.08 + 0.08) =50%

# Bayes' Rule: derivation

Let A and B be two events with P(B) ≠ 0. The conditional probability of A given B is:

$$\mathbf{P}(\mathbf{A}|\mathbf{B}) = \frac{\mathbf{P}(\mathbf{A}\cap\mathbf{B})}{\mathbf{P}(\mathbf{B})}$$




The idea: if we are given that the event B occurred, the relevant sample space is reduced to B  (i.e., P(B)=1 because we know B is true) and conditional probability becomes a probability measure on B.

# Bayes' Rule
# + Law of Total Probability:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}$$

# Bayes' Rule:

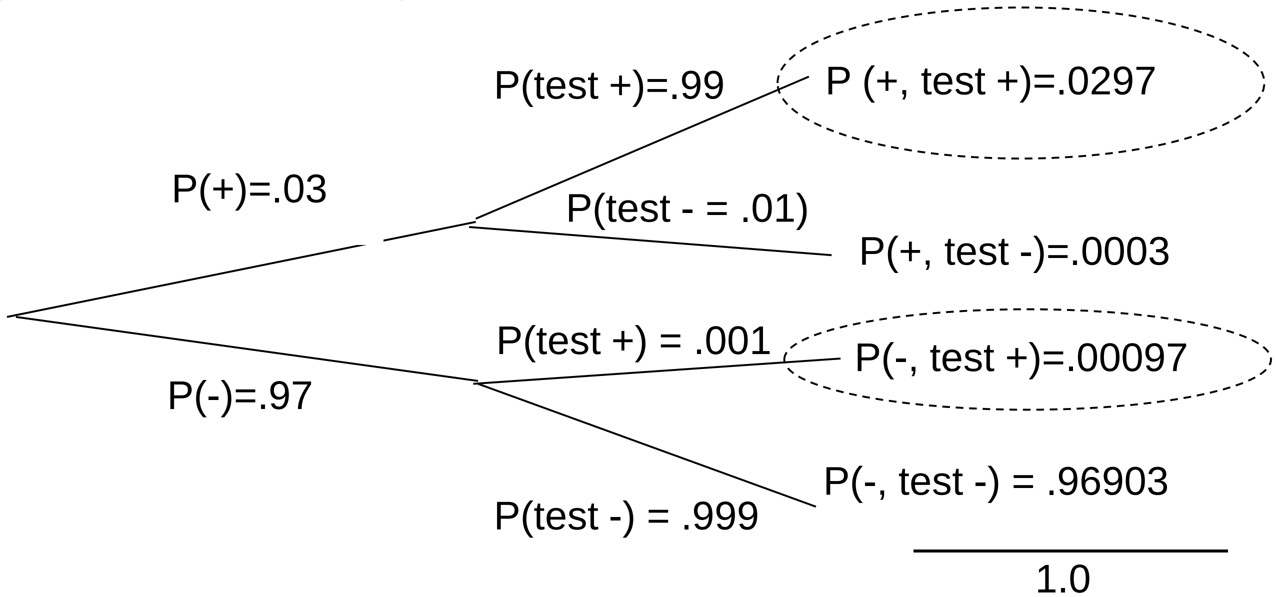
- Why do we care??
- Why is Bayes' Rule useful??
- It turns out that sometimes it is very useful to be able to "flip" conditional probabilities. That is, we may know the P(A|B), but P(B|A) may not be obvious.  Revisiting an example from before will show this…

# In-Class Exercise

- If HIV has a prevalence of 3% in San Francisco, and a particular HIV test has a false positive rate of .001 and a false negative rate of .01, **<u>what is the probability that a random person who tests positive is actually infected (also known as “positive predictive value”)</u>**?

# Answer: using probability tree

Q: What is the probability that a random person who tests positive is actually infected?




A positive test places one on either of the two "test +" branches. But only the top branch also fulfills the event "true infection."

$$P(\mathrm{HIV}+|\mathrm{test}+) = \frac{P(\mathrm{test}+\cap\mathrm{HIV}+)}{P(\mathrm{test}+)} = \frac{0.0297}{0.0297+0.0097} = 96.8\%$$

# Answer: using Bayes' rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^C)P(A^C)}$$

A: Probability HIV+
B: Test Result is Positive

P(A) = 0.03, P($A^c$) = 0.97

P(B|A) = 0.99, P(B|$A^c$) =0.001

$$P(\mathrm{HIV}+|\mathrm{test}+) = \frac{0.99(0.03)}{0.99(0.03)+0.001(0.97)} = 96.8\%$$

# Another Example

- Three jars contain colored balls as described in the table below.
  - One jar is chosen at random and a ball is selected. If the ball is red, what is the probability that it came from the $2^{nd}$ jar?

| Jar # | Red | White | Blue |
|---|---|---|---|
| 1 | 3 | 4 | 1 |
| 2 | 1 | 2 | 3 |
| 3 | 4 | 3 | 2 |

# Describe Our Event Space

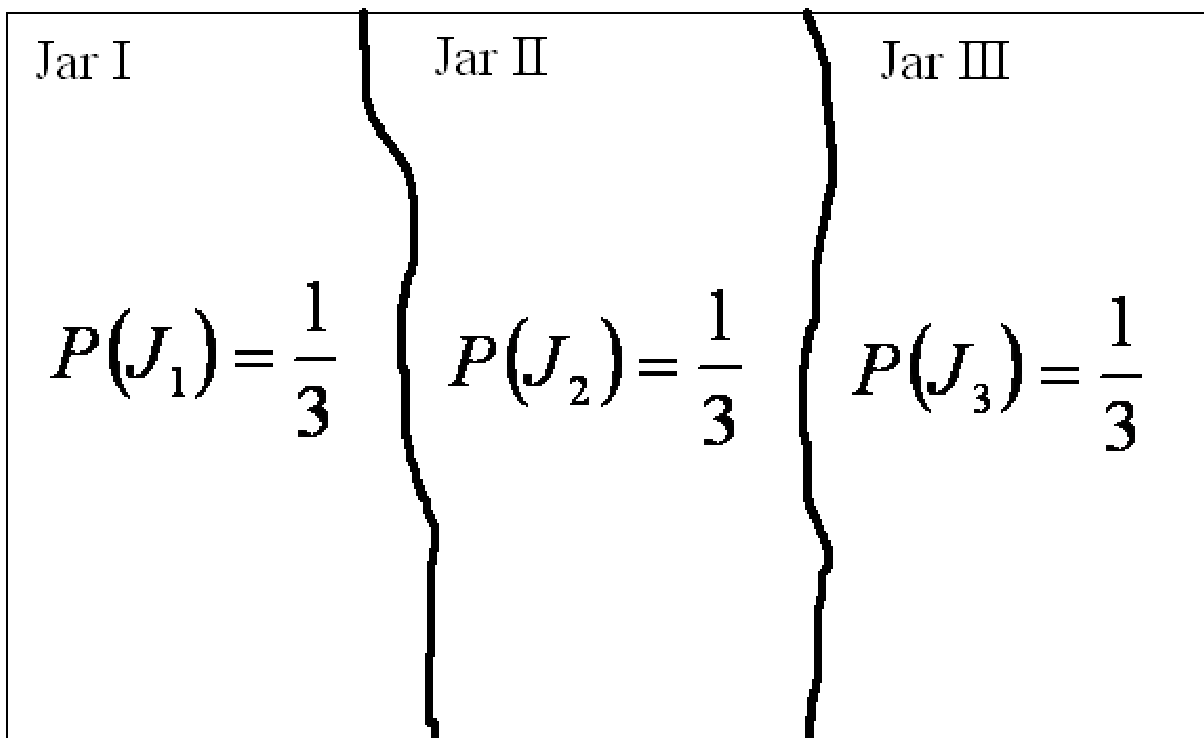
- We will define the following events:
  - $J_1$ is the event that *first* jar is chosen
  - $J_2$ is the event that *second* jar is chosen
  - $J_3$ is the event that *third* jar is chosen
  - $R$ is the event that a *red* ball is selected

# Partitioning Sample Space

- The events $J_1$, $J_2$, and $J_3$ mutually exclusive.(WHY?)
  - *You can't chose two different jars at the same time*
- Because of this, our sample space has been divided or *partitioned* along these three events

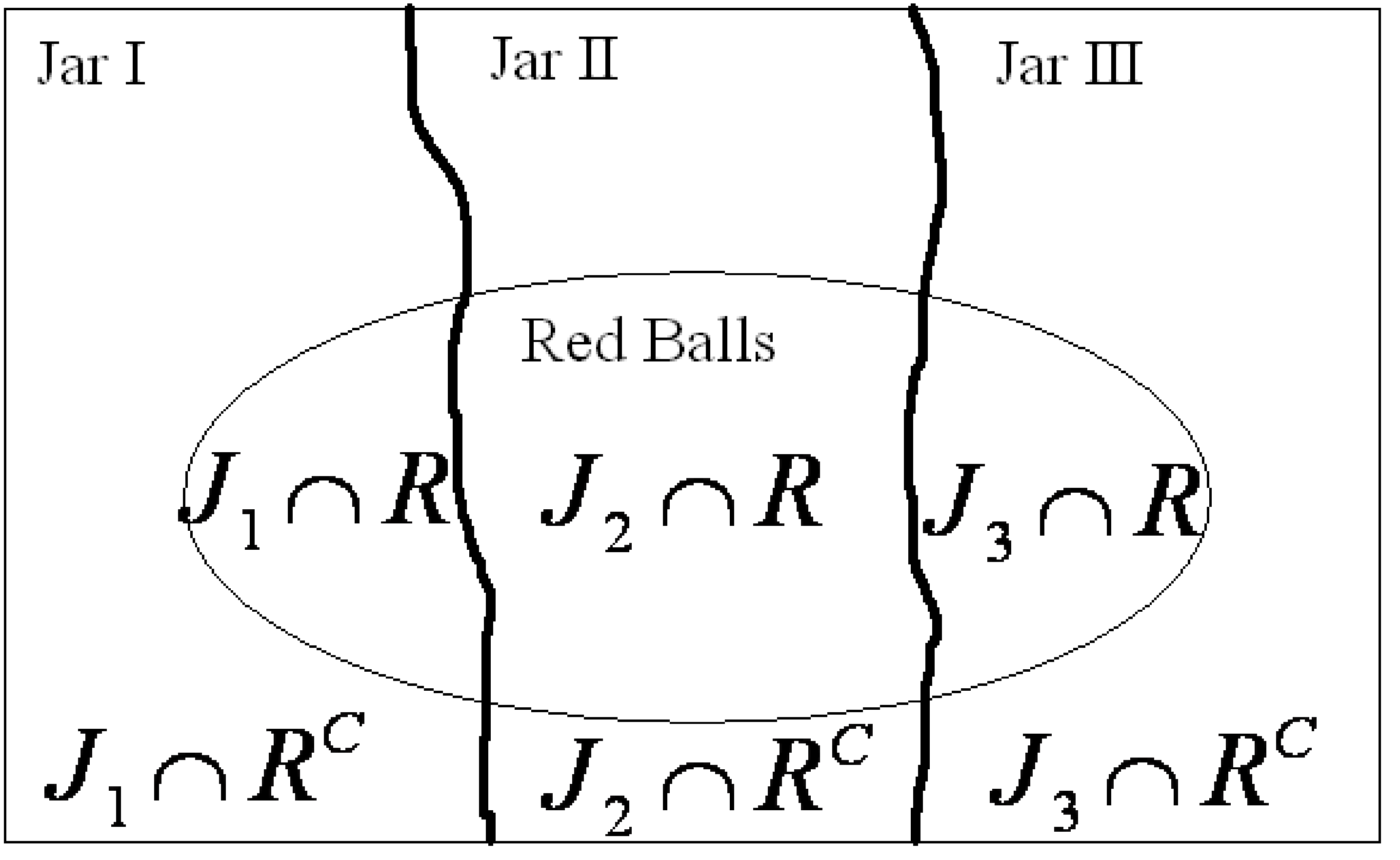
# Sample Space: {Jar x Color}

- $\{J_1,\text{red}\},\{J_1,\text{white}\},\{J_1,\text{blue}\}$
- $\{J_2,\text{red}\},\{J_2,\text{white}\},\{J_2,\text{blue}\}$
- $\{J_3,\text{red}\},\{J_3,\text{white}\},\{J_3,\text{blue}\}$



# Sample Space

- All of these three jars have red balls, so the set of red balls overlaps all three sets of our partition



# Finding Probabilities

- What are the probabilities for each of the events in our sample space?
- How do we find them?

$$P(A \cap B) = P(A \mid B)P(B)$$

# Computing Probabilities

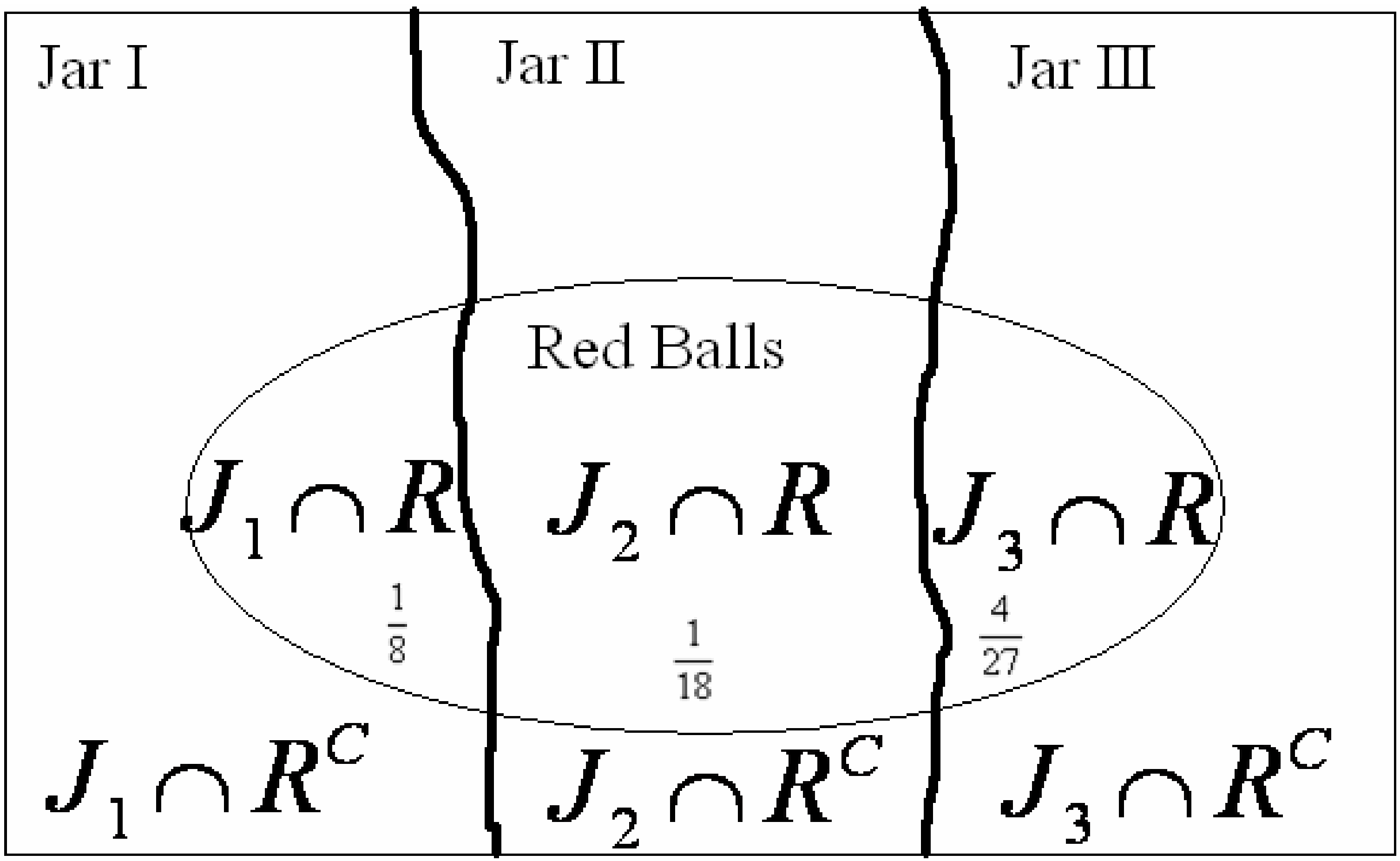
| Jar # | Red | White | Blue |
|---|---|---|---|
| 1 | 3 | 4 | 1 |
| 2 | 1 | 2 | 3 |
| 3 | 4 | 3 | 2 |

$$P(J_1 \cap R) = P(R \mid J_1)P(J_1) = \frac{3}{8} \cdot \frac{1}{3} = \frac{1}{8}$$

$$P(J_2 \cap R) = P(R \mid J_2)P(J_2) = \frac{1}{6} \cdot \frac{1}{3} = \frac{1}{18}$$

$$P(J_3 \cap R) = P(R \mid J_3)P(J_3) = \frac{4}{9} \cdot \frac{1}{3} = \frac{4}{27}$$

# Update Partitioned Sample Space



# Where are we going with this?

- Our original problem was:
  - One jar is chosen at random and a ball is selected. If the ball is red, what is the probability that it came from the 2nd jar?
- In terms of the events we've defined we want:

$$P(J_2 \mid R) = \frac{P(J_2 \cap R)}{P(R)}$$

# Finding our Probability

Use the expanded Bayes' Rule (i.e., our friend the Law of Total Probability)

$$P(J_2 \mid R) = \frac{P(J_2 \cap R)}{P(R)}$$

$$= \frac{P(J_2 \cap R)}{P(J_1 \cap R) + P(J_2 \cap R) + P(J_3 \cap R)}$$

# Plug and Chug!

- Plugging in the appropriate values:

$$P(J_2 \mid R) = \frac{P(J_2 \cap R)}{P(J_1 \cap R) + P(J_2 \cap R) + P(J_3 \cap R)}$$

$$= \frac{\left(\frac{1}{18}\right)}{\left(\frac{1}{8}\right) + \left(\frac{1}{18}\right) + \left(\frac{4}{27}\right)} = \frac{12}{71} \approx 0.17$$

# In Summary

- Reviewed conditional probabilities
- Established Bayes’ Rule with 2 examples.
- Next Time:
  - Random variables
  - Probability functions