ISTA 116: Statistical Foundations for the Information Age

Probability Intro

19 October 2011

Outline

1 Why Learn About Probability?
2 What is Probability?
- Interpreting Probabilities

3 Sample Space and Events
- Set Operations

4 Axioms of Probability and Their Corollaries

- Saw last time some “real world” uses for probability:
  - Evaluating medical evidence
  - Detecting “Foul Play”
  - Deciding what action will give the best chance of a good result
  - Determine whether correlations / group differences in data are “real”, or due to “chance” (sampling error)

Medical Testing

- Suppose a test for a somewhat rare disease (affecting 1 in 10,000 people) is 99% accurate: 99% of sick people test positive, and 99% of healthy people test negative.
- If you test positive, what is the probability you have the disease?
- This is a question of **Conditional Probability**:
  - What is the probability of disease, *conditioned on* testing positive? (Recall tornado example)

Is A Difference Real?

- Black defendants got the death penalty more often than white defendants.
  - Are juries racially biased? What's the probability of a difference this big happening just by "random chance"?
  - If this probability is low enough, we say that the difference is **statistically significant**.
  - Note that this is a *different* question from "Is the relationship *causal*".

Is A Correlation Real?

- Frost and Illiteracy were negatively correlated among U.S. states.
- Can debate why this might be: does frost *cause* illiteracy? Does some other factor lead to increases in both?
- But before we even do that, we should ask "How likely are we to get this or a bigger correlation value *by random chance*?"
- If this probability isn't very low, there might be nothing to explain.
- Again, if the probability of this happening by random chance is low enough, then the correlation is **statistically significant**.

Conditional Probability is Not Symmetric

- We can state all of these things as conditional probabilities:
  - What is the probability of a difference this large in a **sample** *conditioned on* the hypothesis that there is really **no difference** in the population?
  - What is the probability of a correlation this large in a sample, *conditioned on* the hypothesis that there is **no relationship** in the population?

Conditional Probability is Not Symmetric

- This is *not* the same thing as the reverse:
  - What is the probability that there is no difference in the population, *conditioned on* having a difference in the sample?
  - What is the probability that there is no relationship in the population, *conditioned on* having a particular correlation in the sample?
- Example: 99% of people with the disease test positive, but only 1% of people who test positive have the disease.

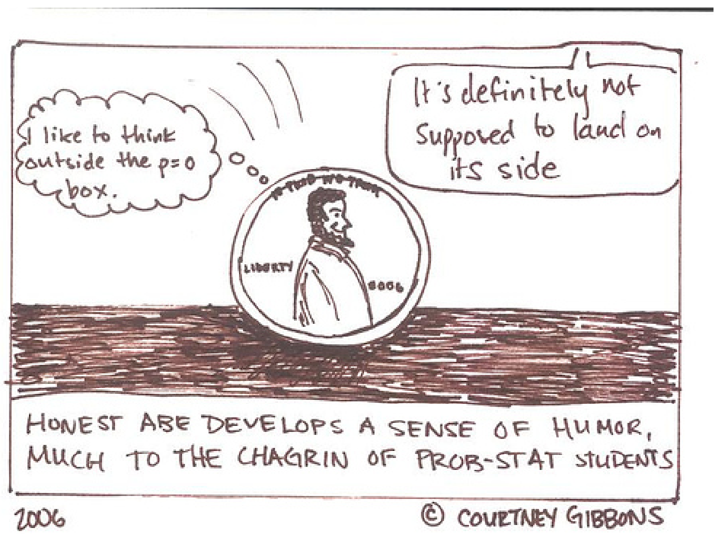
What is Probability?

- If I flip a coin, what is the probability that it will come up heads?
- Most people say $1/2$, but why is that?
- What is the probability that the coin will come up *either* heads *or* tails?
- Which is more likely?
- Intuitively we have two equations:

$$\begin{aligned} P(\text{Heads}) + P(\text{Tails}) &= 1 \\ P(\text{Heads}) &= P(\text{Tails}) \\ \Rightarrow 2P(\text{Heads}) = 2P(\text{Tails}) &= 1 \\ \Rightarrow P(\text{Heads}) = P(\text{Tails}) &= 1/2 \end{aligned}$$

Interpreting Probabilities

- What does this $1/2$ mean?
- Basically two schools of thought on this:
  - Objective probability
  - Subjective probability



Reminders/Announcements

- Lab 5 posted; due next Monday
- Extra assignments posted soon
  - Can use as a quiz grade

Objective Probabilities

- Probabilities are properties of the external world.
- The probability of an event represents the **long run proportion** of the time the event occurs under repeated, controlled experimentation.
- If I flip the coin a million bajillion times, the probability of heads is the proportion of the flips that would come up heads.

$$P(\text{Heads}) \approx \frac{\#\text{ of flips that came up heads}}{\text{Total }\#\text{ of flips}}$$

Objective Probabilities

- On this view, probabilities only truly apply to outcomes of processes that can be repeated indefinitely.
- In particular, can't ascribe a probability to something that has already happened
- Also can't ascribe probabilities to events that may only happen once.
- Objective probability is associated with **classical** or **frequentist statistics**.

Objective Probabilities

- Advantages
  - Probabilities are determined solely by observation
  - Everyone should assign the same probability to a (well defined) experimental outcome
- Criticisms
  - Restrictive set of events that can have probabilities
  - Can never assign an exact probability to anything; only an approximation

Subjective Probabilities

- Probabilities aren't in the world itself; they're in our knowledge/beliefs about the world
- I have no reason to believe heads and tails have different probabilities, so I assign them both $1/2$.
- More generally, can assign probabilities by the betting odds you'd be willing to accept:

$$\frac{P(\text{Not Heads})}{P(\text{Heads})} = \text{Minimum betting odds to bet on Heads}$$

Subjective Probabilities

- You can bet on a lot of things in principle, even things that have already happened, or would only happen once, but which you don't have complete information about.
- Moreover, different people have different information, and might ascribe different probabilities to the same event.
- But as we gain new information, we can use mathematical rules of probability to *update* our beliefs (i.e., condition on new information) in a rational way.
- This process, of subjective initial probabilities combined with rational updating, is associated with **Bayesian statistics**

Bayesian Statistics

Figure: Reverend Thomas Bayes (1701-1761)

Figure: Pierre-Simon Laplace (1749-1827)

The Dutch Book

Figure: A Dutch Book

- A "Dutch Book" is a set of bets that are guaranteed to make money regardless of the outcome
- Example: 2:1 odds on Heads *and* 2:1 odds on Tails
- Examples: Race tracks, credit default swaps, ...
- Can rule out Dutch Book probabilities as "irrational".

Subjective Probability

- Advantages
  - Can assign probabilities to a lot more propositions
  - Use Bayesian math to update beliefs as evidence comes in
  - Allow us to state questions intuitively: "What is the probability that you have disease X, *given* that you tested positive?"
- Criticisms
  - Too much "wiggle room" to set initial probabilities
  - Different people can assign different probabilities to the same thing, with both equally "rational"

The Sample Space

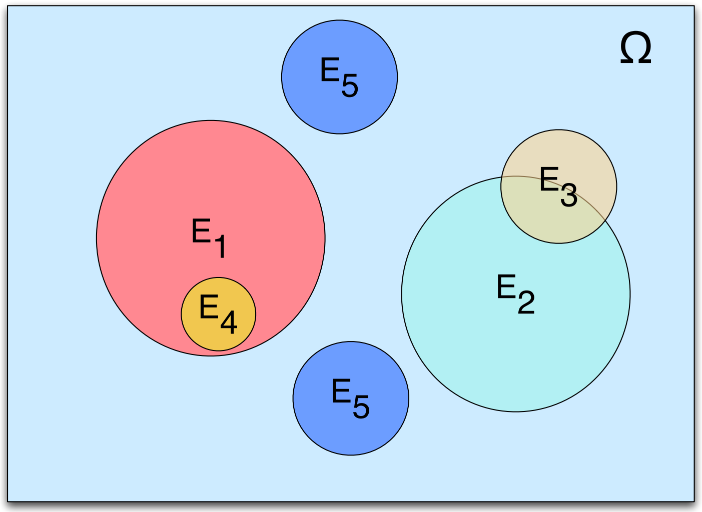
- Probability is very closely tied to area, so we use lots of spatial metaphors
- The set of all possible distinguishable outcomes of a random experiment is called the **sample space**. Often written Ω.
  - What's the sample space for an individual coin flip?
  - What's the sample space for rolling a die?
  - For pulling a ball out of a bag?
  - For randomly choosing a student?
  - For flipping two different coins?
  - Flipping one coin twice?
  - Observing the number of earthquakes in San Francisco in a particular year?
  - Observing someone's height?

A Sample Space

Events

- Formally, an **event** E is a subset (or "sub-region") of the sample space. When we make a particular observation, we are either "in" E or not.
- It is sometimes helpful to think about events as propositions ("sentences") that are either True or False on a particular occasion.
  - "The coin comes up heads"
  - "The die comes up an even number"
  - "There are more than 20 earthquakes"
  - "I choose a sophomore"
- Notice that Ω itself is an event: "something possible happens".
- The **empty set**, $\emptyset$ is also an event: "nothing possible happens".

Some Events



Events

- It is events that have probabilities.
  - $P(\text{Coin comes up heads}) = 1/2$
  - Or, for short, $P(\text{Heads}) = 1/2$
  - $P(\text{Die comes up an even number}) = 3/6$
  - etc.
- Probabilities correspond to the "size" of the event set.

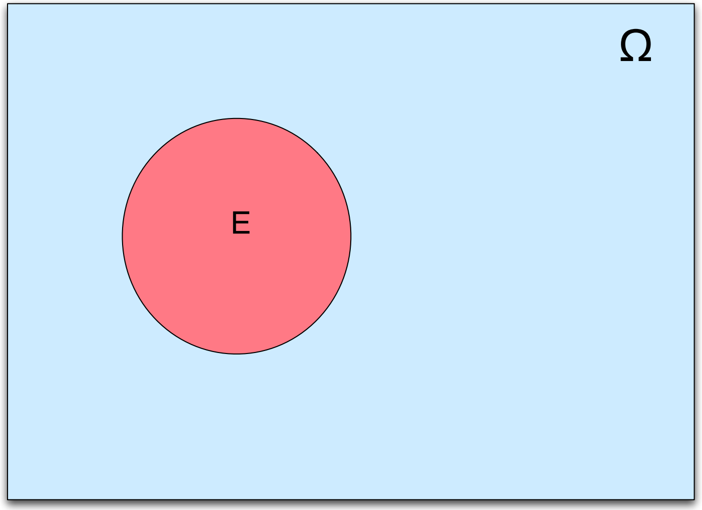
Set Operations

- Since events are sets of outcomes, we can apply **set operations** on them to get new events.
- If we think about events as True/False propositions, the basic set operations correspond to AND, OR and NOT.

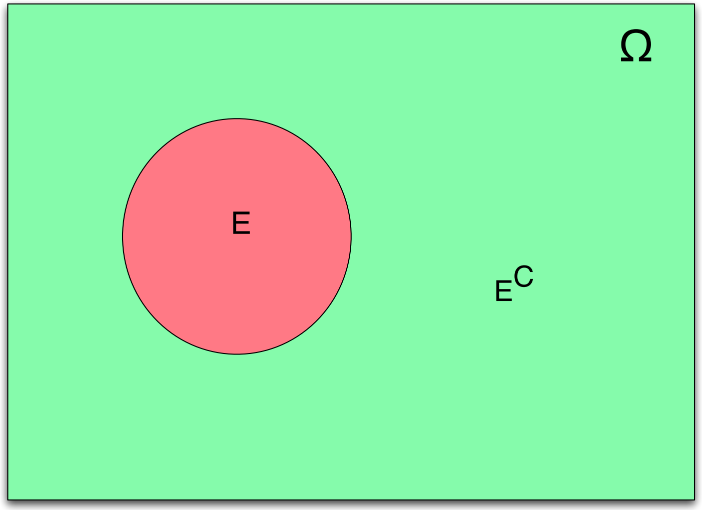
Set Complement = NOT

- If E is an event, its **complement**, written E^C is the negation of E: it is the event representing everything in the sample space outside E that can occur.
- For each E, identify E^C in words:
  - E_1 = Coin comes up heads
  - E_2 = Die comes up even
  - E_3 = I draw a yellow ball
  - E_4 = There are more than 20 earthquakes
  - E_5 = A person is between 5 and 6 feet tall (inclusive)
  - E_6 = Die comes up between 1 and 6 (inclusive)
- What is $(E^C)^C$?

The Complement



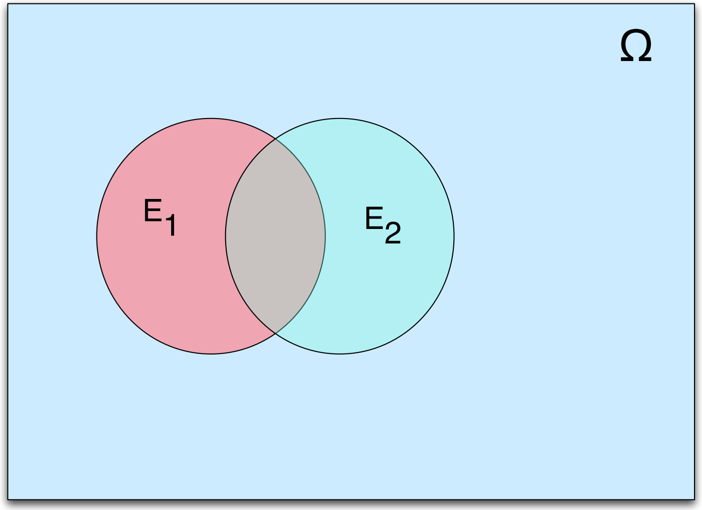
The Complement



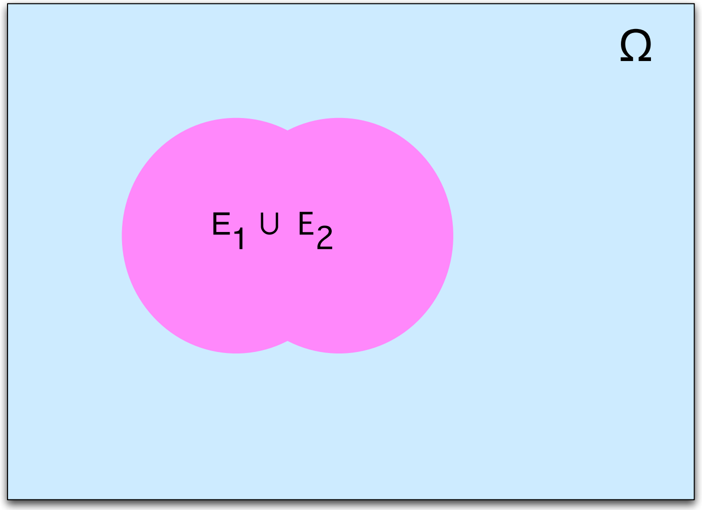
Set Union $= OR$

- If E_1 and E_2 are events, their **union** (written $E_1 \cup E_2$) represents the event that happens when *either* E_1 *or* E_2 (or both) occur.
- Identify $E_1 \cup E_2$ in words:
  - $E_1 =$ Coin comes up heads; $E_2 =$ Coin comes up tails
  - $E_1 =$ Die comes up 2 or less;
    $E_2 =$ Die comes up 4 or more
  - $E_1 =$ I pick a freshman; $E_2 =$ I pick a sophomore.
  - $E \cup E^C$

The Union



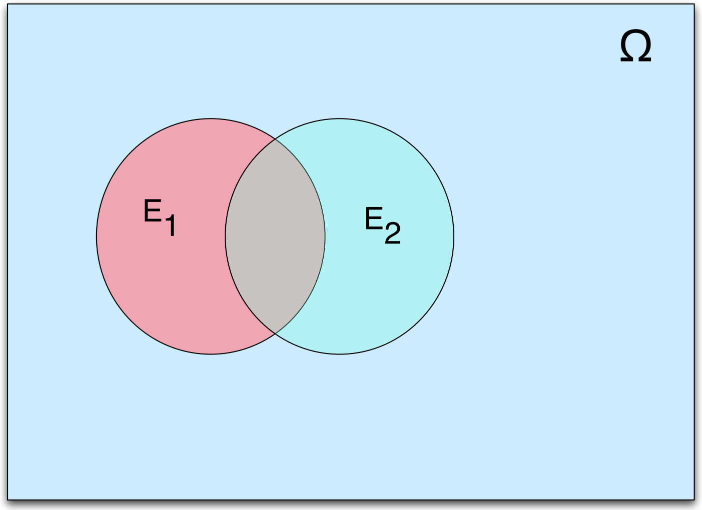
The Union



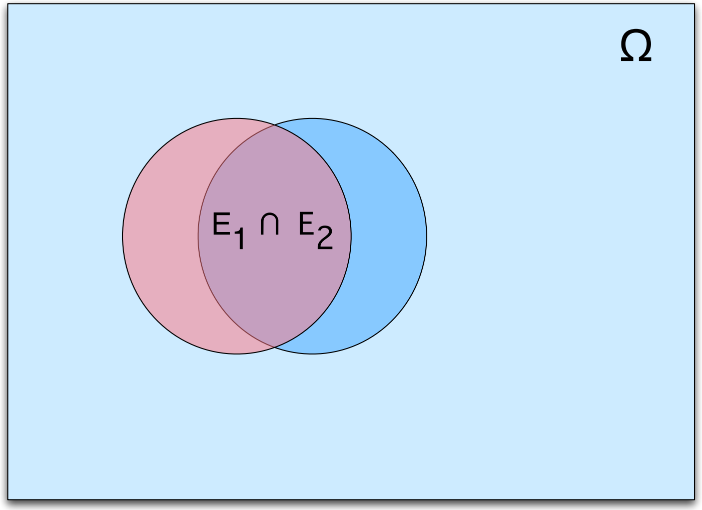
Set Intersection = AND

- If E_1 and E_2 are events, their **intersection** (written $E_1 \cap E_2$) represents the event that happens when *both* E_1 *and* E_2 occur.
- Identify $E_1 \cap E_2$ in words:
  - $E_1 =$ Coin comes up heads; $E_2 =$ Coin comes up tails
  - $E_1 =$ Die comes up 2 or more;
    $E_2 =$ Die comes up 4 or less
  - $E_1 =$ I pick a lefty; $E_2 =$ I pick an upperclassman
  - $E \cap E^C$
  - $E_1 \cap (E_1 \cup E_2)$

The Intersection



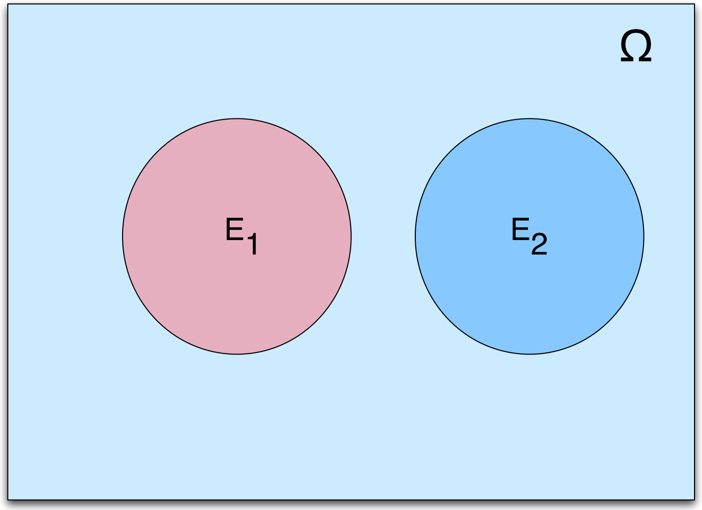
The Intersection



Disjoint Events

- Two events, E_1 and E_2 are said to be **disjoint** if their intersection is the empty set, i.e., $E_1 \cap E_2$
- What does this mean in words?

Disjoint Events

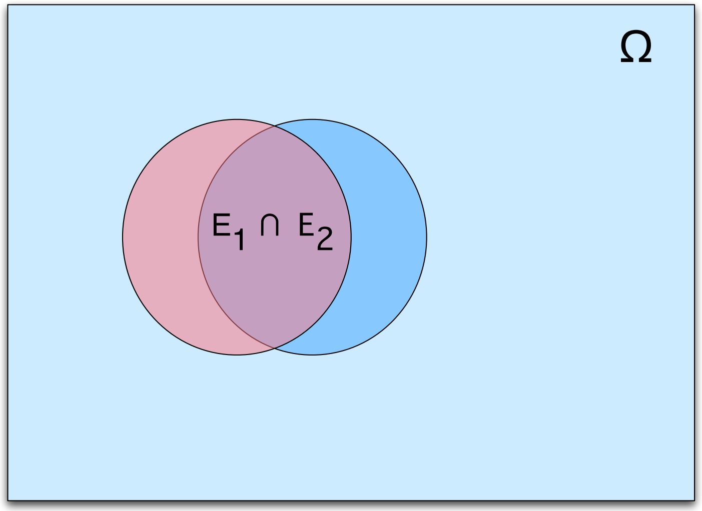


DeMorgan's Laws

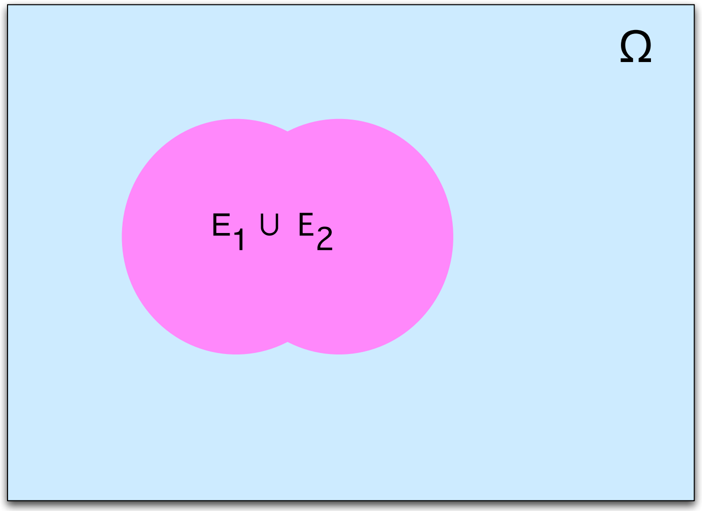
There are some basic identities called **DeMorgan's Laws** that tell you what you get when you combine set operations in certain ways

1. $(E_1 \cap E_2)^C = E_1^C \cup E_2^C$
   - The complement of the intersection is the union of the complements.
   - "If not both, then either not one, or not the other"
2. $(E_1 \cup E_2)^C = E_1^C \cap E_2^C$
   - The complement of the union is the intersection of the complements.
   - "If neither, then both not one and not the other"

The Intersection



The Union



Three Axioms of Probability

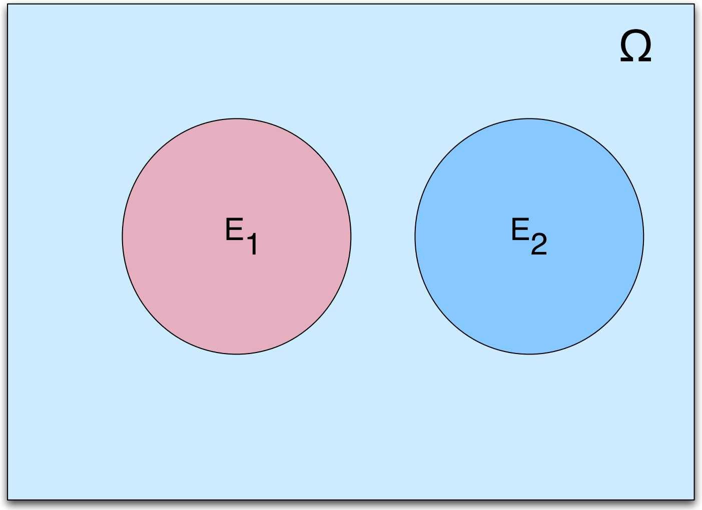
Regardless of your preferred interpretation of probabilities, they have to follow three simple rules:

(Axioms of Probability)

1. **Nonnegativity**: $P(E) \geq 0$ for any event E
2. **Disjoint Additivity**: If E_1 and E_2 are **disjoint**, then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$.
3. **Unity of the Sample Space**: $P(\Omega) = 1$

Everything else follows from these three.

Disjoint Events



Consequences of the Axioms

- How would I calculate $P(E^C)$? (For example, suppose E is "There are fewer than 20 earthquakes")
- How would I calculate $P(E_1 \cup E_2)$ if E_1 and E_2 might *not* be disjoint? (Take E_1 = "I pick a lefty", E_2 = "I pick an upperclassman")
  - What else do I need to know?

Two Corollaries

(Probability Corollaries)

1. **The Complement Rule**: $P(E^C) = 1 - P(E)$
   - Either E happens or it doesn't, so these two probabilities add to one
   - Proof: E and E^C are disjoint, and their union is Ω, so $P(E) + P(E^C) = P(\Omega) = 1$.
2. **Probability of General Unions**: $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$
   - If we add up the probabilities, we are double counting the intersection, so subtract one out.
   - Four disjoint possibilities: $E_1 \cap E_2$, $E_1 \cap E_2^C$, $E_1^C \cap E_2$, and $E_1^C \cap E_2^C$. Now construct the pieces out of disjoint unions.