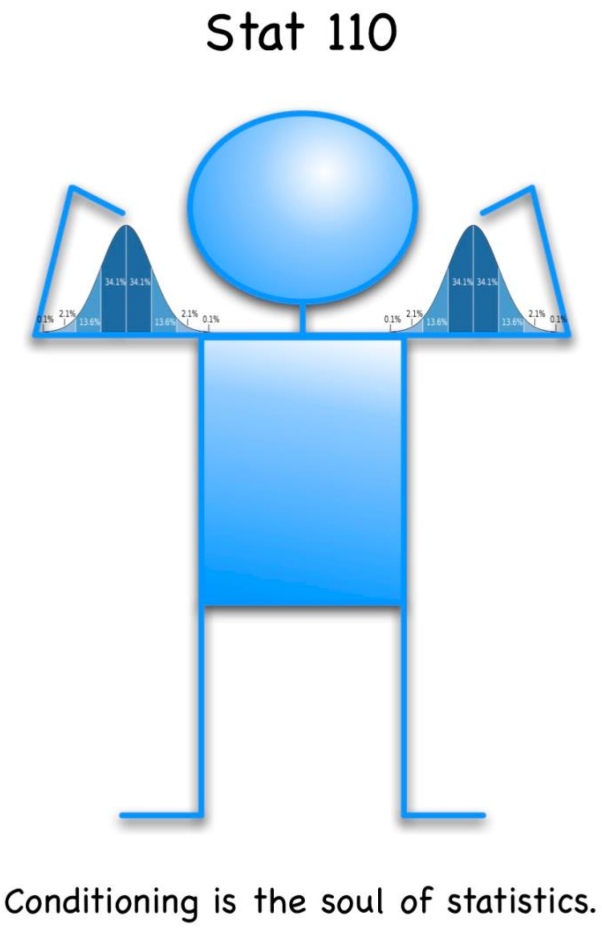
Welcome to Stat 110:
Introduction to Probability



Unit 1: Intro to Probability and Counting

Unit 1 Outline

- Course logistics and details
- What is Probability?
- Probability Axioms/Rules
- Counting
- The Birthday Problem
- Story Proofs

Stat 110

- Prereq's:
  - Multivariable Calculus (equivalent to Math 21a…can be taken concurrently)

- The course material is nothing like what you may have seen in an introductory statistics course (like stat 104). The material will be quite *mathy* and calculus-based, eventually.

- It will be example and problem driven, and will not include many theoretical proofs, but will include many heuristic and intuitive mathematical justifications: often referred to as "story proofs".

Kevin’s Contact Info

- My office: Science Center, Room SC-614
- Office Hours:
    - Tues 2:30-3:30pm, and Thurs11:45am-12:45pm
    - **Also happily by appointment (via email)**

- Phone numbers:
    - Statistics Department: (617) 495-5496
    - My office (SC-614): (617) 495-8711

- Email: krader@fas.harvard.edu (best way to get a hold of me)

Teaching Staff

- Teaching Fellows:

  Head TF: Viviana Garcia (vgarcia@fas.harvard.edu)

  Other TFs: too many to list. See course website "Teaching Staff and Office Hours" for finalized list by the end of the week:
  http://isites.harvard.edu/icb/icb.do?keyword=k104821&pageid=icb.page690517

- Teaching fellows will be teaching sections, holding office hours, answering questions via email, and grading HW's and exams.

Course Website

Course website:
http://isites.harvard.edu/icb/icb.do?keyword=k104821

- There you will find (eventually):
  - Syllabus
  - Administrative Announcements
  - Lecture Notes and Videos
  - Section Material
  - Assigned Homeworks
  - Other Study Material (practice exams, web links, etc...)

Lectures

- Meeting Times:
  - Tues & Thurs, 1–2:30pm in Sanders Theater in Memorial Hall.
- Lectures are not mandatory. But the easiest way to succeed in this course (and in college) is to attend class and **pay attention**!
- Lectures will be videotaped, and the videos will be posted on the course website.
- Exams will be based on the lectures and HWs.

Sections

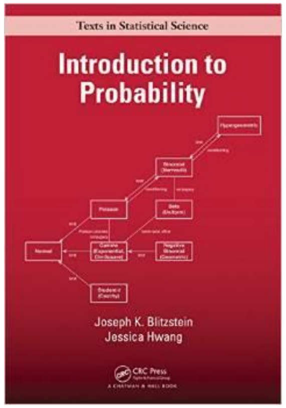
- Optional (but strongly recommended) weekly section to discuss homework, do extra problems, and review difficult concepts.
- Meeting times: all throughout the week (mostly Tuesday through Friday, afternoons and evenings)
- Sections will begin next week (Sept 8-12) as the first HW is due Fri, Sept 12.
- Look for announcement on the course website for permanent times (OH's too).

Lecture Notes

- Paper copies will NOT be handed out at the beginning of lecture after Unit #2 (we will provide copies up til that point).
- They'll be organized unit-by-unit (which will roughly follow the chapters in the text). Each unit is about 3 hours of lecture time (with random variation around that average time).
- Future lecture notes will be posted online about 24 hours in advance. An email will be sent when they are posted.
- Notes are very concise – you are encouraged to add your own annotations and develop your own notes.
- Occasionally mistakes appear in lecture notes; corrected versions will be posted after class.

Recommended Textbook

(not required)



- *Introduction to Probability*, Blitzstein & Hwang, 1st edition. Amazon Link:

  www.amazon.com/Introduction-Probability-Chapman-Statistical-Science/dp/1466575573

  Publisher Direct, 20% off as of Aug 27…may now be expired: http://www.crcpress.com/product/isbn/9781466575578

- Some of the assigned homework problems will be assigned from the text, but will always be reproduced for you on the assignment itself.

- Exams will be based on the lectures directly, and nothing new from the text not seen in the lectures, notes, or HW's.

- It's a great reference for more details on what is seen in the lectures. The lectures will follow the text pretty closely.

Computing (not required at all)

- Statistical Computing Package: ***R***
  - Can be downloaded from: http://cran.us.r-project.org/
  - Rstudio program will be encouraged.

    http://www.rstudio.com/
- Some demonstrations and calculations will be presented in class, but *R* will not be used besides that. It is a great tool to learn if you are going to practice statistics in the future.
- Some calculations on HW will asked to be simplified numerically. Any of *R*, google, or an online calculator (or anything else) can be used for these.

Exams

- Midterm: Thurs, Oct 23, 1-2:30pm (location TBD).
- Final Exam: Exam Group 8. Date and location TBD.
- You will be given a page of distributions on the exam (see online for the exact sheet).
- You will be allowed one sheet of notes for the midterm and two sheets for the final exam. Front-and-back OK.
- Exams are difficult. The median is often below 70.

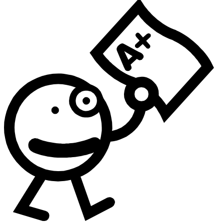
Homeworks

- Posted to course website one week in advance. There will be 11 of them, and they will be due at 1:10pm the following Friday.
  - HW #1 will be posted by Friday at noon, and will be due the next Friday, Sept 12th, at 1:10pm.
- Submission: paper copy to the Stat 110 HW collection box (location TBD). Directions will be at the top of the assignment.
- No HWs are dropped ☹. You will be allowed one late HW (up to 96 hours later) with no questions asked. You will be allowed to submit late HW with an official excuse (note from UHS or your resident dean).

HW Collaboration

- You are encouraged to discuss homework with other students (and with the instructor and TAs, of course), but you must write your final answers yourself, in your own words.
- Solutions prepared “in committee” or by copying or paraphrasing someone else’s work are not acceptable; your handed-in assignment must represent your own thoughts.
- **Please indicate on your problem sets the names of the students with whom you worked.**

Course Grading



Component	*Weighting1*	*Weighting2*
Homework	30%	30%
Midterm	20%	35%
Final Exam	50%	35%
Total	**100%**	**100%**

Your overall score for the course will be the higher of the 2 weighting schemes presented above. Final course letter grades are not assigned according to a fixed percentages of A's, B's, etc…

Course Description

A comprehensive introduction to probability, as a language and set of tools for understanding statistics, science, risk, and randomness. Basics: sample spaces and events, conditional probability, and Bayes' theorem. Univariate distributions: density functions, expectation and variance, Normal, *t*, Binomial, Negative Binomial, Poisson, Beta, and Gamma distributions. Multivariate distributions: joint and conditional distributions, independence, transformations, and Multivariate Normal. Limit laws: law of large numbers, central limit theorem. Markov chains: transition probabilities, stationary distributions, convergence.

Unit 1 Outline

- Course logistics and details
- What is Probability?
- Probability Axioms/Rules
- Counting
- The Birthday Problem
- Story Proofs

Probability is the basis of Statistics

- Statistics is the field that studies uncertainty (while Mathematics studies certainty). Probability is useful to help explain this uncertainty and variation/variability.
- Probability is used in many other fields:
- Biology: used all the time in Genetics (and with Bayes theorem, which we will study extensively).
- Medicine: clinical trials are rooted in probability.
- Physics: quantum mechanics is based almost completely on probability (of where an electron is).
- Gambling: Historically, probability was first studied based on games of chance. Bring down the house!
- Finance: “betting” on the stock market’s uncertainty

Probability Terminology

Terminology

- ***random experiment*** – an experiment whose *individual outcomes* are uncertain but there is a regular distribution in a large number of repetitions.
    - Examples:
        - Coin tossing and dice rolling
        - The lottery and other games of chance
        - Results of taking a drug in a clinical trial.

- ***outcome*:** the value of one replication of a random experiment. Written as little *s*.
    - Coin Tossing:
        - $s = H$ with one toss of a coin
        - $s = HTT$ with three tosses

Probability Terminology (cont.)

- ***sample space* (labeled *S*):** is the set of all possible outcomes of a random experiment
- Examples:
    1. Toss a coin three times: $S = \{HHH, THH, HTH, \ldots, TTT\}$
    2. Face showing when rolling a six-sided die: $S = \{1,2,3,4,5,6\}$
    3. Pick a **real** number between 1 and 20: $S = \{[1,20]\}$
    4. Sex of a randomly selected person: S = {Male, Female}

- ***event* (labeled *A*, *B*, etc…):** a set of outcomes of a random experiment.
- Examples:
    1. The event A that *<u>exactly</u> two heads* are obtained when a coin is tossed three times: $A = \{HHT, HTH, THH\}$
    2. The result of the toss of a fair die is an even number: $A = \{2,4,6\}$
    3. The number chosen from the set of all real numbers between 1 and 20 is at most 8.23: $A = \{[1,8.23]\}$
    4. A randomly selected person is Female: A = {Female}

Combining Two or More Events (more Terminology)

- The ***union*** of two events *A* and *B* is the event that either *A* occurs or *B* occurs or both occur:

$$C = (A\ or\ B) = (A \bigcup B)$$

- The ***intersection*** of two events *A* and *B* is the event that both *A* and *B* occur.

$$C = (A\ and\ B) = (A \bigcap B) = AB$$

- The ***complement*** of an event *A*, A^c, is the event that *A* does not occur and thus consists of outcomes that are not in *A*

Note: the "or" in statistics is always the inclusive or.

More Set Terminology

English:	Set Notation:
empty set	$\varnothing$
a possible outcome	$s \in S$
A occurred	$s_{actual} \in A$
something must occur	$s_{actual} \in S$
A or B, but not both	$(A \cap B^C) \cup (A^C \cap B)$
at least one of $A_1, \ldots, A_n$	$A_1 \cup A_2 \cup \cdots \cup A_n$
all of $A_1, \ldots, A_n$	$A_1 \cap A_2 \cap \cdots \cap A_n$
A implies B	$A \subseteq B$
A and B are disjoint (mutually exclusive)	$A \cap B = \varnothing$
$A_1, \ldots, A_n$ are a partition of S.	$(A_1 \cup \cdots \cup A_n) = S,\ A_i \cap A_j = \varnothing$ for $i \neq j$

Naïve Definition of Probability

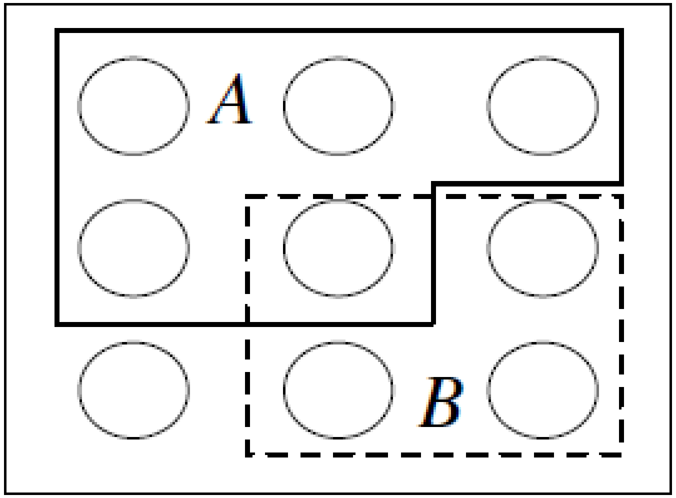
- For a random phenomenon, if the sample space is finite and **if all of the individual outcomes have the same probability**, then the probability of an event A (written $P(A)$) is the ratio

$$P_{naive}(A) = \frac{\#\text{ of elements in } A}{\#\text{ of elements in } S} = \frac{|A|}{|S|}$$

Use this formula to determine the probability of getting two heads in three tosses of a coin? Probability of getting an even number in one roll of a die?

Illustration of Probability: Pebble World

- Sample spaces, outcomes, events, and probability can be illustrated through a "Pebble World" view [if S is finite].
- Every outcome is represented by a "pebble" in the sample space. And events are just collections of these pebbles:



- Probabilities then can be calculated based on counting pebbles.

Interpretation of Probability

- There are 2 main schools of thought on how to interpret probability:
- ***<u>Frequentist</u>*** <u>approach:</u> probability represents a long-run frequency of occurrences in a large number of repetitions. In a coin-flipping example, if we say P(Heads) = ½, then we would expect to observe the coin landing heads up 50% of the time if we tossed it over and over and over…
- ***<u>Bayesian</u>*** <u>approach:</u> probability represents a degree of belief about an event's possibility of occurring. This allows us to assign probabilities to events that simply cannot be repeated (think: predicting the results of an election before it has occurred).
- We'll talk more about these differences throughout the course.

Standard Deck of Cards

(like for poker)

Standard Deck of Cards Example

- You pick a card from a standard deck of 52 cards. The sample space S is the set of all 52 cards.
  - What are some examples of events in this setting?
  - What are the union, intersection, and complement of these events? How do they relate?
  - What are the probabilities of these events and the combination of them?
- Be careful with notation: you can only take probabilities of events, which are comprised of outcomes. You cannot take probabilities of numbers. P(7) doesn't really make sense (bad notation), but P(card is a 7) = P(A) where A = "card is a seven" makes perfect sense!

Standard Deck of Cards Example (cont.)

- Let:
  A = card is an ace
  K = card is a King
  F = card is a face card (Jack, Queen, King)
  D = card is a diamond
  H = card is a heart
- What are the union, intersection, and complement of these events?
- How do these events relate?
- What are the probabilities of these events and the combination of them?

Unit 1 Outline

- Course logistics and details
- What is Probability?
- Probability Axioms/Rules
- Counting
- The Birthday Problem
- Story Proofs

Rules (axioms) of Probability

Non-naïve definition of probability: probability is a function that takes an event from a sample space and assigns to it a real number between 0 and 1. This probability function must satisfy three rules:

Axiom 1: $P(A) \geq 0$

Axiom 2: $P(S) = 1$.

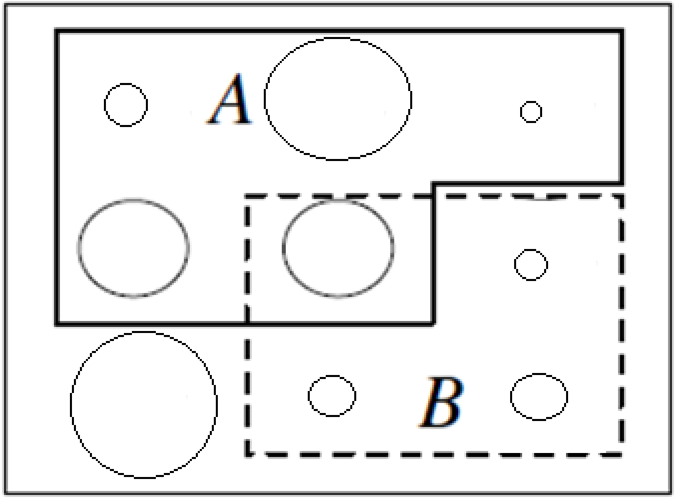
Axiom 3: If $A_1, A_2, \ldots$ are disjoint events, then:

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j)$$

All of probability can be derived based on these rules!

Non-naïve Probability in Pebble World

- We can illustrate this probability mapping in Pebble World (again, for finite sample spaces).
- In stead of calculating probabilities simply by counting pebbles, we instead can make these pebbles have mass (change their size).
- Key: The mass must sum to one. And then probability is just the total mass of pebbles inside the event.



Some results of these Probability Axioms

Result 1: $P(A^c) = 1 – P(A)$. The probability of an event not happening is 1 – (Probability of event happening). Proof:

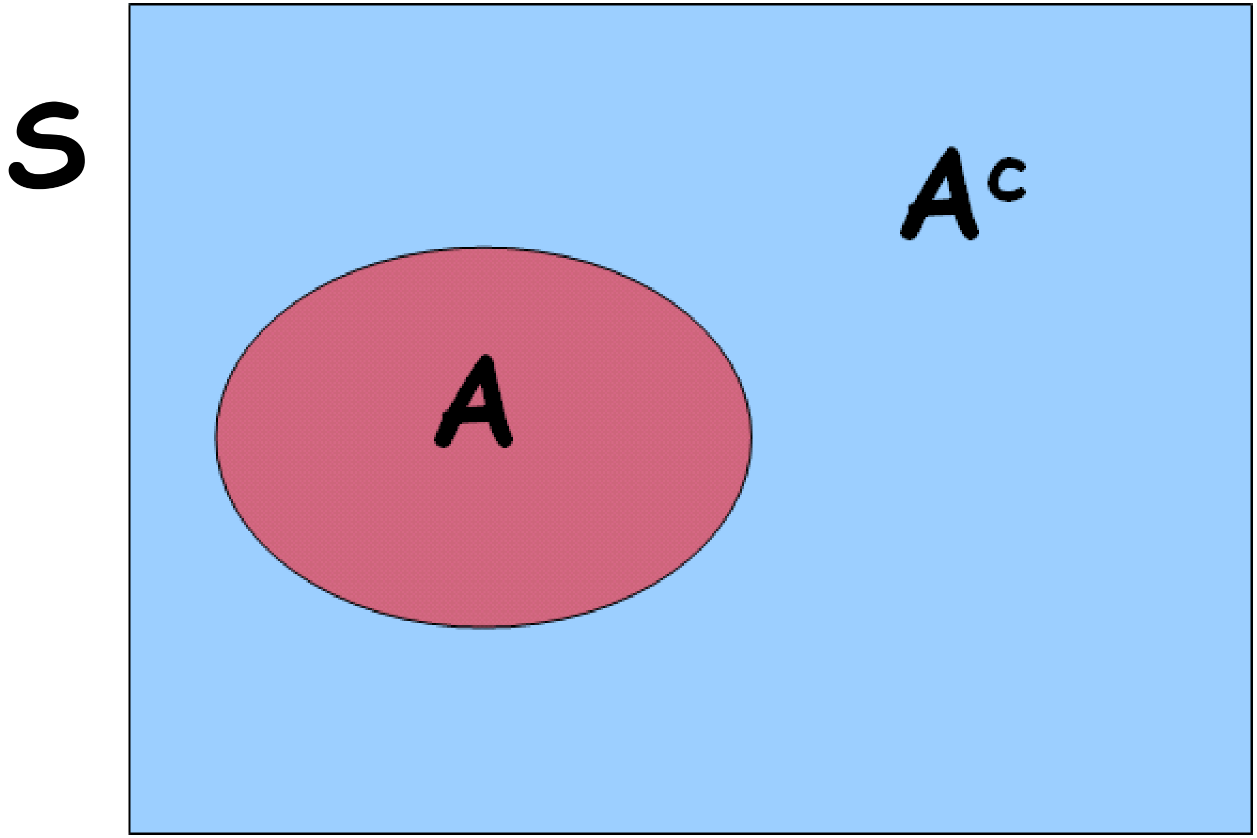
Result 2: If $A \subseteq B$, then $P(A) \leq P(B)$. If A is contained in B, then A's probability cannot be greater then B's probability.

Result 3: $P(A \cup B) = P(A) + P(B) – P(A \cap B)$. The probability of A or B happening is the sum of their probabilities, minus the overlap.

*Note: If A and B are *disjoint* or *mutually exclusive* events then $P(A \cup B) = P(A) + P(B)$.

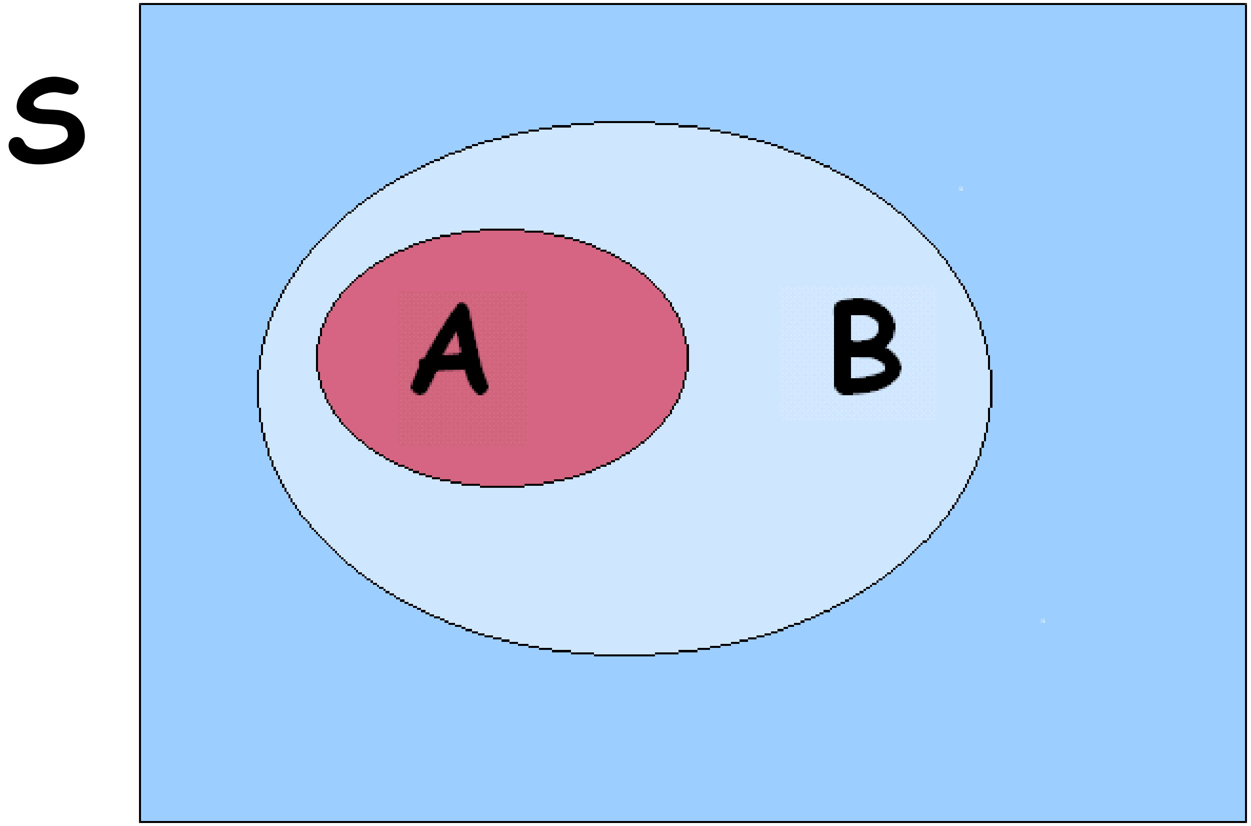
Justification for these can be seen in Venn Diagrams…

Result 1. For any event A, $P(A^c) = 1 - P(A)$



Venn Diagram

Result 2. If $A \subseteq B$, then $P(A) \leq P(B)$

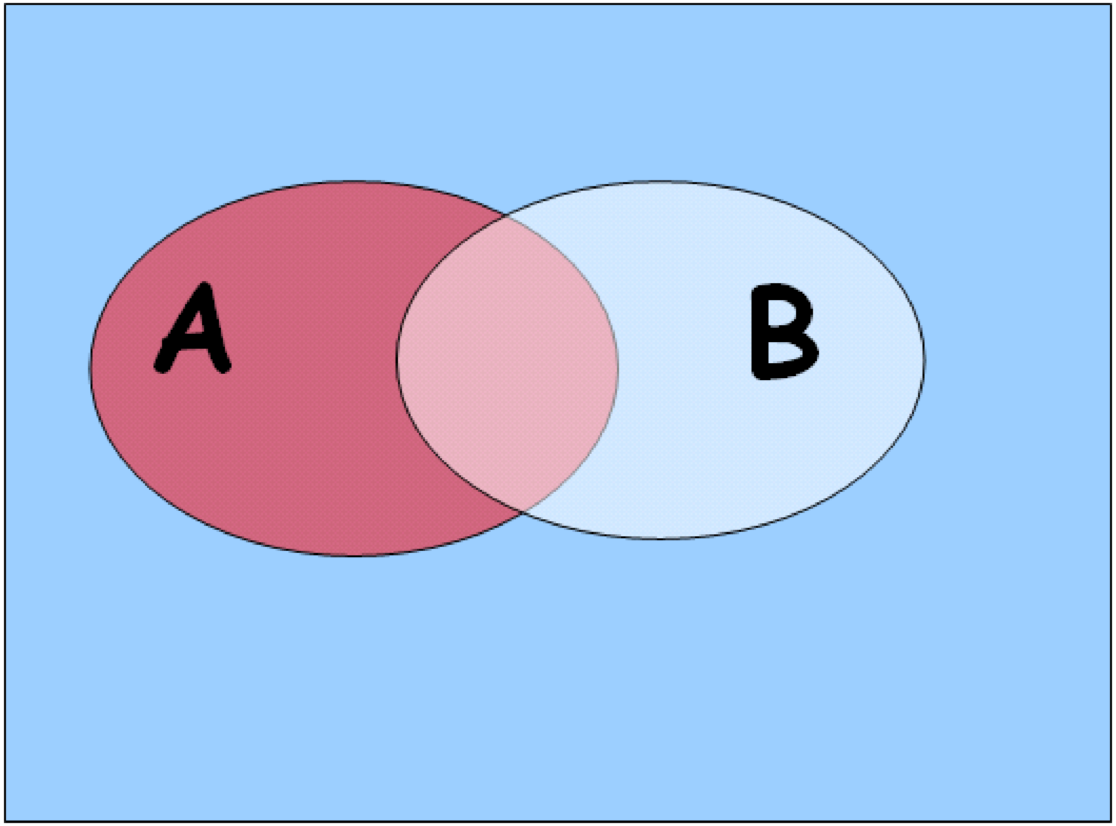


Venn Diagram

Result 3. In general for any events A and B:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

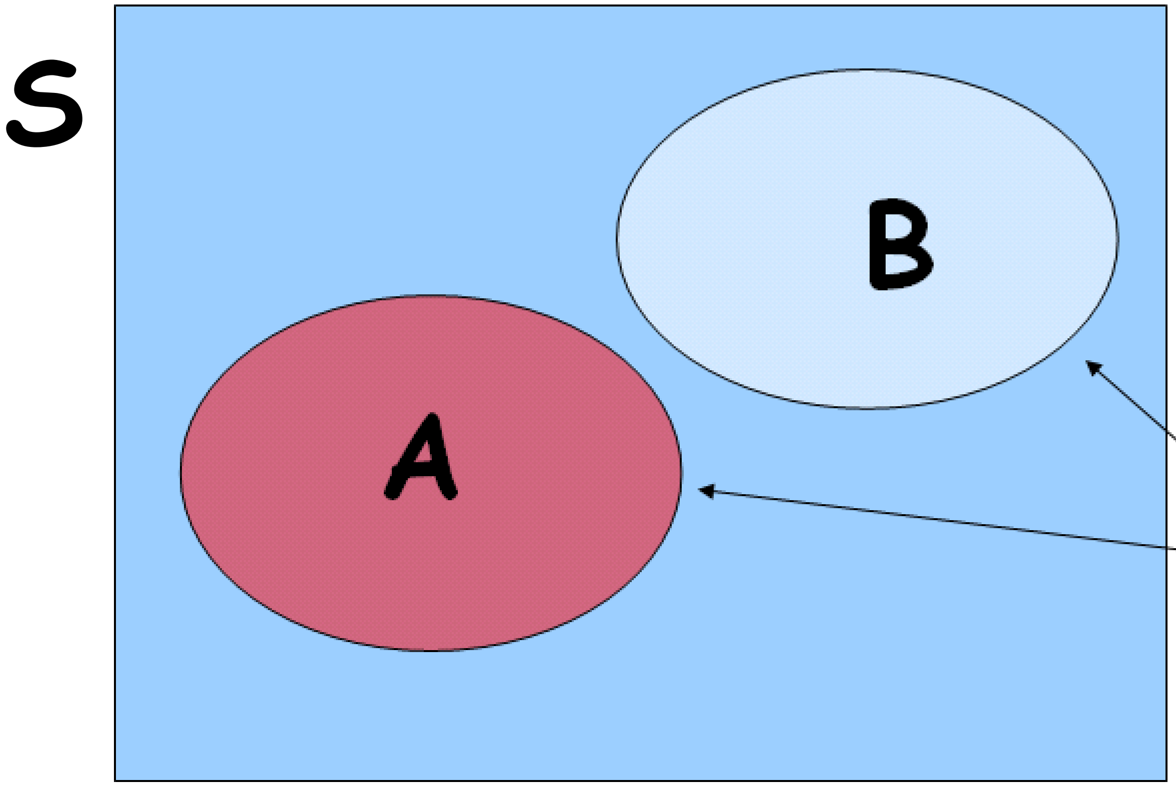
S



Venn Diagram

If A and B are disjoint events then

$P(A \cup B) = P(A) + P(B)$



A and B are mutually exclusive (aka disjoint)

-No outcomes in common, they cannot happen at same time

Venn Diagram

Inclusion-Exclusion: an extension

For 3 events, the probability of their union is (Diagram helps):

$$P(A \bigcup B \bigcup C) = P(A) + P(B) + P(C)$$
$$- P(A \bigcap B) - P(A \bigcap C) - P(B \bigcap C)$$
$$+ P(A \bigcap B \bigcap C)$$

In general for *n* events, the probability of their union is:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j}^{n} P(A_i \cap A_j)$$
$$+ \sum_{i<j<k}^{n} P(A_i \cap A_j \cap A_k) - \dots$$
$$+ (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n)$$

Deck of *n* Cards Example:

- Consider a well-shuffled deck of *n* cards, labeled 1 through *n*. You flip over the cards one at a time saying the numbers 1 through *n* as you are flipping. You win the game if at any point the number on the card matches the number you say aloud (for example, the 3rd card flipped over is labeled #3).
  [t*his is called de Montmort's matching problem*]

- What is the probability of winning?

- Key: write this union of events as an inclusion-exclusion of intersection of events…we'll come back to solving this to completion.

Unit 1 Outline

- Course logistics and details
- What is Probability?
- Probability Axioms/Rules
- Counting
- The Birthday Problem
- Story Proofs

Counting

- In order to calculate probabilities, we must learn how to count the events in a sample space.

- At first this is very easy (think coin flips), but it can get very difficult (think poker hands from a deck of cards).

Multiplication Rule

- Consider a compound experiment consisting of 2 sub-experiments: experiment A and experiment B. Suppose experiment A has *a* possible outcomes and experiment B has *b* possible outcomes. Then the overall compound experiment has $a \cdot b$ possible outcomes.
- This result can easily be seen based on a tree diagram.

Multiplication Rule Example:

- Example: an ice cream company sells 4 flavors (vanilla, chocolate, strawberry, mint) in one of 3 different types of cones (cake, sugar, waffle).
  - How many possibilities for cone and ice cream flavor combinations are there?
  - Now suppose you order 2 ice cream cones in a weekend (one on Sat and one on Sun). How many possibilities of pairs of ice cream cones are there?
  - Now suppose you don't care what order you ate the cones, just which flavors/combos you ate. How many possibilities of pairs are there now?
  - What if you decide that you will not order the same thing twice?

Sampling with Replacement

- Consider n objects and making k choices from them, one at a time with replacement (that is, choosing a certain object does not preclude it from being chosen again).
- How many possibly outcomes are there for this sample?
- There are n^k possible outcomes.

Sampling without Replacement

- Consider n objects and making k choices from them, one at a time *without replacement* (that is, an object can only be chosen once).
- How many possibly outcomes are there for this sample?
- There are $n(n-1)(n-2)\ldots(n-k+1)$ possible outcomes
- Note: $k \leq n$

Permutations and Factorials

- A permutation of the numbers 1,2,3…,n is an arrangement of all of the numbers in some order. For example, 3,1,5,4,2 is a permutation of the number 1,2,3,4,5.
- Based on the sampling without replacement theorem, there are $n! = n(n-1)(n-2)\ldots 1$ ways to order these numbers.
  - Recall: $0! = 1$ (by definition)
- Example: how many ordered ways can n people form a line?

Forming Groups

- Suppose we would like to select a sample of size *k* from a population of size *n*, and the order of the individuals selected in the sample does not matter!
- How many different samples can be selected?
- Sometimes it's helpful to keep it simple: let's say $n = 5$ and $k = 2$:

- In general, this can be counted with the *binomial coefficient* (sometimes called a *combinatoric* operator).

Binomial Coefficient

- For any non-negative integers k and n, the binomial coefficient, $\binom{n}{k}$, read as "*n* choose *k*", is the number of subsets of size *k* from a set of size *n*.

- For $k \leq n$, the binomial coefficient is defined as:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\ldots(n-(k-1))}{k!}$$

Binomial Coef. Example #1

- Consider a group of four people:
  - How many ways are there to choose a 2-person committee?
  - How many ways are there to form two teams of two (and we do no care which team is which)?

More Binomial Coef. Examples

- Consider a group of 100 people.
  - How many ways are there to choose a 2-person committee?
  - How many ways are there to choose a 5-person committee?
- How many ways are there to permute the letters in the word WAWAWAAA?
- How many ways are there to permute the letters in the word STATISTICS?

Poker Hand Example:

- You pick a hand of 5 cards from a standard deck of 52 cards. The sample space *S* is the set of all possible hands of 5 cards.
- How many different hands of cards can you get?
    - Note: each of the 52 cards is unique, but the order you draw the cards does not matter.
- What is the probability of each of these hands?
- What are some examples of events?
- How many ways can you be dealt a *full house*: which is when you have three of the same number and two of another number (Ex: 7♣, 7♥, 7♦, K♠, K♣).
- What is the probability of being dealt a full house?

Deck of *n* Cards Example Revisited

- Consider a well-shuffled deck of *n* cards, labeled 1 through *n*. You flip over the cards one at a time saying the numbers 1 through *n* as you are flipping. You win the game if at any point the number on the card matches the number you say aloud (for example, the 3rd card flipped over is labeled #3).
  [t*his is called de Montmort's matching problem*]

- What is the probability of winning?

- Key: write this union of events as an inclusion-exclusion of intersection of events…

Deck of *n* Cards Example, Solution

Let A_i be the event that the i^{th} card has the number i on it.
What is $P(A_i)$? Why? $P(A_i) = 1/n$

For $i \neq j$, what is $P(A_i \cap A_j)$? Why?

$$P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$$

For $i \neq j \neq k$, what is $P(A_i \cap A_j \cap A_k)$? Why?

$$P(A_i \cap A_j \cap A_k) = \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)}$$

Deck of *n* Cards Example, Solution

So what is $P(A_1 \cup \ldots \cup A_n)$?

$$P(A_1 \cup \cdots \cup A_n) = \frac{n}{n} - \frac{\binom{n}{2}}{n(n-1)} + \frac{\binom{n}{3}}{n(n-1)(n-2)} - \ldots$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \ldots$$

Recall, the Taylor Series expansions for $1/e = e^{-1}$ around $a = 0$:

$$e^{-1} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \ldots$$

So the probability of winning converges to $1 - e^{-1} \approx 0.63$ for big n.

Unit 1 Outline

- Course logistics and details
- What is Probability?
- Probability Axioms/Rules
- Counting
- The Birthday Problem
- Story Proofs

The Birthday Problem

- There are *k* people in a room. Assume each person's birthday is equally likely to be any of the 365 days of the year (we are ignoring leap years), and that the people's birthdays are independent/unrelated (no sets of twins).
- What is the probability that two or more people in the room have the same birthday?
- How many people do you need before there is at least 0.50 probability of a shared birthday?

The Birthday Problem (cont.)

- To begin, let's start simple (a common strategy in probability):
- What if k = 1? k = 2? k = 3?
- Another strategy: it is often easier to find $P(A^c)$ than it is to find $P(A)$. And recall: $P(A) = 1 - P(A^c)$

- To generalize this:

P(at least 1 birthday match) = 1 – P(no birthday matches)

$$= 1 - \frac{365 \cdot 364 \cdot 363 \cdots (365 - (k-1))}{365^k}$$

The Birthday Problem (cont.)

- Here is a table of probabilities of matching for different values of k:

k	P(at least 1 birthday match)
5	0.027
10	0.117
20	0.411
22	0.476
23	0.507
50	0.970
100	0.9999997

Unit 1 Outline

- Course logistics and details
- What is Probability?
- Probability Axioms/Rules
- Counting
- The Birthday Problem
- Story Proofs

Story Proofs

- As mentioned earlier, we will not be doing many analytical and algebraic proofs. They get too burdensome too quickly.
- We instead will provide what we will call *story proofs*, which are intuitive explanations for results/theorems. Or proof by interpretation.
- It may not be mathematically rigorous, but goes further in explaining ***why*** the result is actually true.
- Best to show some examples.

Story Proof #1

- For any non-negative integers n and k with $k \leq n$, we can show:

$$\binom{n}{k} = \binom{n}{n-k}$$

- This is easy algebraically. But the intuitive explanation is more important:
- Consider choosing a committee of size k in a group of n people. There are $\binom{n}{k}$ way to do this. But another way to choose this committee is to specify the n-k people not in the committee, and there are $\binom{n}{n-k}$ ways to do this. This fully specifies the k people who are in the committee, and vice-versa. Thus it is two way of counting the same thing.

Story Proof #2

- For any non-negative integers n and k with $k \leq n$, we can show:

$$n\binom{n-1}{k-1} = k\binom{n}{k}$$

- This is also easy algebraically. But the story proof is better:
- Suppose you are trying to select a team of k people from n people, one of whom will be selected as team captain. We could first choose the team captain, and the remaining $k - 1$ team members (left-hand side). Equivalently, we could first choose the k team members, and the choose the captain among those k (right-hand side).

Story Proof #3

- *Vandermonde's identity*:

$$\binom{m+n}{k} = \sum_{j=0}^{k} \left[\binom{m}{j} \binom{n}{k-j} \right]$$

- This is NOT easy algebraically. But the story proof is:
- Suppose there are *m* men and *n* women, from which a group of size *k* will be selected => $\binom{m+n}{k}$ possibilities. If there are *j* men, then there must be $k - j$ women. The right-hand side sums up the ways to make these selections for all the cases of *j*.

Story Proofs in HW

- We encourage you to use English and story proofs/explanations in your HW.
- The final answer is not as important as the ideas behind how you got there.
- It is not always clear just from the math that you took the correct approach.
- Adding in some explanations in English can help
- Think about how best to explain it to your grader so that he/she understands YOUR thought process.
- One other HW pitfall: do not use “run-on” math equations!

Last Word: Why Statistics?

Because it's Sexy!!!

"*I keep saying the* ***sexy*** *job in the next ten years will be* ***statisticians****…I do think those skills - of being able to access, understand, and communicate the insights you get from data analysis - are going to be extremely important.*"

- Hal Varian, Google's Chief Economist, in 2009