INTRODUCTION TO PROBABILITY

First we have to start with probability

- Probability is the "language of statistics."
- Some of the probability concepts we talk about may not seem relevant to regression, ANOVA, etc.
- I have tried to select only the *necessary* probability concepts required to support the statistical inference in the second part of the semester.

Statistics is “probability backwards”

- When studying probability, we know something about a population and say something about a (potential) sample.
- When doing statistics, we observe a sample and try to say something about a population.

Calculus

- To do “mathematical statistics”, you need to know some calculus.
- This might require you to do some review (on your own).
- Do your best!

Basic definitions

- PROBABILITY (in common conversation): a measure of belief that a future event will occur
- RANDOM EVENT: an event whose occurrence cannot be predicted with certainty
  - Toss of a coin
  - Rain tomorrow (in some specified area)
  - Patient’s response to treatment
  - Number of fatal car crashes (in some area)

How to approach probability?

- Subjectively
  - Everybody may have a different probability!
- Using relative frequency
  - Based on *data*!
  - For example, 240 of 300 patients respond to this treatment.
  - The probability YOU respond to this treatment is 240/300=0.8=80%
- Axiomatically
  - Mathematically principled
  - Will rely on some assumption

More formal definitions for probability

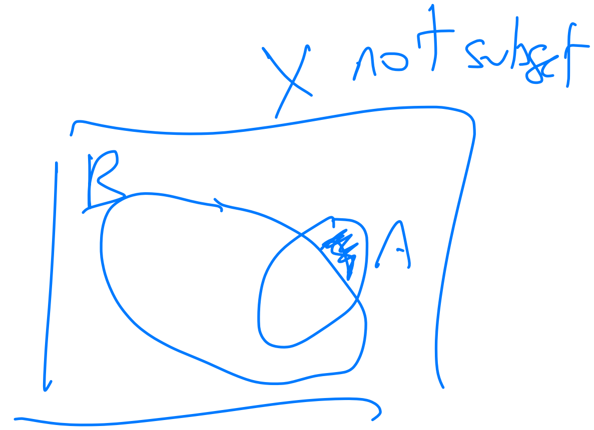
- EXPERIMENT (or RANDOM EXPERIMENT): process by which some observation is made
  (a) Coin toss
  (b) Systolic blood pressure of a patient
  (c) Number of tumors in a tissue sample
  (d) Number of accidental poisonings in some area
- EVENT (or OUTCOME): Result of an experiment
  (a) Tails
  (b) 110-115 mg Hg
  (c) 0 tumors
  (d) <20 poisonings

Relative frequency

- Useful when it's possible to conceive of many identical repetitions of an experiment.
- Example: with a "fair" (six-sided) die, $P(\{1\})=P(\{2\})=P(\{3\})=P(\{4\})=P(\{5\})=P(\{6\})=1/6$
- If you roll a die 30 times, you would expect about five of each.
- What if you roll a die 30 times and there are no 1's?
- Q: Is that <u>possible</u> if the die is fair?
- Q: If the die is really fair, how probable is such an outcome?
- Q: Would that outcome lead you to doubt whether the die is really fair?

Set notation

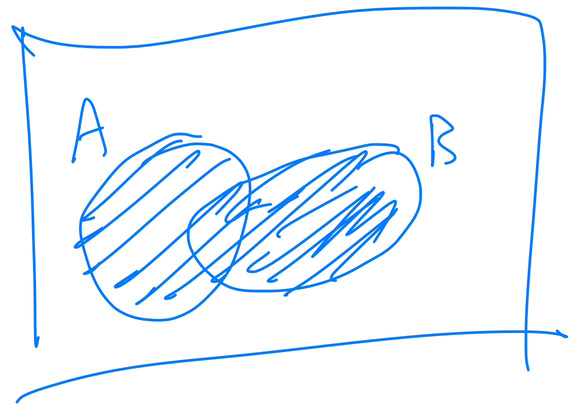
- We can only calculate probabilities on sets (of possible outcomes), so we need to review some basic concepts.
- We will use capital letters to denote sets.
- Example: $A=\{a_1, a_2, a_3, a_4\}$
- Two "special" sets:
  - S is the "universal" set – in probability, S is the collection of all possible outcomes.
  - $\varnothing$ (or $\{\}$) is the "null" or "empty" set.

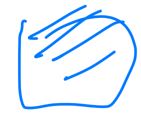


Subsets

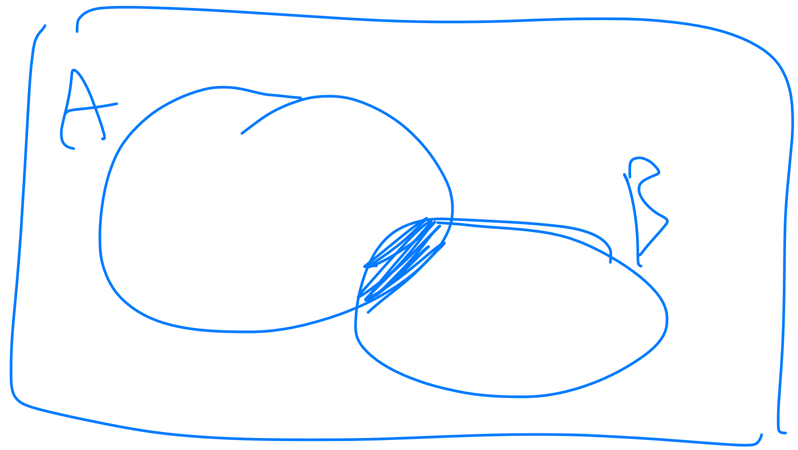
- SUBSET: A is a subset of B if every element of A is also in B.
- Example: A={a_1, a_2, a_3, a_4}, B={a_2, a_4}, C= {a_1, a_4, a_5}.
  - Q: Is B a subset of A? Yes
  - Q: Is C a subset of A? No
  - Q: Is B a subset of C? No
  - Q: Is ∅ a subset of A? Yes

Unions



 $A \cup B$

- UNION: The union of two sets A and B is the set that contains all elements that are either in A *or* in B (or in both).
- Notation: The union of A and B is denoted A ∪ B.
- Example: A={a_1, a_2, a_3, a_4}, B={a_2, a_4}, C= {a_1, a_4, a_5}.
  - Q: What is A ∪ B? $\{a_1, a_2, a_3, a_4\} = A$
  - Q: What is B ∪ C?
  - Q: What is C ∪ Ø?



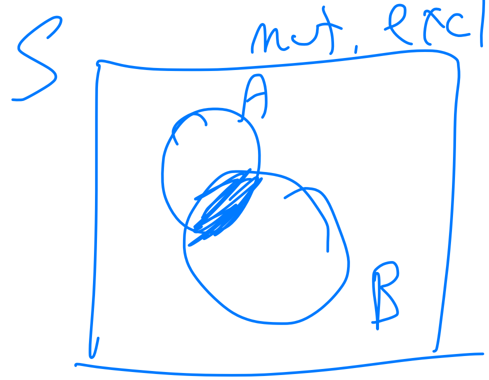
Intersections

- <u>INTERSECTION</u>: The intersection of two sets is the set of all elements that are both in A *and* in B.
- Notation: The intersection of A and B is denoted $A \cap B$.
- Example: A={a_1, a_2, a_3, a_4}, B={a_2, a_4}, C= {a_1, a_4, a_5}.
  - Q: What is $A \cap B$? $\{a_2, a_4\} = B$
  - Q: What is $B \cap C$?
  - Q: What is $C \cap \emptyset$?

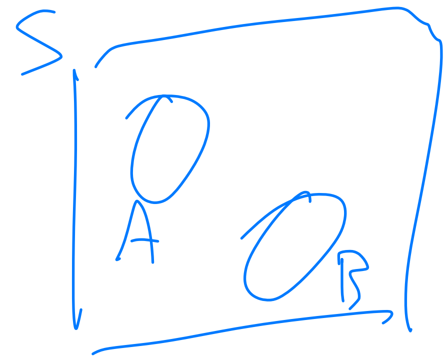
Complements

- <u>COMPLEMENT</u>: If A is a subset of S, then the complement of A is the set of all elements that are in S but *not* in A.
- Notation: The complement of A is denoted $\bar{A}$.
- Example: Let S={$\cancel{a_1}$, a_2, $\cancel{a_3}$, a_4, $\cancel{a_5}$, a_6, a_7}.
  - Q: What is the complement of {a_1, a_3, a_5}? $\{a_2, a_4, a_6, a_7\}$
  - Q: What is the complement of S? $\emptyset$
  - Q: What is the complement of Ø? S

A & B not mut. excl.



A&B mut. excl.



Mutual exclusivity

- <u>MUTUALLY EXCLUSIVE</u>: Sets A and B are mutually exclusive if $A \cap B = \emptyset$.
- Mutually exclusive sets have no elements in common.

Example: rolling a single die

$A = \{1, 2\}$

$B = \{1, 3, 5\}$

$C = \{2, 4, 6\}$

- We will roll a die one time.
- Q: What is S? $S = \{1, 2, 3, 4, 5, 6\}$
- Let A={outcome is less than 3}, B={outcome is odd}, C={outcome is even}. events
- Q: What is $A \cap B$? $\{1\}$
- Q: What is $A \cap C$? $\{2\}$
- Q: What is $A \cup C$? $\{1, 2, 4, 6\}$
- Q: What is $\bar{A}$? $\{3, 4, 5, 6\}$
- Q: Are there any mutually exclusive pairs? Yes B&C are mut. exc!

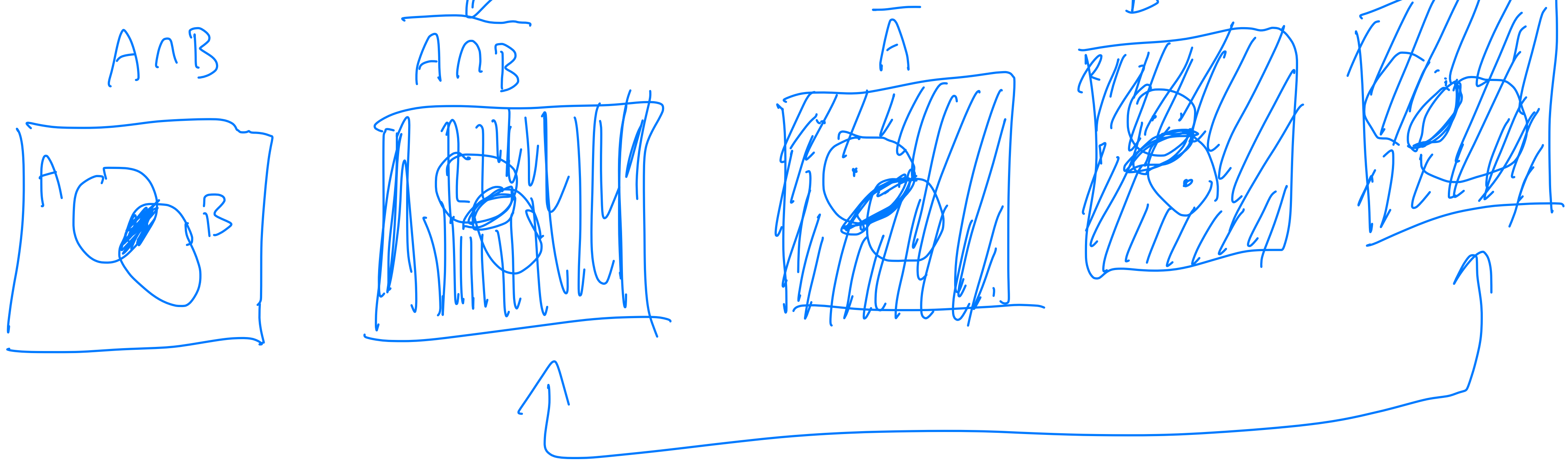
Some fundamental laws

- Distributive laws:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

- De Morgan's laws:

$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$
$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$



Probability model for discrete outcomes

- DISCRETE SAMPLE SPACE: If the sample space S contains a (finite or) countable number of outcomes, it is discrete.
- Examples:
  - Roll a single die. S= $\{1, 2, 3, 4, 5, 6\}$
  - Toss a coin repeatedly until observing “heads”, count the number of tosses. S= $\{1, 2, 3, 4, 5, \ldots\}$ "countable" $\Rightarrow$ discrete
  - Count the number of positive COVID-19 cases for people who appeared during a specified 1-hour period at a testing center. S= $\{0, 1, 2, 3, \ldots, 10{,}000{,}000\}$

Events

- We previously defined event as the outcome of a (random) experiment.
- We refine that now
- EVENT: A *set* of some possible outcomes of a random experiment.
- Notation: We will use capital letters for events (and use set notation).
- SIMPLE EVENT: An event that has no subsets (other than itself and the null set), i.e., a set with exactly one possible outcome)
- COMPOUND EVENT: An event with two or more possible outcomes.

Example: Rolling a single die

- Roll one die
- Let A={odd number} $= \{1, 3, 5\}$ compound
- Let B={2} simple

Example: Tossing a coin

- Toss a coin repeatedly until "heads" is observed, count the number of tosses $S = \{1, 2, 3, 4 \ldots \}$
- Let C="heads on first toss" $C = \{1\}$ simple
- Let D="heads before third toss" $D = \{1, 2\}$ compound

$H \rightarrow 1$

$TH \rightarrow 2$

$TTH \rightarrow 3$

$\{H\}$

$\{H, TH\}$