STAT 100: Week 8

Ricky’s Section

Introductions and Attendance

Introduction: Name

Question of the Week: If you could add one consistent item to HUDS, what would it be?

Important Reminders

Anonymous Feedback

https://docs.google.com/forms/d/e/1FAIpQLSfKv_FGvsooqm-IvtxKx3Vf6bBzSJE2jamK1gklAzL6NkXE8w/viewform

Midterm Recap

Overall...

- We're very proud of all your learning/growth!
  - Many of you came in with no coding background
- Be sure you're making the most of all the resources
  - Workshop, section, OH, 1-on-1 OH, Slack
- Improvement is taken into consideration

Some Important Things I Noticed

- Set a timer on the Oral Component
- Don't "over-explain" an answer—just say what you need
- **The p-value formula**: "If {the null hypothesis were true}, then the probability of observing {our test statistic} or {more extreme} would be {p-value}."
  - *Ex: If the coin was fair so that the true probability of heads is 50%, then the probability of observing 80% heads of more would be 0.01. Because 0.01 is such a low probability, we have evidence to reject the null hypothesis.*

Debrief

- Thoughts? Questions? Comments? Concerns?
- Anything you found surprising?
- Any concepts you want me to go over?
- We also can skip this if everyone just wants to move on

Content Review: Week 8

Foundations of Probability

- **Probability**: A value between 0 and 1 (intuitively, a "long-term frequency")
  - **Naive probability** is all favorable outcomes / all possible outcomes
  - *Ex: Probability of getting dealt an ace is 4/52 = 0.077*
- **Outcome**: Result after conducting an **experiment**
  - *Ex: After the experiment, I get dealt the ace of hearts*
- **Sample space**: Set of all possible **outcomes** of **experiment**
  - *Ex: There are 52 cards I could've been dealt*
- **Event**: Collection of **outcomes**
  - *Ex: The event I get dealt an ace is the collection of 4 specific outcomes*
  - *If A = the event I get dealt an ace, then P(A) = 0.077*

More on Probability

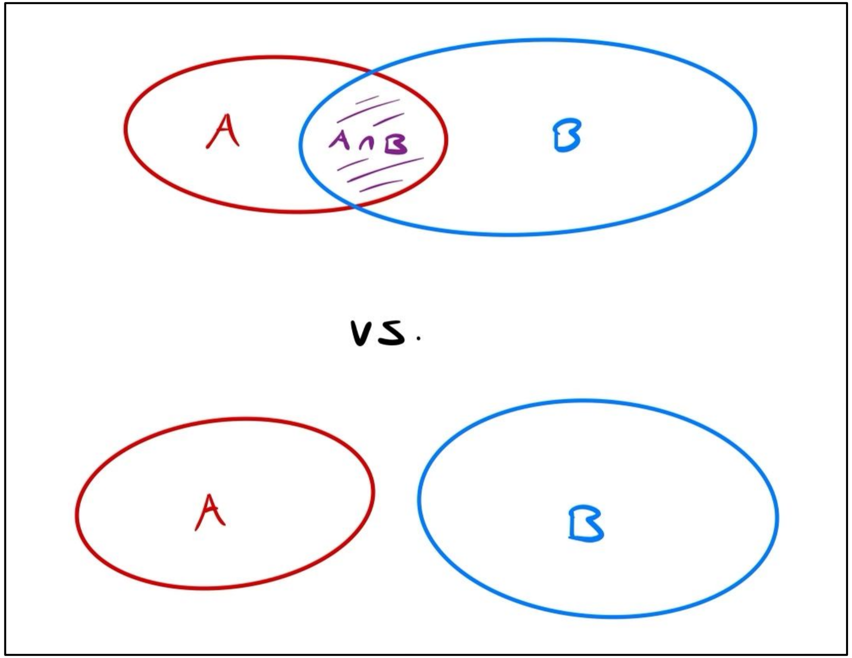
- **Disjoint events**: Events that CANNOT occur at the same time
  - *Ex: The event I get dealt an ace and the event I get dealt a king are disjoint*
  - *Now, the event I get dealt an ace and the event I get dealt a red card are NOT disjoint. Why?*
- **Independent events**: Knowing one event happens gives no info on the other
  - *Ex: If I flip a fair coin twice, the event I get heads on the 1st flip and the event I get heads on the 2nd flip are independent*
  - *Now, the event I get dealt an ace and the event I get dealt another ace afterwards (assuming no reshuffling) are NOT independent. Why?*
- **Conditional probability**: P(A | B) is probability of A, knowing B occured
  - *Ex: Given I got dealt a red card, what is the probability I got dealt the ace of hearts? It's NOT 1/52 anymore. Why?*

Our Toolkit

- **Union**: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- **Intersection**: $P(A \cap B) = P(A)\, P(B \mid A) = P(B)\, P(A \mid B)$
- **Complement Rule**: $P(A) = 1 - P(A^C)$, $P(A \mid B) = 1 - P(A^C \mid B)$
- **Def. of Conditional Probability**: $P(A \mid B) = P(A \cap B) / P(B)$
- **Bayes' Rule**: $P(A \mid B) = P(B \mid A)\, P(A) / P(B)$
- **LOTP**: $P(A) = P(A \mid B)\, P(B) + P(A \mid B^C)\, P(B^C)$

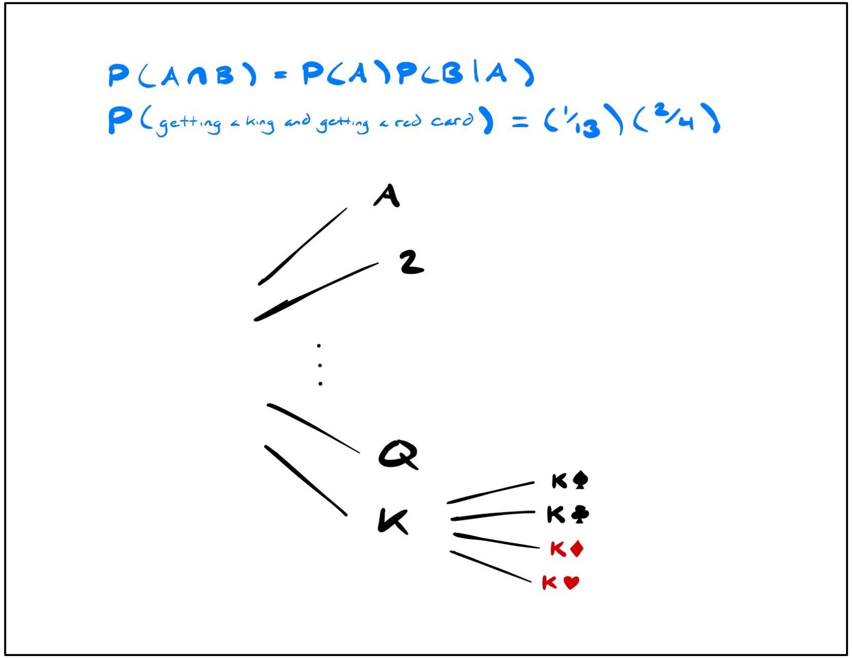
Union and Disjoint Events

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - *Ex: Probability of getting a king or a red card is 4/52 + 26/52 - 2/52 = 28/52*
- What happens if A and B are **disjoint**?
  - $P(A \cap B) = 0$ because A and B cannot occur simultaneously
  - Thus, for **disjoint events**, $P(A \cup B) = P(A) + P(B)$
- Use Venn diagrams



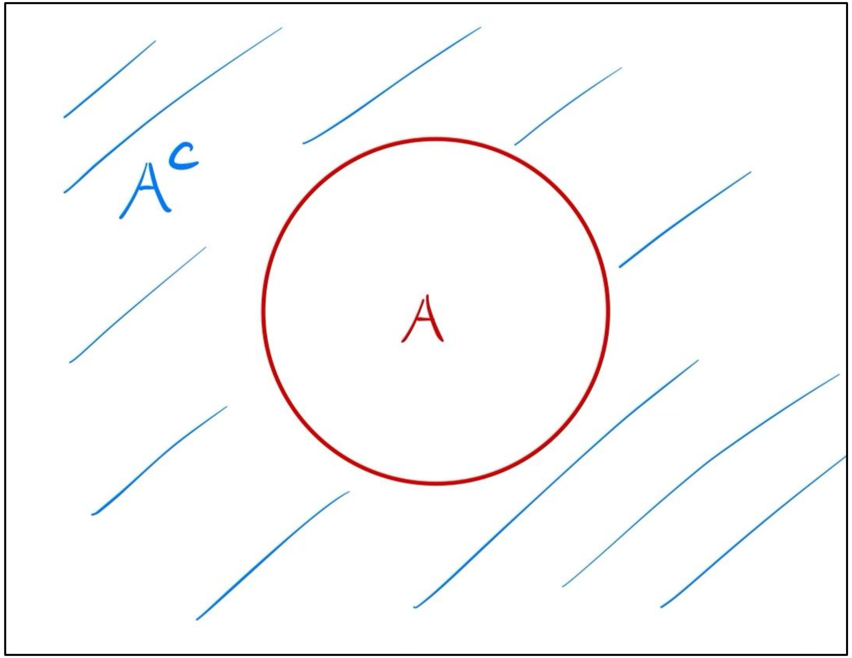
Intersection and Independent Events

- $P(A \cap B) = P(A)\,P(B \mid A) = P(A \mid B)\,P(B)$
    - *Ex: Probability of getting a king and a red card is (1/13)(2/4) = 2/52*
- What happens if A and B are **independent**?
    - $P(B \mid A) = P(B)$ because A happening gives no information on B
    - Thus, for **independent events**, $P(A \cap B) = P(A)\,P(B)$
- $P(A \cap B)$ is $P(A)$ and $P(B$, given $A)$



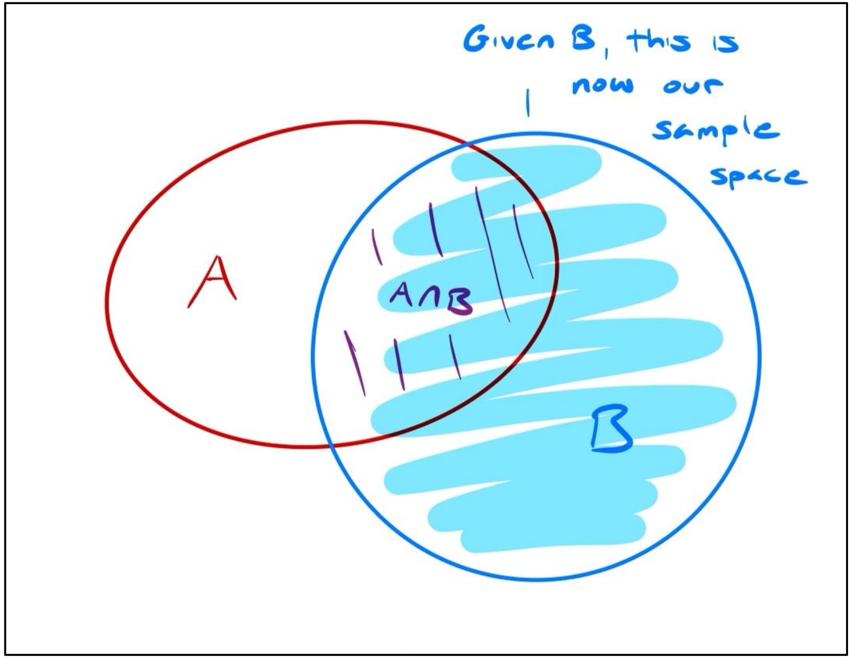
Complement Rule

- $P(A) = 1 - P(A^C)$
  - *Ex: Probability of rolling a 6 (⅙) is 1 minus probability of NOT rolling a 6 (1 - ⅚)*
  - This is because all possible outcomes in sample space sum to 1
- **Conditional probabilities** are still probabilities, so…
  - $P(A \mid B) = 1 - P(A^C \mid B)$
- Use Venn diagrams



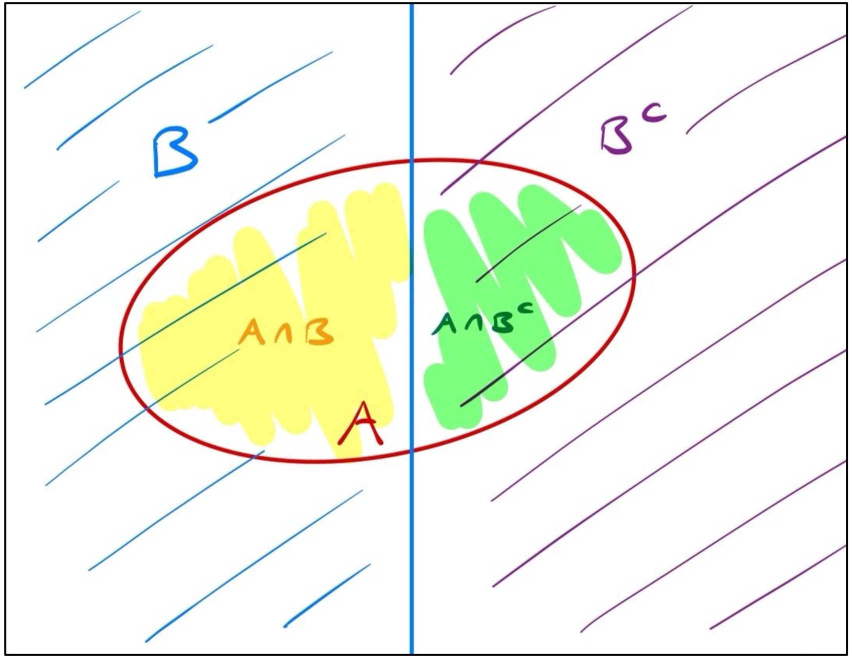
Conditional Probability

- Two ways to tackle this
- $P(A \mid B) = P(A \cap B) / P(B)$ by **definition**
  - **Conditioning** on B means we now live in B—it's our new sample space
  - Divide by P(B) so the total prob. is 1
- $P(A \mid B) = P(B \mid A)\, P(A) / P(B)$ by **Bayes' Rule**
  - Remember the definition of **intersection** from two slides ago?
- Use Venn diagrams



Law of Total Probability (LOTP)

- $P(A) = P(A \mid B)\, P(B) + P(A \mid B^C)\, P(B^C)$
- What are the two ways P(A) can happen?
  - We can partition the **sample space** with B and B^C
  - $P(A) = P(A \cap B)$ or $P(A \cap B^C)$
  - By **intersection** and **disjoint**, $P(A) = P(A \mid B)\, P(B) + P(A \mid B^C)\, P(B^C)$
- Use Venn diagrams



Recapping Our Toolkit: Notes

- **Union**: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
    - For **disjoint** events, $P(A \cup B) = P(A) + P(B)$ because $P(A \cap B) = 0$
- **Intersection**: $P(A \cap B) = P(A)\, P(B \mid A) = P(B)\, P(A \mid B)$
    - For **independent** events, $P(A \cap B) = P(A)\, P(B)$ because $P(A \mid B) = P(A)$
- **Complement Rule**: $P(A) = 1 - P(A^C)$, $P(A \mid B) = 1 - P(A^C \mid B)$
    - Use when you see "**at least**" (e.g., "Find the probability of rolling a 5+ at least once in 3 rolls")
- **Def. of Conditional Probability**: $P(A \mid B) = P(A \cap B) / P(B)$
- **Bayes' Rule**: $P(A \mid B) = P(B \mid A)\, P(A) / P(B)$
- **LOTP**: $P(A) = P(A \mid B)\, P(B) + P(A \mid B^C)\, P(B^C)$
    - Use for **wishful thinking** (e.g., "I really wish I knew which factory the cone came from")
- In general with probability, **start by defining events**

What are the 2 main strategies for finding unconditional probability, such as P(A)?

Question:

What are the 2 main strategies for finding unconditional probability, such as P(A)?

Complement rule and LOTP.

We often use complement for "at least" (e.g., "Find the probability of rolling a 5+ at least once in 3 rolls").

We often use LOTP for "wishful thinking" (e.g., "I really wish I knew which factory the cone came from").

One More Probability!

- **Positive predictive value (PPV)**: In a diagnostic test, the probability that a person has the disease, given that they tested positive for it (**true positive**)
- PPV = $P(D \mid T^+)$, where D is event of having disease and T^+ is event of testing positive

$$P(D|T^+) = \frac{P(D \cap T^+)}{P(T^+)} = \frac{P(D \cap T^+)}{P(D \cap T^+) + P(D^C \cap T^+)} = \frac{P(T^+|D)P(D)}{P(T^+|D)P(D) + P(T^+|D^C)P(D^C)}$$

$$= \frac{(\text{sens})(\text{prev})}{(\text{sens})(\text{prev}) + (1 - \text{spec})(1 - \text{prev})}$$

If I roll a fair 6-sided dice twice, what is the probability that I land 5 or higher at least once?

Question:

If I roll a fair 6-sided dice twice, what is the probability that I land 5 or higher at least once?

By **complement rule**, P(rolling 5+ at least once) = 1 - P(not rolling 5+ either turn).

Since rolls are **independent**, P(not rolling 5+ either turn) = P(not rolling 5+ on a turn) × P(not rolling 5+ on a turn).

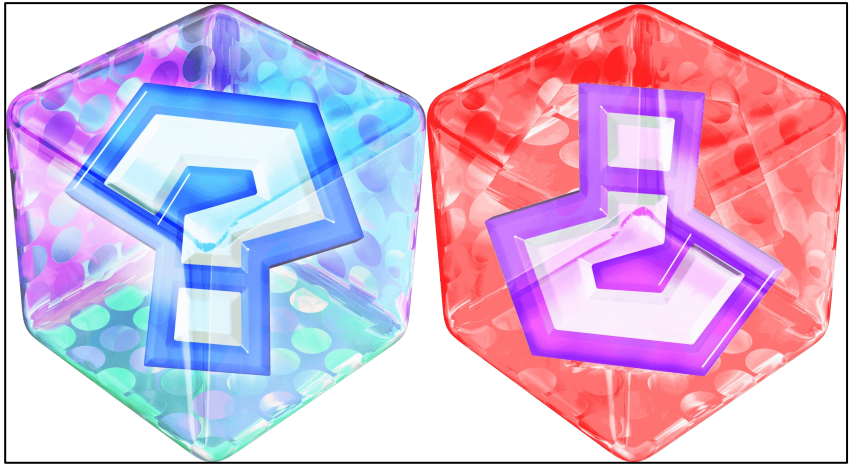
We know P(not rolling 5+ on a turn) = 4/6 = ⅔. Thus, 1 - (⅔)(⅔) = **5/9.**

Random Variables

- **Random variable**: A function that maps each **event** in the **sample space** to a number
- Intuitively, think of a r.v. as an unknown value that "crystallizes" to a certain number AFTER an **experiment**
  - *Ex: X is a r.v. for the number of heads I get after flipping 10 coins. X could be 0, 1, …, or 10. After the experiment, it "crystallizes" to one of those numbers.*

A Silly (but Useful) Intuition for Random Variables

- Think of **random variables** as mystery boxes in Mario Kart
- It's unknown what it will crystallize to, but we can still describe the random variable will probabilities
  - For example, there's a pretty low probability this random variable will crystallize to a bullet bill



Two Types of Random Variables

- **Discrete r.v.s**: Can crystallize to countable numbers
- Usually, **discrete r.v.s** are counted
  - *Ex: The number of people who show up to a party tomorrow (could be 0, 1, 2, …)*

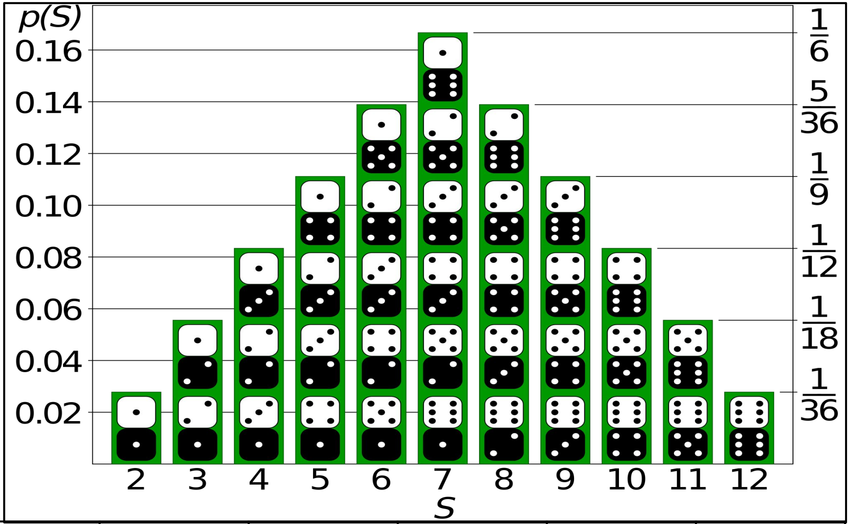
- **Continuous r.v.s**: Can crystallize to any real number in an interval
- Usually, **continuous r.v.s** are measured
  - *Ex: The temperature at noon tomorrow (could be any real number above absolute zero)*

Probability Distributions

- **Probability distributions**: Functions that give probabilities of all possible **outcomes** for a **r.v.**
  - Intuitively, it describes a **r.v.** through its probabilities
  - We can learn a lot about a **r.v.** by its **probability distribution**
- For **discrete r.v.s**, we use **Probability Mass Functions (PMFs)**
  - $f(x) = P(X = x_i)$
  - "Probability of big X (r.v.) crystallizing to little x (a certain value)"
- For **continuous r.v.s**, we use **Probability Density Functions (PDFs)**
  - $f(x)$, where $P(a \leq X \leq b) = \int_a^b f(x)dx$
  - For continuous r.v.s, the probability of X crystallizing to a certain value is 0, so we're concerned with X crystallizing to any value within some interval

PMFs for Discrete Random Variables

- **PMF**: $f(x) = P(X = x_i)$
  - "Probability of big X (r.v.) crystallizing to little x (a certain value)"
- **PMF must sum to 1**
  - Intuitively, all possible probabilities should sum to 1



x_i	2	3	4	5	6	7	8	9	10	11	12
$P(X = x_i)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Expected Value, Variance, and SD

- These are useful **summary statistics** to describe a **r.v.**
- **Expected value**: Weighted **mean** of a **r.v.**
    - $E(X) = \Sigma\, x_i\, P(X = x_i) = \mu$
    - We're weighing each possible crystallization by its probability
- **Variance**: Measure of **spread** of a **r.v.**
    - $Var(X) = \Sigma\, (x_i - \mu)^2\, P(X = x_i) = \sigma^2$
- **SD**: Average distance of all points from the **mean** of a **r.v.**
    - $SD(X) = \sqrt{Var(X)} = \sigma$

What is the expected value of a dice roll? Interpret the meaning in context.

Question:

What is the expected value of a dice roll? Interpret the meaning in context.

Formula: $E(X) = \Sigma\, x_i\, P(X = x_i)$.

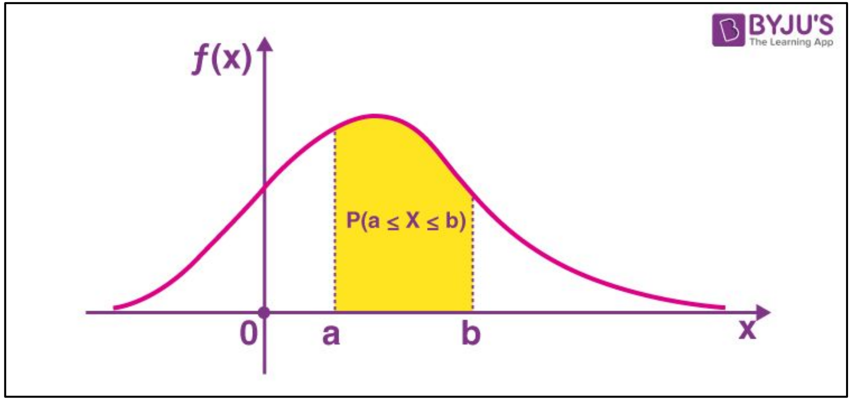
X, the **r.v.** for the value of a dice roll, can "crystallize" to 1, 2, 3, 4, 5, or 6 (with ⅙ probability of each).

$E(X) = 1(⅙) + 2(⅙) + 3(⅙) + 4(⅙) + 5(⅙) + 6(⅙) = 3.5$

This is the weighted **mean**. On average, we expect the value of our roll to be 3.5.

PDFs for Continuous Random Variables

- **Continuous r.v.s** are trickier because the probability X crystallizes to any one value is 0
- **PDF**: f(x), where $P(a \leq X \leq b) = \int_a^b f(x)dx$
- Intuitively and visually, think of **PDF** as **a shape whose area represents probability**
    - Thus, the area of the entire shape is 1
    - f(x) evaluated at any certain point is NOT probability; here, probability is AREA

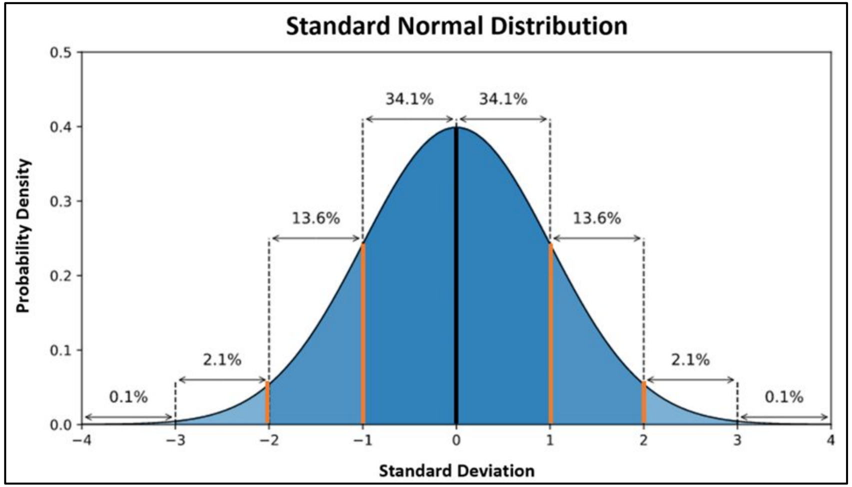


Special Types of Random Variables

- The really important types of **r.v.s** (which show up often) have names
- If your r.v. matches the "story" of a named random r.v., it makes your life easier
- X ~ Name(Values of Key Parameters)
  - *Ex: X ~ Bin(100, 10) is read as "X is distributed binomial with parameters 100 and 10"*
- **Parameters**: Named **r.v.s** are families, so **parameters** specify the **distribution** with a certain shape/center/spread

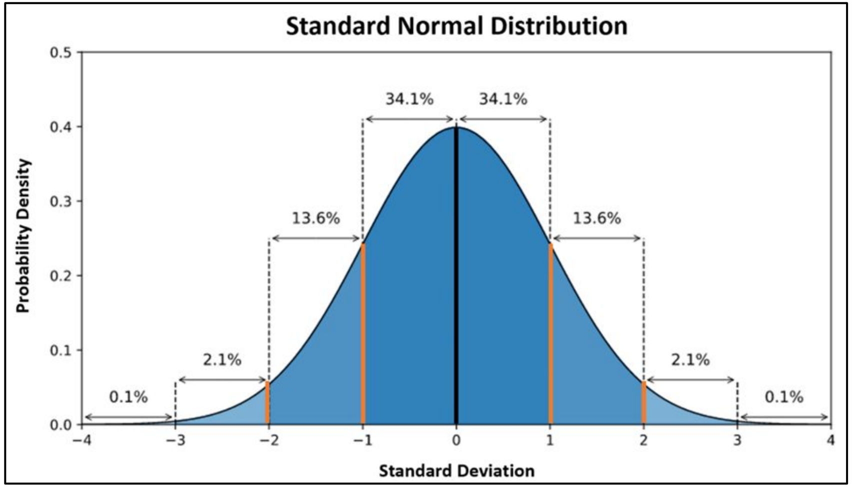
Normal Distribution

- **Normal distribution**: A symmetric and unimodal "bell shape" that approximates many **distributions**
- N(μ, σ) has 2 **parameters**
  - μ **is mean**
  - σ **is standard deviation**
- Z(0, 1) is **Standard Normal**
  - **0 is mean**
  - **1 is standard deviation**



Standardizing and Z-Scores

- **Standardizing**: Transforming **normal r.v. (X)** into **standard normal r.v. (Z)**
  - Comparing in terms of **Z-scores** (**standard deviations**) is easier
- **Z-score**: Measure of how many **SDs** the **sample statistic** is away from its **mean**
  - Z-score = $(X - \mu) / \sigma$
  - Z-score for test statistic = $(\text{statistic} - \mu) / \sigma$



The Important Functions for Normal Distribution

- **pnorm()**: Used to calculate **probabilities** on a **normal distribution** (often, for **p-value** during **hypothesis test**)
  - *Ex: What is the **probability** a student scores an 1800 on the SAT if the scores are N(1500, 300)?*
- pnorm(q = TEST-STAT, mean = MEAN, sd = STAN-DEV)
  - *Ex: pnorm(q = 1800, mean = 1500, sd = 300) = 0.8413447*
- **qnorm()**: Used to calculate **quantiles** on a **normal distribution** (often, for **critical value** during **confidence interval**)
  - *Ex: What score on the SAT would put a student in the 99th **quantile** (percentile)?*
- qnorm(p = QUANTILE, mean = MEAN, sd = STAN-DEV)
  - *Ex: qnorm(p = 0.99, mean = 1500, sd = 300) = 2197.904*

Why Does Any of This Matter?

- **Central Limit Theorem (CLT)**: For **random samples** and a **large sample size**, the **sampling distribution** of many **sample statistics** is approximately **normal**
  - Thus, when assumptions are met, we can conduct inference using the normal distribution as a good approximation
  - We will revisit inference next week through this lens!

In Closing...

- Probability is hard
  - Don't feel bad if this takes a bit to click
  - Probability is important, but it's not the focus of this course—after this p-set, it should be more chill
- If you're interested in more probability, consider STAT 110!

Questions?

P-Set 5

Have a great rest
of your week!