STA257 (Probability and Statistics I) Lecture Notes, Fall 2025

by Jeffrey S. Rosenthal, University of Toronto, www.probability.ca

(Last updated: September 10, 2025)

Note: I will update these notes regularly, posting them on the course web page each evening after lectures (though without annotations). However, they are just rough, point-form notes, with no guarantee of completeness or accuracy. They should in no way be regarded as a substitute for attending and learning from all the lectures, studying the course textbook, and doing the suggested homework exercises.

Introduction

• Course Information: See the course web page at: probability.ca/sta257

• Register for PollEverywhere: probability.ca/sta257/pollinfo.html

→ **USE YOUR REGULAR UofT EMAIL!**

• Who here is doing a specialist or major program involving: Statistics / Data Science? Mathematics? Actuarial Science? Computer Science? Economics/Commerce? Physics/Chemistry/Biology? Education? Psychology/Sociology? Engineering? Other?

• Who here has seen probabilities in elementary school? high school? STA130?

→ Don't worry, we will start from scratch. (Just need math and logic.)

• Life is full of randomness and uncertainty: lotteries, card games, computer games, gambling, weather, TTC, airplanes, friends, jobs, classes, science, finance, elections, diseases, safety/risk, demographics, internet routing, legal cases, ... whenever we're not sure of the outcome or what will happen next.

• Lots of interesting probability questions to solve! Such as ...

→ What's the probability you'll win the Lotto Max jackpot, i.e. that you will choose the correct 7 distinct numbers between 1 and 50?

→ If 200 students each flip a fair coin, then how many Heads is the most likely? How likely? What's the probability of more than 150 Heads?

→ If you repeatedly roll a fair 6-sided die [show], then how many rolls will there be on average before the first time you roll a 3?

→ At a party of 40 people, what is the probability that some pair of them have the same birthday?

→ If a disease affects one person in a thousand, and a test for the disease has 99% accuracy, and you test positive, then what is the probability you have the disease?

→ If you pick a number uniformly at random between 0 and 1, then what is the probability that you pick exactly the number 3/4?

→ Three-Card Challenge. [demonstration] What are the probabilities of the initial (front) colour? Then, what are the probabilities of the back colour?

• History of Mathematical Probability Theory (in brief):

→ Mathematics is very precise and certain. For thousands of years, it simply ignored the uncertainty of probabilities.

→ Then, in 1654, the French writer Antoine Gombaud (the "Chevalier de Méré") asked the mathematician Pierre de Fermat some gambling questions:

→ Which is more likely (or are they the same) (and are they more than 50%):
(a) Get at least one six when rolling a fair six-sided die 4 times; or
(b) Get at least one pair of sixes when rolling two fair six-sided dice 24 times?

→ He thought (a) was $4 \times (1/6) = 2/3$, and (b) was $24 \times (1/36) = 2/3$. Correct?

→ Also: (c) Suppose a gambler is playing a best-of-seven match, where whoever wins 4 (fair) games first in the winner, and so far they have won 3 times and lost 1, but then the match gets interrupted. What is the probability that they would have won the match, if it had been allowed to continue?

→ Fermat then corresponded with the mathematician Blaise Pascal to find solutions to these questions (later!), and mathematical probability theory was born!

POLL: If you have independent probability 1/2 of winning each game, and you are up 3 games to 1, what do you think is the probability that you will win 4 games first? **(A)** 1/2. **(B)** 2/3. **(C)** 3/4. **(D)** 7/8. **(E)** No idea. [Best guess only – later.]

- So, can probabilities be studied mathematically?

→ Can we use certain mathematics to study the uncertainty of probabilities?

→ Yes! That's why we're here! To be certain about our uncertainty!

→ But we have to define our terms carefully ...

Sample Space (§1.2) (i.e. Section 1.2 of the textbook)

- The first part of any probability model is the sample space, written S, which is the set of all possible outcomes.

→ e.g. flip a coin: $S = \{\text{Heads, Tails}\}$, or $S = \{H, T\}$.

→ e.g. flip a coin three times in a row:
$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.

→ Or, if we only care about the number of Heads: $S = \{0, 1, 2, 3\}$.

→ e.g. tonight's dinner: $S = \{\text{Beef, Chicken, Fish}\}$. (Assume one.)

→ e.g. the number of bees I will see on my walk home: $S = \{0, 1, 2, 3, \ldots\}$.

→ e.g. the price of IBM stock next month: $S = [0, \infty)$.

→ e.g. the height (in cm) of the next student I meet: $S = (0, \infty)$.

→ e.g. your grade in this class: $S = \{0, 1, 2, 3, \ldots, 100\}$.

→ e.g. roll one six-sided die: $S = \{1, 2, 3, 4, 5, 6\}$.

→ e.g. roll two six-sided dice: $S = \{1, 2, 3, 4, 5, 6\}^2$, i.e.
$S = \{11, 12, 13, 14, 15, 16,\ 21, 22, 23, 24, 25, 26,$
$31, 32, 33, 34, 35, 36,\ 41, 42, 43, 44, 45, 46,$
$51, 52, 53, 54, 55, 56,\ 61, 62, 63, 64, 65, 66\}$.

→ Or, if we only care about the sum, instead maybe take $S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.

→ e.g. "Pick any integer between 1 and 10": $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$.

→ e.g. "Pick any number between 0 and 1": $S = [0, 1]$. (important case!)

- Summary: The sample space S can be <u>any</u> non-empty set which contains <u>all</u> of the possible outcomes. Simple!

- But it gets more interesting when we also have …

Probabilities and Events (§1.2)

- An event A is "any" subset $A \subseteq S$.

- For any event A, we can define the probability $\mathrm{P}(A)$ that it will occur.

→ e.g. flip a "fair" coin: $\mathrm{P}(H) = \mathrm{P}(T) = 1/2$.

→ (Note: We often use e.g. "$\mathrm{P}(H)$" as shorthand for "$\mathrm{P}(\{H\})$", etc.)

→ e.g. roll a fair six-sided die: $\mathrm{P}(1) = \mathrm{P}(2) = \mathrm{P}(3) = \mathrm{P}(4) = \mathrm{P}(5) = \mathrm{P}(6) = 1/6$.

→ e.g. tonight's dinner: maybe P(Beef)=0.40, P(Chicken)=0.15, and P(Fish)=0.45.

→ (Note: We could also write P(Fish) = 45%, etc. Usually percentages are good for intuition, but pure probabilities (not percentages) are better for calculation.)

→ e.g. flip three fair coins: $\mathrm{P}(HHH) = \mathrm{P}(HHT) = \ldots = \mathrm{P}(TTT) = 1/8$.

→ e.g. roll two fair dice: $\mathrm{P}(11) = \mathrm{P}(12) = \ldots = \mathrm{P}(65) = \mathrm{P}(66) = 1/36$.

→ e.g. Pick any integer between 1 and 10. [Try it!]
<u>Could</u> be "uniform", i.e. $\mathrm{P}(1) = \mathrm{P}(2) = \ldots = \mathrm{P}(10) = 1/10$. Or <u>instead</u>, maybe …
P(3)=P(6)=P(7)=0.2, and P(5)=0.1, and P(1)=P(2)=P(4)=P(8)=P(9)=P(10)=0.05.

→ e.g. Pick any number between 0 and 1, "uniformly" ("Uniform[0,1]"):
$\mathrm{P}([0, 1/2]) = 1/2$, $\mathrm{P}([1/2, 1]) = 1/2$, $\mathrm{P}([0, 1/3]) = 1/3$, $\mathrm{P}([1/3, 2/3]) = 1/3$,
and in general $\mathrm{P}([a, b]) = b - a$ whenever $0 \le a \le b \le 1$. Diagram:

- Or maybe instead $\mathrm{P}([a, b]) = b^2 - a^2$ whenever $0 \le a \le b \le 1$. Valid?

- Or maybe instead $\mathrm{P}([a, b]) = (b - a)^2$ whenever $0 \le a \le b \le 1$. Valid?

Basic Properties of Probabilities (§1.2)

- Let's begin with a specific example (and then we will generalise):

- e.g. tonight's dinner, with P(Beef)=0.40, P(Chicken)=0.15, and P(Fish)=0.45.

→ Probability of Beef <u>or</u> Chicken = P({Beef, Chicken}) = P({Beef}) + P({Chicken}) $= 0.40 + 0.15 = 0.55$.

→ Probability of <u>any</u> dinner = Probability of Beef or Chicken or Fish = P({Beef, Chicken, Fish}) = P({Beef}) + P({Chicken}) + P({Fish}) $= 0.40 + 0.15 + 0.45 = 1$.

→ Probability dinner is <u>not</u> Beef <u>nor</u> Chicken <u>nor</u> Fish $= \mathrm{P}(\emptyset) = 0$.

- In <u>general</u>, certain properties <u>must</u> hold for <u>any</u> probability model ("axioms"):

- If A is an event, then $0 \le \mathrm{P}(A) \le 1$.
- If $A = S$ is the event corresponding to all outcomes, then $\mathrm{P}(A) = \mathrm{P}(S) = 1$.
- Or, if $A = \emptyset$ is the event corresponding to no outcomes, then $\mathrm{P}(A) = \mathrm{P}(\emptyset) = 0$.
- Additivity: If A and B are disjoint events (i.e. $A \cap B = \emptyset$), e.g. $A = \{\text{Beef}\}$ and $B = \{\text{Chicken}\}$, then $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$.
- More generally, if $A_1, A_2, A_3, \ldots$ are any sequence (finite or infinite) of disjoint events (i.e. $A_i \cap A_j = \emptyset$ whenever $i \ne j$), then $\mathrm{P}\left(\bigcup_i A_i\right) = \sum_i \mathrm{P}(A_i)$.

→ So, in particular, since $\mathrm{P}(S) = 1$, all of the probabilities have to add up to 1.

→ e.g. P(Heads) + P(Tails) = 0.5 + 0.5 = 1.

→ e.g. P(Beef) + P(Chicken) + P(Fish) = 0.40 + 0.15 + 0.45 = 1.

Suggested Homework: 1.2.1, 1.2.2, 1.2.3, 1.2.4, 1.2.8, 1.2.9, 1.2.10, 1.2.11, 1.2.12, 1.2.13, 1.2.14, 1.2.15.

END WEDNESDAY #1

Derived Properties of Probabilities (§1.3)

- Once we know the above properties, then we can use them to prove others too:
- Fact: If A^C is the complement of A, i.e. the set of all outcomes which are not in A, then $\mathrm{P}(A^C) = 1 - \mathrm{P}(A)$. (Important! Remember this! Use this!)

→ Proof: Note that A and A^C are disjoint, so $\mathrm{P}(A \cup A^C) = \mathrm{P}(A) + \mathrm{P}(A^C)$. But $\mathrm{P}(A \cup A^C) = \mathrm{P}(S) = 1$, so $1 = \mathrm{P}(A) + \mathrm{P}(A^C)$, i.e. $\mathrm{P}(A^C) = 1 - \mathrm{P}(A)$. ∎

→ e.g. P(Fish) = P(not Beef or Chicken) = 1 − P(Beef or Chicken) = 1 − 0.55 = 0.45.

- Fact: For any events A and B, $\mathrm{P}(A) = \mathrm{P}(A \cap B) + \mathrm{P}(A \cap B^C)$. (∗)

Diagram:

→ Proof: The events $A \cap B$ and $A \cap B^C$ are disjoint, and $(A \cap B) \cup (A \cap B^C) = A$, so by additivity, $\mathrm{P}(A \cap B) + \mathrm{P}(A \cap B^C) = \mathrm{P}(A)$. ∎

→ e.g. integer between 1 and 10: P(even) = P(even and ≤ 4) + P(even and ≥ 5) $= \mathrm{P}(\{2, 4\}) + \mathrm{P}(\{6, 8, 10\})$.

- Re-arranging (∗) also gives that: $\mathrm{P}(A \cap B^C) = \mathrm{P}(A) - \mathrm{P}(A \cap B)$. (∗∗)
- Fact: If $A \supseteq B$, then $\mathrm{P}(A) = \mathrm{P}(B) + \mathrm{P}(A \cap B^C)$. (∗∗∗)

→ Proof: This follows from (∗), since if $A \supseteq B$, then $A \cap B = B$. ∎

→ e.g. integer between 1 and 10: P(≤ 7) = P(≤ 4) + P(≤ 7 but ≥ 5).

- Monotonicity: If $A \supseteq B$, then $\mathrm{P}(A) \ge \mathrm{P}(B)$. (Remember this!)

→ Proof: We must have $P(A \cap B^C) \geq 0$, so from (∗∗∗), $P(A) = P(B) + P(A \cap B^C) \geq P(B) + 0 = P(B)$. ∎

→ e.g. $P(\{\text{Beef, Chicken}\}) = 0.55 \geq 0.40 = P(\{\text{Beef}\})$.

• Law of Total Probability – Unconditioned Version: Suppose $A_1, A_2, \ldots$ are a sequence (finite or infinite) of events which form a partition of S, i.e. they are disjoint ($A_i \cap A_j = \emptyset$ for all $i \neq j$) and their union equals the entire sample space ($\bigcup_i A_i = S$), and let B be any event. Diagram:

Then $P(B) = \sum_i P(A_i \cap B)$. That is: $P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \ldots$.

→ Proof: Since the $\{A_i\}$ are disjoint, and $A_i \cap B \subseteq A_i$, therefore the $\{A_i \cap B\}$ are also disjoint. Furthermore, since $\bigcup_i A_i = S$, therefore $\bigcup_i (A_i \cap B) = S \cap B = B$. Hence, $P(B) = P\Big(\bigcup_i (A_i \cap B)\Big) = \sum_i P(A_i \cap B)$. ∎

→ e.g. integer between 1 and 10: Suppose $A_1 = \{\leq 4\} = \{1,2,3,4\}$, and $A_2 = \{\geq 5\} = \{5,6,7,8,9,10\}$, and $B = \{\text{even}\} = \{2,4,6,8,10\}$. Then P(even) = P(even and ≤ 4) + P(even and ≥ 5), i.e. $P(\{2,4,6,8,10\}) = P(\{2,4\}) + P(\{6,8,10\})$.

• Principle of Inclusion-Exclusion: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

→ (Of course, if they're disjoint ($A \cap B = \emptyset$), then $P(A \cup B) = P(A) + P(B)$.)

→ Intuition: $P(A) + P(B)$ counts each element of $A \cap B$ twice, so we have to subtract one of them off.

→ Proof: The events $A \cap B$, and $A \cap B^C$, and $A^C \cap B$, are all disjoint, and their union is $A \cup B$. Diagram:

Hence, $P(A \cup B) = P(A \cap B) + P(A \cap B^C) + P(A^C \cap B)$.

But from (∗∗), $P(A \cap B^C) = P(A) - P(A \cap B)$ and $P(A^C \cap B) = P(B) - P(A \cap B)$.

Hence, $P(A \cup B) = P(A \cap B) + \big[P(A) - P(A \cap B)\big] + \big[P(B) - P(A \cap B)\big]$ $= P(A) + P(B) - P(A \cap B)$. ∎

→ e.g. integer between 1 and 10: P(even or ≤ 4) = P(even) + P(≤ 4) − P(even and ≤ 4) = $P(\{2,4,6,8,10\}) + P(\{1,2,3,4\}) - P(\{2,4\})$.

→ Or, P(even or perfect square) = P(even) + P(perfect square) − P(even and perfect square) = $P(\{2,4,6,8,10\}) + P(\{1,4,9\}) - P(\{4\})$.

• Optional: A more general Inclusion-Exclusion formula is in Challenge 1.3.10.

• Now, $P(A \cap B) \geq 0$, so $P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$. (!)

• Subadditivity: For any sequence of events $A_1, A_2, \ldots$, not necessarily disjoint, we still always have $P(A_1 \cup A_2 \cup \ldots) \leq P(A_1) + P(A_2) + \ldots$.

→ (Of course, it would be equal if they are disjoint.)

→ Proof (§1.7): Let $B_1 = A_1$, and $B_2 = A_2 \cap (A_1)^C$, and $B_3 = A_3 \cap (A_1 \cup A_2)^C$, and $B_4 = A_4 \cap (A_1 \cup A_2 \cup A_3)^C$, and so on. (That is, each new B_n is the part of A_n which is not already part of $A_1, \ldots, A_{n-1}$.) Diagram:

Then the $\{B_i\}$ are disjoint by construction, and $\bigcup_i B_i = \bigcup_i A_i$.
[Formally, the above construction ensures that $\bigcup_{i=1}^n B_i = \bigcup_{i=1}^n A_i$ for each finite n. Then, in the infinite case, $\bigcup_{i=1}^\infty B_i = \bigcup_{n=1}^\infty (\bigcup_{i=1}^n B_i) = \bigcup_{n=1}^\infty (\bigcup_{i=1}^n A_i) = \bigcup_{i=1}^\infty A_i$.]
Also $B_i \subseteq A_i$ so $\mathrm{P}(B_i) \le \mathrm{P}(A_i)$. Hence, $\mathrm{P}(A_1 \cup A_2 \cup \ldots) = \mathrm{P}(B_1 \cup B_2 \cup \ldots) = \mathrm{P}(B_1) + \mathrm{P}(B_2) + \ldots \le \mathrm{P}(A_1) + \mathrm{P}(A_2) + \ldots$. ∎

→ Alternative proof (for a finite number of events): Use induction! For $n = 2$ events, this follows from Inclusion-Exclusion. Then for $n \ge 3$ events, $\mathrm{P}(A_1 \cup \ldots \cup A_n) = \mathrm{P}\big((A_1 \cup \ldots \cup A_{n-1}) \cup A_n\big)$, which by Inclusion-Exclusion is $\le \mathrm{P}\big(A_1 \cup \ldots \cup A_{n-1}\big) + \mathrm{P}(A_n)$, which by induction is $\le \big(\mathrm{P}(A_1) + \ldots + \mathrm{P}(A_{n-1})\big) + \mathrm{P}(A_n)$. ∎

→ e.g. integer between 1 and 10: P(even or ≤ 4) $\le$ P(even) + P(≤ 4), i.e. $\mathrm{P}(\{1,2,3,4,6,8,10\}) \le \mathrm{P}(\{2,4,6,8,10\}) + \mathrm{P}(\{1,2,3,4\})$.

[Note that we do not have "uncountable" subadditivity, e.g. for uniform on $S = [0,1]$, if $A_x = \{x\}$ for each $x \in S$, then $\mathrm{P}(\bigcup_{x \in S} A_x) = \mathrm{P}(S) = \mathrm{P}([0,1]) = 1$, even though $\mathrm{P}(A_x) = \mathrm{P}(\{x\}) = 0$ for each individual $x \in S$, so also $\sum_{x \in S} \mathrm{P}(A_x) = \sum_{x \in S}(0) = 0$.]

Suggested Homework: 1.3.1, 1.3.2, 1.3.3, 1.3.4, 1.3.5, 1.3.7, 1.3.8, 1.3.9.

Uniform Probabilities on Finite Spaces (§1.4)

• Suppose $S = \{s_1, s_2, \ldots, s_n\}$ is some finite sample space, of finite size $|S| = n$, and each element is equally likely.

→ Then $\mathrm{P}(s_1) = \mathrm{P}(s_2) = \ldots = \mathrm{P}(s_n) = 1/n$. ("discrete uniform distribution")

→ And for any event $A = \{a_1, a_2, \ldots, a_k\}$, by additivity we have

$$\mathrm{P}(A) \;=\; \mathrm{P}(a_1) + \mathrm{P}(a_2) + \ldots + \mathrm{P}(a_k) \;=\; \frac{1}{n} + \frac{1}{n} + \ldots + \frac{1}{n} \;=\; \frac{k}{n} \;=\; \frac{|A|}{|S|}\,.$$

→ So, in this case, we just need to count the number of elements in A, and divide that by the number of elements in S. Easy!?! Sometimes!

• e.g. Roll a fair six-sided die. What is $\mathrm{P}(\ge 5)$?

→ Here $S = \{1,2,3,4,5,6\}$ so $|S| = 6$. All equally likely.

→ Also $A = \{5,6\}$ so $|A| = 2$.

→ So, $\mathrm{P}(\ge 5) = \mathrm{P}(A) = |A| \,/\, |S| = 2/6 = 1/3$. Easy!

• Flip two fair coins. What is P(# Heads = 1)?

POLL: **(A)** 1/4. **(B)** 1/3. **(C)** 1/2. **(D)** 3/4. **(E)** 1. **(F)** No idea.

→ Here $S = \{HH, HT, TH, TT\}$, all equally likely. So, $|S| = 4$.

→ And, $A = \{HT, TH\}$. So, $|A| = 2$.

→ Hence, $\mathrm{P}(A) = |A| \,/\, |S| = 2/4 = 1/2$. Easy!

- e.g. Roll one fair six-sided die, and flip two fair coins.

What is P(# Heads = Number Showing On The Die)? (Best guess?)

POLL: **(A)** 1/6. **(B)** 1/8. **(C)** 1/12. **(D)** 1/16. **(E)** 1/24. **(F)** No idea.

→ Here $S = \{1HH, 1HT, 1TH, 1TT, 2HH, \ldots, 6TT\}$. All equally likely.

→ But what is $|S|$?

→ Multiplication Principle: If S is made up by choosing one element of each of the subsets $S_1, S_2, \ldots, S_k$, i.e. if $S = S_1 \times S_2 \times \ldots \times S_k$, then what is $|S|$? Well, ... $|S| = |S_1|\,|S_2|\, \ldots\, |S_k|$.

→ In our example, $S_1 = \{1, 2, 3, 4, 5, 6\}$, and $S_2 = \{H, T\}$, and $S_3 = \{H, T\}$, so $|S| = |S_1|\,|S_2|\,|S_3| = 6 \cdot 2 \cdot 2 = 24$.

→ And what about A? Well, think about the possibilities ...
$A = \{1HT, 1TH, 2HH\}$. (No other combination works. Why?) So, $|A| = 3$.

→ Hence, P(# Heads = Number Showing On The Die) $= |A| \,/\, |S| = 3/24 = 1/8$.

→ [Alternatively (later): $(1/6)(1/2)+(1/6)(1/4) = (1/12)+(1/24) = 3/24 = 1/8$.]

- e.g. Roll three fair six-sided dice. What is $\mathrm{P}(\text{sum} \geq 17)$?

→ Here $S = \{1, 2, 3, 4, 5, 6\}^3$ so $|S| = 6^3 = 216$. All equally likely.

→ But what is A? Think about it ...
Here $A = \{666, 566, 656, 665\}$ (why?), so $|A| = 4$.

→ So, $\mathrm{P}(\text{sum} \geq 17) = \mathrm{P}(A) = |A| \,/\, |S| = 4/216 = 1/54$.

→ Exercise: What about $\mathrm{P}(\text{sum} \geq 16)$? $\mathrm{P}(\text{sum} \geq 15)$?

- Chevalier de Méré's historical 1654 questions:

- (a) What is P(get at least one six when rolling a fair six-sided die 4 times)?

→ Here $S = \{1, 2, 3, 4, 5, 6\}^4$, so $|S| = 6^4 = 1296$. All equally likely.

→ And what is $|A|$? Tricky. Easier to consider ...

→ $A^C = \{\text{no sixes in four rolls}\} = \{1, 2, 3, 4, 5\}^4$, so $|A^C| = 5^4 = 625$.

→ So, $\mathrm{P}(A^C) = |A^C| \,/\, |S| = 5^4 \,/\, 6^4 = 625 \,/\, 1296 \doteq 0.482$.

→ So, $\mathrm{P}(A) = 1 - \mathrm{P}(A^C) \doteq 1 - 0.482 = 0.518$. More than 50%.

→ (Alternatively: By "independence" [later], $\mathrm{P}(A) = 1 - (5/6)^4 \doteq 0.518$.)

- (b) What is P(get at least one pair of sixes when rolling a pair of fair six-sided dice 24 times)?

→ Here $S = \left(\{1, 2, 3, 4, 5, 6\}^2\right)^{24}$, so $|S| = (6^2)^{24} = 6^{48}$ ($>10^{37}$). All equally likely.

→ And what is $|A|$? Tricky. Again, easier to consider ...

→ $A^C = \{\text{no pair of sixes in 24 rolls}\} = \{11, 12, 13, \ldots, 64, 65\}^{24}$, so $|A^C| = 35^{24}$.

→ So, $P(A^C) = |A^C| \,/\, |S| = 35^{24}/6^{48} \doteq 0.509$.

→ So, $P(A) = 1 - P(A^C) \doteq 1 - 0.509 = 0.491$. Less than 50%.

→ (Again, alternatively by independence [later], $P(A) = 1 - (35/36)^{24} \doteq 0.491$.)

Suggested Homework: 1.4.1, 1.4.9, 1.4.10, 1.4.11, 1.4.12, 1.4.13.

END MONDAY #1

• (c) In a best-of-seven match with fair (50%) games, if a player has won 3 games and lost 1, then what is the probability they will win the match?

→ Various paths to victory: win right away, lose then win, etc. Tricky.

→ One solution: Pretend 3 more games will always be played. (Result unchanged.)

→ Then $S = \{\text{Win, Lose}\}^3$, so $|S| = 2^3 = 8$, all equally likely.

→ What about A? Well, here $A^C = \{\text{Lose, Lose, Lose}\}$, so $|A^C| = 1$.

→ Hence, $P(A^C) = |A^C|/|S| = 1/8$, and so $P(A) = 1 - P(A^C) = 7/8$.

→ Exercise: What if the player has won just 2 games and lost 1? (Trickier.)

Warning about Non-Uniform Probabilities

• e.g. Roll two fair dice. What is P(sum is ≤ 3)?

→ POSSIBLE SOLUTION: The sum is in $S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. So, $|S| = 11$. And, the event "≤ 3" corresponds to $A = \{2, 3\}$, so $|A| = 2$. Hence, P(sum is $\le 3) = |A|/|S| = 2/11$. Right?

→ WRONG! These sums are not all equally likely, i.e. it is not uniform! So, $P(A) \ne |A|/|S|$. That formula is only when all outcomes are equally likely. Important!

→ INSTEAD: Let $S = \{\text{all ordered pairs of two dice}\}$, i.e. $S = \{11, 12, 13, \ldots, 65, 66\}$. Then $|S| = 36$. Now each outcome in S is equally likely. And, now $A = \{11, 12, 21\}$. So, $P(A) = |A|/|S| = 3/36 = 1/12$. Correct!

• Also, note that sometimes the sample space S is a discrete infinite set:

→ e.g. $S = \mathbf{N} := \{1, 2, 3, \ldots\}$, with $P(i) = 2^{-i}$ for each $i \in S$.

→ Valid? Yes, since $2^{-i} \ge 0$, and $\sum_{i=1}^{\infty} 2^{-i} = \frac{2^{-1}}{1-2^{-1}} = 1$. (Geometric series.)

→ Then e.g. P(Even Number) $= \sum_{i=2,4,6,\ldots} 2^{-i} = \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \ldots = \frac{1/4}{1-(1/4)} = 1/3$.

→ And, $P(\le 10) = \sum_{i=1}^{10} 2^{-i} = \frac{2^{-1}-2^{-11}}{1-2^{-1}} = \frac{(1/2)-(1/2048)}{1-(1/2)} = 1023/1024$. Close to 1.

→ But on a discrete infinite space, cannot ever have a uniform distribution!

• Summary: Don't assume it's uniform when it isn't!

More Finite Uniform Probabilities (§1.4)

• Distinct, in order: e.g. Suppose there are ten people at a party, and you randomly pick three of the people, in order (1-2-3). What is the probability that your choices will also be the three richest people at the party, in the same order?

→ S is the set of all ways of picking three people, in order. All equally likely.

→ But what is $|S|$?

→ The first person can be picked in 10 different ways.

→ Then, the second person can be picked in 9 different ways.

→ Then, the third person can be picked in 8 different ways.

→ So, $|S| = 10 \cdot 9 \cdot 8 = 720$.

→ Also, $|A| = 1$ since there is only one matching choice.

→ So, P(you picked the three richest, in order) $= |A|/|S| = 1/720$.

• More generally, the number of ways of picking k distinct items, <u>in order</u>, out of n items total, is equal to $n(n-1)(n-2)\ldots(n-k+1) = n!/(n-k)!$. ("permutations")

→ In particular, if $k = n$, then the number of ways of picking <u>all</u> n items in order is equal to $n(n-1)(n-2)\ldots(1) = n!$. ("$n$ factorial")

• "The Birthday Problem": Suppose 40 (say) people at a party are each equally likely to be born on any one of 365 days of the year. Then what is the probability that at least one <u>pair</u> of them have the same birthday? (Any guesses?)

→ Here, S is the set of all 40-tuples of possible birthdays. All equally likely.

→ (List their birthdays in <u>order</u>, since they might not all be distinct.)

→ So, by the Multiplication Principle, $|S| = 365^{40}$.

→ What about $|A|$? Not easy ...

→ Instead, consider A^C. (Then can use that $\mathrm{P}(A) = 1 - \mathrm{P}(A^C)$.)

→ A^C is the set of all ways of picking 40 <u>distinct</u> birthdays, in order.

→ So, $|A^C| = 365 \cdot 364 \cdot 363 \cdot \ldots \cdot 326 = 365! \,/\, 325!$.

→ So, $\mathrm{P}(A^C) = \big(365!/325!\big) \,/\, 365^{40} \doteq 0.109$.

→ So, $\mathrm{P}(A) = 1 - \mathrm{P}(A^C) \doteq 0.891$. Over 89%. Very likely! (Make a bet?)

→ Intuition: Even with just 40 people, have $\binom{40}{2} = 780$ <u>pairs</u> of people – lots!

→ Or, if 23 people, $\mathrm{P}(A^C) = \big(365!/342!\big)\,/365^{23} \doteq 0.493$, so $\mathrm{P}(A) \doteq 0.507 > 50\%$.

<u>POLL:</u> With <u>60 people</u>, what is P(some pair have same birthday)? (guess)
(A) 92.8%. **(B)** 95.1%. **(C)** 99.4%. **(D)** 99.86%. **(E)** 99.993%.

→ With 60 people: $\mathrm{P}(A^C) = \big(365!/305!\big)\,/365^{60} \doteq 0.059$; $\mathrm{P}(A) \doteq 0.994 = 99.4\%$.

→ (For discussion with "C" people, see the textbook's Challenge 1.4.21.)

• Distinct, unordered: Suppose we are still picking k distinct objects, but now we don't care about the <u>order</u>. Then, we have to <u>divide</u> by the number of different <u>orderings</u> of k items, which is: $k! = k(k-1)(k-2)\ldots(2)(1)$.

→ So, the number of ways of picking k distinct items out of n items total, <u>ignoring</u> order, is equal to $n(n-1)(n-2)\ldots(n-k+1) \,/\, k! = n!/(n-k)!\,k!$. ("combinations"; "choose formula", or "binomial coefficient") Also written as: $\binom{n}{k}$.

POLL: Suppose there are ten people at a party, and you randomly pick a collection of three of the people, but ignoring order. What is the probability that your choices will also be the three richest people at the party (in any order)?
(A) 1/60. **(B)** 1/120. **(C)** 1/240. **(D)** 1/360. **(E)** 1/720. **(F)** No idea.

→ Here S is all ways of picking three people (ignoring order). All equally likely.

→ But what is $|S|$?

→ Here $|S| = \binom{10}{3} = \frac{10!}{7!\,3!} = 120$.

→ And, again $|A| = 1$ since there is only one matching choice.

→ So, P(you picked the three richest, ignoring order) $= |A|/|S| = 1/120$.

→ Six times as large as before! Makes sense since $3! = 6$.

- e.g. Lotto Max jackpot:

→ Here $S = \{$all choices of 7 distinct numbers between 1 and 50$\}$.

→ All equally likely. And, we do not care about the order.

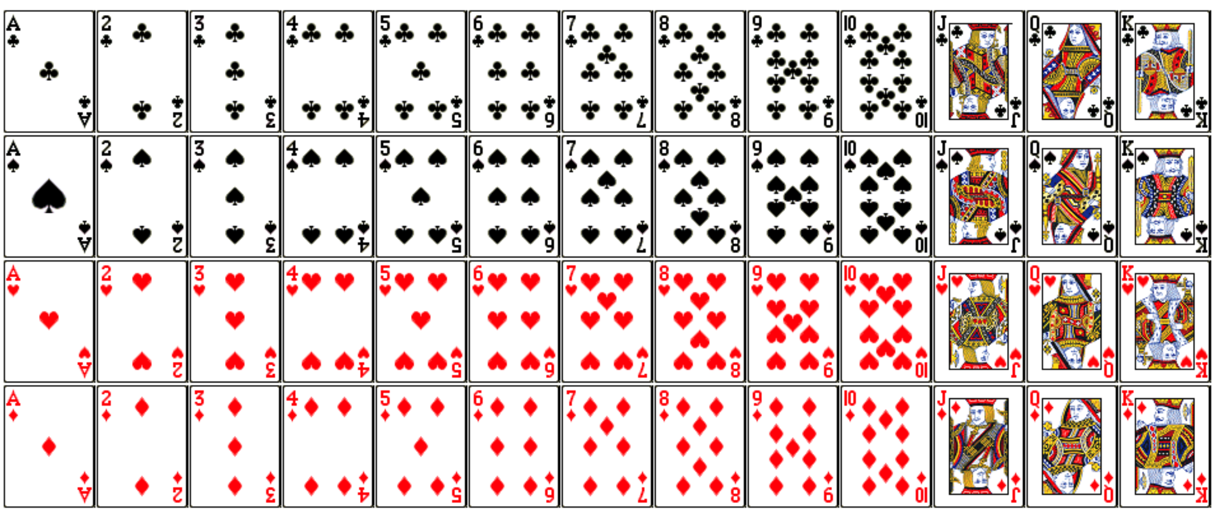
→ So, $|S| = \frac{50!}{43!\,7!} = 99,884,400 \doteq 100$ million.

→ Also, A is the one correct choice. So, $|A| = 1$.

→ So, P(jackpot) = P(choose the correct 7 distinct numbers between 1 and 50) $= |A| \,/\, |S| = 1/99,884,400 \doteq 1/100,000,000 = 0.000001\%$. Very small!

→ (For \$5, you get three separate choices of 7 numbers, which increases P(jackpot) to 3 / 99,884,400 = 1 /33,294,800 ... still very small ...)

- Recall that a standard deck of playing cards has four suits (Clubs, Spades, Hearts, Diamonds), and each suit has 13 ranks (A,2,3,4,5,6,7,8,9,10,J,Q,K), so 52 cards total:



- A card's value is its number, counting A as 1, J as 11, Q as 12, and K as 13.

- Suppose we pick one playing card from a standard deck, uniformly at random.

→ So S is the set of all cards in the deck, with $|S| = 52$, all equally likely.

→ Then what is P(Club or 7)? Can solve this directly, or ...

→ Here P(Club) $= 13/52 = 1/4$, and P(7) $= 4/52 = 1/13$.

→ Also, P(Club and 7) = P(7-of-Clubs) $= 1/52$.

→ So, by Inclusion-Exclusion, P(Club or 7) = P(Club) + P(7) − P(Club and 7) $= 1/4 + 1/13 - 1/52 = 16/52 = 4/13$.

Suggested Homework: 1.3.6, 1.4.4, 1.4.6, 1.4.7, 1.4.8. Trickier: 1.4.5.

POLL: Suppose we draw a pair of distinct cards uniformly from a standard deck. What is P(both are Face Cards), i.e. P(both are J/Q/K)?
(A) $(3/52)^2$. **(B)** $(12/52)^2$. **(C)** $12/\binom{52}{2}$. **(D)** $\binom{12}{2}/\binom{52}{2}$. **(E)** No idea.

END WEDNESDAY #2