

CSCI 3022

intro to data science with probability & statistics

September 12, 2018

1. HW 1 due this Fri. 9/14/18
2. Mathematical probability

← 5 P.M. Moodle.
OH today till 5:30

Last Time

- **Sample Space Ω** : set of all possible outcomes of an experiment.
- **Event**: a set of one or more outcomes.
- **Probability Function P** : assigns value in $[0, 1]$ to each outcome or event.
- Two requirements:
 1. Probability of the sample space is 1. $P(\Omega) = 1$
 2. $P(A \cup B) = P(A) + P(B)$ ^{only} if A and B are disjoint.
- If A and B are not disjoint then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- If results of two trials don't affect each other, we say they are **independent**.

Warming up

- Suppose you draw one card from a standard 52-card deck

Question: What is the probability that the card is **A**♦?

$$\frac{1}{52}$$

Question: What is the probability that the card is an **A** or a ♦?

$$\frac{1}{4} + \frac{1}{13} - \frac{1}{52} = \boxed{\frac{16}{52}}$$

$$P(\spadesuit) + P(A) - P(A \cap \spadesuit)$$

A rigorous way to compute probabilities

to be clear:
disjoint

- Suppose we know $P(\omega)$ for each outcome ω in Ω .
- We can compute the probability of an event A (1 or more outcomes) as the sum of the probabilities of the outcomes in A

$$A = \{\omega_1, \omega_2, \omega_3\} \quad \text{then} \quad P(A) = P(\omega_1) + P(\omega_2) + P(\omega_3)$$

Question: suppose we flip a *biased* coin $P(\{H, T\}) = \{p, 1-p\}$ exactly 3 times. What is the probability that we get two or more T?

T T H,	ω_1	$P(\omega_1) = (1-p)^2 p$
T H T,	ω_2	$P(\omega_2) = (1-p)^2 p$
H T T,	ω_3	$P(\omega_3) = (1-p)^2 p$
T T T,	ω_4	$P(\omega_4) = (1-p)^3$

$$P(A) = 3(1-p)^2 p + (1-p)^3$$

To Infinity and Beyond!

- Suppose you flip a biased coin until a H comes up. Prove that the probability that you flip a H eventually is 1.

discrete.

$$|\Omega| = \infty$$

Question 1: what is the sample space for this experiment?

$$\Omega = \{H, TH, TTH, TTTT, TTTTH, \dots\}$$

Question 2: what is the probability that you flip a heads eventually?

$$P(E) = P(H) + P(TH) + P(TTH) + P(TTTTH) + \dots$$

$$= p + (1-p)p + (1-p)^2 p + (1-p)^3 p + \dots$$

$$= \sum_{i=0}^{\infty} p (1-p)^i = p \sum_{i=0}^{\infty} (1-p)^i \stackrel{\text{geom series}}{=} \frac{p}{1-(1-p)} = \frac{p}{1-1+p} = \frac{p}{p} = 1$$

Conditional probability

Question: Stop a random person on the street and ask them ⁱⁿ what month they were born. What is the probability that they were born in a 31-day month?

$$E = \{ \overset{1}{\text{Jan}}, \overset{2}{\text{Mar}}, \overset{3}{\text{May}}, \overset{4}{\text{June}}, \overset{5}{\text{Aug}}, \overset{6}{\text{Oct}}, \overset{7}{\text{Dec}} \} \quad P(E) = \frac{7}{12}$$

Question: What is the probability that they were born in a month with an r in the name?

$$R = \{ \overset{1}{\text{Jan}}, \overset{2}{\text{Feb}}, \overset{3}{\text{Mar}}, \overset{4}{\text{Apr}}, \overset{5}{\text{Sept}}, \overset{6}{\text{Oct}}, \overset{7}{\text{Nov}}, \overset{8}{\text{Dec}} \}$$

$$P(R) = \frac{8}{12}$$

Conditional probability

Question: Suppose that the person tells you that they were born in a 31-day month. Now what is the probability that they were born in a month with an r in it?

$$P(R) = \frac{8}{12}$$

$$P(E_{31}) = \frac{7}{12}$$

we know this event E happened.
 E is true.

mid

"given"
assume to be true

want to know:

$$P(R | E)$$

Conditional probability

1 1 1 1 1
 1 1 1 0 2
 1 1 0 1 3
 1 1 0 0 4
 1 0 1 1 5
 1 0 1 0
 1 0 0 1
 1 0 0 0

Definition: The conditional probability of A given C is

$$P(A|C) = \frac{P(A \cap C)}{P(C)}$$

provided that

0 1 1 1 6
 0 1 1 0 7
 0 1 0 1
 0 1 0 0
 0 0 1 1 8
 0 0 1 0
 0 0 0 1
 0 0 0 0

Example: A bit string (1s and 0s) of length 4 is generated at random so that each of the 16 possible bit strings is equally likely. What is the probability that it contains *at least* two consecutive 1s given that the first bit is a 1?

8 contain 2 consec. ones

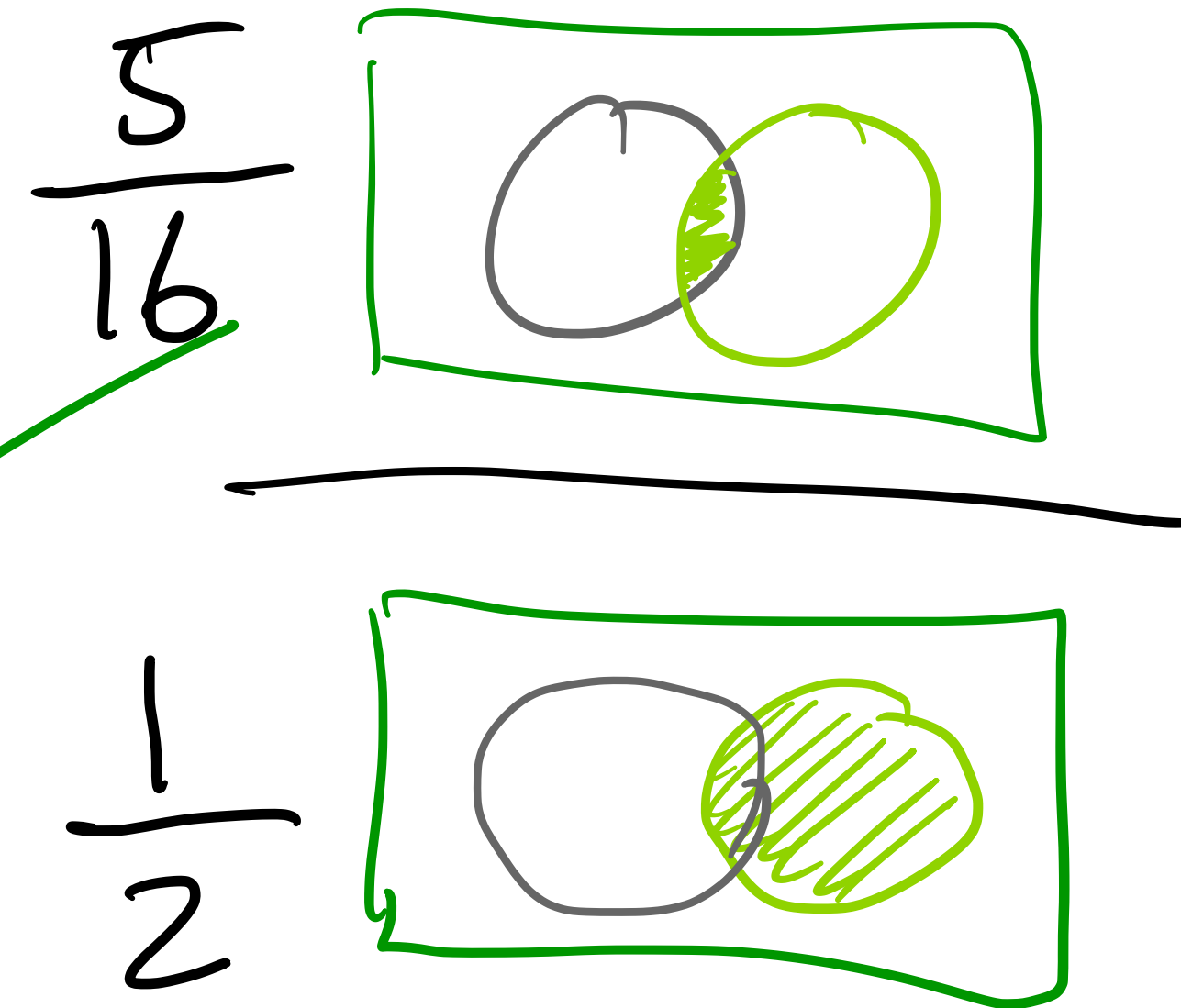
8 start with ones.

5 start w/ one, have two consec 1s.

$$P(\text{start w/ 1} \cap \text{2 in a row}) = \frac{5}{16}$$

$$P(\text{start w/ 1}) = \frac{1}{2}$$

$$\frac{5/16}{1/2} = \frac{10}{16} = \frac{5}{8}$$



Conditional probability

Example2: A bit string (1s and 0s) of length 4 is generated at random so that each of the 16 possible bit strings is equally likely. What is the probability that it **does not** contain two consecutive 1s given that the first bit is a 1?

$$P(E|C) = 1 - P(\text{not } E|C) = 1 - \frac{5}{8} = \frac{3}{8}$$

↑
prev slide

NB: the conditional probability $P(\cdot | C)$ is a valid probability function.

$$P(A|C) + P(\text{not } A|C) = 1$$

$$\text{i.e. } P(\Omega) = 1$$

The product rule of probability

- The definition of conditional probability can be manipulated!
- In this form, we call this the **product rule**:

$$P(A \cap C) = P(A|C)P(C)$$

- The product rule is useful when the conditional probability is easy to compute, but the probability of intersections of events is difficult.

Example: you draw 2 cards from a deck. What is the probability that they are both black?

$$\begin{aligned} P(1^{\text{st}} \text{ card black} \cap 2^{\text{nd}} \text{ card black}) &= P(2^{\text{nd}} \text{ black and } 1^{\text{st}} \text{ card black}) P(1^{\text{st}} \text{ card black}) \\ &= \frac{25}{51} \times \frac{1}{2} = \frac{25}{102} \end{aligned}$$

Independent events - intuition

Example: you draw 2 cards from a deck. What is the probability that they are both black?

Are the events of drawing the first black card and the second black card **independent**?

NO 😞

Independent events - math

Definition: An event A is said to be independent of event B if $P(A|B) = P(A)$

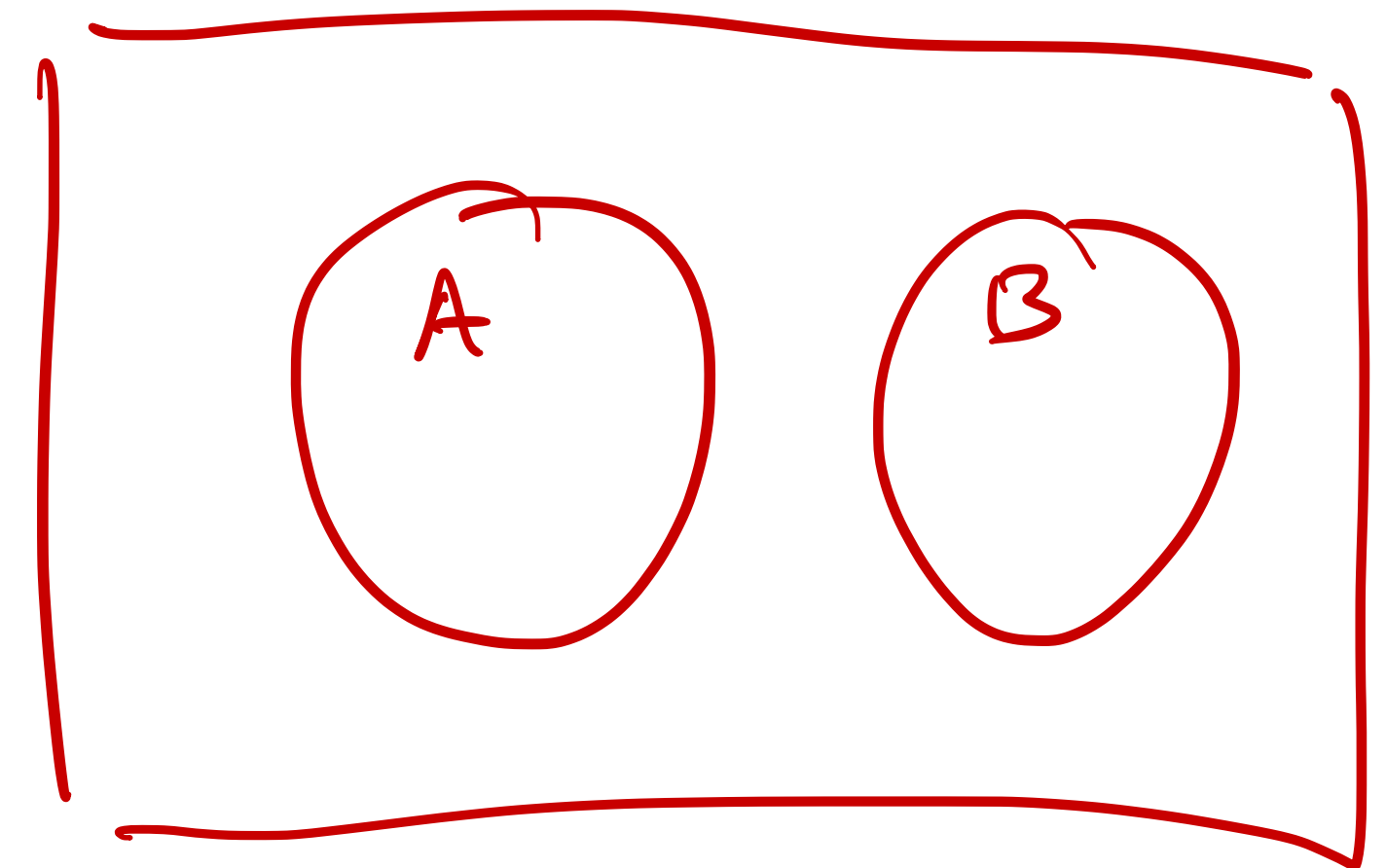
This definition, combined with the product rule (or the definition of conditional probability) gives us many equivalent definitions (or tests!) for independence:

→ 1. $P(A|B) = P(A)$ \Leftrightarrow indep.

→ 2. $P(B|A) = P(B)$

→ 3. $P(A \cap B) = P(A)P(B)$

$$P(B|A)P(A) = P(A \cap B) = P(A|B)P(B)$$



Subtleties of Independence

- **Def:** Events A_1, A_2, \dots, A_m are independent if

$$P(A_1 \cap A_2 \cap \dots \cap A_m) = P(A_1)P(A_2) \dots P(A_m)$$

Question: is independence of A, B, and C the same as: A and B are independent; B and C are independent; A and C are independent?

Example: Flip a fair coin twice. Let A be "Heads on flip 1", let B be "Heads on flip 2", let C be "the two flips match".

$$P(A) = \frac{1}{2}$$

$$P(B) = \frac{1}{2}$$

$$P(C) = \frac{1}{2}$$

$$P(A \cap B \cap C) = \frac{1}{4}$$

$$P(A) \cdot P(B) \cdot P(C) = \frac{1}{8}$$

$$P(C|B) = \frac{\frac{1}{2}}{\frac{1}{2}} = \frac{1}{2} = P(C)$$

Law of Total Probability

6,4

3,7

- Suppose I have a bag of marbles. The first bag contains 6 white marbles and 4 black marbles. The second bag contains 3 white marbles and 7 black marbles.
- Now suppose I put the two bags in a box. If I close my eyes, grab a bag from the box, and then grab a marble from that bag, what is the probability that it is black?

$$P(\text{marble is black } B) = P(B | \text{bag 1}) P(\text{bag 1}) + P(B | \text{bag 2}) P(\text{bag 2})$$
$$P(B \cap \text{bag 1}) + P(B \cap \text{bag 2})$$

$$= \frac{4}{10} \cdot \frac{1}{2} + \frac{7}{10} \cdot \frac{1}{2}$$
$$= \frac{11}{20}$$

Law of Total Probability

bag 1
 $2x$

bag 2
 x

$$2x + x = 1 \Rightarrow x = \frac{1}{3}$$

- Same as before: The first bag contains 6 white marbles and 4 black marbles. The second bag contains 3 white marbles and 7 black marbles. But suppose the first bag is much larger than the second bag—I'm twice as likely to grab the first bag as the second bag. What's the probability that I get a black marble?

$$P(B) = P(B|\text{bag 1})P(\text{bag 1}) + P(B|\text{bag 2})P(\text{bag 2})$$

$$= \frac{4}{10} \cdot \frac{2}{3} + \frac{7}{10} \cdot \frac{1}{3}$$

$$= \frac{8}{30} + \frac{7}{30} = \frac{1}{2}$$

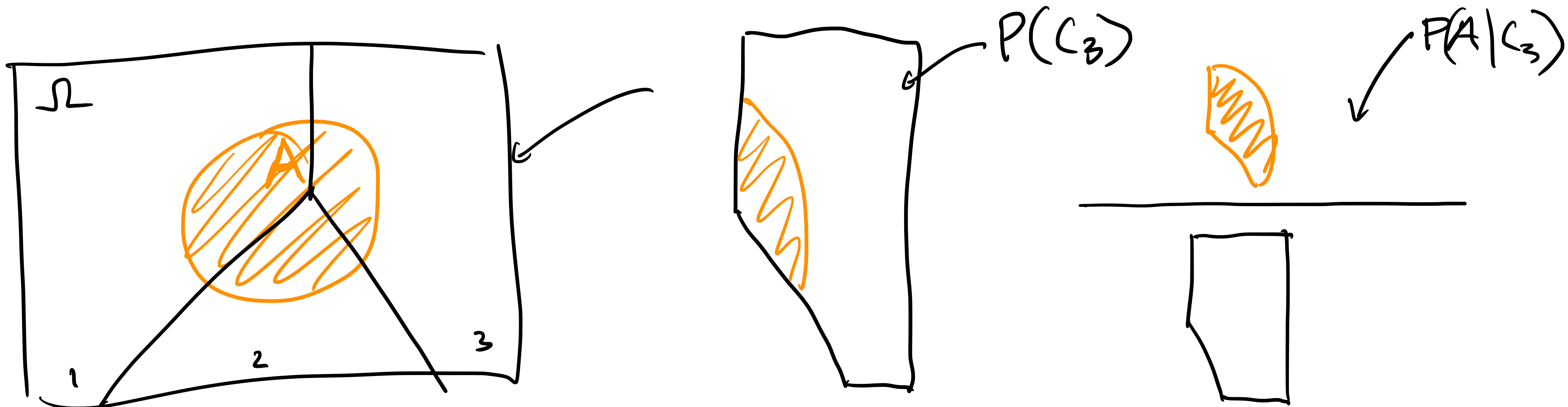
Law of Total Probability

- **Definition:** Suppose C_1, C_2, \dots, C_m are disjoint events such that

$$C_1 \cup C_2 \cup \dots \cup C_m = \Omega$$

Then the probability of an arbitrary event A can be expressed as:

$$P(A) = P(A|C_1)P(C_1) + P(A|C_2)P(C_2) + \dots + P(A|C_m)P(C_m)$$



Data Party!

- <https://tinyurl.com/octopusbirthday>

Quiz let on Moodle.

417 till 5:30!

The Birthday Paradox

- Say there are two random people in a room. What is the probability that they have different birthdays?

The Birthday Paradox

- Say there are n random people in a room. What is the probability that they all have different birthdays?

The Birthday Paradox

- 23 people in the room: $P(\text{shared bday}) = 0.5973$
- 58 people in the room: $P(\text{shared bday}) = 0.9917$
- 120 people in the room: $P(\text{shared bday}) = 0.999999999998$

