CSCI 3022-002 Intro to Data Science
Conditional Probability

recall HW:

H: { get 18" rain }

T: { regular September }

Opening **Example**: Suppose we flip a coin with a 1% chance per flip of landing on heads. Define $X =$ the number of tails flips before we see a heads. What is $P(X = 0)$? $P(X = 1)$? $P(X = i)$? Verify that $P(X) = 1$ over all of Ω.

random variable X

Opening Example Sol'n

$X = \{0, 1, 2, 3, 4, \ldots\}$

Suppose we flip a coin with a 1% chance per flip of landing on heads. Define X = the number of tails flips before we see a heads. What is $P(X = 0)$? $P(X = 1)$? $P(X = i)$? Verify that $P(X) = 1$ over all of Ω.

Poll

$\frac{1}{100}\left(1 + \frac{99}{100}\right) \approx 2/100$

$P(X=0) = P(\{H\}) = 1/100$

$P(X=1) = P(\{T_1 \; H_2\}) = P(\{T\}) \cdot P(\{H\}) = \frac{99}{100} \cdot \frac{1}{100}$

1 AND 2

$P(X=2) = P(\{TTH\}) = P(\{T\})^2 \cdot P(\{H\}) = \left(\frac{99}{100}\right)^2 \cdot \frac{1}{100}$

$P(X=k) = \left(\frac{99}{100}\right)^k \left(\frac{1}{100}\right)$

$1 = 1/100 + \frac{99}{100} \cdot \frac{1}{100} + (99/100)^2 \cdot 1/100 + \ldots = \frac{1}{100}\left(1 + \frac{99}{100} + \left(\frac{99}{100}\right)^2 + \left(\frac{99}{100}\right)^3 + \ldots\right)$

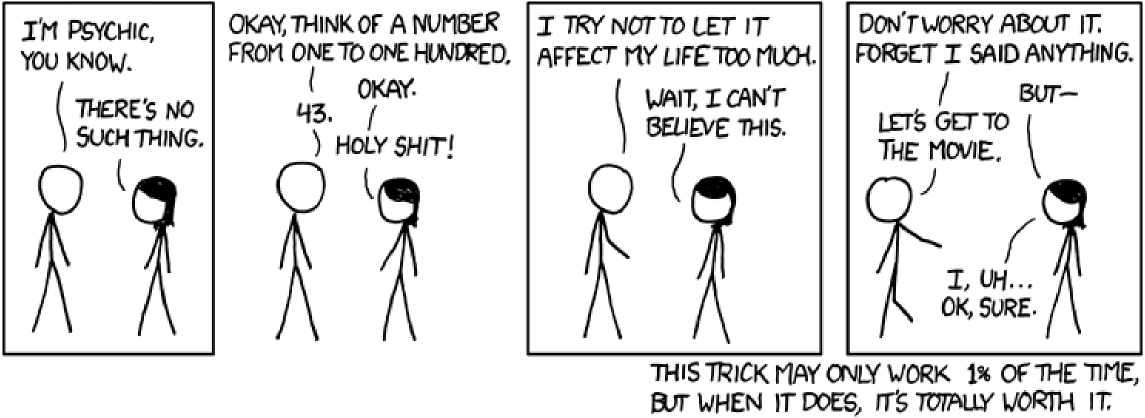
Opening Example Sol'n

Suppose we flip a coin with a 1% chance per flip of landing on heads. Define $X =$ the number of tails flips before we see a heads. What is $P(X = 0)$? $P(X = 1)$? $P(X = i)$? Verify that $P(X) = 1$ over all of Ω.

$a + ar + ar^2 + ar^3 + \ldots = \frac{a}{1-r}$

1. $P(X = 0) = P(\{H\}) = .01$.
2. $P(X = 1) = P(\{TH\}) = P(\{T\})P(\{H\}) = .99 \cdot .01$.
3. $P(X = 2) = P(\{TTH\}) = P(\{T\})^2 P(\{H\}) = .99^2 \cdot .01$.
4. $P(X = i) = P(\{T \ldots TH\}) = P(\{T\})^i P(\{H\}) = .99^i \cdot .01$.
5. $\sum_{i=0}^{\infty} P(X = i) = \sum_{i=0}^{\infty} .99^i \cdot .01 = \frac{.01}{1-.99} = 1$. Sanity check passed!

$= \frac{1}{100} \sum_{i=0}^{\infty} \left(\frac{99}{100}\right)^i = \frac{1}{100} \cdot \frac{1}{1 - 99/100} = \frac{1}{100} \cdot \frac{1}{1/100}$ $\boxed{= 1}$


I'M PSYCHIC, YOU KNOW.
THERE'S NO SUCH THING.
OKAY, THINK OF A NUMBER FROM ONE TO ONE HUNDRED.
OKAY.
43.
HOLY SHIT!
I TRY NOT TO LET IT AFFECT MY LIFE TOO MUCH.
WAIT, I CAN'T BELIEVE THIS.
DON'T WORRY ABOUT IT. FORGET I SAID ANYTHING.
BUT—
LET'S GET TO THE MOVIE.
I, UH... OK, SURE.
THIS TRICK MAY ONLY WORK 1% OF THE TIME, BUT WHEN IT DOES, IT'S TOTALLY WORTH IT.

Announcements and To-Dos

Announcements:

1. HW 2 Posted, due Monday! (check your version!)
most recent/final posted this am
2. Another nb day this Friday!

Last time we learned:

1. Basics of Probability in review.

To do:

1. Start your HW! Ensure you can load the data and work with it. Read problems to think about what's involved.

Last Time...

A few big takeaways from our first lecture on probability.

- A *sample space* (denoted Ω) of a probabilistic process is the set of all possible outcomes of that process.
- An *event* is any collection (subset) of outcomes from the sample space.
- *Probability* is a function that takes in events and random variables and outputs numbers in $[0, 1]$.
- **Idea:** two or more trials are *independent* if they don't affect each other.

Last Time...

A few big takeaways from our first lecture on probability.

- A *sample space* (denoted Ω) of a probabilistic process is the set of all possible outcomes of that process.
- An *event* is any collection (subset) of outcomes from the sample space.
- *Probability* is a function that takes in events and random variables and outputs numbers in $[0, 1]$.
  1. If A and B are disjoint (mutually exclusive) sets, $P(A \cup B) = P(A) + P(B)$.
     inclusive OR
  2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. if overlap
- **Idea:** two or more trials are *independent* if they don't affect each other.

Formal Probability

events may include
many outcomes

event

Suppose we know $P(\omega)$ for each outcome ω in Ω.

We can compute the probability of an event A which may include one or more outcomes as the sum of all of the probabilities of the (distinct) outcomes in A:

$$P(A) = \sum_{\omega \in A} P(\omega)$$

Example: Suppose we flip a biased coin with a probability function given by $P(\{H, T\}) = \{p, 1 - p\}$ three times. What is the probability we get two or more tails?

Adding Outcomes

$P(\# T \geq 2)$

Example: Suppose we flip a biased coin with a probability function given by $P(\{H, T\}) = \{p, 1-p\}$ three times. What is the probability we get two or more tails?

Samp space: $\{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$

$$P(E) = \underbrace{P(TTH) + P(THT) + P(HTT)}_{P(\text{exactly } 2\ T\text{'s})} + P(TTT).$$

$$= \qquad \text{indep} \qquad P(\text{exactly } 2\ T\text{'s}) + P(\text{exactly } 3\ T\text{'s})$$

$$= P(T)\cdot P(T)\cdot P(H) + P(T)\cdot P(H)\cdot P(T) + P(H)\cdot P(T)\cdot P(T) + P(T)P(T)\cdot P(T)$$

Adding Outcomes

Example: Suppose we flip a biased coin with a probability function given by $P(\{H,T\}) = \{p, 1-p\}$ three times. What is the probability we get two or more tails?

- A is the event that that we see two or more tails. It includes the following elements of Ω: $\{\{TTH\}, \{THT\}, \{HTT\}, \{TTT\}\}$.
- $P(A) = \sum_{\omega \in A} P(\omega) = P(\{TTH\}) + P(\{THT\}) + P(\{HTT\}) + P(\{TTT\})$ because of these outcomes are *disjoint.*
- These probabilities are the products of probabilities of the 3 flips within each, because each flip is *independent* and the probabilities are identical. As a result:

$$\begin{aligned} P(A) &= (1-p)\cdot(1-p)\cdot p + (1-p)\cdot p\cdot(1-p) + p\cdot(1-p)\cdot(1-p) \\ &+ (1-p)\cdot(1-p)\cdot(1-p) \\ &= 3p(1-p)^2 + (1-p)^3 \end{aligned}$$

Conditional Probability

Example: If you stop a random person on the street and ask them what month they were born, what is the probability they were born in a long month? (i.e., with 31 days in it)

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

Conditional Probability

Example: If you stop a random person on the street and ask them what month they were born, what is the probability they were born in a long month? (i.e., with 31 days in it)
Lazy answer: let L be the event that their birth month has 31 days in it.
$\{Jan, Mar, May, Jul, Aug, Oct, Dec\}$ are the elements in L out of 12 months total, so $P(L) = \frac{7}{12}$ if all months are equally likely.

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

Conditional Probability

Example: If you stop a random person on the street and ask them what month they were born, what is the probability they were born in a long month? (i.e., with 31 days in it)

Slightly less lazy answer: let L be the event that their birth *day* in a month with 31 days in it. The months in L, now span $7 \cdot 31 = 217$ days out of 365 (.2422) total, so $P(L) = \frac{217}{365}$ if all days are equally likely.

wrong!

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

Conditional Probability

Example: If you stop a random person on the street and ask them what month they were born, what is the probability they were born in a long month? (i.e., with 31 days in it)

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

(Only the lazy answer): Let R be the event that their birth month has an 'r' in the name. $\{Jan, Feb, Mar, Apr, Sep, Oct, Nov, Dec\}$ are the elements in R, so $P(R) = \frac{8}{12}$ if all months are equally likely.

Conditional Probability

Example: Now suppose this person tells you that they were born in a 31 day month. What, then, is the probability that they were born in a month with an 'r' in it?

event: 31 day $\quad P(\text{"31"}) = 7/12 \quad P(\text{"r"}) = 8/12$

$P(\text{"r" given 31-day month})$

Conditional Probability

Example: Now suppose this person tells you that they were born in a 31 day month. What, then, is the probability that they were born in a month with an 'r' in it?

→ 7 items

Recall $\{Jan, Mar, May, Jul, Aug, Oct, Dec\} \in L$ and
$\{Jan, Feb, Mar, Apr, Sep, Oct, Nov, Dec\} \in R$.

4/12 have both!

Conditional Probability

Example: Now suppose this person tells you that they were born in a 31 day month. What, then, is the probability that they were born in a month with an 'r' in it?

Recall $\{Jan, Mar, May, Jul, Aug, Oct, Dec\} \in L$ and $\{Jan, Feb, Mar, Apr, Sep, Oct, Nov, Dec\} \in R$.
Our *given* knowledge has reduced the sample space to just those elements in L! Now that $\Omega = L$, we are only interested in the elements in R that are also in L. (4 of)

$$P(R \textbf{ given } L) = \frac{\# event\, in\, both}{\# events\, in\, L}$$

$$P(R|L) = \frac{P(R \cap L)}{P(L)}$$

"given"

$$= \frac{4/\cancel{12}}{7/\cancel{12}}$$

$$= 4/7$$

Conditional Probability

Notation:
We will use the notation $P(A|B)$ to represent the conditional probability of event A *given* that the event B has occurred. B is the "conditioning event."

Definition: *Conditional Probability* is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad \frac{\text{both}}{B}$$

provided that $P(B) > 0$.

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 2 consecutive 1s, given that the first bit is a 1?

Conditional Probability

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 2 consecutive 1s, given that the first bit is a 1?

1101
1000
1001
1010
1100
1110
1011
1111

5/8

Conditional Probability

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 2 consecutive 1s, given that the first bit is a 1?

List all 16 strings is an option... Or consider the event A : that there are consecutive 1's. Maybe it's easier to find A^C! The strings without consecutive 1's that start with a 1 are $\{1010, 1000, 1001\}$. Let C be the event that the first bit is a 1.

10 10
01
00

$$P(A^C|C) = \frac{P(A^C \cap C)}{P(C)}$$
$$= \frac{3/\cancel{16}}{8/\cancel{16}} = 3/8$$

= P (not 2 conse. 1's given starts w/ 1)

Conditional probability $P(\cdot|C)$ is a valid probability function, so the complementation property $P(A|C) = \boxed{1 - P(A^C|C) = \frac{5}{8}}$ holds.

The Multiplication Rule

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

The definition of conditional probability yields the following result:

Multiplication Rule:

- $P(A \cancel{\cup} B) =$ $\cap$ $P(A|B) \cdot P(B)$

The multiplication rule is particularly useful when conditional probabilities are easier to calculate than intersections.

The Multiplication Rule

The definition of conditional probability yields the following result:
Multiplication Rule:

- $P(A \cancel{\cup} B) = P(A|B)P(B)$
  $\cap$

The multiplication rule is particularly useful when conditional probabilities are easier to calculate than intersections.

The Multiplication Rule

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

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

The definition of conditional probability yields the following result:

Multiplication Rule:

- $P(A \underset{\cap}{\cancel{\cup}} B) = P(A|B)P(B)$
- $P(A \underset{\cap}{\cancel{\cup}} B) = P(B|A)P(A)$

The multiplication rule is particularly useful when conditional probabilities are easier to calculate than intersections.

Example: You draw two cards from a standard playing deck. What is the probability that they are both red?

$$P(R_1 \text{ and } R_2) = P(R_1) \cdot P(R_2 | R_1)$$

$$= 26/52 \quad 25/51$$

Independence, formally

"knowledge of B
gives no info.
about A"

Definition: Two events A and B are said to be *independent* if $P(A|B) = P(A)$.

This definition, combined with the product rule give us three equivalent tests for independence:

1. $P(A|B) = P(A)$
2. $P(B|A) = P(B)$
3. $P(A \cap B) = P(A)P(B)$

$$= P(A\cancel{|B}) \cdot P(B)$$

$$= P(B\cancel{|A}) \cdot P(A)$$

Independence, in detail!

We don't have to stop at two sets. Sometimes we have lots of outcomes we want to be unrelated.

Events $A_1, A_2, \ldots, A_n$ are *mutually independent* if for every $k = 2, 3, \ldots, n$ and every subset of indices $i_1, i_2, \ldots, i_k$

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k})$$

In other words, for any selection of A mutually independent event, the probability of their intersection is equal to the product of their individual probabilities.

Independence, in detail!

Why does this matter? Consider the following

Example: Flip a fair coin twice, and define

1. A : "heads on flip 1" $P(A) = 1/2$
2. B : "heads on flip 2" $P(B) = 1/2$
3. C : "same outcomes on both flips" $P(C) = 1/2$

HH
HT
TH
TT

each $P = .25 = 1/4$

What are $P(A), P(B), P(C), P(A|B), P(A|C), P(B|C)$?

What about $P(A \cap B \cap C?)$

$= 1/4 \neq P(A) \cdot P(B) \cdot P(C)$

$P(A \cap B) = P(A) \cdot P(B)$?

$1/4 = 1/2 \cdot 1/2$

A & B are indep

$P(A \cap C) = 1/4 = 1/2 \cdot 1/2$

Independence, in detail!

Why does this matter? Consider the following
Example: Flip a fair coin twice, and define

1. A : "heads on flip 1"
2. B : "heads on flip 2"
3. C : "same outcomes on both flips"

What are $P(A), P(B), P(C), P(A|B), P(A|C), P(B|C)$?
What about $P(A \cap B \cap C?)$
Any *pair* of A, B, C looks independent, since
$P(A) = P(B) = P(C) = P(A|B) = P(A|C) = P(B|C) = 1/2$.
However, $P(A \cap B \cap C?) = P(\{HH\}) = 1/4$ which is not the same as the triple product $P(A)P(B)P(C) = \frac{1}{8}$.
Ultimately, event C is determined by the combination of A and B.

The Law of Total Probability

Example: Suppose I have a couple of bags 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?

either black from bag 1 or black marble bag 2

$P(B) = P(B_1) + P(B_2)$

The Law of Total Probability

Example: Suppose I have a couple of bags 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?

Equally likely

There are two 'ways' we get a black marble: from bag 1 or from bag 2. We just have to add both up!

$$\begin{aligned} P(\mathbf{B}) &= P(\mathbf{B\ from\ 1}) + P(\mathbf{B\ from\ 2}) \\ &= P(\mathbf{B} \cap \mathbf{1}) + P(\mathbf{B} \cap \mathbf{2}) \\ &= P(\mathbf{B}|\mathbf{1})P(\mathbf{1}) + P(\mathbf{B}|\mathbf{2})P(\mathbf{2}) \\ &= \frac{4}{10} \cdot \frac{1}{2} + \frac{7}{10} \cdot \frac{1}{2} \\ &= \frac{11}{20} \end{aligned}$$

← LTP

← Conditional Prob.

The Law of Total Probability

Example: 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 what if bag 1 is made of a much nicer material to touch, so I am twice as likely to randomly select bag 1 from between the 2 bags?

$P(1) + P(2) = 1$

$P(1) = 2\,P(2)$

$P(1) = 2/3 \qquad P(2) = 1/3$

$$P(B) = P(B \cap 1) + P(B \cap 2)$$
$$= P(B|1)\,P(1) + P(B|2) \cdot P(2)$$

The Law of Total Probability

Example: 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 what if bag 1 is made of a much nicer material to touch, so I am twice as likely to randomly select bag 1 from between the 2 bags?

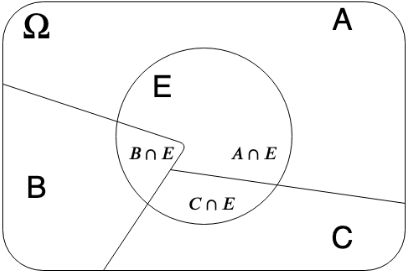
Same solution!:

$$\begin{aligned} P(\textbf{B}) &= P(\textbf{B from 1}) + P(\textbf{B from 2}) \\ &= P(\textbf{B} \cap \mathbf{1}) + P(\textbf{B} \cap \mathbf{2}) \\ &= P(\textbf{B}|\mathbf{1})P(\mathbf{1}) + P(\textbf{B}|\mathbf{2})P(\mathbf{2}) \\ &= \frac{4}{10} \cdot \frac{2}{3} + \frac{7}{10} \cdot \frac{1}{3} \\ &= \frac{1}{2} \end{aligned}$$

The Law of Total Probability

Definition: A *Partition* of Ω is a set of disjoint events $E_1, E_2, \ldots E_k$ such that $E_1 \cup E_2 \cup \cdots \cup E_k = \Omega$. Given such a partition, any event A can be decomposed into:

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \cdots + P(A \cap E_k)$$



$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_k)P(E_k)$$

"Counting" Examples

Example 1: Suppose that your music app contains 100 songs, 10 of which are by the Beatles. Using the shuffle feature, what is the probability that the first Beatles song heard is the fifth song played?

"Counting" Examples

Example 1: Suppose that your music app contains 100 songs, 10 of which are by the Beatles. Using the shuffle feature, what is the probability that the first Beatles song heard is the fifth song played? **Solution:** Two ways

1. Count all possible shuffles of the first 5 songs: $P(100, 5)$ - then count all possible selections where 1-4 are not B and 5 is B: $P(90, 4)$ and also $P(5, 1)$

$$P(E) = \frac{P(90, 4)C(4, 1)}{P(100, 5)}$$

2. Do things *conditionally*:

$$\begin{aligned} P(\mathit{NNNNB}) &= P(N_1N_2N_3N_4)P(B_5 \textbf{ GIVEN } N_1N_2N_3N_4) \\ &= P(N_1N_2N_3N_4)\frac{5}{90} = P(N_1N_2N_3)P(N_4 \textbf{ GIVEN } N_1N_2N_3)\frac{5}{90} \\ &= \cdots = \frac{90}{100}\frac{89}{99}\frac{88}{98}\frac{87}{97}\frac{10}{96} \end{aligned}$$

"Counting" Examples

Example 2: What is the probability of being dealt all 4 kings in poker (five cards)?

"Counting" Examples

Example 2: What is the probability of being dealt all 4 kings in poker (five cards)?
Solution: Easiest way is to just count:

1. Count all possible selections of five cards - $C(52,5)$ - then count all possible hands with all 4 kings: they only differ by their last card, which has 48 possible values. Then

$$P(\textit{4 Kings}) = \frac{48}{C(52,5)} = \frac{48!5!}{52!}$$

2. We could also try to do things *conditionally* by adding up the 5 ways to get this hand by drawing cards in order: $\{KKKKN, KKKNK, KKNKK, KNKKK, NKKKK\}$ (note that each are equally likely).

$$\begin{aligned}
&P(\textit{5 Kings}) = 5 * P(\textit{KKKKN}) \\
&= 5 * P(\textit{cards 1-4 all Kings})P(\textit{ card 5 is not GIVEN cards 1-4 are kings}) \\
&= 5 * P(\textit{cards 1-3 all Kings})P(\textit{ card 4 is a King GIVEN cards 1-3 are kings}) \\
&= 5 * P(\textit{cards 1-3 all Kings})\frac{1}{49} = \cdots = 5 * \frac{4}{52}\frac{3}{51}\frac{2}{50}\frac{1}{49}
\end{aligned}$$

"Counting" Examples

Example 3: What is the probability of being dealt a flush in poker (five cards)?

"Counting" Examples

Example 3: What is the probability of being dealt a flush in poker (five cards)?

Solution: Two ways

1. Count all possible selections of five cards - $C(52,5)$ - then count all possible selections of flushes: $C(13,5)$ for the values on the flush and $C(4,1)$ for the possible suits. Then

$$P(\textit{flush}) = \frac{C(13,5)C(4,1)}{C(52,5)}$$

2. Do things *conditionally*:

$$P(\textit{all 5 cards same suit})$$
$$= P(\textit{cards 1-4 match suit AND card 5 matches that suit})$$
$$= P(\textit{cards 1-4 match suit})P(\textit{ card 5 matches that suit GIVEN cards 1-4 match suit})$$
$$= \cdots = \frac{52}{52}\frac{12}{51}\frac{11}{50}\frac{10}{49}\frac{9}{48}$$

Daily Recap

Today we learned

1. More on probability theory
2. Breaking problems down into smaller parts using *conditioning*

Moving forward:

- nb day Friday!
- Monday: HW 2 due: make sure you have current version

Next time in lecture:

- We probably talk even more about probability!