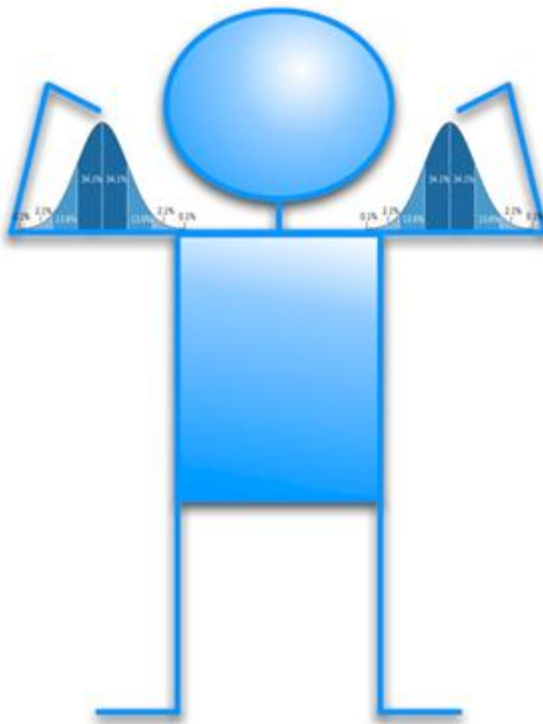# Stat S110

Conditioning is the soul of statistics.

Unit 2: Conditional Probability
Chapter 2 in the text

Unit 2 Outline

- Conditional Probability Definition

- Bayes' Rule

- Independence

- The Gambler's Ruin

- Conditionally Paradoxical

Why is conditional probability important?

- Conditional probability always us to update a probability statement given other information that is related.

- Suppose you have plans to attend the Red Sox game (or a picnic, an outdoor concert, etc…) tonight, and you are trying to determine whether it will rain later in the day.

- Before any information is gathered, you may put a probability on the event that is will rain: $P(R)$.

- However, you may look at the window and see that it is very cloudy. You certainly would want to update the probability of rain given this information: $P(R \mid C)$.

- You may gather more information throughout the day, so your conditional probability may change: $P(R \mid A \cap B \cap C)$.

- By conditioning on known information, we can improve our probability predictions!

3

Definition: Conditional Probability

- *conditional probability*: the probability of one event occurring under the condition that we know the outcome of another event

- Let A and B be two events in a sample space, with $P(B) > 0$. The conditional probability of event A, given that B has occurred, written $P(A/B)$, is:

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

- $P(A/B)$ is read as "probability of A, given B" has happened, or probability of A if B is true.

- $P(A)$ ignoring B is often called the *prior* probability, and $P(A/B)$ is often called the *posterior* probability incorporating knowledge/information of event B.

- We've essentially been using conditional probability already (sampling without replacement).

4

Conditional probability in Pebble World
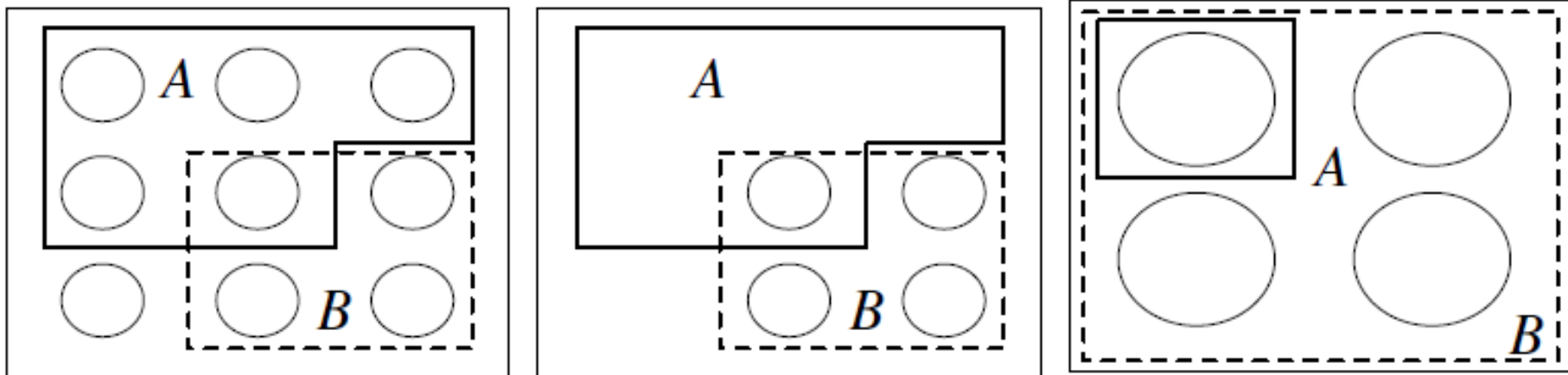
- We are trying to determine $P(A|B)$.
- On the left is the entire sample space. But we do not need to incorporate anything in B^C to determine $P(A|B)$.
- So in the end, we just restrict ourselves to only considering the outcomes in B and asking: how much of B does A take up?

Key Concept

- Conditional probability is probability!
- So all of the rules/axioms, results, and properties hold.
- For example:

$$P(S \mid E) = 1, \quad P(\varnothing \mid E) = 0$$

$$P(A \mid E) = 1 - P(A^C \mid E)$$

$$P(A \bigcup B \mid E) = P(A \mid E) + P(B \mid E) - P(A \bigcap B \mid E)$$

$$P(A \mid B \bigcap E) = \frac{P(A \text{ and } B \mid E)}{P(B \mid E)}$$

- And many more…

6

Simplest Example

- There is a bag with 3 balls in it: 1 is red, and 2 are black
- You draw two balls out of the bag, one at a time (without replacement). What is the probability the second ball is black given the first ball is black?
- Define the events:

  A: the first ball drawn is black
  B: the second ball drawn is black

First Ball ⟶

Second Ball ⟶

B1	B1	B2	B2	R	R
B2	R	B1	R	B1	B2

$$P(B\,|\,A) = \frac{P(A \cap B)}{P(A)} = \frac{P(Both\ Balls\ are\ Black)}{P(First\ Ball\ is\ Black)} = \frac{2/6}{4/6} = \frac{1}{2}$$

- What is the probability the 1st ball was black given the 2nd is black?

Very tricky…

- The Monty Hall Problem
  - There are prizes behind 3 doors: two are 'worthless' (an ant farm) and one is expensive (like a new car)
  - You are asked to choose one of the 3 doors
  - Then, Monty Hall (from Let's Make a Deal) opens one of the other 2 doors and shows you a worthless prize

- **Should you switch doors?**

- **NYTimes take:**
    http://www.nytimes.com/2008/04/08/science/08monty.html

A general multiplication rule (from conditional probability)

- Suppose A and B are two events in a sample space (not necessarily independent). Then

  - $P(A \cap B) = P(B \mid A)P(A) = P(A)P(B \mid A)$

  - $P(A \cap B \cap C) = P(A) P(B \mid A)P(C \mid A \cap B)$

  - In fact, for general $A_1,\ldots,A_n$:

  $P(A_1 \cap \ldots \cap A_n) = P(A_1) P(A_2/A_1)P(A_3/A_1 \cap A_2)\ldots P(A_n/A_{n-1} \cap \ldots \cap A_1)$

- The first relationship is a simple algebraic rearrangement of the definition of conditional probability:

$$P(B \mid A) = \frac{P(B \text{ and } A)}{P(A)}$$

9

A Personable Example

It is known that approximately 20% of men and 3% of women are taller than 6 feet in the *US*.

Let F = the event that someone is *female* and T = *taller than 6 feet*.

a) What is P(T | F)? What is P(T | F^C)?

b) What is the probability that the next person walking through the door is a woman **and** 6 feet tall?

c) What is the probability that the next person walking through the door is 6 feet tall?

Example (cont.)

c) What is the probability that the next person walking through the door is 6 feet tall?

Two ways for this to happen: (T and F) or (T and F^c) **[Think Venn Diagrams]**

$$P(T) = P(T \cap F) + P(T \cap F^C)$$
$$= P(T \mid F)P(F) + P(T \mid F^C)P(F^C)$$
$$= 0.03 \times 0.50 + 0.20 \times 0.50 = 0.115$$

2-way tables can help organize your thinking…

		Tall (6' or more)	
		Yes	No
Female	Yes	P(F)*P(T | F) = (0.5)*(0.03) = 0.015	P(F)*P(not T | F) = (0.5)*(0.97) = 0.485
	No	P(not F)*P(T | not F) = (0.5)*(0.20) = 0.100	P(not F)*P(not T | not F) = (0.5)*(0.80) = 0.400

P(F ∩ T)

Law of Total Probability

- Let $A_1, \ldots, A_n$ be a partition of the sample space S (that is, the A_i are disjoint and their union makes up the entire S) with $P(A_i) > 0$ for all A_i. Then:

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i)$$

- Illustration of law of total probability:
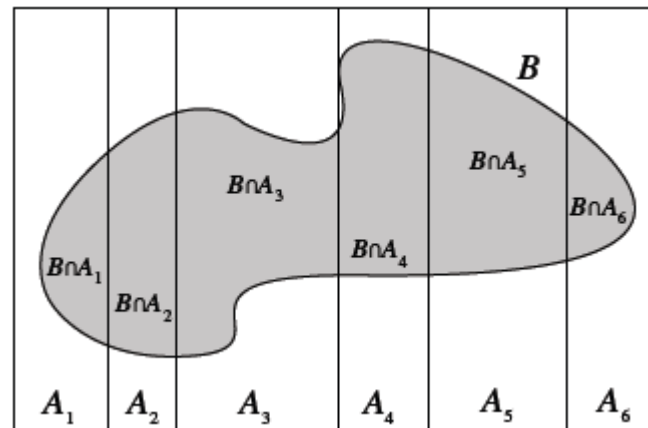
Unit 2 Outline

- Conditional Probability

- Bayes' Rule

- Independence

- The Gambler's Ruin

- Conditionally Paradoxical

Bayes' Rule

- Bayes' rule (formula) provides a way to go from $P(B\,|\,A)$ to $P(A\,|\,B)$ (they are in general not equal…)

- If A and B are two events whose probabilities are not 0 or 1, then:

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

- It is often written using the law of total probability:

$$P(A\,|\,B) = \frac{P(B\,|\,A)P(A)}{P(B\,|\,A)P(A) + P(B\,|\,A^{C})P(A^{C})}$$

- Or:

$$P(A_1\,|\,B) = \frac{P(B\,|\,A_1)P(A_1)}{\sum_{i=1}^{n} P(B\,|\,A_i)P(A_i)}$$

15

d) What is the probability that a person known to be 6 feet tall is a woman?

Bayes Rule Directly:

$$P(F \mid T) = \frac{P(T \mid F)P(F)}{P(T \mid F)P(F) + P(T \mid F^C)P(F^C)}$$

$$= \frac{0.03(0.5)}{0.03(0.5) + 0.2(0.5)} = 0.130$$

Pretty Simple from the 2x2 table…

$$P(F \mid T) = \frac{P(F \ and \ T)}{P(T)} = \frac{0.015}{0.015 + 0.100} = 0.130$$

Bayes' Rule Example: Random Coin

- You have one fair coin, and one biased coin which lands Heads with probability 3/4. You pick one of the coins at random and flip it three times. It lands Heads all three times. Given this information, what is the probability that the coin you picked is the fair one?

$$P(F \mid A) = \frac{P(A \mid F)P(F)}{P(A \mid F)P(F) + P(A \mid F^C)P(F^C)}$$

$$= \frac{(1/2)^3(1/2)}{(1/2)^3(1/2) + (3/4)^3(1/2)} \approx 0.23$$

- So what?

Random Coin Continued

- Suppose that we have now seen our chosen coin land Heads three times. If we toss the coin a fourth time, what is the probability that it will land Heads once more?

$$P(H \mid A) = P(H \mid F, A)P(F \mid A) + P(H \mid F^C, A)P(F^C \mid A)$$
$$= (1/2)0.23 + (3/4)(0.77) \approx 0.69$$

Unit 2 Outline

- Conditional Probability Definition

- Bayes' Rule

- **Independence**

- The Gambler's Ruin

- Conditionally Paradoxical

Independent events

- Two events *A* and *B* are *independent* if and only if knowing that one event occurs does not change the probability that the other event occurs. What does that mean for the conditional probabilities?

$$P(A \mid B) = P(A)$$

- The following are equivalent definitions of independence:

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

$$P(A \mid B) = P(A \mid B^C)$$

$$P(B \mid A) = P(B)$$

Note: it takes care to draw independence in a Venn Diagram…

20

An Example

- There is a bag with 3 balls in it: 1 is red, and 2 are black

- You draw two balls out of the bag, one at a time (without replacement). Define the events:

  A: the first ball drawn is black

  B: the second ball drawn is black

- Are *A* and *B* independent? How do you know?

A practice problem

Ultrasound is often used to determine the sex of an unborn baby. However, because the procedure relies on visual detection of anatomic differences between male and female babies, the error rates differ according to whether the baby is a boy or a girl.

- P(ultrasound predicts male | baby is male) = 75%
- P(ultrasound predicts female | baby is female) = 90%

(a) Consider an individual woman who comes to the clinic for an ultrasound to predict her baby's sex. What is the probability that the ultrasound gives the wrong result?

(b) Suppose a clinic performs 10 ultrasounds a day, and the ultrasound results are independent. What is the probability of one or more incorrect sex determinations?

(c) What is the probability that a baby is male, given that the ultrasound predicts a male?

22

Solution

Let $B = \{\text{baby is a boy}\}$, $B^C = \{\text{baby is a girl}\}$

$A = \{\text{ultrasound predicts boy}\}$, $A^C = \{\text{ultrasound predicts girl}\}$

(a) $P(Wrong\ result) = P(A\ and\ B^C) + P(A^C\ and\ B)$

$$= P(A | B^C)P(B^C) + P(A^C | B)P(B)$$

$$= 0.10(0.50) + 0.25(0.50) = 0.175$$

(b) $P(\text{at least 1 out of 10 are wrong predictions})$

$$= 1 - P(0\ \text{wrong predictions}) = 1 - P(10\ \text{correct predictions})$$

$$= 1 - (0.825)^{10} = 1 - 0.146 = 0.854$$

(c) $P(boy | ultrasound\ predicts\ boy) = P(B | A)$

$$= P(B\ and\ A) / P(A) = P(B\ and\ A) / [P(B\ and\ A) + P(B^C\ and\ A)]$$

$$= P(A | B)P(B) / [P(A | B)P(B) + P(A | B^C)P(B^C)]$$

$$= [0.75(0.50)] / [0.75(0.50) + 0.10(0.50)] = 0.375 / 0.425 = 0.882$$

23

Independence of 3 events

- Three events *A*, *B,* and *C* are ***independent*** if and only if all of the following hold:

$$P(A \bigcap B) = P(A)P(B)$$
$$P(A \bigcap C) = P(A)P(C)$$
$$P(B \bigcap C) = P(B)P(C)$$
$$P(A \bigcap B \bigcap C) = P(A)P(B)P(C)$$

- If the first 3 conditions hold, then we say *A*, *B,* and *C* are ***pairwise independent***, but this does not mean that 4th condition holds. Example: *A* = first toss is heads, *B* = second toss is heads, *C* = both tosses have the same result.

- Independence of 4+ events gets unwieldy very quickly…

Conditional Independence

- Two events *A* and *B* are ***conditionally independent given E*** if and only if:

$$P(A \bigcap B \mid E) = P(A \mid E)P(B \mid E)$$

- Two events can be conditionally independent given E, but not independent themselves. And vice versa. It's easy to make either mistake.

- For example (conditional independence given *E* vs. given E^C): Suppose there are 2 types of classes: good classes (*G*) and bad classes. In a good class, if you work hard (*W*) you are likely to get an A (*A*). In bad classes the professor randomly assigns grades to students. Then *W* and *A* are conditionally independent of G^C, but not given *G*.

25

Unit 2 Outline

- Conditional Probability Definition

- Bayes' Rule

- Independence

- The Gambler's Ruin

- Conditionally Paradoxical

The Gambler's Ruin

- Suppose there are two gamblers, A and B (B could be the *house*), and they make a sequence of $1 bets. For each bet, gambler A has probability p of winning, and gambler B has probability $q = 1 - p$ of winning. Gambler A starts with i dollars and gambler B starts with $N - i$ dollars.

- The amount of money that gambler A has can be viewed as a *random walk* on the integers 0 through N, with probability p of going to the right in a given step and probability $q = 1 - p$ of going to the left. Visually:
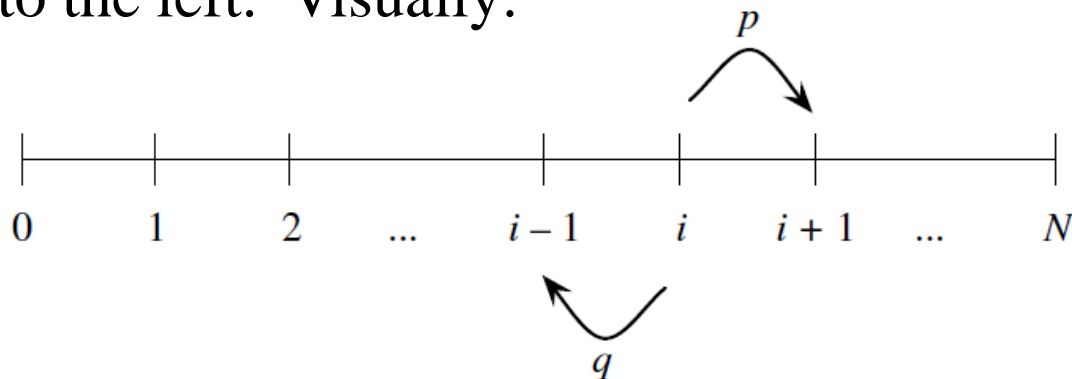
The Gambler's Ruin

- What is the probability that A walks away with all the money?

- Key: define W = event that A wins all the money, p_i = probability that A wins all the money for a value of i, condition on the first step, use LOTP, and *reset* your starting point.

- Thus:

$$p_i = P(W \mid A \text{ starts at } i, \text{wins } 1^{st} \text{ round}) \cdot p + P(W \mid A \text{ starts at } i, \text{loses } 1^{st} \text{ round}) \cdot q$$

$$= P(W \mid A \text{ starts at } i+1) \cdot p + P(W \mid A \text{ starts at } i-1) \cdot q$$

$$= p_{i+1} \cdot p + p_{i-1} \cdot q$$

The Gambler's Ruin

$$p_i = p_{i+1} \cdot p + p_{i-1} \cdot q$$

- This is a *difference equation*, which can be solved based on the fact that p_i will have the form of x^i (see Math Review for more details). This can then be reduced to the polynomial:

- $px^2 - x + q = 0$, which has two roots: 1 and q/p. So unless q = p = ½, the general solution is:

$$p_i = a \cdot 1^i + b \cdot \left(\frac{q}{p} \right)^i$$

- Since we know that $p_0 = 0$ and $p_N = 1$, we can solve the above to find the values of a and b. Specifically:

$$a = -b = \frac{1}{1 - \left(\dfrac{q}{p} \right)^N}$$

29

The Gambler's Ruin

- What about when $p = ½$? There is only one root, and it is of the form:

$$p_i = a \cdot 1^i + b \cdot i \cdot 1^i$$

- And again solving for the boundary condition, we find that $a = 0$ and $b = 1/N$.

- So the probability that A wins when starting with i dollars is:

$$p_i = \begin{cases} \left(1 - \left(\dfrac{q}{p}\right)^i\right) \Big/ \left(1 - \left(\dfrac{q}{p}\right)^N\right) & \text{if } p \neq 1/2 \\ i/N & \text{if } p = 1/2 \end{cases}$$

- How do these two conditions make sense? In what way are they consistent with each other?

30

The Gambler's Ruin: Take-Home

- So what's the take home message?
- Well first let's explore the probability of B winning all the money. By symmetry, we just need to change the starting wealth to $N - i$ and switch the roles of q and p.
- So for all values of both i and p, it can be shown that P(A wins) + P(B wins) = 1.
- What does this mean?
- This means the game is guaranteed to end! The game can never be *stuck* in neverending oscillation!

Unit 2 Outline

- Conditional Probability Definition

- Bayes' Rule

- Independence

- The Gambler's Ruin

- **Conditionally Paradoxical**

Prosecutor's Fallacy
(http://en.wikipedia.org/wiki/Sally_Clark)

- In 1998 Sally Clark was arrested and tried for murder after her two sons died shortly after birth (possibly due to sudden infant death syndrome, or SIDS). During the trial an expert witness testified that 1/8500 newborns die of SIDS, so the probability of two babies in the same fanily dying in a row of SIDS was $(1/8500)^2 \approx 1$ in 73 million. The expert continued and said the probability of Clark's innocence was 1 in 73 million.

- What are two major problems with this line of reasoning?

  1) Independence of children dying of SIDS

  2) Incorrectly interpreting conditional probability (not using Bayes rule to flip the conditioning).

Simpson's Paradox

- Two doctor's, Dr. Hibbert and Dr. Nick, each perform two types of surgeries: heart surgery and band-aid removal, with the following results:

	Heart	Band-Aid
Success	70	10
Failure	20	0

Dr. Hibbert

	Heart	Band-Aid
Success	2	81
Failure	8	9

Dr. Nick

- What is the overall probability of success for each doctor?
- What is the probability of success of each surgery type separately, for each doctor?
- Which doctor would you see?

Simpson's Paradox

- For events *A*, *B*, and *C*, we have a ***Simpson's paradox*** if:

$$P(A \mid B, C) < P(A \mid B^C, C)$$

$$P(A \mid B, C^C) < P(A \mid B^C, C^C)$$

but,

$$P(A \mid B) > P(A \mid B^C)$$

- What are events *A*, *B*, and *C* in the Dr. Hibbert v. Dr. Nick example?

UC Berkeley 1973 Sex Bias Case

http://en.wikipedia.org/wiki/Simpson%27s_paradox

- UC Berkeley sued in 1973 for sex bias in admissions to graduate schools.
- Women appeared to have lower acceptance rates than men.
- Key idea: important to condition on the correct events/variables. UC Berkeley was admitting students based on School/Major, so that should be taken into account.

	Applicants	% admitted
Men	8442	44%
Women	4321	35%

Major	Men		Women	
	Applicants	% admitted	Applicants	% admitted
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	272	6%	341	7%

Last Word: The other Simpsons' Paradox

Could Jesus microwave a burrito so hot, that he himself could not eat it?
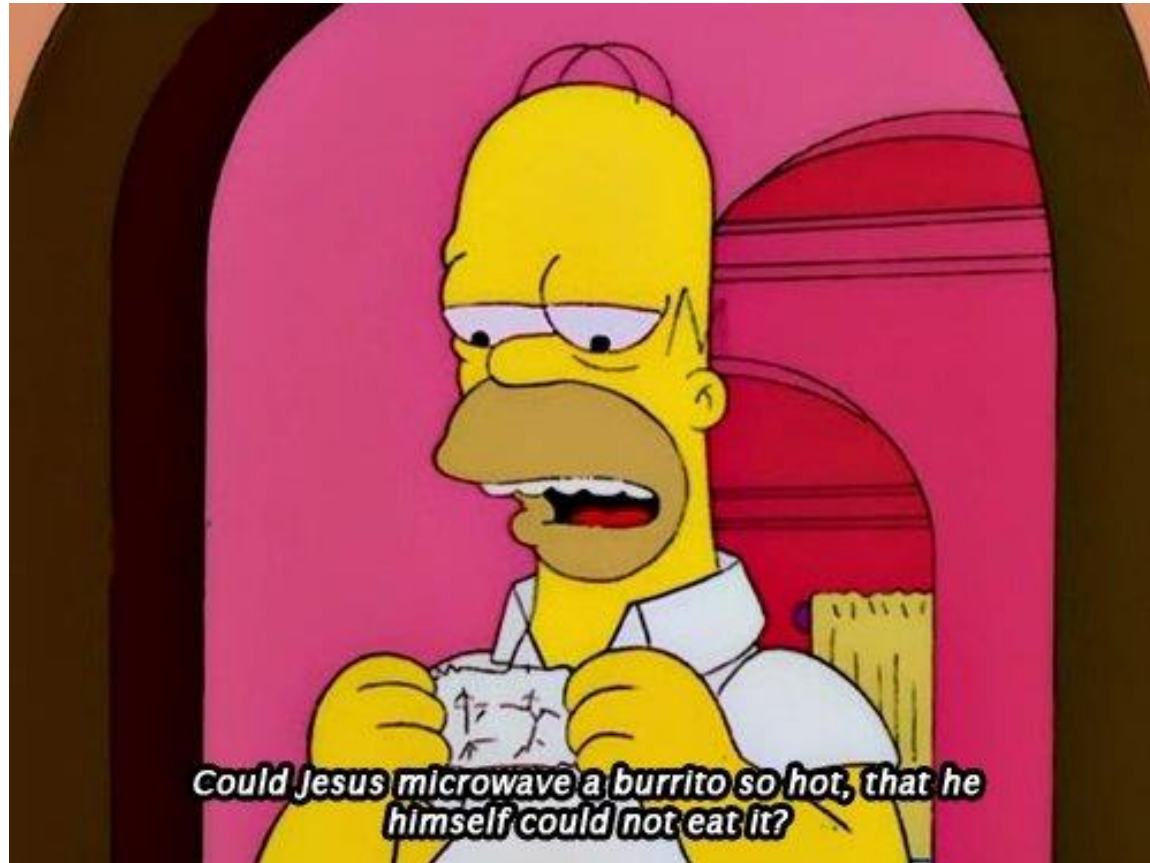
HOMER: Hey, I got a question for you. *(pulls out a piece of paper)* "Could Jesus microwave a burrito so hot that he himself could not eat it?"
NED: Well sir, of course, he could, but then again... wow, as melon-scratchers go that's a honey-doodle.