CS/Biol 123B
Deck 2: The Statistics Lecture

Probability and Statistics without numbers

"Casablanca", 1942

The story of statistics is a gambling story.

"Thunderball", 1965

Believe it or not, some gambling establishments are dishonest!

- "Is this a game of chance, Mr. Fields?"
- "Not the way I play it, no."

"My Little Chickadee", 1940

Believe it or not, some gambling establishments are dishonest!

- "You do know how to play poker don't you?"
- "You look around the table and if you can't spot the sucker, you're the sucker."

"Quiz Show", 1994

Blaise Pascal 1623-1662

- A brilliant mathematician with wealthy friends.
- Gambling was popular among the super-wealthy in Europe.
- Gamblers have 2 kinds of question for mathematicians:
  - Is a bet wise?
  - Is the game fair?

Probability: What is the chance of some "event"

- P(an odd number in roulette) = ?

- P(roll total 7 or 11 with 2 dice) = ?

- P(get a royal flush) = ?

What we say and what we mean

- We say: "What is P(event)?"
- We mean: "What is P(event), ***assuming the game is fair***"?

"Conditional" notation: P(event | condition)

- P(odd number in roulette | honest wheel) = ?

- P(roll total 7 or 11 with 2 dice | fair dice) = ?

- P(get a royal flush | a fair deck) = ?

The 2 kinds of roulette question

"Casablanca", 1942

- While you're playing: "What is the probability that the next roll will be red? (Assuming a fair wheel)"
  - P(red | fair wheel) = ?

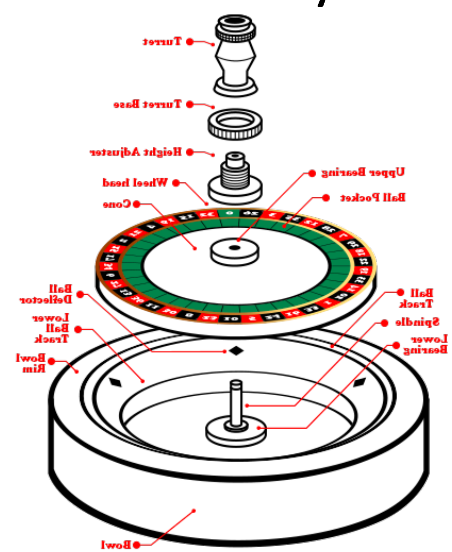
- Walking home because you can't afford Uber fare because you lost all your money because you bet on red 500 times and *never* won: "What is the <u>probability</u> that the wheel is unfair?"
  - P(unfair wheel | 500 not-reds) = ?

P(red | fair wheel)

- Completely knowable by observing what is accessible to you
- Look at the wheel and do the math
- 18 of the 38 slots are red => P(red|fair) = 18/38

P(unfair wheel | 500 not-reds)

- "Unfair wheel" can't be known by observing what's accessible and doing the math:
  - You can't recognize an unfair wheel by looking at it.
  - The casino management won't let you handle it or take it apart.
  - You can't interview the manufacturer, because the casino won't tell you who made it and anyway the makers won't talk to you.
- You can't know, you can only *hypothesize*.



2 very different issues:

- P(an outcome | an explanation about how the universe works)
  - We can rely on the explanation

- P(an explanation about how the universe works | an outcome)
  - The explanation is a *hypothesis* about the unknowable.
  - We call P "the likelihood" rather than "the probability".
  - We accept a hypothesis if its likelihood is high compared to likelihoods of all alternative hypotheses.
  - High = ???? Depends on the situation and your inclination.

- As scientists, we're in the business of explanations about how the universe works

2 very different issues:

- P(an outcome | ~~an explanation about how the universe works~~)
  - We can rely on the ~~explanation~~

- P(~~an explanation about how the universe works~~ | an outcome)
  - The ~~explanation~~ is a *hypothesis* about the unknowable.
  - We call P "the likelihood" rather than "the probability".
  - We accept a hypothesis if its likelihood is high compared to likelihoods of all alternative hypotheses.
  - High = ???? Depends on the situation and your inclination.

- As scientists, we're in the business of ~~explanations about how the universe works~~

2 very different issues:

- P(an outcome | hypothesis)
  - We can rely on the hypothesis

- P(hypothesis | an outcome)
  - The hypothesis is a *hypothesis* about the unknowable.
  - We call P "the likelihood" rather than "the probability".
  - We accept a hypothesis if its likelihood is high compared to likelihoods of all alternative hypotheses.
  - High = ???? Depends on the situation and your inclination.

- As scientists, we're in the hypothesis business.

Hypothesis (plural = hypothes<u>e</u>s)

- A statement about how the universe works
- Consistent with evidence
- Subject to disproof or confirmation
  - A single negative observation is enough to disprove
    - All sheep are white

→ disproved

  - A positive observation supports, increases confidence, doesn't prove
    - All sheep are white

→ supported

So you observed 6 white sheep...

- Hypothesis: all sheep are white
- There are ~1 billion sheep
- Exactly how much confidence should we gain from our observation?
- Can we confidently make a claim about the other 999,999,994 sheep?
- How many white sheep do we have to observe, in order to be 90% confident in the hypothesis? 95% 99.99999%?

Reverend Thomas Bayes (1701-1761)

- Presbyterian minister, philosopher, statistician.
- Bayes' Theorem: simple math, complicated philosophical implications.

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

Wikipedia

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

- In general, P(A|B) != P(B|A)
- If we know one and want the other, we need Bayes' Theorem [and also P(A), P(B)] … then it's just arithmetic.
- Examples:

A	B
Randomly picked card is a 10	That card is a ♡
A person smokes	That person has cancer
A student will ride the bus to SJSU on the 1st day of F21	It will rain in San Jose on the 1st day of F21

And here it gets weird

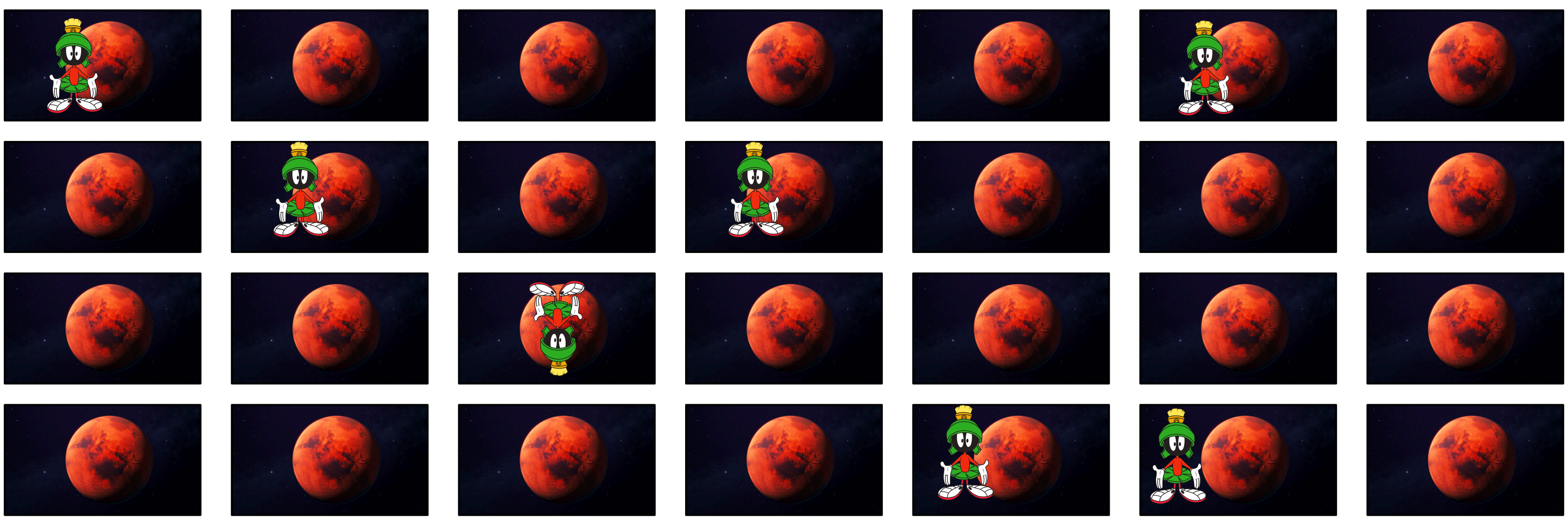
$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

- Bayes' formula doesn't care what kinds of thing A and B represent.
- B might be a hypothesis
  - P(A | hypothesis) makes sense
  - P(hypothesis | A) makes no sense … or maybe it makes a different kind of sense
- Ordinary conditional probabilities makes sense in terms of *expectations*:
  - P(rides bus | rain) = .75 means if it rains on 8/18/21 and we observe 100 SJSU students that morning, we expect approximately .75 * 100 = 75 of them to ride the bus.
  - If it rains on 8/18/21 and we observe 1000 SJSU students that morning, we expect approximately .75 * 1000 = 750 of them to ride the bus.

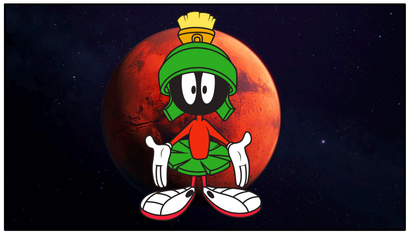
What does P(hypothesis) mean? Or P(hypothesis|observations)?

- Example:
  - Hypothesis: There is life on Mars
  - P(There is life on Mars) = .25

- What it doesn’t mean: If we observe 28 Marses, we expect there will be life on ~7 of them.



What does P(life on Mars) = .25 mean?



- A number but not a probability.
- A measure of my confidence, the strength of my belief.
- What can influence strength of belief?
  - Evidence ← The only scientifically valid influence
  - Trust in the source
  - Hope
  - Culture
  - Fear of the alternative
- We're getting into philosophy here.
- P(hypothesis|observation) is a different kind of thing from P(observation|hypothesis) → We use a different word for it: "Likelihood".

One huge difference: interpreting P

- With probabilities, just think about the value.
  - P(Win Powerball jackpot) ~= 1/300,000,000 ➔ keep your dollar.
  - P(Survive a plane trip) ~= 10,999,999/11,000,000 ➔ don't worry.
- With likelihoods, you usually can't interpret a value on its own.
- A tiny likelihood doesn’t always mean what it seems to mean. And likelihoods are often tiny.
- 2 situations when you reason with likelihoods:
  - If you have 1 hypothesis, and some data
    - Hypothesis testing
  - If you have 2 or more competing hypotheses, and some data
    - E.g. multiple Markov Models or Hidden Markov Models

Hypothesis testing: just the basics

- "Null hypothesis", written H_0, pronounced "H null".
- Roughly defined as the prevailing / accepted explanation for some part of the universe.
  - The last dragon died 100 years ago. It is known, Khaleesi.

    "Game of Thrones", Season 1
  - The mean height of an SJSU student is 5'7".

    I made it up
  - The mean age of a star in our galaxy is 11.4 billion years

    https://www.quora.com/How-old-is-the-average-star-in-our-galaxy

Hypothesis testing: collect data, try to overthrow H_0

- I'm sorry to break it to you like this, but GoT is *not* realistic.
- H_0: Dragons are extinct.
- Daenerys engaged in risky behavior, found 3 counterexamples, and decisively disproved H_0.
- It's almost never decisive.

In the real world, we rarely *disprove* H_0

- H_0 is often a statistical statement, e.g. about mean height of people or mean age of stars. Disproof by counterexample doesn't apply to such hypotheses.
- As in a legal trial, we produce evidence that suggests H_0 is wrong *beyond a reasonable doubt*.
- "*Beyond a reasonable doubt*" is rigorously defined and quantified using statistical reasoning. (But not in 123B.)
- If that happens, we say that we ***reject*** H_0.
- H_0 = The mean height of an SJSU student is 5'7"
  - To prove or disprove, measure all 30,000 students.
  - Not practical.
  - Measure a reasonable number of students. Reject H_0 if the heights cast doubt (by certain rigorous standards) on H_0.
- H_0 = age of a star in our galaxy is 11.4 billion years
  - To prove or disprove, measure all 100 billion stars.
  - Seriously not practical.
  - Measure a reasonable number of stars. Reject H_0 if the ages cast doubt (by certain rigorous standards) on H_0.

To decide about rejecting H_0, think about the p value

- The p value measures the the strength of the support that your data gives to H_0.
- You challenge a hypothesis because you suspect there are forces at work beyond what the hypothesis explains.
- Your suspicion comes from some pattern you saw in your data. The pattern seems inconsistent with H_0.
- Possible reasons for what you saw:
  - A force of nature that nobody noticed until now. It's not explained by H_0. You have made a genuine discovery.
  - Bias in your data. When you measured some SJSU students, you were outside the gym just after basketball practice. Oops.
  - A weird random coincidence. Everyone you measured was < 5'3", for no particular reason.

p-value answers the question "How weird is *that*?"

- I just measured 20 SJSU students, and they are all over 6'3". How weird is *that*?
- I just computed the age of 100 stars in our galaxy, and 5 of them are younger than a million years. How weird is *that*?
- "How weird is *that*?" means "How small is P(*my data*|H_0)?"
- For reasons we won't go into, P(*my data*|H_0) isn't useful, and "How weird is *that*?" isn't the best question to ask.
- Instead we ask a more complicated question: If H_0 is true, what's the probability of seeing *my data* or anything even more unusual?
- p-value = P(Observing any X) , where P(X|H_0) <= P(*my data*|H_0).

Example

- H_0: Only 1 Galapagos tortoise out of 100 lives longer than 225 years.
- You just observed 10 Galapagos tortoises, and their mean age was 268 years old. *How weird is that?*
- The not useful number: P(your observation|H_0) = P(observe 10 tortoises where mean age = 268|H_0)
- The useful number: p-value =

  P(observe 10 tortoises, where mean age >= 268|H_0)

Example

- H_0: Spaceflight doesn't affect expression of heat shock protein 8 in mice.
- Heat shock protein expression in a space mouse was 15% of the expression in a ground control mouse. *How weird is that?*
- The not useful number: P(your observation|H_0) = P(space mouse expresses heat shock protein 8 at exactly 15% of ground-control expression|H_0).
- The useful number: p-value =

P (space mouse expresses heat shock protein 8 at <= 15% of ground-control expression|H_0).

Using p-value

- Intuitively, p-value is the probability of rejecting the null hypothesis by mistake.
- Reject H_0 if p-value is < some threshold.
- Threshold of .05 is common.
- Higher cost of being wrong ➔ need more confidence ➔ use a smaller threshold.

E-values: like p-values for blast

- The strict definition of the E-value of a blast hit:
  - with a query sequence of length L
  - against some database DB
  - that hits a subject with % identity x
- E-value = probability that
  - blasting a random query sequence of length L
  - against the same database DB
  - will hit a subject with % identity >= x
- An imperfect interpretation (but not bad): E-value is the probability that the similarity between the query and the subject is due to coincidence, rather than evolutionary relationship.

E-values: like p-values for blast

- If you hypothesize that the similarity between the query and the subject is coincidence, then you reject that hypothesis if the E-value is small.
- If you reject the hypothesis, the only remaining possibility is evolutionary relationship, i.e. descent with mutation from a common ancestor.
- A threshold of 0.01 is common. As with p-values, different situations call for different thresholds.