Introduction to Probability

Grinnell College

October 27, 2025

Motivation

It's 3:30am on a Sunday in a seedy, smoke-filled, underground bar in Port Arthur TX

You've drank a bottle and some change of rye since the night started

You are playing a texas hold 'em against some biker who looks like he did time and you got a lot of money riding on this hand

Cards dealt are:

- You: A♠ A♣
- Community: K♦ 10♦ 8♦ 4♠ A♥
- Biker: ? ?

What is the probability you have gas money back to Iowa?

Solutions

There are several strategies to get an answer to this question. We could...

- find the proportion of cards that can be dealt that would beat our hand
  - Theoretical probability
- guesstimate
  - Subject probability
- use previous hands dealt in this exact situation to use our data to say something
  - Empirical probability
  - Not applicable in this example
- simulate what possible hands could have been dealt and finding the proportion of the simulations that beat our hand
  - No name that I'm aware of (simulated probabilities?)
  - Touched on this one already

But what is probability?

Guesses?

What is Probability?

The **probability** of an event is the long term frequency of that event happening (contentious; frequentist school of thought)

- Eg Rolling a five on a dice has a $\frac{1}{6}$ probability of happening because we expect roughly a sixth of our rolls to be 5's in th elong run
- measured between 0 and 1
- closer to zero $=$ less likely
- closer to one $=$ more likely
- No applicability on a single, non-repeating event
  - Eg I will or will not wear a costume for Halloween this year to class
  - No repitions -$>$ no frequencies -$>$ no probabilities

(The Bayesian school of thought thinks of probability as a personal belief an event will happen or as the shared expectation that an event will happen)

Uniform Distribution

When multiple outcomes are equally likely, the probability for each outcome is

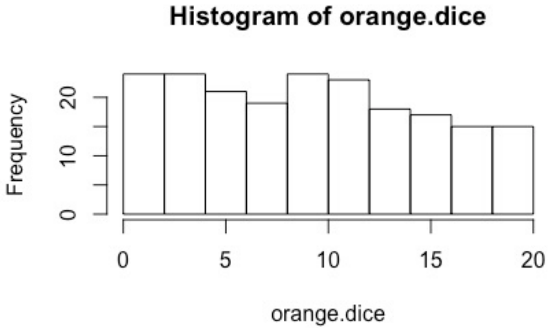
$$\frac{1}{\text{\# of all possible outcomes}}$$

These are called uniform distributions and include rolling a die, flipping a coin, etc...

Examples:
Flipping a coin and getting a head: 1 out of 2 possibilities -> P(heads) = $1/2 = 0.5$

Probability of rolling a crticial failure (1) on a 20-sided dice?

Uniform Distribution



Back in the day we would say this doesn't really seem to have a mode (uniform), no outliers, and symmetrical

Probability more broadly

If multiple outcomes produce the same event, the probability of that event is...

$$\frac{\text{\# of outcomes that cause event}}{\text{\# of all possible outcomes}}$$

My barbarian DnD character can get a critical hit if I roll a 19 or 20. What is the probability I get a critical hit on a given roll?

Probability more Broadly

If multiple outcomes produce the same event, the probability of that event is...

$$\frac{\#\text{ of outcomes that qualify}}{\#\text{ of all possible outcomes}}$$

My barbarian DnD character can get a critical hit if I roll a 19 or 20. What is the probability I get a critical hit on a given roll?

$\frac{\text{I roll either a 19 or 20}}{\text{20 possible rolls}} = 2/20 = 1/10$

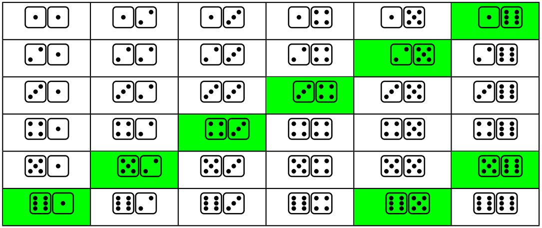
Dice Game: Casino Craps

In the game of craps in a casino, on the first roll, if the player (“shooter”) rolls a 7 or 11 they win. What proportion of dice rolls can do this?

⚀⚀	⚀⚁	⚀⚂	⚀⚃	⚀⚄	⚀⚅
⚁⚀	⚁⚁	⚁⚂	⚁⚃	⚁⚄	⚁⚅
⚂⚀	⚂⚁	⚂⚂	⚂⚃	⚂⚄	⚂⚅
⚃⚀	⚃⚁	⚃⚂	⚃⚃	⚃⚄	⚃⚅
⚄⚀	⚄⚁	⚄⚂	⚄⚃	⚄⚄	⚄⚅
⚅⚀	⚅⚁	⚅⚂	⚅⚃	⚅⚄	⚅⚅

Dice Game: Casino Craps

In the game of craps in a casino, on the first roll, if the player ("shooter") rolls a 7 or 11 they win. What proportion of dice rolls can do this?



8/36 possible rolls, or .2222.

Here, the proportion of equally likely outcomes is exactly the probability we "win" on the first roll

Types of Probability

Subjective Probability:

- How likely an event is to happen based on someone's personal belief / experience / feelings
- Most likely different answers from different people
- Ex: prob. of a sports team winning their next game?

Types of Probability

Theoretical Probability:

- Based on formulas or assumptions about the event
- Eg the dice game from above
- Common assumption is the probabilities are equal between cateogries

Example: Suppose there are 20 marbles in a bag. 2 marbles are red, 6 are blue, and 12 are green.

- prob. of pulling red marble?
- prob. of blue?
- prob. of green?
- prob. of orange?
- prob. of pulling a marble?

Types of Probability

Empirical Probability:

- How likely an event is to happen based on collected data
- Sometimes we estimate the probability with data in the form of a table
- Ex: flip a coin 1000 times and find the 'empirical' probability of getting a Heads

Law of Large Numbers:
If you repeat trials a whole bunch (and they don't affect each other) then the empirical probability will converge to the "true" probability

Empirical Examples

A report published in 1988 summarizes results of a Harvard Medical School clinical trial determining effectiveness of asprin in preventing heart attacks in middle-aged male physicians

	Heart Attack		
Treatment	Attack	No Attack	Total
Placebo	189	10,845	11,034
Aspirin	104	10,933	11,037
Total	293	21,778	22,071

Marginal vs Conditional

We have two other types of probabilities people talk about but they are different from the previous three.

- The previous three were ways of estimating/calculating the probability of some event

Instead, we are going to talk a way to distingish between a statistic that is about the entire population (marignal) or about a subpopulation (conditional)

Marginal Probability

A **marginal probability** is the probability associated with a variable, ignoring all other variables.

- Ie the probability is calculated for/with the entire population
- Eg the probability a randomly selected college would have a tuition rate higher than Grinnell's for all colleges and universities in the US
- Usually the default when people say "the probability"
  - "The probability you get into a car wreck on the way to the airport is higher than the probability you get into a plane crash" doesn't account for if you've been drinking or not (or if your pilot has been drinking...)

What is the probability a randomly selected physician had a heart attack?

Conditional Probability

A **conditional probability** is the probability associated with variable A given variable B

- Ie a probability is calculated only for a particular subpopulation
- We have already seen this!! Graphs broken apart by public vs private colleges
- Eg the probability a randomly selected college would have a tuition rate higher than Grinnell's for ONLY private colleges and universities in the US

What is the probability a physician had a heart attack given they took aspirin? What about if they didn't take aspirin?

Notation

P(E) is used to denote the probability of some event, E

- Often use a letter or abbreviation for the event
- P(patient having a heart attack) $\rightarrow$ P(heart attack) $\rightarrow$ P(H)
- Marginal probabilities are always(?) written like this

P(A|B) is used write the conditional probability of A given B

- P(patient has a heart attack given they take aspirin) $=$ P(H | Aspirin) $=$ P(H|A)

Try on your own!

The following is a table displaying the Marriage vs Divorced/Seperated vs Other (never married, widowed, no answer) for Protestants and Catholics in surveys taken over 2000-2014 period from the General Social Survey (source: "gss_cat" data set in R)

Branch	Other	Div/Sep	Married	Total
Catholic	1717	864	2543	5124
Protestant	3302	2165	5379	10846
Total	5019	3029	7922	$N = 15970$

1. What is the probability we randomly grab a Catholic?
2. What is the probability we randomly someone who is married?
3. What is the probability we randomly grab a divorced or separated Protestant?
4. Given we are sampling only from Catholics, what is the probability a randomly chosen Catholic is divorced?

Operations for Probability

Broardly follow set theory notation if you know that...

- **Union**: The probability that either A or B happens (inclusive "or"!!)
  - $P(A \cup B)$ or P(A or B)
  - Eg what is the probability someone is Catholic or was in "Other"
- **Intersection**: The probability that both A and B happen (both have to happen at the same time)
  - $P(A \cap B)$ or P(A and B)
  - What is the probability we randomly grab a divorced Protestant?
- **Compliment**: The probability of NOT A
  - $P(A^c) = 1 - P(A)$
  - Eg what is the probabiity someone hasn't gone through a divorce?

Operations of Probability 2

- **Independence**: Two events don't affect one another
  - $P(A \cap B) = P(A) * P(B)$
- **Disjoint Events**: Only one of the possible events is possible
  - $P(A \cap B) = 0$
  - I can roll a 4 or a 5 on a six-sided dice but not both at the same time
- **Additive Rule**: The probability of a union of A and B is NOT P(A) + P(B)
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - Otherwise $P(A \cap B)$ is included twice
  - If A and B are disjoint, $P(A \cup B) = P(A) + P(B)$

Operations of Probability 3: Conditional

Conditional Probabilities have their own formula...

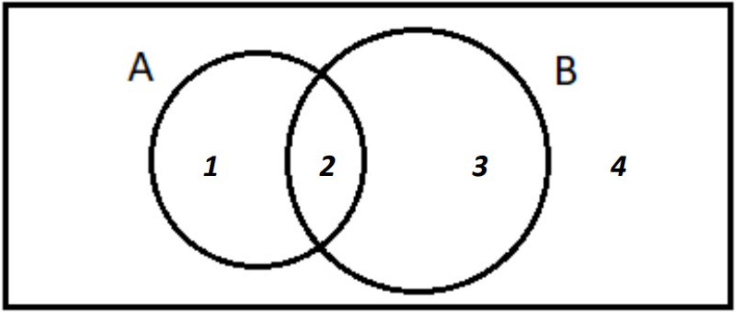
- $P(A \mid B)$ = Probability of A occurring if B occurred
- $P(A \mid B) = P(A \cap B) \,/\, P(B)$

Multiplicative Rule

- Formula for finding the probability of the intersection of two events
- Reorganized version of the above
- $P(A \cap B) = P(A \mid B) * P(B)$

Venn Diagrams

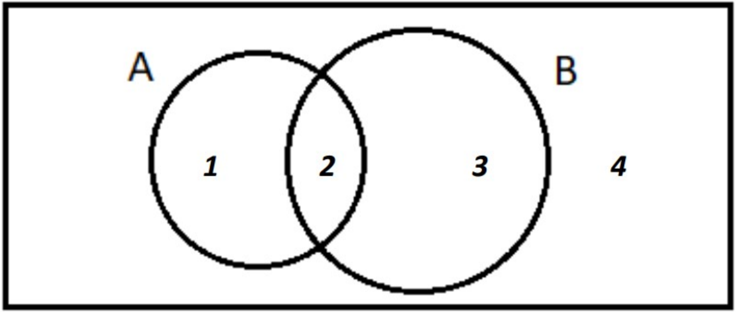
Venn Diagrams help us imagine situations of how two events might be related



1. What is the Probability of A?
2. What is the probabililty of the intersection of A and B?
3. $P(B^c)$?
4. $P(A \cup B)$?

Venn Diagrams

Venn Diagrams help us imagine situations of how two events might be

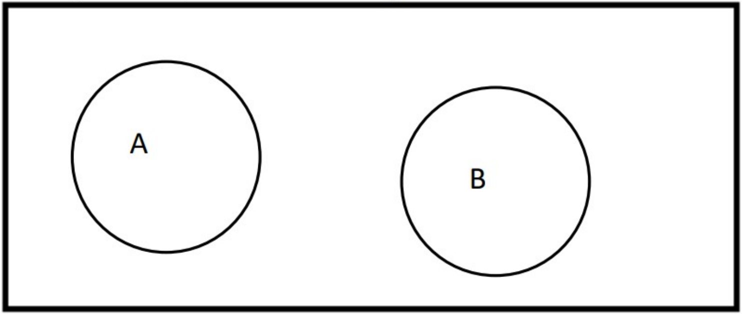


related

1. What is the Probability of A? Areas 1 and 2
2. What is the probabililty of the intersection of A and B? Area 2
3. P(B^c)? Areas 1 and 4 (everything outside of "B")
4. P($A \cup B$)? Areas 1, 2, and 3

Venn Diagrams 2

And here is an example of disjoint events



Probabilities for Continuous Distributions

Thus far we have basically been talking about probability for categorical or discreet data but continuous data gives a different challenge.

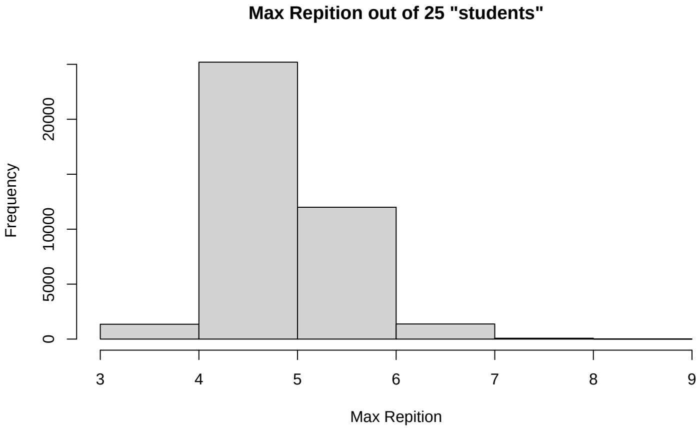
Basically, we talk about the area under the curve of a histogram (with extremely small bins). The proportion of area under the curve is our probability.

More on this in the next few lessons.

Why did we learn this?

Let's try a game: write down ten digits (0-9)

Why did we learn this?



Probability that a class that randomly generates it's numbers wouldn't have 4 repitions or higher is .0338.

This is the basis of statistical testing.