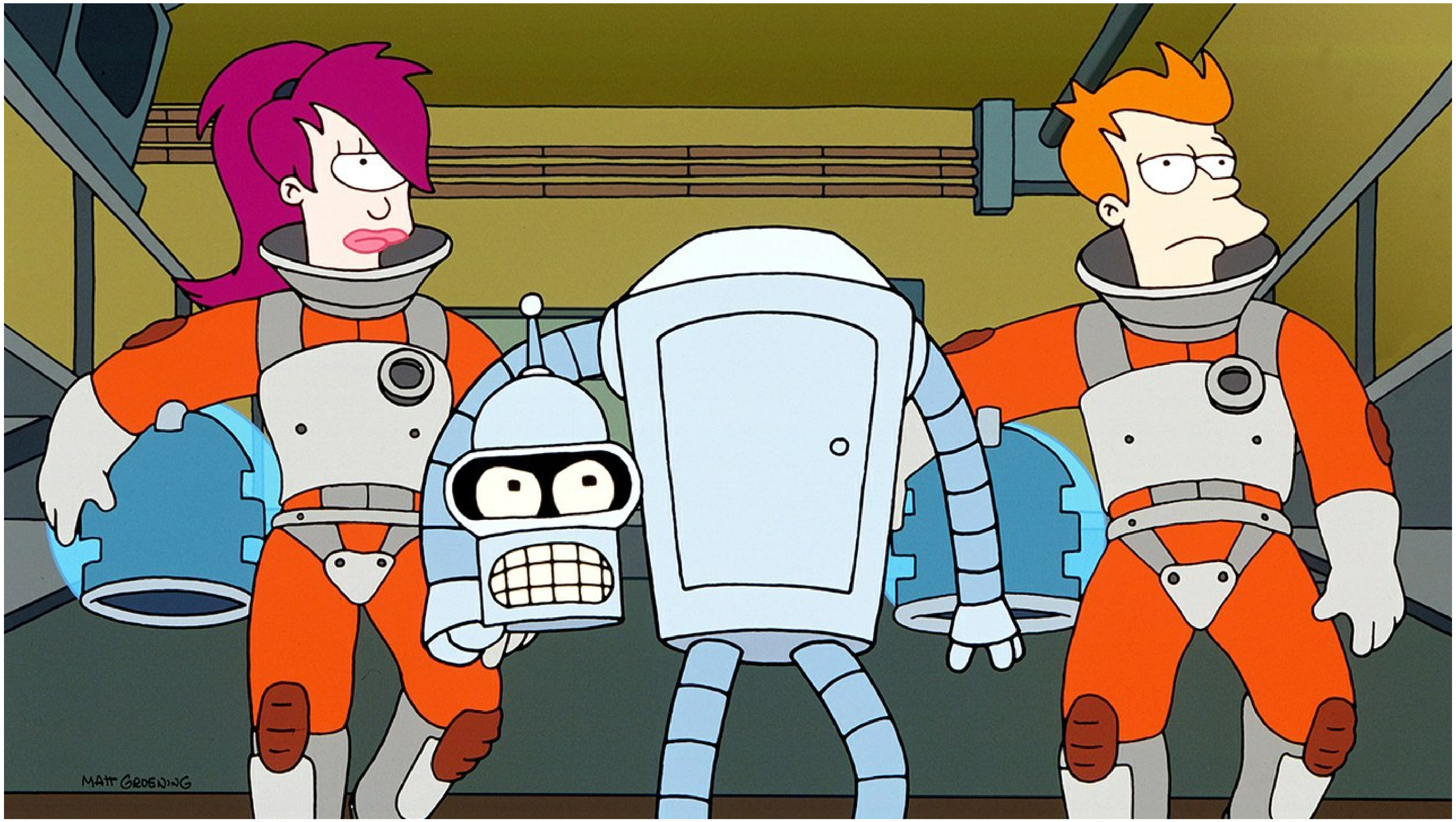
Math Camp: Lesson 4

Statistics and Probability

UW–Madison Political Science

August 22 & 23, 2019

Hang in there



Why do we mess with statistics?

Why statistics?

There is uncertainty in real-world data

- We are interested in x, but lots of forces affect x
- Maybe we are interested in understanding the sizes of these forces
- How do you do that?

Your data are affected by forces that you can't see

- Models that describe our data have unknown variables ("parameters")
- How do we estimate those parameters?

Let's flip a coin

If we flip a fair coin, what is the probability that it lands heads up?

Let's say we got Heads

What's going on here?

Observed data are influenced by underlying processes

- The coin flip (data) is influenced by an underlying probability of Heads
- There is a systematic component and a random component
- Statistical modeling is (in part) distinguishing systematic forces from random forces

Statistics

The mathematical study of data

- Data come from some underlying, unknown process
- Descriptive statistics: describe the data (mean, standard deviation, correlations)
- Inferential statistics: describe the underlying process (as best we can)

We do plenty of both in political science, but the big focus is on inference

- We theorize about how politics works
- We collect data
- We make inferences about the processes that influence the data
- Are those inferences consistent with our theories?

Statistics and probability

To make inferences about data generating process, we use probability

- We have some data: e.g. election outcomes, intensity of social movements, etc.
- But we don't know which process they come from (what are the factors leading to these outcomes?)
- We entertain a few possible different models for the process: model A, model B...
- Data may be more probable under one model or another
- We can calculate the probability of the data under each model to pick the best model
- And then evaluate how certain (or rather, uncertain) we are about our findings

But before we can do any of that

We have to learn some basic math of probability

Agenda

- Counting
- Set theory
- Probability
- Independence, joint probability
- Bayes' Theorem
- Looking ahead

Helpful vocabulary

A random variable is a realization of a process that is at least partially random (i.e. unpredictable)

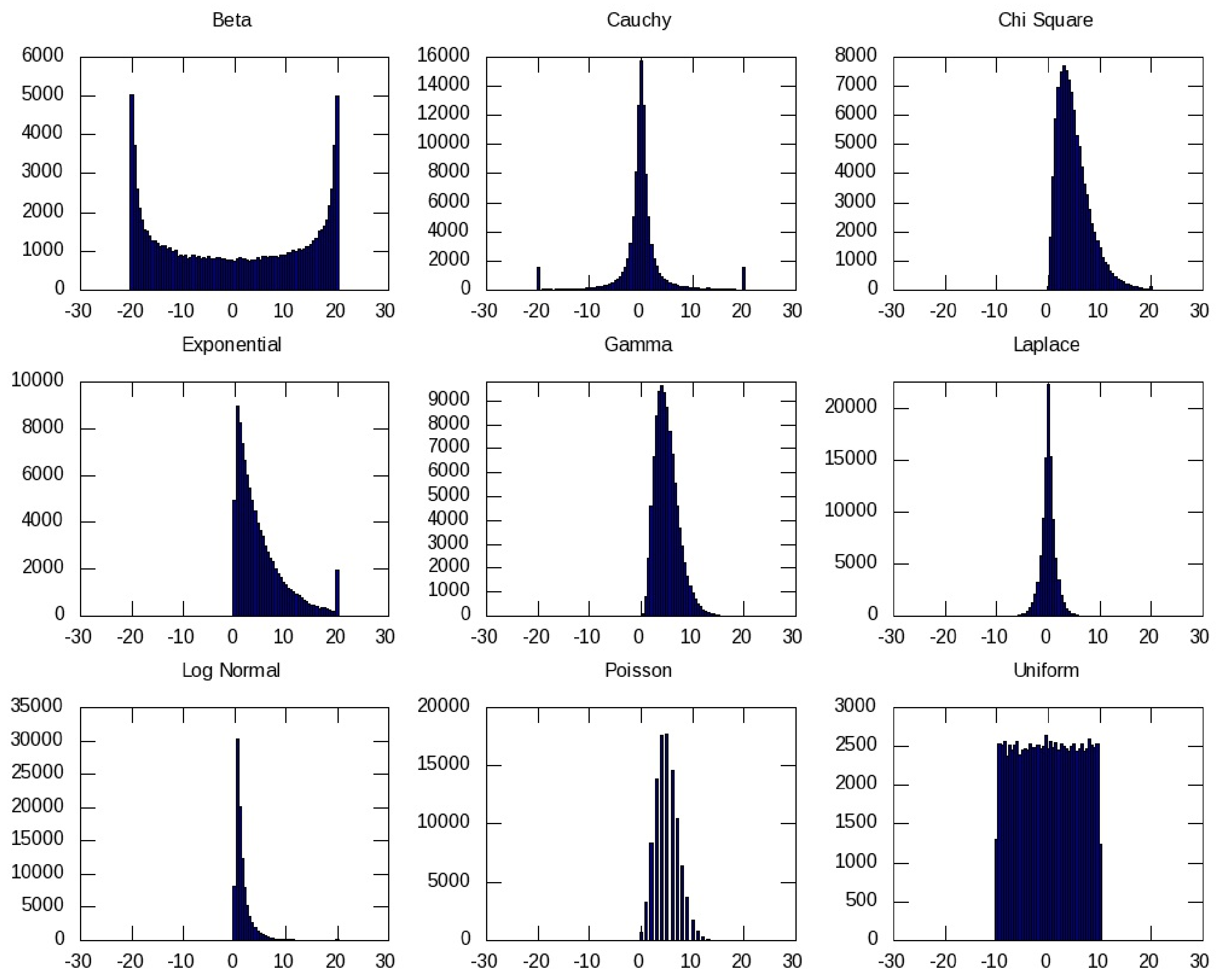
- e.g. coin flip, dice roll
- probability enters statistics through the assumptions we make about the nature of randomness in a random variable

A random variable could have many different potential outcomes (e.g. heads vs. tails). The probability of these outcomes could be unequal (e.g. war vs. no war).

If we wanted to describe the probability of each potential outcome, we would do so with a probability distribution.

- A probability distribution is a function that maps potential outcomes to the probability of those outcomes
- x = potential outcome
- $f(x) =$ probability of x
- These also matter for formal (non-statistical) models (e.g. utility shocks)

Some (empirical) distributions...



https://sysplay.in/blog/tag/probability-density-function/

More about probability distributions

Probability distributions can describe discrete outcomes (coin flips) or continuous outcomes (height, vote margin)

Probability distributions always sum to 1

- "The law of total probability"
- Discrete distributions: sum the probabilities of all potential outcomes
- Continuous distributions: integrate over the continuous space of outcomes

Probability distributions are the basis for statistical inference

- z-scores, p-values
- Prior and posterior beliefs

Counting

Counting

First, what is probability?

- The ratio of an event's expected frequency to the number of possible events
- We need to count the events in question & total possible outcomes

Suppose an event is described by K different component parts. (E.g. we roll a die K many times.) Each component $k = \{1, 2, \ldots, K\}$ has n_k possible values. What is the number of distinct outcomes we could get?

$$\prod_{k=1}^{K} n_k$$

(multiply the n_k's)

I roll a 6-sided die 4 times. How many unique sets of 4 rolls can I obtain (assuming that different orderings of the same 4 numbers are different events)?

Complex counting considerations

Does the order of selection matter? (Is $\{1, 2\} = \{2, 1\}$?)

Are selected objects replaced (able to be selected again) or not replaced?

Ordering with replacement

This is easiest because (a) no need to adjust for "double-counting" and (b) the number of possibilities is always constant.

The number of possible ways to select k elements from a larger pool of n is

$$n \times n \times n \times \ldots \times n = n^k$$

Intuition: in each draw, there are n possibilities. Each of n outcomes in one draw can be combined with the n outcomes in any (and all) other draws.

Example: rolling the dice

Order, no replacement

Also called permutation.

The number of ways to select k objects from a pool of n possible objects, where order matters but replacement does not occur.

Intuition: each draw removes the object from the larger pool. Subsequent draws have one less element to choose from.

$$n * (n-1) * (n-2) * \ldots * (n-k-1) = \frac{n!}{(n-k)!}$$

For example: number of possible ways to deal a card game, winning lottery numbers

Unordered, no replacement

Also called combinations: The number of possible ways to select k objects from a pool of n possible objects, where order does not matter and replacement does not occur

Intuition: we have fewer possibilities than before, substantively identical elements (A and then B; B and then A) are not double counted

$$\frac{n!}{k!(n-k)!} = \binom{n}{k}$$

For example: survey samples, raffles, possible groups of 2 in a classroom

Unordered, with replacement

The number of possible ways to select k elements from a larger pool of n possible elements, where order does not matter and replacement does occur

$$\frac{(n+k-1)!}{(n-1)!k!} = \binom{n+k-1}{k}$$

Example: the number of heads if you flip a coin n times

Exercises

Imagine we rank the 3 top swimmers in this room.

- Is this a situation where order matters?
- Is this a situation with replacement, or no replacement?
- How many different possible rankings could there be?

Imagine we have 4 different scholarships for students in this room. You can win more than one scholarship. How many different combinations of winners can there be?

Imagine we have 5 identical candies for students in this room. You can win more than 1 candy. How many different combinations of winners can there be?

Imagine we have 2 identical bicycles for students in this room. You can only win 1 bicycle. How many combinations of winners?

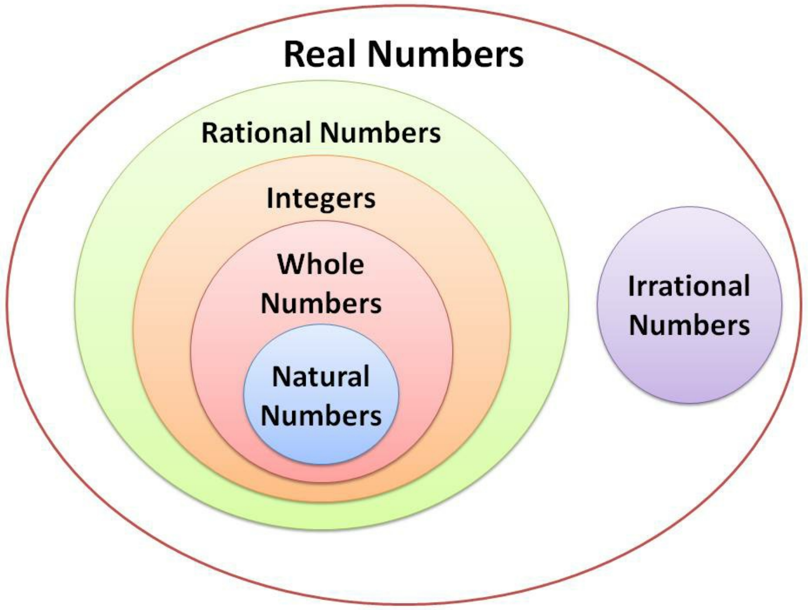
Set theory

Sets

Remember: a set is a collection of elements. Could be numbers, units, areas in space...

- $F = \{1, 2, 3, 4\}$
- $G = \{1, 3, 5\}$
- $H = [0, 1] \cup (2, 3)$

What are unions? Intersections? Disjoints? Subsets? Supersets?



- $P = \{\text{Reagan}, \text{Bush41}, \text{Clinton}, \text{Bush43}, \text{Obama}, \text{Trump}\}$
- $D = \{\text{Carter}, \text{Mondale}, \text{Dukakis}, \text{Clinton}, \text{Gore}, \text{Kerry}, \text{Obama}, \text{HRC}\}$
- $R = \{\text{Reagan}, \text{Bush41}, \text{Dole}, \text{Bush43}, \text{McCain}, \text{Romney}, \text{Trump}\}$
- $I = \{\text{Perot}, \text{Nader}\}$

The sample space

The sample space (denoted S or Ω) is the set that contains all elements in question.

Sometimes called the universal set

Not the same as the set that contains everything. Only the relevant things for what we're currently talking about.

Complementary sets

The complement of set A (denoted as A^C) is the set of all elements in the sample space that are not contained in A

$$A^C \equiv X \text{ such that } X \notin A$$

Example, $\Omega = [0, 1]$

- If $X = (0, 0.5]$, what is X^C?
- $X^C = \{0\} \cup (0.5, 1]$

What is Ω^C?

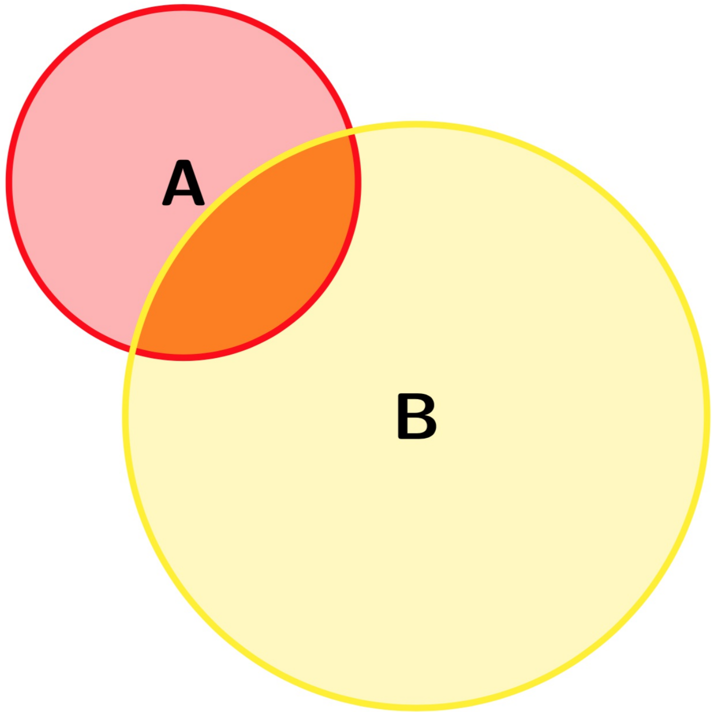
- $\emptyset$

Making sense?

Probability (beginning with sets)

Probability as Sets

We can use sets to represent the probability of events. Total area represents total probability of all events (equal to 1).



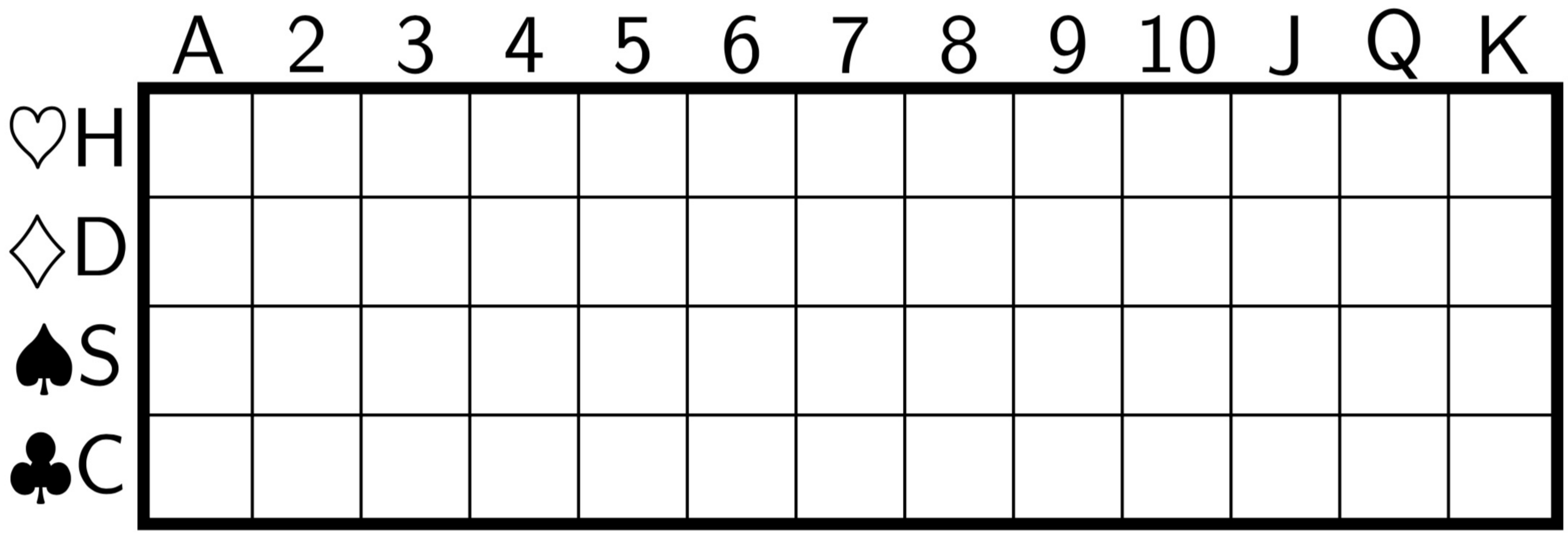
A is an event, and its area is a subset of the total area.

$\Pr(A)$?

$\Pr(A^C)$?

Let's play cards

We have 4 suits (hearts, diamonds, spades, clubs) and 13 card values (Ace, 2, 3, ..., Jack, Queen, King). Suits and values can both be sets.



Total area = 1

Probability of an individual card: $\frac{1}{52}$

Properties of probabilities

Probabilities are strictly bounded on the closed interval $[0, 1]$

- $p(A) \in [0, 1]$
- A is either impossible ($p = 0$), certain ($p = 1$), or in between (possible, $p \in (0, 1)$)

If we had N many exhaustive and mutually exclusive sets of potential outcomes, their probabilities sum to 1. Which is to say, something must happen.

$$\sum_{n=1}^{N} p(A_n) = 1$$

Probability of complements

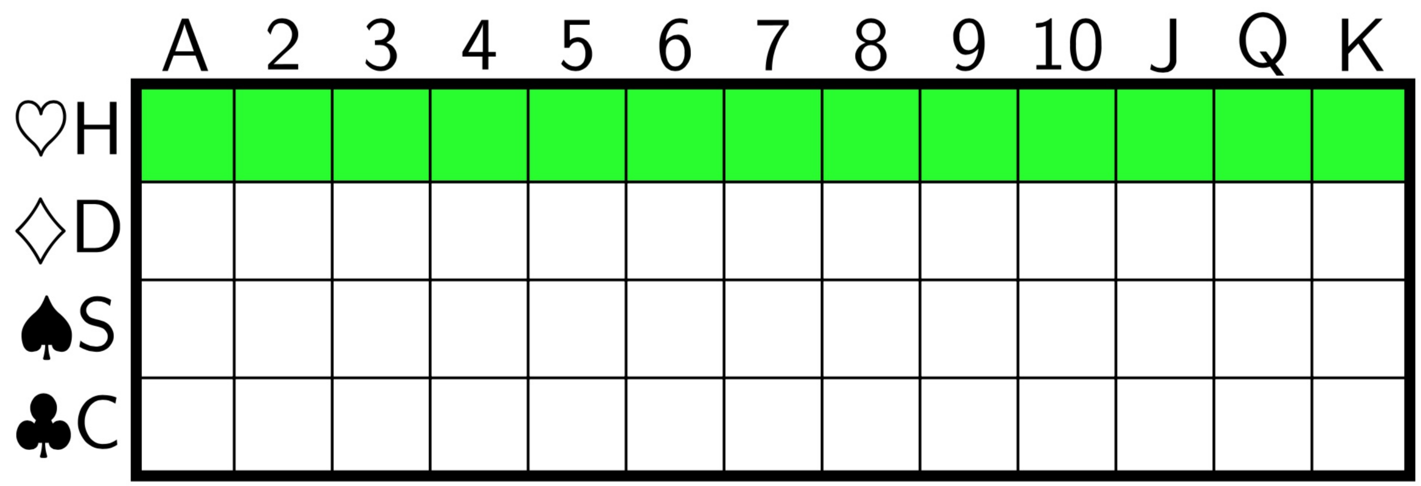
If Ω contains the set of all potential outcomes, and A is an event that is a subset of the outcome space that occurs with $p(A)$

- What is $p(A^C)$?
- $1 - p(A)$

The intuition: Something must happen, either A or not A

Example of complements

Probability that a random card is a Heart? $p(H) = \frac{1}{4}$

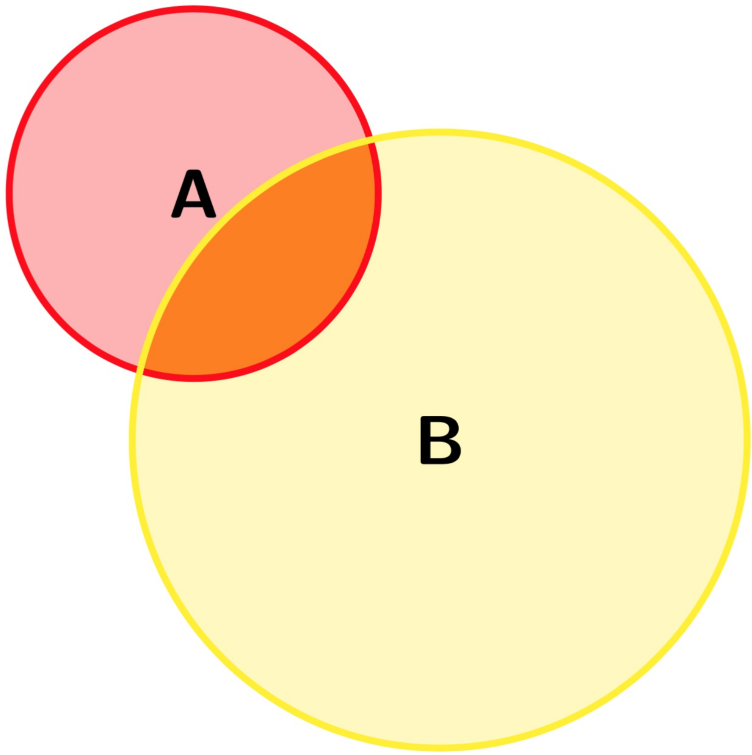


Probability that a card is not a heart? $1 - p(H) = \frac{3}{4}$

Probability of unions

The probability of $A \cup B$

The probability that either A or B occurs

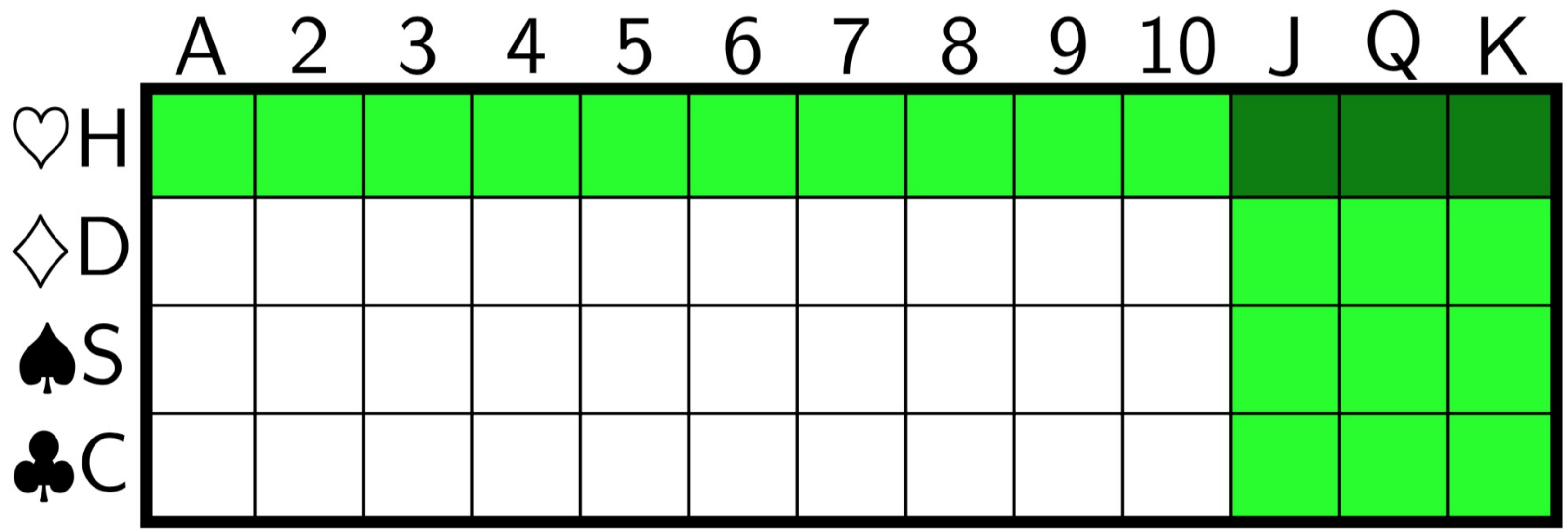


$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

The intuition: the sum of A and B will double count $A \cap B$, so we need to subtract one instance of $A \cap B$

Probability of Unions

What is the probability that we draw a card that is either a heart or a face card?



$p(H) = ?$

$p(F) = ?$

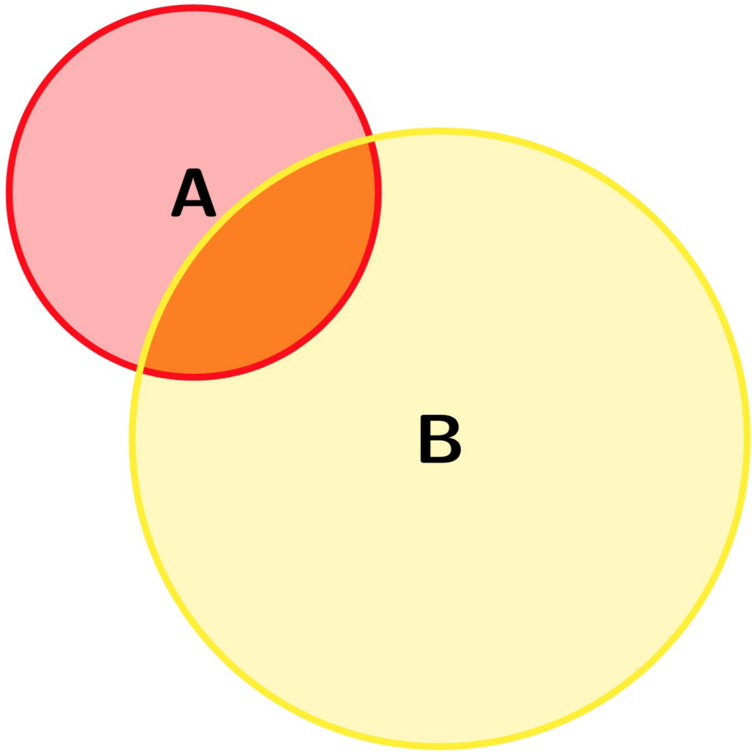
$p(H \cap F) = ?$

$p(H \cup F) = \frac{1}{4} + \frac{12}{52} - \frac{3}{52} = \frac{22}{52}$

Probability of intersections

The probability of $A \cap B$

The probability that both A and B occur

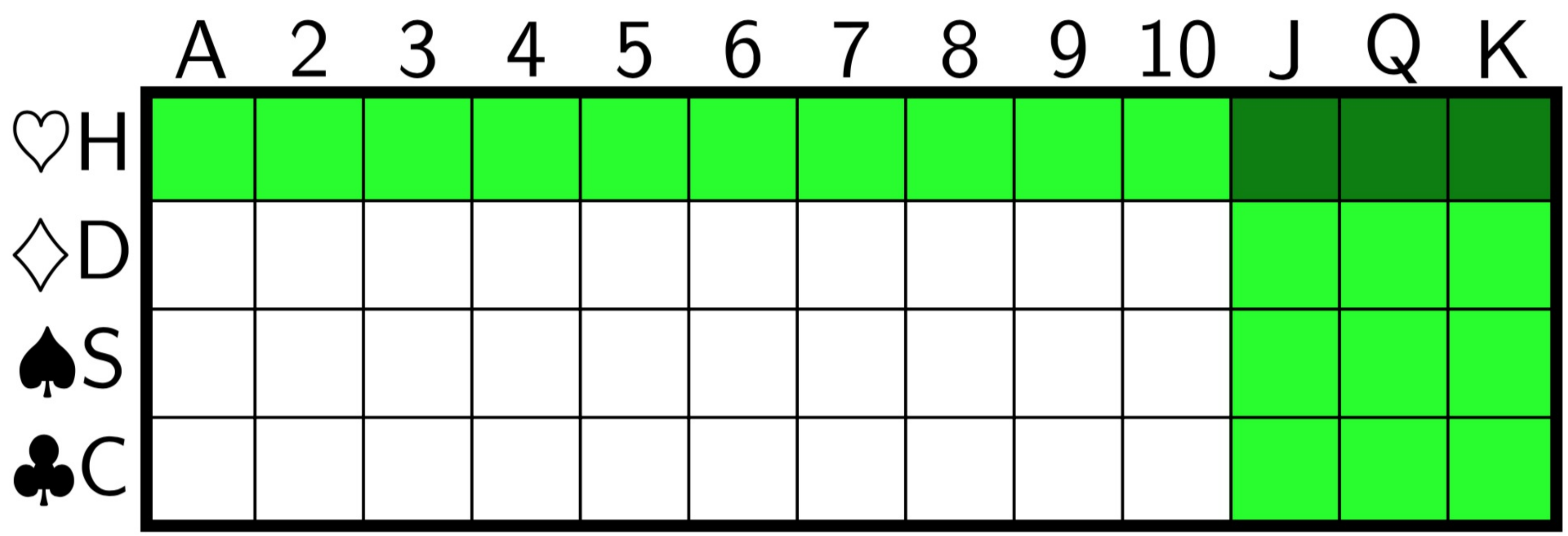


$$p(A \cap B) = p(A) + p(B) - p(A \cup B)$$

The intuition: We care only about the component that we double counted

Probability of intersections

What is the probability that we draw a card that is both a heart and a face card?



$p(H) = ?$

$p(F) = ?$

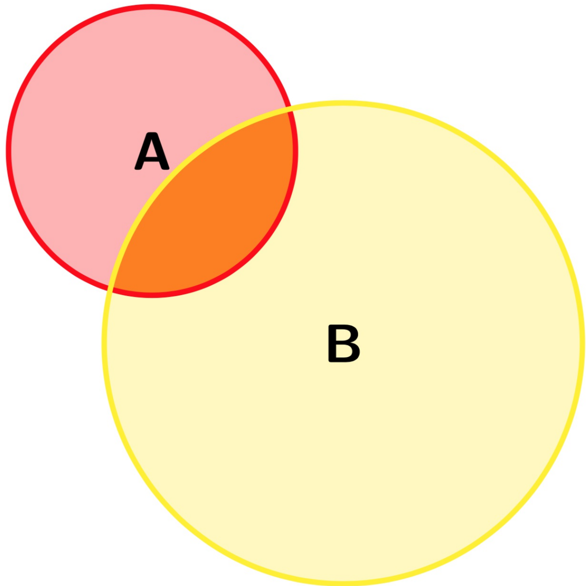
$p(H \cup F) = ?$

$p(H \cap F) = \frac{1}{4} + \frac{12}{52} - \frac{22}{52} = \frac{3}{52}$

Conditional probability

The probability of A, given B, is expressed as $p(A \mid B)$

What is the probability of A, given that B also occurs?

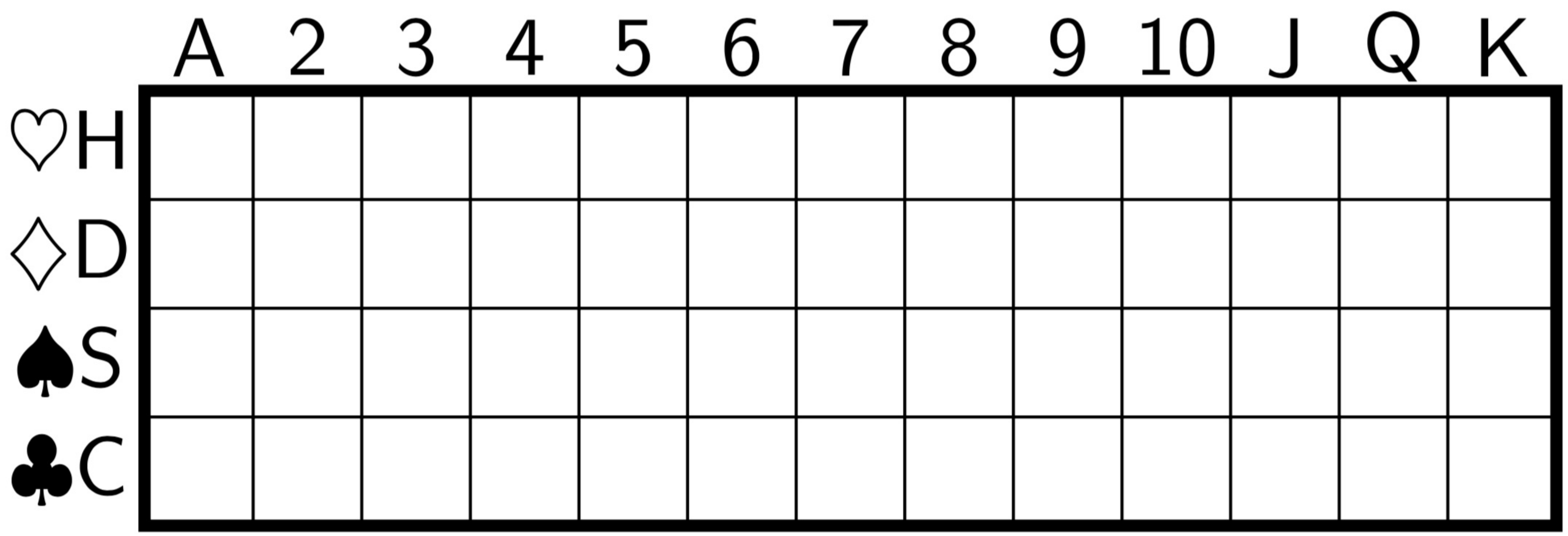


$$p(A \mid B) = \frac{p(A \cap B)}{p(B)}$$

The intuition:

- If we know that B happened, we only care about the space within B
- the probability that both A and B happen, divided by the probability of B
- $p(\text{intersection})\ /\ p(\text{conditioning event})$

Conditional probability



What is the probability of drawing the Ace of Diamonds?

What is the probability of drawing the Ace of Diamonds, given that we have have drawn an Ace?

- $p(\text{Ace of Diamonds}) = \frac{1}{52}$
- $p(\text{Ace}) = \frac{4}{52}$
- $p(\text{Ace of Diamonds} \mid \text{Ace}) = \frac{1/52}{4/52} = \frac{1}{4}$

What's the probability?

	A	2	3	4	5	6	7	8	9	10	J	Q	K
♡H													
♢D													
♠S													
♣C													

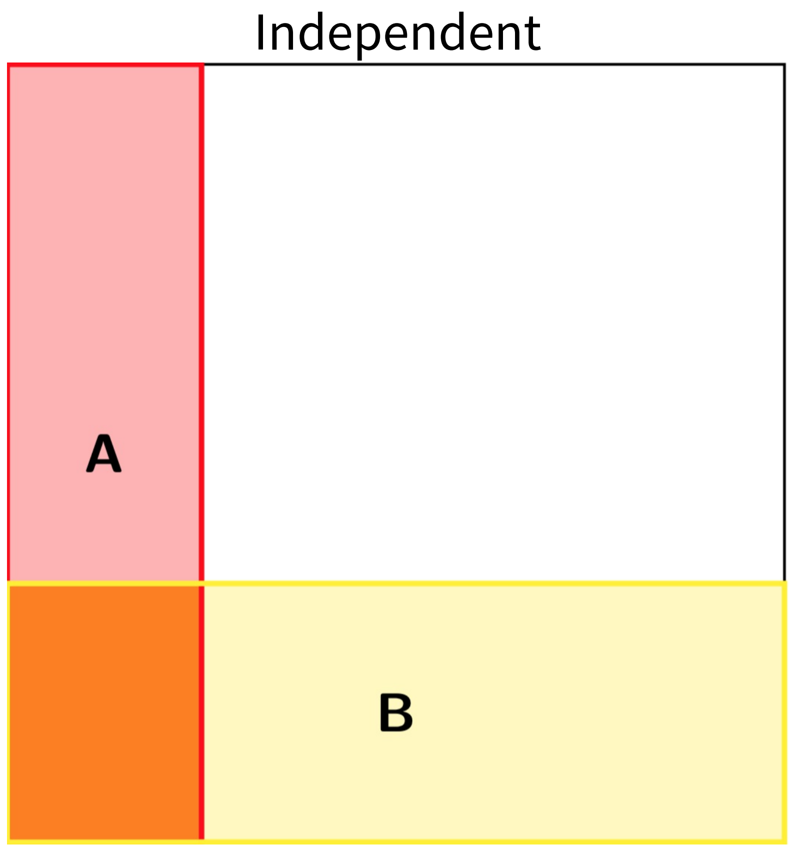
$p(\{8, 9, 10\})$

$p(\{5, 6\} \cup \{6, 10\})$

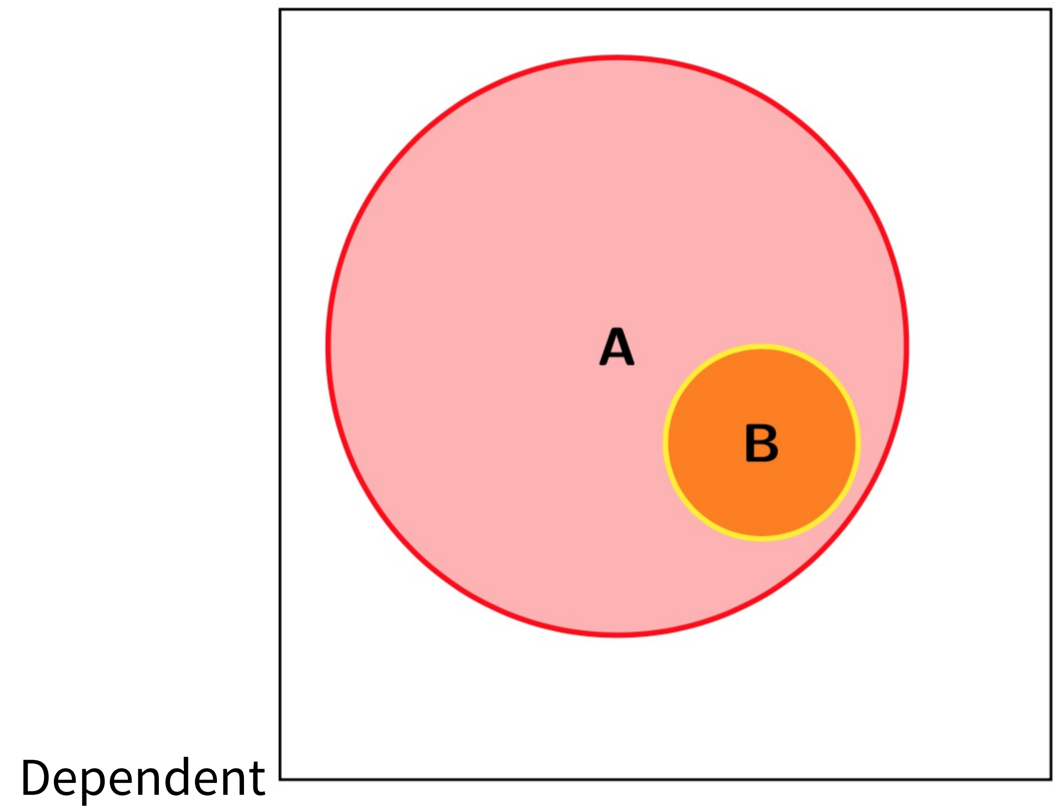
$p(A \mid H^C)$

The notion of independence

Two events are independent if knowing the outcome of one event does not change the probability of the other



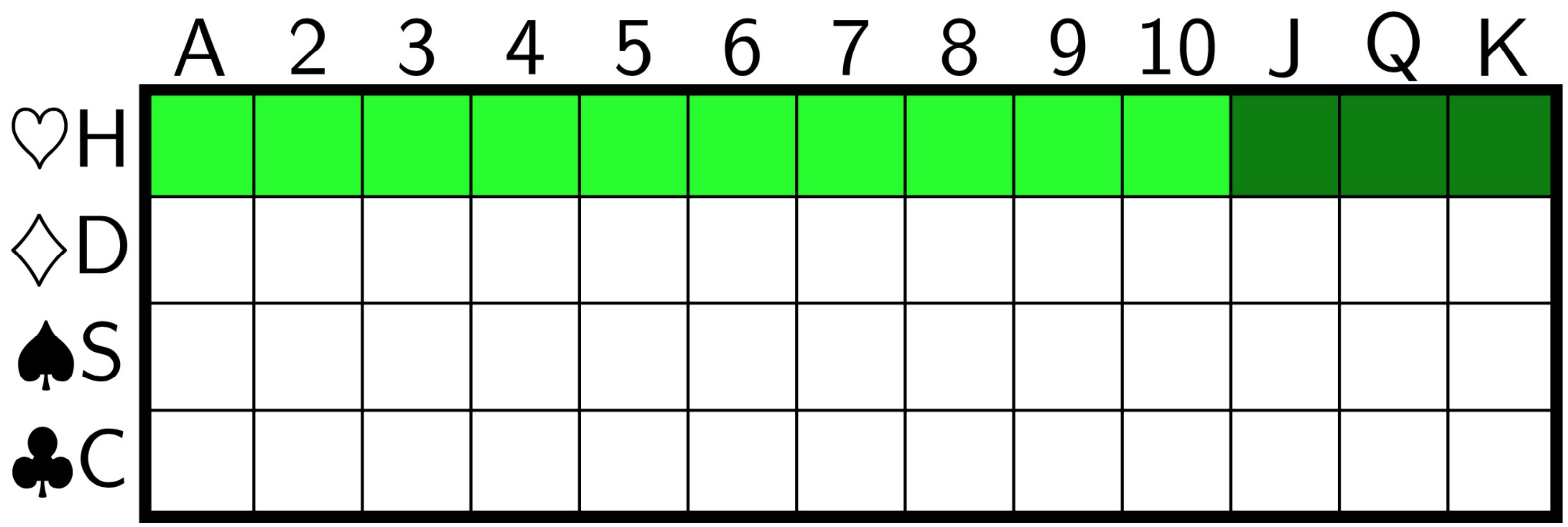
$$p(B) = p(B \mid A)$$



$$p(B) \neq p(B \mid A)$$

Independence of Events

Is drawing a face card independent of drawing a Hearts card?

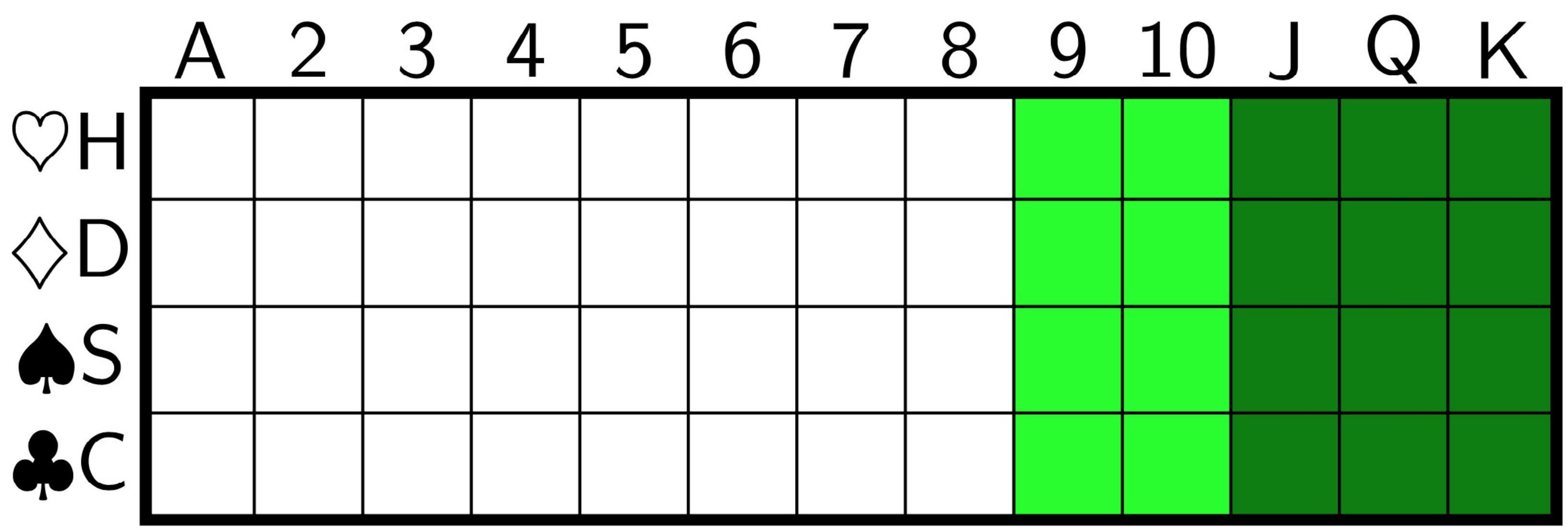


$p(F \mid H) = \frac{3}{13}$

$p(F) = \frac{12}{52} = \frac{3}{13}$

Independence of Events

What about drawing a face card independent of drawing a card greater than 8?



$p(X = F \mid X > 8) = \frac{12}{20} = \frac{3}{5}$

$p(F) = \frac{12}{52} = \frac{3}{13}$

Joint probability

What we're doing here is considering the probability of multiple events

Joint probability: the probability of more than one event occurring simultaneously

$$p(A, B) \equiv p(A) \cap p(B)$$

The exact equation for the joint probability depends on whether the events are independent

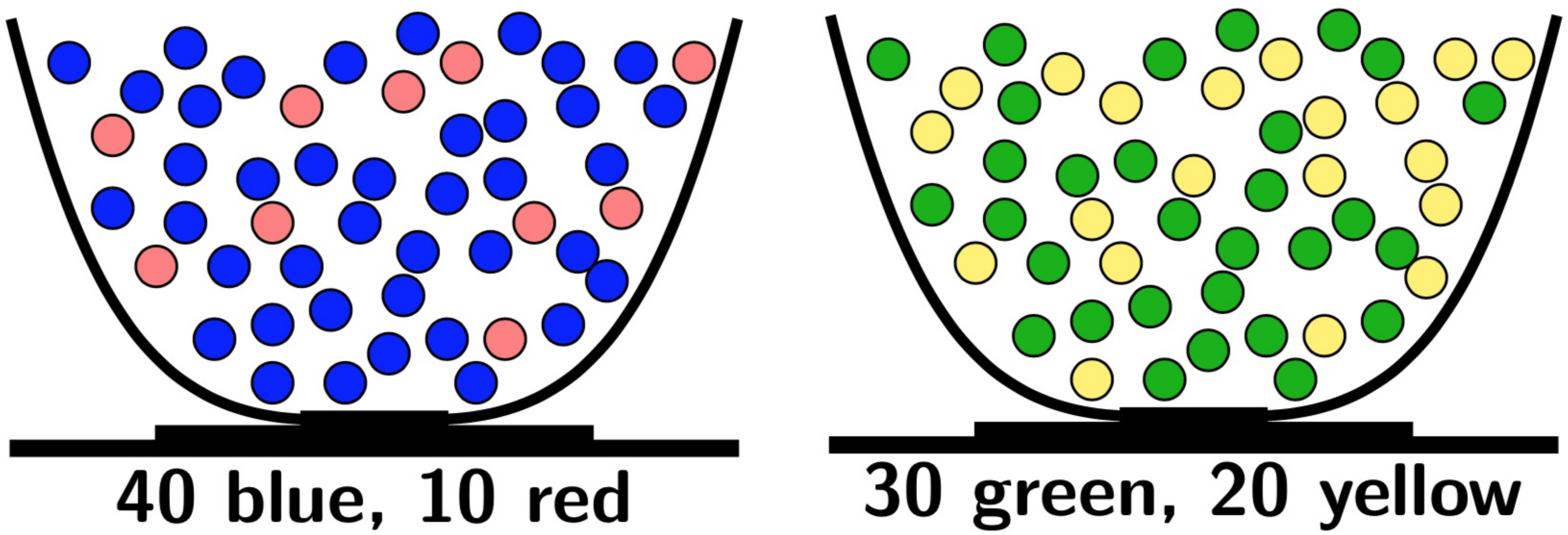
Joint probability of independent events

If multiple events are independent of one another, the joint probability of all events is the product of the individual probabilities.

Example: we flip three coins independently of one another. What's the probability of the sequence $\{H, H, H\}$?

$$\begin{aligned} p(H) \times p(H) \times p(H) &= .5 \times .5 \times .5 \\ &= 0.125 \end{aligned}$$

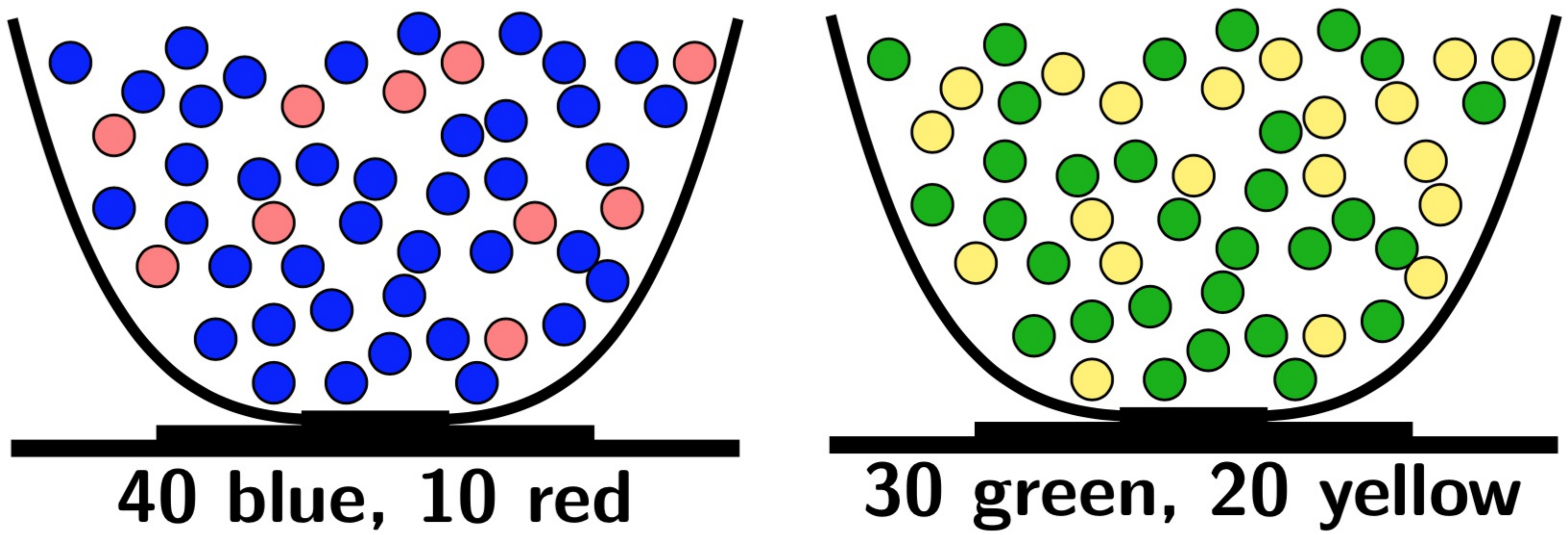
So we've got two bowls



If we draw a ball from each bowl, what is the joint probability of...

- $p(\text{blue}, \text{green}) = ?$
- $p(\text{blue}, \text{yellow}) = ?$
- $p(\text{red}, \text{green}) = ?$
- $p(\text{red}, \text{yellow}) = ?$

So we've got two bowls

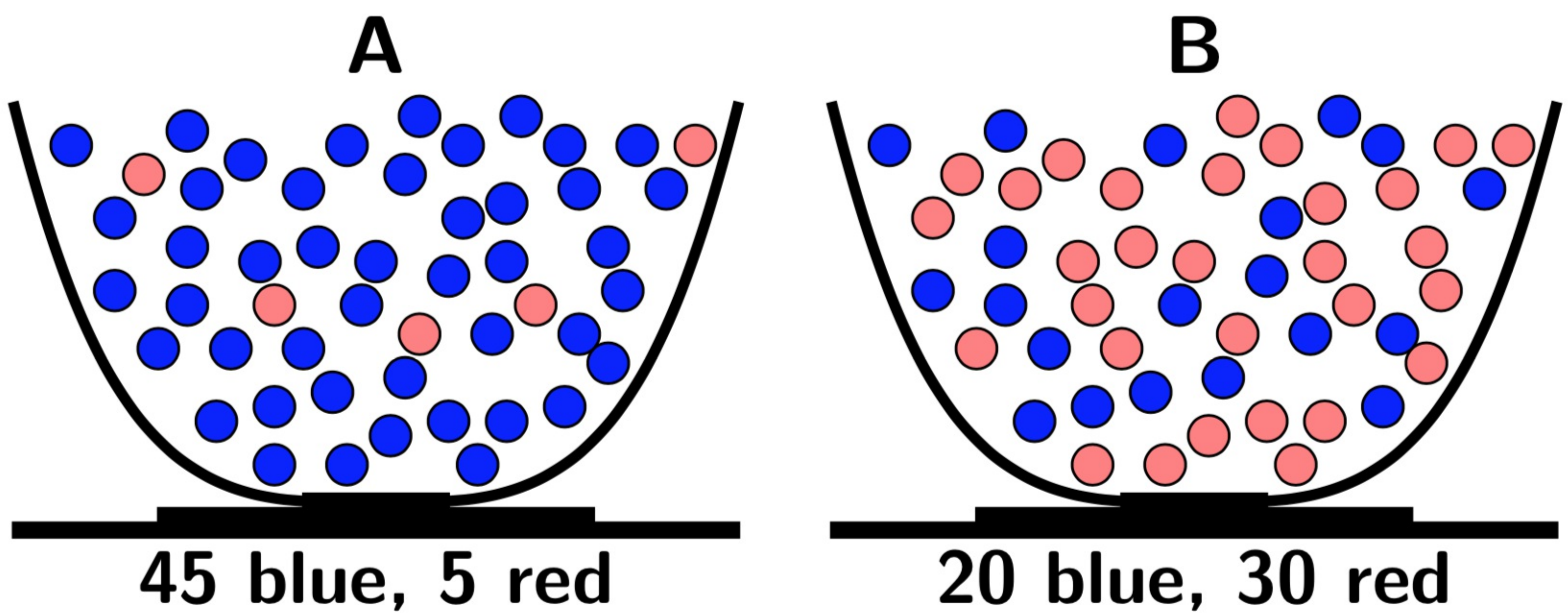


If we draw a ball from each bowl, what is the joint probability of...

- $p(\text{blue}, \text{green}) = \left(\frac{40}{50}\right)\left(\frac{30}{50}\right) = (.8)(.6) = .48$
- $p(\text{blue}, \text{yellow}) = \left(\frac{40}{50}\right)\left(\frac{20}{50}\right) = (.8)(.4) = .32$
- $p(\text{red}, \text{green}) = \left(\frac{10}{50}\right)\left(\frac{30}{50}\right) = (.2)(.6) = .12$
- $p(\text{red}, \text{yellow}) = \left(\frac{10}{50}\right)\left(\frac{20}{50}\right) = (.2)(.4) = .08$

Because these are mutually exclusive and exhaustive events, probabilities sum to 1

Imagine we flip a coin. If heads, we draw a ball from the left urn. If tails, we draw from the right.



This means there are two ways to choose a blue ball: $\{A, \text{blue}\}$ and $\{B, \text{blue}\}$

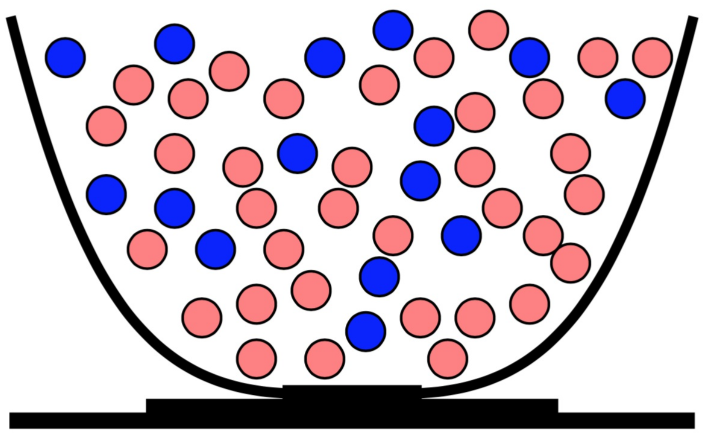
- $p(A, \text{blue}) = 0.5 * \frac{45}{50} = 0.45$
- $p(B, \text{blue}) = 0.5 * \frac{20}{50} = 0.20$

Total probability of blue is the sum of the joint probabilities (a very useful principle...)

$$\begin{aligned} p(\text{blue}) &= p(\text{blue} \mid A) + p(\text{blue} \mid B) \\ &= p(\text{blue} \mid A) + p(\text{blue} \mid A^C) \end{aligned}$$

Thinking about order and replacement

We draw 5 balls from one urn, replacing each time. We get the following sequence:



$\{\text{blue}, \text{red}, \text{blue}, \text{blue}, \text{red}\}$

The probability of this specific sequence is $.3 * .7 * .3 * .3 * .7 = 0.01323$,

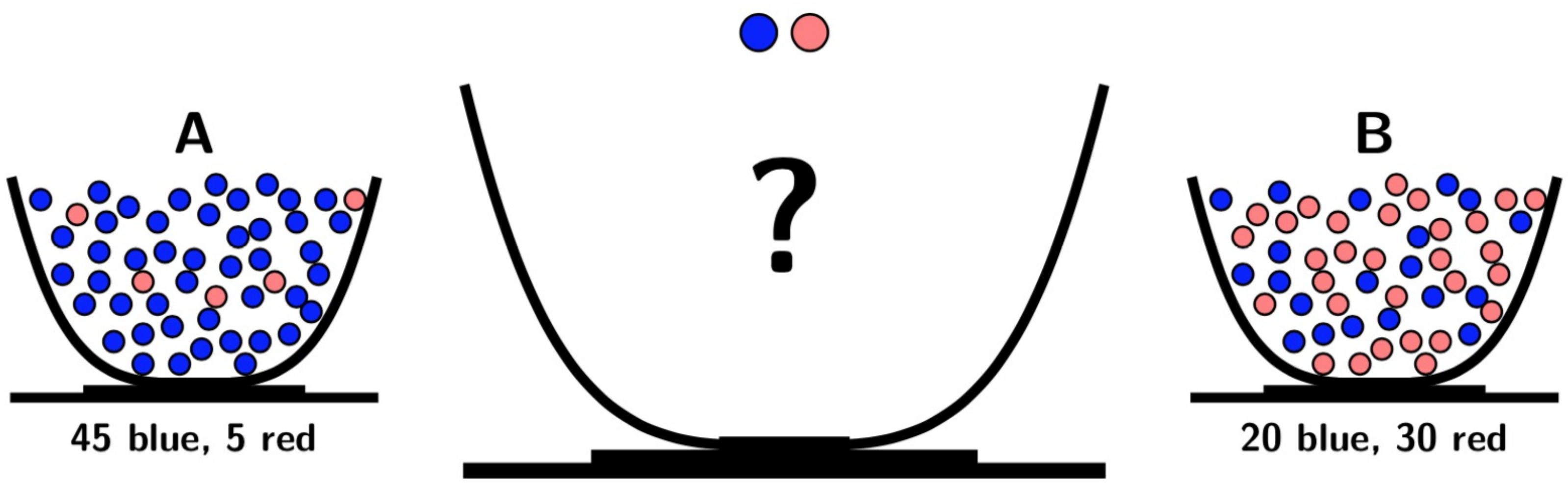
or if we simplify: $0.3^3 0.7^2$

Imagine we don't care about the order, just the probability of three blues (which implies two reds)

- The total probability of 3 blues: sum the p of every sequence that has 3 blues
- Any individual sequence that with 3 blues has probability 0.01323 (above)
- We just need the number of ways to get 3 blues with 5 draws

$$\left(\frac{5!}{3!(5-3)!}\right)(.3)^3(.7)^2 = \binom{5}{3}(.3)^3(.7)^2 = (10)(.01323) = .1323$$

Inverse conditional probability



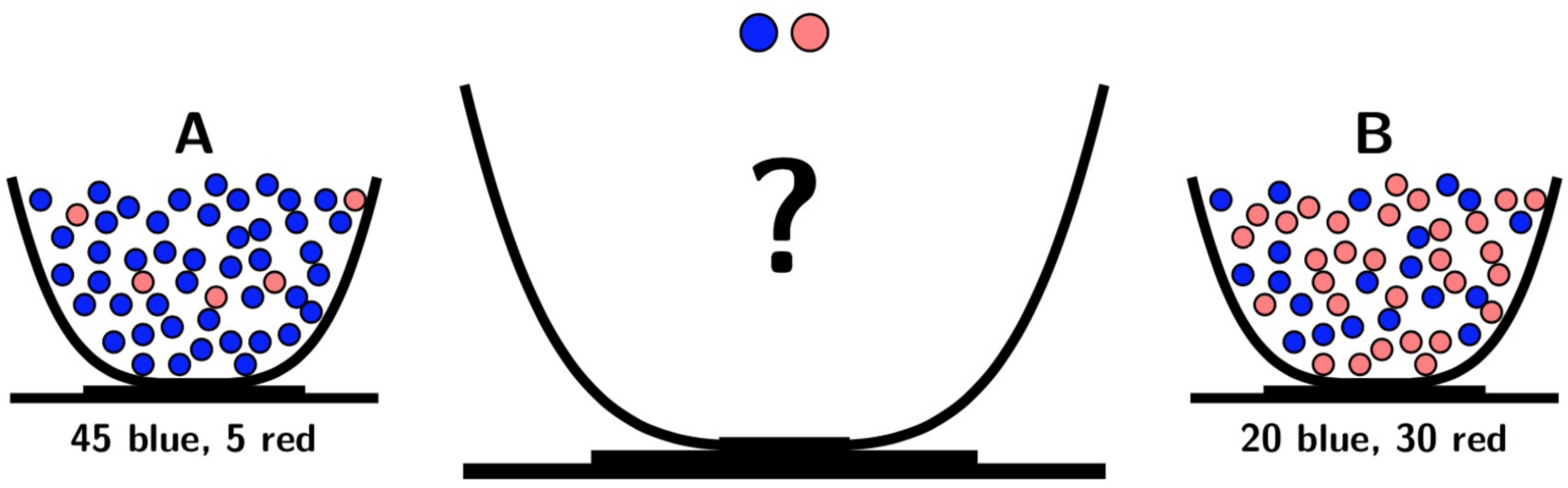
Someone flips a coin to decide whether to draw a ball from bowl A or B (each with 50% probability), but the bowl is hidden from us.

- What is the probability of drawing from bowl A?
- We've drawn a blue ball. What's the probability that we drew from A?

"Inverse" conditional probability problem:

- It's easy to find $p(\text{blue} \mid A)$,
- but how can we invert it to find $p(A \mid \text{blue})$?

Find $p(A \mid \text{blue})$



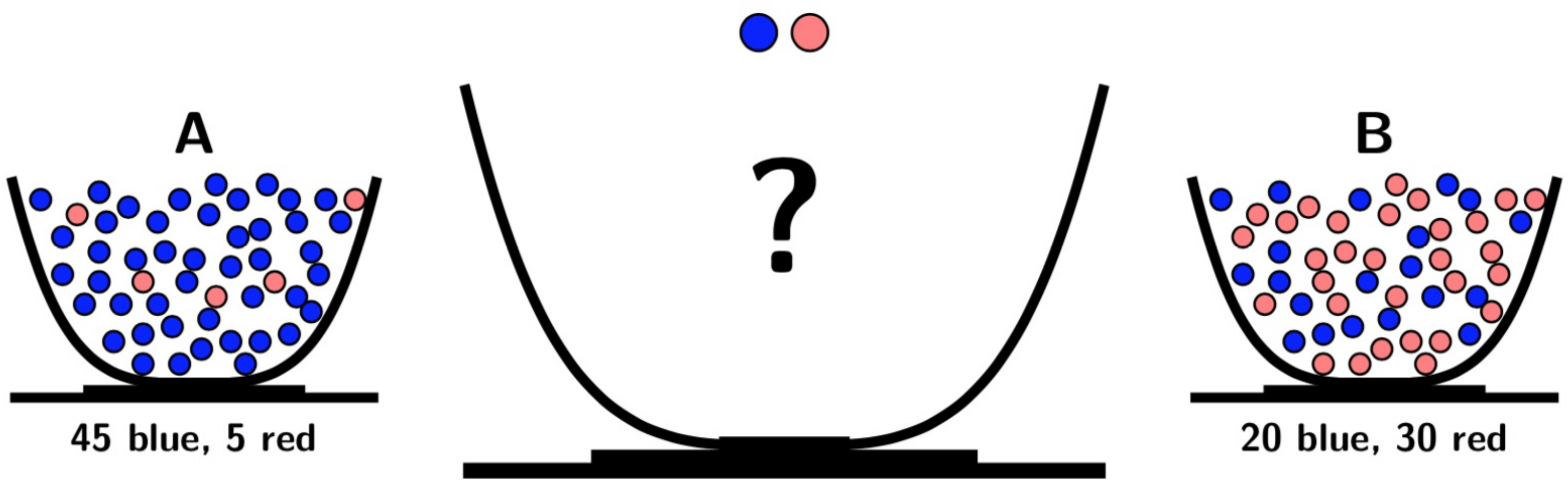
How do we approach any conditional probability problem?

$$p(y \mid x) = \frac{p(y \cap x)}{p(x)}$$

So what do we need for $p(A \mid \text{blue})$?

- $p(A \cap \text{blue})$
- $p(\text{blue})$

Find $p(A \mid \text{blue})$



$p(A \cap \text{blue})$?

- $(0.5)(0.9) = 0.45$
- This is (associatively) the same as $p(\text{blue} \mid A)p(A)$

$p(\text{blue})$?

- $p(A \cap \text{blue}) + p(B \cap \text{blue})$
- $(0.5)(0.9) + (0.5)(0.4) = 0.45 + 0.20 = 0.65$

Find $p(A \mid \text{blue})$

$$p(A \mid \text{blue}) = \frac{p(A \cap \text{blue})}{p(\text{blue})}$$

$$p(A \mid \text{blue}) = \frac{p(\text{blue} \mid A)p(A)}{p(\text{blue})}$$

$$p(A \mid \text{blue}) = \frac{0.45}{0.65} \qquad \approx 0.69$$

This is inverse conditional probability: how we find $p(A \mid \text{blue})$ by starting with $p(\text{blue} \mid A)$.

We just did Bayes' theorem

Bayes' Theorem

Generally it's true that $p(x \mid y) = \frac{p(x \cap y)}{p(y)} = \frac{p(x \cap y)}{p(y \cap x) + p(y \cap x^c)}$

Bayes' Theorem describes how to solve equation by beginning with its inverse

$$p(x \mid y) = \frac{p(y \mid x)p(x)}{p(y)}$$

Or, more generally

$$p(x \mid y) = \frac{p(y \mid x)p(x)}{p(y \mid x)p(x) + p(y \mid x^c)p(x^c)}$$

A common Bayes example

A rare disease occurs in .01% of the population. We have a test for it, but it isn't perfect. 98% of individuals with the condition will test positive (2% false negative). 97% of those without the condition test negative (3% false positive).

You get the test done. The test is positive.

What's the probability that you have the disease?

- Prior probability: .01% you have the disease
- What is the updated (posterior) probability that you have the disease, given that you test positive

Applying Bayes

$$\text{Posterior probability} = \frac{p(\text{data} \mid \text{prior}) \times \text{prior}}{p(\text{data})}$$

$$p(\text{disease}|+) = \frac{p(+ \cap \text{ disease})}{p(+)}$$

$$p(\text{disease}|+) = \frac{p(+ \cap \text{ disease})}{p(+ \cap \text{ disease}) + p(+ \cap \text{ disease}^c)}$$

$$p(\text{disease}|+) = \frac{p(+ \mid \text{disease})p(\text{disease})}{p(+ \mid \text{disease})p(\text{disease}) + p(+ \mid \text{disease}^c)p(\text{disease}^c)}$$

$$.003 \approx \frac{(.98)(.0001)}{(.98)(.0001) + (.03)(.9999)}$$

If our prior is .01% chance of disease, a positive test revises the probability to .3%.

This is called Bayesian updating

Why Bayesian statistics is hard

Take a look at the denominator of Bayes' theorem

$$\text{Posterior probability} = \frac{p(\text{data} \mid \text{prior}) \times \text{prior}}{p(\text{data} \mid \text{prior})\text{prior} + p(\text{data} \mid \text{prior}^c)\text{prior}^c}$$

Imagine we have a continuous prior. E.g. we believe that the probability of a coin flip giving us heads (or tails) is close to 0.5 but we are a little uncertain (due to the weight of the coin sides).

In situations like this, our prior takes the form of a continuous probability distribution where each potential value has an associated probability.

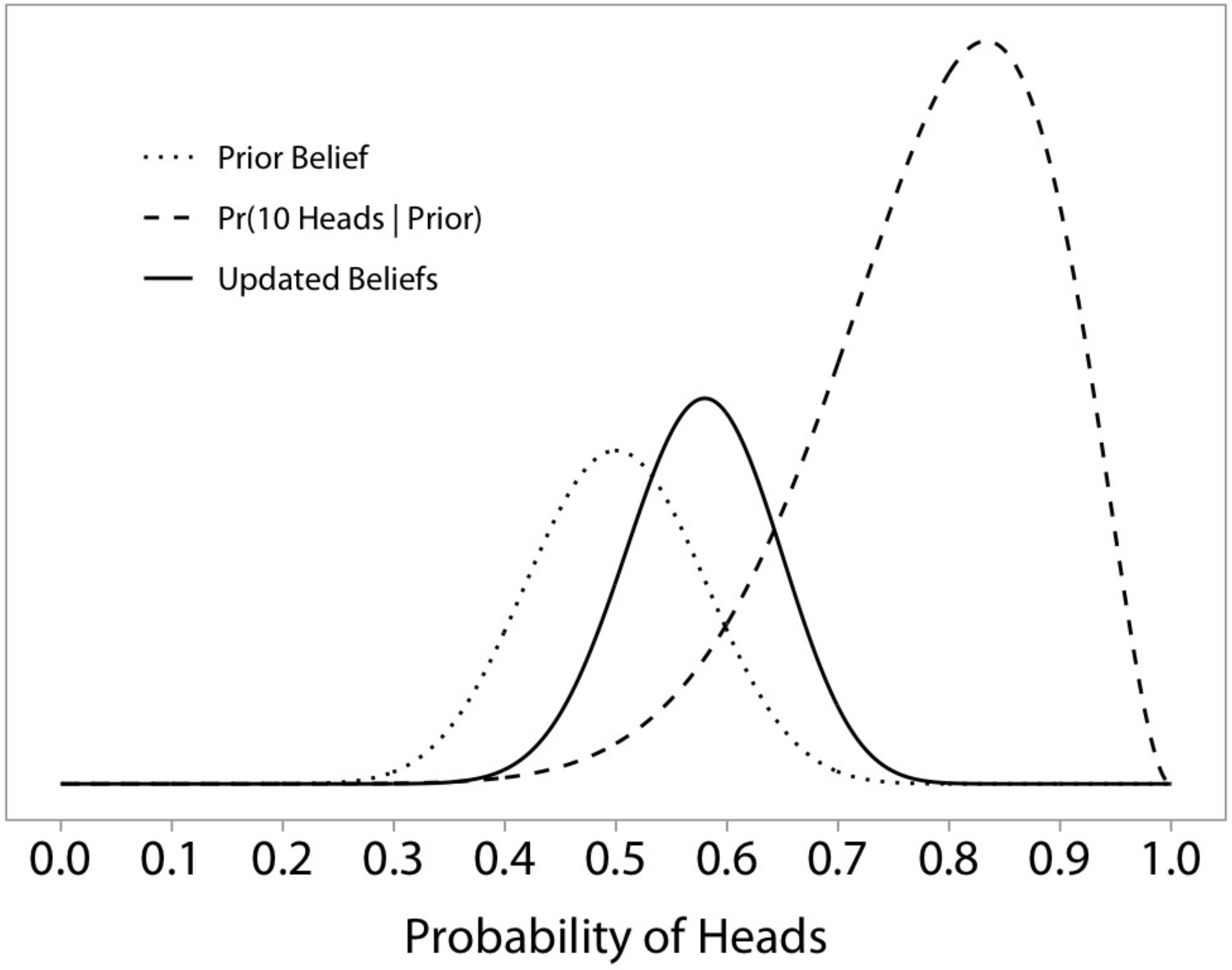
If we have a continuous parameter, summing anything across all values (e.g. prior^c) means integrating. And integrating is hard.

As a result, if you ever want to do Bayesian analysis for your own research, it tends to be computationally expensive (slow) and somewhat approximate.

Moreover, our results are distributions, not just single estimates.

A continuous example

We think that the probability of a "heads" on a coin is most likely 0.5, but we aren't certain about that. We flip the coin 12 times and find 10 heads. What is our revised belief?



Wait, how can "beliefs" affect probability?

Depends what we're talking about. There are two big ways to think about probability:

1. Over a large number of repeated, controlled trials, probability is the fraction of trials in which an event occurs.

2. Probability is never something that we know, only something that we learn about. Given our most reasonable information and evidence about the setting, how likely is an event?

More generally, we are interested in estimating unknown parameters in a mathematical model of a social interaction.

1. The parameter is unknown but fixed. There is an honest-to-god true value, and we are estimating it using data.

2. The parameter is unknown, and our information about it will always be imperfect. The information we obtain (conditioning on our model, on our data) can only approximate a distribution of possible values that are more or less plausible.

The two statistical genders

"Frequentism"

- Statistical properties come from repeated sampling assumptions
- There exists a true parameter, which we estimate
- We can calculate probability that our data were created by different assumed parameter values
- Low probability of data can be used to reject parameter values
- Focus is on the probability of the data, assuming a fixed parameter

"Bayesianism"

- Statistical properties come from posterior distribution
- Parameters are "random" (not fixed), only approximated with a distribution
- We have prior notions about plausible parameter values
- We can estimate the likelihood of data at different prior values
- Data updates our prior to form posterior beliefs
- Focus is on the probability of the parameter, updating a prior with data

Looking ahead

Methods courses

If you want to understand statistical work in political science, you should do:

- 812, 813, MLE
- Empirical methods (817)

Formal theory courses:

- 835 (game theory)
- Formal models of domestic (836) and international (837) politics

Advanced methods courses include

- Multilevel modeling, Time series, Panel data, Bayesian analysis, Experimental methods, Event history

Courses outside the department:

- Ag econ: applied regression, choice models
- Sociology: causal inference, survey methods
- Statistics: networks, machine learning

Methods pathways

Take the foundations courses no matter what

First field: "I want to study how to study politics". You still need a substantive interest

Second field: "I want to teach and research about/use new methods," not just, "I can do statistics okay"

Minor: 3 courses (see reqs)

Advice for methods courses

Take as many as you feasibly can.

Don't delay MLE.

Even if you a qualitative researcher, the epistemological lessons of large-N analysis are valuable.

If you're going to read empirical social science, you should take empirical social science courses.

Pick something you like and get good at it

- Time series, Bayes, text as data, causal inference, experiments

Do replication projects

Advice for methods in the discipline

Learn an unfamiliar method from a different field/subfield and apply it to your interests

Take the open science and the "replication crisis" seriously

Take math seriously (it helps you ride the learning curve)

Be a plain text social scientist (take your computer seriously)

Learn $\LaTeX$, learn R, learn `git`. Stata works but we may be way past the inflection point

If you might leave academia for data science, consider Python and machine learning