Statistical Rethinking

Thoughts, Notes, Takeaways

Video 1. The Golem of Prague [link]

Feb 4, 2022

Motivation in science comes from fascination with the natural world. Data is our way to make our interest in these things quantitative. Statistics are our way to make sense of those data. P-values are fine, but they give no indication of causation.

GOLEMS:

more powerful than its creators, but has no intent of its own. Golems have no wisdom or foresight, so without clear instructions, can destroy their creator.

Process models vs statistical models? Can we define the two?

Falsifying the null hypothesis is not actually that useful, because we basically never think the null hypothesis isn't true.

If we start with process models (theoretical model explain the causal relationship between variables), and translate into statistical models, two different process models can lead to the same statistical model, and are therefore not very accurate.

Sometimes, models are not possible or unique. For example, there is no Null ecological community or social network.

OWLS:

Representative of the technological guides to building good statistical models; step 1: draw some circles, step 2: draw the owl. In this course, we'll learn all the steps in between, identifying assumptions, connections between variables, and hidden steps that are not usually considered.

Outline of the scientific workflow;

1. **Theoretical Estimand:** what are you trying to do in the first place? The first step to good statistical modelling is defining the goal that your models will produce.
2. **Scientific Causal Model:** building models that generate synthetic observations. Forward simulating, logical models that produce data and let us design statistical procedures.
3. **Use 1&2 to build statistical models:** build statistical procedures to "get at" the estimand, or tell us that it's not possible, and we need to find a new way to get at the estimand.
4. **Simulate from 2 to validate that 3 yields 1:** check back against causal models and real data to ensure that statistical models are valid. This will enhance our trust in our models, as well as our colleagues'.

5. **Analyze Real Data:** finally, once we know we have good models, apply them to real data to test our models against nature.

Why Bayes?

When Galileo viewed Saturn, he saw a circle with ears. No matter how many times he looked, he saw the same thing. Is this a statistical problem? yes! A Bayesian one. With a Bayesian approach, we have a permissive and flexible approach to modelling. We can express uncertainty at all levels from sampling to inference bias. We also have direct solutions for measurement error and missing data, and finally, Bayes allows focus on scientific models without worrying what kinds of statistical estimators, sampling distributions, or etc that you are going to use. Only estimator is our posterior distribution.

Process error, - warblers respond to more things than temperature

Observation error - you don't see all the warblers

Measurement error - your scale is bad

Frequentists think if you keep doing this same thing (looking at Saturn), then you should be able to predict an outcome. Bayesians may have an expectation that the planet might not be correct, so maybe another error is

DAGS:

Throughout the 20th century, there has been conflict between Frequentists and Bayesians towards statistical inference, but that's not our fight. Both are capable frameworks, we just need better connection of causal models and statistical procedures.

For *Statistical Models* to produce scientific insight, they require additional *Scientific (Causal) Models.* The reasons for statistical analysis are not found in the data themselves, but rather the causes of the data. The Causes of the data can't be extracted from the data alone. **No causes in, no causes out.**

Causes are not optional

Even when the goal is descriptive, we need casual models. The sample differs from the population: describing the population requires causal thinking. You need causal information about how the sample differs from the population.

Causal Inference:

- The attempt to understand the scientific causal model using data that may have been produced.
- Correlation =/= Causation and Causation =/= Correlation
- Causal inference is a **prediction** of intervention
- Causal inference is **imputation** of missing observations.

Causal Prediction

- Knowing a cause means being able to predict the consequences of an intervention.
  - If I do this, then that.

- trees moving and wind blowing are always correlated, so it is not absolutely clear from data alone which causes the other. If we were to intervene, e.g. get people to climb a bunch of trees and shake them, wind wouldn't blow, which indicates that trees moving are not the cause of wind.

Causal Imputation

- Knowing a Cause means being able to **construct** unobserved **counterfactual outcomes**
  - By understanding causal relationships, we can simulate answers to situations that have not occurred.

DAGs: Directed Acyclic Graphs
Heuristic models, analyzable by eyeballs, that help you clarify your scientific thinking and analyze statistical approaches.

It's absolutely not safe to add every possible compound to a model. Bad controls can make things worse.

Recap:

Golems:

Brainless, powerful statistical models

Owls:

Documented, objective procedures

DAGs:

Transparent scientific assumptions to
- Justify scientific effort
- expose it to useful critique
- connect theories to golems

Video 2. Bayesian Inference [link]

Predicting the surface of Earth

Imagine a globe, being thrown around a classroom. Each time it is caught, mark down if the pointer finger is on land or on water. We'll use these occurrences to make a land/water dataset to eventually predict the surface makeup of earth.

If we do this virtually, we get:

L W L L W W W L W W

from this data:
- How should we use this sample to make an inference about the Earth?
- How do we produce a summary?
- How do we represent uncertainty?

Bayesian Inference:

For **each possible explanation** of the data, we are going to **count all the ways that data can happen.**

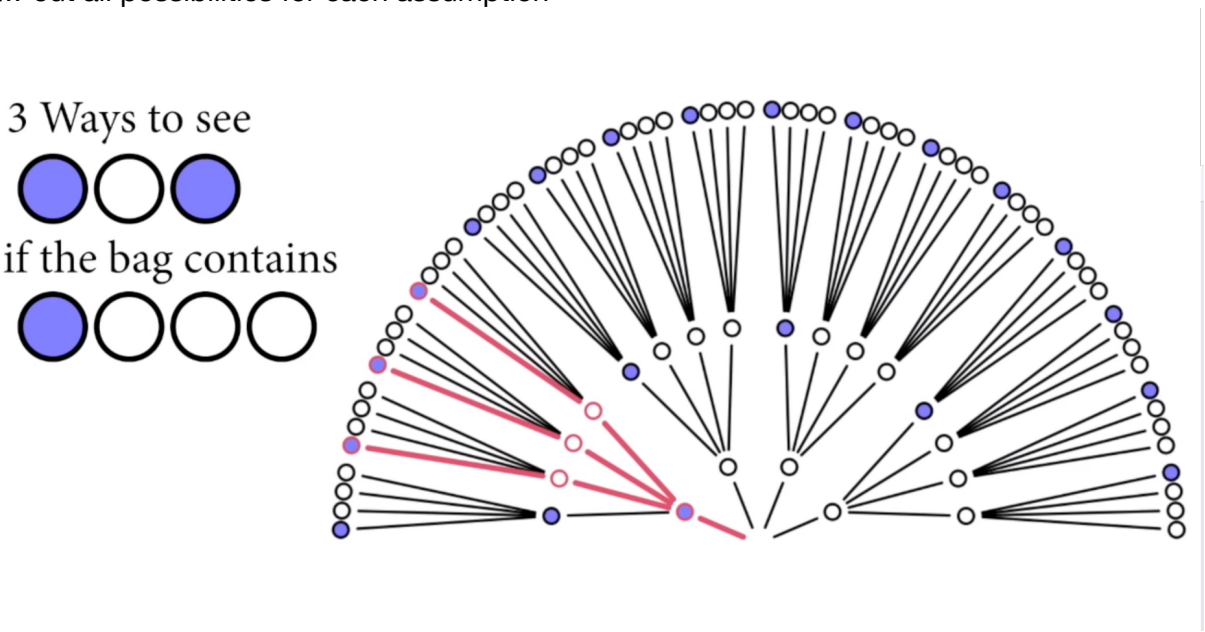
Explanations with more ways to produce the data = more plausible.

4 marbles in a bag. Some blue, some white. We will pick marbles one by one, and use what we have to infer how many are blue and how many white.

If we've sampled with replacement 3 times, and received Blue, White, Blue, how can we use that to infer the contents of the bag.

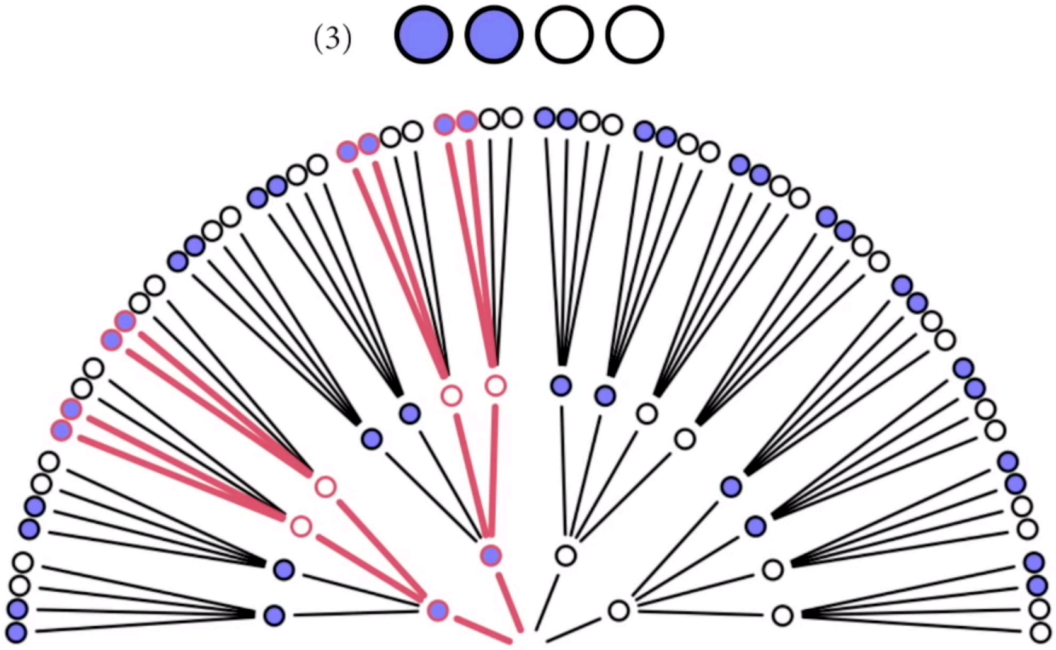
Garden of forking data

Draw out all possibilities for each assumption

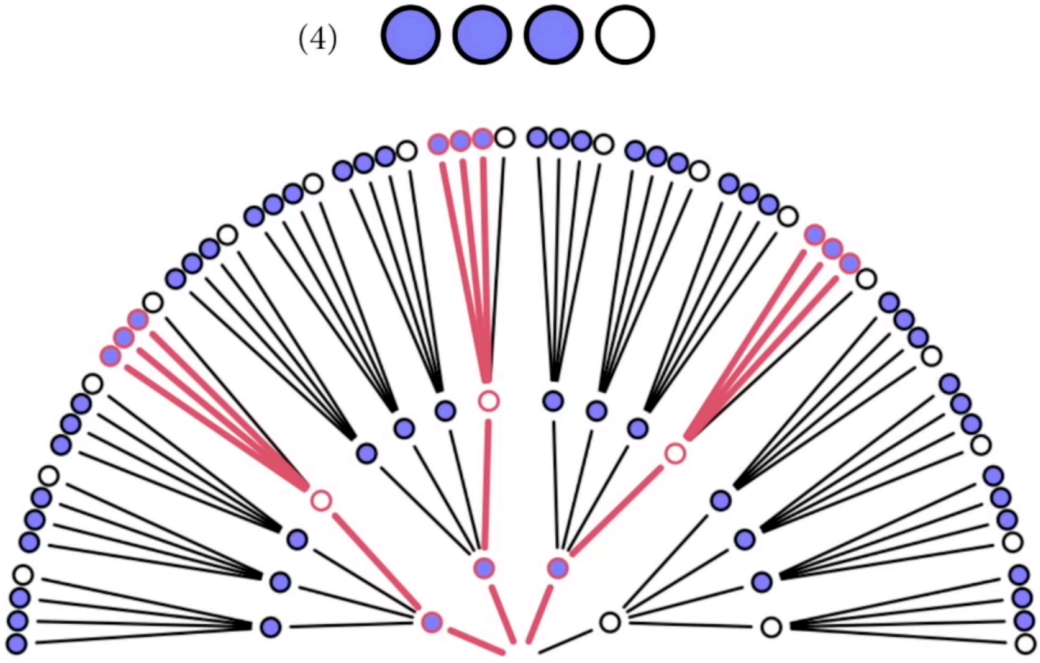


Here, if we start with the **assumption that there is one blue marble and three white marbles,** how many ways can we possibly reach our drawing of **Blue, White, Blue?**

so **if the bag contains 1 blue and 3 white marbles, there are 3 possible ways to get Blue, White, Blue**



If the bag contains 2 blue, 2 white, there are 8 possible ways to get Blue, White, Blue



If the bag contains 3 blue, 1 white, there are 9 possible ways to get Blue, White, Blue

According to Bayesian inference, it is most plausible that the bag has 3 blue, 1 white marble, because *things that can happen in more ways are more plausible.*

Possible composition	p	ways to produce data	plausibility
[○○○○]	0	0	0
[●○○○]	0.25	3	0.15
[●●○○]	0.5	8	0.40
[●●●○]	0.75	9	0.45
[●●●●]	1	0	0

```
ways <- c( 3 , 8 , 9 )
ways/sum(ways)
```

R code 2.1

```
[1] 0.15 0.40 0.45
```

WHAT IF WE DRAW ANOTHER MARBLE FROM THE BAG?

Bayesian Updating: If we continue data collection, we can multiply the probability of the next step by the existing probabilities of the previous steps.
If we draw another blue marble, it is now much more likely that there are 3 blue marbles in the bag.

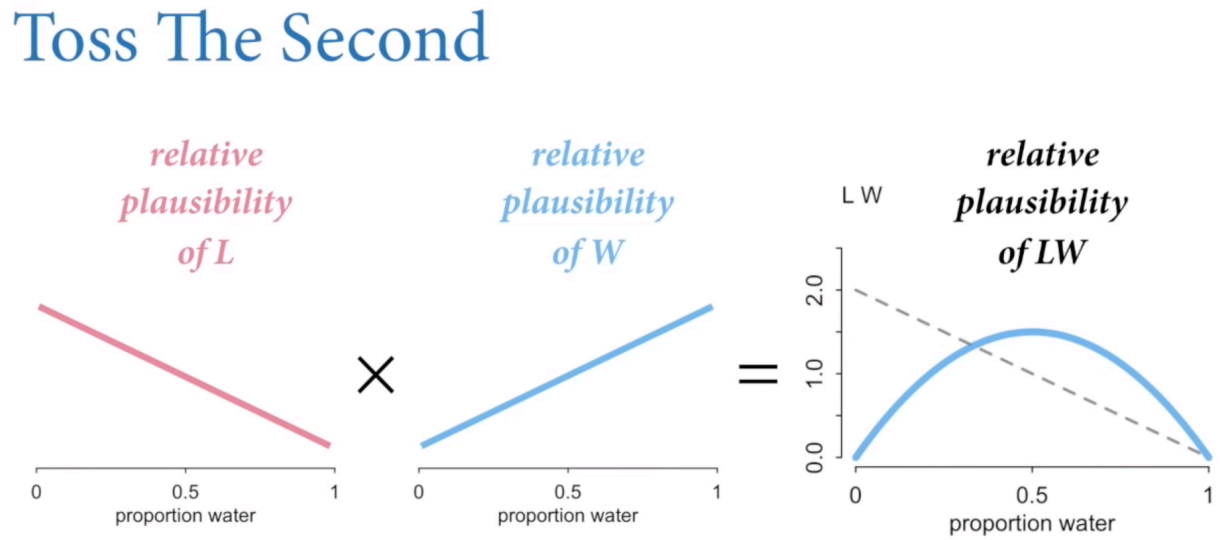
Another draw from the bag: ●

Conjecture	Ways to produce ●	Previous counts	New count
[○○○○]	0	0	$0 \times 0 = 0$
[●○○○]	1	3	$3 \times 1 = 3$
[●●○○]	2	8	$8 \times 2 = 16$
[●●●○]	3	9	$9 \times 3 = 27$
[●●●●]	4	0	$0 \times 4 = 0$

Ok, so if we apply these principles back to the globe question - each new time we collect a datapoint, we change a density curve guessing the proportion of water on Earth.

On the First toss, if we get Land, our prior guesses that it's most likely that the Earth is all Land, but the posterior is a diagonal line from 0 proportion water to 1 proportion water, because if the Earth was 100% water, our observation of Land would be impossible. The first prior/posterior are created totally by **logic.**

Also, **what is density, anyway?**
Density is relative plausibility of the proportion water on earth. For now, the actual number does not matter.



As we gain data, we multiply the diagonal curves of each individual toss together to get the 2-toss curve.

All that Bayesian Updating is, is multiplying a bunch of these diagonal curves as you keep gaining data. Each new curve comes solely from multiplying individual diagonal curves for each new datapoint.

HOW IS THIS DIFFERENT FROM OTHER ANALYSES?

1. there is no minimum sample size!
   - Even with 1 datapoint, we have more idea as to what the truth is as we did in the beginning.

Question: Is there a minimum sample size for what is *useful?* For example, in this first example, our single datapoint prediction predicts the actual truth of proportion water (~.7) as *less plausible* than random. Is this useful?? - zero-informed prior is a very stupid point of departure. Common sense facts can change your analyses a lot.

2. Shape embodies sample size!
   - instead of calculating degrees of freedom or other quantitative measures of sample size, the curve is build *out of* your sampling that has already been done

3. No point estimates!
   - Once you build a plausibility curve, you don't use a single point from the curve, because no point on the curve is more valid than the others. In Bayesian Inference, ***the distribution is the estimate***. Always use the *entire distribution.*

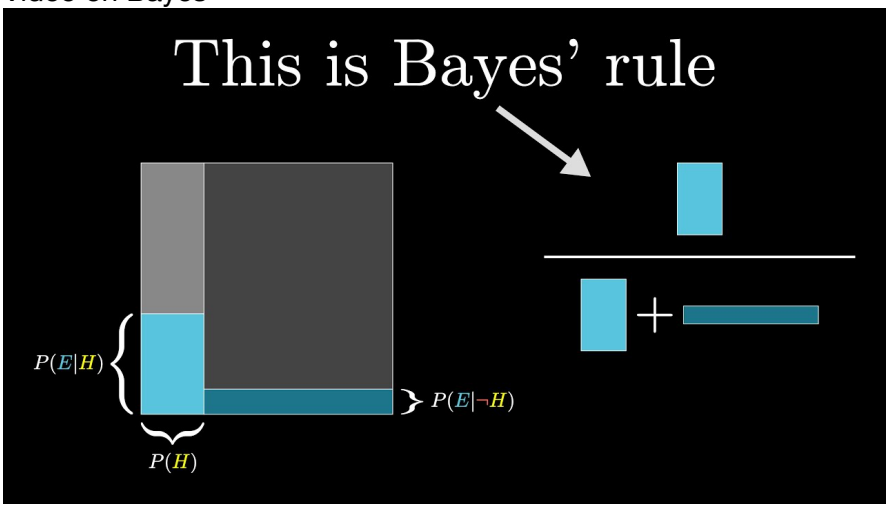
4. No one interval is true.

- Intervals communicate the shape of the posterior, but there are an infinite number of intervals you could draw. For example, you might use the 50% of centre of the sample, but there is actually no logical reason that you would use this rather than anything else.
- 89% interval is a class favourite. Nothing actually changes at the end of the intervals, except that those answers are less plausible.
- 95% interval is an obvious superstition. Nothing magical happens at the boundary.

- Question: at 52:00, why is the 5 values curve not even on .25 and .75?

FROM POSTERIOR TO PREDICTION

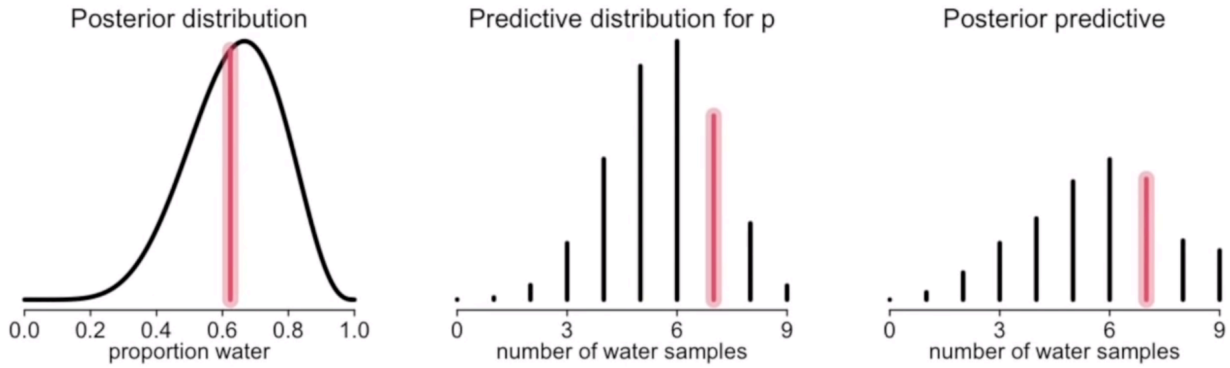
Video on Bayes



Implications of the model depend on the **entire** posterior, we **must** average any inference over the entire posterior.

Usually, this requires integral calculus (this will average over the entire posterior distribution, OR we can just take samples from the posterior to do calculus without realizing it.

Uncertainty ⇨ Causal model ⇨ Implications



Pink line on the left is our posterior distribution. If we sample a random place on that distribution (pink line), we will get a predictive distribution for p.

The posterior predictive is our calculation of collected predictions, averaged over the uncertainty of the entire posterior distribution.

Question: will the posterior distribution and posterior predictive ever be different? what does it mean if they are??

OVERVIEW

Bayesian data analysis is
1. find all possible explanations of the data
2. count all the ways that data can happen
3. find which answer has the most possible ways to get there - that's your prediction!

There are no guarantees in Bayesian analysis, except that it's **logical.** Probability theory is a method logically deducing **implications of data** under the assumptions that you choose.

Any framework selling you more power than this is hiding some assumptions!

Question: does this mean that Bayes can't be mechanistic? Does it tell us reasons, or just answers??

Video 3. Intro to Regressions [link]

CONCEPT INTRO: MERCURY IN RETROGRADE

Orbits are weird, and when we watch planets go by, there's a point where they move backwards (e.g. "Mercury in Retrograde"). This is due to the fact that Earth orbits faster than Mars. So, if I'm an ancient astronomer, how do I figure out when it will be moving backwards?

Model 1. Real model: Earth and Mars both revolving around the sun, different speeds, so at some point, because Earth is fast,
Model 2. Geocentric model, whereby other planets revolve around Earth, but also orbit themselves. These models are descriptively accurate, but mechanistically wrong. It's possible to make good and accurate predictions, but without understanding *why* things are actually happening. Prediction Without Explanation!

Similarly, **Linear Regression** cab be mechanistically wrong, but still descriptively accurate, and still useful in predicting one variable using others.

WHAT ARE LINEAR REGRESSIONS?

- simple Golems
- models of Mean and Variance of a variable.
- mean as **weighted sum** of other variables
- many other tests are actually regressions: ANOVA, ANCOVA, etc etc etc.

Enter: Normal Distribution

A normal distribution gives you the most common number of ways that the answer is possible.

Example of students lining up on a football field and flipping coins- heads = step right, tails = step left. As coin tosses continue, students keep spreading, and your position is the net sum of those fluctuations in your coin flips. If we do this with a few hundred students, and take a histogram, the histogram will always be highest on the midfield, because **there are many more ways (or sequences of throws) that result in a net-zero difference than ways that result in a +10 difference.**

Why are normal distributions so normal???

Generative Argument: they arrive more commonly in nature, because they arrive from summed fluctuations, and summed fluctuations tend to cancel.

Statistical Argument: for estimating mean and variance, normal distribution is the least informative distribution. The ONLY info in it is that it has a mean and a variance.

Question: can we discuss the differences between these arguments??

A variable does not have to be normally distributed for normal models to be useful. (for example, you could predict the mean and variance of a uniformly distributed variable)

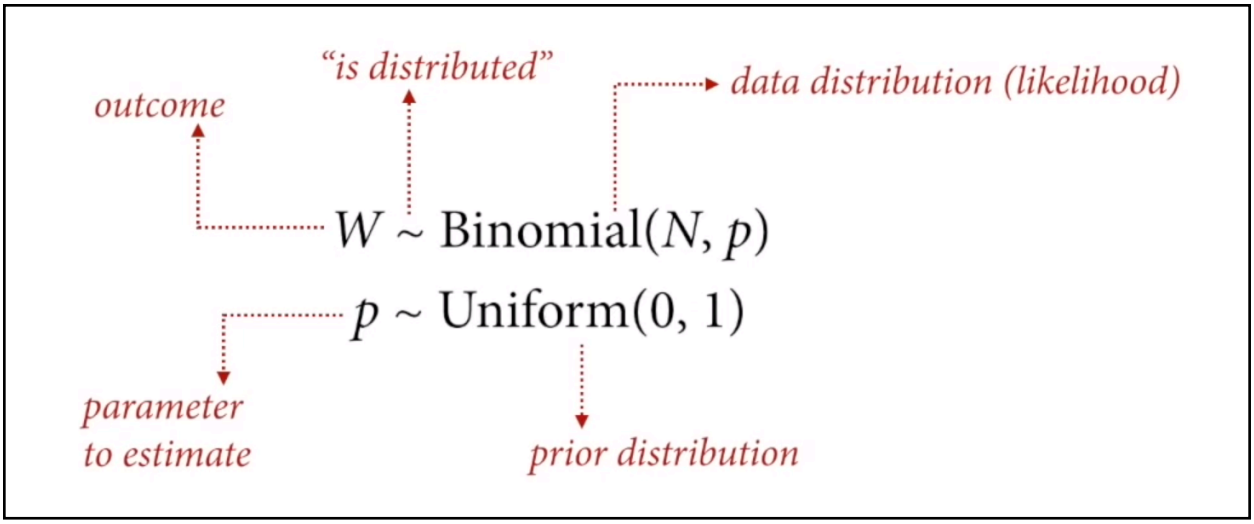
MAKING NORMAL MODELS:

GOALS:

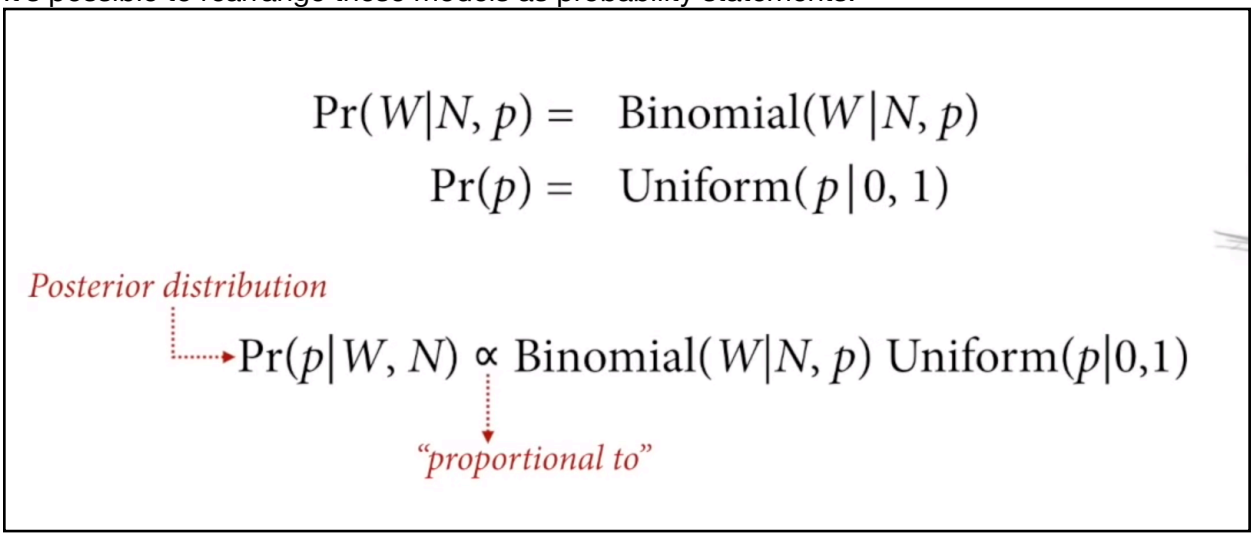
- language for representing models
- how to calculate bigger posterior distributions
- constructing and understanding linear models

LANGUAGE FOR MODELING

Revisit the globe tossing model:



It's possible to rearrange these models as probability statements.



If we think about Height and Weight, we can have a linear set of points, then we can make a DAG, such that H —> W (height influences weight, but not the other way around).

So then we know W = f(H), where weight is some function of height, but we don't yet know what that function is.

When making generative models, there are a couple of kinds:

Dynamic generative models would take measurements of these two variables though time and find rates of change at increments throughout lifespan. (height plateaus, and height velocity decreases to zero). Gaussian variation result of summed fluctuations.

Static generative models simply say changes in height result in changes in weight, but there is no mechanism as to *why.* This is what we'll use today. Gaussian variation result of growth history.

Linear Regression vs Line:
Linear regression does not use the actual Y distributions, but rather the mean Y value, so eventually what you get is "the expectation of Y is condition on X in relation to mu"

Enter Simulations:

Question: what are the units of alpha, beta, and sigma? same as the variable?

Alpha of 0 = if an individual is height of 0, they will also have weight of 0
Beta = slope. If beta is .5, you are gaining .5 unit weight for every unit of height.
Sigma == standard deviation - has to be positive! stretches the distribution, but can't stretch negatively.

Can somebody discuss through the balls-in-a-bowl lines thing? when there are 10 bouncing around?