Introduction to Bayesian inference

Einar Holsbø

Department of Computer Science, University of Tromsø

8th of October, 2021

45 minutes is not enough to introduce a whole field of inference

This is more of a brief glance

On learning statistics

Or: As you are now so once was I

Anthropological observations

Anthropological observations

On principle no two statisticians agree on how to conduct their trade.

Anthropological observations

On principle no two statisticians agree on how to conduct their trade.

If two statisticians seem to agree, ask them a few questions about specifics.

Anthropological observations

On principle no two statisticians agree on how to conduct their trade.

If two statisticians seem to agree, ask them a few questions about specifics.

A statistician on a desert island can always survive by disagreeing with herself.

Anthropological observations

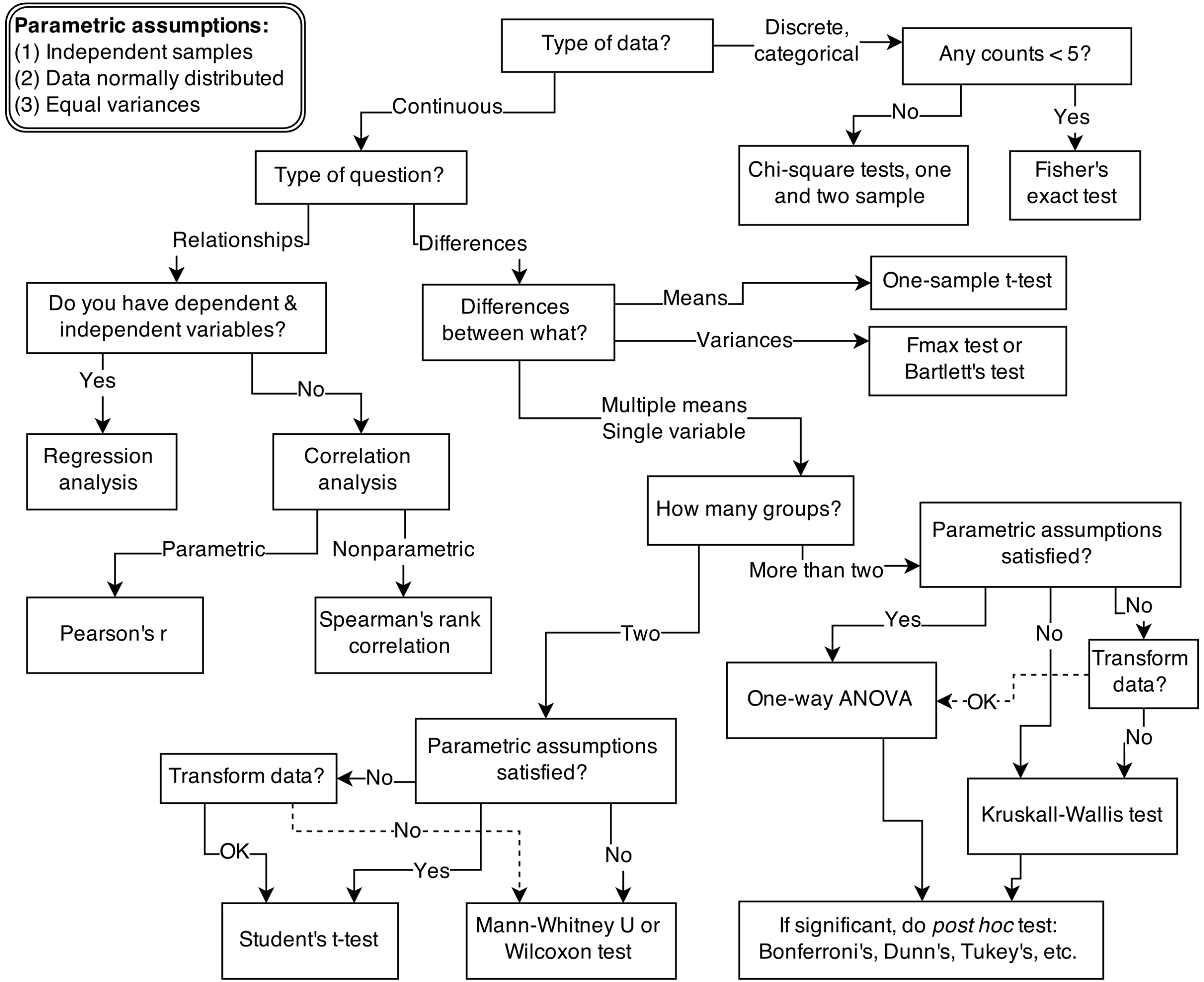
On principle no two statisticians agree on how to conduct their trade.

If two statisticians seem to agree, ask them a few questions about specifics.

A statistician on a desert island can always survive by disagreeing with herself.

The upshot: statistics is very confusing from the outside

A Flowchart from Hell



The Flowchart from Hell is enticing because it (sort of) helps navigate a large set of locally optimal procedures

But it's not statistics

Bayesian inference is about building a model for **your** particular problem

Two situations

You're shopping for a vacuum cleaner online

- Reputable Electronics Company, model A:
  - **Rating:** 3.9/5 (7 users)
  - **Cost:** Pricey
- Shady Practices Inc., model B:
  - **Rating:** 4.8/5 (6 users)
  - **Cost:** Moderate

Which one do you get?

My assessment as a savvy vacuum buyer

My assessment as a savvy vacuum buyer

Shady Practices products tend to catch fire in the first few weeks.

My assessment as a savvy vacuum buyer

Shady Practices products tend to catch fire in the first few weeks.

Usual ratings for SP products are more like 1 — 2 out of 5

My assessment as a savvy vacuum buyer

Shady Practices products tend to catch fire in the first few weeks.

Usual ratings for SP products are more like 1 — 2 out of 5

5 reviews is not a lot. My guess is the final rating will be around 2/5

My assessment as a savvy vacuum buyer

Shady Practices products tend to catch fire in the first few weeks.

Usual ratings for SP products are more like 1 — 2 out of 5

5 reviews is not a lot. My guess is the final rating will be around 2/5

Get the pricey one! Won't have to buy three to last the year.

Your routine screening looks bad

https://www.flickr.com/photos/iloasiapacific/8055935073

- The test came back positive for *spare ribs*
- It's very serious
- We are informed that the test is 99% accurate

Do you worry?

My asesssment as a savvy patient

My asessment as a savvy patient

I happen to know that spare ribs is pretty rare

My asessment as a savvy patient

I happen to know that spare ribs is pretty rare

I guess it affects about 1 out of every 10000 people

My asessment as a savvy patient

I happen to know that spare ribs is pretty rare

I guess it affects about 1 out of every 10000 people

Routine screening: I'd say the probability for spare ribs is like 1%

My asessment as a savvy patient

I happen to know that spare ribs is pretty rare

I guess it affects about 1 out of every 10000 people

Routine screening: I'd say the probability for spare ribs is like 1%

How we interpret data relies on what we know about the world!

These are examples of Bayesian reasoning:

How should the data change our opinions?

To quanitify how our opinions should change we need two pieces of information:

To quanitify how our opinions should change we need two pieces of information:

- A reasonable assumption about plausible effects
  - (what do we know about the world?)

To quanitify how our opinions should change we need two pieces of information:

- A reasonable assumption about plausible effects
  - (what do we know about the world?)
- The degree to which the data support different effects

To quanitify how our opinions should change we need two pieces of information:

- A reasonable assumption about plausible effects
  - (what do we know about the world?)

- The degree to which the data support different effects

The final opinion (inference) is a compromise between these

Vacuum example:

- **Reasonable assumption:** SP makes trash, ratings will be low (1-2)
- **Data support:** high ratings from few people (4.8)
- **Inference:** Perhaps this will be a top tier SP rating (2-ish)

Spare ribs example:

- **Reasonable assumption:** Routine screening; few people get SR (1 in 10k)
- **Data support:** Positive test, quite accurate (99 in 100)
- **Inference:** Probability for SR is low (1 in 100)

The combination of data and assumptions into final inference is fundamentally about counting.

Bayesian inference is counting

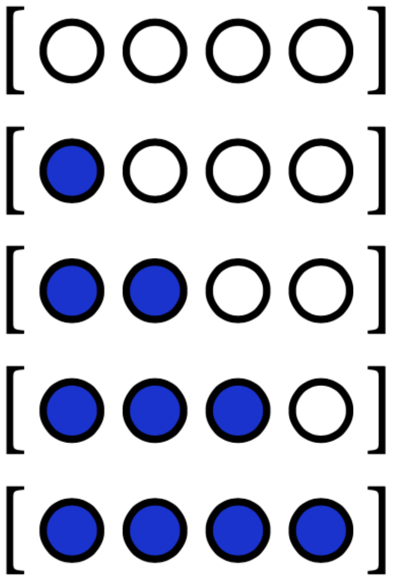
Consider the drawing of marbles from a bag:

- The bag contains four marbles
- A marble is either blue or white

Bayesian inference is counting

Consider the drawing of marbles from a bag:

- The bag contains four marbles
- A marble is either blue or white



- Five possible hypotheses about the bag's contents
- **Reasonable assumption**: they are equally likely
- Let's gather data!

Bayesian inference is counting

Experimental protocol:

- We draw three marbles with replacement
  - Draw a marble
  - Record its color
  - Put it back, shake the bag vigorously
- Resulting data: 

Bayesian inference is counting

What support do the data  lend to our five hypotheses?

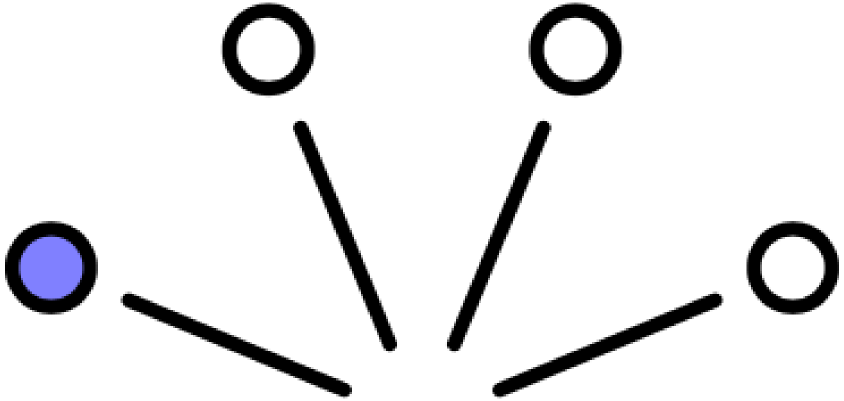
Quantify by **counting** the number of ways in which each hypothesis could generate the observed sequence.

Bayesian inference is counting

What support do the data  lend to our five hypotheses?

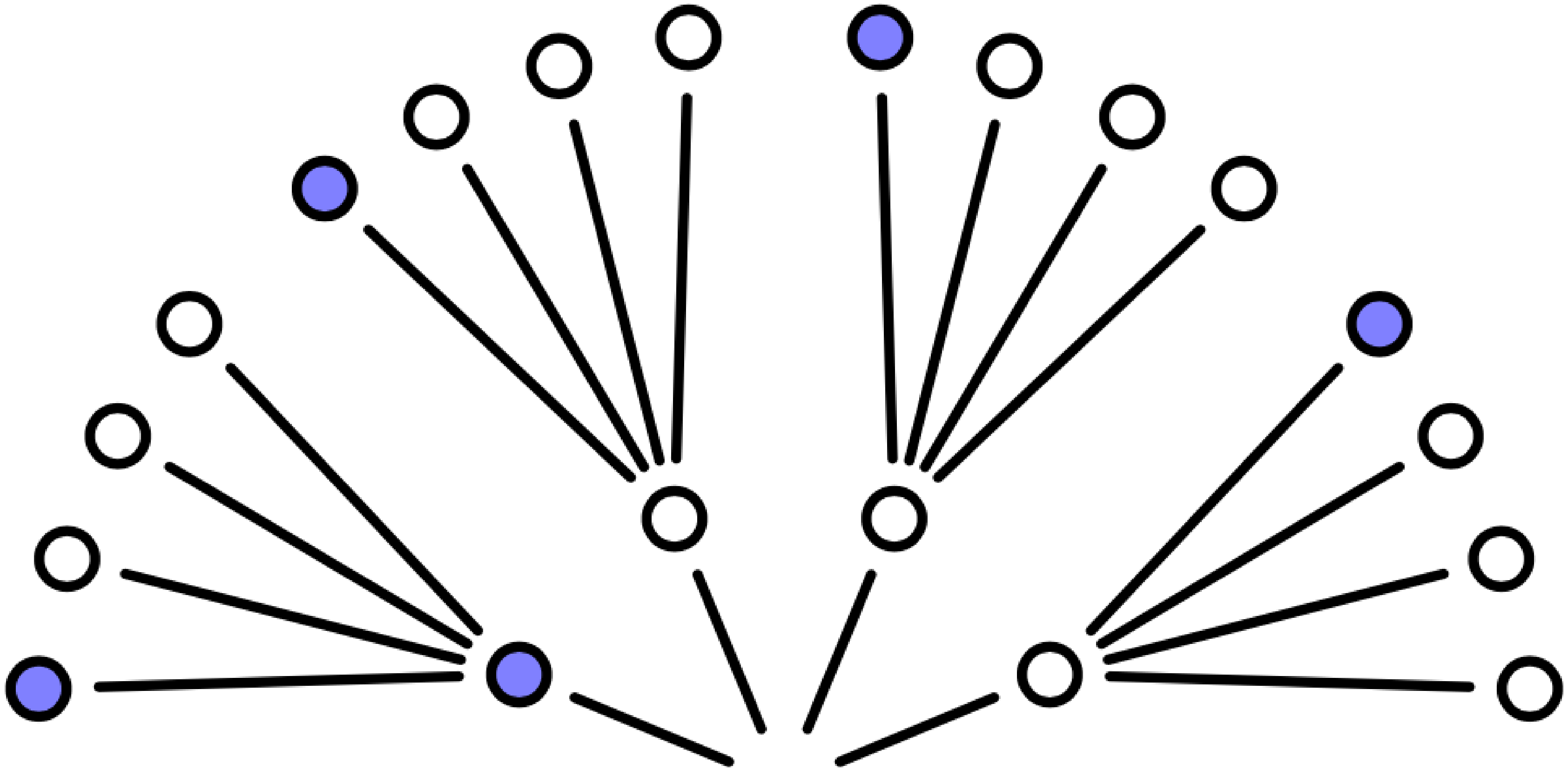
Quantify by **counting** the number of ways in which each hypothesis could generate the observed sequence.

Focus on the  hypothesis. The first draw *could* have happened in four ways:



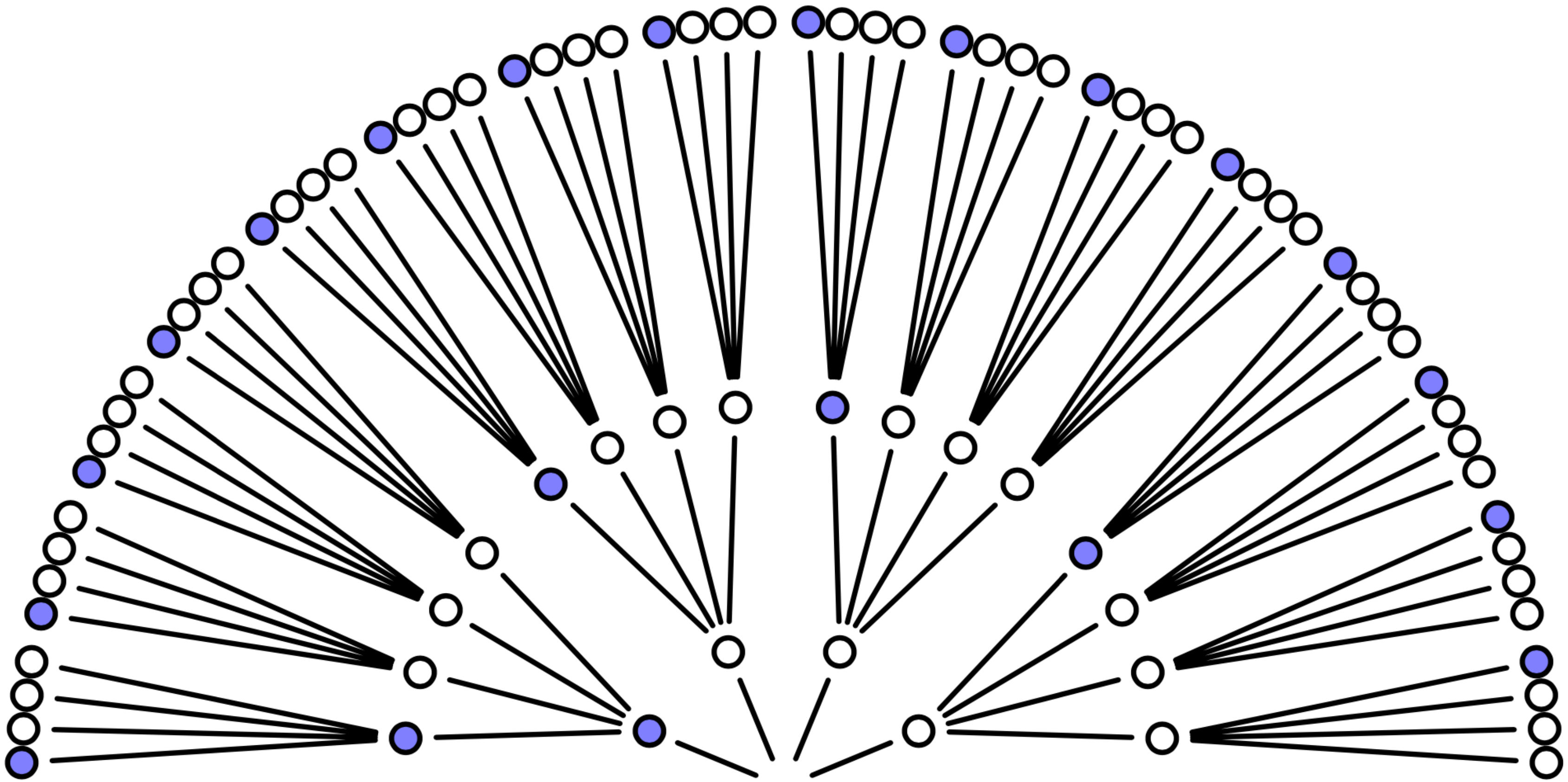
Data: ●○● Hypothesis: [●○○○]

Four possible second draws per first draw:



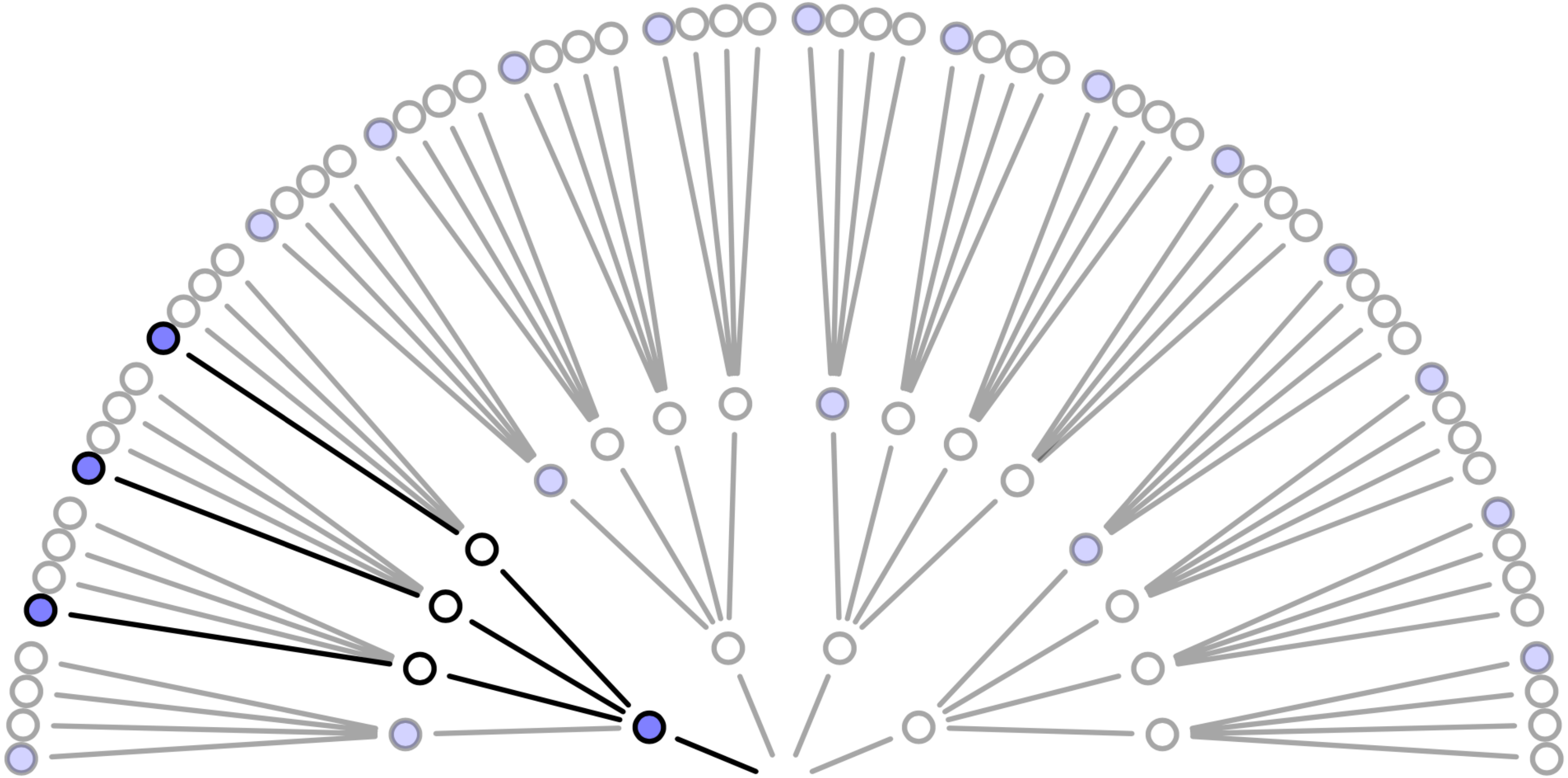
Data: Hypothesis: [ ]

Four possible third draws per second draw:



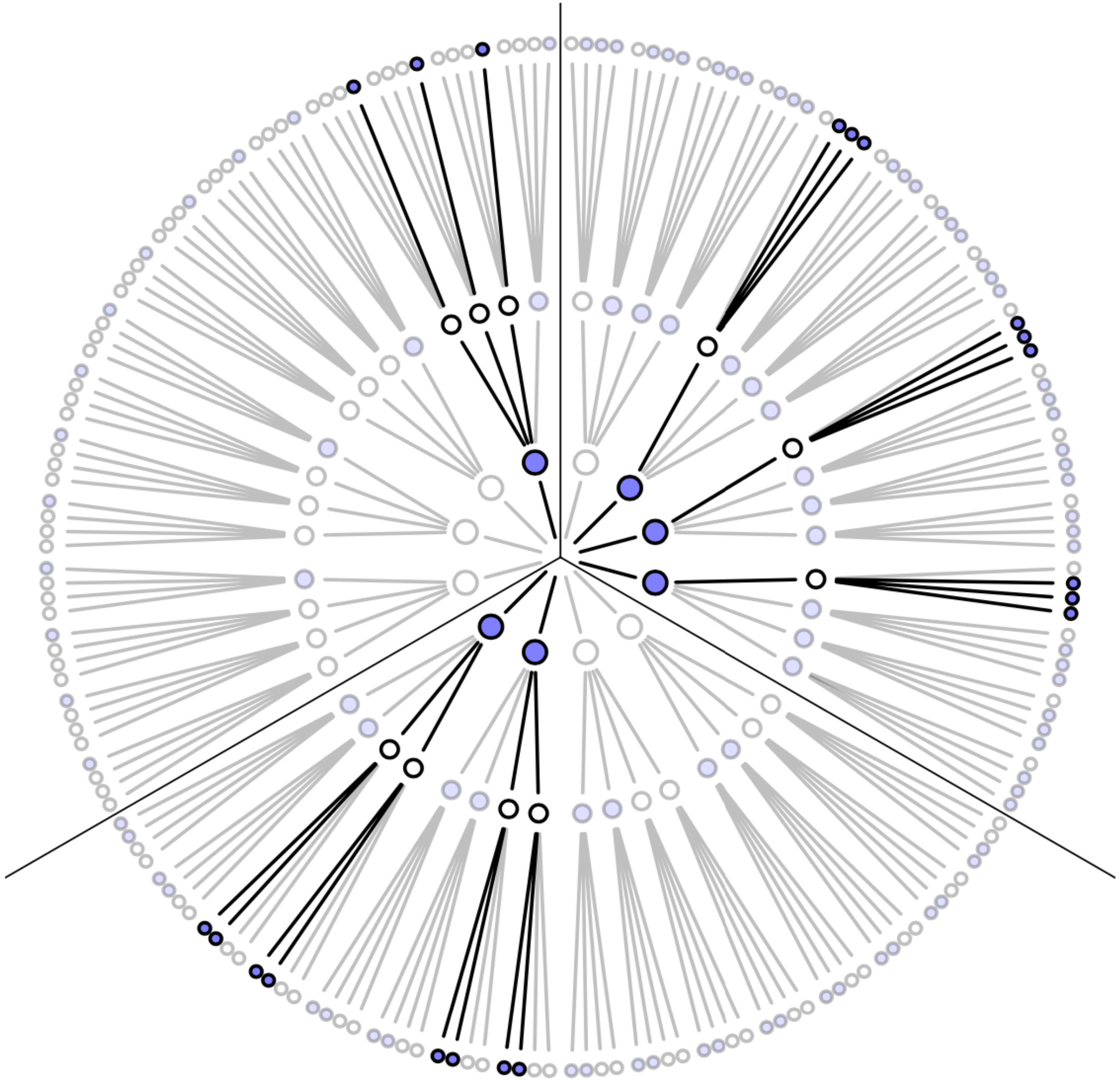
Data: ●○● Hypothesis: [●○○○]

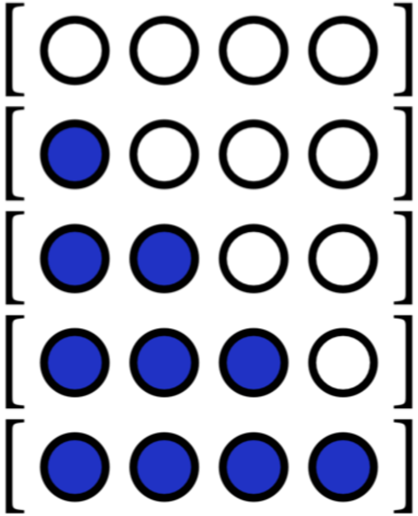
Out of 4x4x4=64 *possible* data sets, only three look like ours:



Two hypotheses are excluded immediately (why?)

Enumerating all possible data sets for the remaining three hypotheses:





Initial “count”
or assumption

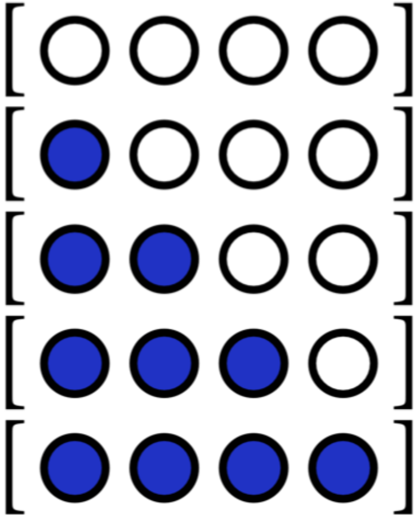
1

1

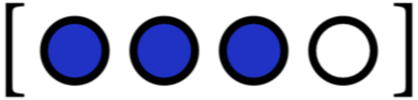
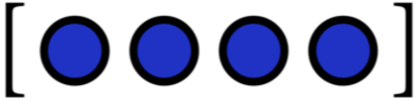
1

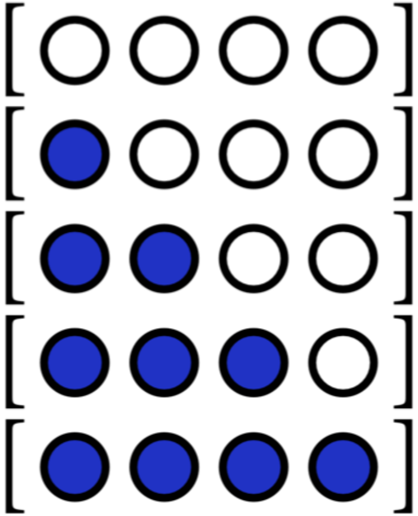
1

1

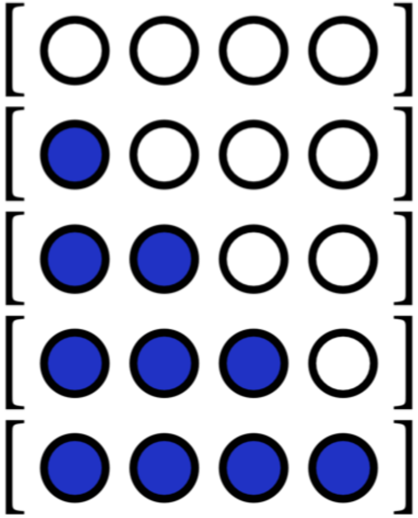


Initial “count” or assumption		Ways to produce data
1	x	0
1	x	3
1	x	8
1	x	9
1	x	0

	Initial “count” or assumption		Ways to produce data		Final “count”
	1	x	0	=	0
	1	x	3	=	3
	1	x	8	=	8
	1	x	9	=	9
	1	x	0	=	0



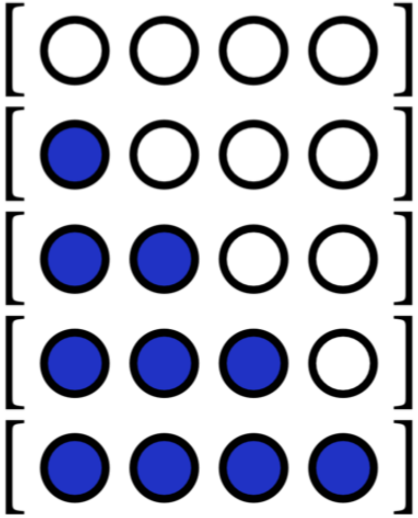
Initial count or assumption		Ways to produce data		Final count	Probability or inference
1	x	0	=	0	0
1	x	3	=	3	.15
1	x	8	=	8	.4
1	x	9	=	9	.45
1	x	0	=	0	0



Initial count or assumption		Ways to produce data		Final count	Probability or inference
1	x	0	=	0	0
1	x	3	=	3	.15
1	x	8	=	8	.4
1	x	9	=	9	.45
1	x	0	=	0	0

Probabilities should sum to 1: divide by total count (20)

Different initial assumption:



Initial count
or assumption

0

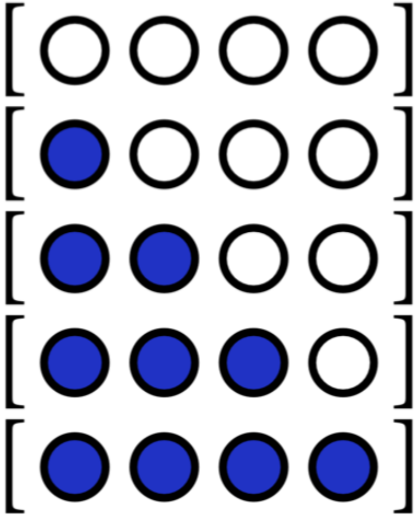
1

2

1

0

Different initial assumption:



Initial count or assumption		Ways to produce data
0	x	0
1	x	3
2	x	8
1	x	9
0	x	0

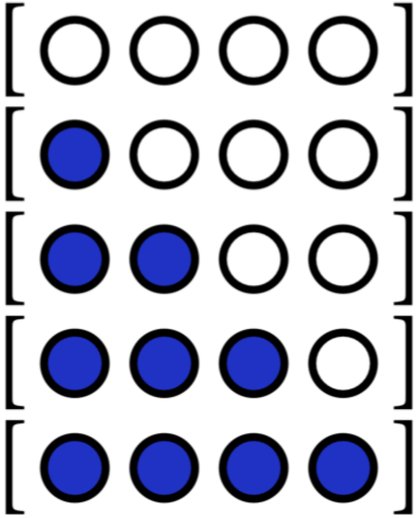
Different initial assumption:

	Initial count or assumption		Ways to produce data		Final count	Probability or inference
[○○○○]	0	x	0	=	0	0
[●○○○]	1	x	3	=	3	.11
[●●○○]	2	x	8	=	16	.57
[●●●○]	1	x	9	=	9	.32
[●●●●]	0	x	0	=	0	0

We draw another marble: ● — the previous counts become the assumption:

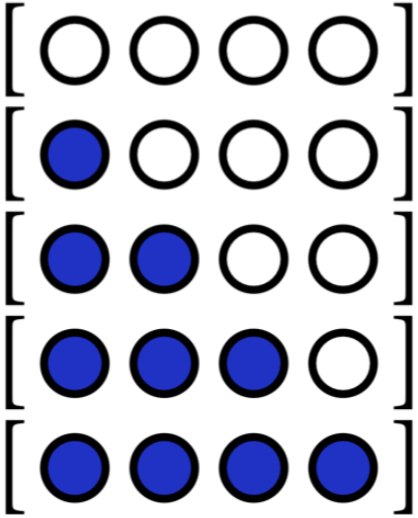
	Initial count or assumption
[○ ○ ○ ○]	0
[● ○ ○ ○]	3
[● ● ○ ○]	16
[● ● ● ○]	9
[● ● ● ●]	0

We draw another marble: — the previous counts become the assumption:



Initial count or assumption		Ways to produce data
0	x	0
3	x	1
16	x	2
9	x	3
0	x	4

We draw another marble:  **— the previous counts become the assumption:**



Initial count or assumption		Ways to produce data		Final count	Probability or inference
0	x	0	=	0	0
3	x	1	=	3	.05
16	x	2	=	32	.52
9	x	3	=	27	.43
0	x	4	=	0	0

We've used **Bayes' rule** from probability theory:

$$P(\text{hypothesis} \mid \text{data}) \propto$$

$$P(\text{hypothesis}) \times P(\text{data} \mid \text{hypothesis})$$

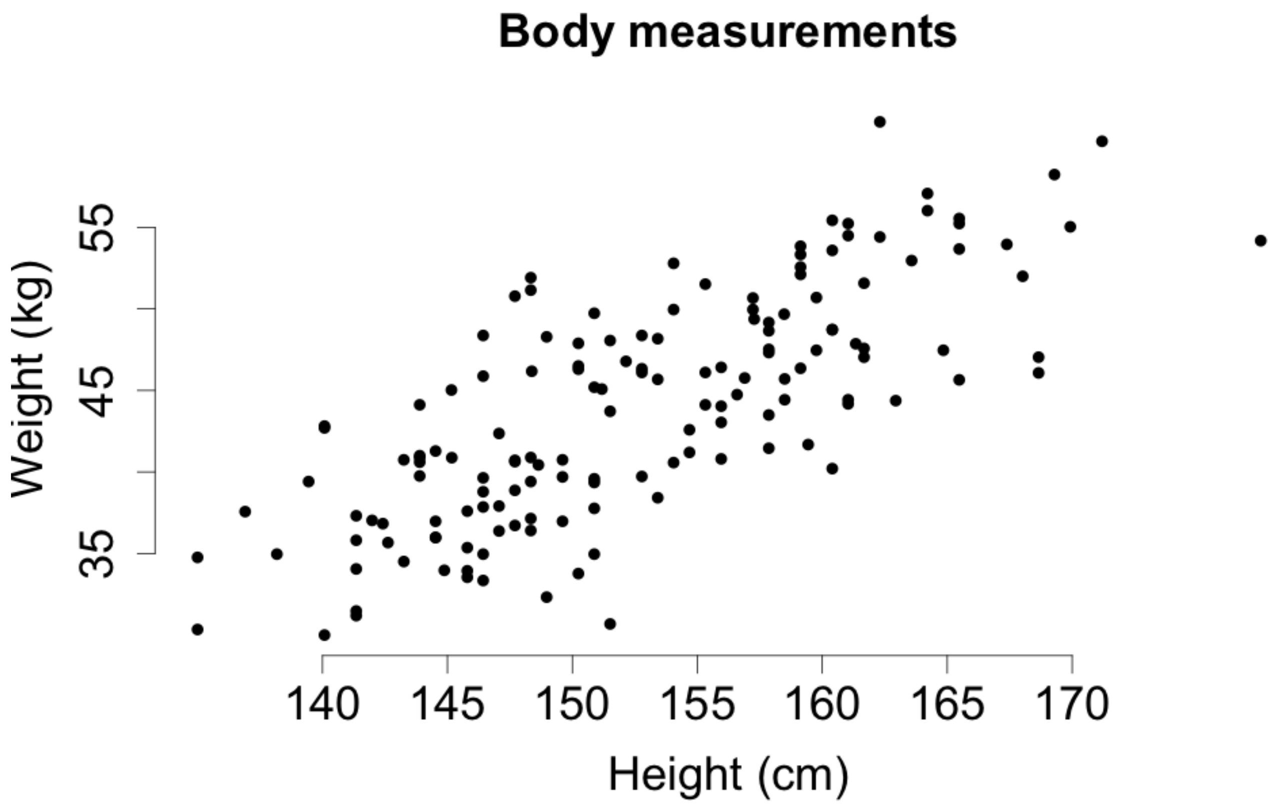
Technical names:

$$\text{posterior } \propto \text{prior} \times \text{likelihood}$$

What we usually count is quite complicated so we get computers to do it

Example: Body weight and height

Some data from a certain African demographic



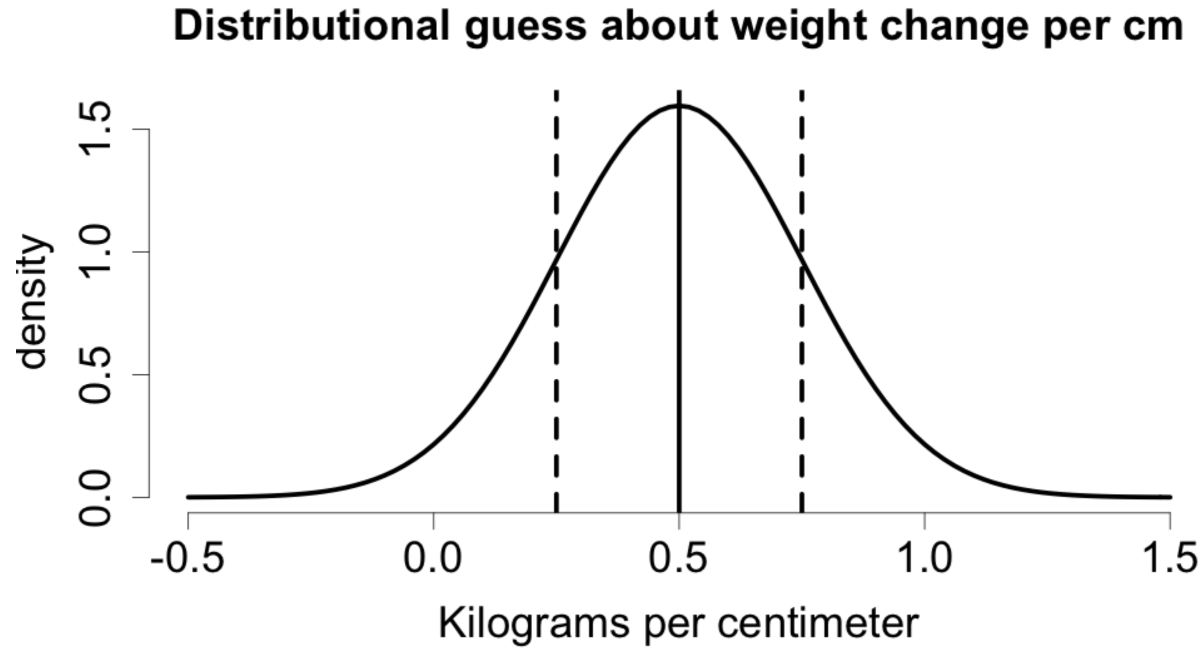
Howell, N. (2010). ***Life Histories of the Dobe !Kung: Food, Fatness, and Well-being over the Life-span.*** **Origins of Human Behavior and Culture. University of California Press.**

What are reasonable assumptions?

- Probably naive to think there is no correlation
- Reasonable to assume weight increases with height?
- How much?

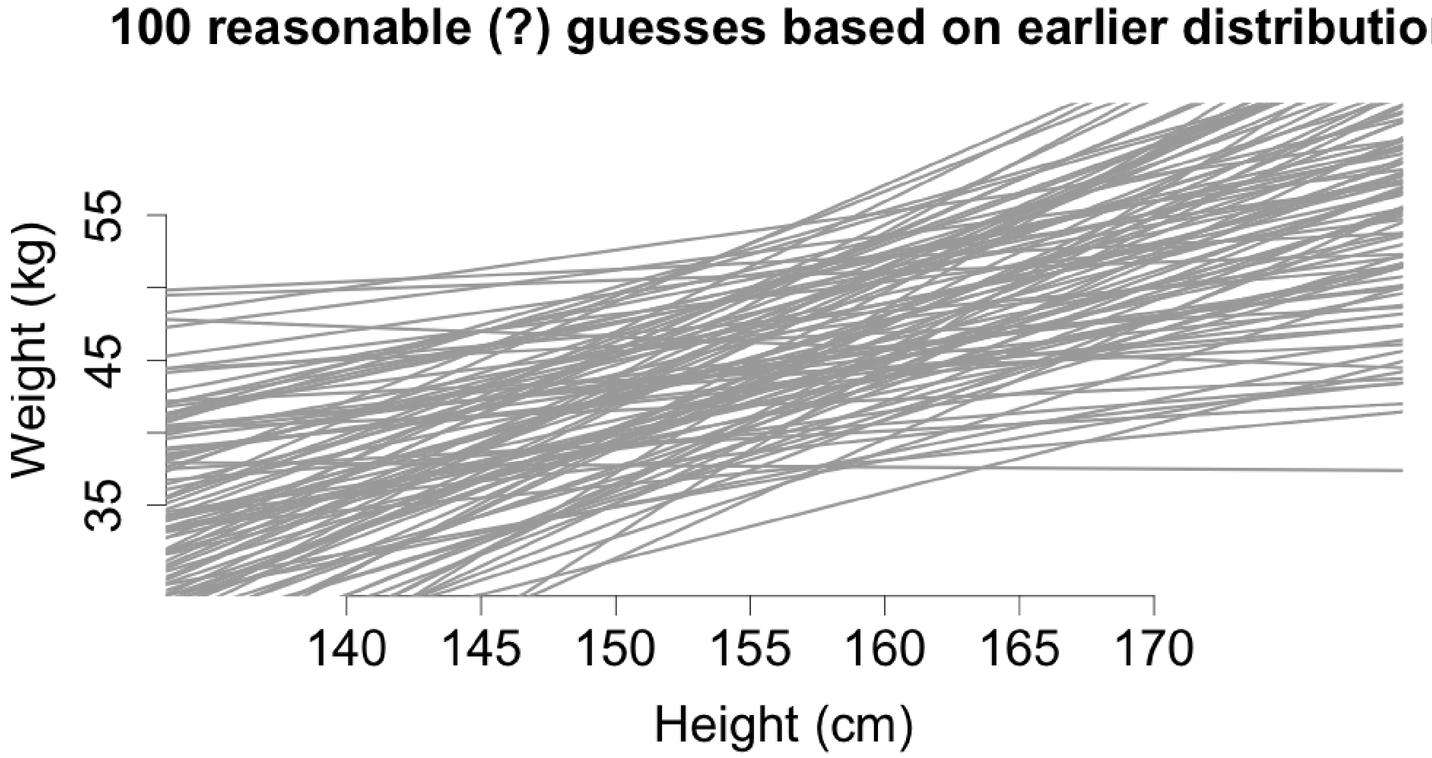
What are reasonable assumptions?

- 1 liter of human weighs about 1kg
- I guess a 1 cm thick cross-section of my trunk is about half a liter
- Rough guess: someone 1 cm taller may on average weigh a half-kilo more
- Plus-minus a quarter-kilo maybe?

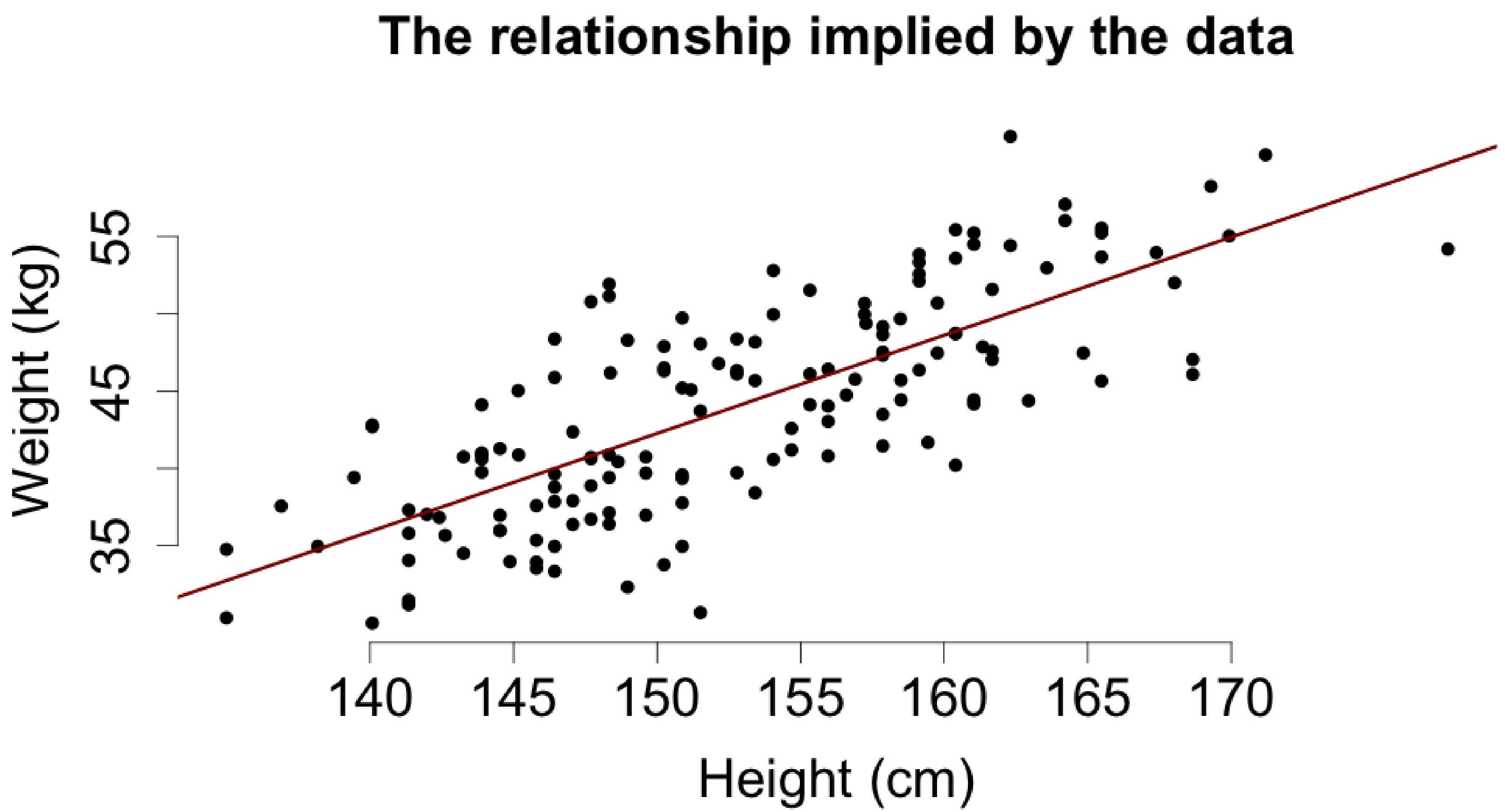


What are reasonable assumptions?

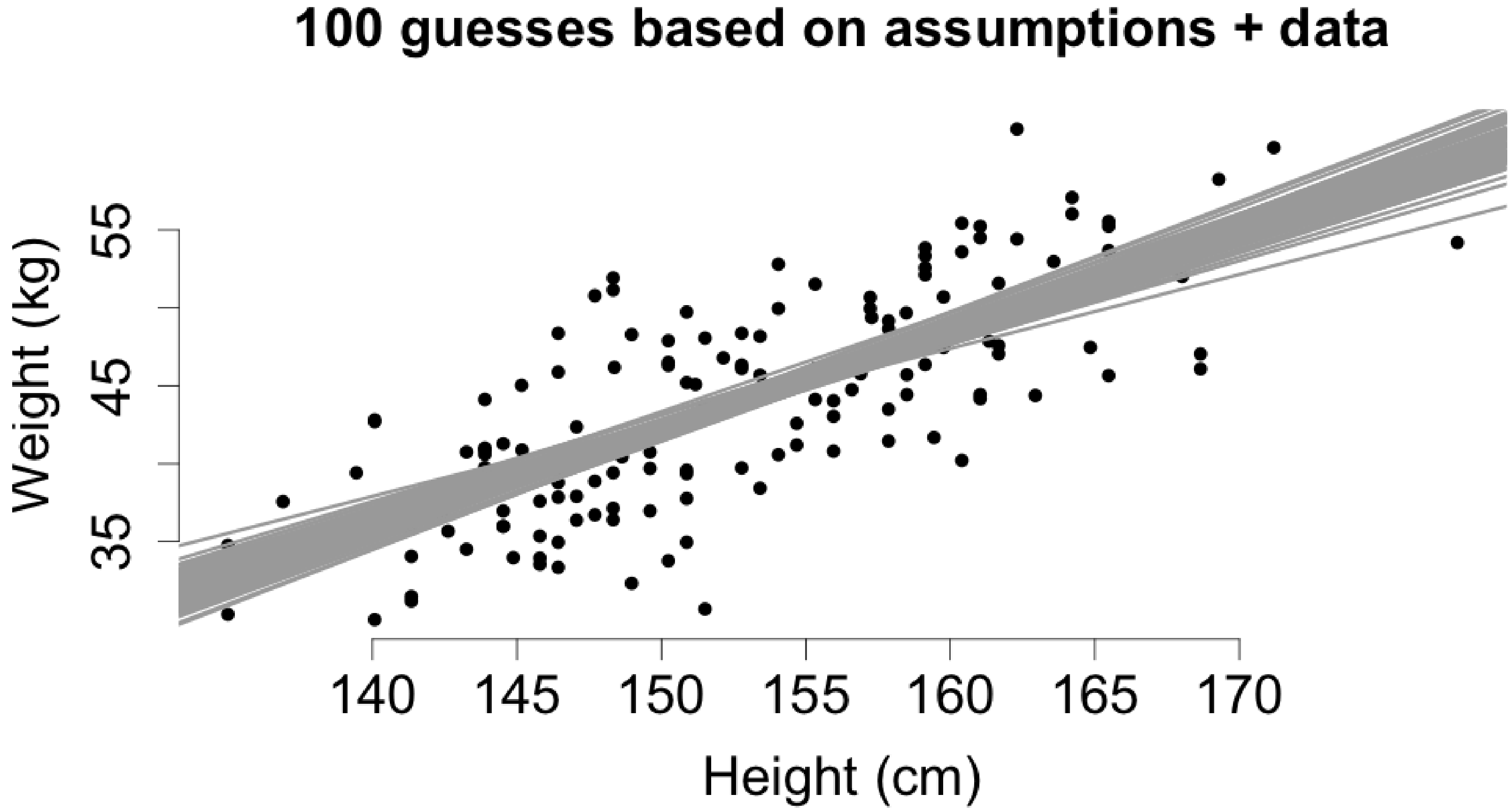
We can easily show the implied relationship by **simulating** from this distribution:

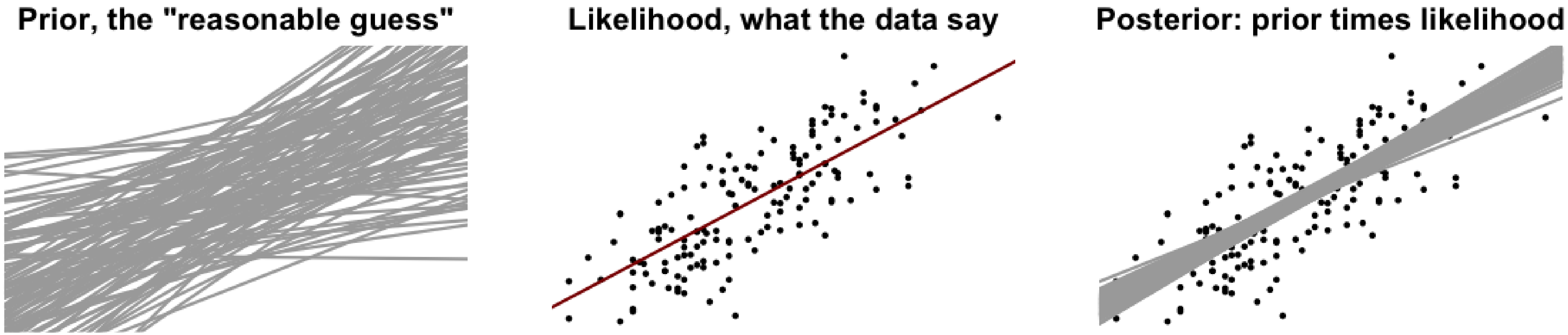


Support of the data for the weight/height relationship:



Final inference

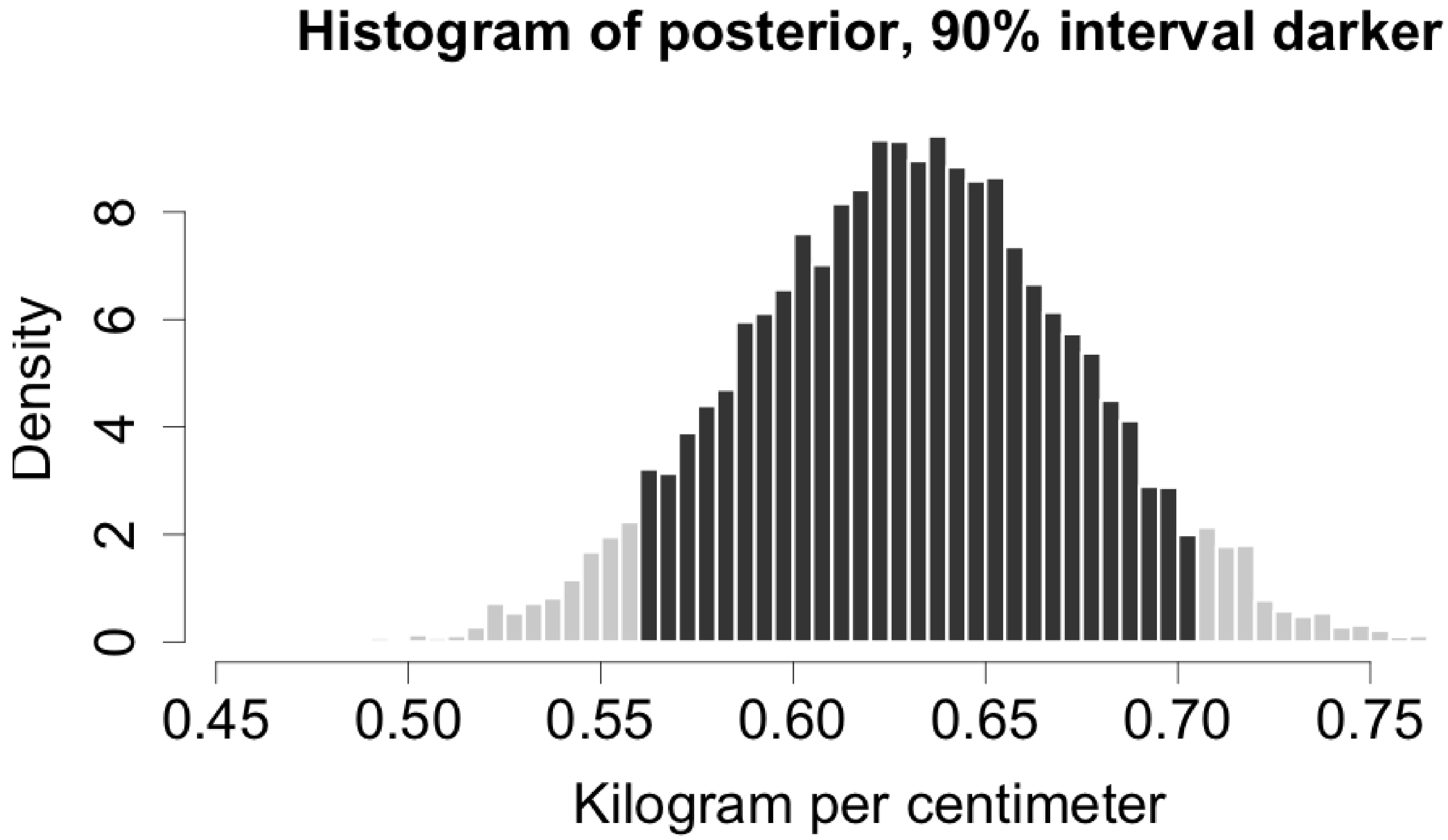



Prior, the "reasonable guess"
Likelihood, what the data say
Posterior: prior times likelihood

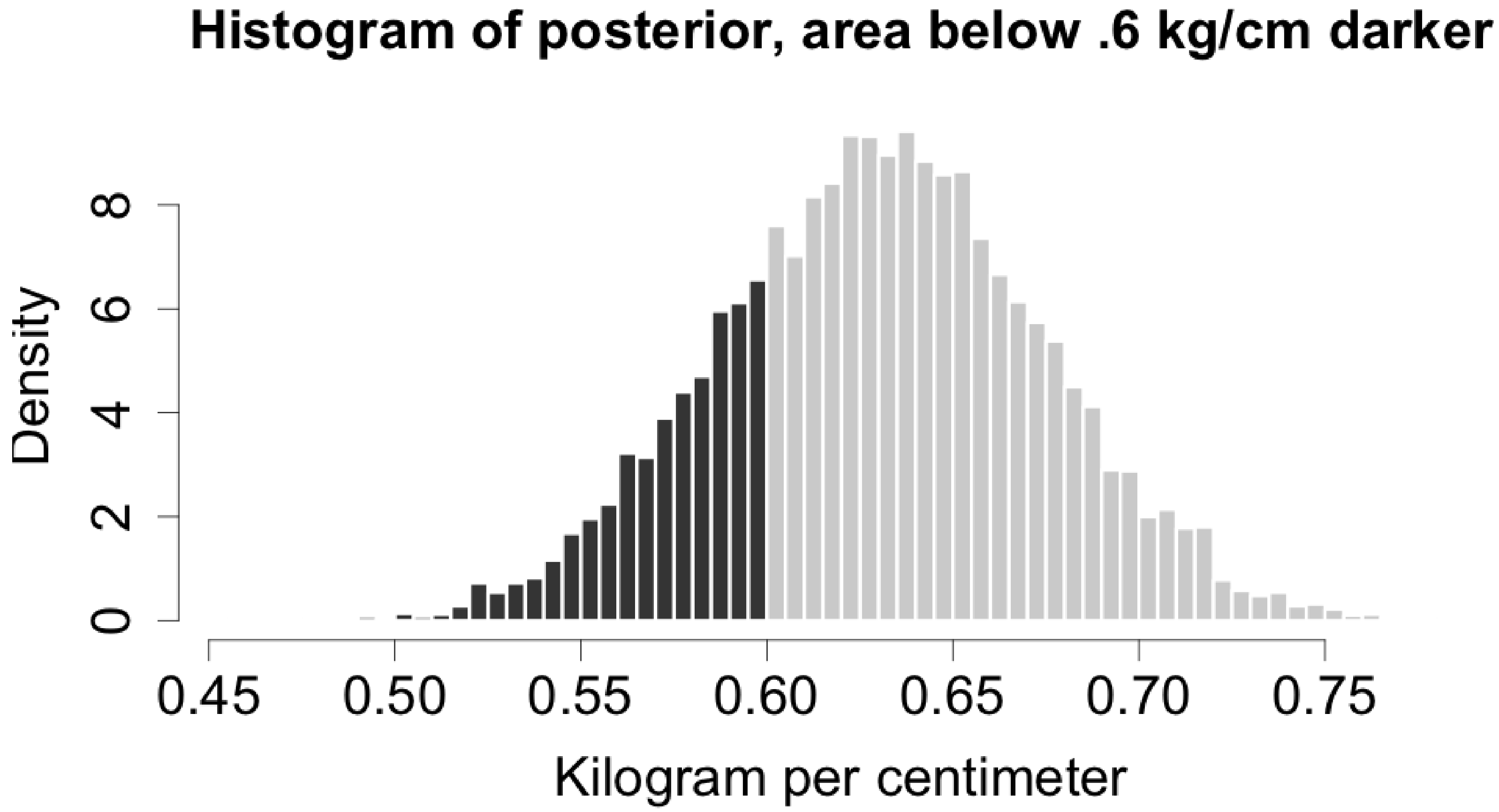
The end-product: (samples from) a distribution

Means we can make probability statements

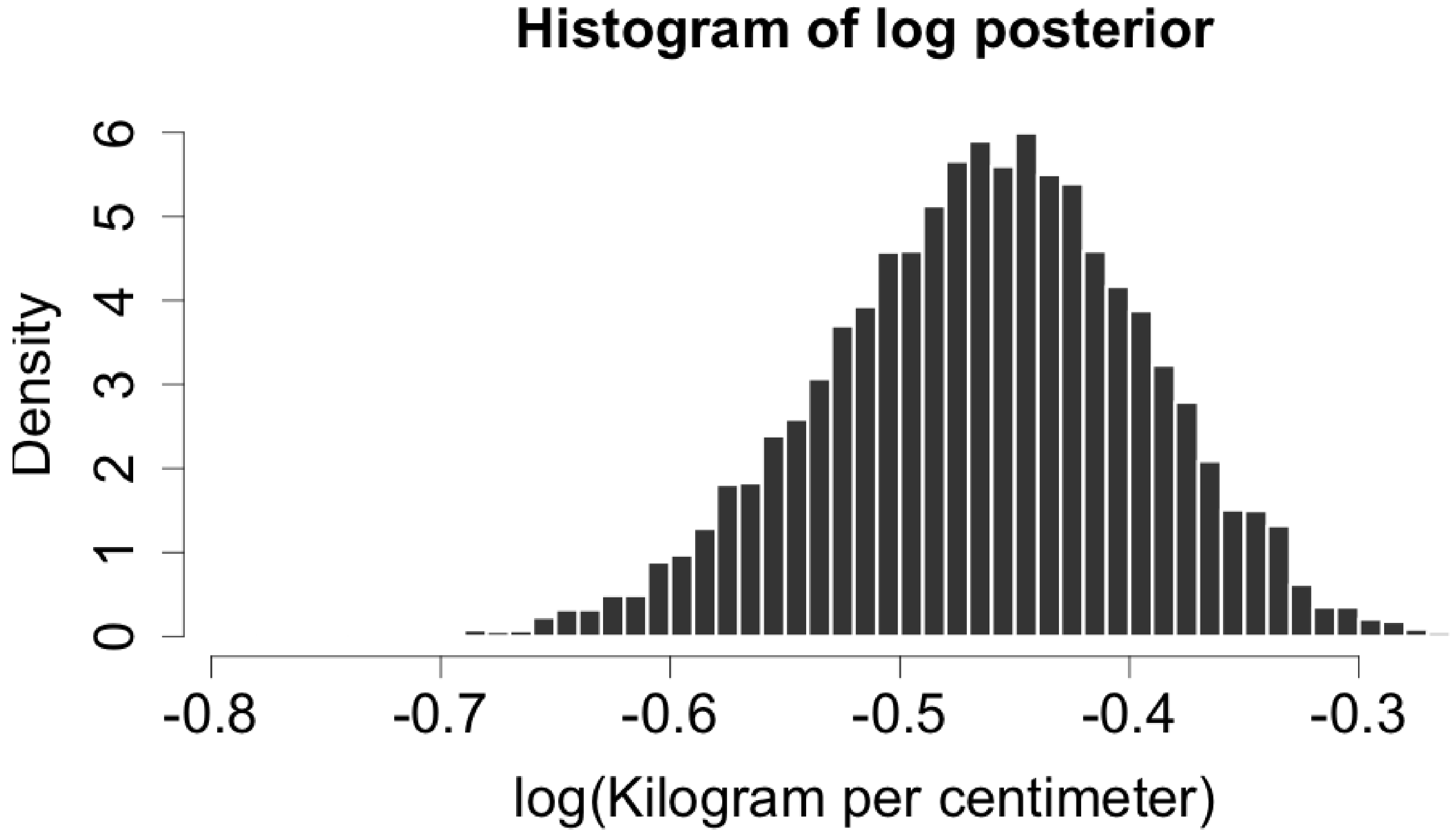
90% probability that the weight/height coef is between 0.56 and 0.7



24% probability that the weight/height coef smaller than 0.6



Transformations: plot shows logarithm of height/weight coef.



In general: posterior inference very flexible, possible to ask diverse questions

The full model looks kind of scary, but highlights flexibility

$$\text{weight } = \alpha + \beta \times \text{height } + \text{ error}$$

$$\alpha \sim \text{Normal}(0, 5)$$

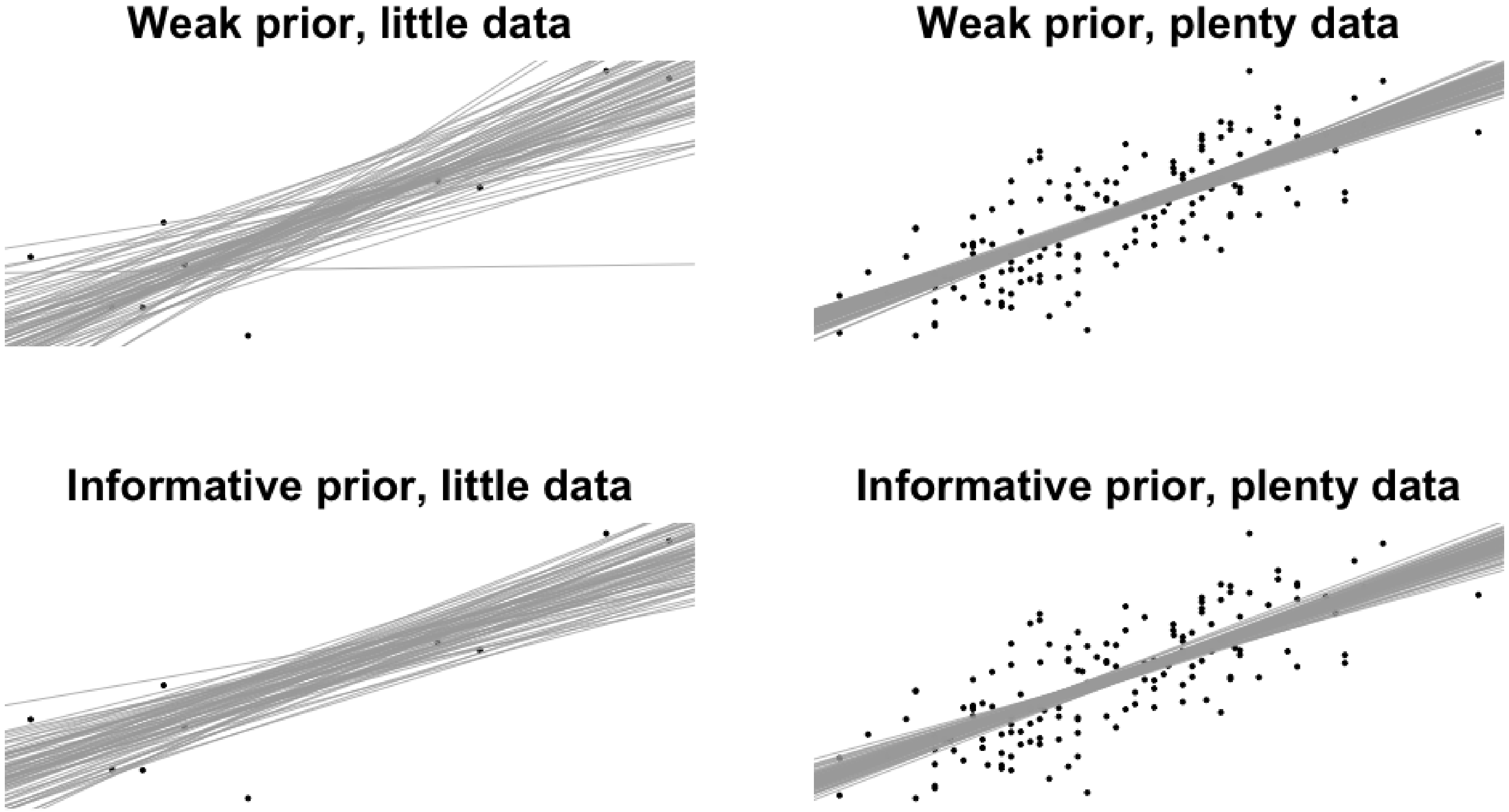
$$\beta \sim \text{Normal}(0.5, 0.25)$$

$$\text{error} \sim \text{Normal}(0, \sigma)$$

$$\sigma \sim \text{Uniform}(0, 50)$$

Bayesian inference is most useful if you have informative prior information and if there is not a lot of data

Variants of the weight/height exercise



If there is lots of data, the prior info becomes irrelevant

Finishing up: some pros/cons

Some pros:

- Very flexible
- Can borrow information in various ways
  - From prior
  - From similar situations (Hospital A similar to Hospital B?)
- I specified all parts of the model
  - I can (maybe) defend it
  - You can more easily critique it

Finishing up: some pros/cons

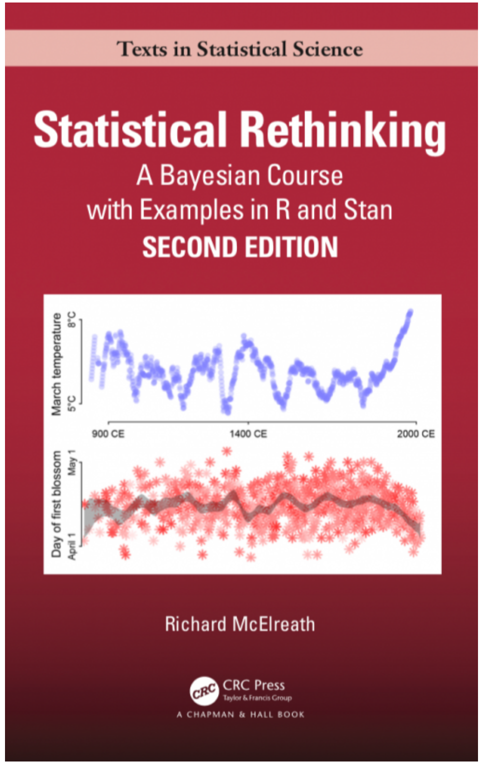
Some pros:

- Very flexible
- Can borrow information in various ways
  - From prior
  - From similar situations (Hospital A similar to Hospital B?)
- I specified all parts of the model
  - I can (maybe) defend it
  - You can more easily critique it

Some cons:

- Priors based on judgment
  - Controversial
- Can be quite technical (tools are improving)
- More work than doing a quick t-test (tools are improving)

Finishing up: due credit



Flowchart from hell, marbles example & figures, weight/height data all from here.

Very pedagogical. Heartily recommended.

Thank you!