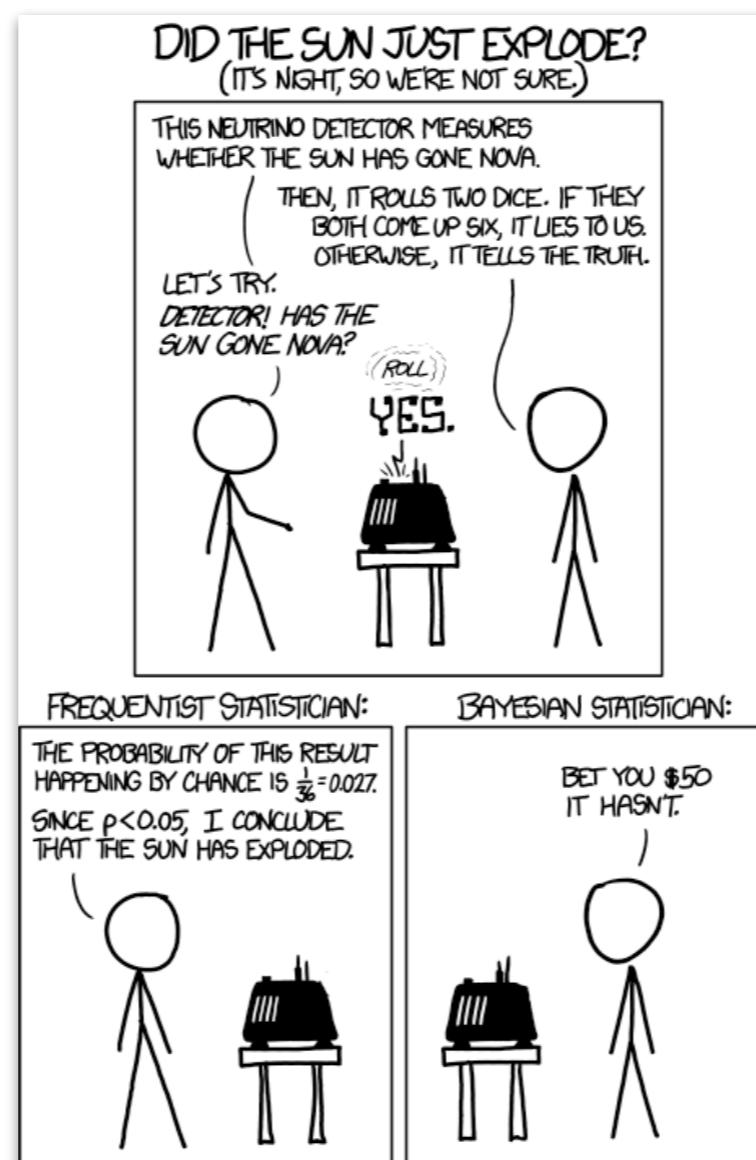


Bayesian data analysis 1



03/04/2024

DID THE SUN JUST EXPLODE?

(IT'S NIGHT, SO WE'RE NOT SURE.)

THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.

THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.

LET'S TRY.

DETECTOR! HAS THE SUN GONE NOVA?

ROLL

YES.



FREQUENTIST STATISTICIAN:

THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS $\frac{1}{36} = 0.027$. SINCE $p < 0.05$, I CONCLUDE THAT THE SUN HAS EXPLODED.



BAYESIAN STATISTICIAN:

BET YOU \$50 IT HASN'T.



Feedback

Your feedback

I feel like I get logistic regressions!

Logistics

Final presentations

Final presentation

Thanks for filling out this survey to help us with planning!

How are you planning to present? *

In class (preferred option if possible)

Remotely (live)

I will record the presentation and submit a video before March 16th.

Other...

What's your name (e.g. Tobias Gerstenberg)? *

Short answer text

What's the name of your team's github repository (e.g. final-project-tobi)? *

Short answer text

How many people are in your team (e.g. 1, 2, or 3)?

Short answer text

<https://tinyurl.com/psych252presentation24>

Plan for today

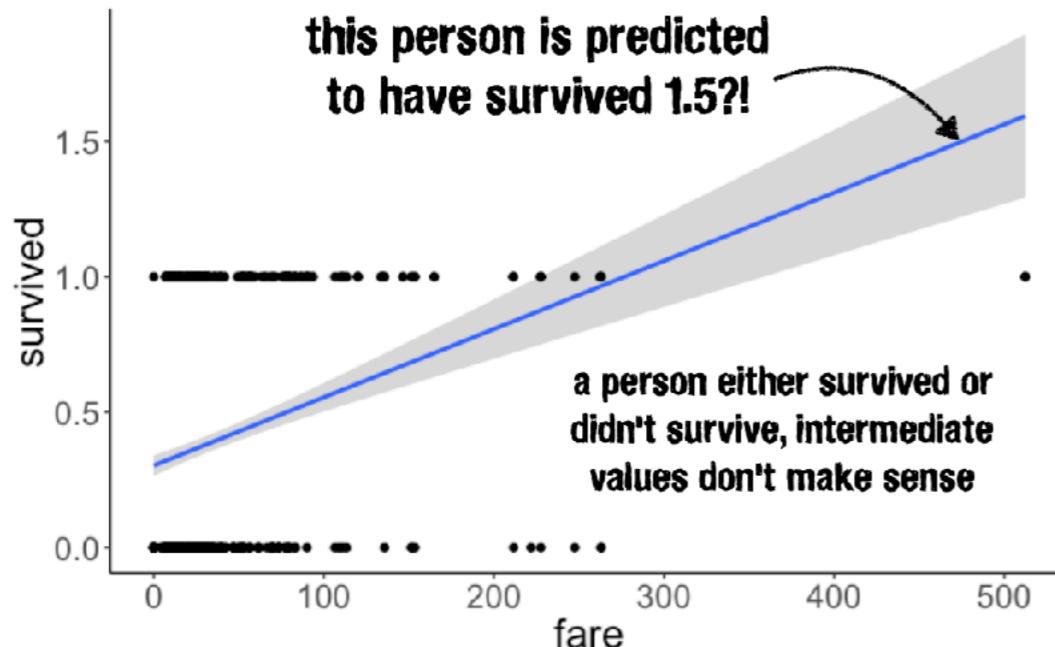
- Quick recap
- Mixed effects logistic regression
- Bayesian data analysis
 - Comparison between frequentist and Bayesian data analysis
 - Quick flash from the past
 - Flipping coins
 - What affects the posterior?
 - Ingredients: likelihood, prior, inference
 - Doing Bayesian data analysis

Quick recap

Quick recap: Logistic regression

Is there a relationship between fare and survived?

```
fit.lm = lm(formula = survived ~ 1 + fare, data = df.titanic)
```



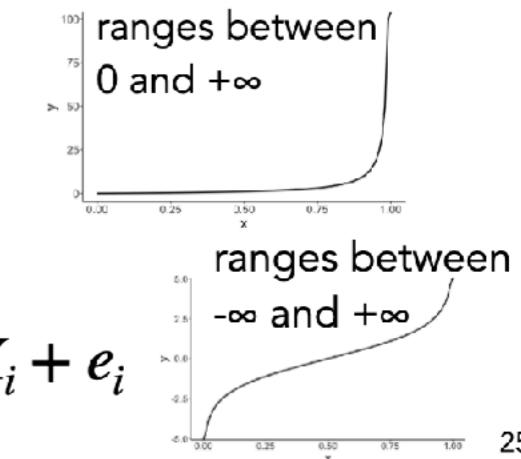
22

Logit transform

$$\pi_i = b_0 + b_1 \cdot X_i + e_i$$

predict the probability of Y

we need to transform the dependent variable so that it can take any value between $-\infty$ and $+\infty$ (we can then transform it back into a probability later)



25

Fitting a logistic regression in R

```
1 fit.glm = glm(formula = survived ~ 1 + fare,
2                      family = "binomial",
3                      data = df.titanic)
4
5 fit.glm %>% summary()
```

```
Call:
glm(formula = survived ~ 1 + fare, family = "binomial", data = df.titanic)

Deviance Residuals:
    Min      10     Median      30      Max 
-2.4906 -0.8878 -0.8531  1.3429  1.5942 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) -0.941330  0.095129 -9.895 < 2e-16 ***
fare        0.015197  0.002232  6.810 9.79e-12 ***
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

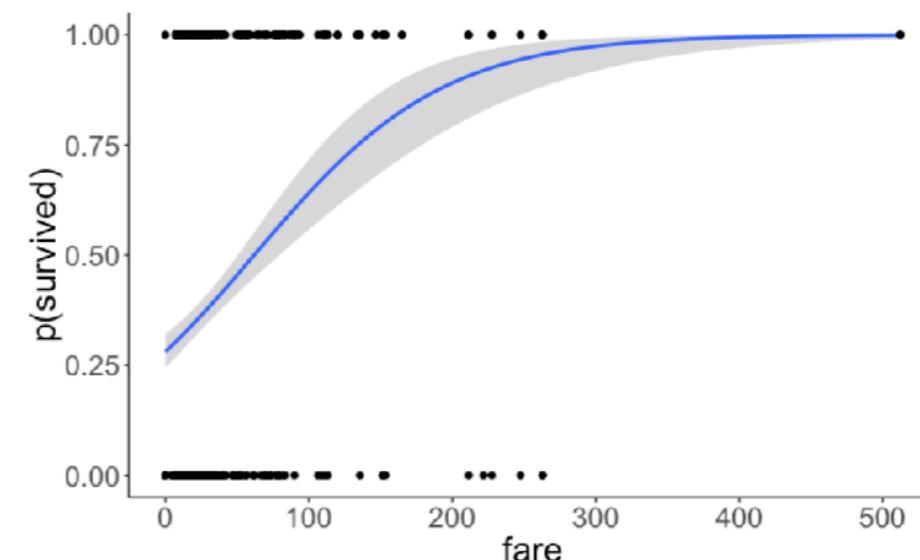
Null deviance: 1186.7 on 890 degrees of freedom
Residual deviance: 1117.6 on 889 degrees of freedom
AIC: 1121.6

Number of Fisher Scoring iterations: 4
```

27

Visualize the model's predictions

```
1 ggplot(data = df.titanic,
2           mapping = aes(x = fare,
3                             y = survived)) +
4   geom_smooth(method = "glm",
5               method.args = list(family = "binomial")) +
6   geom_point() +
7   labs(y = "p(survived)")
```

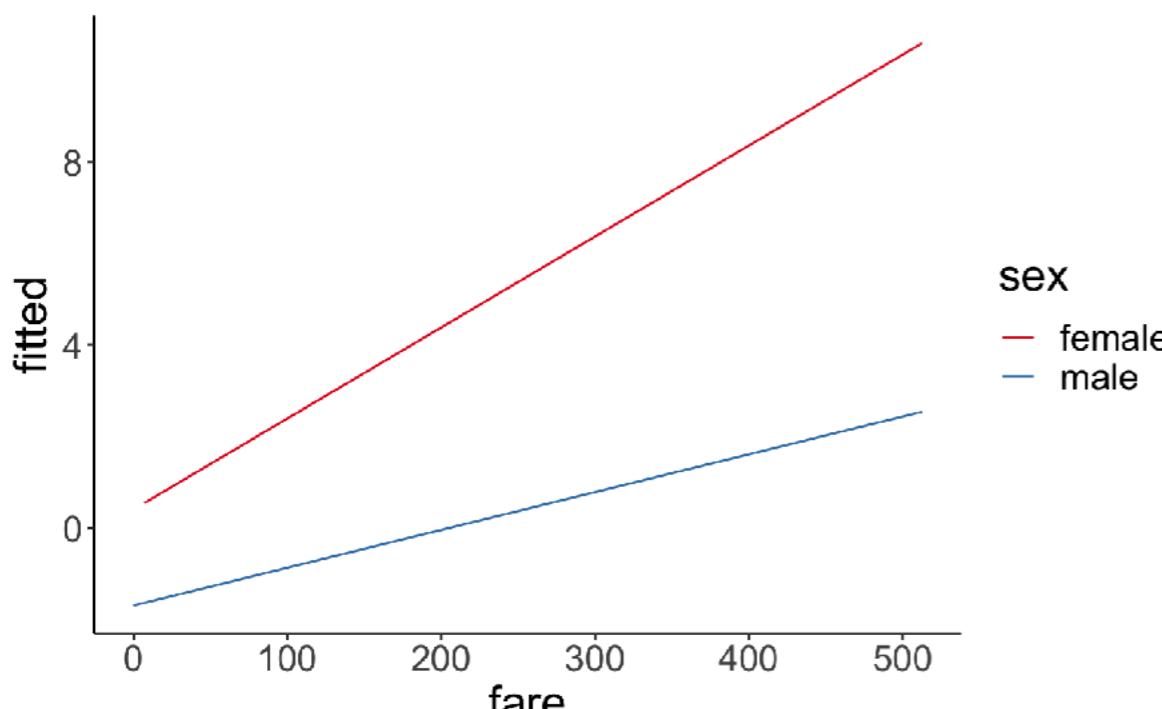


28

9

Quick recap: Logistic regression

```
1 fit.glm3 = glm(formula = survived ~ 1 + sex * fare,  
2 family = "binomial",  
3 data = df.titanic)  
4  
5 fit.glm3 %>%  
6 summary()
```



lines in log odds
space

Call:
`glm(formula = survived ~ 1 + sex * fare, family = "binomial", data = df.titanic)`

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.6280	-0.6279	-0.5991	0.8172	1.9288

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.408428	0.189999	2.150	0.031584 *
sexmale	-2.099345	0.230291	-9.116	< 2e-16 ***
fare	0.019878	0.005372	3.701	0.000215 ***
sexmale:fare	-0.011617	0.005934	-1.958	0.050252 .

Signif. codes:	0 ***	0.001 **	0.01 *	0.05 .
	0.1 ' '	1		

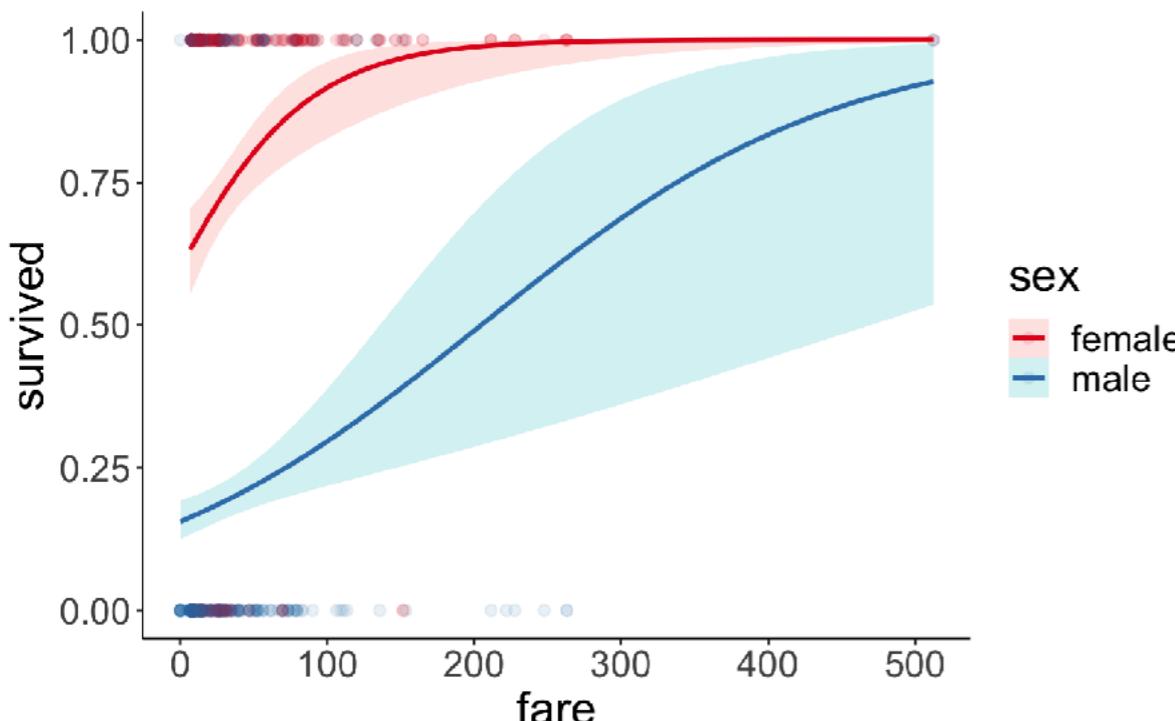
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1186.66 on 890 degrees of freedom
Residual deviance: 879.85 on 887 degrees of freedom
AIC: 887.85

Number of Fisher Scoring iterations: 5

Quick recap: Logistic regression

```
1 fit.glm3 = glm(formula = survived ~ 1 + sex * fare,  
2 family = "binomial",  
3 data = df.titanic)  
4  
5 fit.glm3 %>%  
6 summary()
```



after inverse logit transformation

Call:
`glm(formula = survived ~ 1 + sex * fare, family = "binomial", data = df.titanic)`

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-2.6280	-0.6279	-0.5991	0.8172	1.9288

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.408428	0.189999	2.150	0.031584 *
sexmale	-2.099345	0.230291	-9.116	< 2e-16 ***
fare	0.019878	0.005372	3.701	0.000215 ***
sexmale:fare	-0.011617	0.005934	-1.958	0.050252 .

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1186.66 on 890 degrees of freedom
Residual deviance: 879.85 on 887 degrees of freedom
AIC: 887.85

Number of Fisher Scoring iterations: 5

Quick recap: Interpreting the model output

Let's consider a binary predictor

```
1 fit.glm2 = glm(formula = survived ~ sex,
2                  family = "binomial",
3                  data = df.titanic)
4
5 fit.glm2 %>% summary()
```

```
Call:
glm(formula = survived ~ sex, family = "binomial", data = df.titanic)

Deviance Residuals:
    Min      1Q  Median      3Q     Max 
-1.6462 -0.6471 -0.6471  0.7725  1.8256 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) 1.0566    0.1290   8.191 2.58e-16 ***
sexmale     -2.5137    0.1672 -15.036 < 2e-16 ***
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

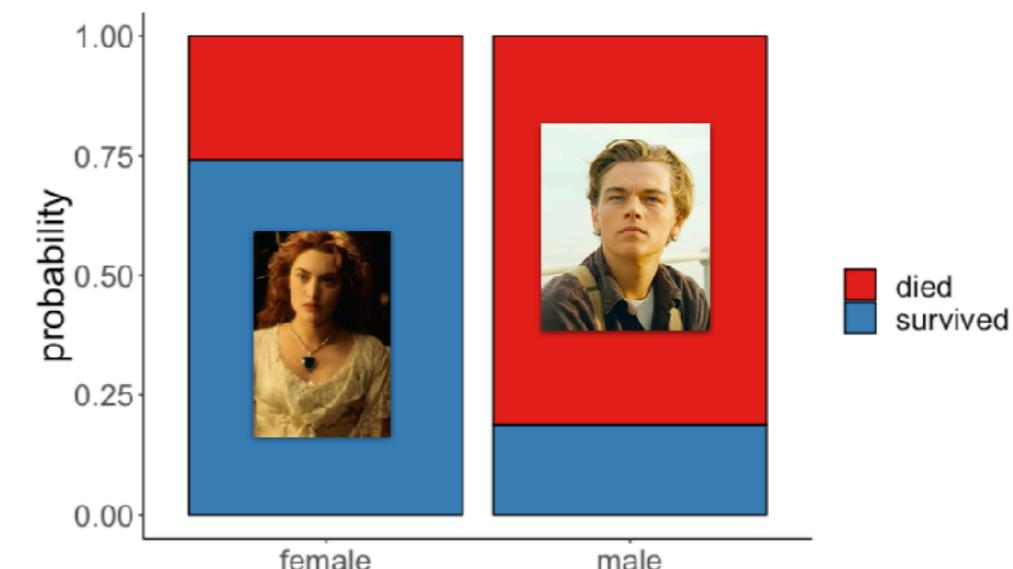
Null deviance: 1186.7 on 890 degrees of freedom
Residual deviance: 917.8 on 889 degrees of freedom
AIC: 921.8

Number of Fisher Scoring iterations: 4
```

sex was significantly associated with survival

35

Was the probability of survival different between female and male passengers on the Titanic?



Let's consider a binary predictor

$$\ln\left(\frac{p(\text{survived})_i}{1 - p(\text{survived})_i}\right) = b_0 + b_1 \cdot \text{sex}_i$$

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.0566	0.1290	8.191	2.58e-16 ***
sexmale	-2.5137	0.1672	-15.036	< 2e-16 ***

sex	survived	n	p	p(survived sex)
female	0	81	0.09	0.26
female	1	233	0.26	0.74
male	0	468	0.53	0.81
male	1	109	0.12	0.19

if sex == 0:

$$\ln\left(\frac{\widehat{p(\text{survived})}_i}{1 - \widehat{p(\text{survived})}_i}\right) = b_0$$

$$p(\text{survived})_i = \frac{e^{b_0}}{1 + e^{b_0}} = 0.74$$

36

Now let's go back to a continuous predictor

$$\ln\left(\frac{p(\text{survived})_i}{1 - p(\text{survived})_i}\right) = b_0 + b_1 \cdot \text{fare}_i$$

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.941330	0.095129	-9.895	< 2e-16 ***
fare	0.015197	0.002232	6.810	9.79e-12 ***

fare	prediction	p(survival)
0	-0.94	0.28
10	-0.79	0.31
50	-0.18	0.45
100	0.58	0.64
500	6.66	1.00

$$\ln\left(\frac{\widehat{p(\text{survived})}}{1 - \widehat{p(\text{survived})}}\right) = -0.94 + 0.015 \cdot 10$$

$$p(\text{survived})_i = \frac{e^{-0.94+0.015 \cdot 10}}{1 + e^{-0.94+0.015 \cdot 10}} = 0.31$$

38

12

Quick recap: Fitting and reporting models

Simulating a logistic regression

```

1 # make example reproducible
2 set.seed(1)
3
4 # set parameters
5 sample_size = 1000
6 b0 = 0
7 b1 = 1
8
9 # generate data
10 df.data = tibble(
11   x = rnorm(n = sample_size),
12   y = b0 + b1 * x,
13   p = inv.logit(y) %>%
14     mutate(response = rbinom(n(), size = 1, p = p))
15
16 # fit model
17 fit = glm(formula = response ~ 1 + x,
18            family = "binomial",
19            data = df.data)
20
21 # model summary
22 fit %>% summary()

```

set some parameters

linear model (y is in log odds)

transform into probability

randomly draw response

fit a logistic regression

summarize the result

45

Assessing the model fit

$$\text{log-likelihood} = \sum_{i=1}^n [Y_i \cdot \ln(P(Y_i)) + (1 - Y_i) \cdot \ln(1 - P(Y_i))]$$

actual value ↘

predicted value ↘

- calculate the probability of the observed response
- take the log of these probabilities
- sum them up to get the log-likelihood of the data (given the model)

response	p(Y = 1)	p(Y = response)	log(p(Y = response))
1	0.34	0.34	-1.07
0	0.53	0.47	-0.75
1	0.30	0.30	-1.20
1	0.81	0.81	-0.22
1	0.56	0.56	-0.58
0	0.30	0.70	-0.36
1	0.60	0.60	-0.52
1	0.65	0.65	-0.43
1	0.62	0.62	-0.48
0	0.41	0.59	-0.54

Testing hypotheses

```

joint_tests()

anova(test = "LRT")

```

not quite the same value

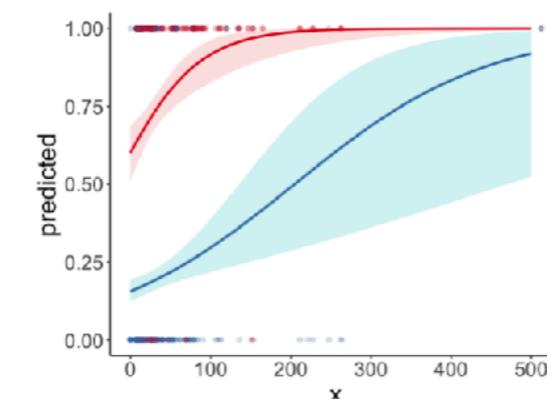
47

<https://stats.stackexchange.com/questions/400101/using-emmeans-with-clmm-to-look-at-joint-effects>

51

Reporting results

- Visualize the data
- Show a table with the regression results
- Report significance of different factors
- Interpreting parameter estimates is tricky -- probably best to report probabilities for a few example cases



Predicted values of survived
sex = male
x Predicted SE 95% CI
0 0.15 0.13 [0.51, 0.69]
100 0.30 0.21 [0.83, 0.96]
200 0.39 0.95 [0.93, 1.00]
300 1.00 1.48 [0.97, 1.00]
400 1.00 2.02 [0.93, 1.00]
500 1.00 2.55 [1.00, 1.00]

sex = female
x Predicted SE 95% CI
0 0.15 0.13 [0.12, 0.19]
100 0.30 0.21 [0.22, 0.39]
200 0.49 0.24 [0.23, 0.70]
300 0.69 0.63 [0.25, 0.90]
400 0.83 0.94 [0.44, 0.97]
500 0.92 1.13 [0.57, 0.99]



13

Assumptions

- linearity (between predictors and log odds)
- independence
- no multi-collinearity
- model fails to converge when there is **complete separation**:
 - if outcome variable can be perfectly predicted by a (combination of) predictor(s)

Different kinds of generalized models

Different linking functions

```
binomial(link = "logit")  
  
gaussian(link = "identity")  
  
Gamma(link = "inverse")  
  
inverse.gaussian(link = "1/mu^2")  
  
poisson(link = "log")  
  
quasi(link = "identity", variance = "constant")  
  
quasibinomial(link = "logit")  
  
quasipoisson(link = "log")
```

**apply different transformations to the
dependent variable**

Mixed effects logistic regression

Mixed effects logistic regression

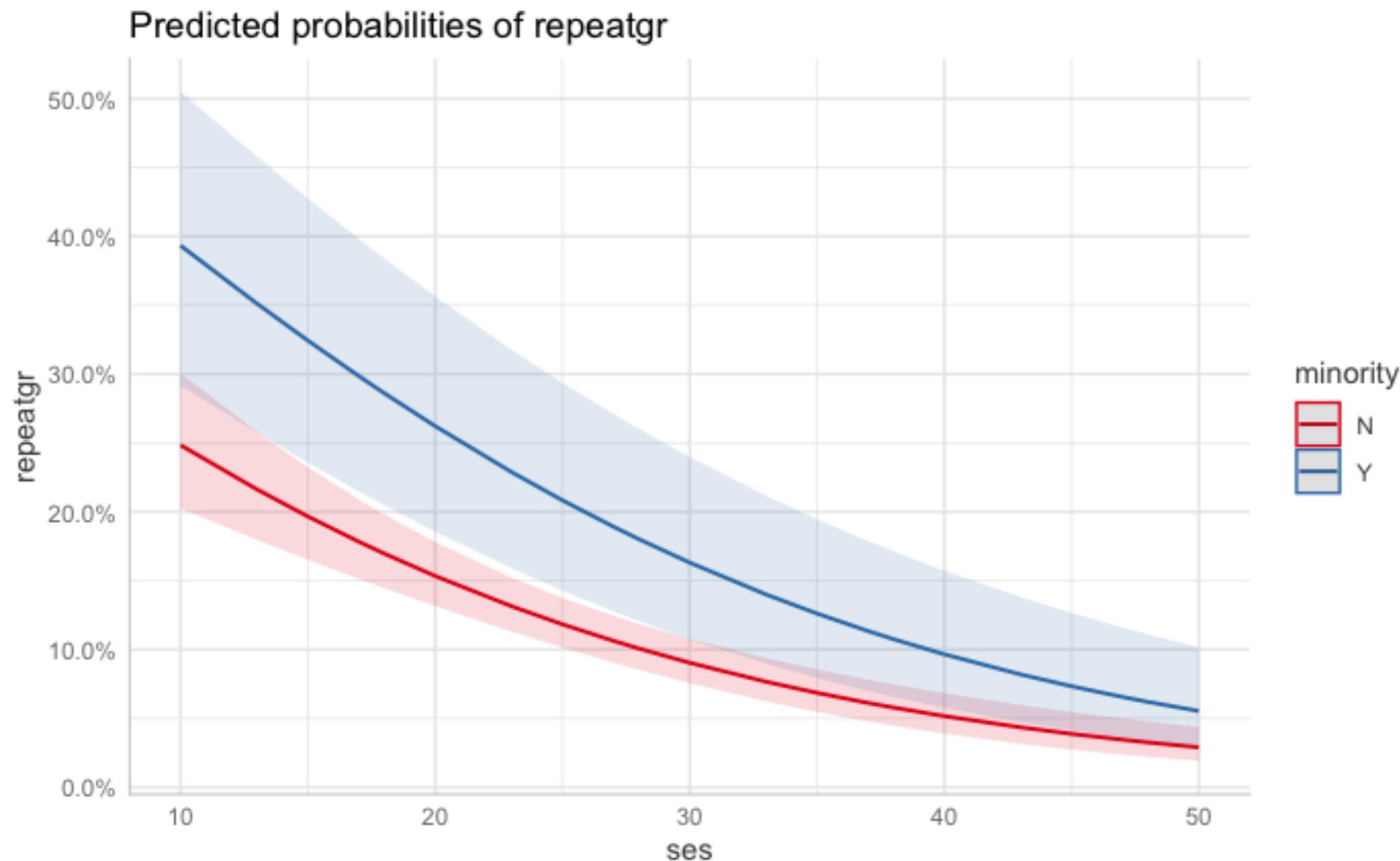
repeated a grade: yes / no

```
1 fit = glmer(repeatgr ~ 1 + ses * Minority + (1 | schoolNR),  
2               data = df.language,  
3               family = "binomial")  
4  
5 fit %>% summary()
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']  
Family: binomial ( logit )  
Formula: repeatgr ~ 1 + ses + minority + (1 | school_nr)  
Data: df.language  
  
AIC      BIC      logLik deviance df.resid  
1659.1  1682.1   -825.6    1651.1     2279  
  
Scaled residuals:  
    Min      1Q  Median      3Q      Max  
-0.9235 -0.4045 -0.3150 -0.2249  5.8372  
  
Random effects:  
 Groups   Name        Variance Std.Dev.  
 school_nr (Intercept) 0.2489   0.4989  
Number of obs: 2283, groups: school_nr, 131  
  
Fixed effects:  
            Estimate Std. Error z value Pr(>|z|)  
(Intercept) -0.506291  0.197570 -2.563  0.01039 *  
ses         -0.060086  0.007524 -7.986 1.39e-15 ***  
minorityY    0.673612  0.238660  2.822  0.00477 **  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Correlation of Fixed Effects:  
          (Intr) ses  
ses       -0.898  
minorityY -0.308  0.208
```

Mixed effects logistic regression

```
1 ggpredict(model = fit,  
2            terms = c("ses [all]", "minority")) %>%  
3   plot()
```



Hypothesis test: anova (test = "LRT")

```
1 fit_a = glmer(repeatgr ~ 1 + ses + minority + (1 | school_nr),  
2                 data = df.language,  
3                 family = "binomial")  
4  
5 # dropping ses as a predictor  
6 fit_c = glmer(repeatgr ~ 1 + minority + (1 | school_nr),  
7                 data = df.language,  
8                 family = "binomial")  
9  
10 anova(fit_a, fit_c, test = "LRT")
```

```
Data: df.language  
Models:  
fit_c: repeatgr ~ 1 + minority + (1 | school_nr)  
fit_a: repeatgr ~ 1 + ses + minority + (1 | school_nr)  
      npar   AIC   BIC  logLik deviance    Chisq Df Pr(>Chisq)  
fit_c     3 1732.5 1749.7 -863.27    1726.5  
fit_a     4 1659.1 1682.1 -825.57    1651.1 75.395  1 < 2.2e-16 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Bayesian Data Analysis

Datacamp course

recommended!!

The screenshot shows a DataCamp course page. At the top left, it says "INTERACTIVE COURSE". The main title is "Fundamentals of Bayesian Data Analysis in R". Below the title is a button labeled "Replay Course". To the right is a circular icon containing a bar chart and the text "FUNDAMENTALS OF BAYESIAN DATA ANALYSIS". Below the title, course details are listed: "4 hours", "23 Videos", "58 Exercises", "6,177 Participants", and "4,450 XP".

Course Description

Bayesian data analysis is an approach to statistical modeling and machine learning that is becoming more and more popular. It provides a uniform framework to build problem specific models that can be used for both statistical inference and for prediction. This course will introduce you to Bayesian data analysis: What it is, how it works, and why it is a useful tool to have in your data science toolbox.



<https://www.datacamp.com/courses/fundamentals-of-bayesian-data-analysis-in-r>

Great online book

An Introduction to Data Analysis

Michael Franke

last rendered at: 2023-02-08 14:03:16



II Data

3 Data, variables & experimental desi...

4 Data Wrangling

5 Summary statistics

6 Data Visualization

III Bayesian Data Analysis

7 Basics of Probability Theory

8 Statistical models

9 Bayesian parameter estimation

10 Model Comparison

11 Bayesian hypothesis testing

IV Applied (generalized) linear mod...

12 Linear regression

13 Bayesian regression in practice

14 Categorical predictors

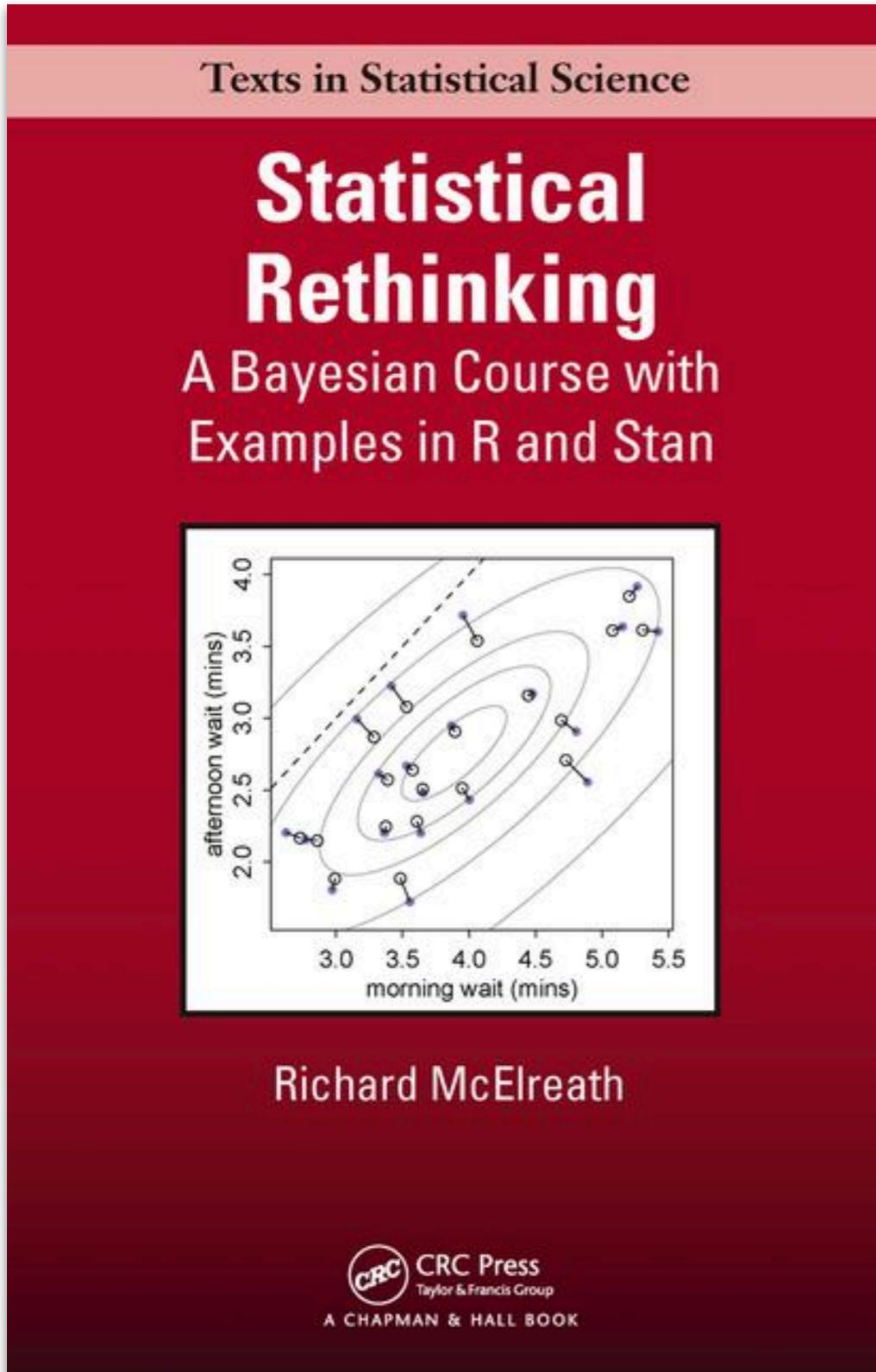
15 Generalized linear model

V Frequentist statistics

16 Null Hypothesis Significance Testing

17 Comparing frequentist and Bayesi...

Great book on Bayesian data analysis



- nice hands-on book (which uses R throughout)
- rewrite of all the code with "tidyverse" and "BRMS" is here: <https://bookdown.org/content/4857/>
- video lectures are available here: https://github.com/rmcelreath/stat_rethinking_2024

video
lectures
available

DAGs

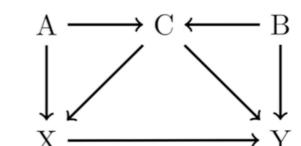
Different queries, different models

Which control variables?

Absolute not safe to add everything — **bad controls**

How to test/refine the causal model?

DAGs are intuition pumps: get head out of data, into science



Comparison between frequentist and Bayesian data analysis

Goal of data analysis: Inference about the world

Frequentist statistics

- generate a sampling distribution of the test statistic assuming H_0
- compare observed value of the test statistic with the sampling distribution
- reject the H_0 if probability of observed value (or more extreme values) is less than α

Bayesian statistics

- directly test hypotheses of interest
- define prior over hypotheses $p(H)$
- compute likelihood of the data for each hypothesis $p(D|H)$
- use Bayes' rule to infer the posterior over hypotheses given the data $p(H|D)$

Objections to frequentist NHST



null hypothesis
significance testing

- p-value is not a measure of evidential support
 - becomes smaller as N increases
- results are often misinterpreted (both p-values and confidence intervals are not particularly intuitive)
- what we want to know: $p(\text{Hypothesis} \mid \text{Data})$
- what we calculate: $p(\text{Data} \mid \text{Null Hypothesis})$

Frequentists vs. Bayesian

- both "want" to evaluate the evidence for a hypothesis using a sample of data $p(H|D)$
- it's often easier to calculate the inverse: the probability of the data given a hypothesis $p(D|H)$
- frequentists use a rule of thumb (p-value) to make a decision
- Bayesians use Bayes' rule

Why don't more people use Bayesian Statistics?

- supposedly more difficult
 - relies on the logic of probability theory
- reliance on a *prior*
- reliance on computing and simulation
 - we can't just use SPSS
 - but we can use JASP (Just Another Statistics Program)

and we've already learned
how to simulate and
visualize data in this class!



What are (some of) the benefits of Bayesian data analysis?

- intuitive model testing and comparison
- straightforward interpretation of results
 - Bayesian credible intervals vs. Confidence intervals
- more model flexibility
 - adequately express assumptions about the data-generating process
- better predictions!

Flash from the past

Clue guide to probability

$$p(B|A) = \frac{p(A|B) \cdot p(B)}{p(A)}$$

we derived this using the definition of conditional probability

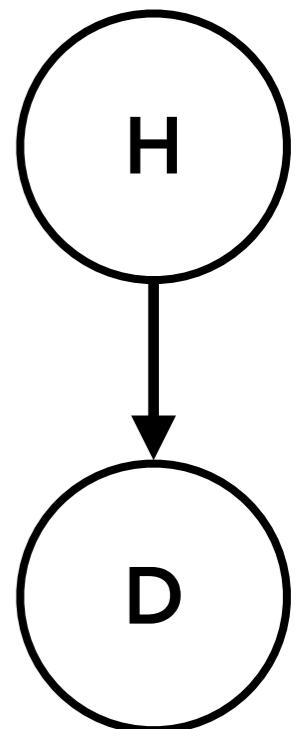
posterior

$$p(H|D) = \frac{\text{likelihood} \quad \text{prior}}{p(D)} \quad \frac{p(D|H) \cdot p(H)}{p(D)}$$

subjective probability interpretation

H = Hypothesis

D = Data



formal framework for learning from data

updating our prior belief $p(H)$, to a posterior belief $p(H|D)$ given some data

Clue guide to probability

A patient named Fred is tested for a disease called *conditionitis*, a medical condition that afflicts **1% of the population**. The test result is positive, i.e., the test claims that Fred has the disease. Let **D** be the event that Fred has the disease and **T** be the event that he tests positive.

Suppose that the test is “95% accurate”; there are different measures of the accuracy of a test, but in this problem it is assumed to mean that **P(T|D) = 0.95** and **P(¬T|¬D) = 0.95**. The quantity $P(T|D)$ is known as the *sensitivity* (= true positive rate) of the test, and $P(\neg T|\neg D)$ is known as the *specificity* (= true negative rate).

Find the conditional probability that Fred has *conditionitis*, given the evidence provided by the positive test result.

Clue guide to probability

what we know

$$P(D) = 0.01$$

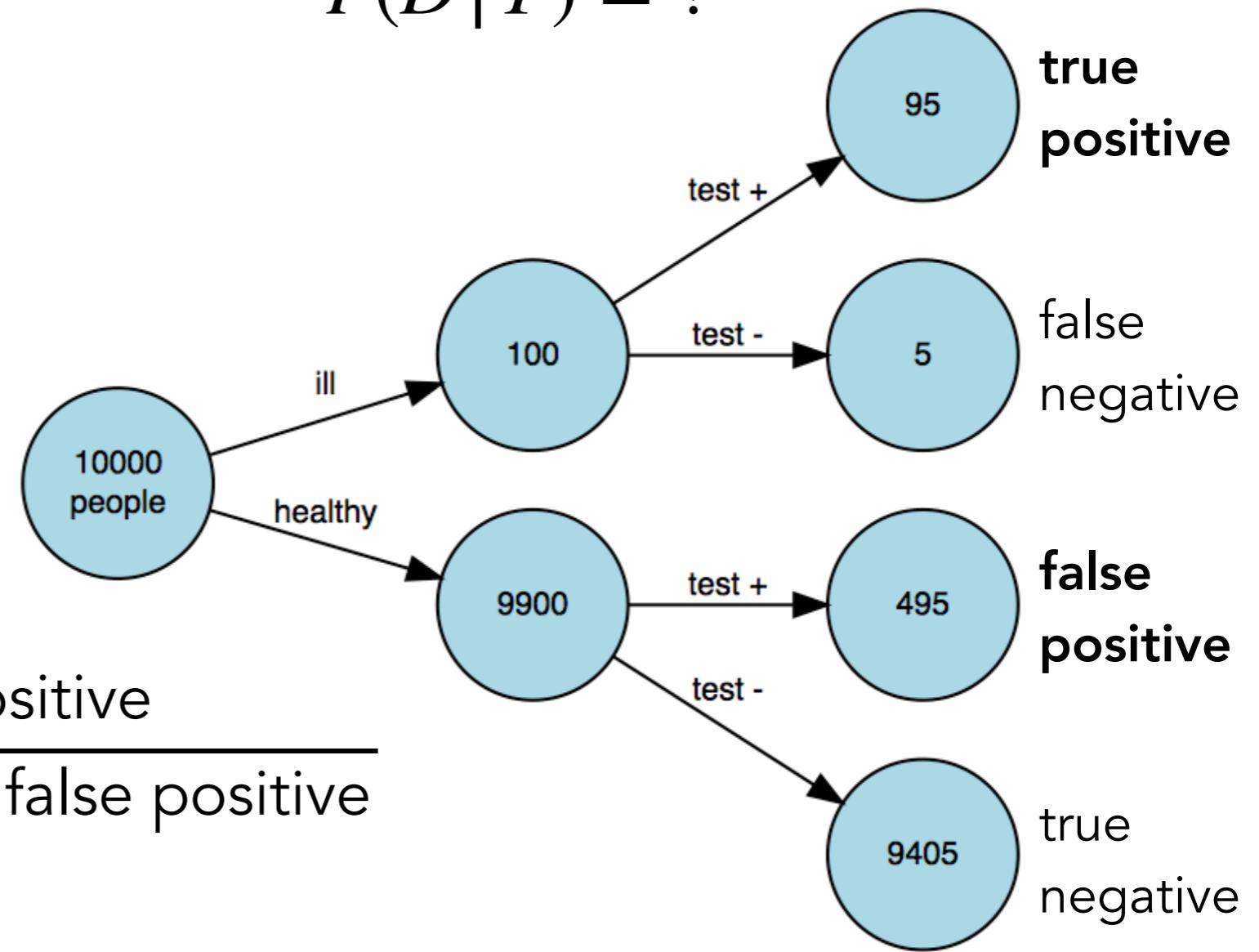
$$P(T|D) = 0.95$$

$$P(T|\neg D) = 0.05$$

$$\begin{aligned} P(D|T) &= \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \\ &= \frac{95}{95 + 495} \\ &\approx 0.16 \end{aligned}$$

what we want to know

$$P(D|T) = ?$$



Summer camp

Register now for Summer Chess Camp!

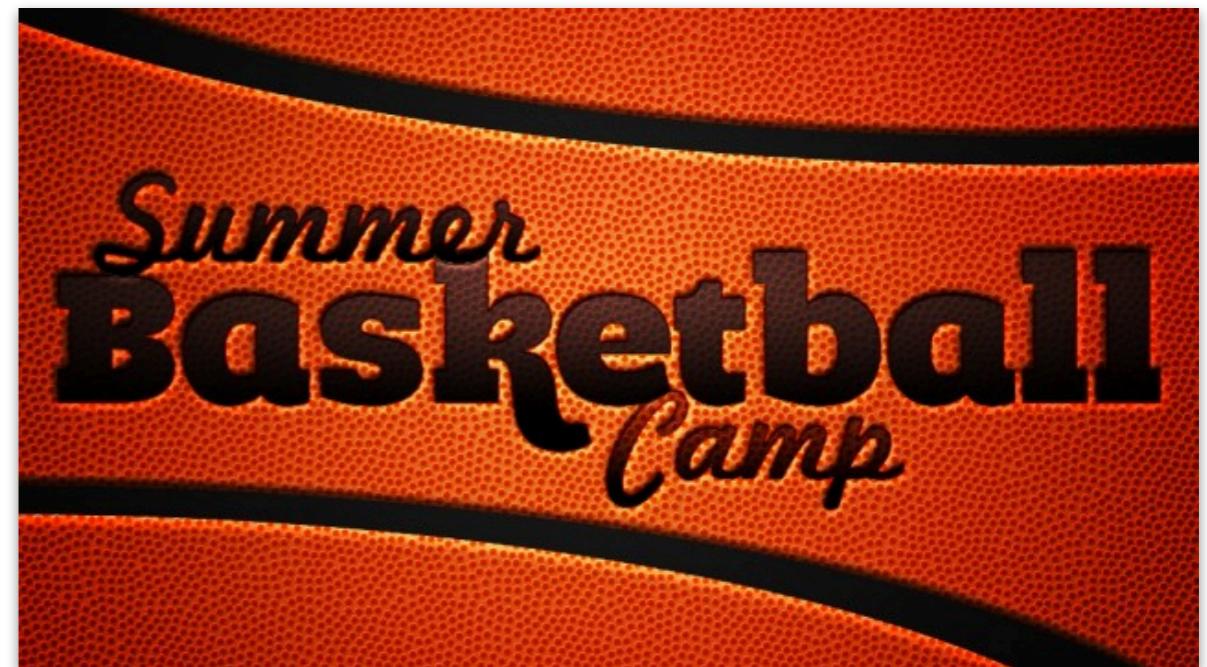


**think
Move**
CHESS ACADEMY

All skill levels
welcome!

July 23 - July 27
and
August 13 - August 17

www.thinkmovechess.com



twice as many kids go to the basketball camp

$X \sim \text{Normal}(\mu = 170, \sigma = 8)$

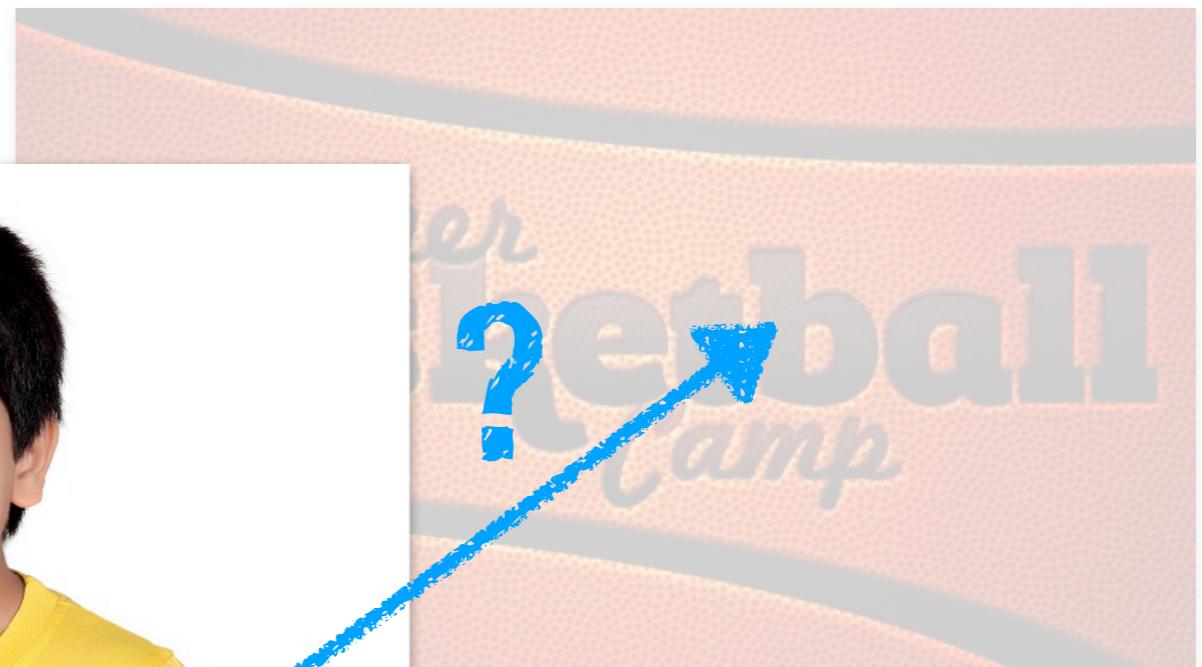
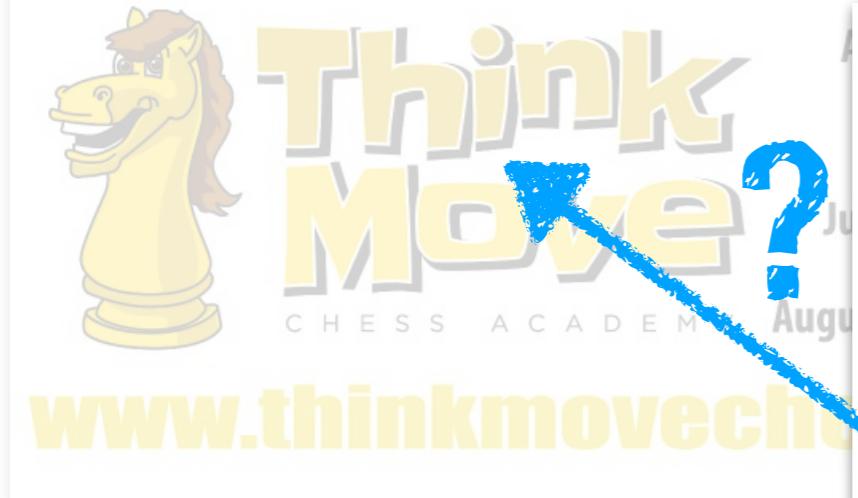


$X \sim \text{Normal}(\mu = 180, \sigma = 10)$



Summer camp

Register now for Summer Chess Camp!



twice as many

$X \sim \text{Normal}(\mu = 170, \sigma = 10)$

basketball camp

$\text{Normal}(\mu = 180, \sigma = 10)$

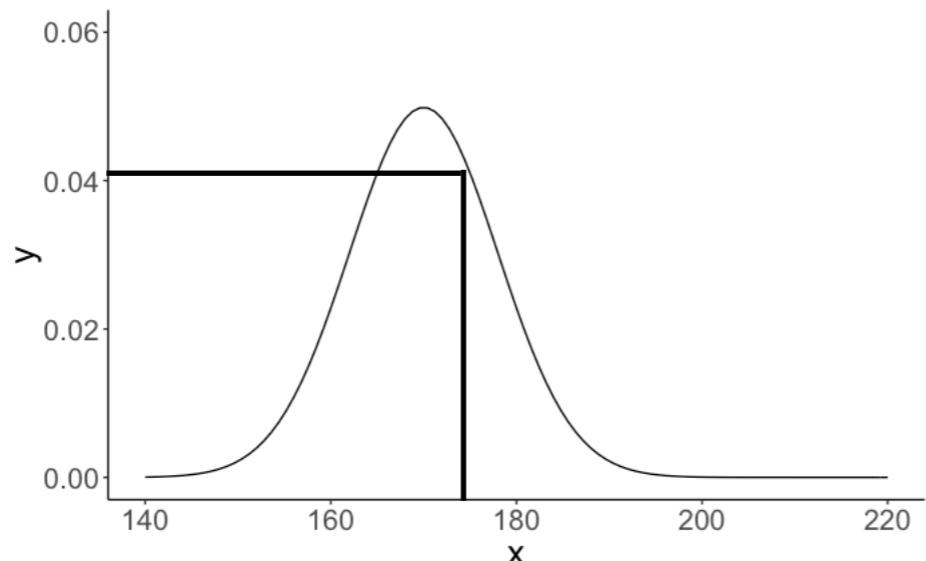
Summer camp

prior

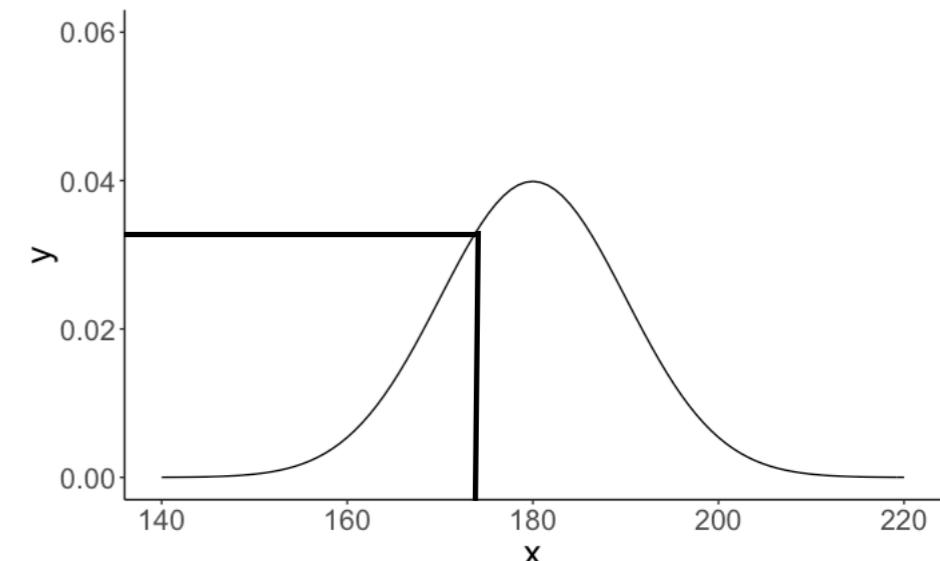
$$p(\text{chess}) = \frac{1}{3}$$

$$p(\text{basketball}) = \frac{2}{3}$$

likelihood



$$\begin{aligned} \text{dnorm}(175, \text{mean} = 170, \text{sd} = 8) \\ = 0.041 \end{aligned}$$



$$\begin{aligned} \text{dnorm}(175, \text{mean} = 180, \text{sd} = 10) \\ = 0.035 \end{aligned}$$

posterior

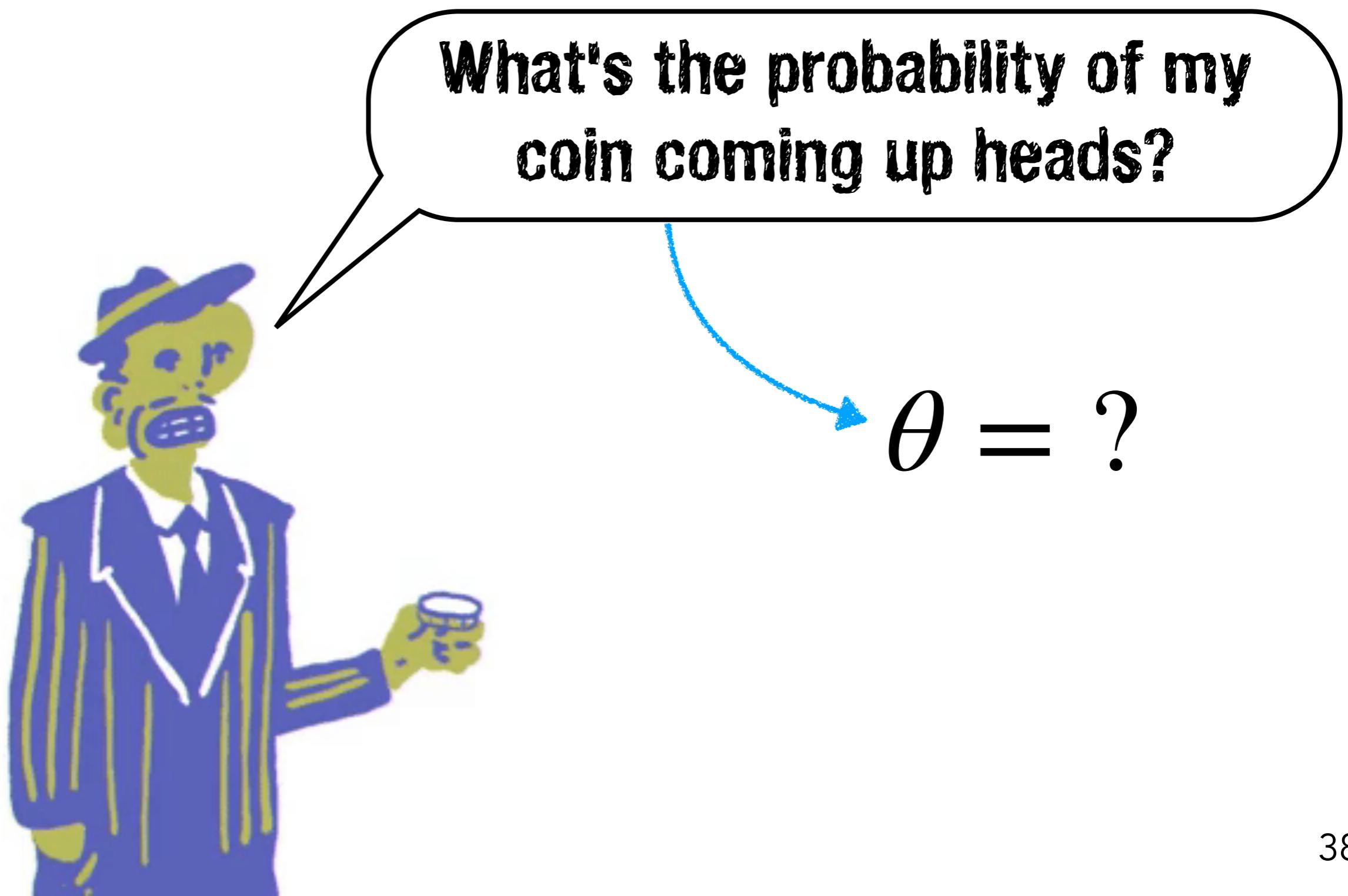
$$p(\text{basketball} | 175) = \frac{p(175 | \text{basketball}) \cdot p(\text{basketball})}{p(175 | \text{basketball}) \cdot p(\text{basketball}) + p(175 | \text{chess}) \cdot p(\text{chess})}$$

$$p(\text{basketball} | 175) = \frac{0.035 \cdot 2/3}{0.035 \cdot 2/3 + 0.041 \cdot 1/3} \approx 0.63$$

send the kid to
the basketball
gym!

Flipping coins

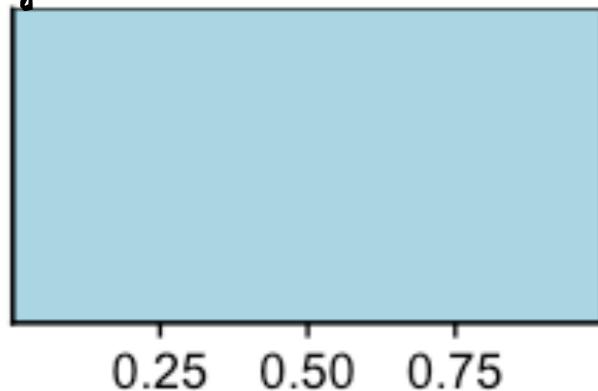
Flipping coins



Learning from data

How does/should our belief change as evidence comes in?

prior



Today's
posterior is
tomorrow's
prior.



$$p(\theta | \text{n}_{\text{success}} = 6, \text{n}_{\text{trials}} = 8)$$

Coin flip example

Which coin did I flip?

Hypotheses

$$\theta = 0.3$$



$$\theta = 0.5$$



$$\theta = 0.9$$



Data



#8 tails, #2 heads

Bayesian Recipe

- Hypotheses
- Prior over hypotheses
- Data
- Likelihood of the data given each hypothesis
- Posterior over hypotheses given the data

**+ a healthy dose
of Bayes' rule**

$$p(H|D) = \frac{p(D|H) \cdot p(H)}{p(D)}$$

Coin flip example

```
1 # data  
2 data = rep(0:1, c(8, 2)) ← 8 tails and 2 heads  
3  
4 # parameters  
5 theta = c(0.1, 0.5, 0.9) ← hypotheses  
6  
7 # prior  
8 prior = c(0.25, 0.5, 0.25) ← prior over the hypotheses  
9  
10 # likelihood  
11 likelihood = dbinom(sum(data == 1), size = length(data), prob = theta)  
12 ← binomial distribution  
13 # posterior  
14 posterior = likelihood * prior / sum(likelihood * prior)
```

$$p(H|D) = \frac{p(D|H) \cdot p(H)}{p(D)}$$

law of total probability:

$$p(D) = \sum_{i=1}^n p(D|H_i) \cdot p(H_i)$$

Coin flip example

```
1 # data  
2 data = rep(0:1, c(8, 2)) ← 8 tails and 2 heads  
3  
4 # parameters  
5 theta = c(0.1, 0.5, 0.9) ← hypotheses  
6  
7 # prior  
8 prior = c(0.25, 0.5, 0.25) ← prior over the hypotheses  
9  
10 # likelihood  
11 likelihood = dbinom(sum(data == 1), size = length(data), prob = theta)  
12 ← binomial distribution  
13 # posterior  
14 posterior = likelihood * prior / sum(likelihood * prior)
```

multiply re-normalize

theta	prior	likelihood	prior_x_like	posterior
0.1	0.25	0.19	0.0475	0.69
0.5	0.50	0.04	0.02	0.31
0.9	0.25	0.00	0.00	0.00

Coin flip example

data: #8 tails, #2 heads

Which coin was flipped?

what the
model knows
before having
seen the data



learning by
conditioning
on the data

what the
model knows
after having
seen the data

$p = 0.1$

$p = 0.5$

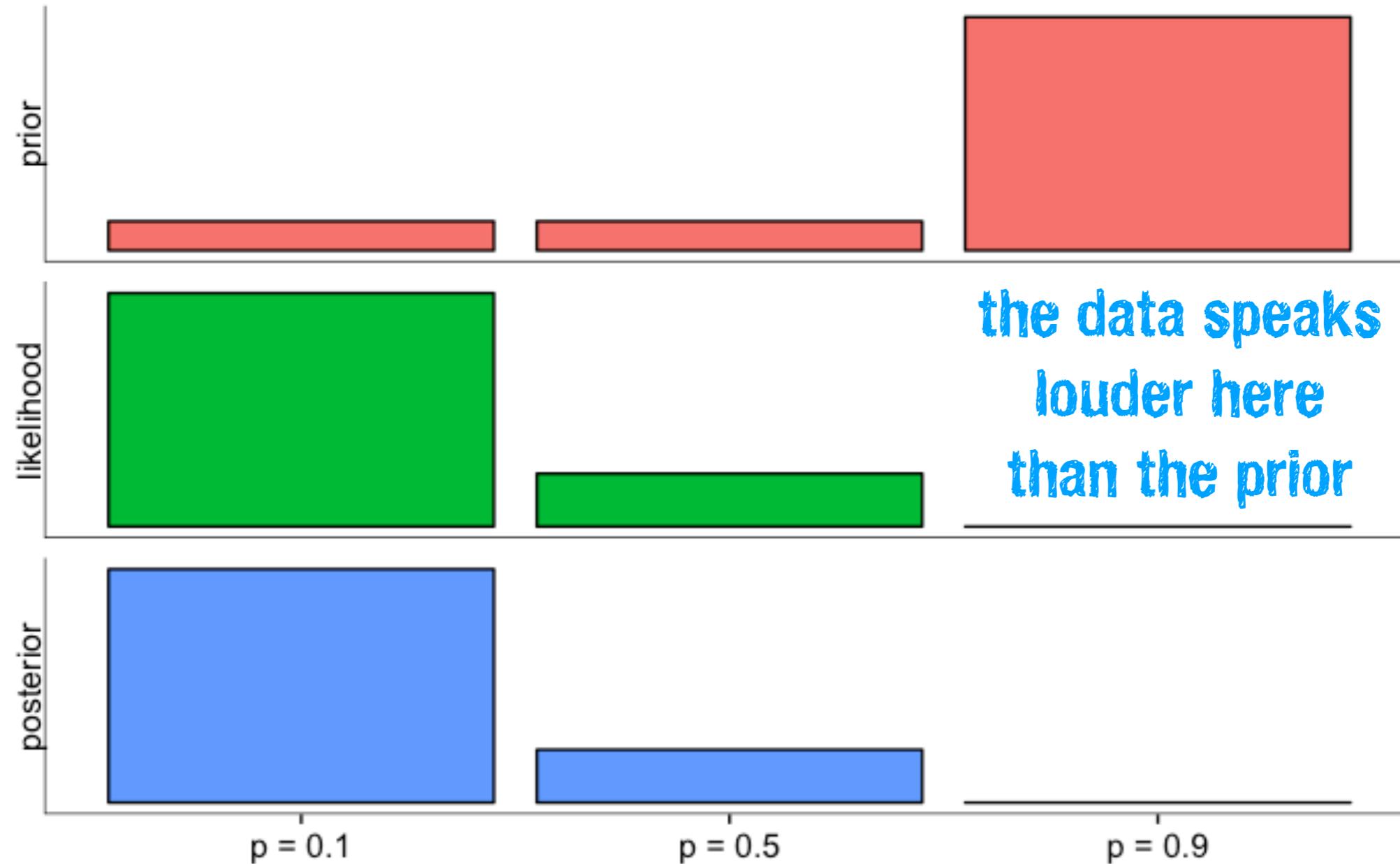
$p = 0.9$

posterior = multiplicative weighting of prior and likelihood

Coin flip example

data: #8 tails, #2 heads

Which coin was flipped?

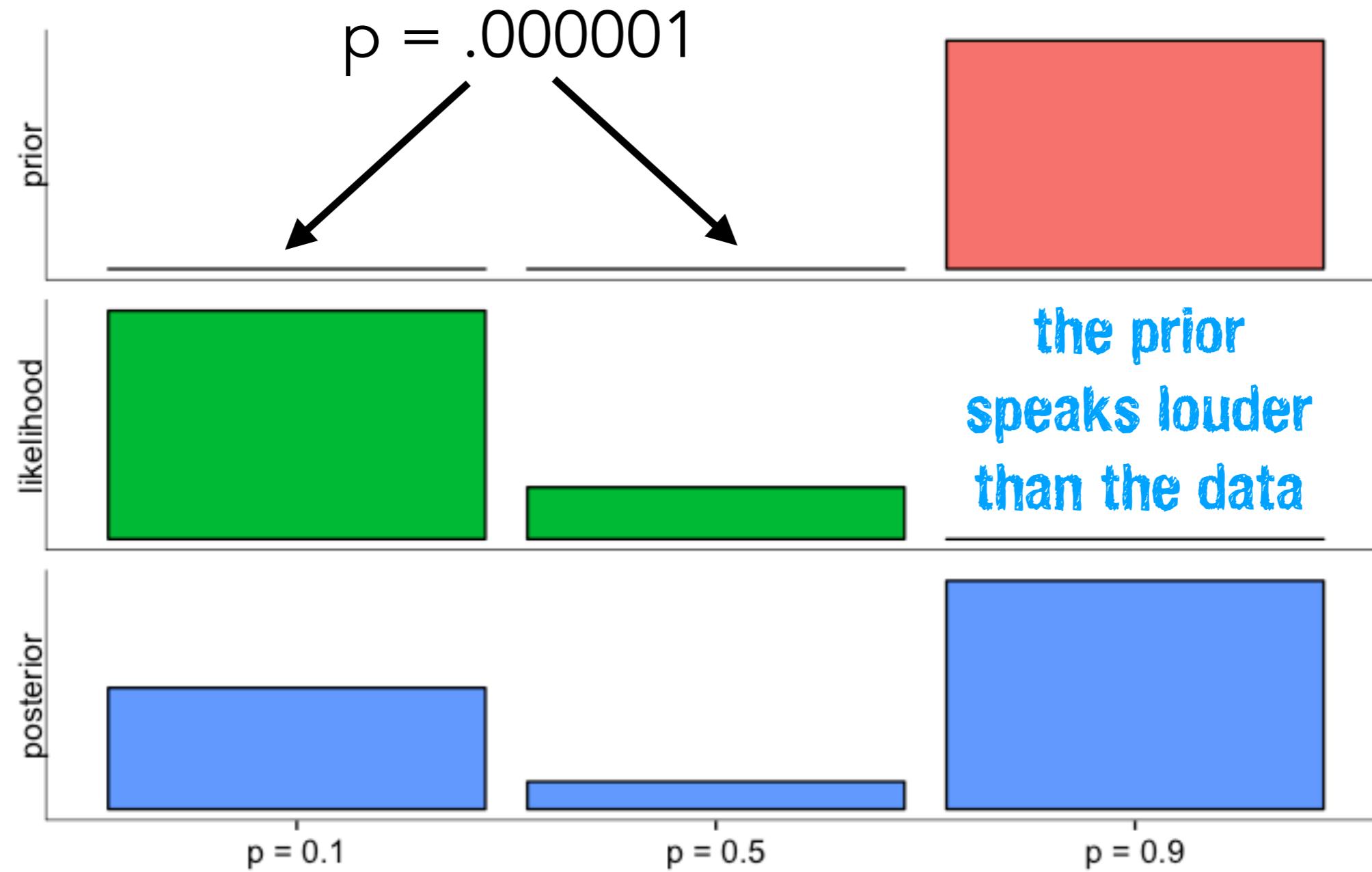


posterior = multiplicative weighting of prior and likelihood

Coin flip example

data: #8 tails, #2 heads

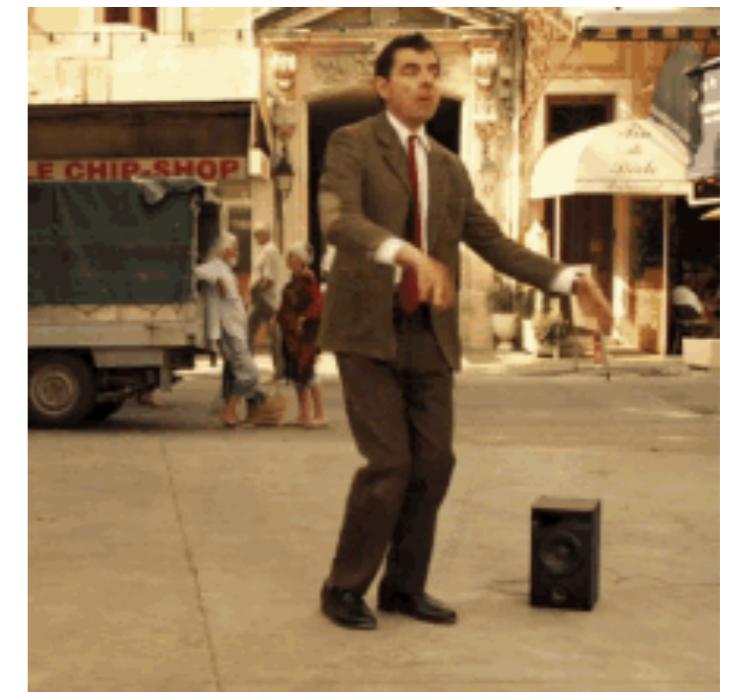
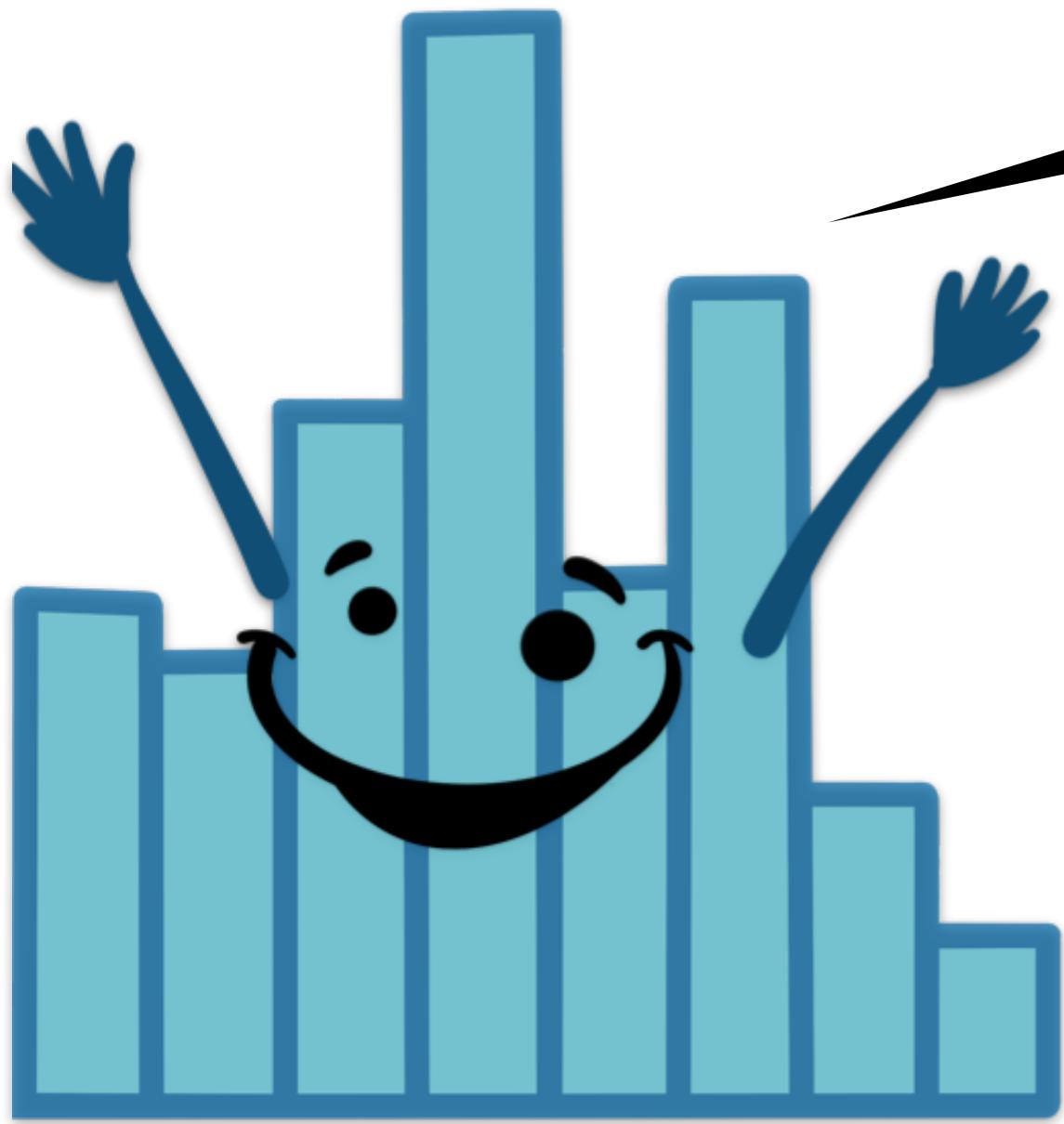
Which coin was flipped?



posterior = multiplicative weighting of prior and likelihood

02:00

stretch break!



What affects the posterior?

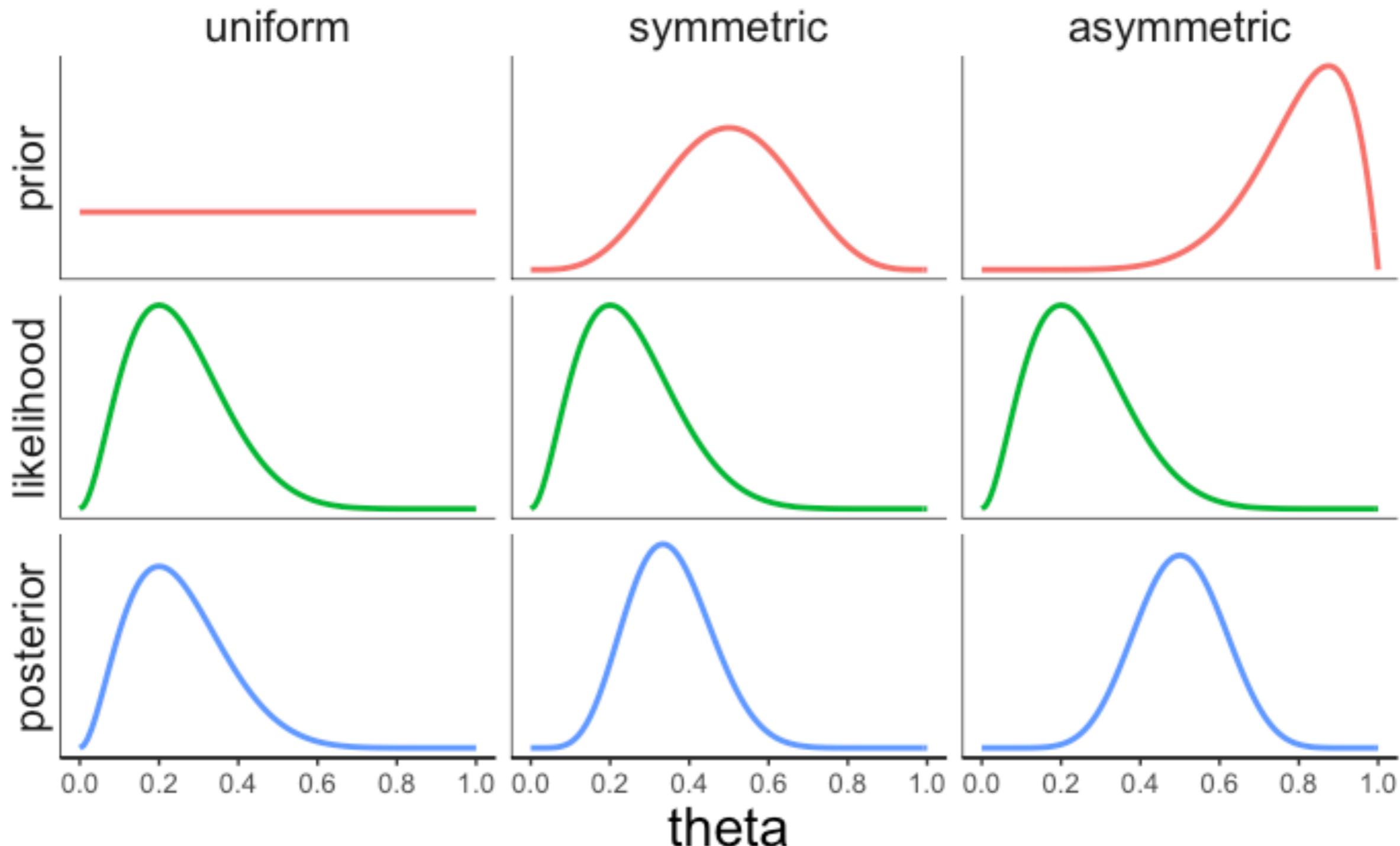
What affects the posterior?

1. the prior over hypotheses
2. the likelihood of the data given each hypothesis

$$p(H | D) = \frac{\text{Likelihood} \cdot \text{Prior}}{\text{Normalizing constant}}$$
$$p(H | D) = \frac{p(D | H) \cdot p(H)}{p(D)}$$

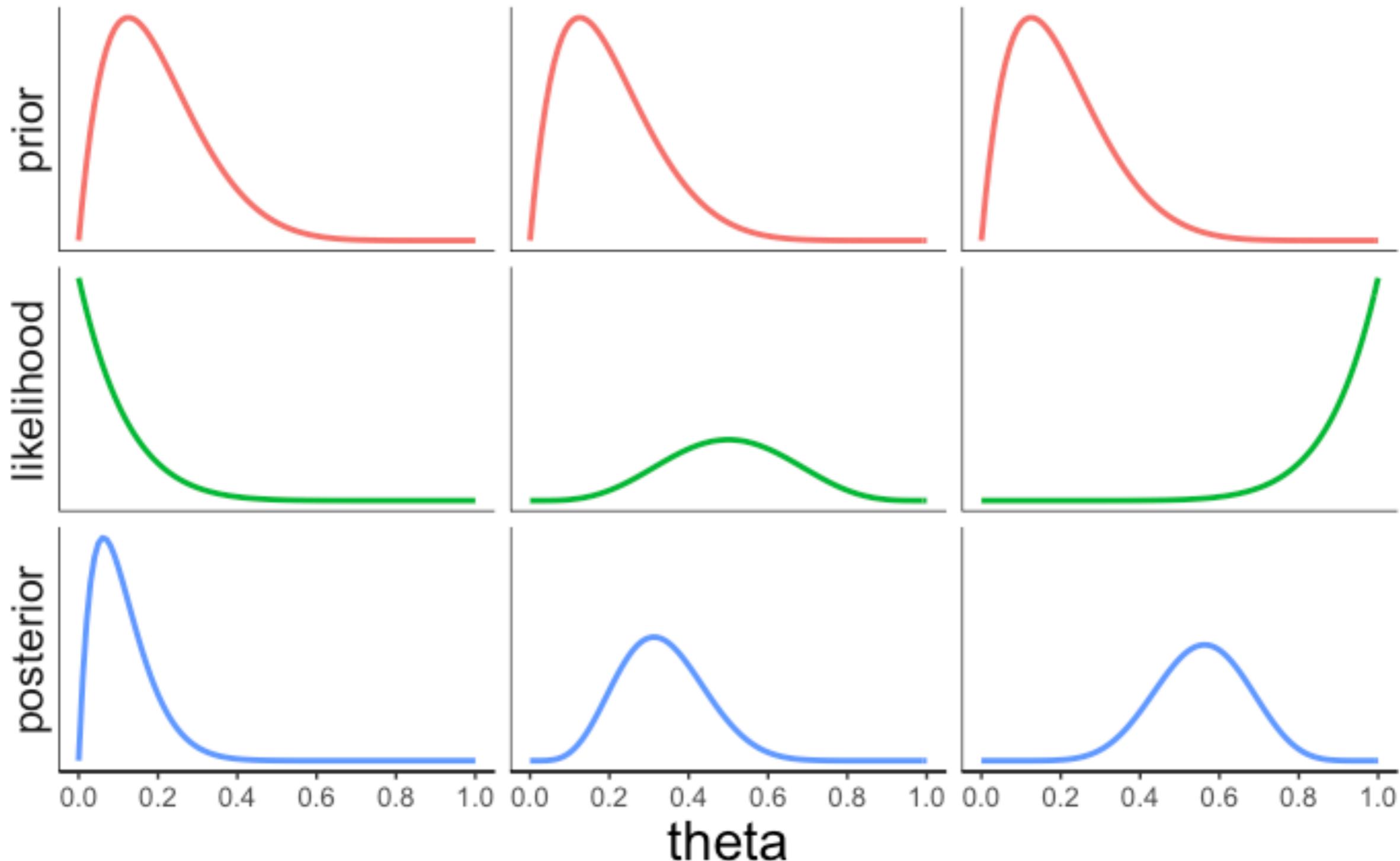
The effect of the prior

same data, different priors

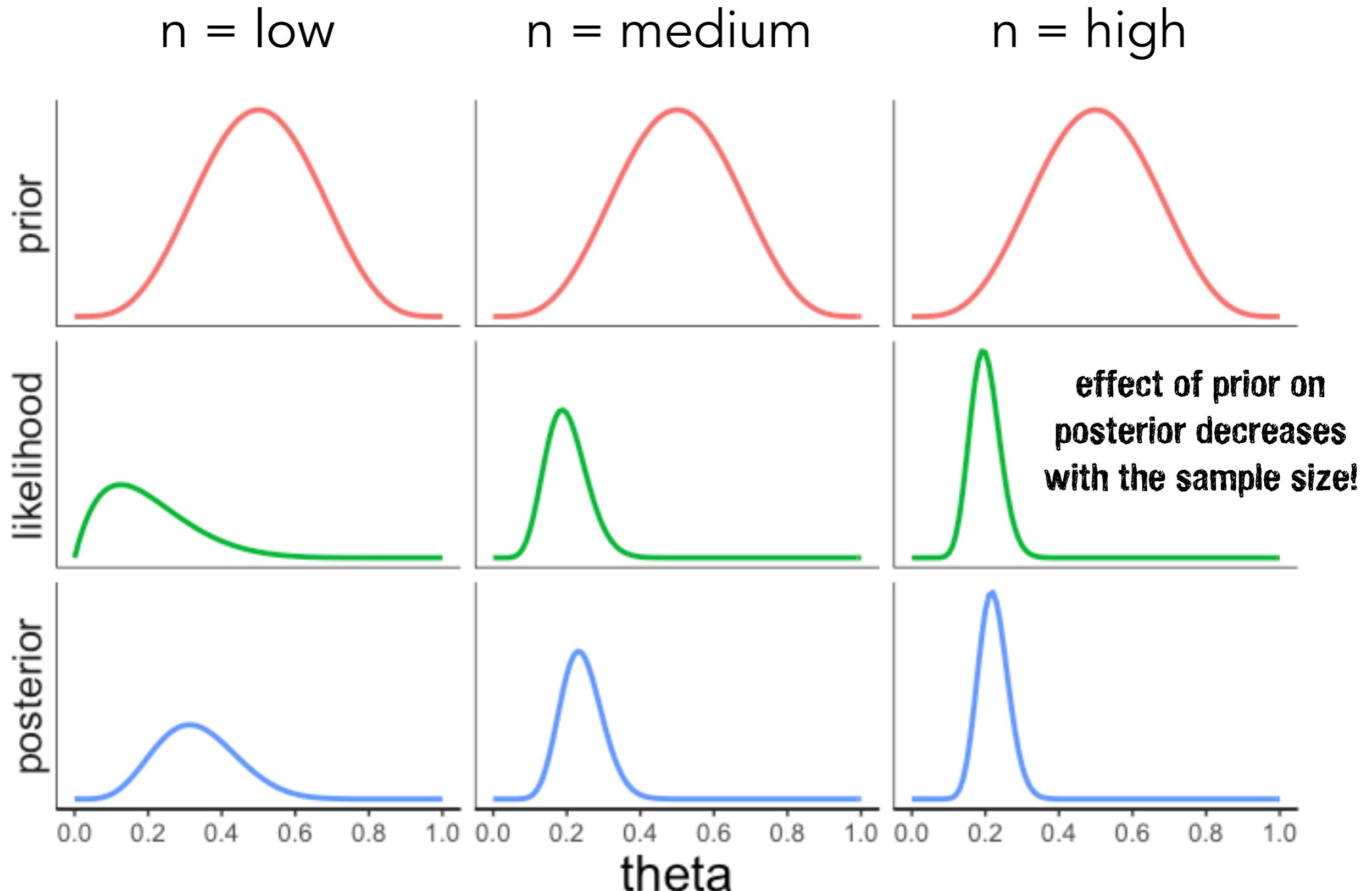


The effect of the likelihood

same prior, different data



The effect of sample size



Ingredients: likelihood, prior, inference

Ingredients

$$p(H | D) = \frac{\text{Likelihood} \quad \text{Prior}}{p(D)}$$

Posterior

Normalizing constant

$$p(H | D) = \frac{\text{Likelihood} \cdot \text{Prior}}{\text{Normalizing constant}}$$

Posterior

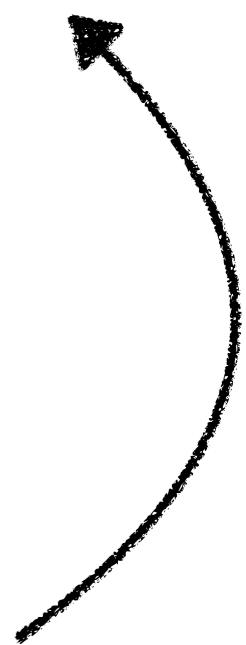
Likelihood **Prior**

$p(D | H) \cdot p(H)$

$p(D)$

Likelihood

- **What probabilistic model describes best how the data were generated?**
 - What assumptions can you make about the data?
 - What's the nature of your dependent variable (e.g. binary, ordered, continuous)?
 - Does the model re-create the behavior of interest?



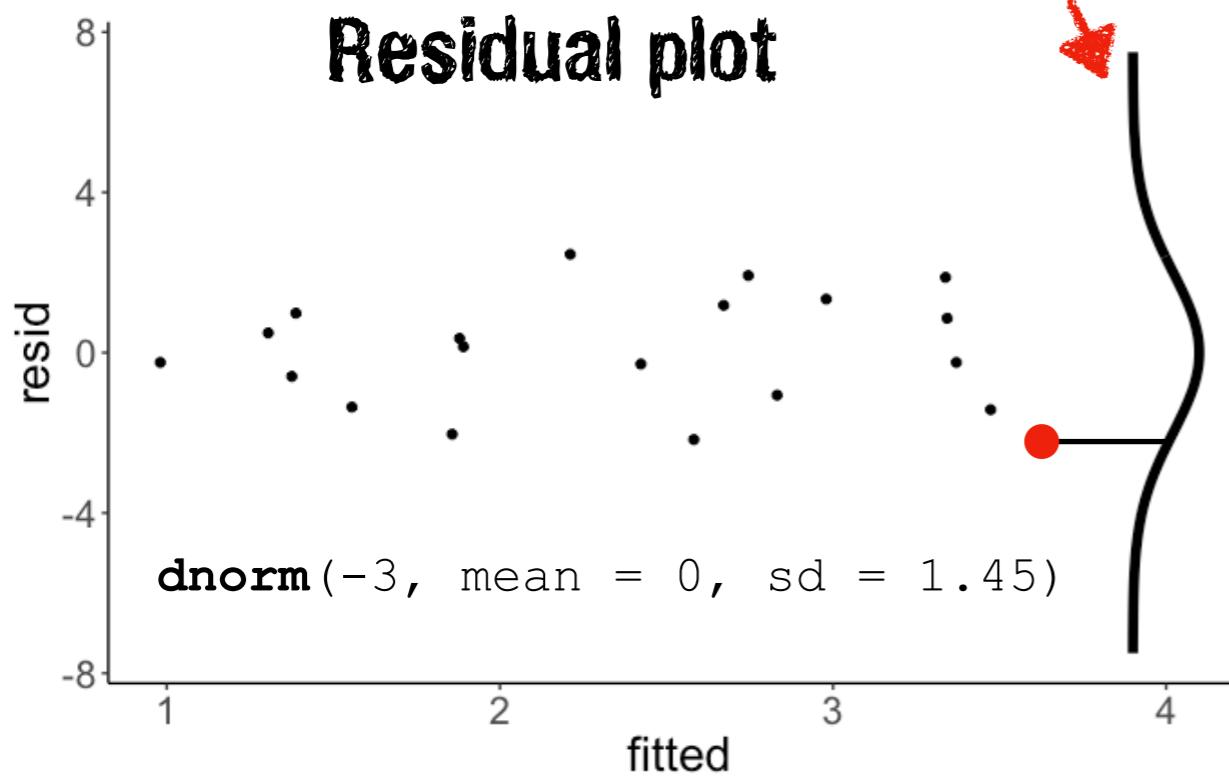
Likelihood

Gaussian distribution

$$Y_i = b_0 + b_1 \cdot x_i + e_i$$

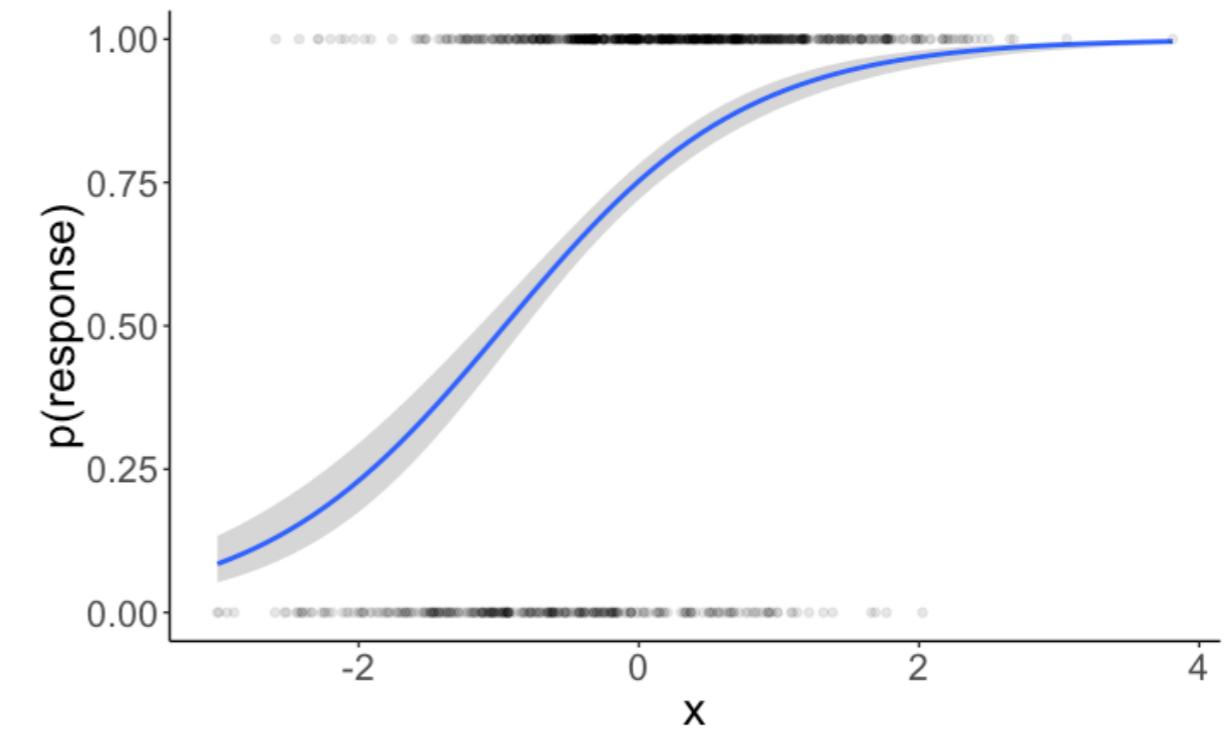
$$e_i \sim \mathcal{N}(0, \sigma)$$

Residual plot



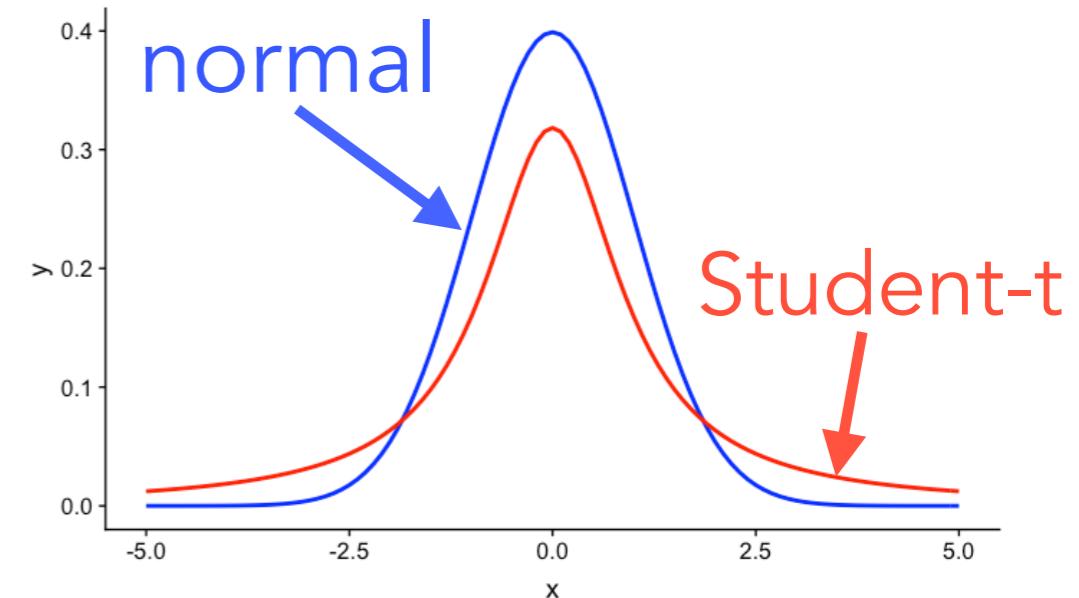
Binomial distribution

```
1 fit.glm = glm(formula = survived ~ 1 + fare,  
2 family = "binomial",  
3 data = df.titanic)
```



Likelihood

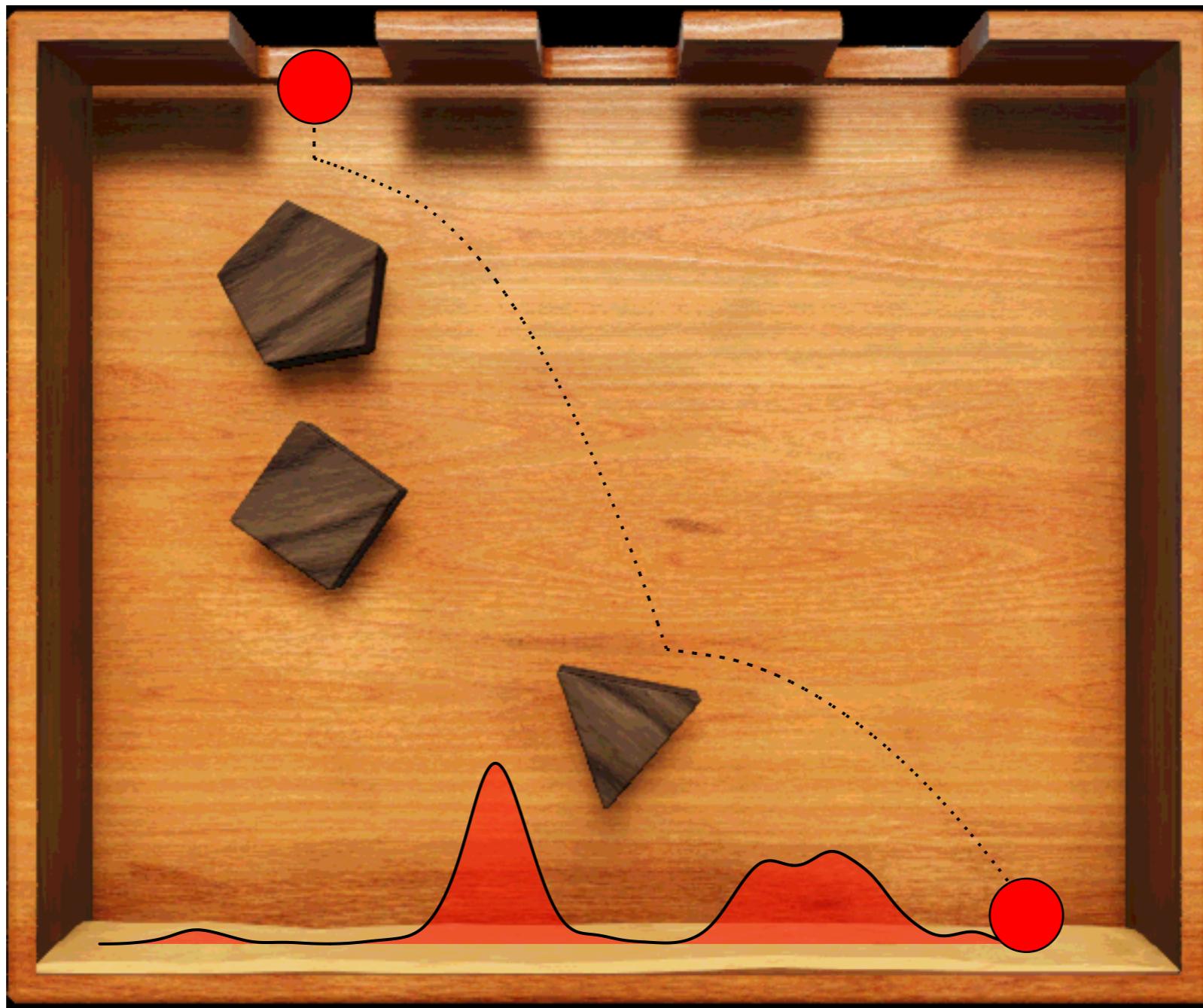
- **Bernoulli:**
 - binary data
 - a single trial
- **Poisson:** count of discrete events
- **Beta-binomial:** like binomial but probability of success may change across trials
- **Student-t:**
 - same as Normal
 - handles greater variability in the data
(distribution has **fat tails**)
- ...







Prediction: Where will the ball land?

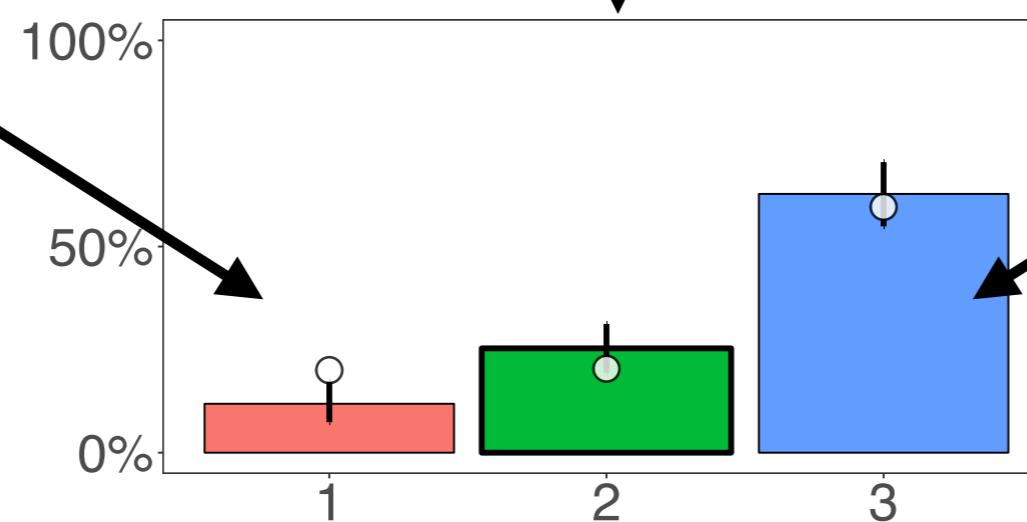
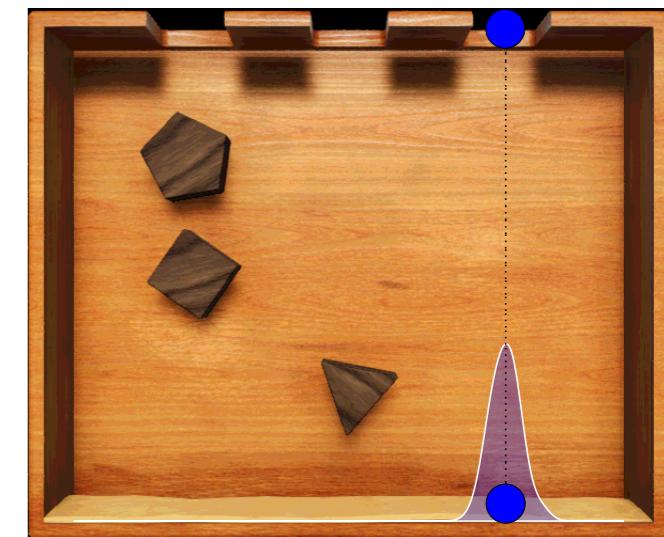
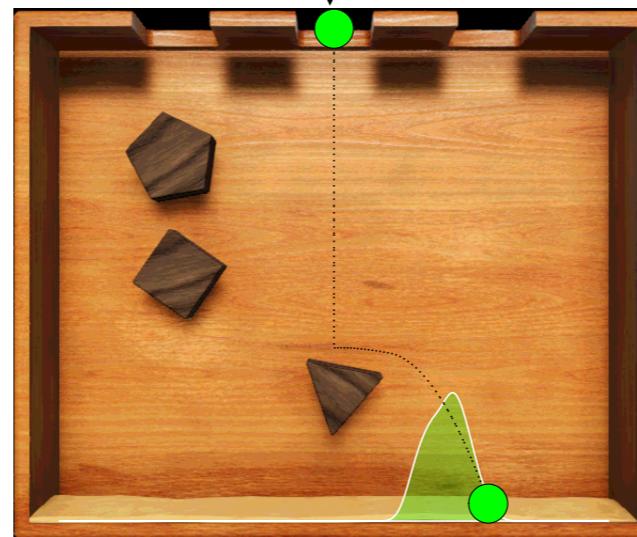
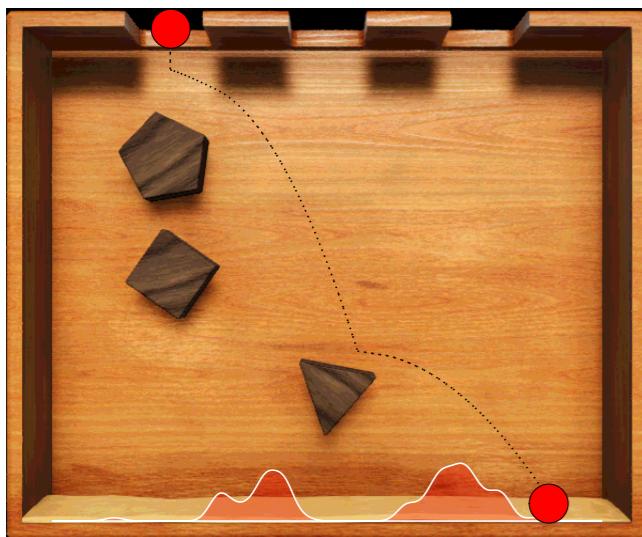
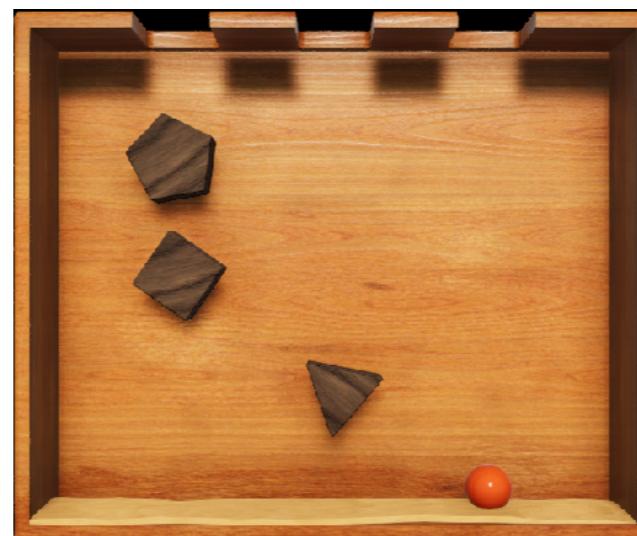


Aggregated responses

Inference: In which hole was the ball dropped?

distance between ball's true x position and x position in sample

$$\exp\left(-\frac{d(\text{ball_x}_{\text{final}}, \text{ball_x}_{\text{hole}})}{2\sigma^2}\right)$$



□ data
○ model prediction

Prior

$$p(H | D) = \frac{\text{Likelihood} \cdot \text{Prior}}{p(D)}$$

Posterior

Likelihood Prior

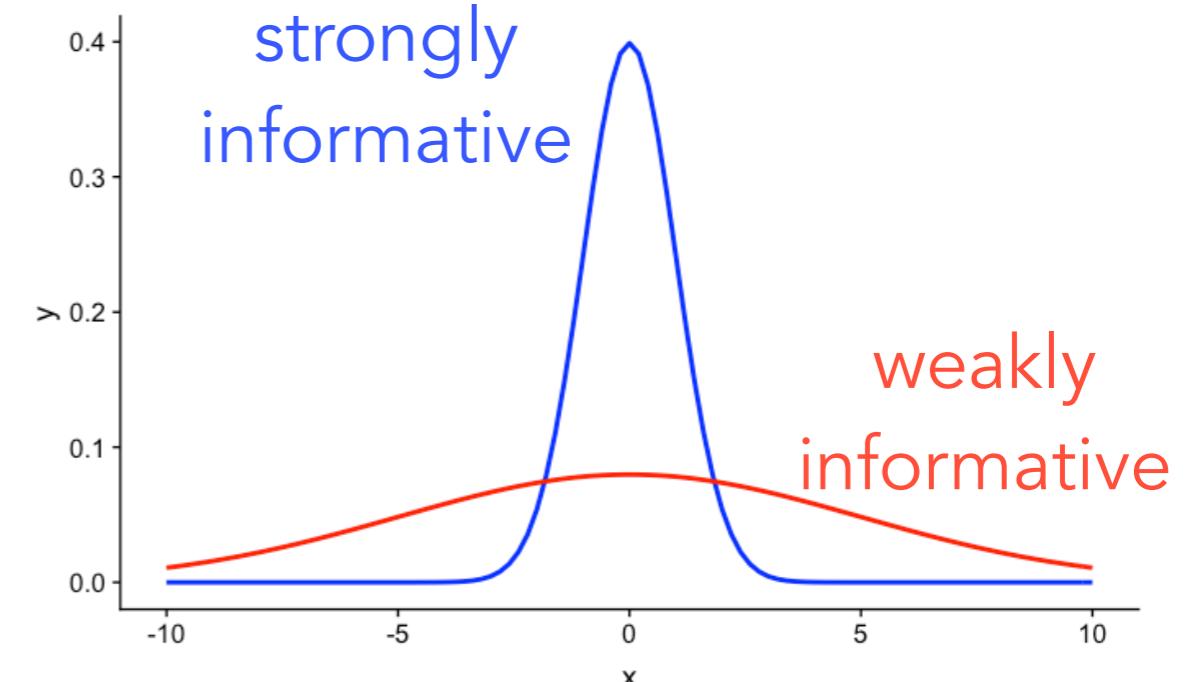
Normalizing constant

Prior

- **uniform:**

- continuous or discrete
- bounded between minimum and maximum

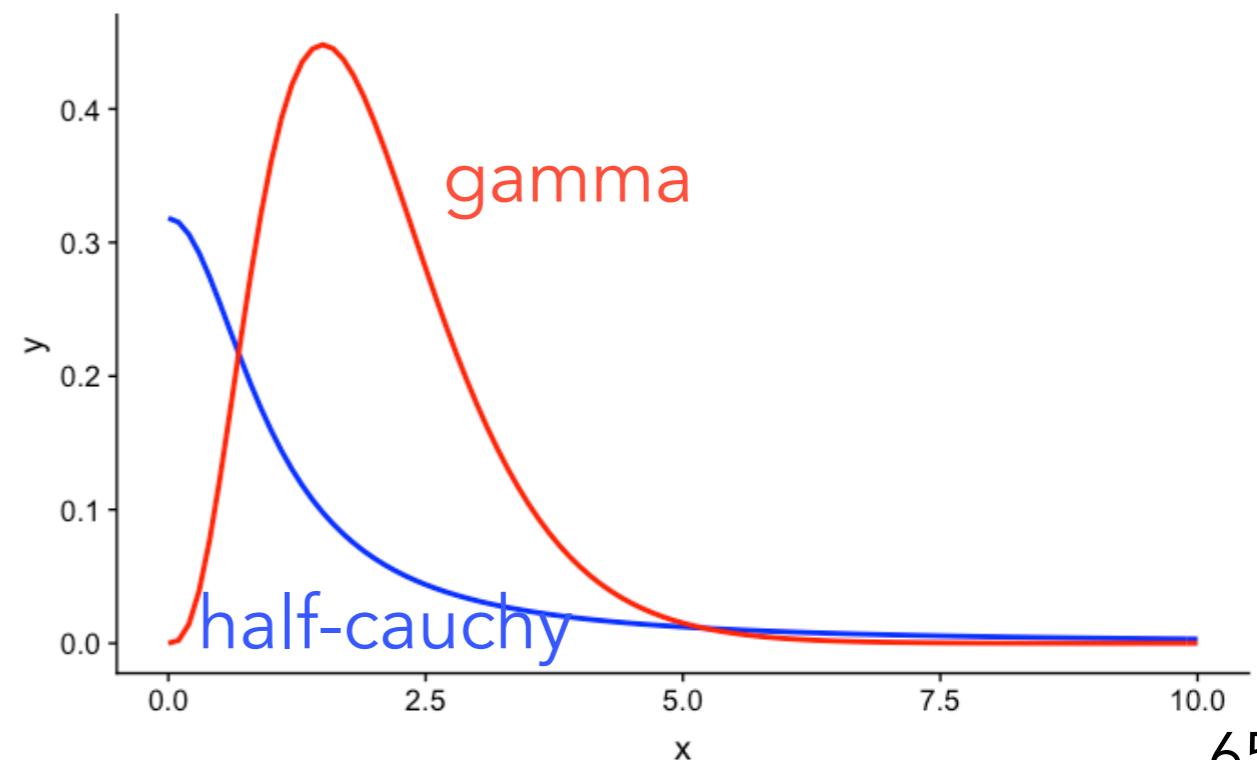
for **beta coefficients** in a regression



- **gaussian:**

- sd determines how informative the prior is

for **standard deviation** of the Gaussian



- **gamma, half-cauchy:**

- for parameters we know are positive

Inference

Posterior

$$p(H | D) = \frac{\text{Likelihood} \quad \text{Prior}}{p(D)}$$

Normalizing constant

the devil is in the denominator ...

Doing Bayesian inference

Discrete hypothesis space

$$p(H|D) = \frac{p(D|H) \cdot p(H)}{\sum_{i=1}^n p(D|H_i) \cdot p(H_i)}$$

sum over all possibilities

Continuous hypothesis space

$$p(H|D) = \frac{p(D|H) \cdot p(H)}{\int_{-\infty}^{\infty} p(D|H_i) \cdot p(H_i) dH_i}$$

integral over all possibilities

Discretizing the parameters

```
1 # grid  
2 theta = seq(0, 1, 0.01) ← 100 discrete values  
3
```

```
4 # data  
5 data = rep(0:1, c(8, 2))  
6  
7 # calculate posterior
```

```
8 df.prior = tibble(theta = theta,  
9 prior_uniform = dbeta(theta, shape1 = 1, shape2 = 1),  
10 prior_normal = dbeta(theta, shape1 = 5, shape2 = 5),  
11 prior_biased = dbeta(theta, shape1 = 8, shape2 = 2)) %>%
```

```
12 pivot_longer(cols = -theta,  
names_to = "prior_index",  
values_to = "prior") %>%
```

```
13 mutate(likelihood = dbinom(sum(data == 1),  
size = length(data),  
prob = theta)) %>%
```

```
16 group_by(prior_index) %>%
```

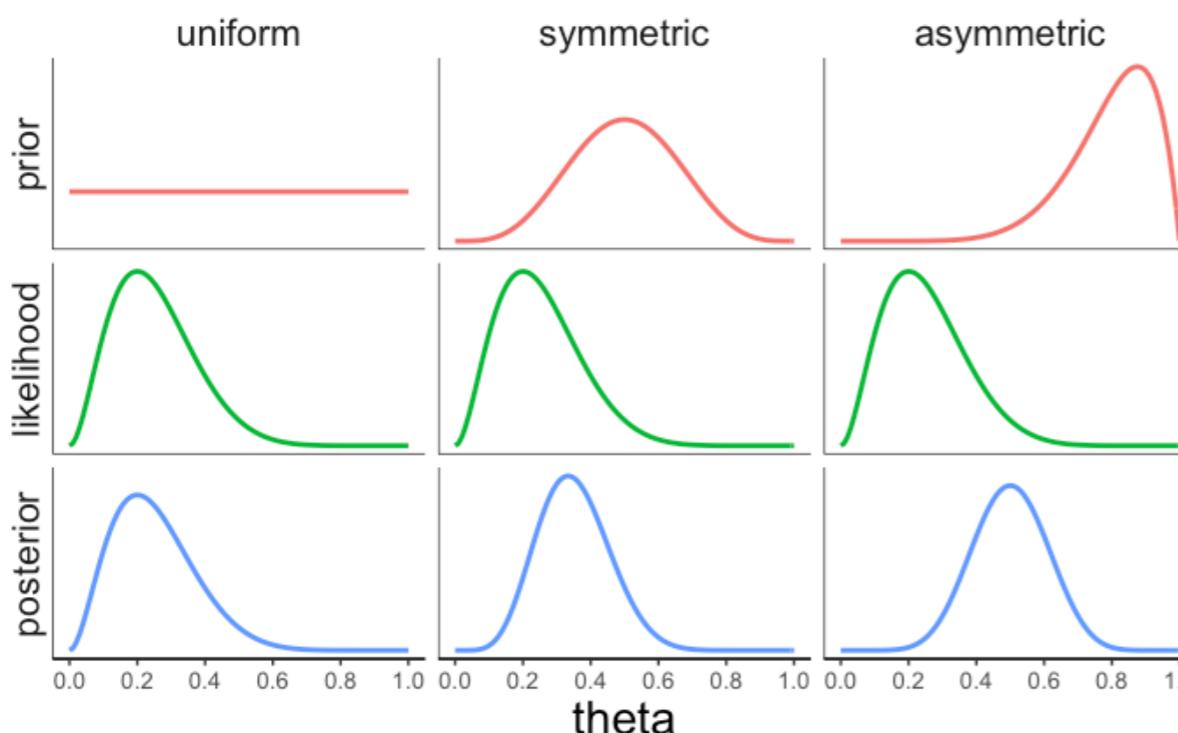
```
17 mutate(posterior = likelihood * prior / sum(likelihood * prior)) %>%
```

```
18 ungroup() %>%
```

```
19 pivot_longer(cols = -c(theta, prior_index),  
names_to = "index",  
values_to = "value")
```

for 3 variables, we would already
need 1 Mio combinations

The CURSE of
dimensionality



Inference via sampling

- we cannot directly calculate the probability of the posterior (because it might have a pretty weird shape)
- **but:** we can draw random samples from the posterior
- we can then use our data wrangling and visualization skills to make inferences based on these samples

It's as if ...

we don't have **pnorm()**

but we do have **rnorm()**

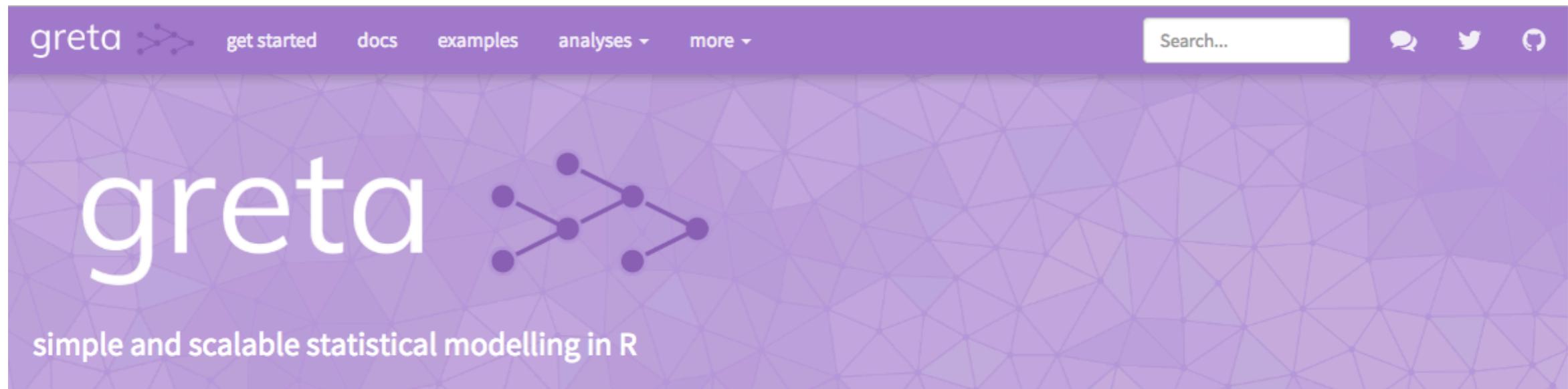
Inference via sampling

- Bayesian data analysis is becoming more popular because:
 - computers are getting more powerful
 - inference techniques are getting better
 - software packages become easier to use

Doing Bayesian data analysis

Software packages

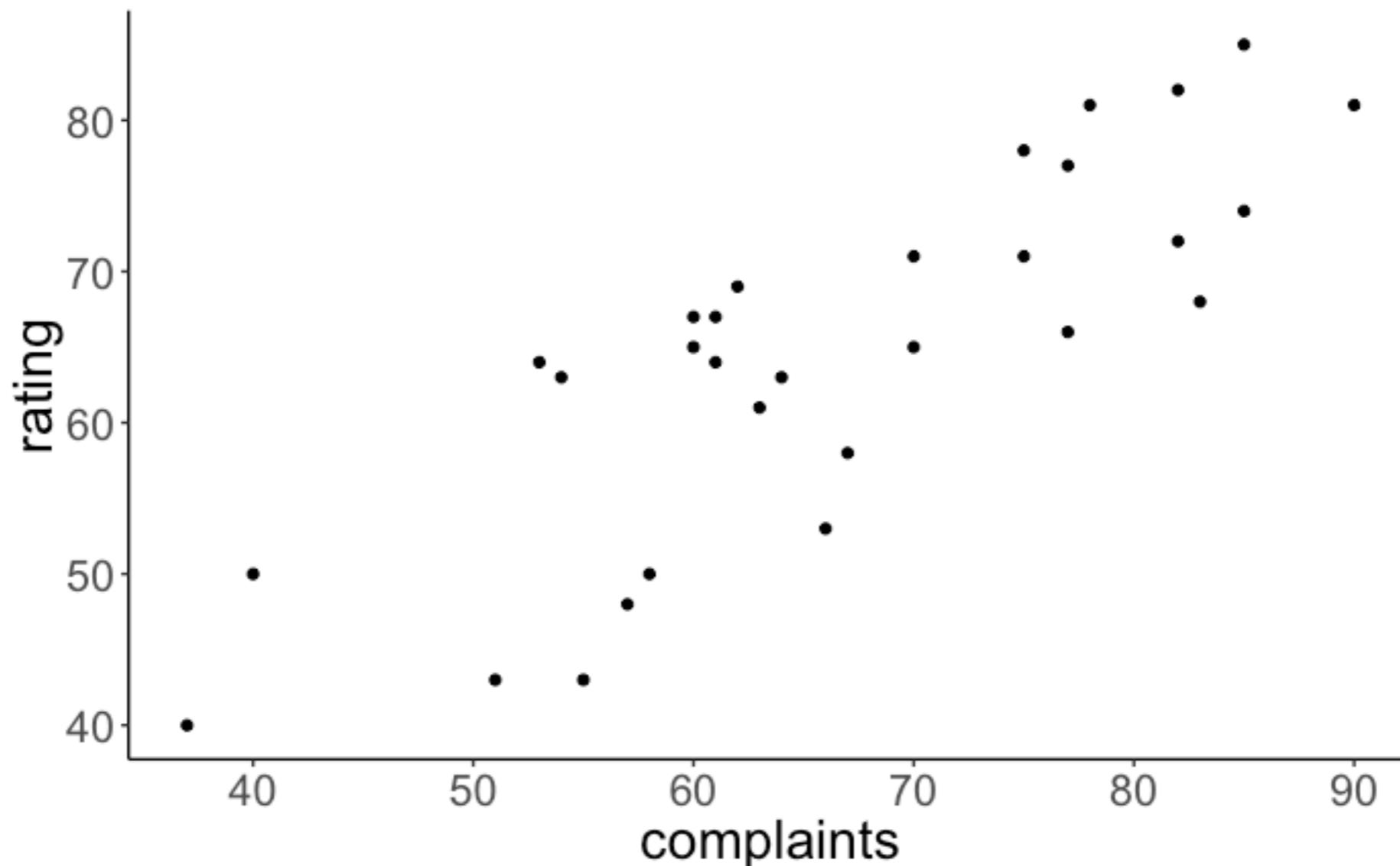
```
library("greta")
```



- let's us write Bayesian models directly in R with a simple syntax
- uses Tensorflow to implement Hamiltonian Monte Carlo sampling (a fast inference algorithm ...)

Attitude data set

What's the relationship between how well an employee handles complaints and their overall rating?



Frequentist analysis

Frequentist analysis

```
1 # fit model
2 fit = lm(formula = rating ~ 1 + complaints,
3           data = df.attitude)
4
5 # print summary
6 fit %>% summary()
```

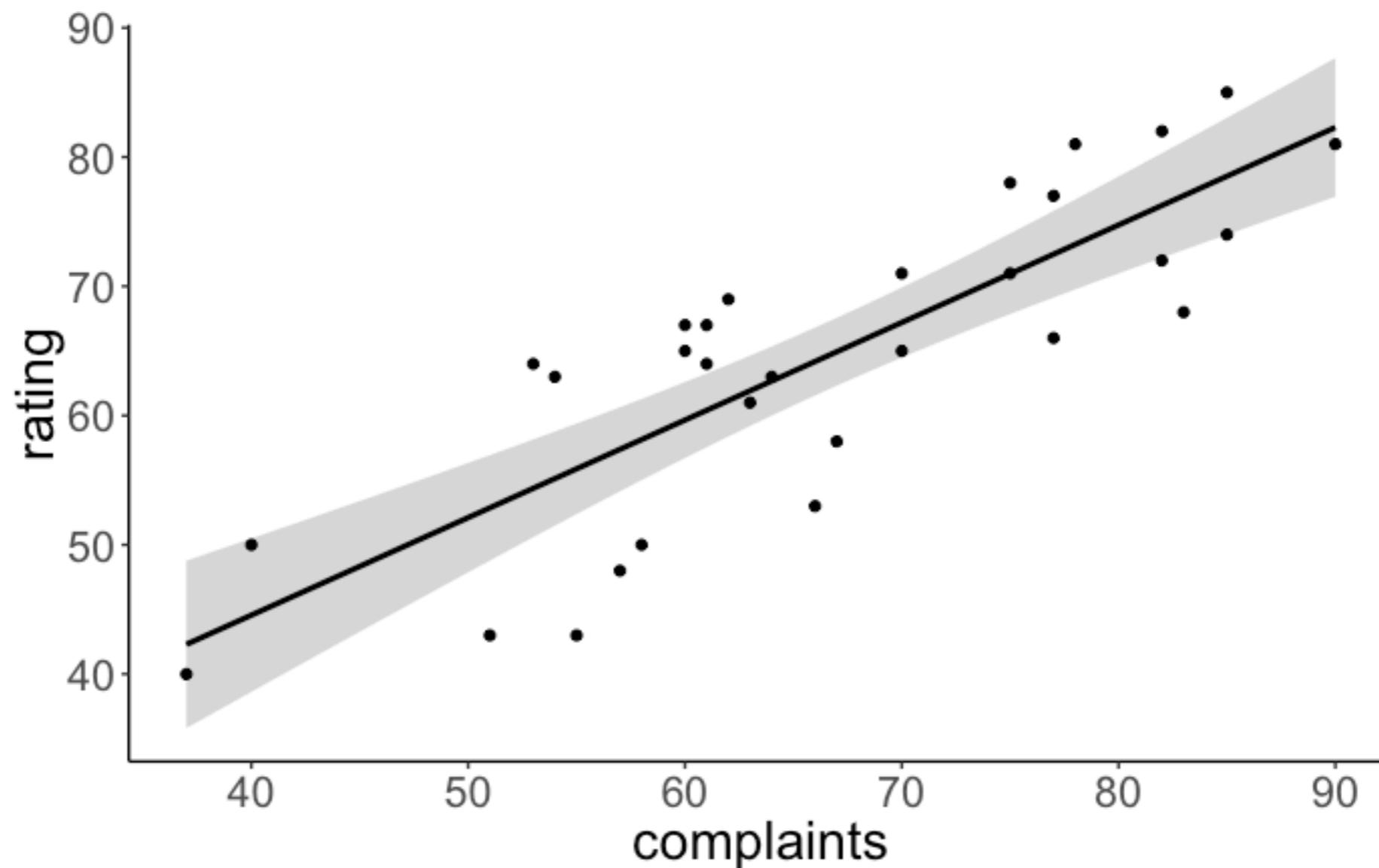
```
Call:
lm(formula = rating ~ 1 + complaints, data = df.attitude)

Residuals:
    Min      1Q  Median      3Q     Max 
-12.8799 -5.9905  0.1783  6.2978  9.6294 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 14.37632   6.61999   2.172   0.0385 *  
complaints   0.75461   0.09753   7.737 1.99e-08 *** 
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.993 on 28 degrees of freedom
Multiple R-squared:  0.6813, Adjusted R-squared:  0.6699 
F-statistic: 59.86 on 1 and 28 DF,  p-value: 1.988e-08
```

Visualize model predictions



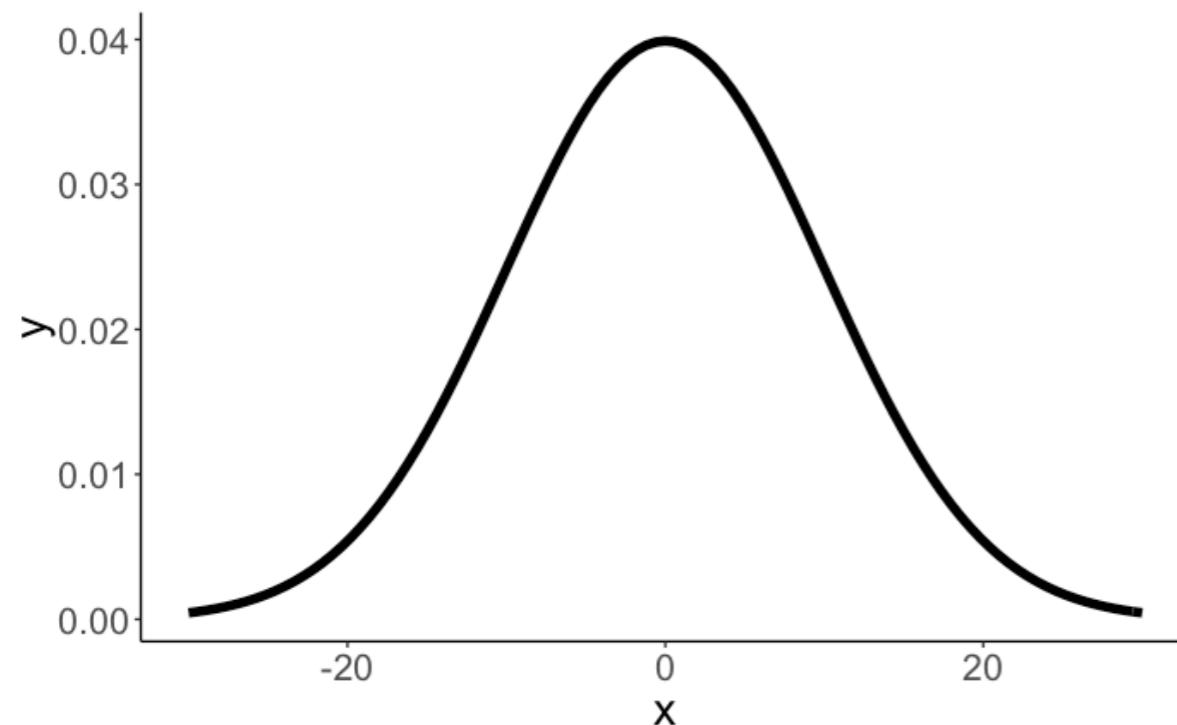
Best-fitting regression line with confidence interval

Bayesian analysis

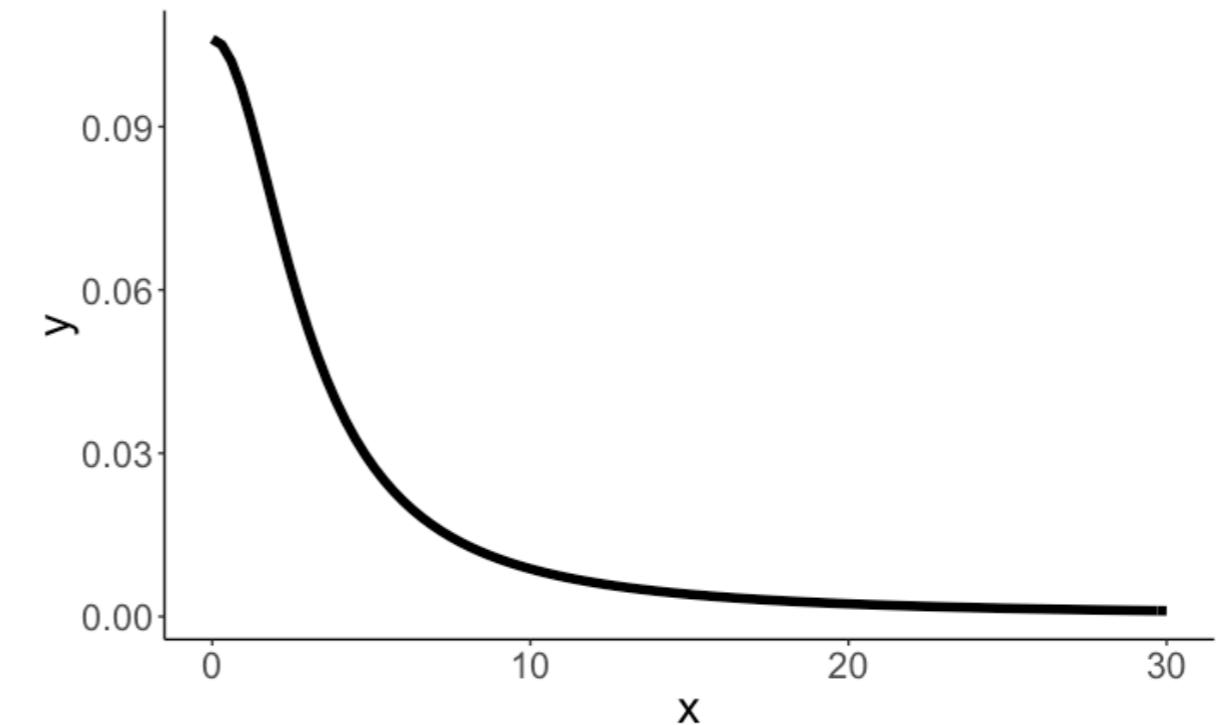
Model specification

```
1 library("greta")
2 library("tidybayes")
3
4 # variables & priors
5 b0 = normal(0, 10) ← priors
6 b1 = normal(0, 10)
7 sd = cauchy(0, 3, truncation = c(0, Inf))
8
9 # linear predictor
10 mu = b0 + b1 * attitude$complaints ← linear combination
11
12 # observation model (likelihood)
13 distribution(attitude$rating) = normal(mu, sd)
14
15 # define the model
16 m = model(b0, b1, sd) ← Gaussian likelihood
                                         ← build the model
```

Priors



**Gaussian prior on
intercept and coefficient**

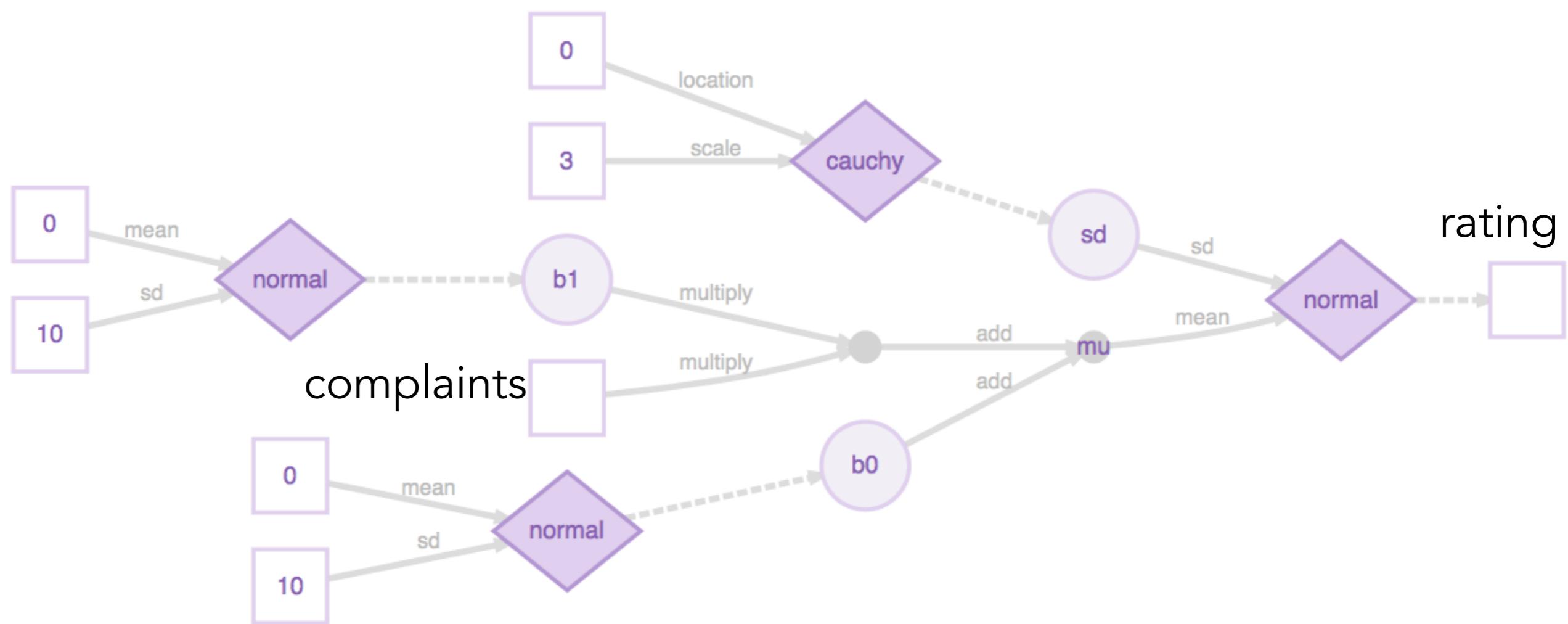


**Truncated Cauchy prior on
the standard deviation**

weakly informative priors (allow for a wide range of possible values)

Graphical representation of the model

```
1 # plotting  
2 plot(m)
```



Inference via sampling

Markov Chain

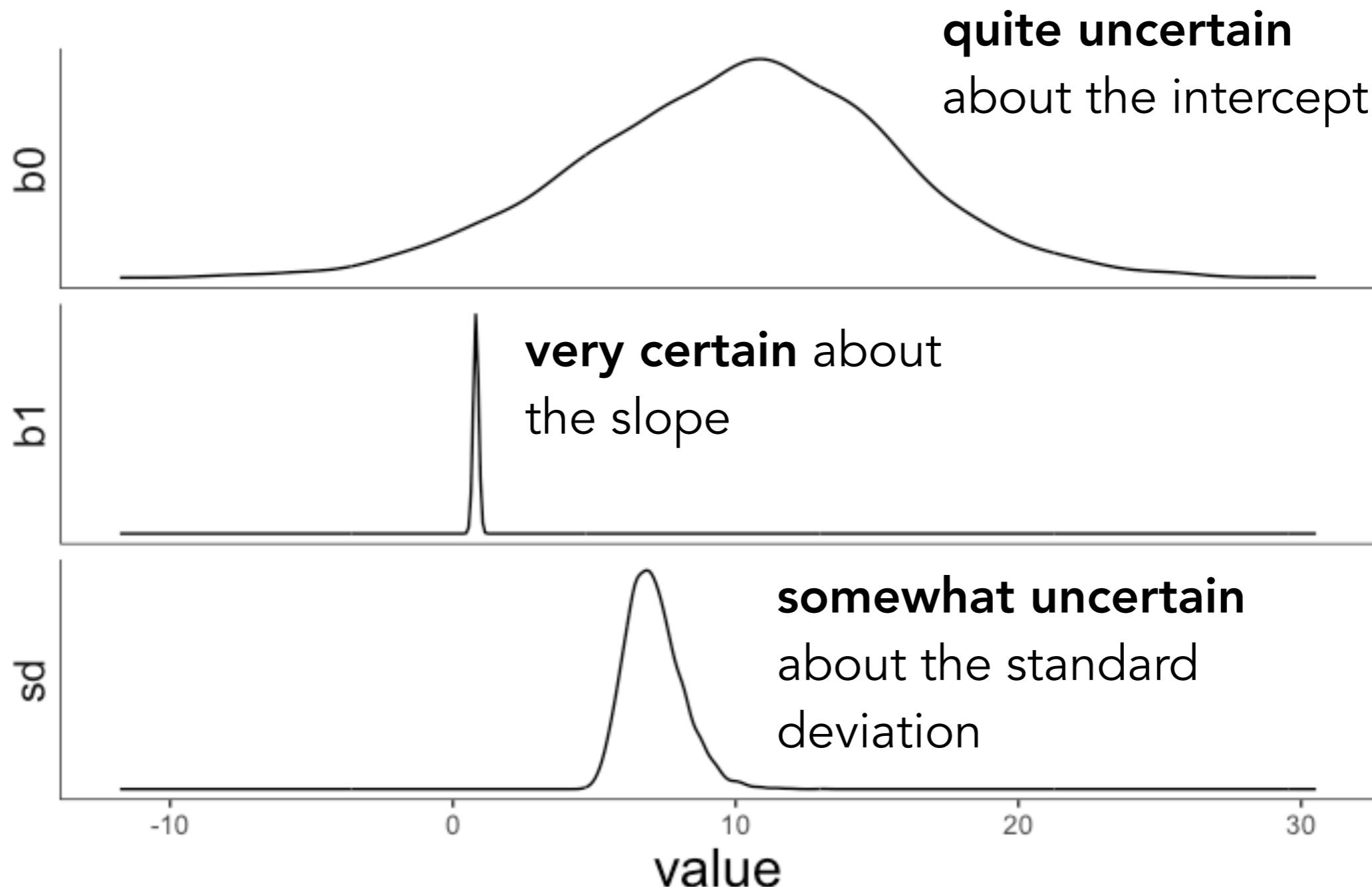
Monte Carlo
inference

```
1 # sampling
2 draws = mcmc(m, n_samples = 1000)
3
4 # tidy up the draws
5 df.draws = tidy_draws(draws) %>%
6   clean_names()
```

chain	iteration	draw	b0	b1	sd
1	1	1	6.08	0.87	7.60
1	2	2	1.12	0.95	7.66
1	3	3	-1.83	0.99	7.01
1	4	4	-4.23	1.02	6.64
1	5	5	3.26	0.87	7.96
1	6	6	-1.04	0.98	7.67
1	7	7	-0.83	0.97	10.12
1	8	8	-1.41	0.97	8.02
1	9	9	9.46	0.81	6.30
1	10	10	10.02	0.84	6.57

each of these is a solution
for explaining the data

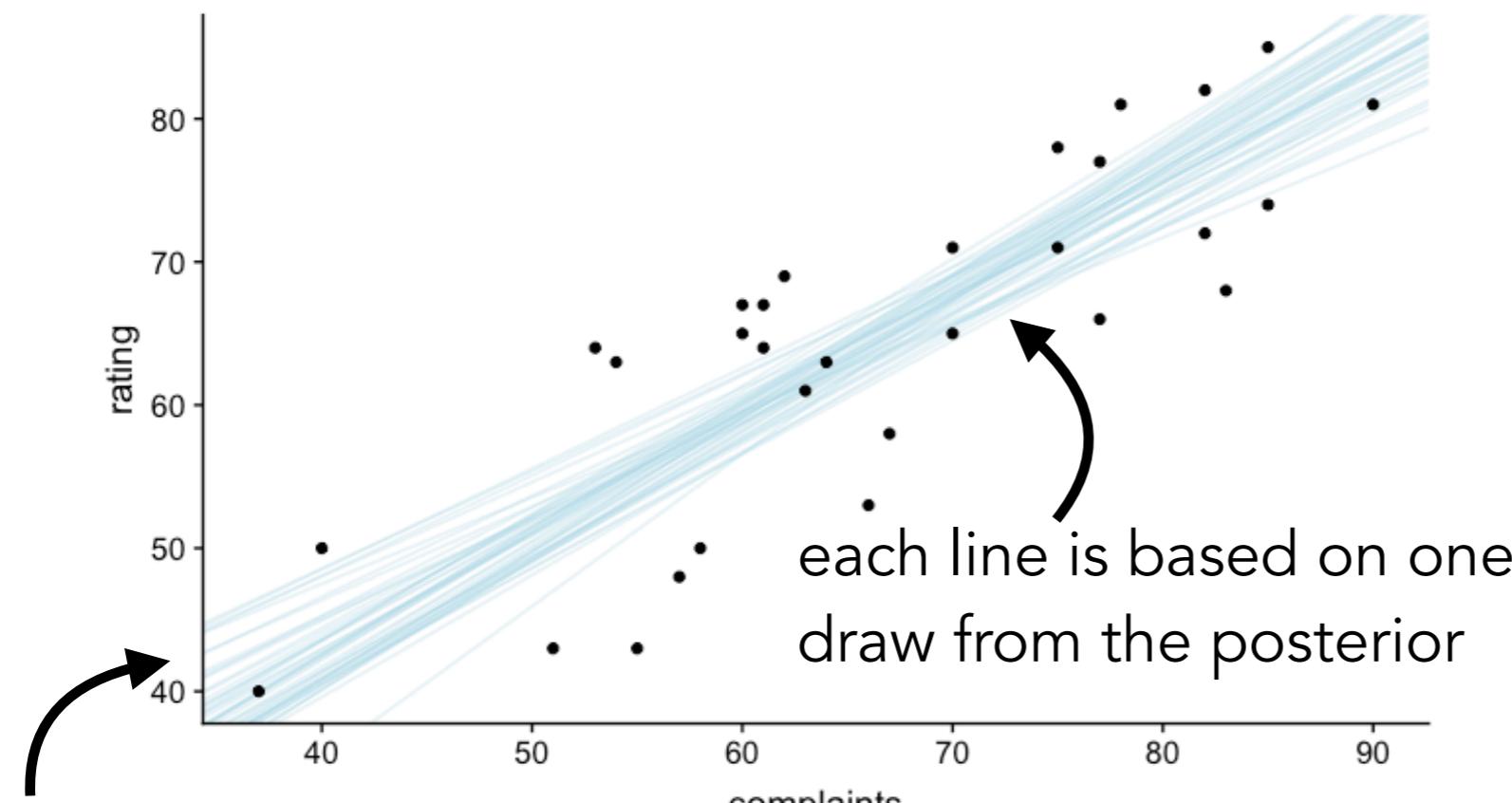
Visualize the posterior



**Posterior distribution over the three
parameters in the model**

Visualize the model predictions

```
1 ggplot(data = df.attitude,
2         mapping = aes(x = complaints,
3                         y = rating)) +
4   geom_abline(data = df.draws %>%
5               sample_n(size = 50),
6               aes(intercept = b0,
7                   slope = b1),
8               alpha = 0.3,
9               color = "lightblue") +
10  geom_point()
```

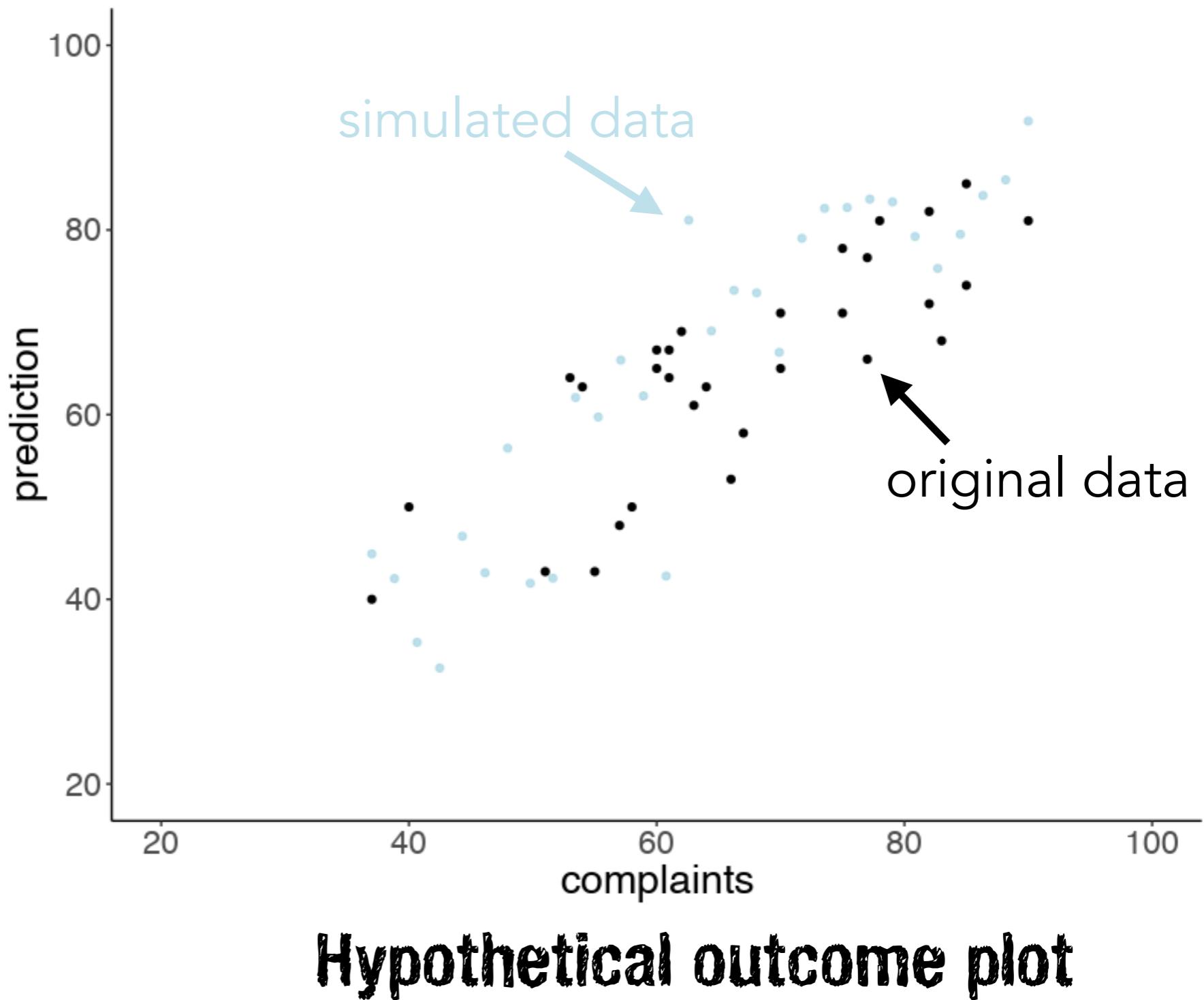


explains the "high" uncertainty about the intercept

chain	iteration	draw	b0	b1	sd
1	1	1	6.08	0.87	7.60
1	2	2	1.12	0.95	7.66
1	3	3	-1.83	0.99	7.01
1	4	4	-4.23	1.02	6.64
1	5	5	3.26	0.87	7.96
1	6	6	-1.04	0.98	7.67
1	7	7	-0.83	0.97	10.12
1	8	8	-1.41	0.97	8.02
1	9	9	9.46	0.81	6.30
1	10	10	10.02	0.84	6.57

Posterior predictive check

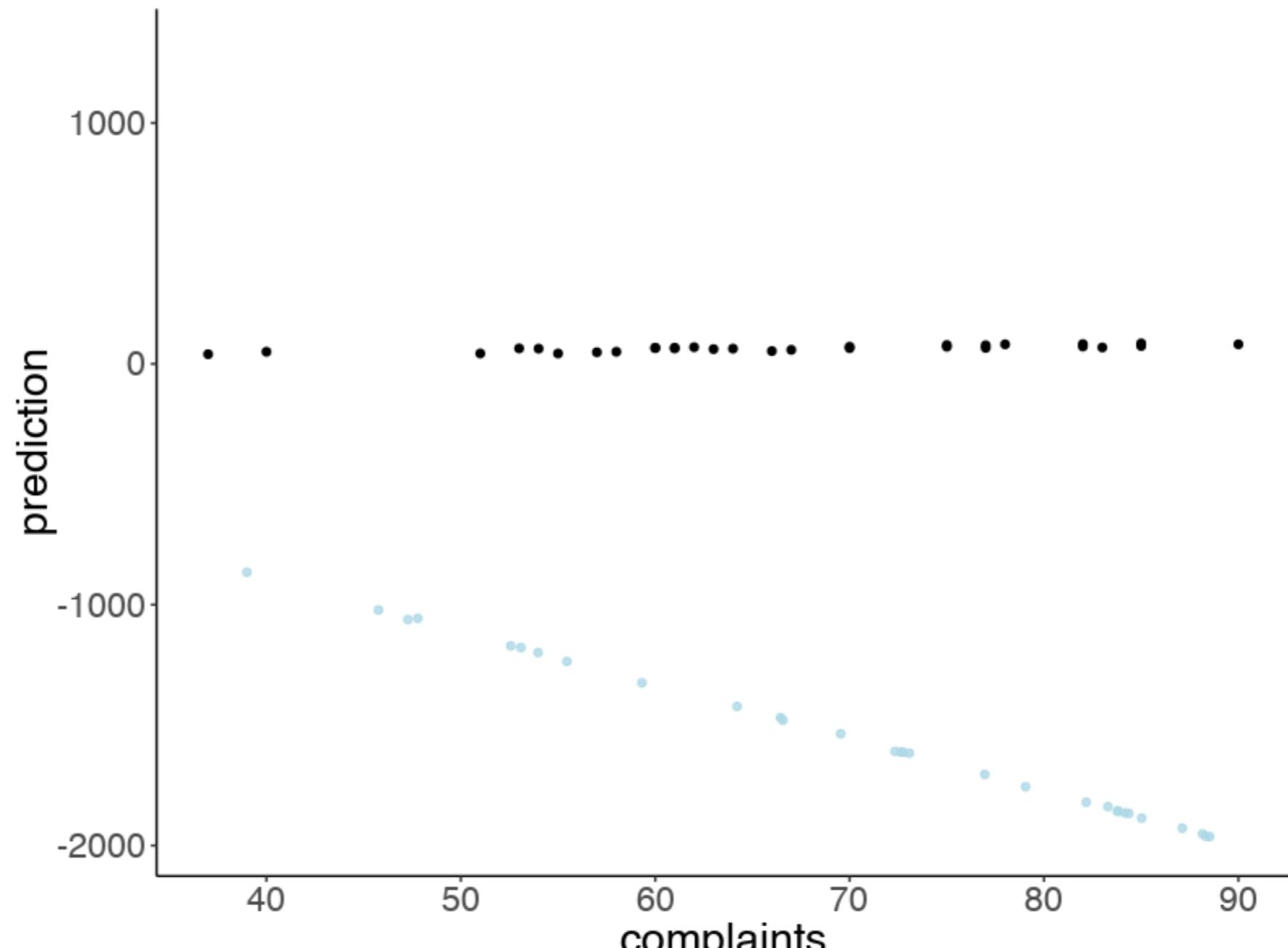
1. sample parameters from the **posterior distribution**
2. generate data using these parameters (using the likelihood function)



Hypothetical outcome plot

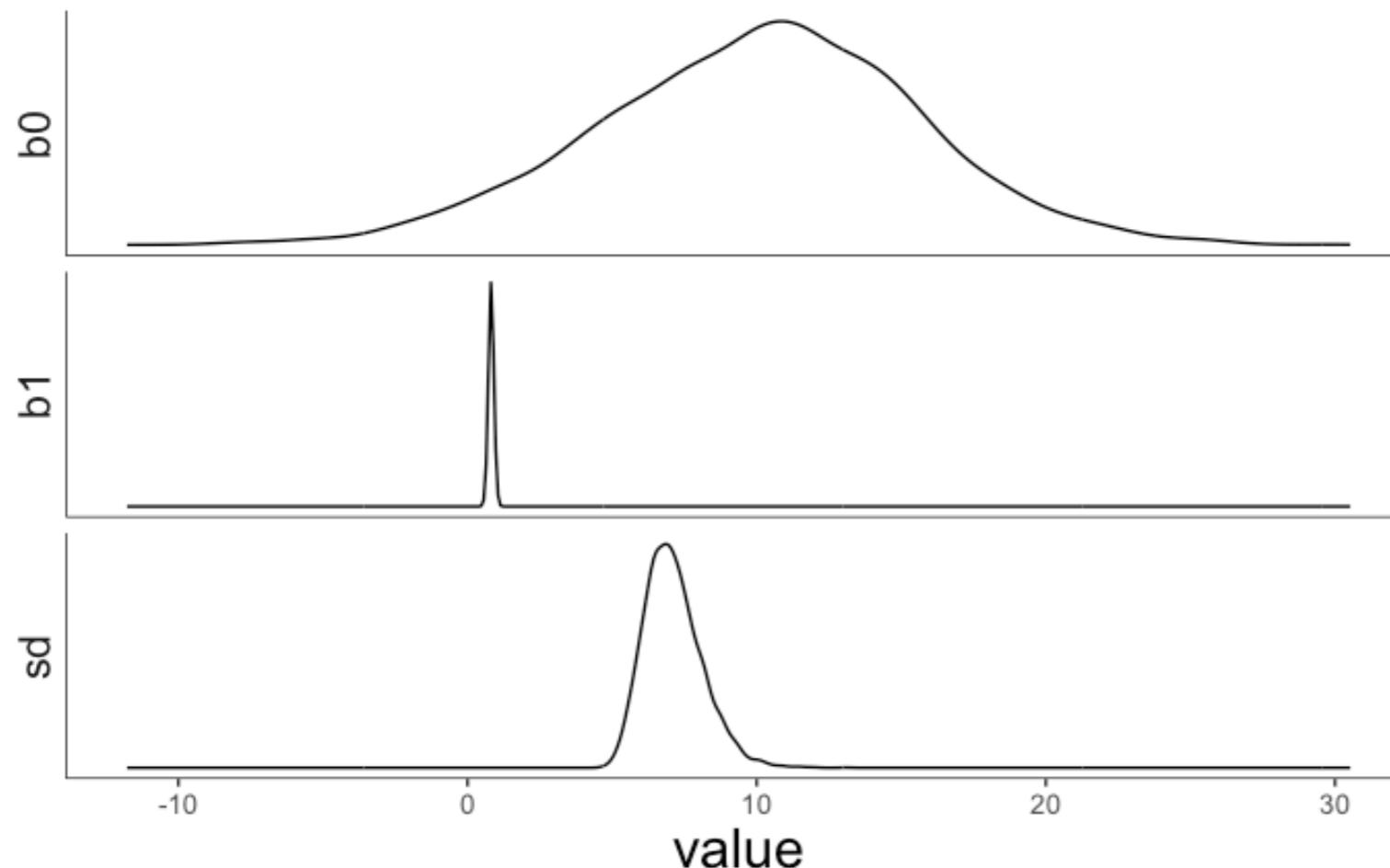
Prior predictive check

1. sample parameters from the **prior distribution**
2. generate data using these parameters (using the likelihood function)



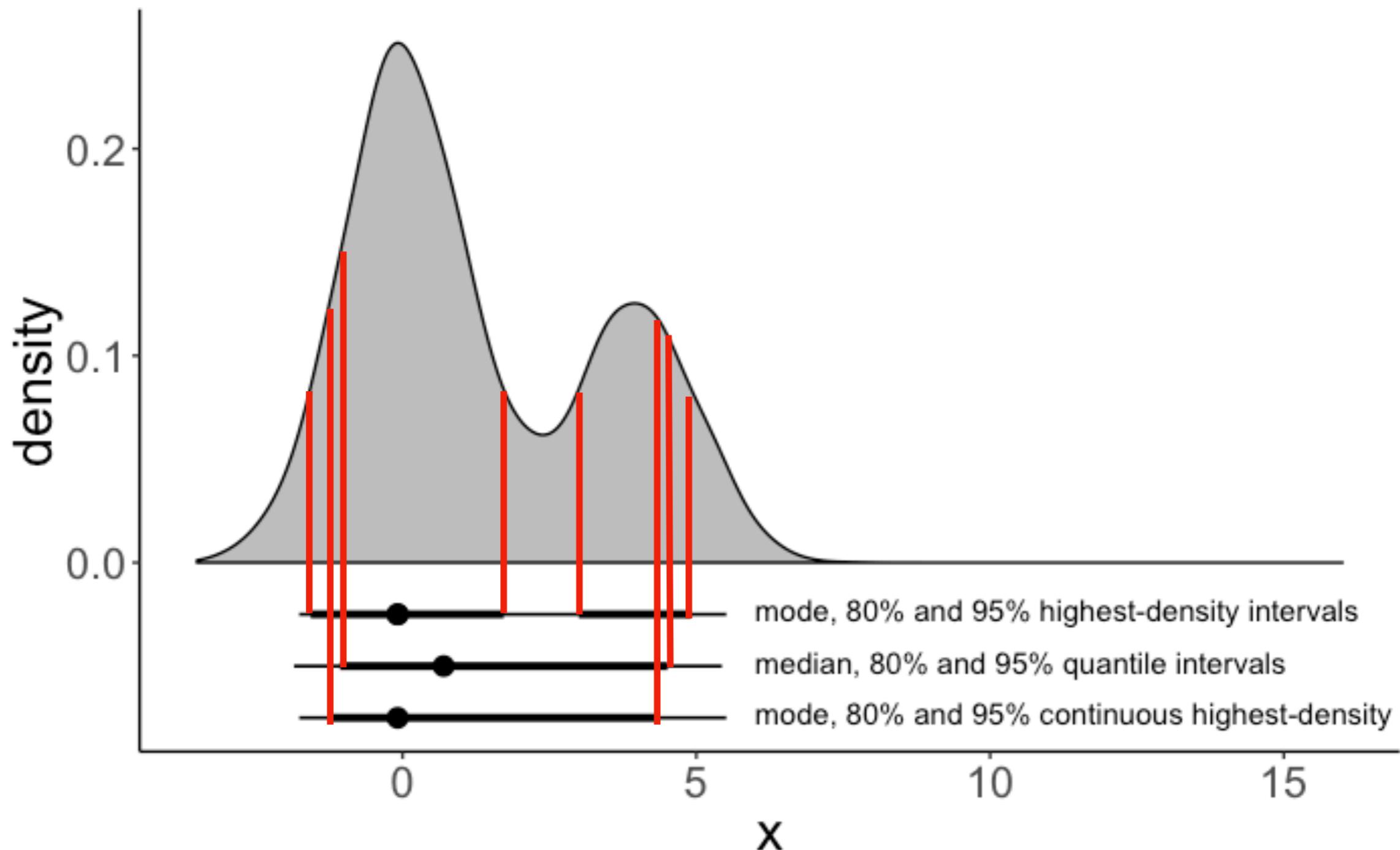
Hypothetical outcome plot

Summarizing results



- Posterior over each parameter is the result of the Bayesian data analysis.
 - no p-values
 - no confidence intervals

Different kinds of credible intervals



Summary

- Quick recap
- Mixed effects logistic regression
- Bayesian data analysis
 - Comparison between frequentist and Bayesian data analysis
 - Quick flash from the past
 - Flipping coins
 - What affects the posterior?
 - Ingredients: likelihood, prior, inference
 - Doing Bayesian data analysis

Feedback

How was the pace of today's class?

much a little just a little much
too too right too too
slow slow

How happy were you with today's class overall?



What did you like about today's class? What could be improved next time?

Thank you!