THE UNIVERSITY of TENNESSEE KNOXVILLE

Lecture 15: Bayesian and Frequentist Worldviews

Instructor: Sergei V. Kalinin

Basic frequentist approach

Views probability as the long-run frequency of events. For example, the probability of getting a 'heads' in a coin toss is interpreted as the long-run frequency of getting 'heads' over many trials.

- **Mean (Expected Value)**:
  - Represents the central tendency of a dataset.
  - It is the sum of all observations divided by the number of observations.

- **Dispersion**:
  - Measures the spread or variability in a dataset.
  - Common metrics include variance (the average squared deviation from the mean) and standard deviation (the square root of variance)

- **Error Bars**:
  - Graphical representation of the variability of data.
  - Commonly used error bars include standard deviation, standard error, and confidence intervals.
  - Help indicate the reliability of a point estimate.
  - Wider error bars typically suggest more variability or uncertainty in the data.

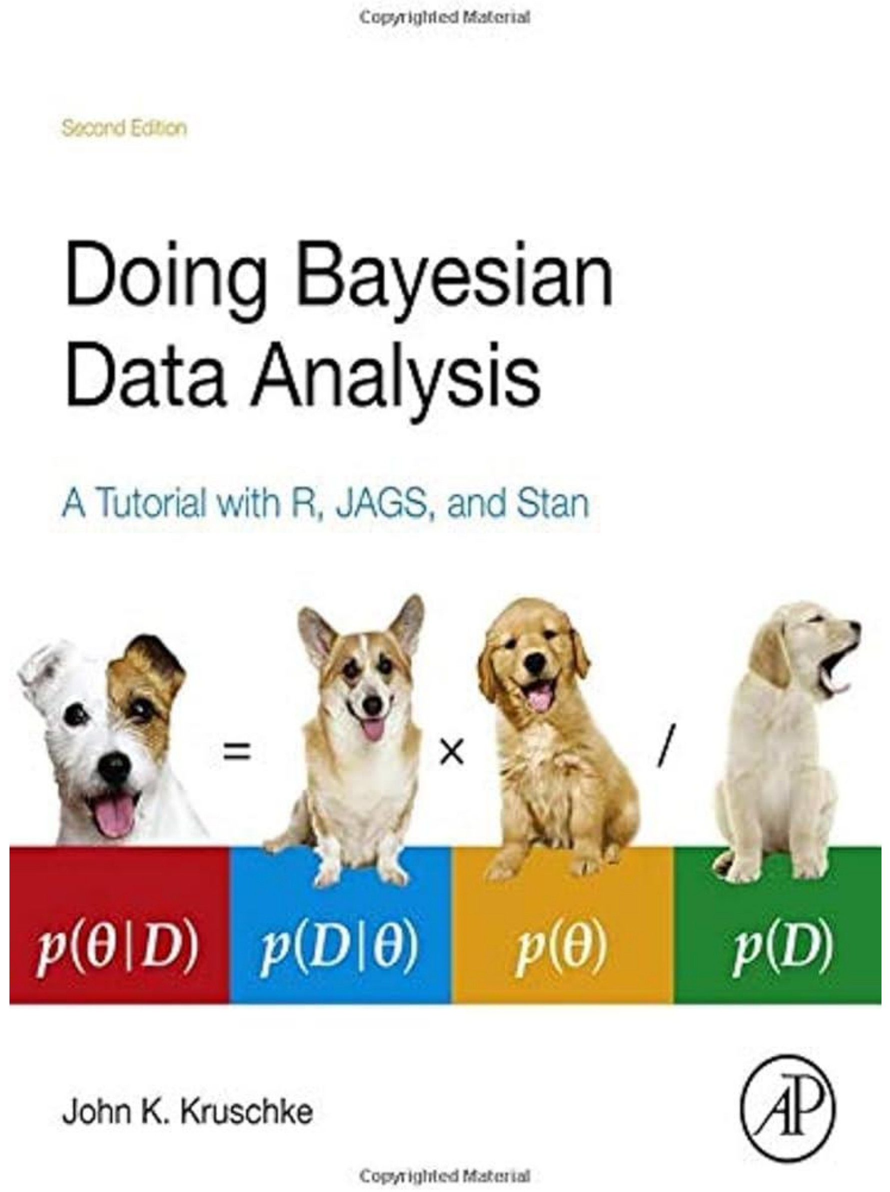
Basic frequentist approach

- **Null Hypothesis (H0)**:
    - A statement or assumption that there is no effect or no difference, serving as a default or starting point
    - It is the hypothesis that researchers typically aim to test against using statistical tests

- **Alternative Hypothesis (H1 or Ha)**:
    - A statement indicating the presence of an effect or difference

- **P-values**:
    - A p-value measures the strength of the evidence against the null hypothesis.
    - It represents the probability of observing data (or something more extreme) given that the null hypothesis is true.
    - A small p-value (typically < 0.05) suggests that the observed data is inconsistent with the null hypothesis, leading researchers to reject H0 in favor of Ha.

If feels non-straightforward, you are not alone…



Setting of the problem

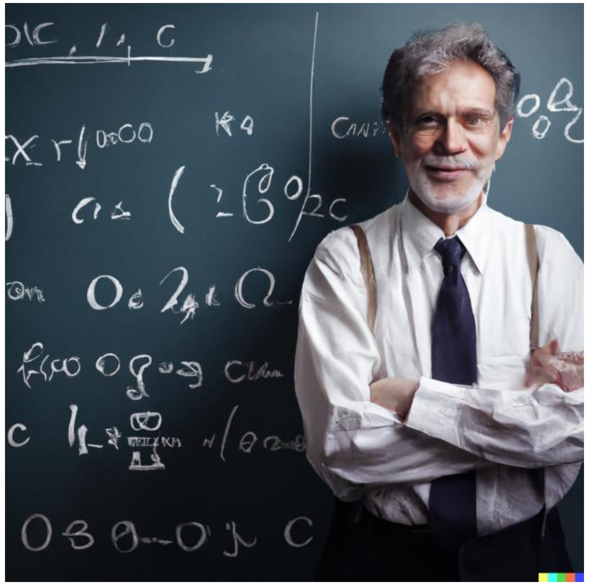
- Imagine that you are tossing a coin with your friend to decide who goes for groceries. What is the probability that the coin will land tails?

- Imagine that previously you have tossed this coin several times, and it landed tails 3 times in a row.

- … what if it was 100 times in a row?

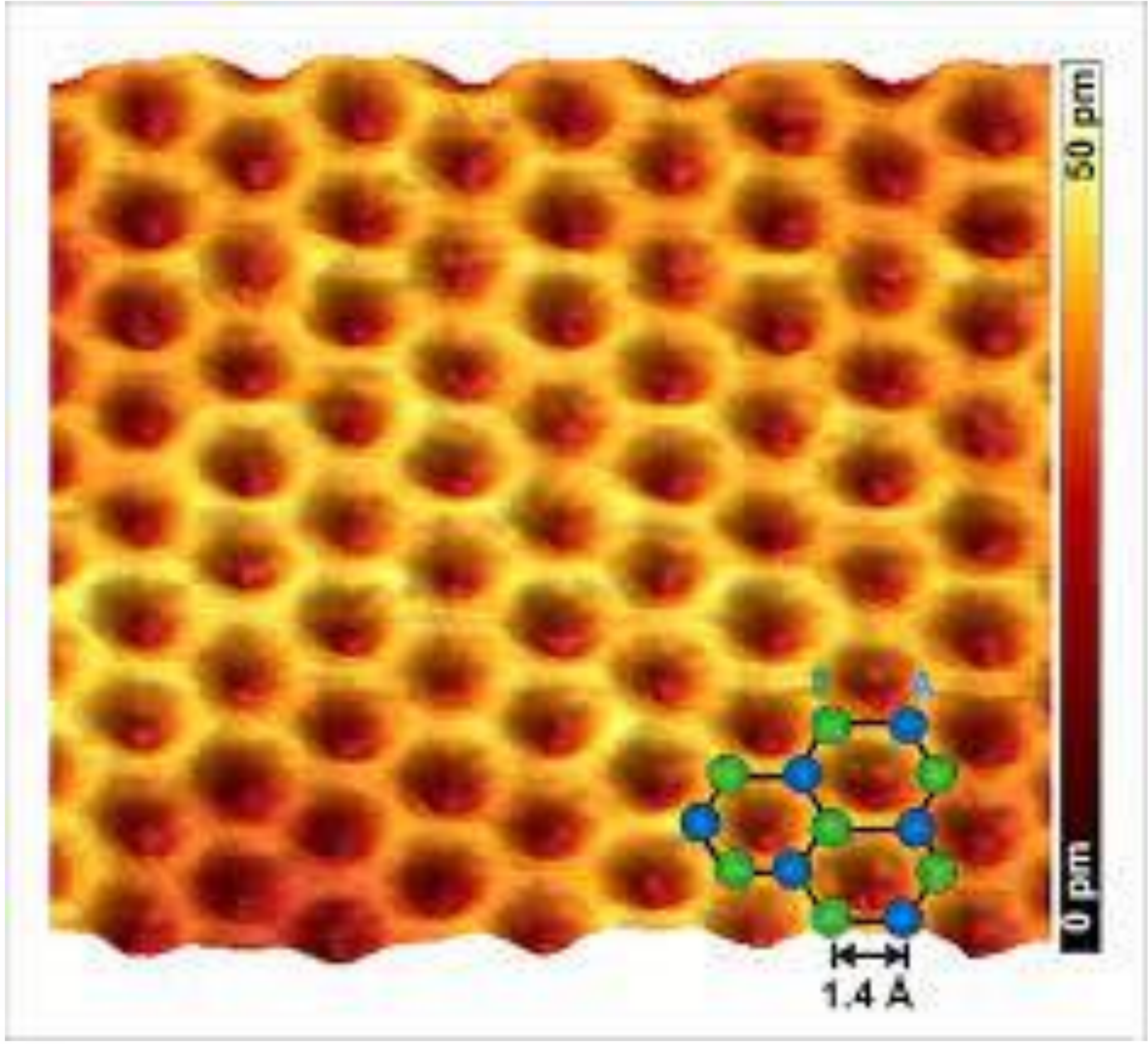
Setting of the problem - 2

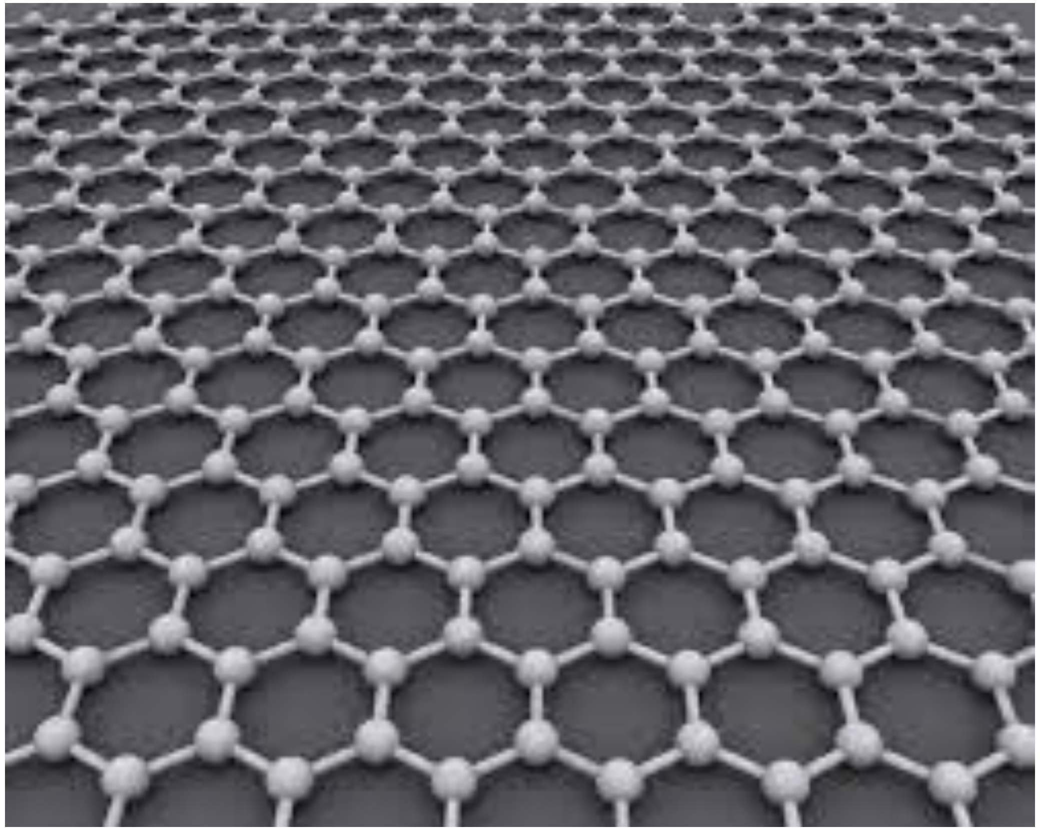
- You have a math exam problem made by Prof. Calculus stating:
  - You have a fair coin.
  - You have tossed it 100 times, and it landed tails.
  - What is the probability that it will land tails during 101 attempt?



- You play the coin toss game with a sketchy person named Joe the Gambler in Las Vegas.
  - He tells you that “We have a fair coin”.
  - You have tossed it 10 times, and it landed tails.
  - What is the probability that it will land tails during 11 attempt?

Off to scientific examples

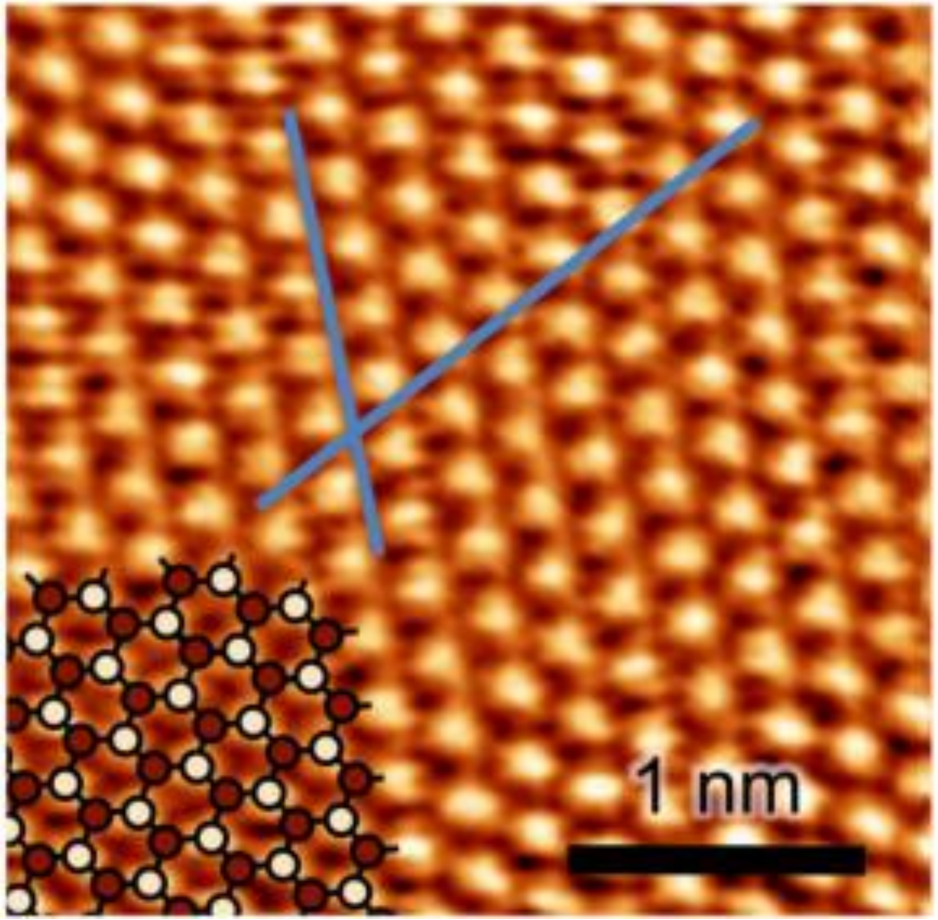


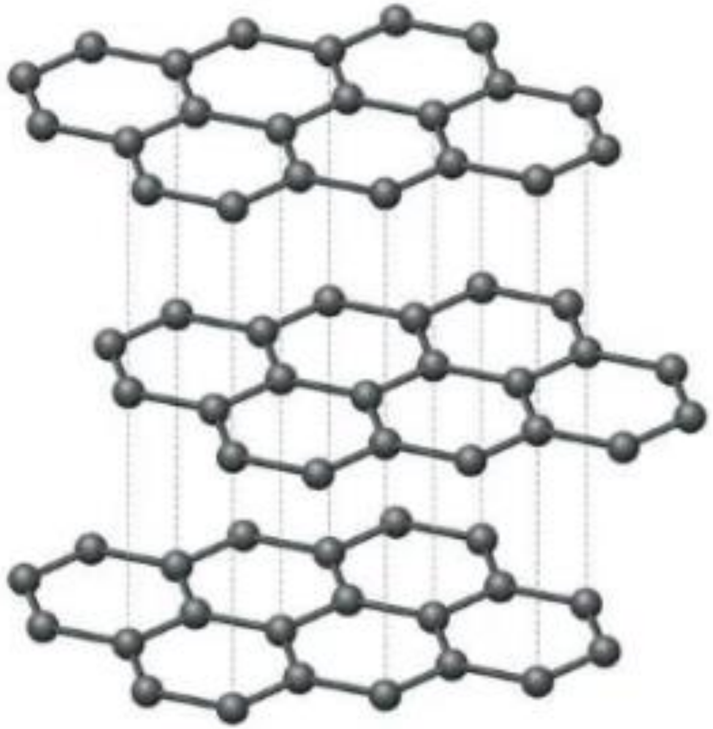


http://www.nanoscience.de/HTML/research/graphene.html

https://en.wikipedia.org/wiki/Graphene

Off to scientific examples





https://unacademy.com/content/jee/study-material/chemistry/structure-of-graphite-and-uses/

Atomically resolved STM image of the graphite surface. A schematic drawing of the hexagonal lattice of graphite is overlaid . Bright and dark filled circles correspond to the two different sublattice sites, and only the white sublattice is observed in STM image.

A. Amend, T. Matsui, H. Sato, and H Fukuyama, STS Studies of Zigzag Graphene Edges Produced by Hydrogen-Plasma Etching, Journal of Surface Science and Nanotechnology 16, 72 (2018) DOI:10.1380/ejssnt.2018.72

Frequentist paradigm

- Defines probability as a long-run frequency independent, identical trials

- Looks at parameters (i.e., the true mean of the population, the true probability of heads) as fixed quantities

This paradigm leads one to specify the null and alternative hypotheses, collect data, calculate the significance probability under the assumption that the null is true, and draw conclusions based on these significance probabilities using size of the observed effects to guide decisions

R. A. Fisher (1890–1962) https://en.wikipedia.org/wiki/Ronald_Fisher

Bayesian paradigm

- Defines probability as a subjective belief (which must be consistent with all of one's other beliefs)

- Looks at parameters (i.e., the true mean population, the true probability of heads) as random quantities because we can never know them with certainty

This paradigm leads one to specify plausible models to assign a prior probability to each model, to collect data, to calculate the probability of the data under each model, to use Bayes' theorem to calculate the posterior probability of each model, and to make inferences based on these posterior probabilities. The posterior probabilities enable one to make predictions about future observations and one uses one's loss function to make decisions that minimize the probable loss

Thomas Bayes, 1701 - 1761 https://en.wikipedia.org/wiki/Thomas_Bayes

Bayesianism vs. frequentism in life

You are waiting on a subway platform for a train that is known to run on a regular schedule, only you don't know how much time is scheduled to pass between train arrivals, nor how long it's been since the last train departed.

As more time passes, do you:

(a) grow **more confident** that the train will arrive soon, since its eventual arrival *can only be getting closer*, not further away, or

(b) grow **less confident** that the train will arrive soon, since the longer you wait, the more likely it seems that either the scheduled arrival times are far apart or else that you happened to arrive just after the last train left – or both.

from Floyd Bullard

From Jeff Witmer, Introduction to Bayesian Statistics, slides 24 February 2016

Bayesianism vs frequentism in life

An opaque jar contains thousands of beads (but obviously a finite number!). You know that all the beads are either red or white but you have no idea at all what fraction of them are red. You begin to draw beads out of the bin at random without replacement. You notice that all of the first several beads have been red. As you observe more and more red beads, is the conditional probability (i.e., conditional upon the previous draws' colors) of the next bead being red

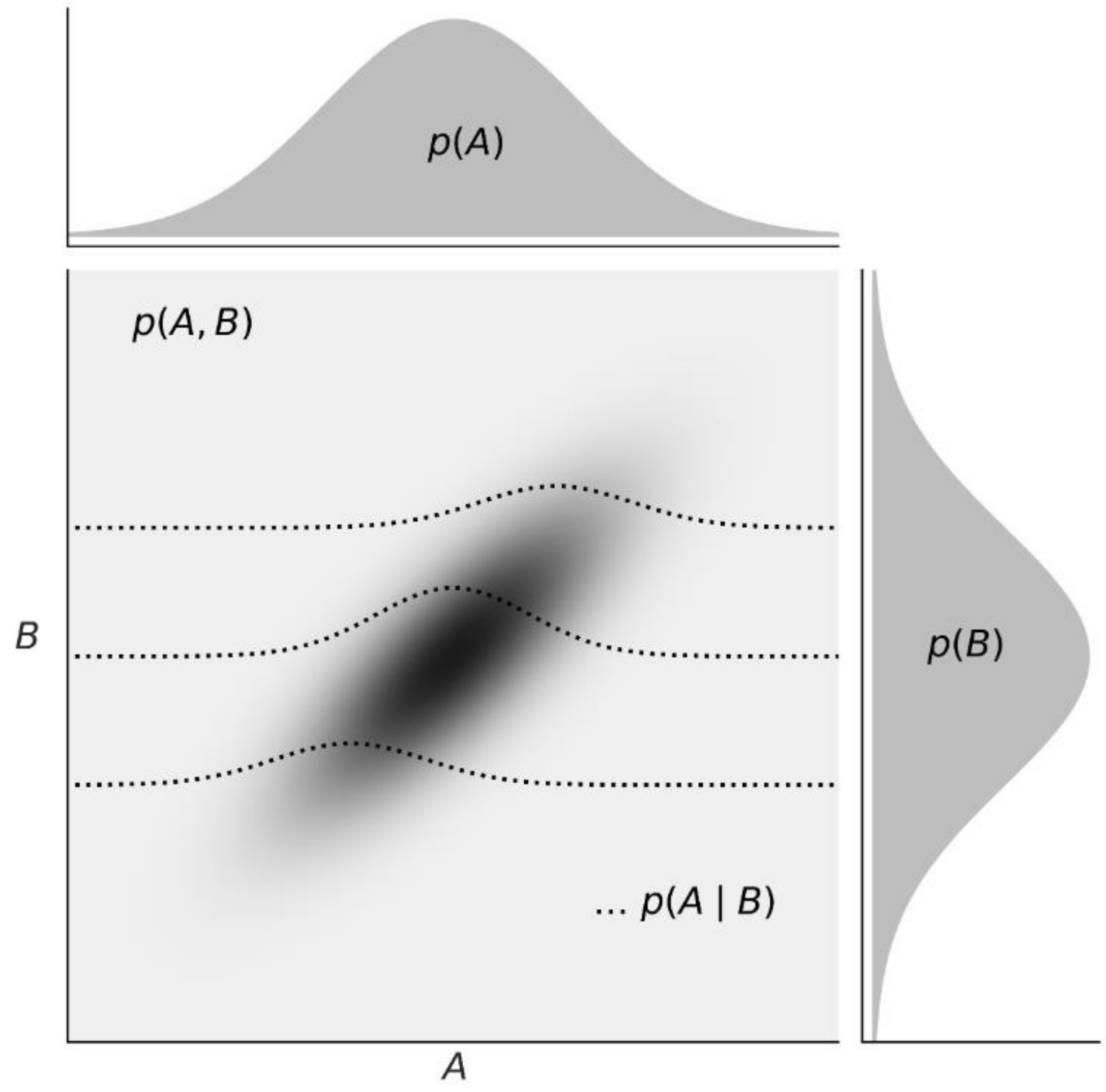
(a) decreasing, ***as it must***, since you're removing red beads from a finite population, or

(b) increasing, because you initially didn't know that there would be so many reds, but now it seems that the jar must be mostly reds.

from Floyd Bullard

From Jeff Witmer, Introduction to Bayesian Statistics, slides 24 February 2016

Probabilities: joint, conditional, marginal



From O. Martin

But what is Bayes' theorem?

Conditional probabilities: $P(a,b) = P(a|b)\,P(b) = P(a)P(b|a)$

Thus, $P(a|b) = P(a,b)/P(b) = P(a)P(b|a)/P(b)$

But what is P(b)? $P(b) = \sum_a P(a,b) = \sum_a P(a)P(b|a)$

$$\boxed{P(a|b) = \frac{P(a)P(b|a)}{P(b)} = \frac{P(a)P(b|a)}{\sum_{a^*} P(a^*)P(b|a^*)}}$$

- **Prior**
- **Posterior**
- **Likelihood**
- **Evidence**

Bayesian paradigm in science

- Bayes' theorem can be usefully re-written for science as:

Posterior: probability of the model given the data

Likelihood: probability of the data given the model

Prior: probability of the model

$$P(\text{model}|\text{data}) = \frac{P(\text{data}|\text{model})\, P(\text{model})}{P(\text{data})}$$

Evidence [can typically be absorbed into the normalization of the posterior]

Bayesian paradigm in science

Typical statistics problem: There is a parameter, θ, that we want to estimate, and we have some data.

Traditional (frequentist) methods: Study and describe $P(\text{data} \mid \theta)$. If the data are unlikely for a given θ, then state "that value of θ is not supported by the data." (A hypothesis test asks whether a *particular* value of θ might be correct; a CI presents a range of plausible values.)

Bayesian methods: Describe the distribution $P(\theta \mid \text{data})$.

A frequentist thinks of θ as fixed (but unknown) while a Bayesian thinks of θ as a random variable that has a distribution.

Bayesian reasoning is natural and easy to think about. It is becoming much more commonly used.

From Jeff Witmer, Introduction to Bayesian Statistics, slides 24 February 2016

Bayesian method

Bayes answers the questions we really care about.

Pr(I have disease | test +) vs Pr(test + | disease)

Pr(A better than B | data) vs Pr(extreme data | A=B)

Bayes is natural (vs interpreting a CI or a P-value)

Note: blue = Bayesian, red = frequentist

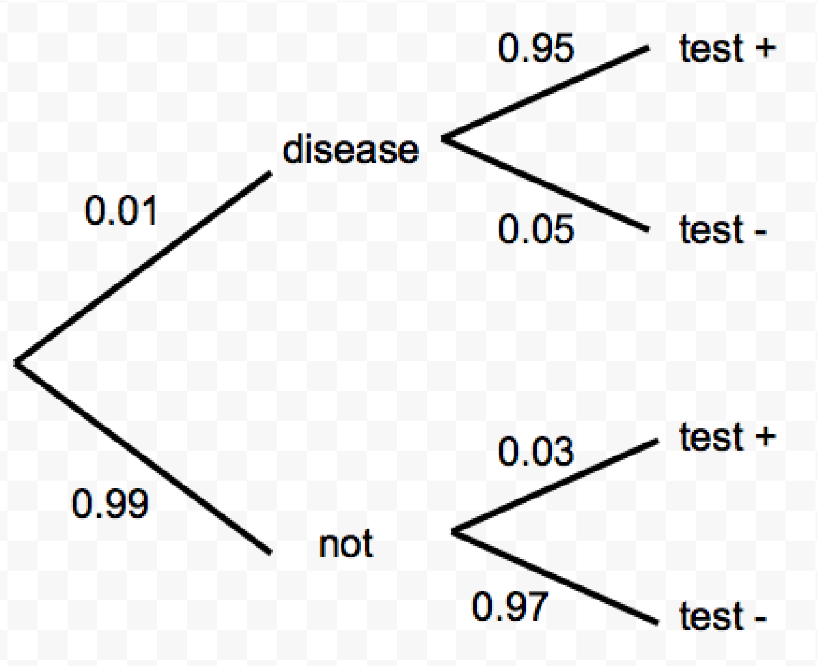
From Jeff Witmer, Introduction to Bayesian Statistics, slides 24 February 2016

Medical test example

Suppose a test is 95% accurate when a disease is present and 97% accurate when the disease is absent. Suppose that 1% of the population has the disease.

What is P(have the disease | test +)?

Medical test example



$$p(\mathrm{dis}\,|\,\mathrm{test+}) = \frac{P(\mathrm{dis})P(\mathrm{test+}\,|\,\mathrm{dis})}{P(\mathrm{dis})P(\mathrm{test+}\,|\,\mathrm{dis}) + P(\text{Ø}\mathrm{dis})P(\mathrm{test+}\,|\,\text{Ø}\mathrm{dis})}$$

$$= \frac{(0.01)(0.95)}{(0.01)(0.95)+(0.99)(0.03)} = \frac{0.0095}{0.0095+0.0297} \text{ » } 0.24$$

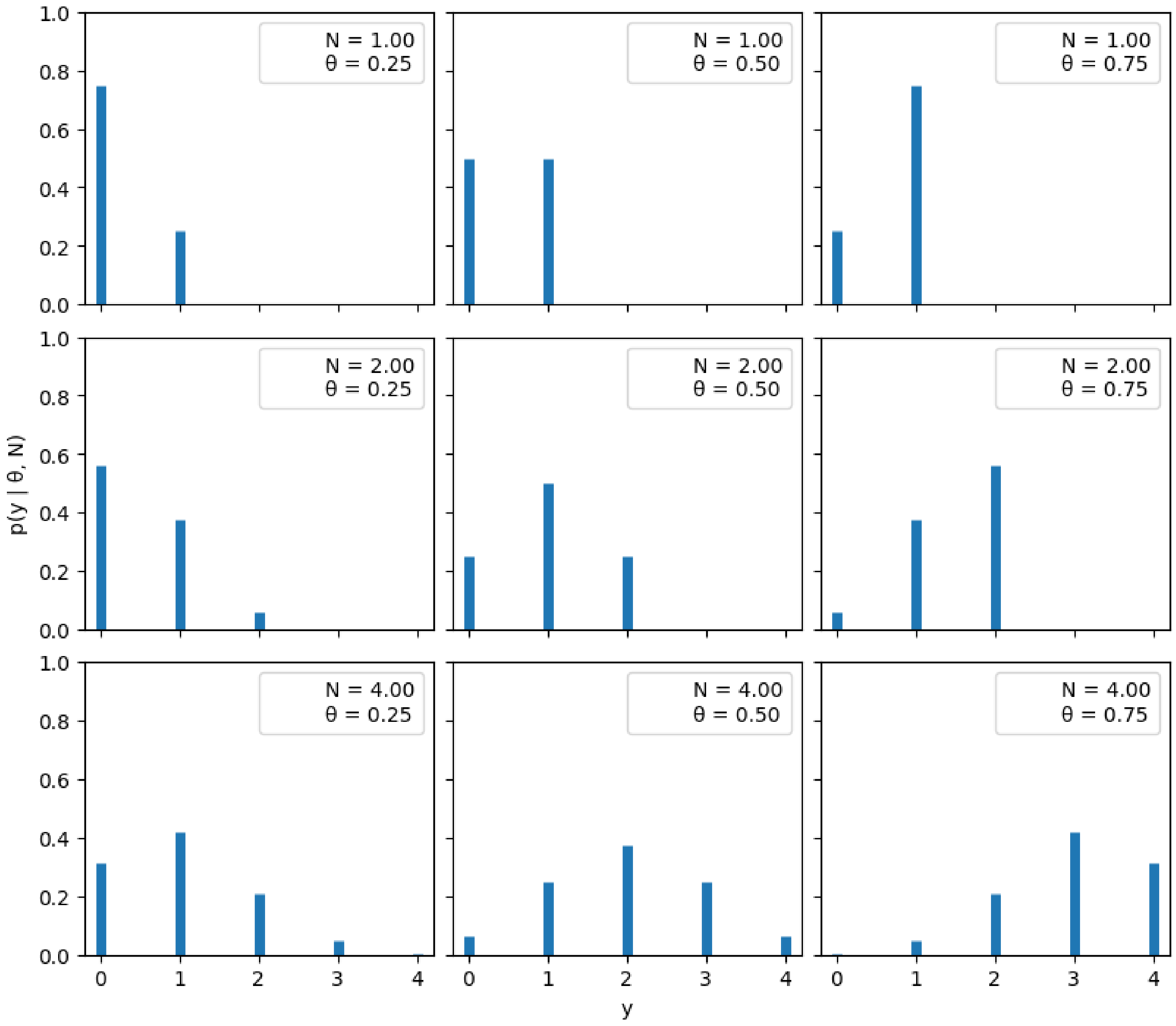
From Jeff Witmer, Introduction to Bayesian Statistics, slides 24 February 2016

Coin Toss

- **Probability of tails:** p
- **Probability of heads:** 1-p

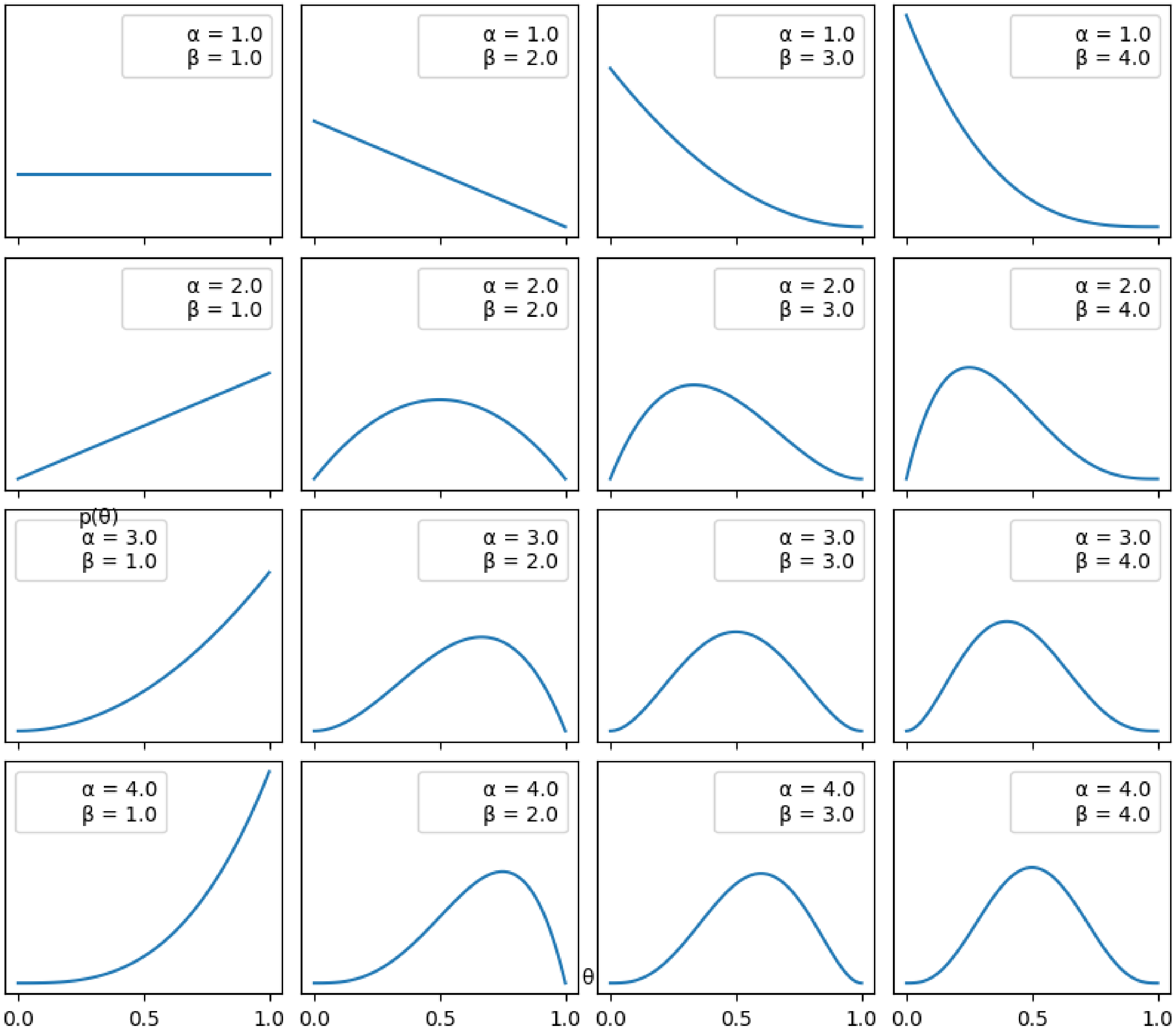
- Probability of getting n tails out of N tosses of coin:

$$C_N^n p^n (1-p)^{N-n}$$



Beta distribution: conjugate to binomial

$$
\begin{aligned}
f(x;\alpha,\beta) &= \mathrm{constant}\cdot x^{\alpha-1}(1-x)^{\beta-1} \\
&= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,x^{\alpha-1}(1-x)^{\beta-1} \\
&= \frac{1}{\mathrm{B}(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}
\end{aligned}
$$



https://en.wikipedia.org/wiki/Beta_distribution

Conjugate distributions

Conjugate Prior: A prior distribution is said to be conjugate to a likelihood function if the resulting posterior distribution is in the same family as the prior. In other words, if you start with a certain type of distribution as your prior, and after observing data and updating your beliefs (via Bayes' theorem), your posterior is still of that same type, then the prior is a conjugate prior for that likelihood function.

• **Computational Convenience**: Using conjugate priors can greatly simplify the mathematical computation required to find the posterior distribution. This can be especially useful in situations where you're continually updating your beliefs with new data; with conjugate priors, you can easily update your posterior without complex integrals or advanced sampling methods.

•**Analytical Solutions**: Many standard problems in Bayesian statistics can be solved analytically using conjugate priors, leading to exact posterior distributions.

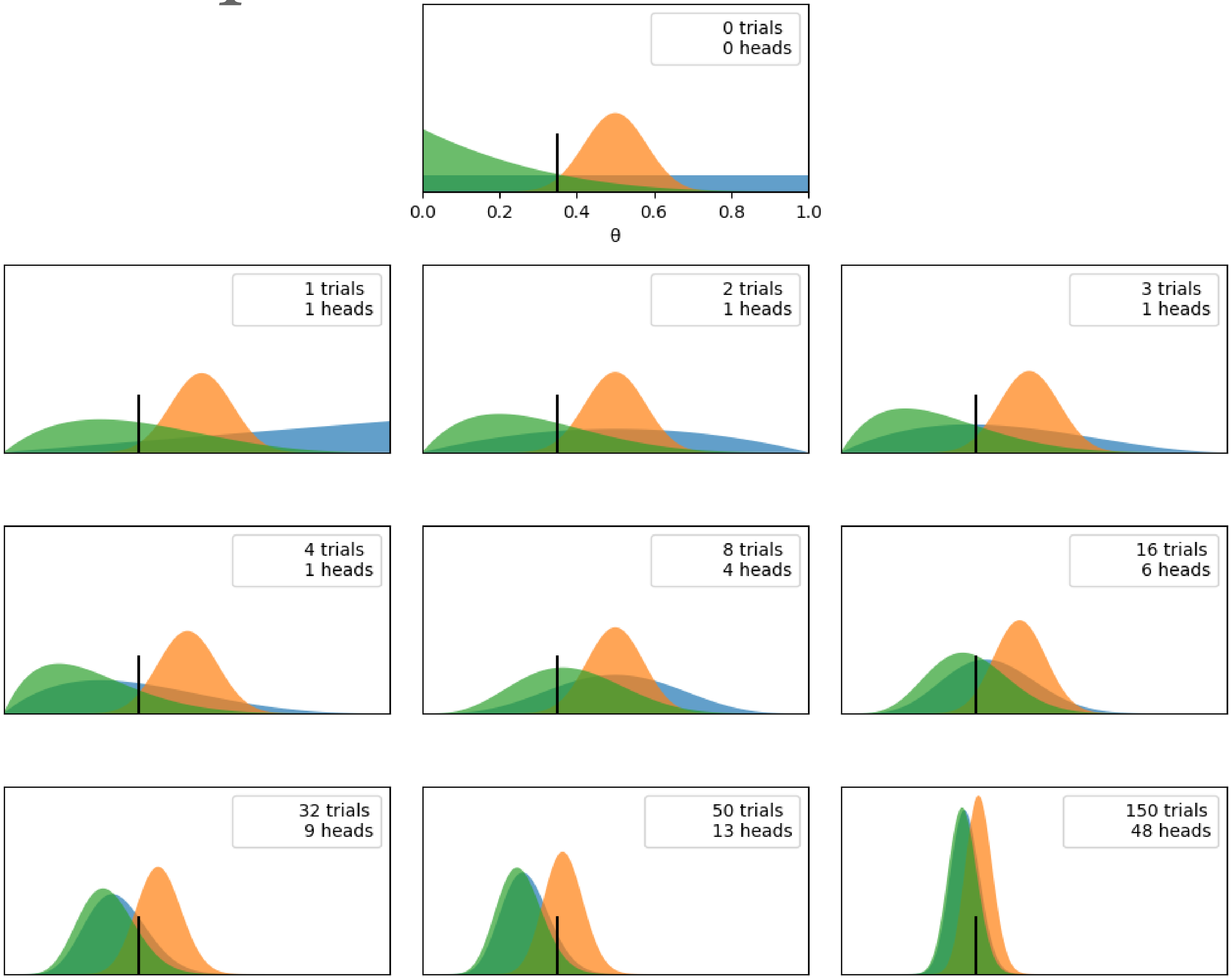
Conjugate distributions

1. **Beta distribution is conjugate to the Binomial likelihood**: This means that if you have a Binomial likelihood (e.g., flipping coins) and a Beta-distributed prior on the probability of heads, the resulting posterior distribution after observing some data will also be a Beta distribution.

2. **Gamma distribution is conjugate to the Poisson likelihood**: If you're observing the number of events occurring in fixed intervals of time or space (modeled by a Poisson distribution) and have a Gamma-distributed prior on the rate parameter, the posterior will also be Gamma-distributed.

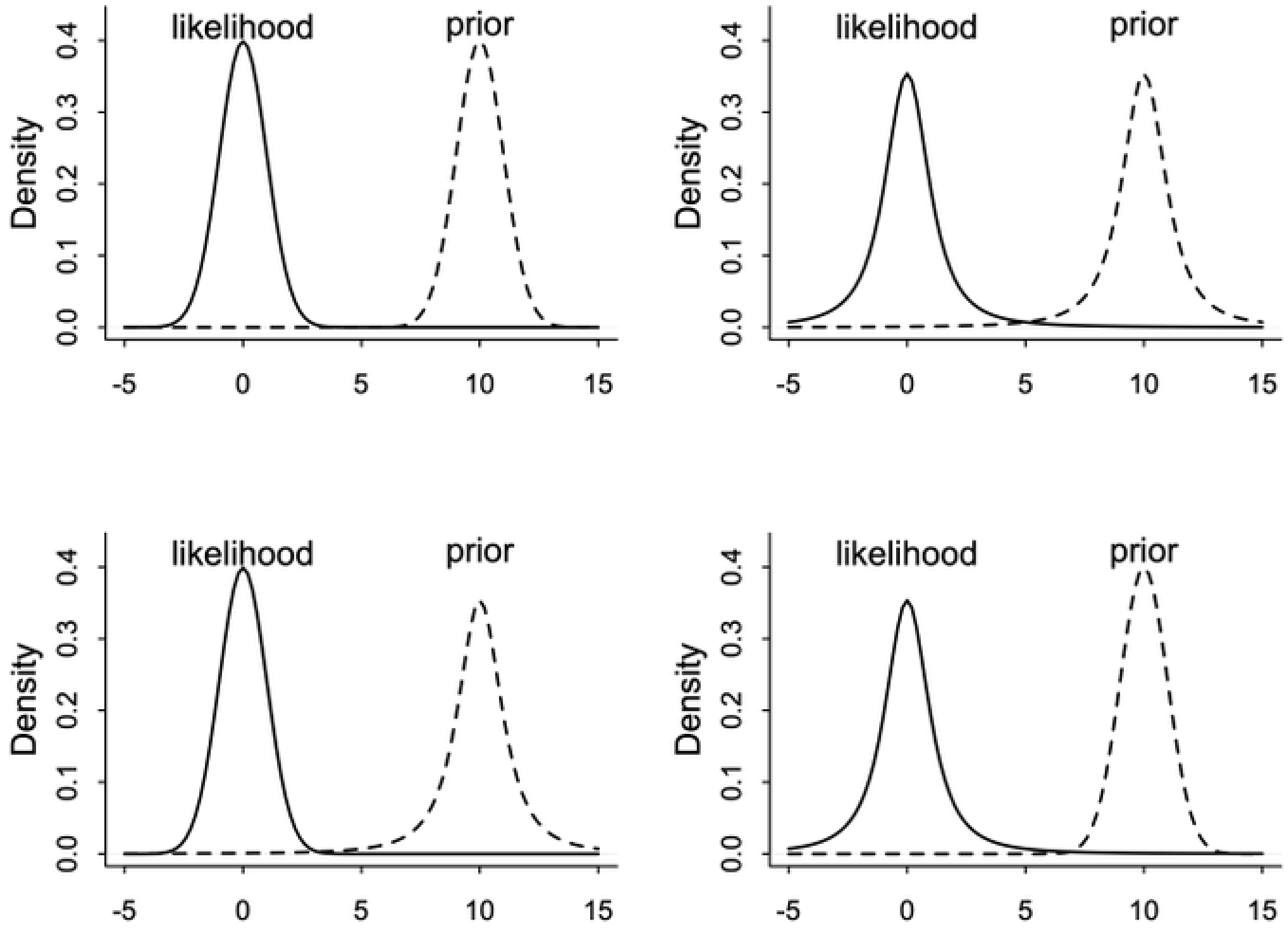
3. **Normal distribution is conjugate to itself**: If both the likelihood and the prior are normally distributed, then the posterior will also be normally distributed.

Can we learn p from several coin tosses?



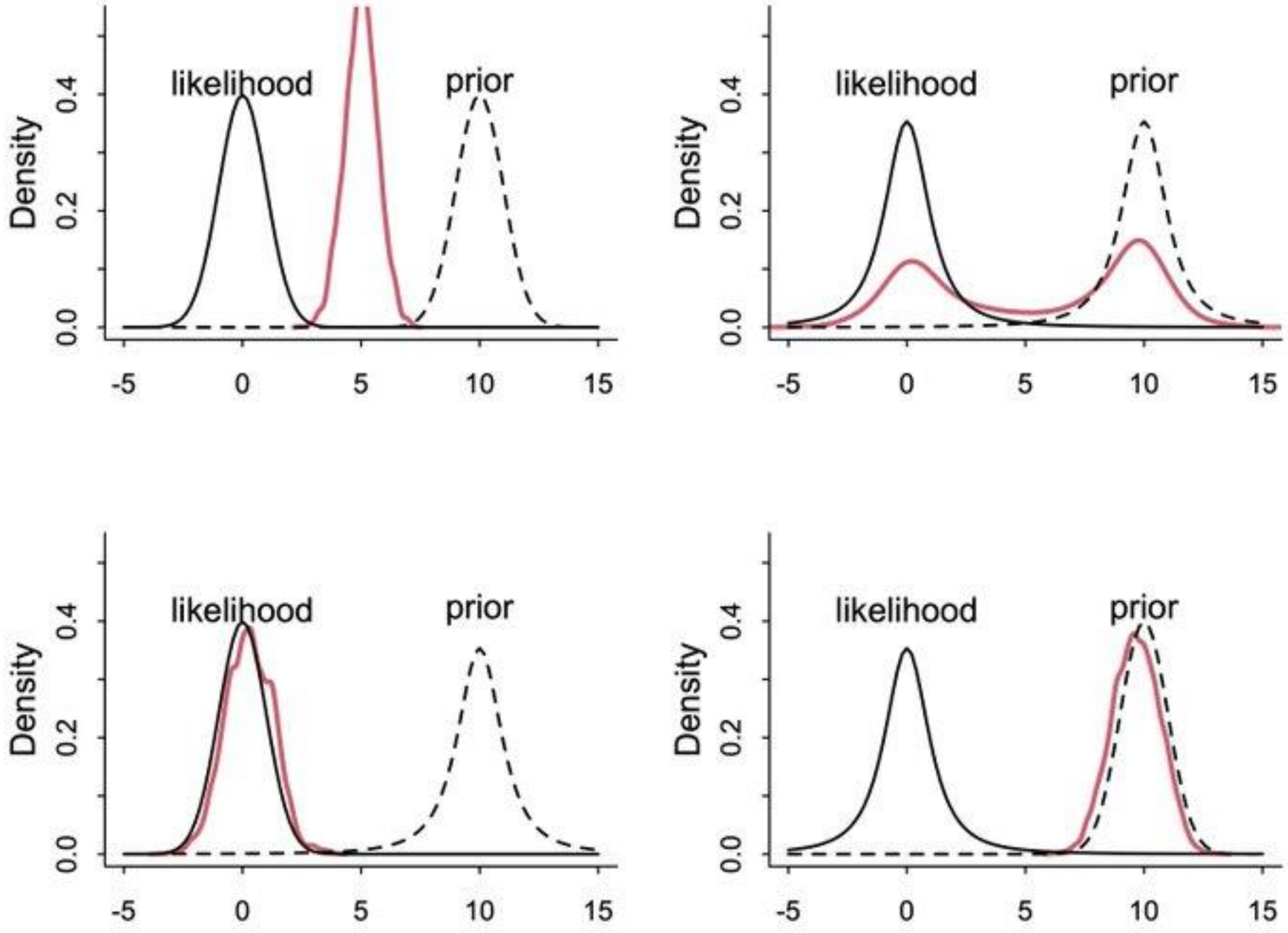
Colab

McElreath Quartet



https://twitter.com/rlmcelreath/status/1701165075493470644

McElreath Quartet



https://twitter.com/rlmcelreath/status/1701165075493470644