

# **Bayesian statistics with R**

## **1. An introduction to Bayesian inference**

---

Olivier Gimenez

November-December 2023

## Credit where credit's due

- Ruth King, Byron Morgan, Steve Brooks (our workshops and [Bayesian analysis for population ecology book](#)).
- Richard McElreath ([Statistical rethinking](#) book and lecture videos).
- Jim Albert and Jingchen Hu ([Probability and Bayesian modelling](#) book).
- Materials shared by [Tristan Marh](#), [Jason Matthiopoulos](#), [Francisco Rodriguez Sanchez](#), [Kerrie Mengerson](#) and [Mark Lai](#).

## Slides, code and data

- All material prepared with R.
- R Markdown used to write reproducible material.
- Dedicated website <https://oliviergimenez.github.io/bayesian-stats-with-R/>.

# Objectives

- Try and demystify Bayesian statistics, and what we call MCMC.
- Make the difference between Bayesian and Frequentist analyses.
- Understand the Methods section of ecological papers doing Bayesian stuff.
- Run Bayesian analyses, safely hopefully.

**BRACE YOURSELF**



## What is on our plate?

1. An introduction to Bayesian inference
2. The likelihood
3. Bayesian analyses by hand
4. A detour to explore priors
5. Markov chains Monte Carlo methods (MCMC)
6. Bayesian analyses in R with the Jags software
7. Contrast scientific hypotheses with model selection
8. Heterogeneity and multilevel models (aka mixed models)

I want mooooore

---

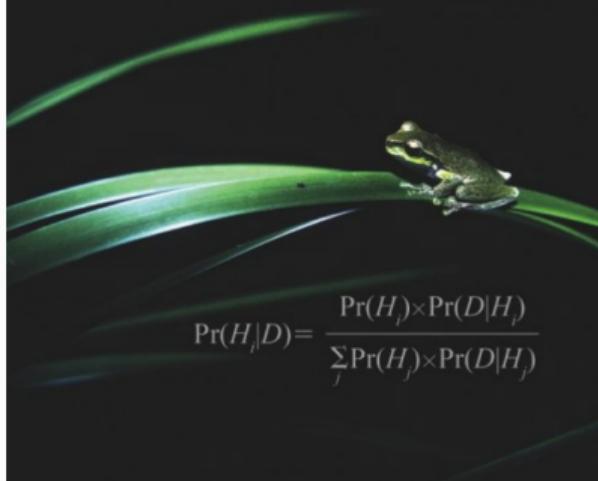
I ONLY LIKE TWO  
THINGS:

THEY'RE BOTH BOOKS



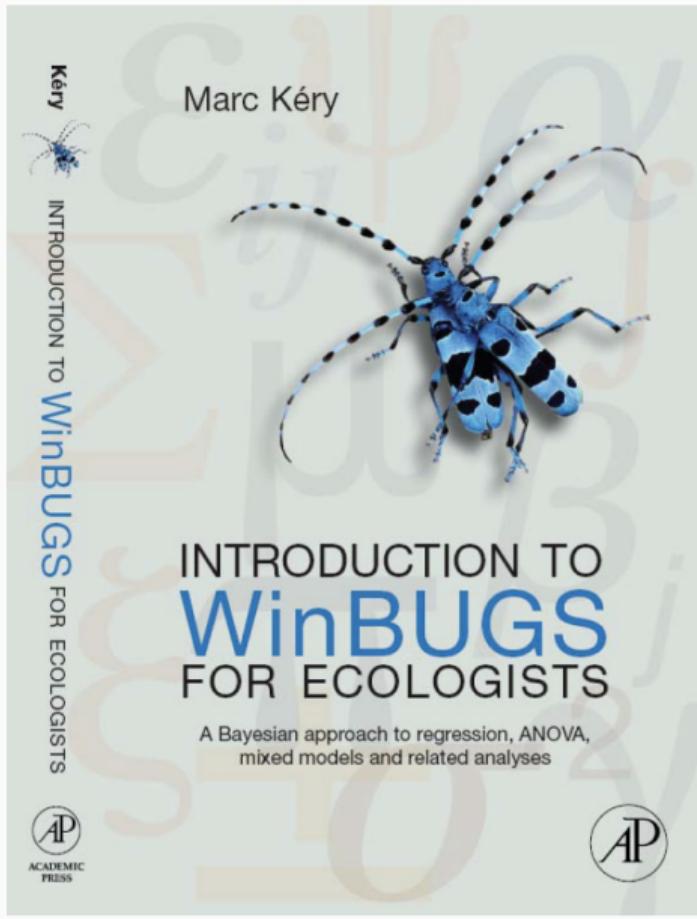
# Bayesian Methods for Ecology

Michael A. McCarthy



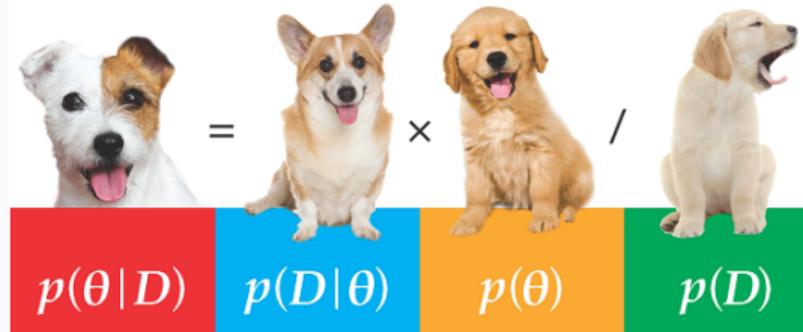
$$\Pr(H_i|D) = \frac{\Pr(H_i) \times \Pr(D|H_i)}{\sum_j \Pr(H_j) \times \Pr(D|H_j)}$$

CAMBRIDGE



# Doing Bayesian Data Analysis

A Tutorial with R, JAGS, and Stan



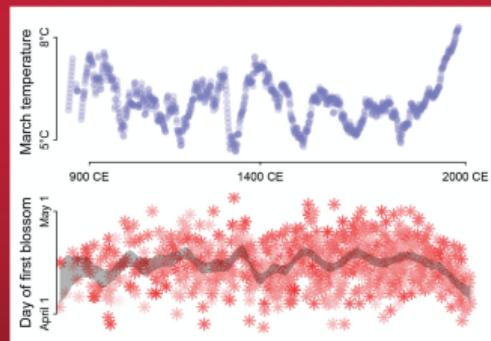
John K. Kruschke



Texts in Statistical Science

# Statistical Rethinking

A Bayesian Course  
with Examples in R and Stan  
**SECOND EDITION**

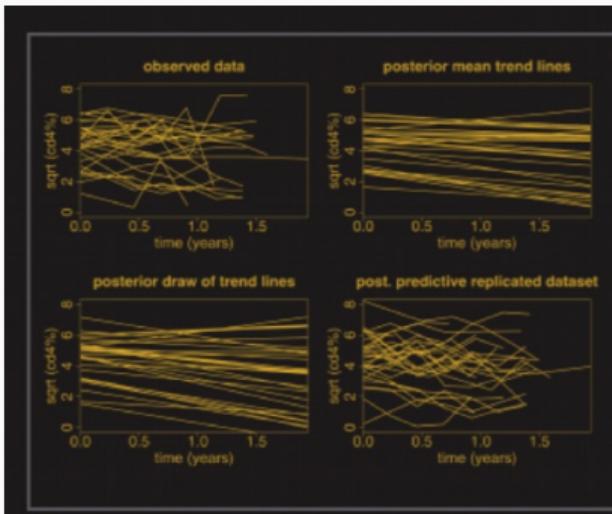


Richard McElreath



Taylor & Francis Group

A CHAPMAN & HALL BOOK

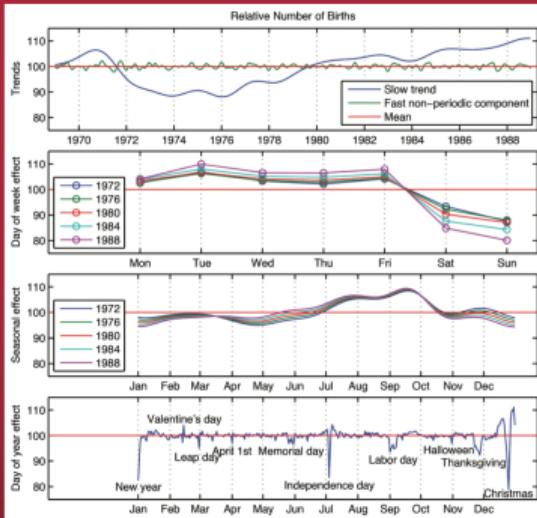


# Data Analysis Using Regression and Multilevel/Hierarchical Models

ANDREW GELMAN  
JENNIFER HILL

# Bayesian Data Analysis

Third Edition



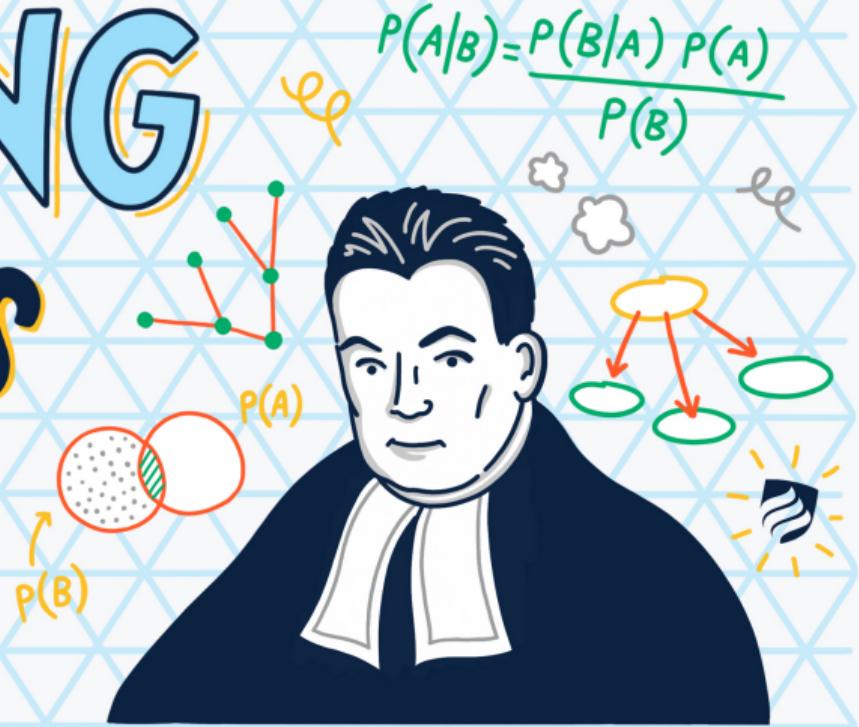
Andrew Gelman, John B. Carlin, Hal S. Stern,  
David B. Dunson, Aki Vehtari, and Donald B. Rubin

Free at <http://www.stat.columbia.edu/~gelman/book/>

## **What is Bayesian inference?**

---

# THE AMAZING Thomas Bayes



## A reminder on conditional probabilities

- $\Pr(A | B)$ : Probability of A given B

## A reminder on conditional probabilities

- $\Pr(A | B)$ : Probability of A given B
- The ordering matters:  $\Pr(A | B)$  is not the same as  $\Pr(B | A)$ .

## A reminder on conditional probabilities

- $\Pr(A | B)$ : Probability of A given B
- The ordering matters:  $\Pr(A | B)$  is not the same as  $\Pr(B | A)$ .
- $\Pr(A | B) = \frac{\Pr(A \text{ and } B)}{\Pr(B)}$



**HOW TO CURE VAMPIRES?**

## Screening for vampirism

- The chance of the test being positive given you are a vampire is  
 $\Pr(+|\text{vampire}) = 0.90$  (**sensitivity**).

## Screening for vampirism

- The chance of the test being positive given you are a vampire is  $\Pr(+|\text{vampire}) = 0.90$  (**sensitivity**).
- The chance of a negative test given you are mortal is  $\Pr(-|\text{mortal}) = 0.95$  (**specificity**).

## What is the question?

- From the perspective of the test: Given a person is a vampire, what is the probability that the test is positive?  $\Pr(+|\text{vampire}) = 0.90$ .

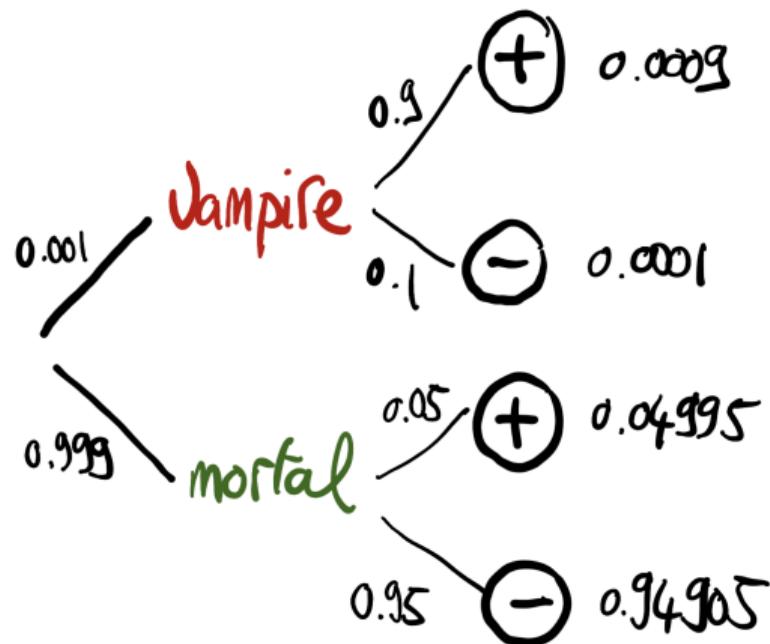
## What is the question?

- From the perspective of the test: Given a person is a vampire, what is the probability that the test is positive?  $\Pr(+|\text{vampire}) = 0.90$ .
- From the perspective of a person: Given that the test is positive, what is the probability that this person is a vampire?  $\Pr(\text{vampire}|+) = ?$

## What is the question?

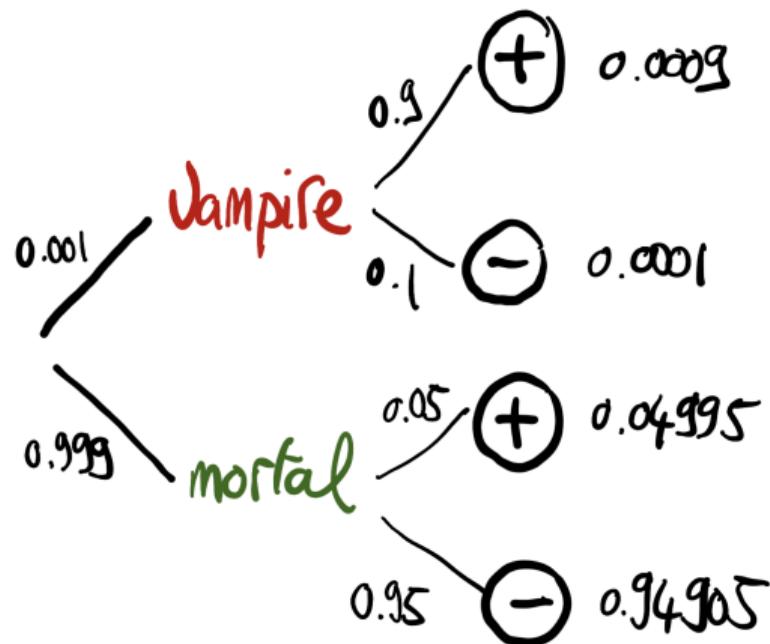
- From the perspective of the test: Given a person is a vampire, what is the probability that the test is positive?  $\Pr(+|\text{vampire}) = 0.90$ .
- From the perspective of a person: Given that the test is positive, what is the probability that this person is a vampire?  $\Pr(\text{vampire}|+) = ?$
- Assume that vampires are rare, and represent only 0.1% of the population. This means that  $\Pr(\text{vampire}) = 0.001$ .

What is the answer? Bayes' theorem to the rescue!



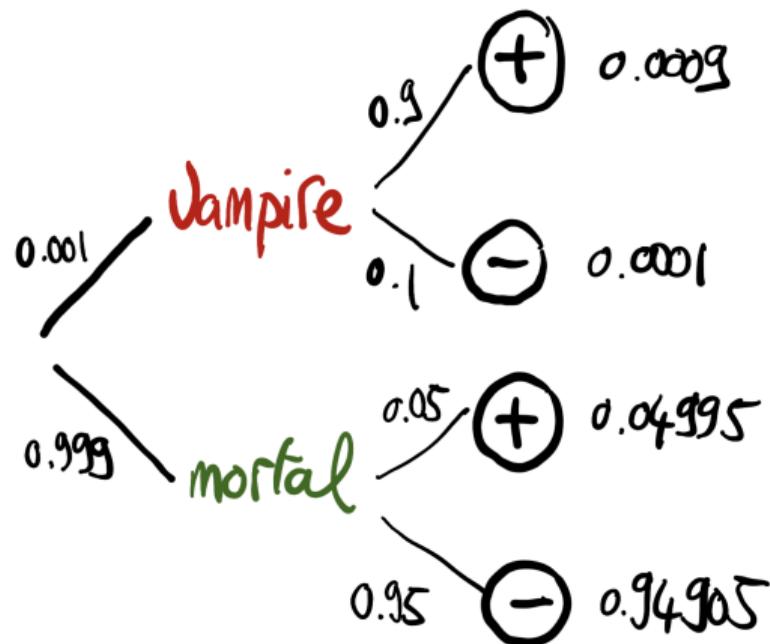
- $\Pr(\text{vampire}|+) = \frac{\Pr(\text{vampire and } +)}{\Pr(+)}$

## What is the answer? Bayes' theorem to the rescue!



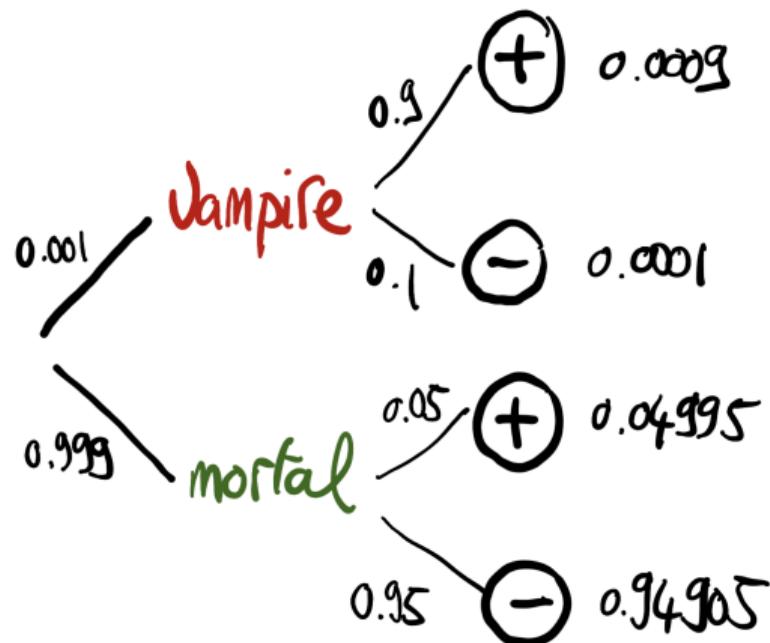
- $\Pr(\text{vampire}|+) = \frac{\Pr(\text{vampire and } +)}{\Pr(+)}$
- $\Pr(\text{vampire and } +) = \Pr(\text{vampire}) \Pr(+|\text{vampire}) = 0.0009$

## What is the answer? Bayes' theorem to the rescue!



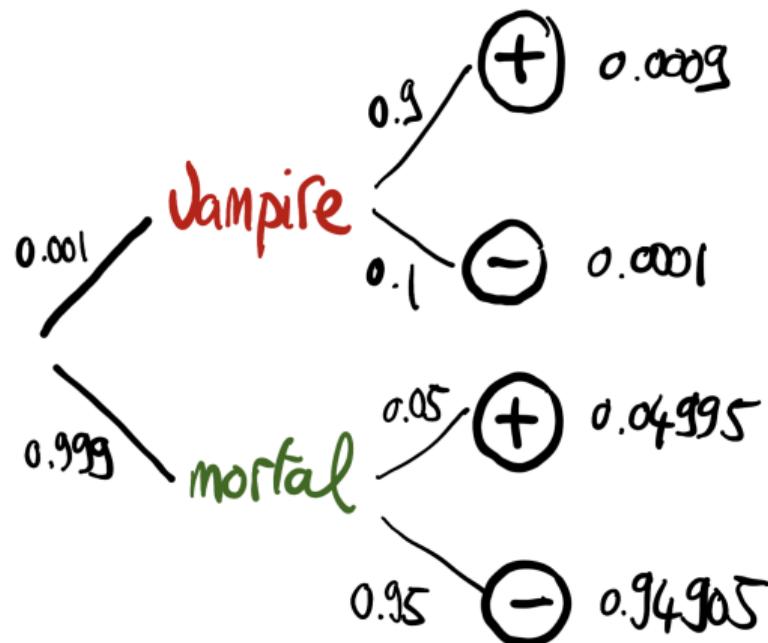
- $\Pr(\text{vampire} | +) = \frac{\Pr(\text{vampire and } +)}{\Pr(+)}$
- $\Pr(\text{vampire and } +) = \Pr(\text{vampire}) \Pr(+|\text{vampire}) = 0.0009$
- $\Pr(+) = 0.0009 + 0.04995 = 0.05085$

## What is the answer? Bayes' theorem to the rescue!



- $\Pr(\text{vampire}|+) = \frac{\Pr(\text{vampire and } +)}{\Pr(+)}$
- $\Pr(\text{vampire and } +) = \Pr(\text{vampire}) \Pr(+|\text{vampire}) = 0.0009$
- $\Pr(+) = 0.0009 + 0.04995 = 0.05085$
- $\Pr(\text{vampire}|+) = 0.0009 / 0.05085 = 0.02$

## What is the answer? Bayes' theorem to the rescue!



- $\Pr(\text{vampire}|+) = \frac{\Pr(\text{vampire and } +)}{\Pr(+)}$
- $\Pr(\text{vampire and } +) = \Pr(\text{vampire}) \Pr(+|\text{vampire}) = 0.0009$
- $\Pr(+) = 0.0009 + 0.04995 = 0.05085$
- $\Pr(\text{vampire}|+) = 0.0009 / 0.05085 = 0.02$

$$\Pr(\text{vampire}|+) = \frac{\Pr(+|\text{vampire}) \Pr(\text{vampire})}{\Pr(+)}$$

## Your turn: Practical 1

---

## Bayes' theorem

- A theorem about conditional probabilities.
- $\Pr(B | A) = \frac{\Pr(A | B) \Pr(B)}{\Pr(A)}$

$$\Pr(A|B) = \frac{\Pr(B|A) \Pr(A)}{\Pr(B)}$$

## Bayes' theorem

- Easy to mess up with letters. Might be easier to remember when written like this:

$$\Pr(\text{hypothesis} \mid \text{data}) = \frac{\Pr(\text{data} \mid \text{hypothesis}) \Pr(\text{hypothesis})}{\Pr(\text{data})}$$

## Bayes' theorem

- Easy to mess up with letters. Might be easier to remember when written like this:

$$\Pr(\text{hypothesis} \mid \text{data}) = \frac{\Pr(\text{data} \mid \text{hypothesis}) \Pr(\text{hypothesis})}{\Pr(\text{data})}$$

- The “hypothesis” is typically something unobserved or unknown. It’s what you want to learn about using the data.

## Bayes' theorem

- Easy to mess up with letters. Might be easier to remember when written like this:

$$\Pr(\text{hypothesis} \mid \text{data}) = \frac{\Pr(\text{data} \mid \text{hypothesis}) \Pr(\text{hypothesis})}{\Pr(\text{data})}$$

- The “hypothesis” is typically something unobserved or unknown. It’s what you want to learn about using the data.
- For regression models, the “hypothesis” is a parameter (intercept, slopes or error terms).

## Bayes' theorem

- Easy to mess up with letters. Might be easier to remember when written like this:

$$\Pr(\text{hypothesis} \mid \text{data}) = \frac{\Pr(\text{data} \mid \text{hypothesis}) \Pr(\text{hypothesis})}{\Pr(\text{data})}$$

- The “hypothesis” is typically something unobserved or unknown. It’s what you want to learn about using the data.
- For regression models, the “hypothesis” is a parameter (intercept, slopes or error terms).
- Bayes theorem tells you the probability of the hypothesis given the data.

## What is doing science after all?

How plausible is some hypothesis given the data?

$$\Pr(\text{hypothesis} \mid \text{data}) = \frac{\Pr(\text{data} \mid \text{hypothesis}) \Pr(\text{hypothesis})}{\Pr(\text{data})}$$

## Why is Bayesian statistics not the default?

- Due to practical problems of implementing the Bayesian approach, and some wars of male statisticians's egos, little advance was made for over two centuries.

## Why is Bayesian statistics not the default?

- Due to practical problems of implementing the Bayesian approach, and some wars of male statisticians's egos, little advance was made for over two centuries.
- Recent advances in computational power coupled with the development of new methodology have led to a great increase in the application of Bayesian methods within the last two decades.

## Frequentist versus Bayesian

- Typical stats problems involve estimating parameter  $\theta$  with available data.

## Frequentist versus Bayesian

- Typical stats problems involve estimating parameter  $\theta$  with available data.
- The frequentist approach (**maximum likelihood estimation** – MLE) assumes that the parameters are fixed, but have unknown values to be estimated.

## Frequentist versus Bayesian

- Typical stats problems involve estimating parameter  $\theta$  with available data.
- The frequentist approach (**maximum likelihood estimation** – MLE) assumes that the parameters are fixed, but have unknown values to be estimated.
- Classical estimates generally provide a point estimate of the parameter of interest.

## Frequentist versus Bayesian

- Typical stats problems involve estimating parameter  $\theta$  with available data.
- The frequentist approach (**maximum likelihood estimation – MLE**) assumes that the parameters are fixed, but have unknown values to be estimated.
- Classical estimates generally provide a point estimate of the parameter of interest.
- The Bayesian approach assumes that the parameters are not fixed but have some fixed unknown distribution - a distribution for the parameter.

## What is the Bayesian approach?

- The approach is based upon the idea that the experimenter begins with some prior beliefs about the system.

## What is the Bayesian approach?

- The approach is based upon the idea that the experimenter begins with some prior beliefs about the system.
- And then updates these beliefs on the basis of observed data.

## What is the Bayesian approach?

- The approach is based upon the idea that the experimenter begins with some prior beliefs about the system.
- And then updates these beliefs on the basis of observed data.
- This updating procedure is based upon the Bayes' Theorem:

$$\Pr(A | B) = \frac{\Pr(B | A) \Pr(A)}{\Pr(B)}$$

## What is the Bayesian approach?

- Schematically if  $A = \theta$  and  $B = \text{data}$ , then

## What is the Bayesian approach?

- Schematically if  $A = \theta$  and  $B = \text{data}$ , then
- The Bayes' theorem

$$\Pr(A | B) = \frac{\Pr(B | A) \Pr(A)}{\Pr(B)}$$

## What is the Bayesian approach?

- Schematically if  $A = \theta$  and  $B = \text{data}$ , then
- The Bayes' theorem

$$\Pr(A | B) = \frac{\Pr(B | A) \Pr(A)}{\Pr(B)}$$

- Translates into:

$$\Pr(\theta | \text{data}) = \frac{\Pr(\text{data} | \theta) \Pr(\theta)}{\Pr(\text{data})}$$

## Bayes' theorem

$$\Pr(\theta \mid \text{data}) = \frac{\Pr(\text{data} \mid \theta) \Pr(\theta)}{\Pr(\text{data})}$$

## Bayes' theorem

$$\Pr(\theta \mid \text{data}) = \frac{\Pr(\text{data} \mid \theta) \Pr(\theta)}{\Pr(\text{data})}$$

- **Posterior distribution:** Represents what you know after having seen the data. The basis for inference, a distribution, possibly multivariate if more than one parameter ( $\theta$ ).

## Bayes' theorem

$$\Pr(\theta \mid \text{data}) = \frac{\Pr(\text{data} \mid \theta) \Pr(\theta)}{\Pr(\text{data})}$$

- **Posterior distribution:** Represents what you know after having seen the data. The basis for inference, a distribution, possibly multivariate if more than one parameter ( $\theta$ ).
- **Likelihood:** We know that quantity, same as in the MLE approach.

## Bayes' theorem

$$\Pr(\theta \mid \text{data}) = \frac{\Pr(\text{data} \mid \theta) \Pr(\theta)}{\Pr(\text{data})}$$

- **Posterior distribution:** Represents what you know after having seen the data. The basis for inference, a distribution, possibly multivariate if more than one parameter ( $\theta$ ).
- **Likelihood:** We know that quantity, same as in the MLE approach.
- **Prior distribution:** Represents what you know before seeing the data. The source of much discussion about the Bayesian approach.

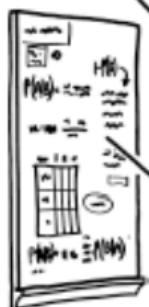
## Bayes' theorem

$$\Pr(\theta \mid \text{data}) = \frac{\Pr(\text{data} \mid \theta) \Pr(\theta)}{\Pr(\text{data})}$$

- **Posterior distribution:** Represents what you know after having seen the data. The basis for inference, a distribution, possibly multivariate if more than one parameter ( $\theta$ ).
- **Likelihood:** We know that quantity, same as in the MLE approach.
- **Prior distribution:** Represents what you know before seeing the data. The source of much discussion about the Bayesian approach.
- **$\Pr(\text{data}) = \int \Pr(\text{data} \mid \theta) \Pr(\theta) d\theta$ :** Possibly high-dimensional integral, difficult if not impossible to calculate. This is one of the reasons why we need simulation (MCMC) methods - more soon.

GIVEN THESE PREVALENCES,  
IS IT LIKELY THAT THE TEST  
RESULT IS A FALSE POSITIVE?

WELL, THIS CHAPTER IS ON  
BAYES' THEOREM, SO YES.



SOMETIMES, IF YOU UNDERSTAND  
BAYES' THEOREM WELL ENOUGH,  
YOU DON'T NEED IT.