# Introduction to Bayesian statistics with R
# 1. Motivations

Olivier Gimenez

last updated: 2025-03-10

# This workshop

# 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.

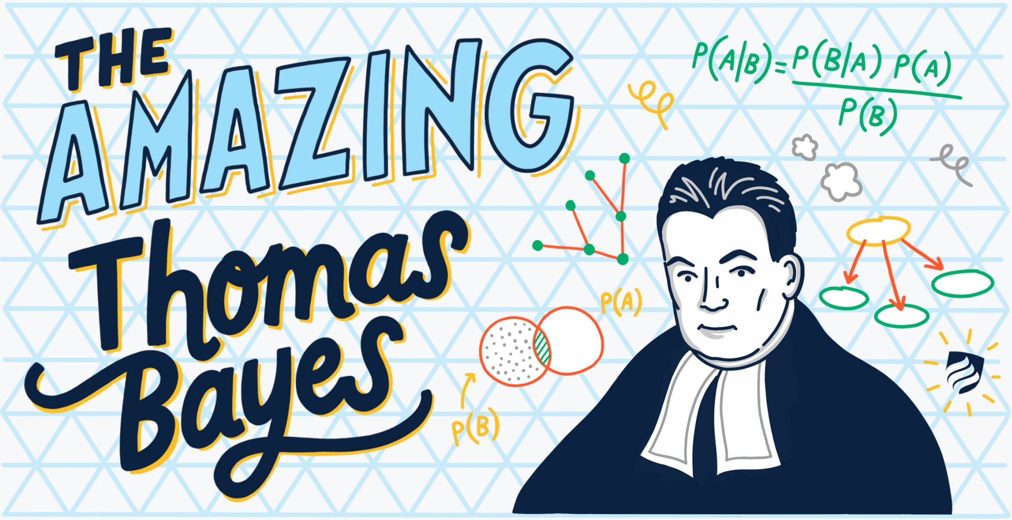
# What's on our plate?

- Section 1 - Motivation and Bayes theorem.
- Section 2 - Markov chain Monte Carlo algorithms (MCMC).
- Section 3 - Introduction to `NIMBLE` and `brms`.
- Section 4 - Priors.
- Section 5 - Case studies and GLMMs.
- Section 6 - Conclusions.

# Material

- Material from previous workshops (https://oliviergimenez.github.io/bayesian-stats-with-R/), a book in progress (https://oliviergimenez.github.io/banana-book/) and a paper on ten quick tips to get you started with Bayesian statistics (https://hal.science/CEFE/hal-04731240v2).
- All material prepared with `R`, with `R Markdown` to ensure reproducibility.
- Slides, data and R codes on GitHub at https://github.com/oliviergimenez/bayes-workshop

# What is Bayesian inference?


THE AMAZING
Thomas Bayes
$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$
P(A)
P(B)

# A reminder on conditional probabilities

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

# A reminder on conditional probabilities

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

# A reminder on conditional probabilities

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


HOW TO CURE VAMPIRES?
imgflip.com

## 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? $\Pr(+|\text{vampire}) = 0.90$.

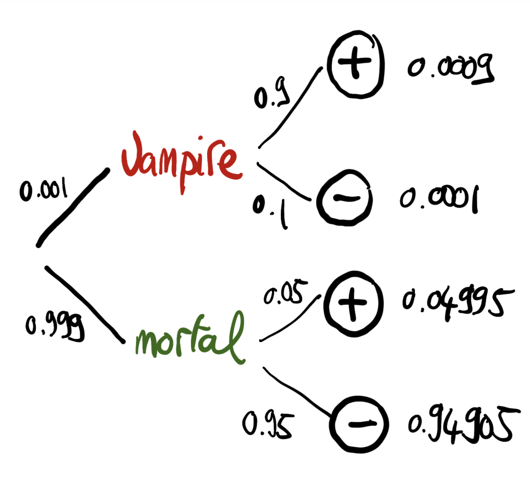
# What is the question?

- 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?

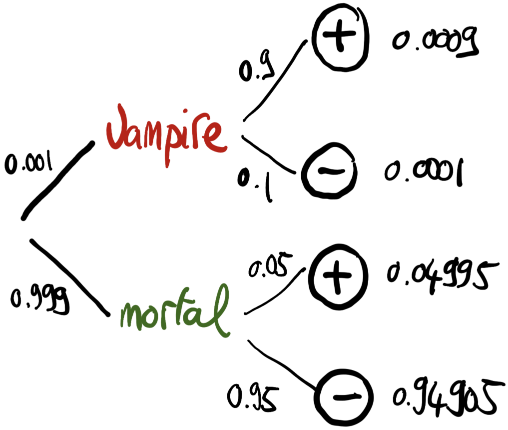
- From the perspective of the test: Given a person is a vampire, what is the probability that the test is positive? $\Pr(+|\text{vampire}) = 0.90$.
- From the perspective of a person: Given that the test is positive, what is the probability that this person is a vampire? $\Pr(\text{vampire}|+) = ?$
-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



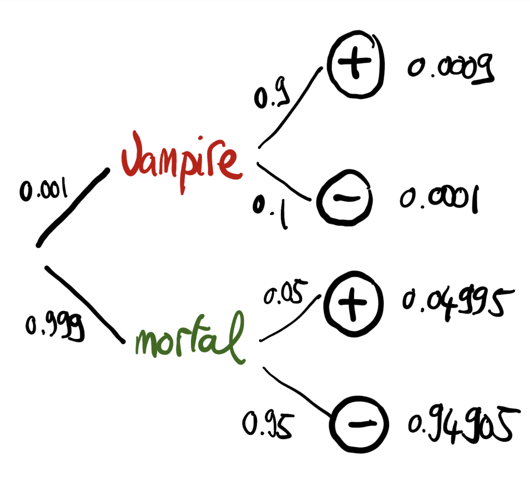

- $\Pr(\text{vampire}|+) = \dfrac{\Pr(\text{vampire and } +)}{\Pr(+)}$

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



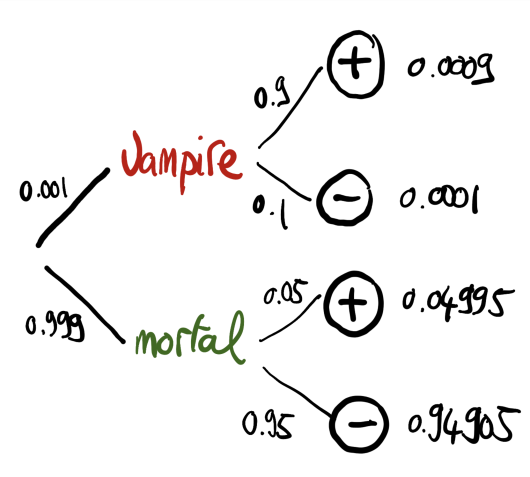

- $\Pr(\text{vampire}|+) = \dfrac{\Pr(\text{vampire and } +)}{\Pr(+)}$
- $\Pr(\text{vampire and } +) =$
  $\Pr(\text{vampire})\ \Pr(+|\text{vampire}) = 0.0009$

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



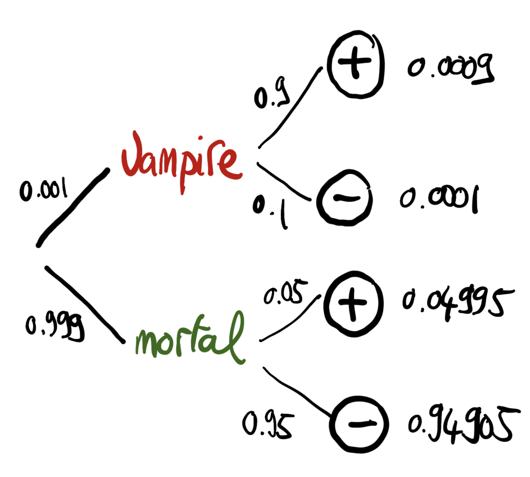

- $\Pr(\text{vampire}|+) = \dfrac{\Pr(\text{vampire and } +)}{\Pr(+)}$
- $\Pr(\text{vampire and } +) =$ $\Pr(\text{vampire})\ \Pr(+|\text{vampire}) = 0.0009$
- $\Pr(+) = 0.0009 + 0.04995 = 0.05085$

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




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

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

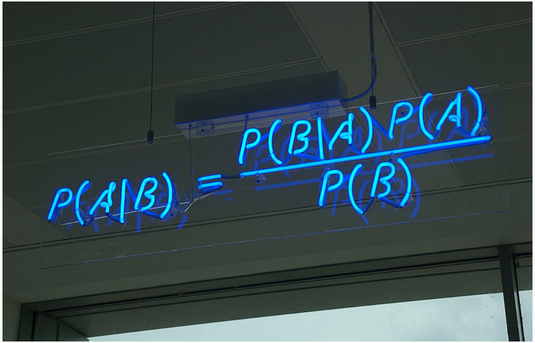



- $\Pr(\text{vampire}|+) = \dfrac{\Pr(\text{vampire and } +)}{\Pr(+)}$
- $\Pr(\text{vampire and } +) = \Pr(\text{vampire})\ \Pr(+|\text{vampire}) = 0.0009$
- $\Pr(+) = 0.0009 + 0.04995 = 0.05085$
- $\Pr(\text{vampire}|+) = 0.0009/0.05085 = 0.02$

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

# Bayes' theorem

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



## 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:

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

# What is the Bayesian approach?

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

# What is the Bayesian approach?

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

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

# What is the Bayesian approach?

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

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

- Translates into:

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

# Bayes' theorem

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

# 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.

# Guiding example

# Survival example

- 120 deer were radio-tracked over winter.
- 61 close to a plant, 59 far from any human activity.
- How to estimate winter survival?

| | Released | Alive | Dead | Other |
|---|---|---|---|---|
| treatment | 61 | 19 | 38 | 4 |
| control | 59 | 21 | 38 | 0 |

- So, $n = 57$ deer were assigned to the treatment group of which $k = 19$ survived the winter.

- So, $n = 57$ deer were assigned to the treatment group of which $k = 19$ survived the winter.
- Of interest is the probability of winter survival, call it $\theta$, for the general population within the treatment area.

- So, $n = 57$ deer were assigned to the treatment group of which $k = 19$ survived the winter.
- Of interest is the probability of winter survival, call it $\theta$, for the general population within the treatment area.
- The obvious estimate is simply to take the ratio $k/n = 19/57$.

- So, $n = 57$ deer were assigned to the treatment group of which $k = 19$ survived the winter.
- Of interest is the probability of winter survival, call it $\theta$, for the general population within the treatment area.
- The obvious estimate is simply to take the ratio $k/n = 19/57$.
- How would the classical statistician justify this estimate?

- Our model is that we have a Binomial experiment (assuming independent and identically distributed draws from the population).

- Our model is that we have a Binomial experiment (assuming independent and identically distributed draws from the population).
- $K$ the number of alive individuals at the end of the winter, so that $P(K = k) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$.

- Our model is that we have a Binomial experiment (assuming independent and identically distributed draws from the population).
- $K$ the number of alive individuals at the end of the winter, so that $P(K=k) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$.
- The classical approach is to maximise the corresponding likelihood with respect to $\theta$ to obtain the entirely plausible MLE:

$$\hat{\theta} = k/n = 19/57$$

.

# The Bayesian approach

- The Bayesian starts off with a prior.

# The Bayesian approach

- The Bayesian starts off with a prior.
- Now, the one thing we know about $\theta$ is that is a continuous random variable and that it lies between zero and one.

# The Bayesian approach

- The Bayesian starts off with a prior.
- Now, the one thing we know about $\theta$ is that is a continuous random variable and that it lies between zero and one.
- Thus, a suitable prior distribution might be the Beta defined on $[0, 1]$.

# The Bayesian approach

- The Bayesian starts off with a prior.
- Now, the one thing we know about $\theta$ is that is a continuous random variable and that it lies between zero and one.
- Thus, a suitable prior distribution might be the Beta defined on $[0, 1]$.
- What is the Beta distribution?

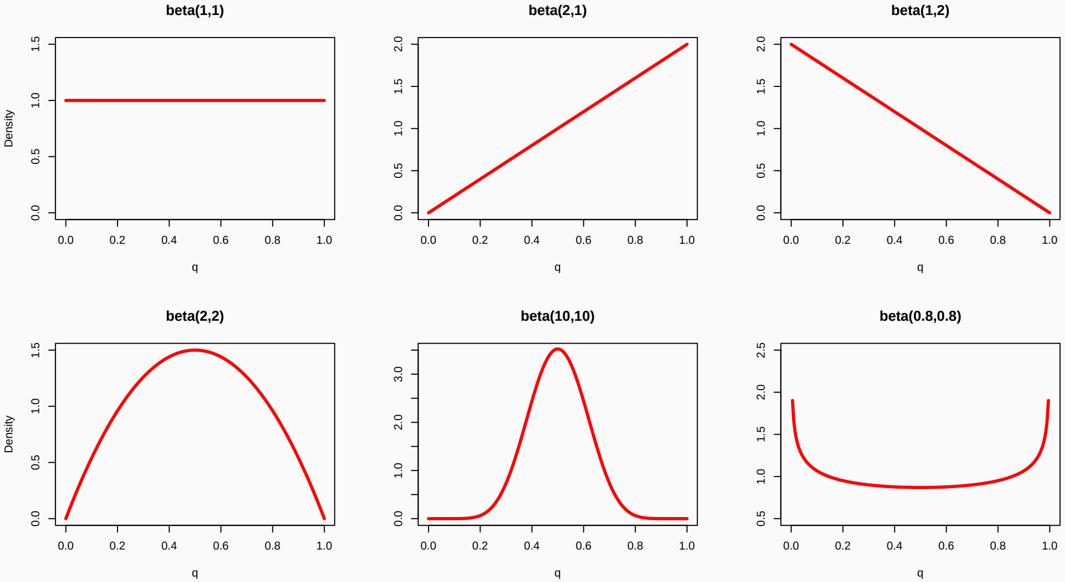
# What is the Beta distribution?

$$q(\theta \mid a, b) = \frac{1}{\mathrm{Beta}(a, b)} \theta^{a-1} (1 - \theta)^{b-1}$$

with $\mathrm{Beta}(a, b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}$ and $\Gamma(n) = (n - 1)!$

# What is the Beta distribution?

- The beta distribution is continuous between 0 and 1, and extends the uniform distribution to situations where not all outcomes are equally likely.
- It has two parameters $a$ and $b$ that control its shape.
- Note that for $a = b = 1$, we get the uniform distribution between 0 and 1.
- The expectation (or mean) of a beta($a$,$b$) is $\frac{a}{a+b}$.


beta(1,1)
beta(2,1)
beta(1,2)
beta(2,2)
beta(10,10)
beta(0.8,0.8)
Density
q

## The Bayesian approach

- We assume a priori that $\theta \sim \mathit{Beta}(a, b)$ so that $\Pr(\theta) = \theta^{a-1}(1-\theta)^{b-1}$

## The Bayesian approach

- We assume a priori that $\theta \sim Beta(a, b)$ so that $\Pr(\theta) = \theta^{a-1}(1-\theta)^{b-1}$
- Then we have:

$$\begin{aligned} Pr(\theta \mid k) &\propto \binom{n}{k}\theta^k(1-\theta)^{n-k}\ \theta^{a-1}(1-\theta)^{b-1} \\ &\propto \theta^{(a+k)-1}(1-\theta)^{(b+n-k)-1} \end{aligned}$$

## The Bayesian approach

- We assume a priori that $\theta \sim Beta(a, b)$ so that $\Pr(\theta) = \theta^{a-1}(1-\theta)^{b-1}$
- Then we have:

$$\begin{aligned} Pr(\theta \mid k) &\propto \binom{n}{k}\theta^{k}(1-\theta)^{n-k}\ \theta^{a-1}(1-\theta)^{b-1} \\ &\propto \theta^{(a+k)-1}(1-\theta)^{(b+n-k)-1} \end{aligned}$$

- That is:

$$\theta \mid k \sim Beta(a+k, b+n-k)$$

## The Bayesian approach

- We assume a priori that $\theta \sim Beta(a, b)$ so that $\Pr(\theta) = \theta^{a-1}(1-\theta)^{b-1}$
- Then we have:

$$\begin{aligned} Pr(\theta \mid k) &\propto \binom{n}{k}\theta^{k}(1-\theta)^{n-k}\ \theta^{a-1}(1-\theta)^{b-1} \\ &\propto \theta^{(a+k)-1}(1-\theta)^{(b+n-k)-1} \end{aligned}$$

- That is:

$$\theta \mid k \sim Beta(a+k, b+n-k)$$

- Take a Beta prior with a Binomial likelihood, you get a Beta posterior (conjugacy)

# Application to the deer example

- Posterior distribution of survival is $\theta \sim Beta(a + k, b + n - k)$.

## Application to the deer example

- Posterior distribution of survival is $\theta \sim Beta(a+k, b+n-k)$.
- If we take a Uniform prior, i.e. $Beta(1,1)$, then we have:

## Application to the deer example

- Posterior distribution of survival is $\theta \sim Beta(a + k, b + n - k)$.
- If we take a Uniform prior, i.e. $Beta(1, 1)$, then we have:
- $\theta_{treatment} \sim Beta(1 + 19, 1 + 57 - 19) = Beta(20, 39)$

# Application to the deer example

- Posterior distribution of survival is $\theta \sim Beta(a + k, b + n - k)$.
- If we take a Uniform prior, i.e. $Beta(1, 1)$, then we have:
- $\theta_{treatment} \sim Beta(1 + 19, 1 + 57 - 19) = Beta(20, 39)$
- Note that in this specific situation, the posterior has an explicit expression, easy to manipulate.

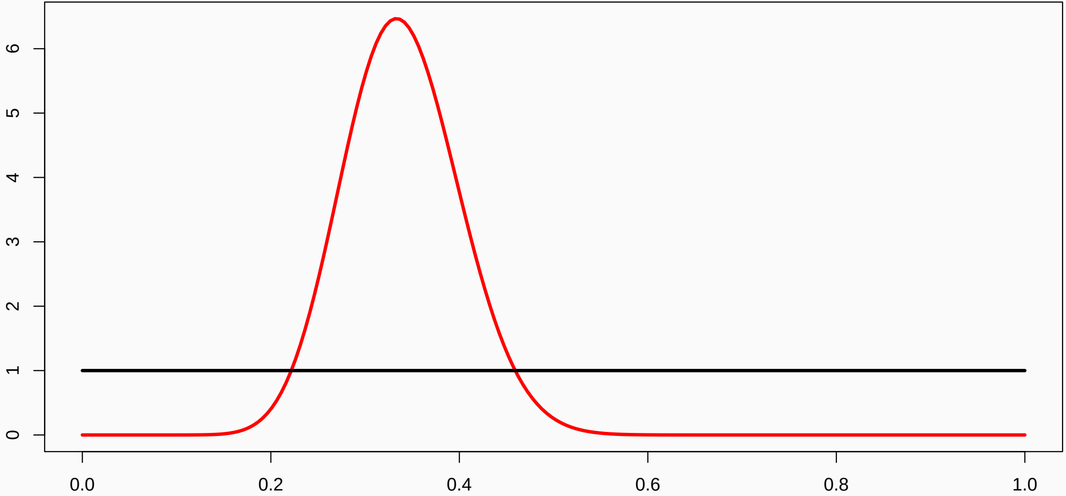
## Application to the deer example

- Posterior distribution of survival is $\theta \sim Beta(a + k, b + n - k)$.
- If we take a Uniform prior, i.e. $Beta(1, 1)$, then we have:
- $\theta_{treatment} \sim Beta(1 + 19, 1 + 57 - 19) = Beta(20, 39)$
- Note that in this specific situation, the posterior has an explicit expression, easy to manipulate.
- In particular, $E(Beta(a, b)) = \dfrac{a}{a + b} = 20/59$ to be compared with the MLE $19/57$.

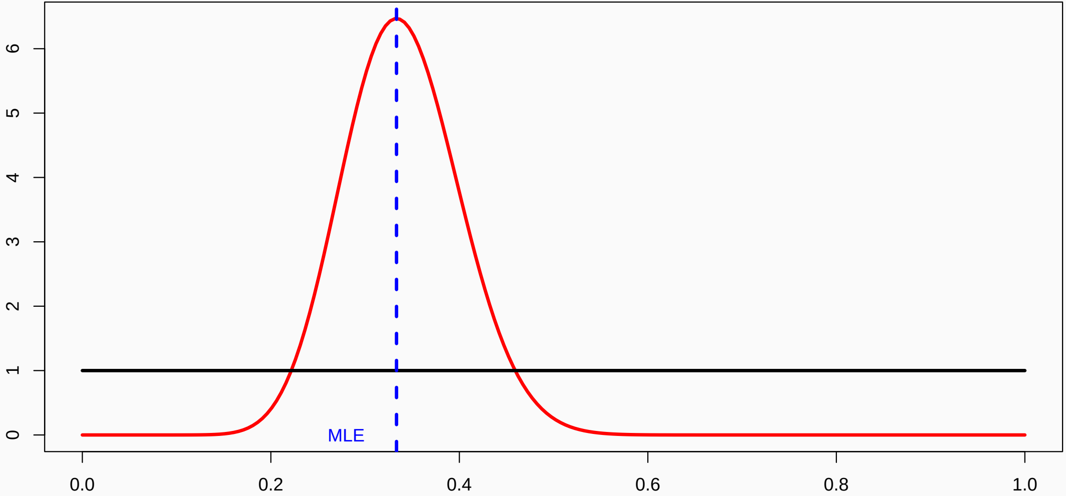
# A general result

**This is a general result, the Bayesian and frequentist estimates will always agree if there is sufficient data, so long as the likelihood is not explicitly ruled out by the prior.**

# Prior $Beta(1, 1)$ and posterior survival $Beta(20, 39)$



# Prior $Beta(1, 1)$ and posterior survival $Beta(20, 39)$

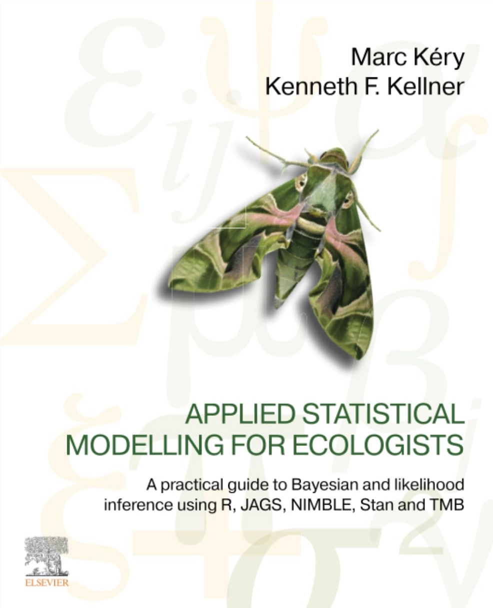


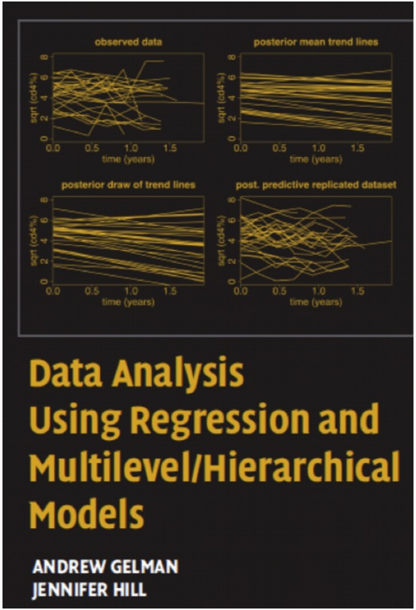
# Summary

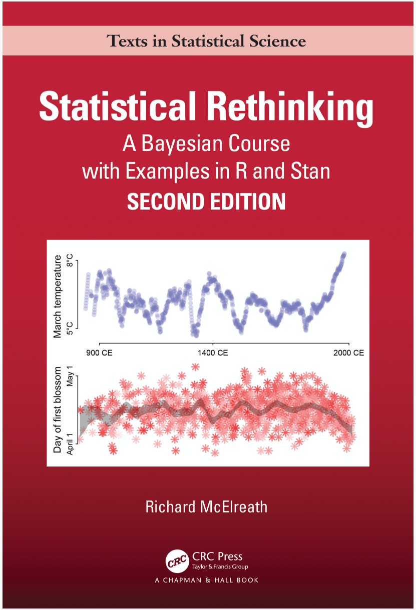
## Summary

- With the Bayes' theorem, you update your beliefs (prior) with new data (likelihood) to get posterior beliefs (posterior): posterior $\propto$ likelihood $\times$ prior.
- In some simple situations, it is possible to explicitly calculate the posterior distribution of model parameters.
- In general though, when applying the Bayes' theorem, you get high-dimensional integral, difficult if not impossible to calculate. This is one of the reasons why we need simulation (MCMC) methods.

# Further reading


Marc Kéry
Kenneth F. Kellner
APPLIED STATISTICAL
MODELLING FOR ECOLOGISTS
A practical guide to Bayesian and likelihood
inference using R, JAGS, NIMBLE, Stan and TMB
ELSEVIER


observed data
posterior mean trend lines
posterior draw of trend lines
post. predictive replicated dataset
sqrt (cd4%)
time (years)
Data Analysis
Using Regression and
Multilevel/Hierarchical
Models
ANDREW GELMAN
JENNIFER HILL


Texts in Statistical Science
Statistical Rethinking
A Bayesian Course
with Examples in R and Stan
SECOND EDITION
March temperature
8°C
5°C
900 CE
1400 CE
2000 CE
Day of first blossom
May 1
April 1
Richard McElreath
CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK