

# Bayesian data analysis

## Outline

- What is probability
- Another definition of probability
- Bayes' Theorem
- Prior probability and posterior probability
- How Bayesian inference is different from what we usually do
- Examples: one species or two; estimating a proportion
- Credible intervals
- Quick notes on Bayesian model selection

## **What is probability**

A way of quantifying uncertainty.

Mathematical theory originally developed to model outcomes in games of chance.

## Definition of probability (frequentist)

The *probability* of an event is the proportion of times that the event would occur if we repeated a random trial over and over again under the same conditions.

A *probability distribution* is a list of all mutually exclusive outcomes of a random trial and their probabilities of occurrence.

## Probability statements that make sense under this definition

If we toss a fair coin, what is the probability of 10 heads in a row?

If we assign treatments randomly to subjects, what is the probability that a sample mean difference between treatments will be greater than 10%?

Under a process of genetic drift in a finite population, what is the probability of fixation of a rare allele?

What is the probability of a result at least as extreme as that observed if the null hypothesis is true?

In these examples, sampling error is the source of uncertainty.

## **Probability statements that don't make sense under this definition**

What is the probability that Iran is building nuclear weapons?

What is the probability that hippos are the sister group to the whales?

What is the probability that the fish sampled from that newly discovered lake represent two species rather than one?

What is the probability that polar bears will be extinct in the wild in 40 years?

## Why they don't make sense

What is the probability that Iran is building nuclear weapons?

[either it is or it isn't – no random trial here]

What is the probability that hippos are the sister group to the whales?

[either it is or it isn't – no random trial here]

What is the probability that the fish sampled from that newly discovered lake represent two species rather than one?

[either there are one or there are two – no random trial]

What is the probability that polar bears will be extinct in the wild in 40 years?

[difficult to cast this as a frequency of occurrence]

In these examples there is no random trial, so no sampling error.

Information is the source of uncertainty.

## **Alternative definition of probability (Bayesian)**

*Probability* is a measure of a degree of belief associated with the occurrence of an event.

A *probability distribution* is a list of all mutually exclusive events and the degree of belief associated with their occurrence.

Bayesian statistics applies the mathematics of probability to subjective degree of belief.

## **Bayesian methods are increasingly used in ecology and evolution**

Is this good?

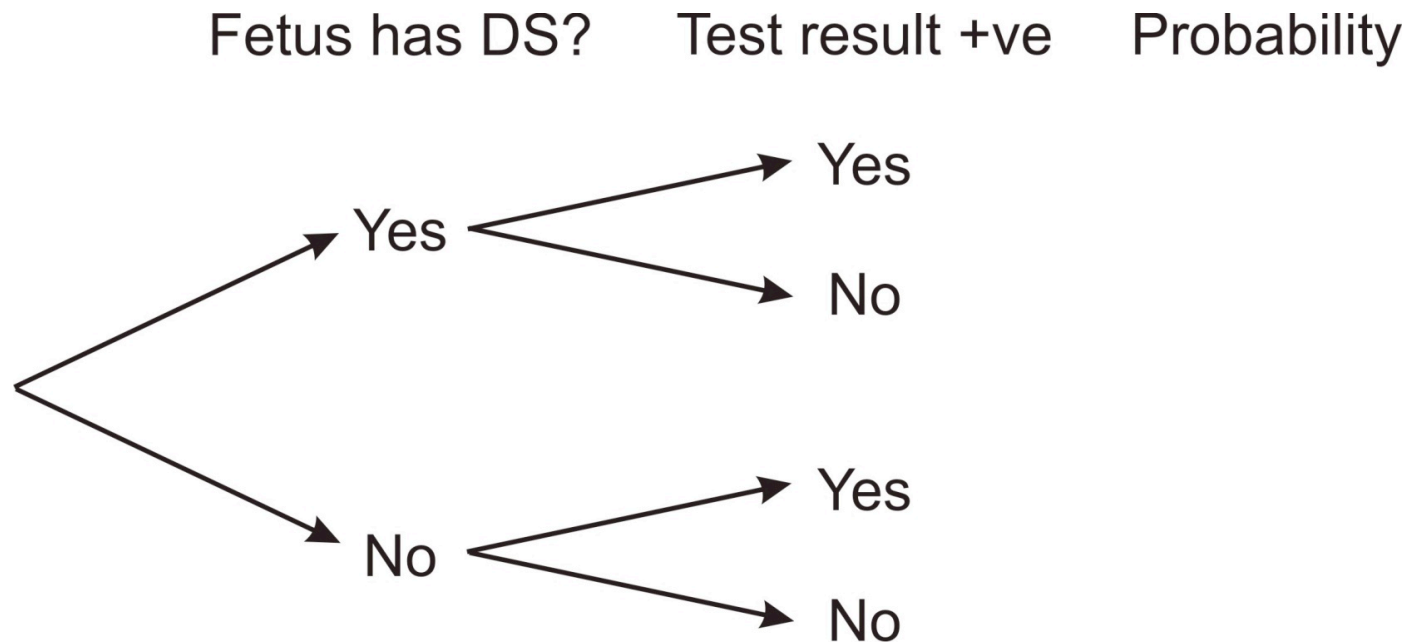
*“Ecologists should be aware that Bayesian methods constitute a radically different way of doing science. Bayesian statistics is not just another tool to be added into the ecologists’ repertoire of statistical methods. Instead, Bayesians categorically reject various tenets of statistics and the scientific method that are currently widely accepted in ecology and other sciences.”* B. Dennis, 1996, *Ecology*



## Bayes' Theorem itself is harmless

Example: detection of Down syndrome (DS).

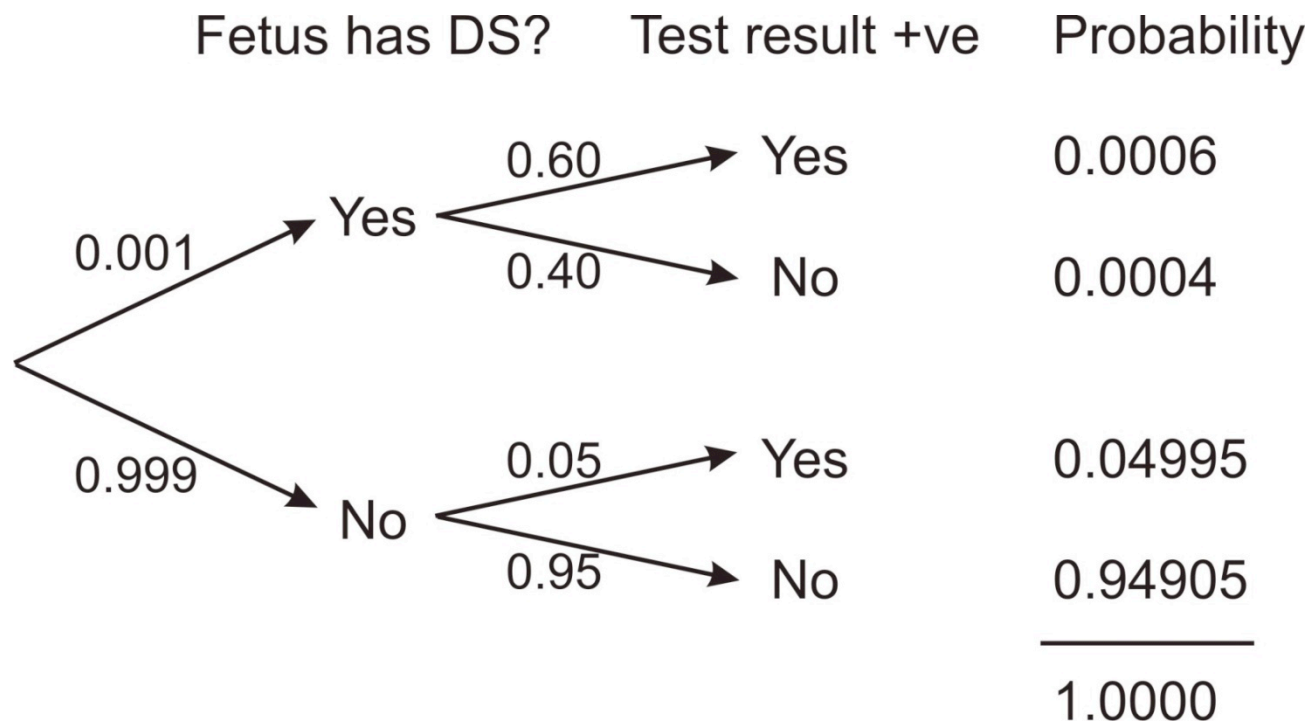
DS occurs in about 1 in 1000 pregnancies. The most accurate test requires amniocentesis, which carries a small risk of miscarriage. The triple test is used first. It is risk-free though less accurate.



## Conditional probability

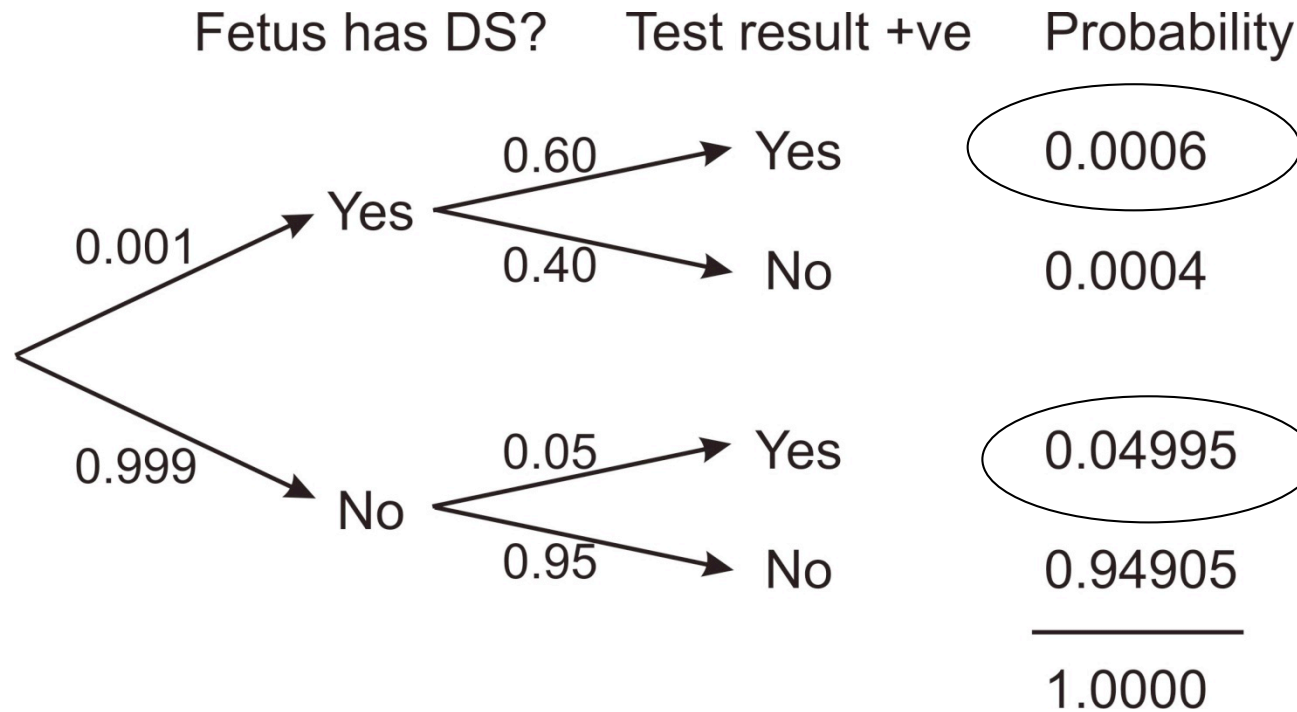
The *conditional probability* of an event is the probability of that event occurring given that a condition is met.

E.g.: the probability of a +ve test result from the triple test, given that a fetus has DS, is 0.6.



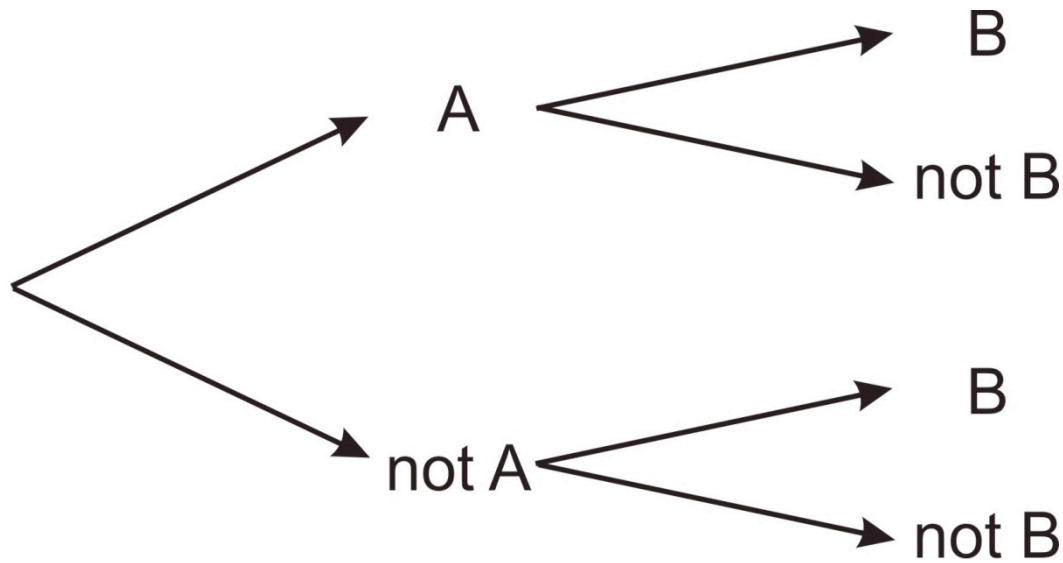
## Conditional probability calculation

What is the probability that a fetus has DS if the test is positive?



$$\Pr[\text{DS} \mid \text{positive}] = \frac{0.0006}{0.0006 + 0.04995} = 0.012, \text{ just } 1.2\%$$

This calculation is formalized in Bayes' Theorem

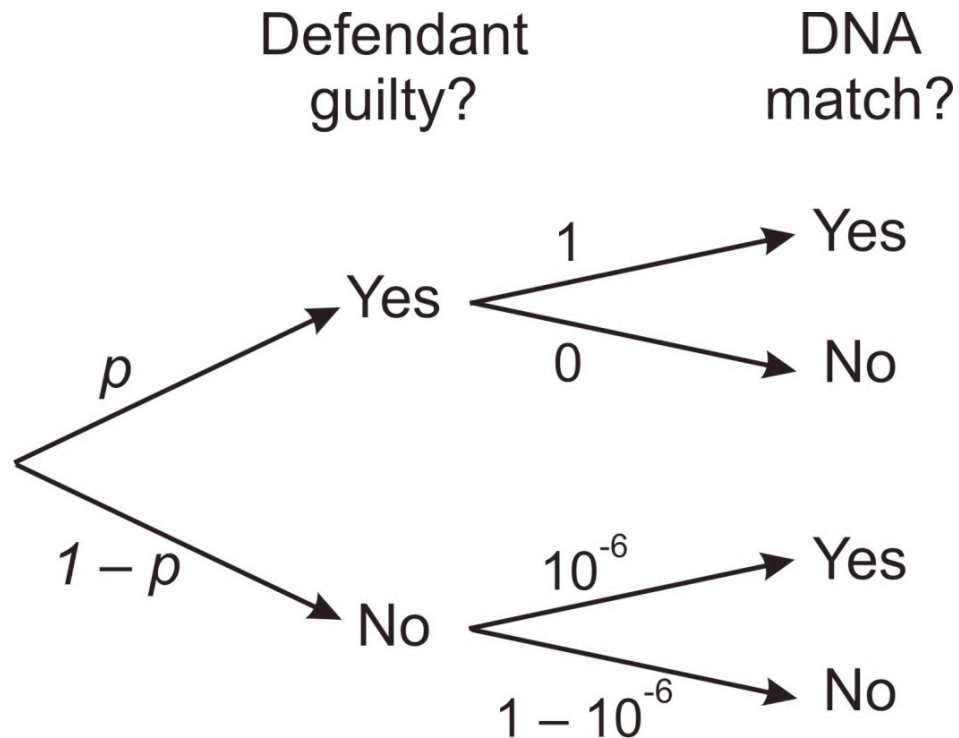


$$\Pr[A | B] = \frac{\Pr[B | A] \Pr[A]}{\Pr[B | A] \Pr[A] + \Pr[B | \text{not}A] \Pr[\text{not}A]}$$

## What's controversial is how Bayes Theorem is used

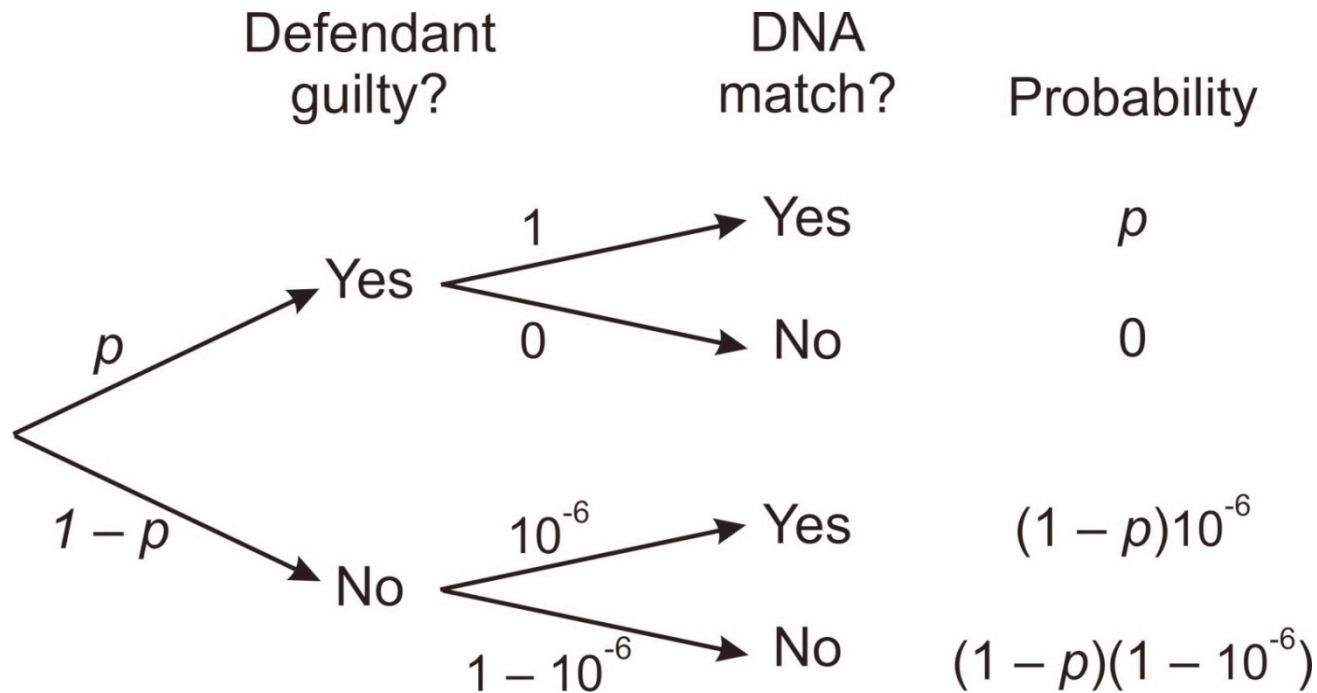
For example: forensic evidence. Bayesian inference can be used in a court to quantify the evidence for and against the guilt of the defendant based on a match with DNA evidence left at the scene of the crime

What is the probability of guilt given a positive DNA match (assuming no contamination of samples)?



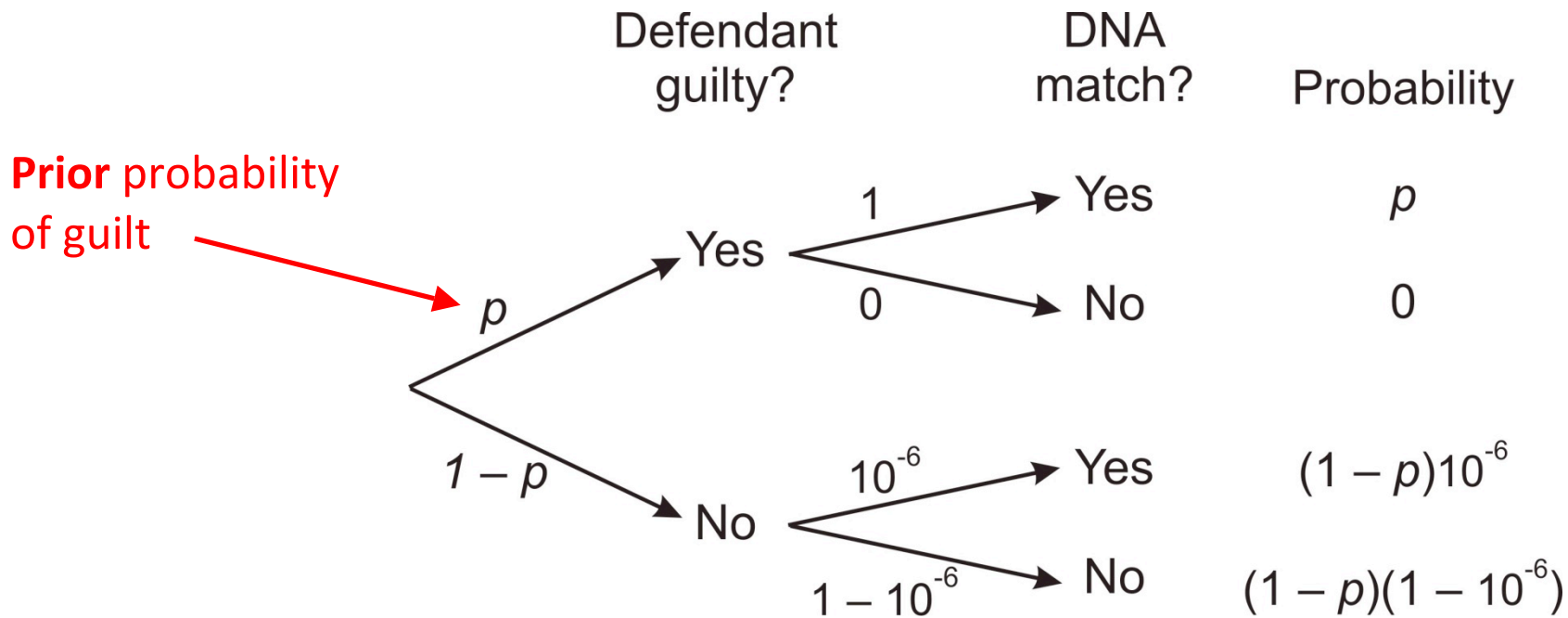
## Bayesian inference in action

What is the probability of guilt given a positive DNA match?



$$\Pr[\text{guilt} \mid \text{match}] = \frac{1(p)}{1(p) + 10^{-6}(1 - p)}$$

## Prior and posterior probability

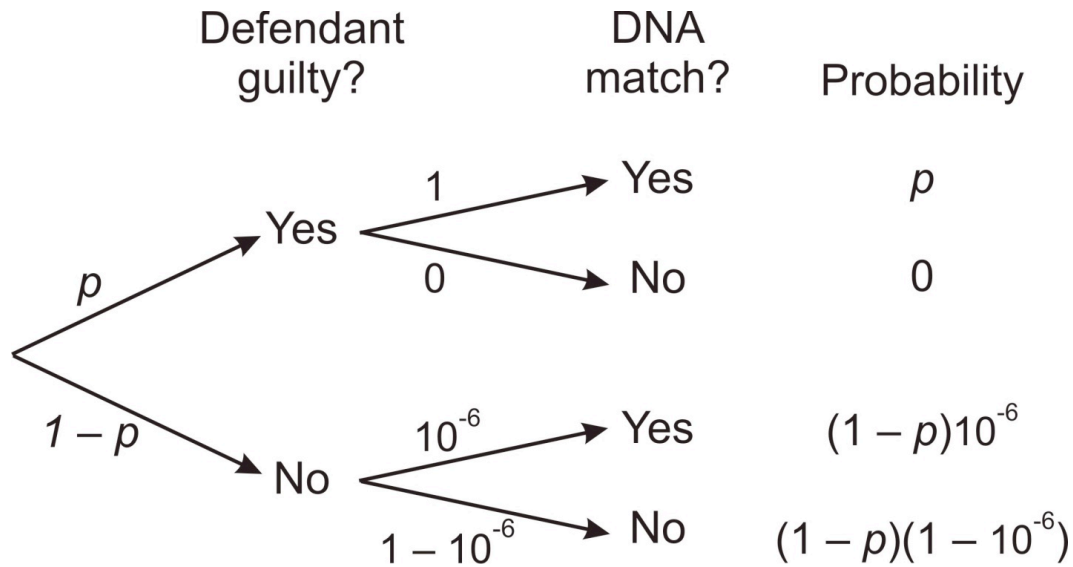


Prior probability of guilt

Posterior probability of guilt

$$\Pr[\text{guilt} \mid \text{match}] = \frac{1(p)}{1(p) + 10^{-6}(1 - p)}$$

## Bayesian inference in action



$$\Pr[\text{guilt} \mid \text{match}] = \frac{1(p)}{1(p) + 10^{-6}(1 - p)}$$

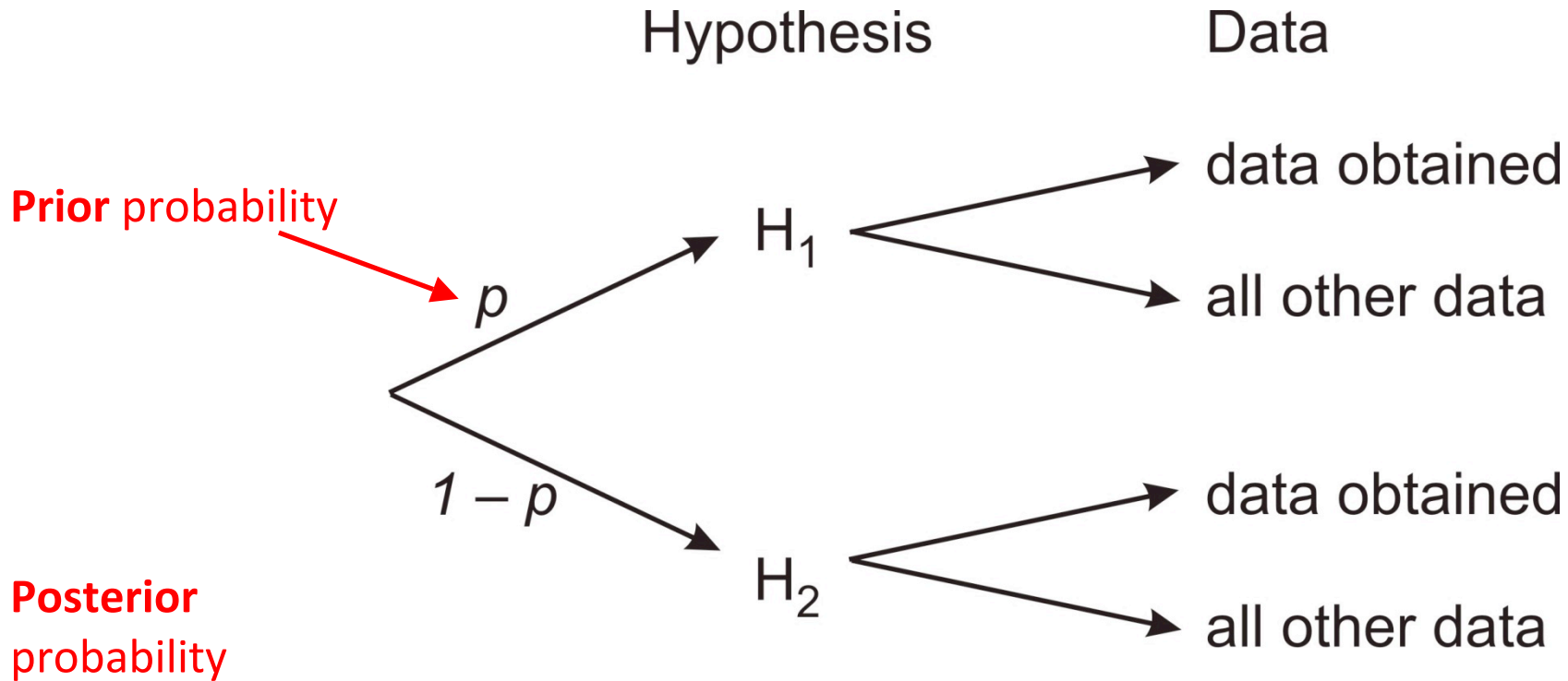
If  $p = 10^{-6}$  then  $\Pr[\text{guilt} \mid \text{match}] = 0.5$

If  $p = 0.5$  then  $\Pr[\text{guilt} \mid \text{match}] = 0.999999$

So, is the defendant guilty or innocent?



## Bayesian inference with data

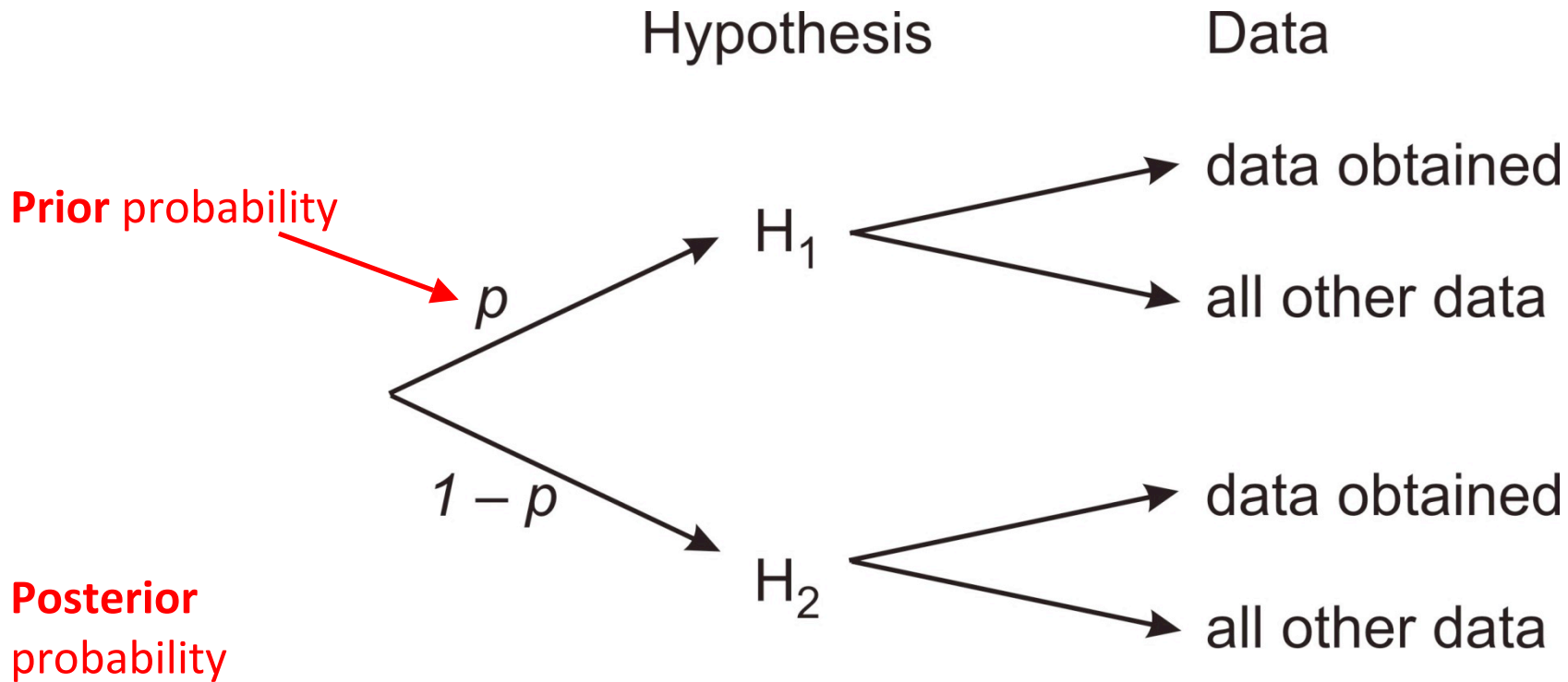


Posterior probability

$$\Pr[H_1 | \text{data}] = \frac{\Pr[\text{data} | H_1] \Pr[H_1]}{\Pr[\text{data} | H_1] \Pr[H_1] + \Pr[\text{data} | H_2] \Pr[H_2]}$$

[break]

## Bayesian inference with data



$$\Pr[H_1 | \text{data}] = \frac{\Pr[\text{data} | H_1] \Pr[H_1]}{\Pr[\text{data} | H_1] \Pr[H_1] + \Pr[\text{data} | H_2] \Pr[H_2]}$$

## **How Bayesian inference is different from what we usually do**

The prior probability represents the investigator's strength of belief about the value of the fixed parameter or hypothesis.

The posterior probability expresses how the investigator's beliefs have been altered by the data.

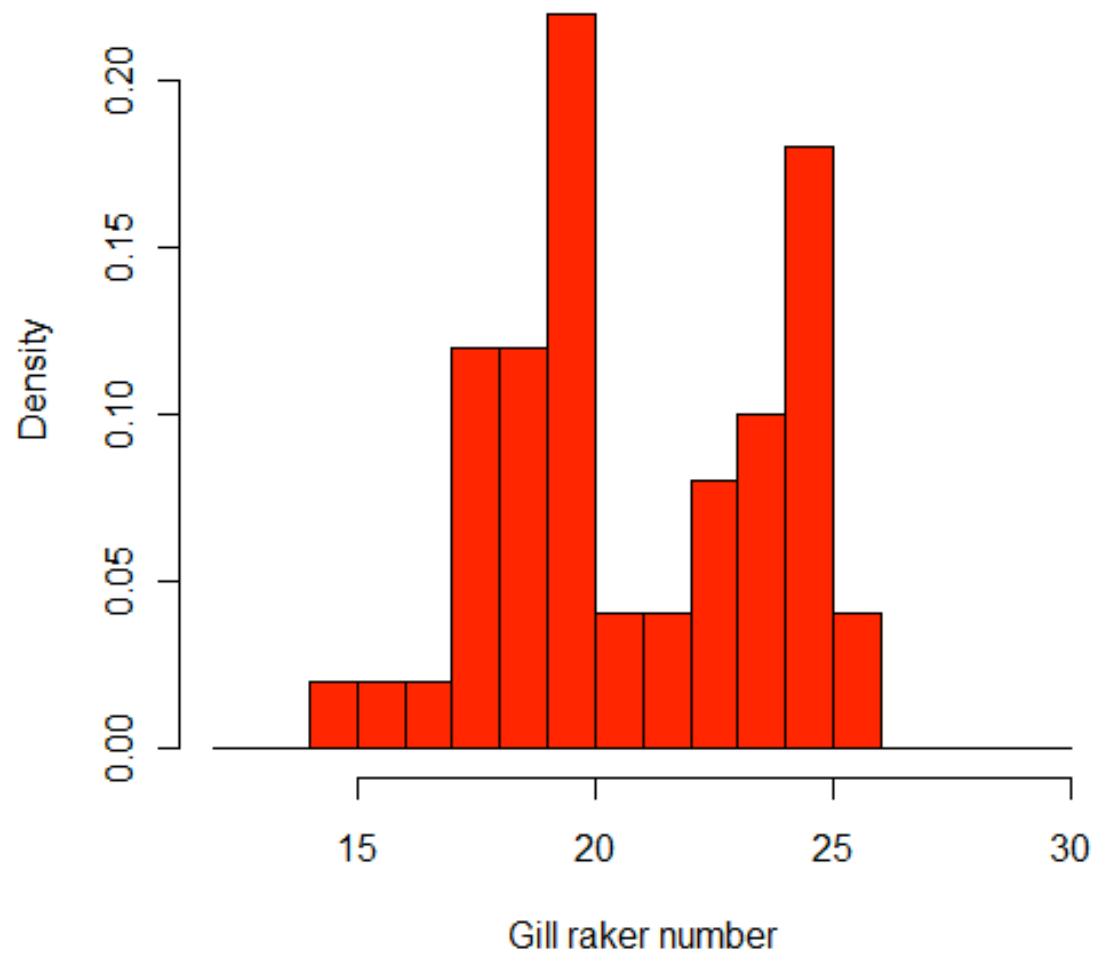
Mathematically, the hypothesis or parameter is treated as though it is random, not fixed.

The only data considered are the data observed, not the data that "might have been"; other possible outcomes of the experiment are not considered.

## Example 1: One species or two

Data: Gill raker number of 50 fish collected from a new lake

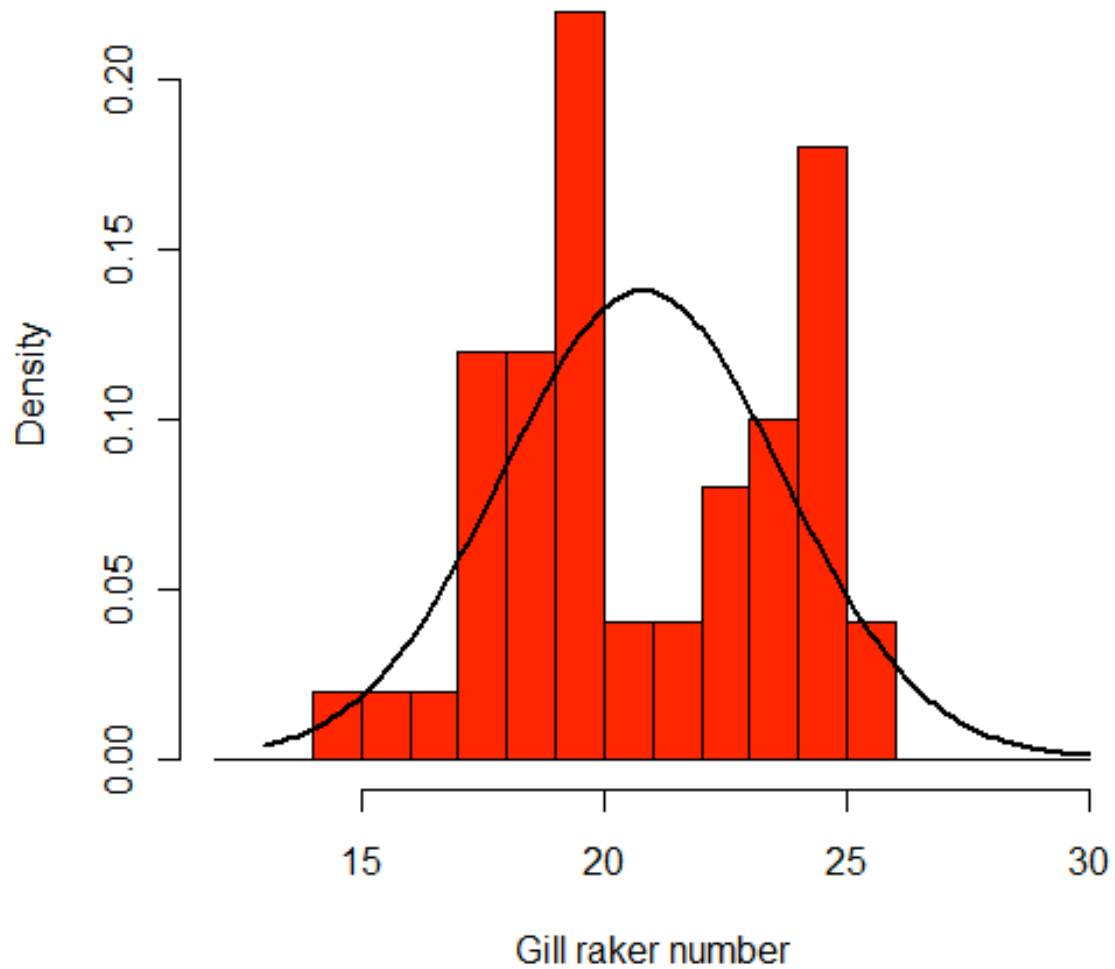
What is the probability that they represent 2 species rather than 1?



## $H_1$ : one species

Assume a normal distribution of measurements

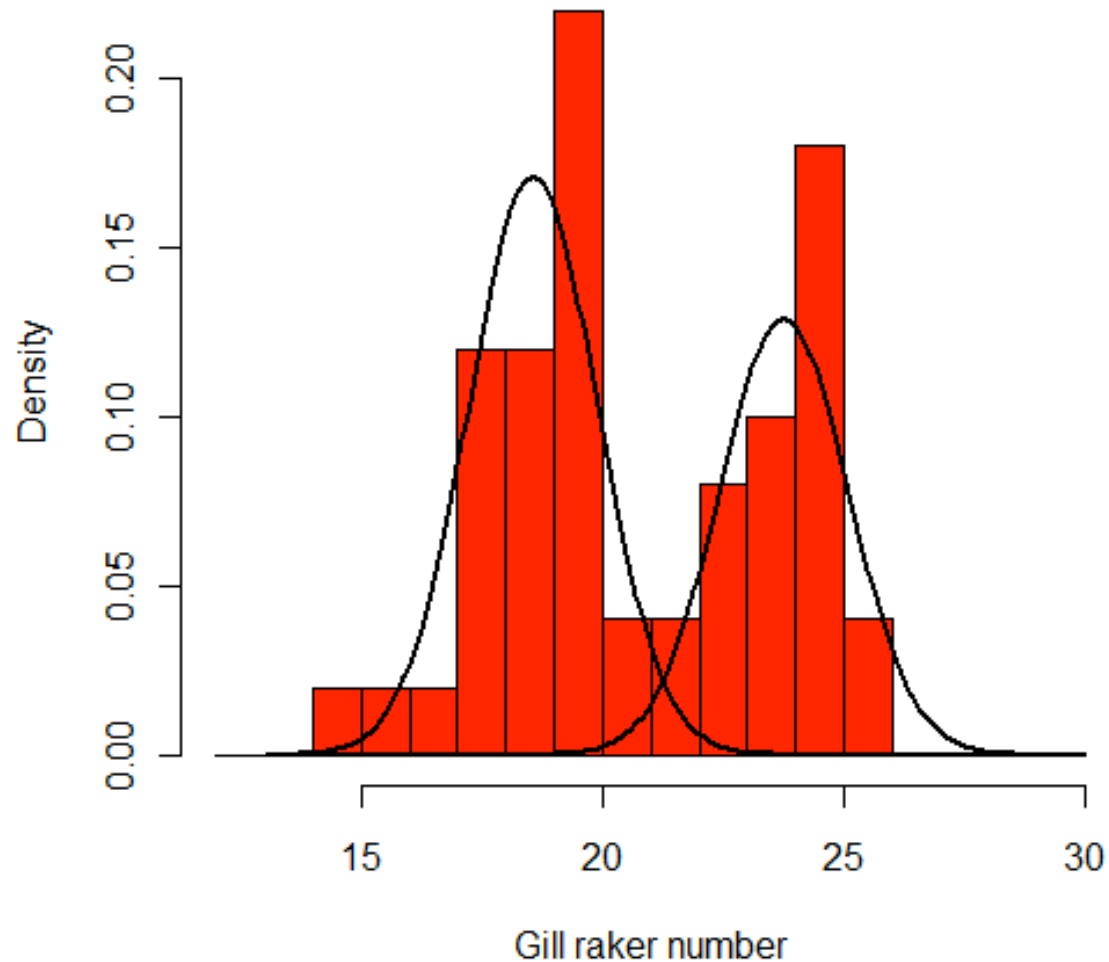
$$\Pr[\text{data} | H_1] = \log L[H_1 | \text{data}] = -124.06$$



## H<sub>2</sub>: two species

Assume normal distributions with equal variance in both groups

$$\Pr[\text{data} \mid H_2] = \log L[H_2 \mid \text{data}] = -116.51$$



## Posterior model probabilities

Plug the likelihoods into Bayes Theorem to calculate the posterior probabilities of each hypothesis given the data

Posterior probability depends on the prior probability

Here is the probability that  $H_2$  is correct (two species are present):

<b>Prior probability</b> $\Pr[H_2]$	<b>Posterior probability</b> $\Pr[H_2 \mid \text{data}]$
0.500	0.99
0.005	0.91
0.001	0.66

If prior is small, need more data to increase posterior probability



## Example 2: Bayesian estimation of a proportion

Study of the sex ratio of the communal-living bee, (Paxton and Tengo, 1996, *J. Insect. Behav.*)

What is the proportion of males in the reproductive adults emerging from colonies?

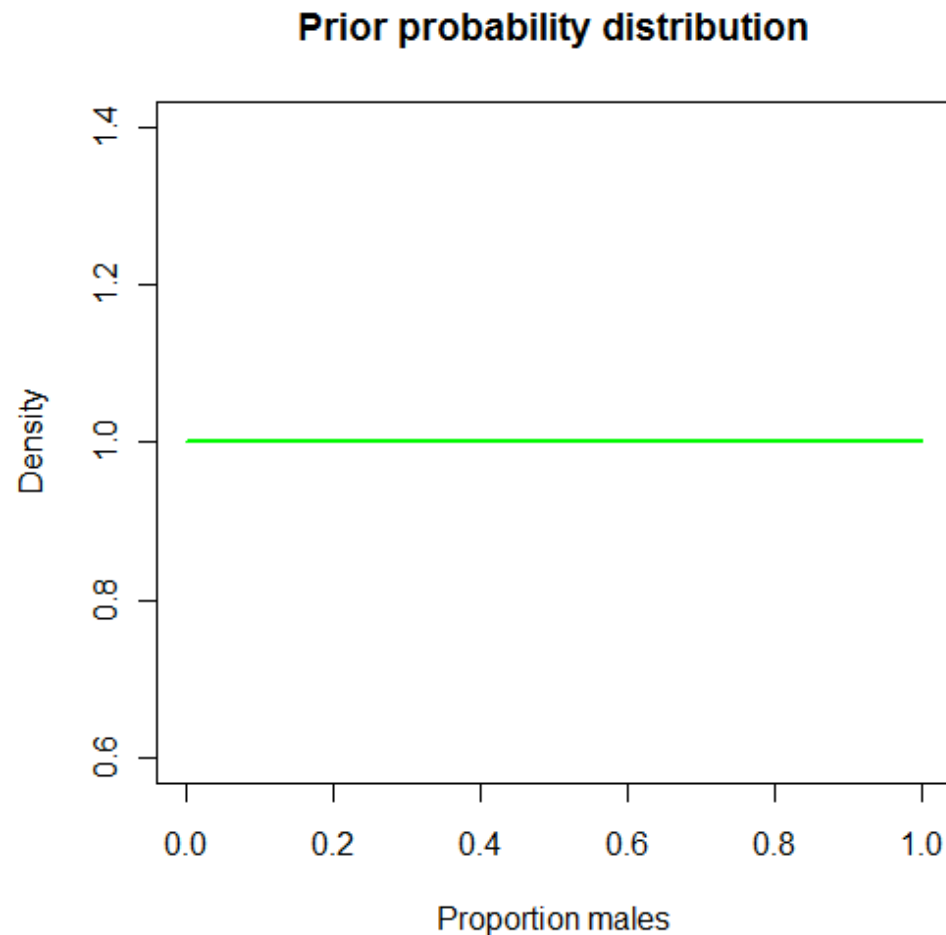


<http://www.flickr.com/photos/90408805@N00/>

## Bayesian estimation of a proportion

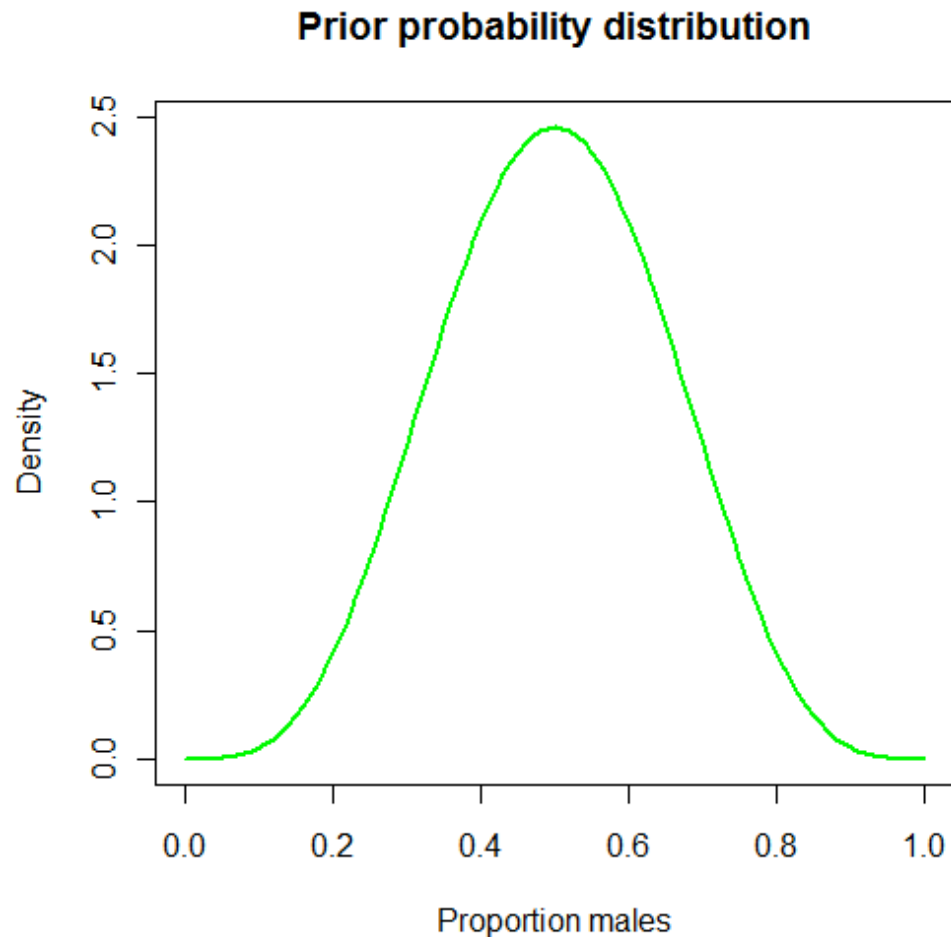
To begin, we need to come up with a prior probability distribution for the proportion.

Case 1: an “noninformative” prior: expression of total ignorance.



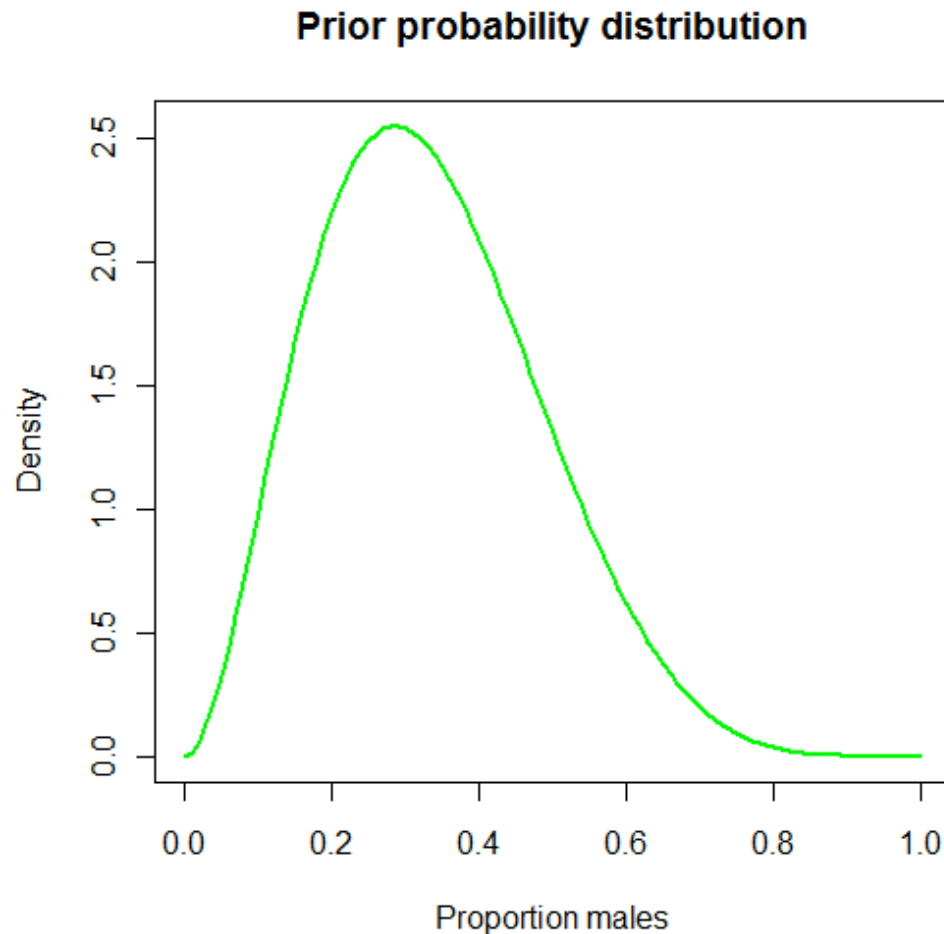
## Bayesian estimation of a proportion

Case 2: Most species have a sex ratio close to 50:50, and this is predicted by simple sex-ratio theory. The following prior attempts to capture this previous information (this is really what priors are for).



## Bayesian estimation of a proportion

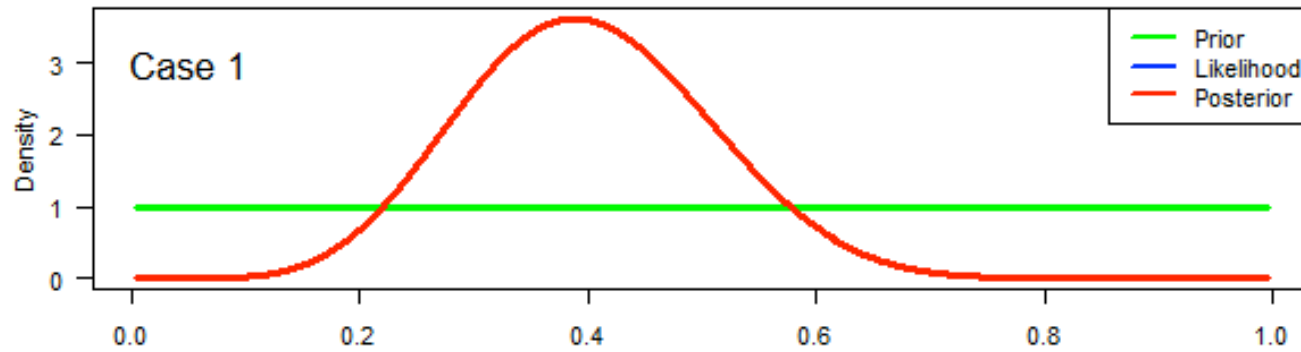
Case 3: Then again, female-biased sex ratios do exist in nature, more than male-biased sex ratios, especially in bees and other hymenoptera. The following prior attempts to capture this previous information.



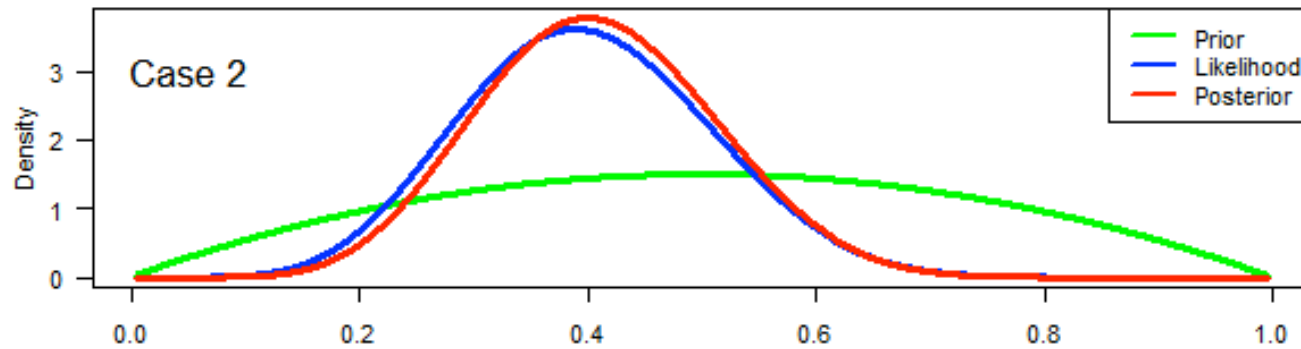
## Bayesian estimation of a proportion

Data: From day 148 at nest S31: 7 males, 11 females  $\hat{p}_{MLE} = 0.39$

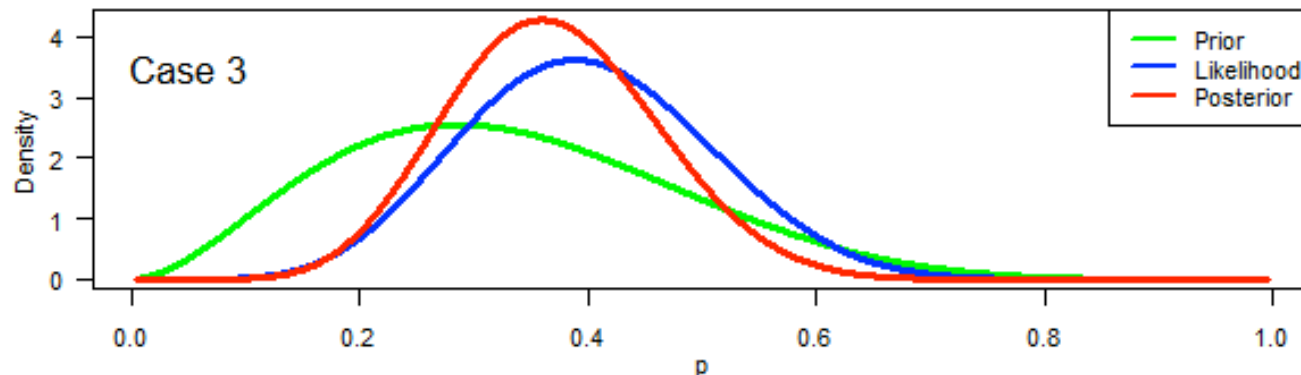
$$\hat{p} = 0.39$$



$$\hat{p} = 0.40$$



$$\hat{p} = 0.36$$



## Bayesian estimation of a proportion

The estimate having maximum posterior probability depends on the prior probability distribution for the estimate.

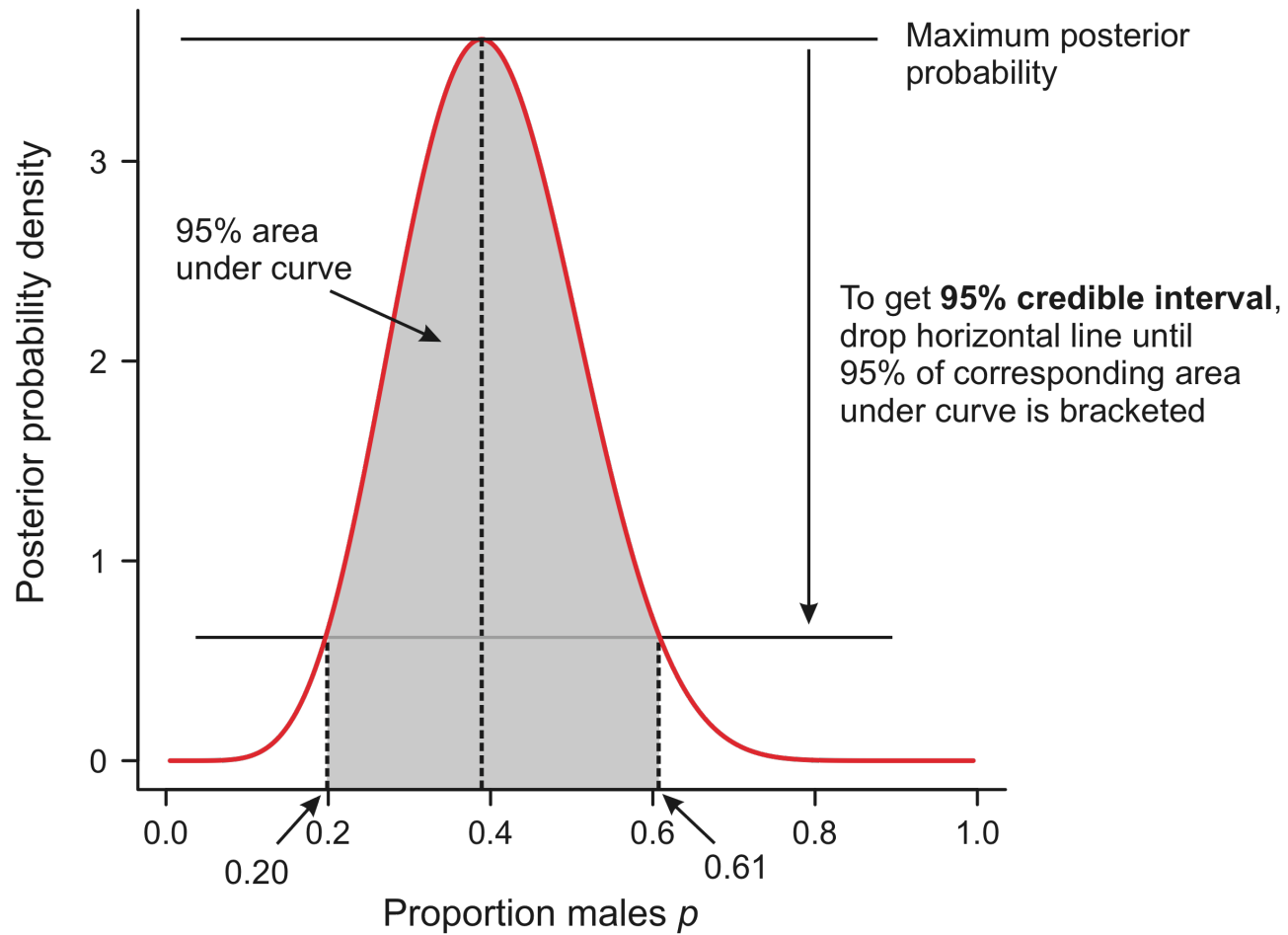
Potential source of controversy: The prior is partly subjective. Different researchers may use different priors, hence obtain different estimates with the same data.

To resolve this we might all agree to use “noninformative” priors. But this prevents us from incorporating prior information, which is regarded as one of the strengths of the Bayesian approach.

Conflict can be resolved if we base the prior on a survey of preexisting evidence (lot of work). Would you agree with this?

# Bayesian estimation of a proportion

## 95% credible interval



## Bayesian estimation of a proportion

Interpretation of the interval estimates

95% likelihood-based confidence interval:

$$0.19 \leq p \leq 0.62$$

Interpretation: Most plausibly,  $p$  is between 0.19 and 0.62.

In approximately 95% of random samples taken from the same population, the likelihood-based confidence interval so calculated will bracket the true population proportion  $p$ .

95% credible interval:

$$0.20 \leq p \leq 0.61 \quad (\text{assuming Case 1, with noninformative prior})$$

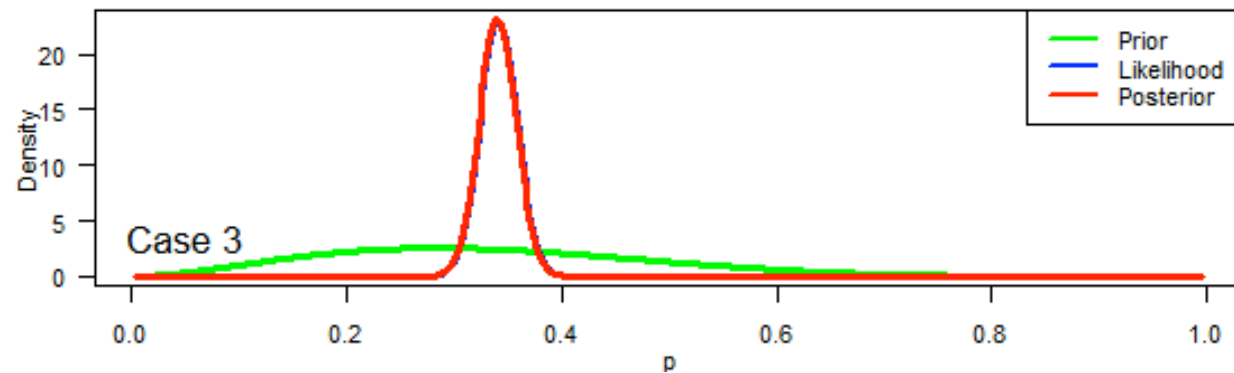
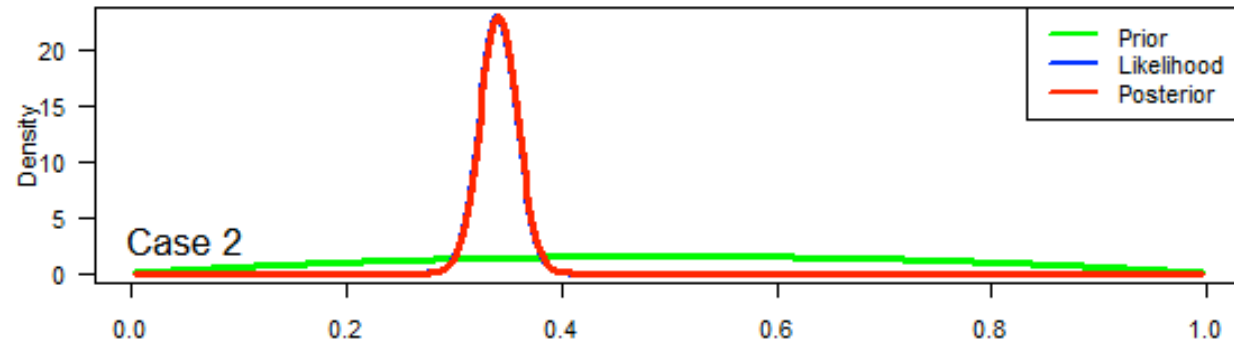
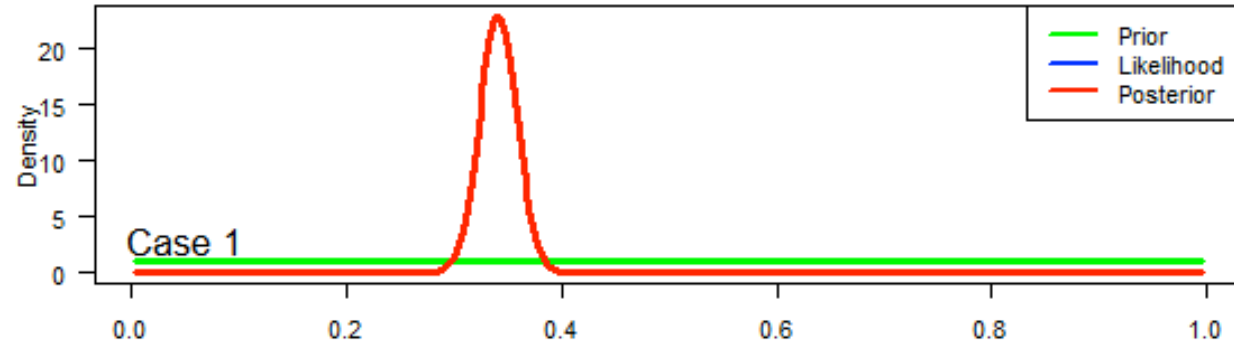
Interpretation: The probability is 0.95 that the population proportion lies between 0.20 and 0.61



## Bayesian estimation of a proportion

All the data: 253 males, 489 females  $\hat{p}_{MLE} = 0.34$

With lots of data, the choice of prior has little effect on the posterior distribution.



## Bayesian model selection (quick notes)

Model selection: the problem of deciding the best candidate model fitted to data

Requires a criterion to compare models, and a strategy for finding the best

One Bayesian approach uses BIC as the criterion (Bayesian Information Criterion). Derived from a wholly different theory, but yields a formula similar to that of AIC. It assumes that the “true model” is one of the models included among the candidates. The approach has a tendency to pick a simpler model than that from AIC (“penalty” is more severe).

$$\text{AIC} = -2 \ln L(\text{fitted model} \mid \text{data}) + 2k$$

$$\text{BIC} = -2 \ln L(\text{fitted model} \mid \text{data}) + k \log(n)$$

$k$  is the number of parameters estimated in the model (including intercept and  $\sigma^2$ ),  $n$  is the sample size.

## Summary

- Bayesian probability is a different concept than the frequentist concept
- Bayes' Theorem can be used to estimate and test hypotheses using posterior probability
- The approach incorporates (requires) prior probability
- The influence of prior probability declines with more data
- The interpretation of interval estimates differs from the frequentist definition
- These ideas are becoming used more in ecology and evolution

**Discussion paper for next week:**

Hobbs & Hilborn (2006) Multiple working hypotheses in ecology.

Download from “**assignments**” tab on course web site.

Presenters: Alice & Norah

Moderators: Natalie & Susannah