Outline

- What is probability
- Another definition of probability
- Bayes’ Theorem
- Prior probability; posterior probability
- How Bayesian inference is different from what we usually do
- Example: one species or two
- Example: estimating a proportion
- Credible intervals
- Bayes hypothesis testing using the Bayes factor
- 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 small 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, <u>sampling error</u> 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 Iran is or isn't – no random trial here]
- What is the probability that hippos are the sister group to the whales?
  [either they are or they're not – 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 is one species or there are two – no random trial]
- What is the probability that polar bears will be extinct in the wild in 40 years?
  [maybe it is possible to state in terms of the accumulation of outcomes of random trials]

In these examples there is no random trial, so no sampling error. I<u>nformation</u> is the source of uncertainty, not sampling error.

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 uncertainty measured as subjective degree of belief.

Bayesian methods are increasingly used in ecology and evolution

"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*

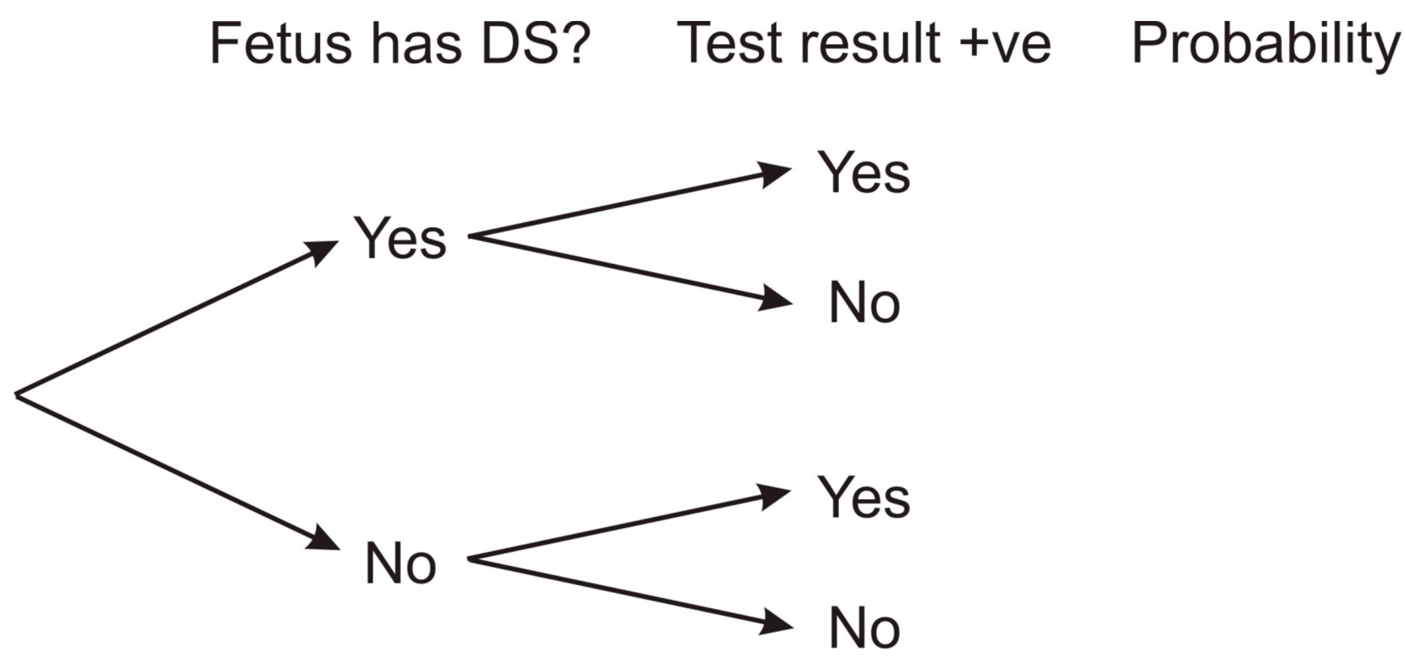
"Ecologists are facultative Bayesians" (M. Mangel, pers. comm. 2013)

Should we be using it?

Bayes' Theorem itself is harmless

Example: detection of Down syndrome (DS).

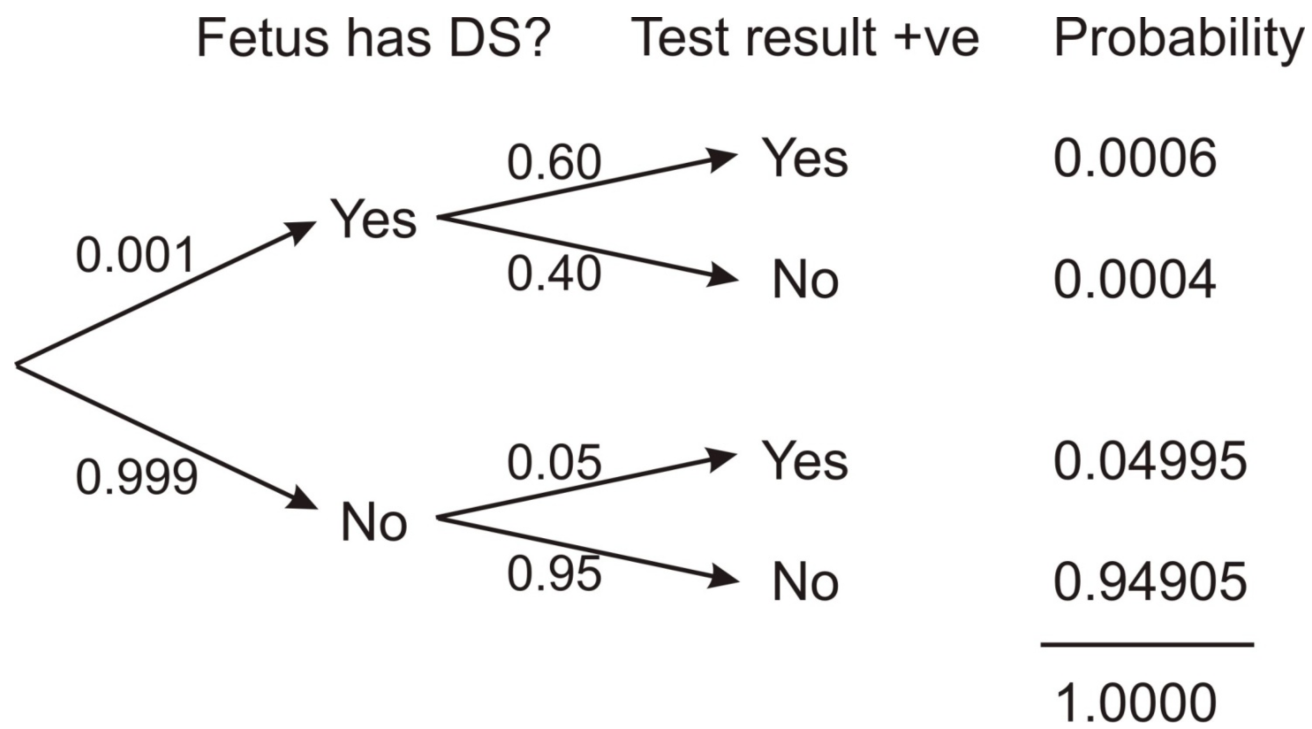
DS occurs in about 1 in 1000 pregnancies. A "triple test" of levels of 3 blood sera (α-fetoprotein, estriol, and β-subunit of human chorionic gonadotropin) is widely used. It is cheap and risk-free. A newer DNA test is more accurate. The most accurate test requires amniocentesis, which carries a small risk of miscarriage.



Conditional probability

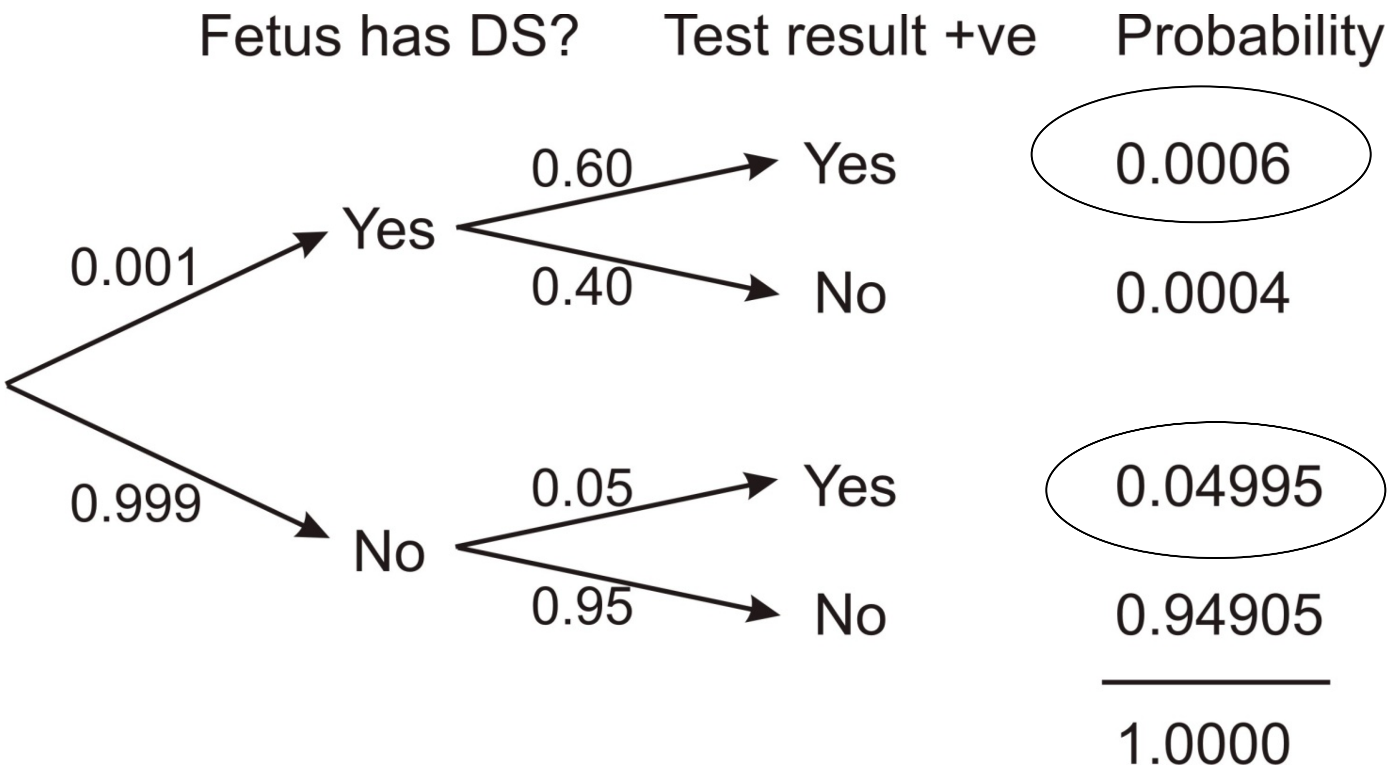
Remember that the *conditional probability* of an event is the probability of that event occurring given that a condition is met.

The probability of a positive test result from the triple test is 0.6, given that a fetus has DS. The probability of a positive result is 0.05, given that a fetus is not DS.



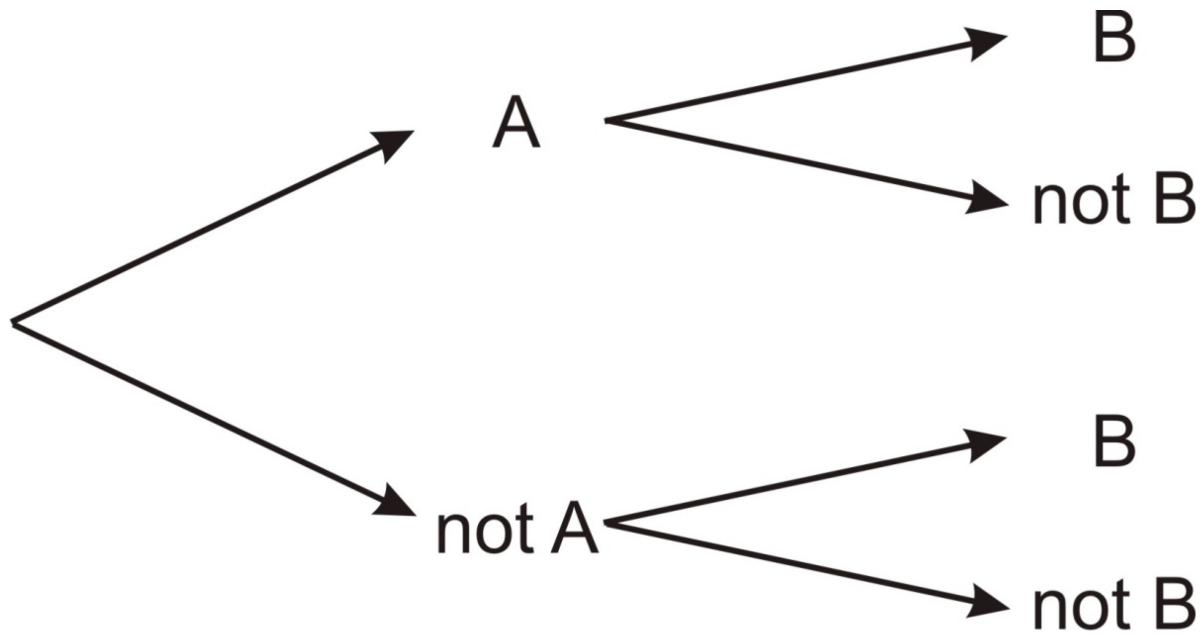
Conditional probability calculation

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



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

This calculation is formalized in Bayes' Theorem

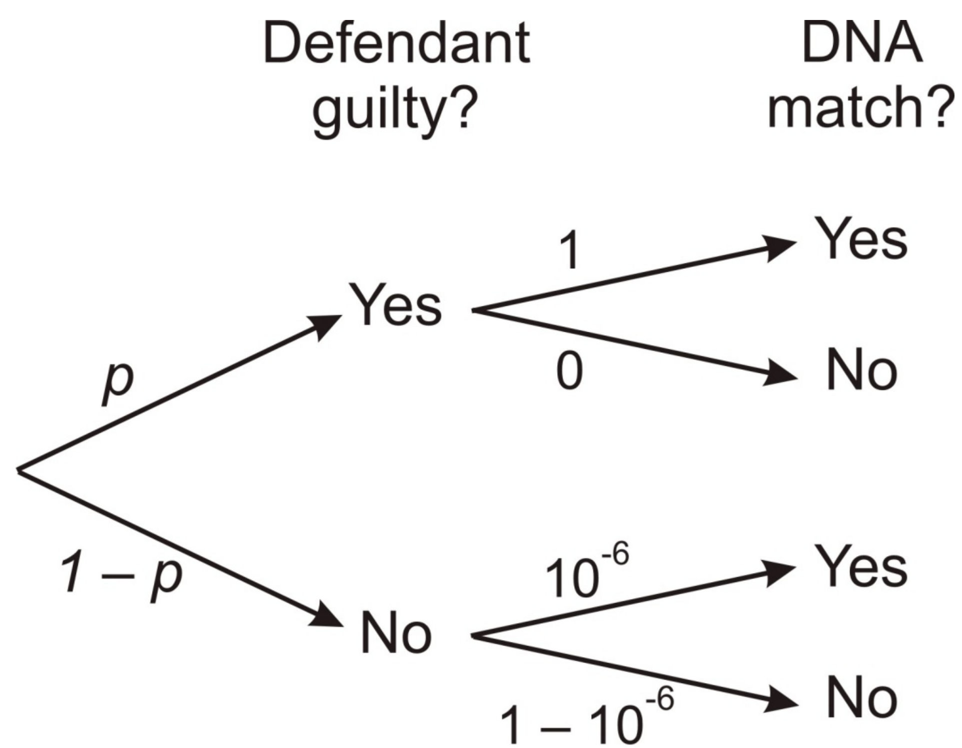


$$\Pr[A \mid B] = \frac{\Pr[B \mid A]\,\Pr[A]}{\Pr[B \mid A]\,\Pr[A] + \Pr[B \mid \text{not}A]\,\Pr[\text{not}A]}$$

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

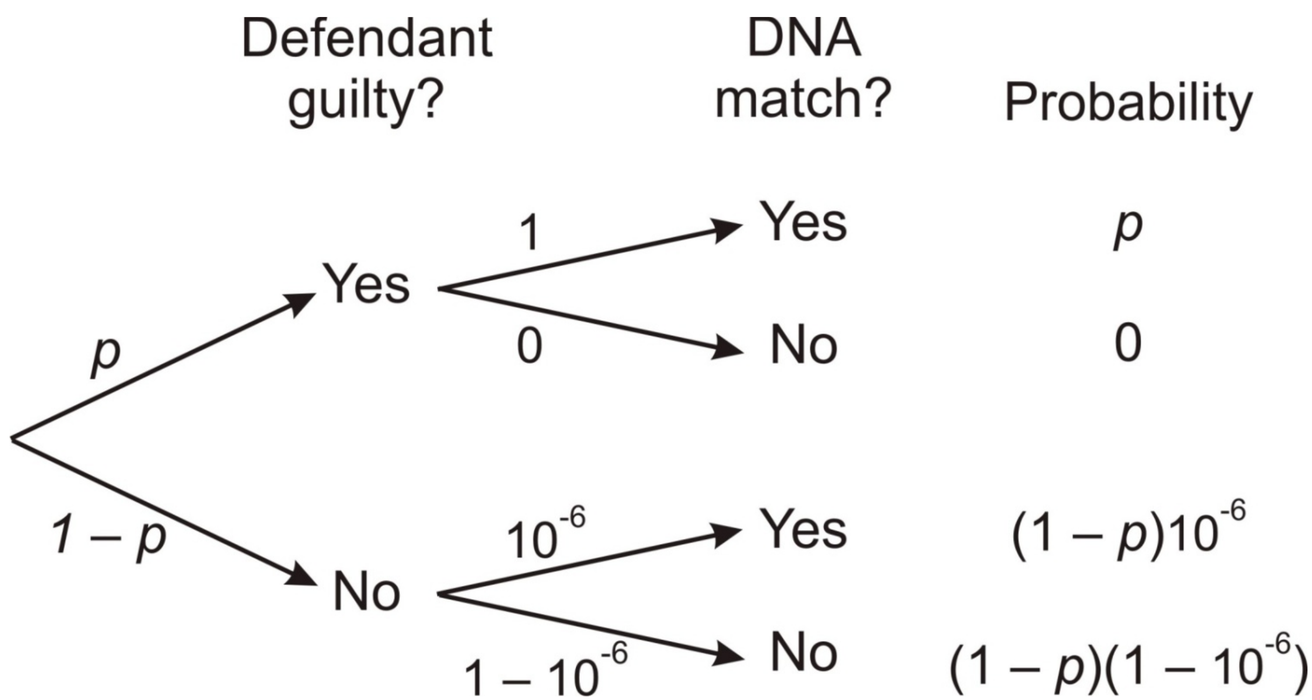
For example: forensic evidence. Bayesian inference can be used in a court of law to quantify the evidence for and against the guilt of the defendant based on a match to DNA evidence left at the crime scene.

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



Bayesian inference in action

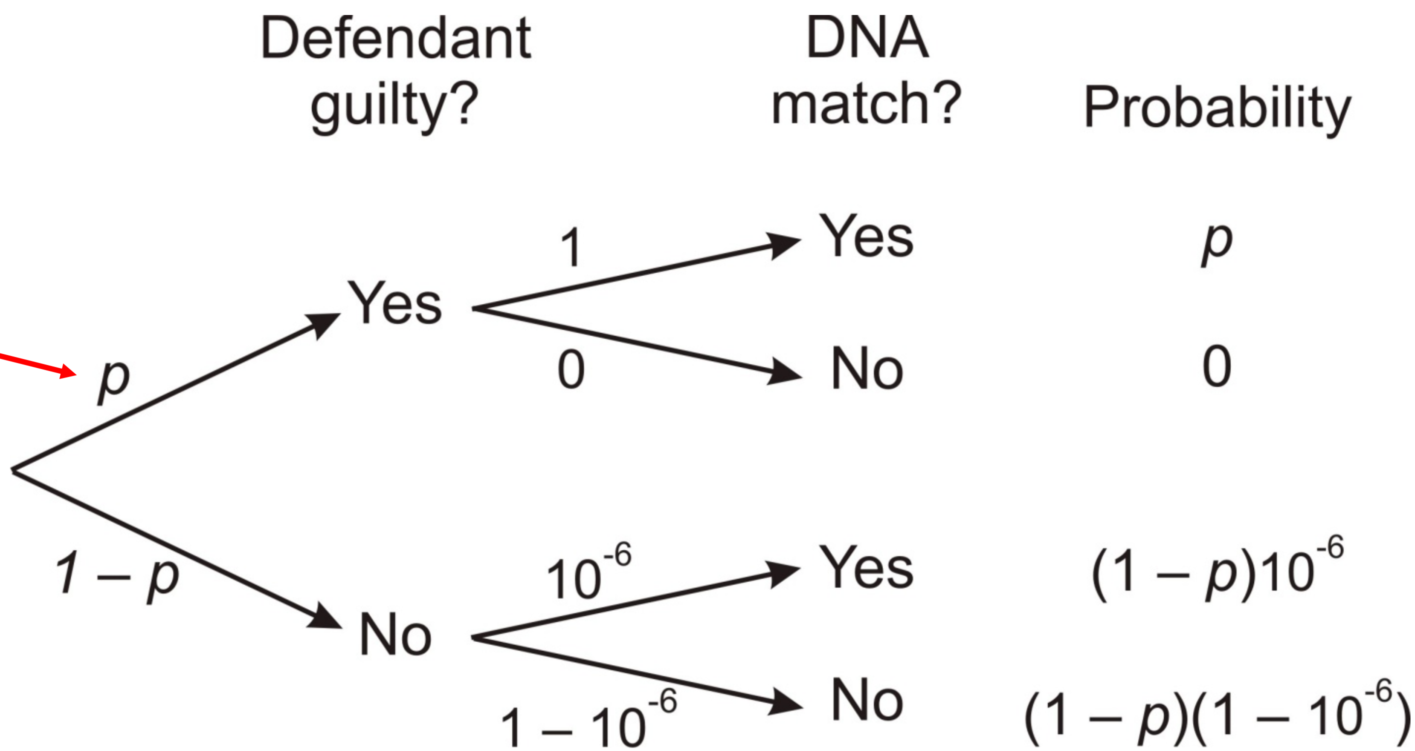
What is the probability of guilt given a positive DNA match?



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

Prior and posterior probability

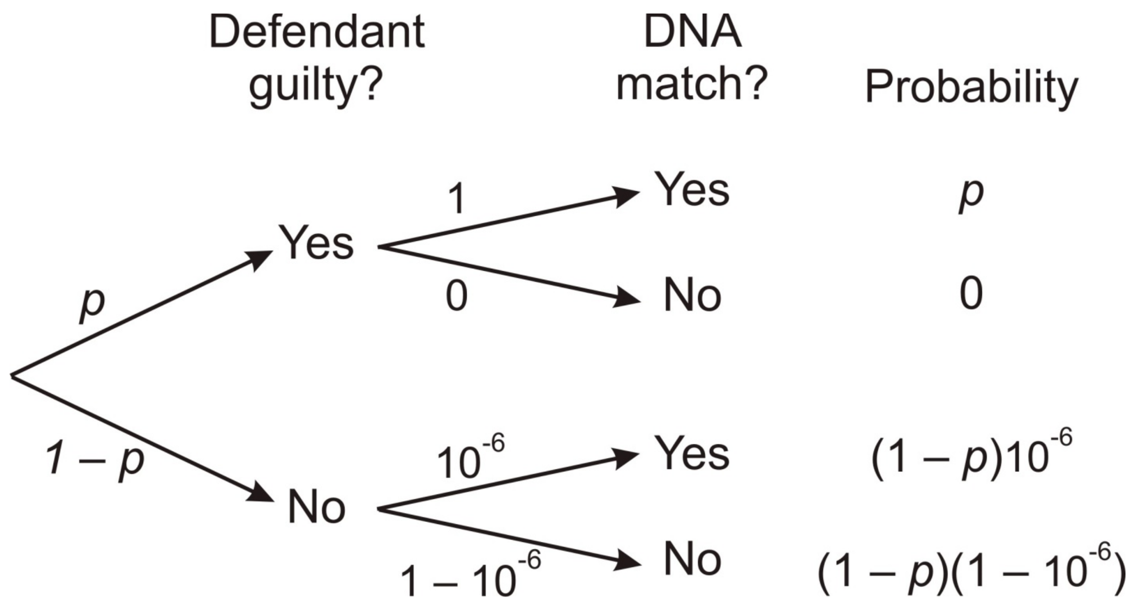
Prior probability of guilt



Posterior probability of guilt

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

Bayesian inference in action



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

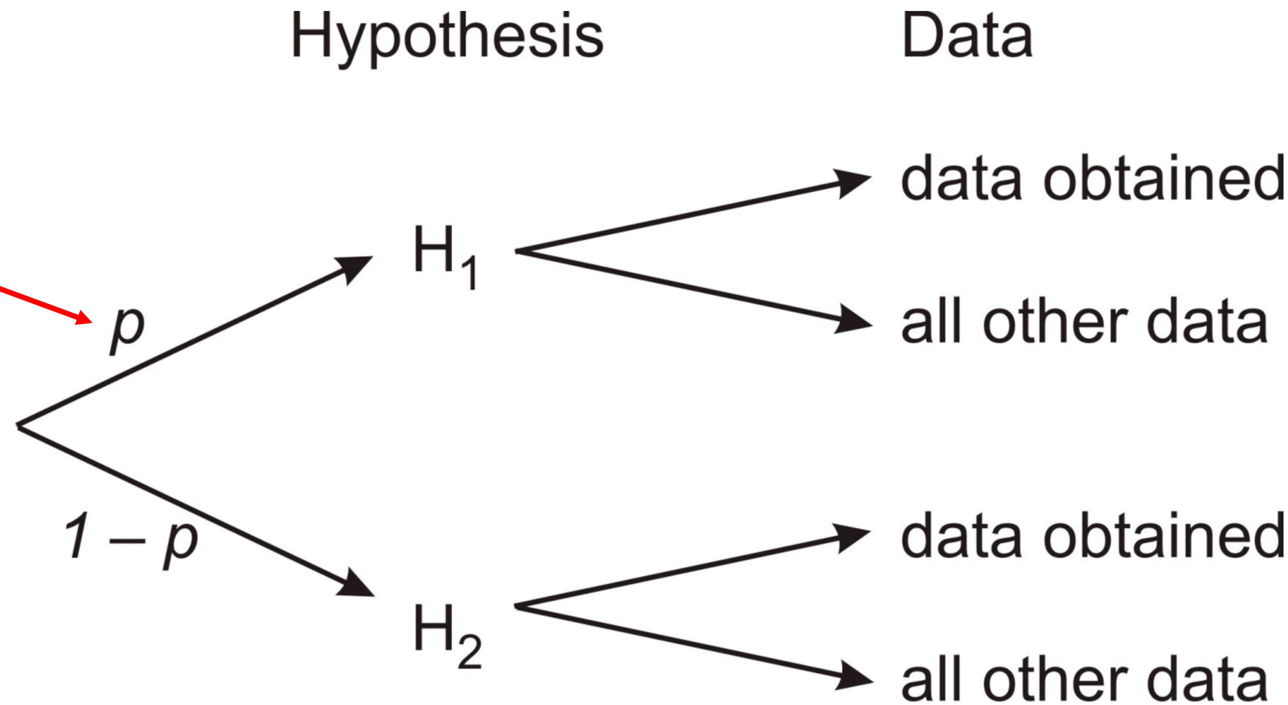
If $p = 10^{-6}$ then $\Pr[\text{guilt} \mid \text{match}] = 0.5$

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

So, is the defendant guilty or innocent?

Bayesian inference with data

Prior probability



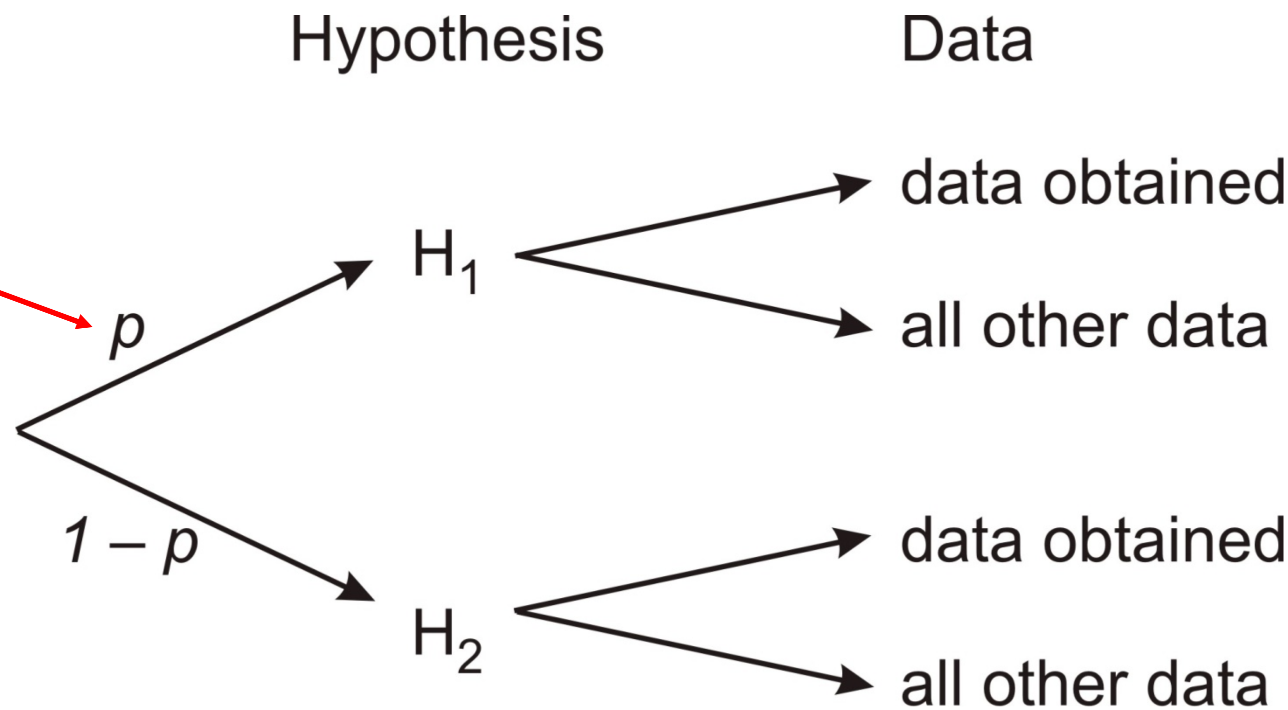
Posterior probability

$$\Pr[\mathrm{H}_1 \mid \mathrm{data}] = \frac{\Pr[\mathrm{data} \mid \mathrm{H}_1]\ \Pr[\mathrm{H}_1]}{\Pr[\mathrm{data} \mid \mathrm{H}_1]\ \Pr[\mathrm{H}_1] + \Pr[\mathrm{data} \mid \mathrm{H}_2]\ \Pr[\mathrm{H}_2]}$$

Bayesian inference goes beyond likelihood

Pr[data|H_1] is the likelihood of H_1 given the data

Prior probability



Posterior probability

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

How Bayesian inference is different from what we usually do

The prior probability represents the investigator's strength of belief about the hypothesis, or strength of belief about the parameter value, before the data are gathered.

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

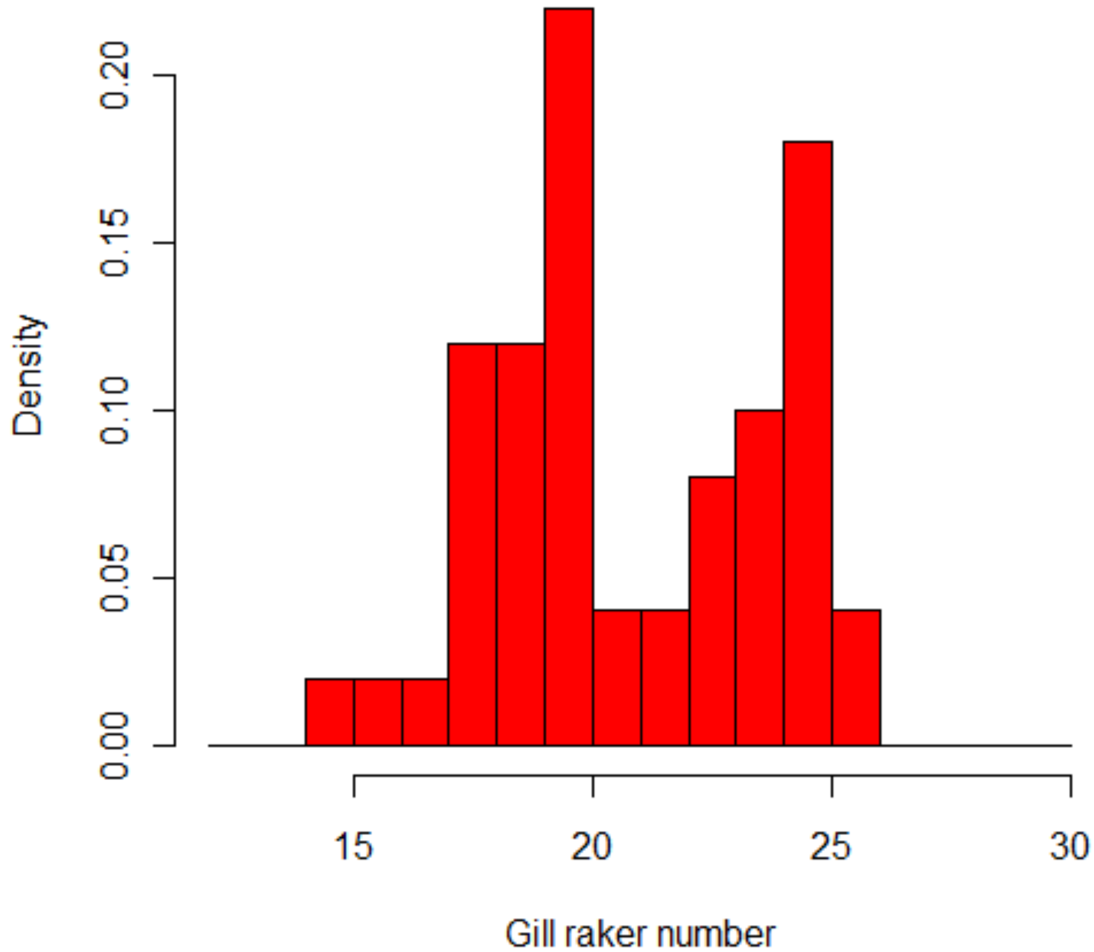
Mathematically, the value of the hypothesis or parameter is treated as though it is a *random variable* that has a probability distribution.

Here are several examples of how it works in practice.

Example 1: One species or two

Data: Gill raker counts for 50 fish collected from a new lake

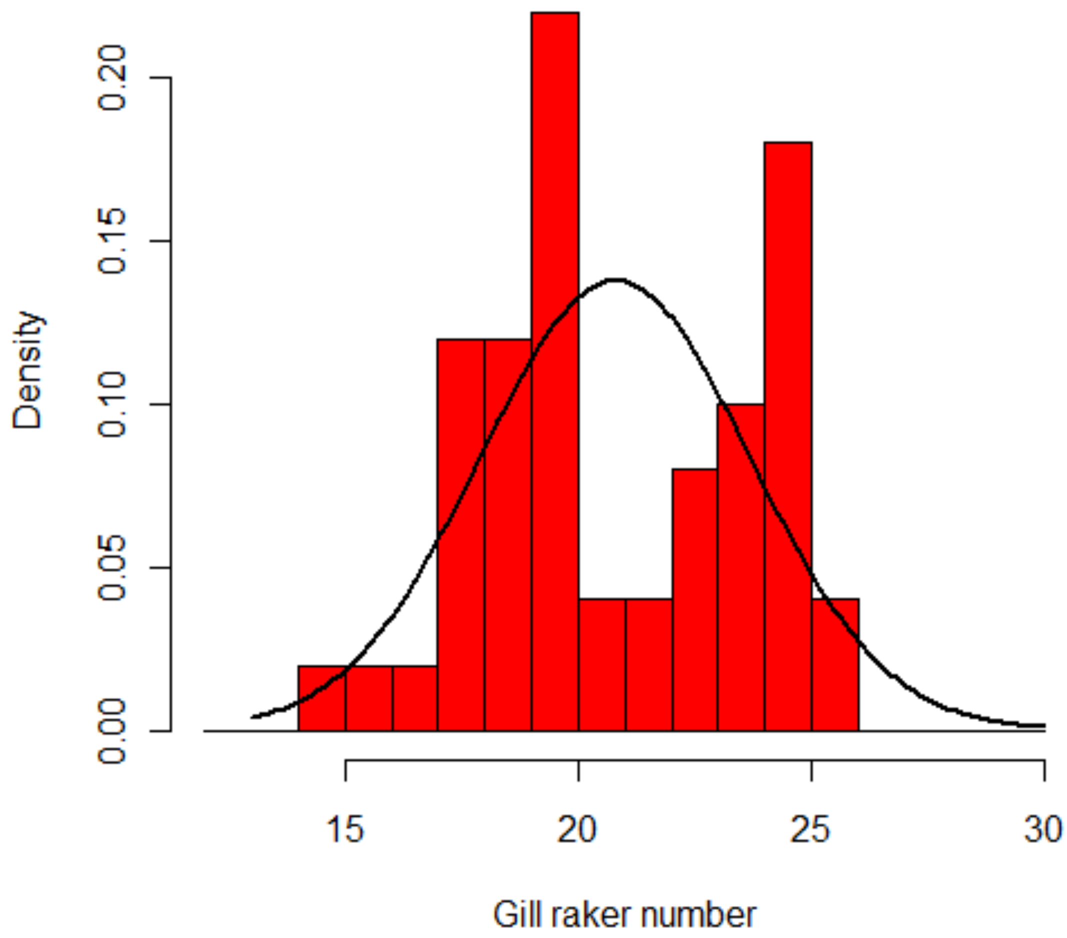
What is the probability that the counts represent 2 species rather than 1?



H_1: one species

Assume a normal distribution of measurements

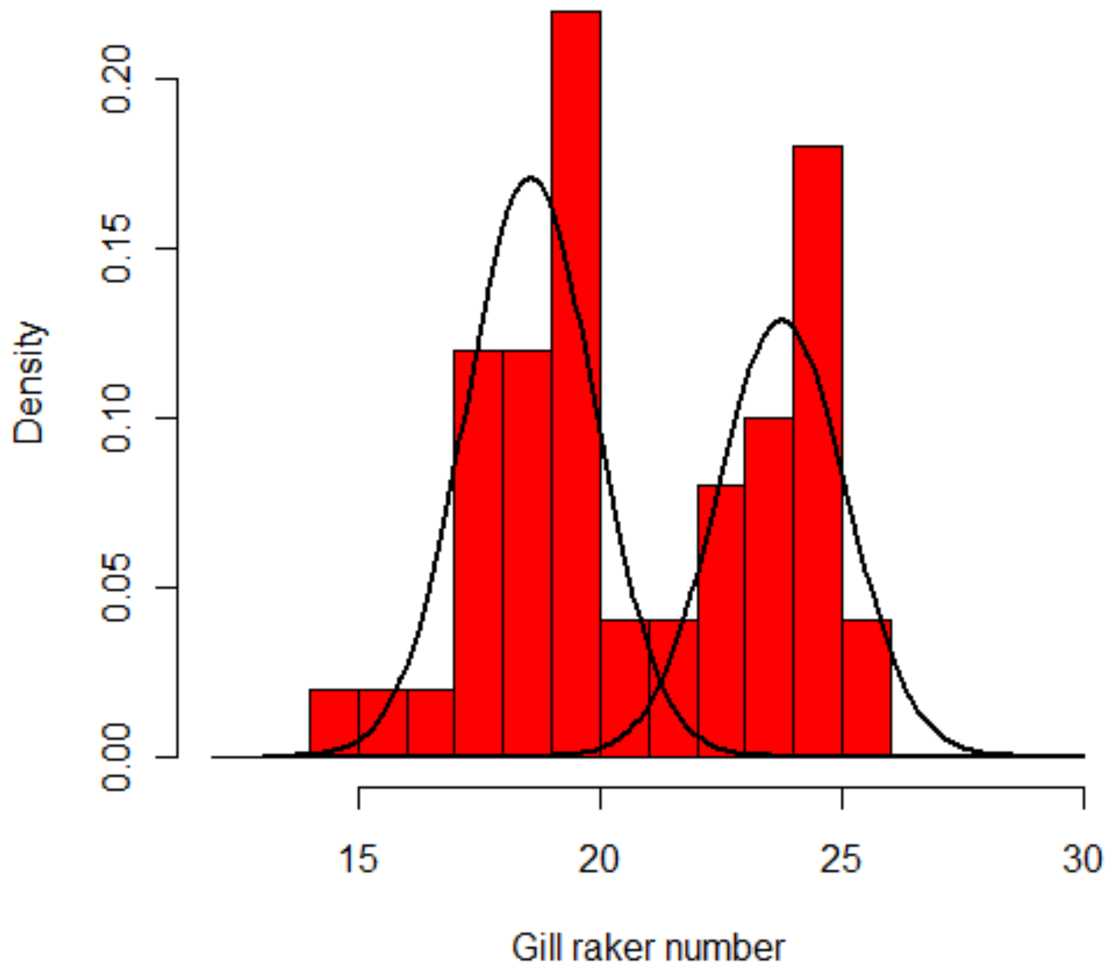
$$\Pr[\text{data} \mid H_1] = L[H_1 \mid \text{data}] = e^{-124.06}$$



H_2: two species

Assume normal distributions with equal variance in both groups

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



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):

Prior probability $\Pr[H_2]$	**Posterior probability** $\Pr[H_2 \mid \text{data}]$
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.*)

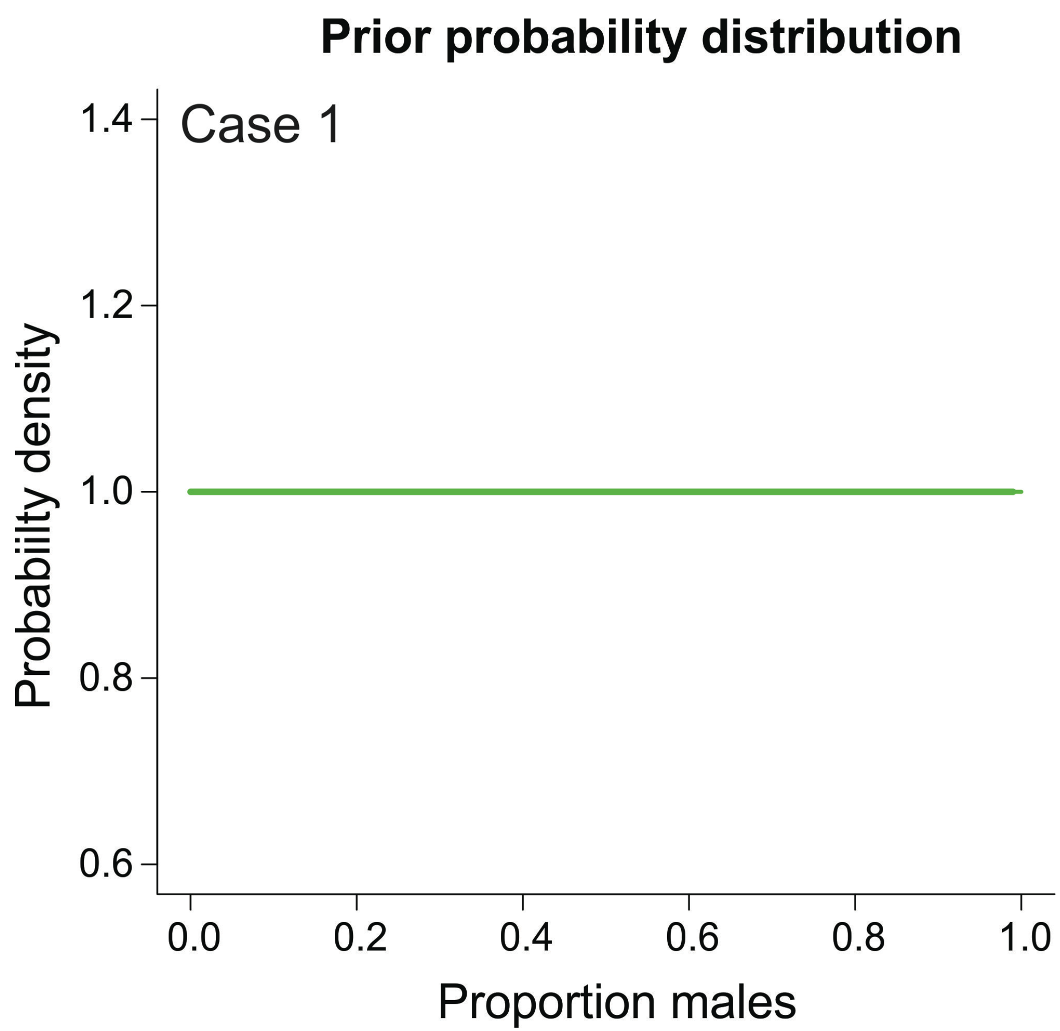
What is the proportion of males among the reproductive adults emerging from colonies?

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

Bayesian estimation of a proportion

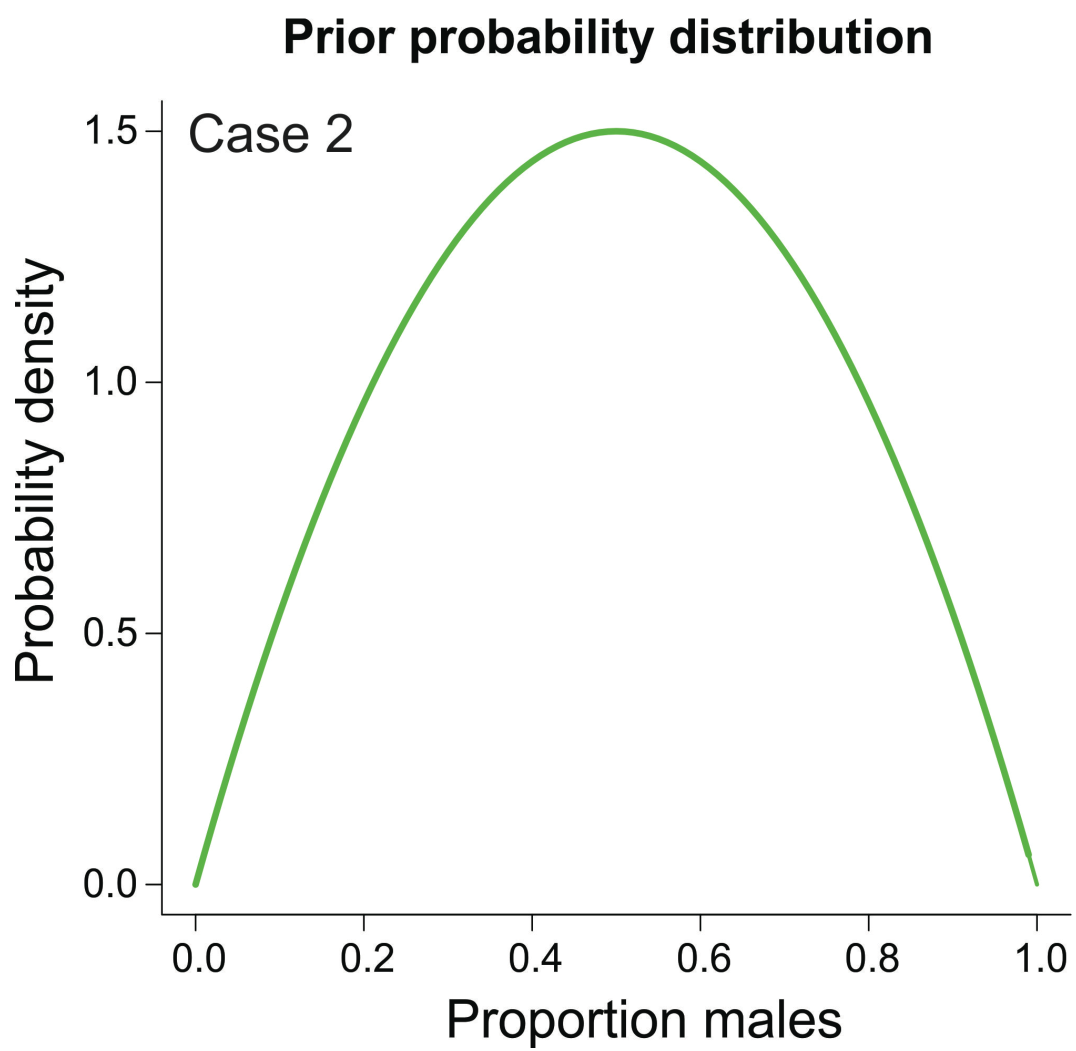
To begin, we need to come up with a prior probability distribution for the proportion.

Case 1: the "noninformative" prior: expression of total ignorance.



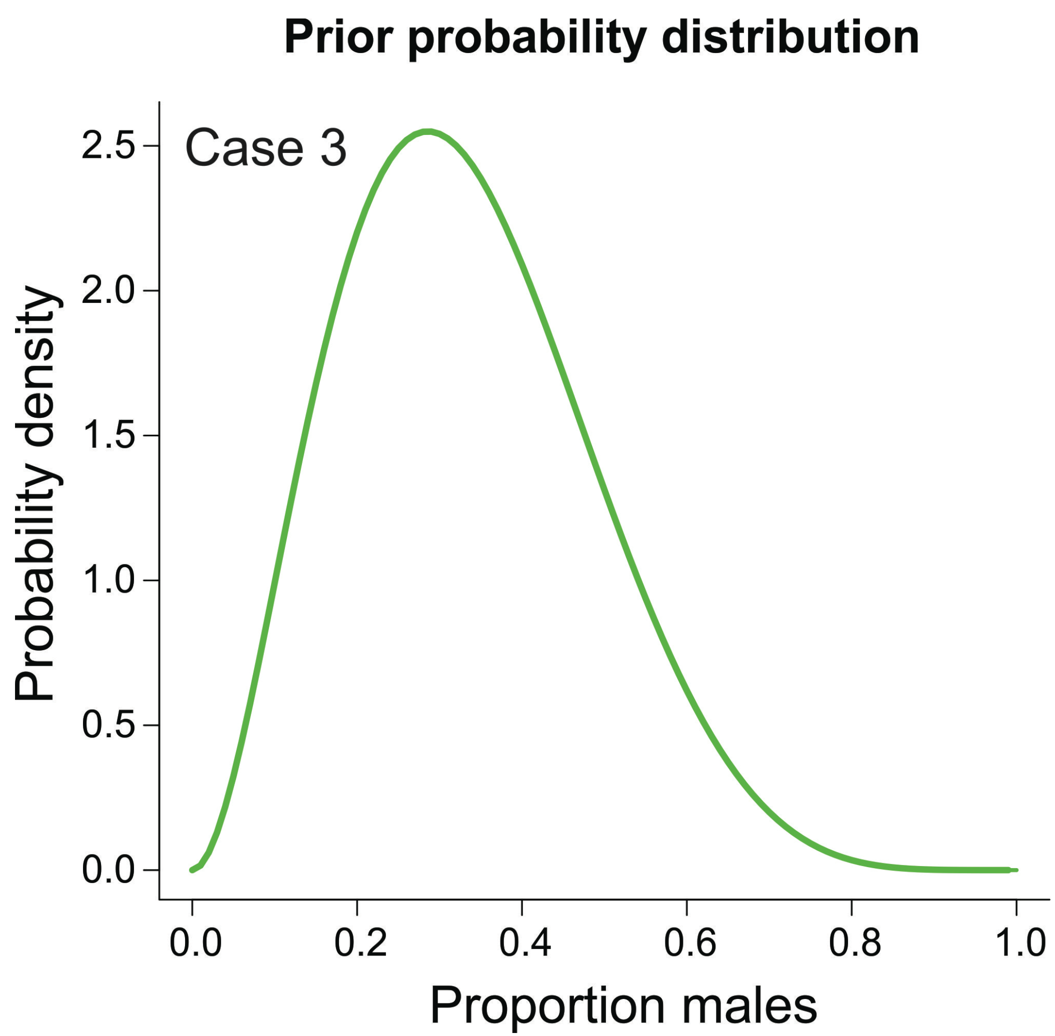
Bayesian estimation of a proportion

Case 2: Most species have a sex ratio close to 50:50, and this is predicted by simple sex-ratio theory. This prior probability distribution attempts to incorporate 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 incorporate this previous information.



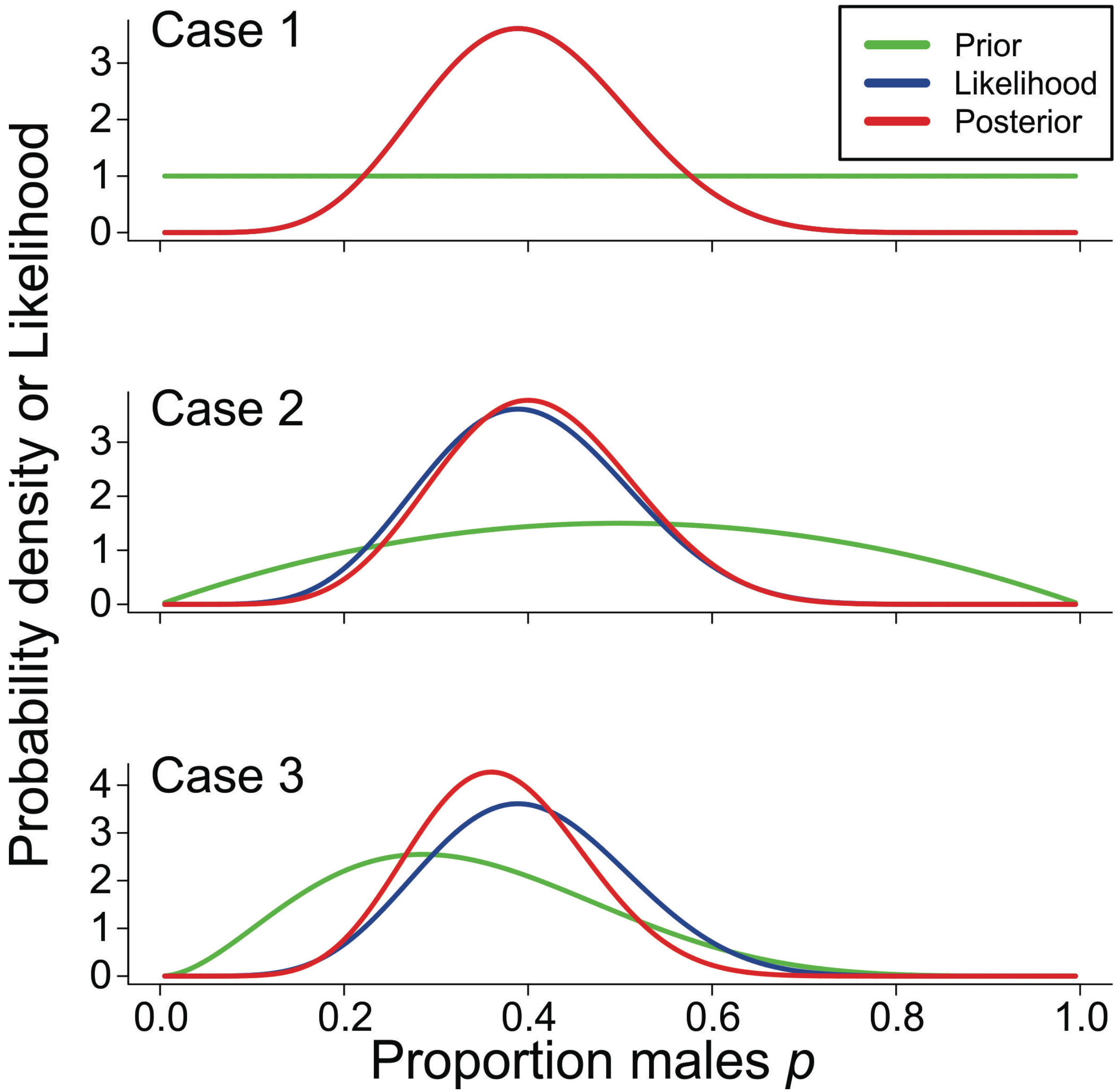
Bayesian estimation of a proportion

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

$\hat{p} = 0.39$

$\hat{p} = 0.40$

$\hat{p} = 0.36$



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 subjective. Different researchers may use different priors, hence obtain different estimates with the same data.

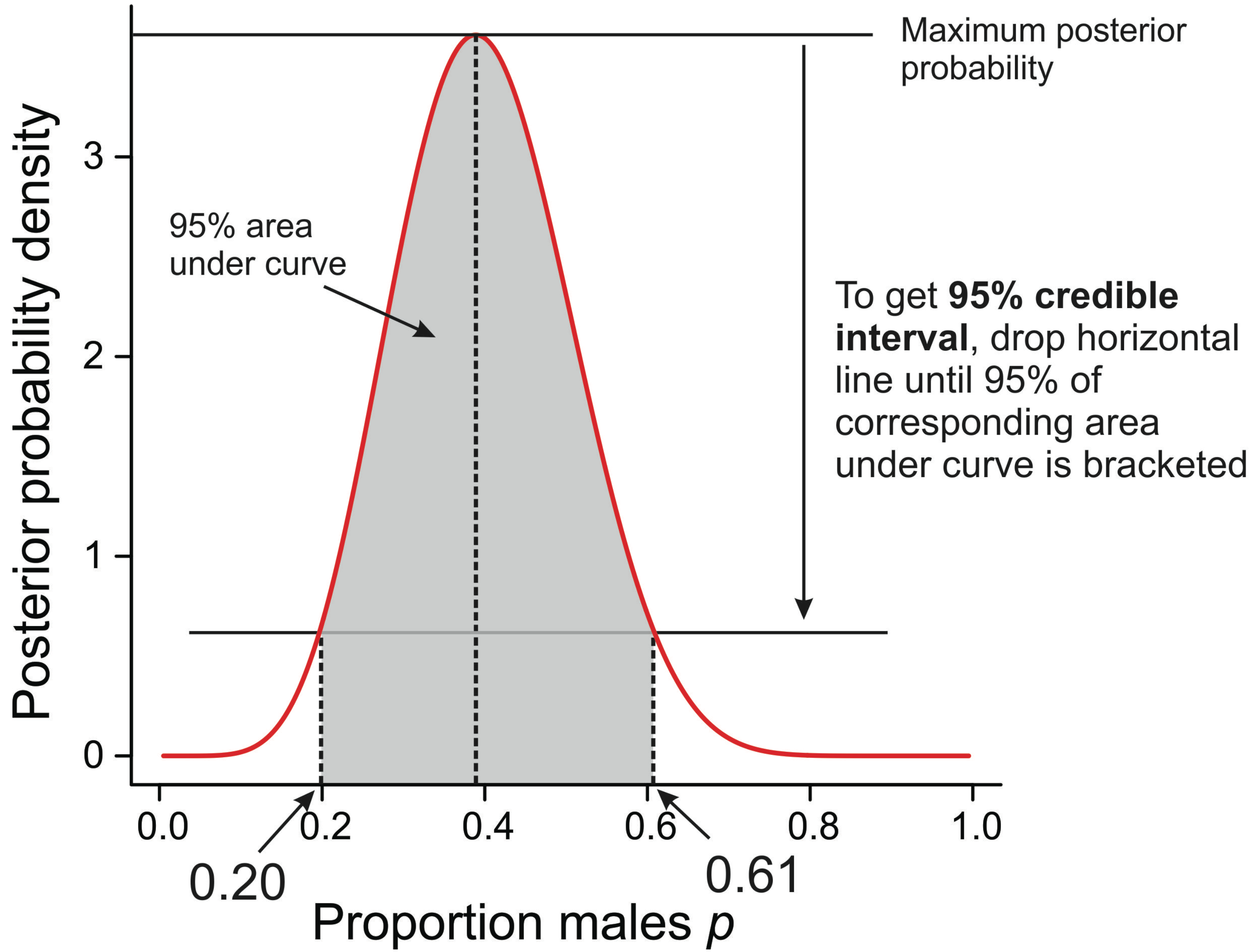
To resolve this we might all agree to use “noninformative” priors. But this stance prevents us from incorporating prior information, which is regarded as one of the strengths of the Bayesian approach.

Maybe the issue about the subjectivity of priors can be resolved if we base the prior on a survey of preexisting evidence (lot of work).

Choice of prior not so important if there is a lot of data.

Bayesian estimation of a proportion

95% credible interval



Bayesian estimation of a proportion
Interpretation of the interval estimates

95% likelihood-based confidence interval: $0.19 < p < 0.62$

Likelihood interpretation:
Most plausibly, *p* is between 0.19 and 0.62. In repeated random samples taken from the same population, the likelihood-based confidence interval so calculated will bracket the true population proportion *p* approximately 95% of the time.

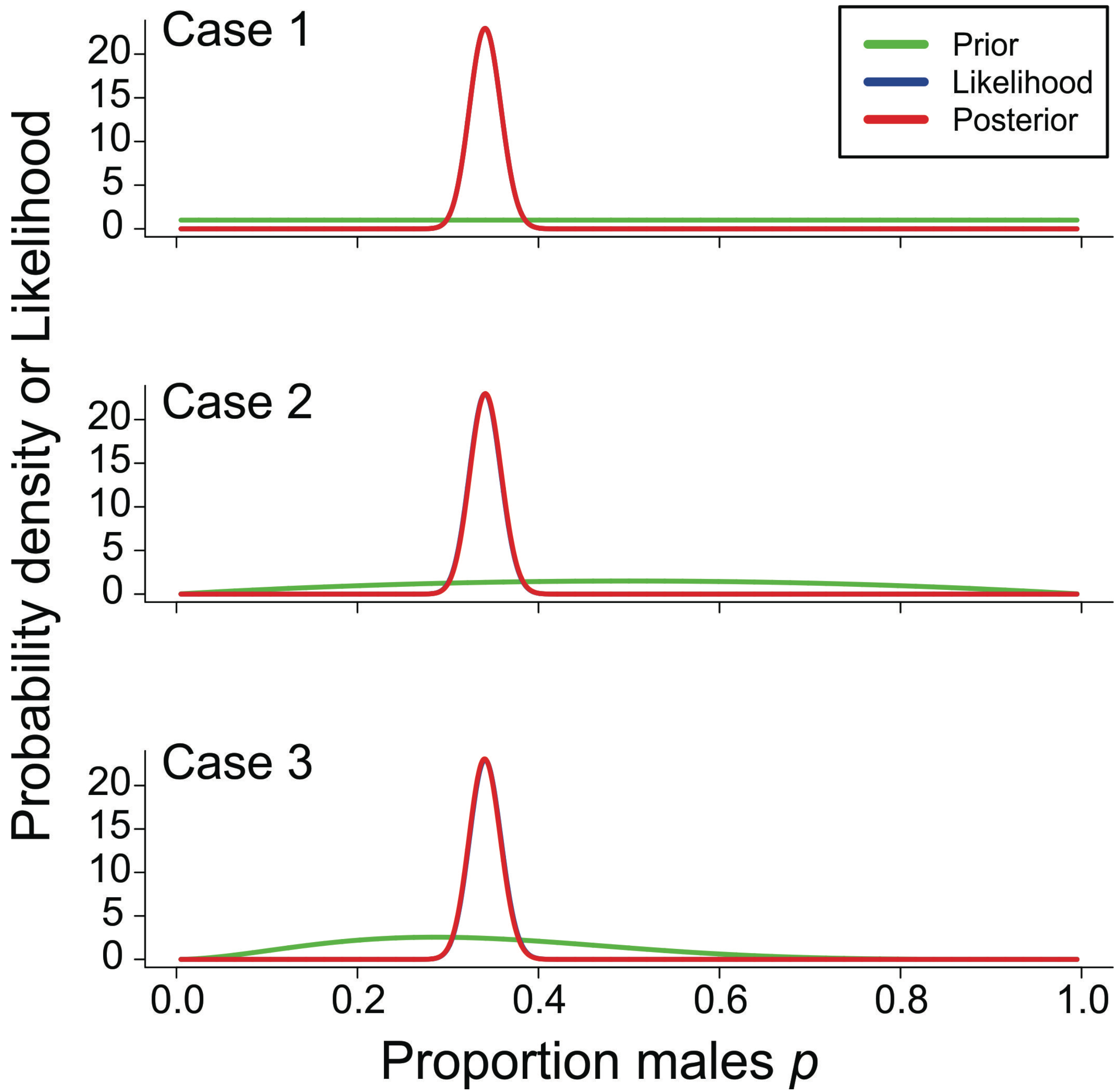
95% credible interval: $0.20 < p < 0.61$
(assuming Case 1, with non-informative prior)

Bayesian 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 $\hat{p}_{\text{MLE}} = 0.34$

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



Bayesian hypothesis testing using the Bayes factor

Bayesian methods can be used to quantify the strength of evidence for one hypothesis relative to another using a quantity called the *Bayes factor*. This represents a Bayesian alternative to null hypothesis significance testing.

Within the Bayesian framework, one can calculate the weight of evidence for one hypothesis relative to another, given the data. The Bayes factor is commonly used to quantify this.

For example, when comparing means of two groups, we can still consider a null and alternative hypothesis.

$\mathrm{H}_0: \mu_1 - \mu_2 = 0$

$\mathrm{H}_\mathrm{A}: \mu_1 - \mu_2 \neq 0$

The difference from conventional statistics is that here the null hypothesis has no precedence.

Bayesian hypothesis testing using the Bayes factor

Before seeing the data, the two hypotheses H_0 and H_A are given the prior probabilities $Pr(H_0)$ and $Pr(H_A)$.

Once the data are observed, the Bayes factor is the ratio of the posterior probabilities

$$\text{Bayes factor} = \frac{\Pr[\mathrm{H_A}|\mathrm{data}]}{\Pr[\mathrm{H_0}|\mathrm{data}]} = \frac{\Pr[\mathrm{data}|\mathrm{H_A}]}{\Pr[\mathrm{data}|\mathrm{H_0}]} \times \frac{\Pr[\mathrm{H_A}]}{\Pr[\mathrm{H_0}]}$$

$\Pr[\mathrm{data}|\mathrm{H_0}]$ is the straightforward likelihood of $\mathrm{H_0}]$, since $\mu_1 - \mu_2$ under $\mathrm{H_0}$ is a single point (0). It is simple to compute.

$\Pr[\mathrm{data}|\mathrm{H_A}]$ is a little more complicated because we have to integrate the likelihood over the probability distribution of values for $\mu_1 - \mu_2$ (computer packages do this).

Bayesian hypothesis testing using the Bayes factor

If two hypotheses H_0 and H_A have the same prior probability, i.e., $\Pr(H_0) = \Pr(H_A)$, then the Bayes factor is just

$$\text{Bayes factor} = \frac{\Pr[\text{data}|\text{H}_\text{A}]}{\Pr[\text{data}|\text{H}_0]}$$

The remaining question is then to decide what constitutes strong evidence in favor of the alternative hypothesis.

A Bayes factor of 1 – 3 is considered “anecdotal evidence” for H_A
A Bayes factor of 3 – 10 is considered “substantial evidence” for H_A
A Bayes factor of 10 – 30 is considered “strong evidence” for H_A

Bayesian hypothesis testing

Weight of evidence, comparing *P*-values from 855 *t*-tests in the psychology literature with corresponding Bayes factors (Wetzels et al. 2011).

Weight of evidence from the two approaches (*t*-test vs Bayes factor) is strongly correlated.

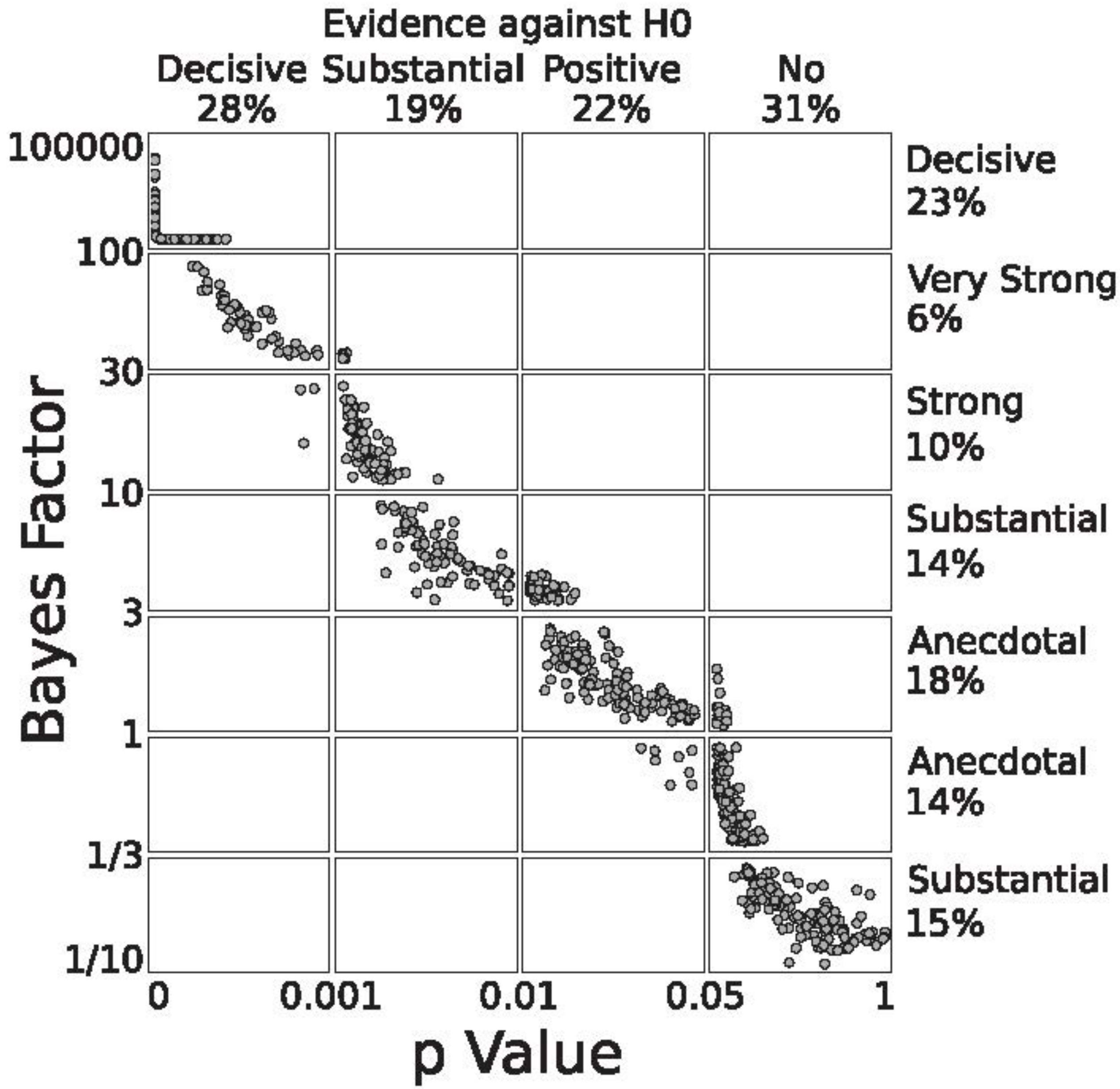


Fig. 3. The relationship between Bayes factor and *p* value. Points denote comparisons (855 in total). The scale of the axes is based on the decision categories, as given in Table 1.

Bayesian hypothesis testing

Weight of evidence, comparing *P*-values from 855 *t*-tests in the psychology literature with corresponding Bayes factors (Wetzels et al. 2011).

But notice how weak is the criterion $P = 0.05$ by the standard of the Bayes factor.

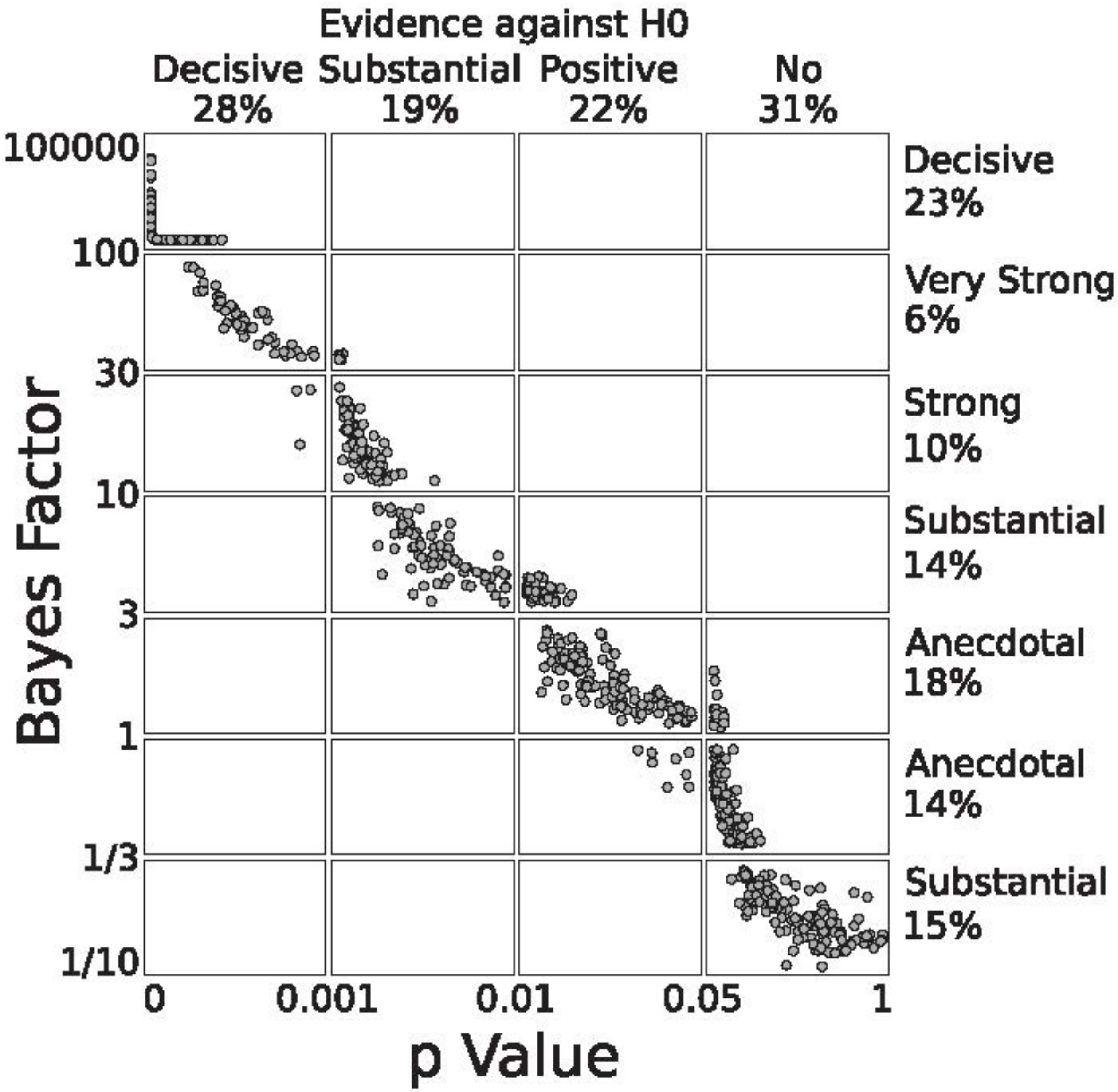


Fig. 3. The relationship between Bayes factor and *p* value. Points denote comparisons (855 in total). The scale of the axes is based on the decision categories, as given in Table 1.

Should we re-think conventional standards?

Revised standards for statistical evidence

Valen E. Johnson[1]

Department of Statistics, Texas A&M University, College Station, TX 77843-3143

Edited by Adrian E. Raftery, University of Washington, Seattle, WA, and approved October 9, 2013 (received for review July 18, 2013)

Recent advances in Bayesian hypothesis testing have led to the development of uniformly most powerful Bayesian tests, which represent an objective, default class of Bayesian hypothesis tests that have the same rejection regions as classical significance tests. Based on the correspondence between these two classes of tests, it is possible to equate the size of classical hypothesis tests with evidence thresholds in Bayesian tests, and to equate *P* values with Bayes factors. An examination of these connections suggest that recent concerns over the lack of reproducibility of scientific studies can be attributed largely to the conduct of significance tests at unjustifiably high levels of significance. To correct this problem, evidence thresholds required for the declaration of a significant finding should be increased to 25–50:1, and to 100–200:1 for the declaration of a highly significant finding. In terms of classical hypothesis tests, these evidence standards mandate the conduct of tests at the 0.005 or 0.001 level of significance.

the average value of the sampling density of the observed data under each of the two hypotheses, averaged with respect to the prior density specified on the unknown parameters under each hypothesis.

Paradoxically, the two approaches toward hypothesis testing often produce results that are seemingly incompatible (13–15). For instance, many statisticians have noted that *P* values of 0.05 may correspond to Bayes factors that only favor the alternative hypothesis by odds of 3 or 4–1 (13–15). This apparent discrepancy stems from the fact that the two paradigms for hypothesis testing are based on the calculation of different probabilities: *P* values and significance tests are based on calculating the probability of observing test statistics that are as extreme or more extreme than the test statistic actually observed, whereas Bayes factors represent the relative probability assigned to the observed data under each of the competing hypotheses. The latter

Bayesian model selection

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.

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

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

Summary

- Bayesian probability is a different concept than frequentist probability
- 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 (credible interval) differs from the frequentist definition (confidence interval)
- Bayesian hypothesis testing using the Bayes factor suggests that we need to raise our standards of evidence.
- Bayesian ideas are becoming used more in ecology and evolution

Discussion paper for next week:

Dochtermann & Jenkins (2011) Multiple hypotheses in behavioral ecology.

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

Presenters: Cam B & Eric

Moderators: Melody & ___________