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

- Hypothesis testing using the Bayes factor

- Bayesian model selection

Elephant population size estimation

Mark-recapture

This example continues one from the likelihood workshop. Eggert et al. (2003. Molecular Ecology 12: 1389-1402) used mark-recapture methods to estimate the total number of forest elephants inhabiting Kakum National Park in Ghana by sampling dung and extracting elephant DNA to uniquely identify individuals. Over the first seven days of collecting the researchers identified 27 elephant individuals. Refer to these 27 elephants as *marked*. Over the next eight days they sampled 74 individuals, of which 15 had been previously marked. Refer to these 15 elephants as *recaptured*.

Probability model

Provided the assumptions are met (no births, deaths, immigrants, or emigrants while the study was being carried out; the dung of every elephant had an equal chance of being sampled, and selection of each individual was independent of the others), then the number of recaptured (previously marked) individuals X in the second sample should have a hypergeometric distribution with parameters k (the size of the second sample of individuals), m (total number of marked individuals in the population when the second sample was taken), and n (total number of unmarked individuals in the population at the time of the second sample).

Likelihood

1. Using the hypergeometric distribution, calculate the log-likelihood of each of a range of possible values for the number of elephants in the Park. Note that the total number of elephants is $n + m$, and that m is known ($m = 27$). Note also that only integer values for n are allowed, and that n cannot be smaller than $k - X$, the observed number of unmarked individuals in the second sample.

2. Plot the log-likelihood function. Find the value of n that maximizes the likelihood. Add m to this number to obtain the maximum likelihood estimate for population size.*

3. Calculate the likelihood-based 95% confidence interval for the total number of elephants.**

* 133

** $104 < N < 193$

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 <u>random trial</u> 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 1 standard deviation?

- 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, the source of uncertainty is <u>sampling error</u>.

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?
  [not sure – could this be viewed as a cumulative outcome of random trials?]

In these examples there is no random trial, so no sampling error. The source of uncertainty is lack of <u>information</u>—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)

Bayes' Theorem itself is harmless

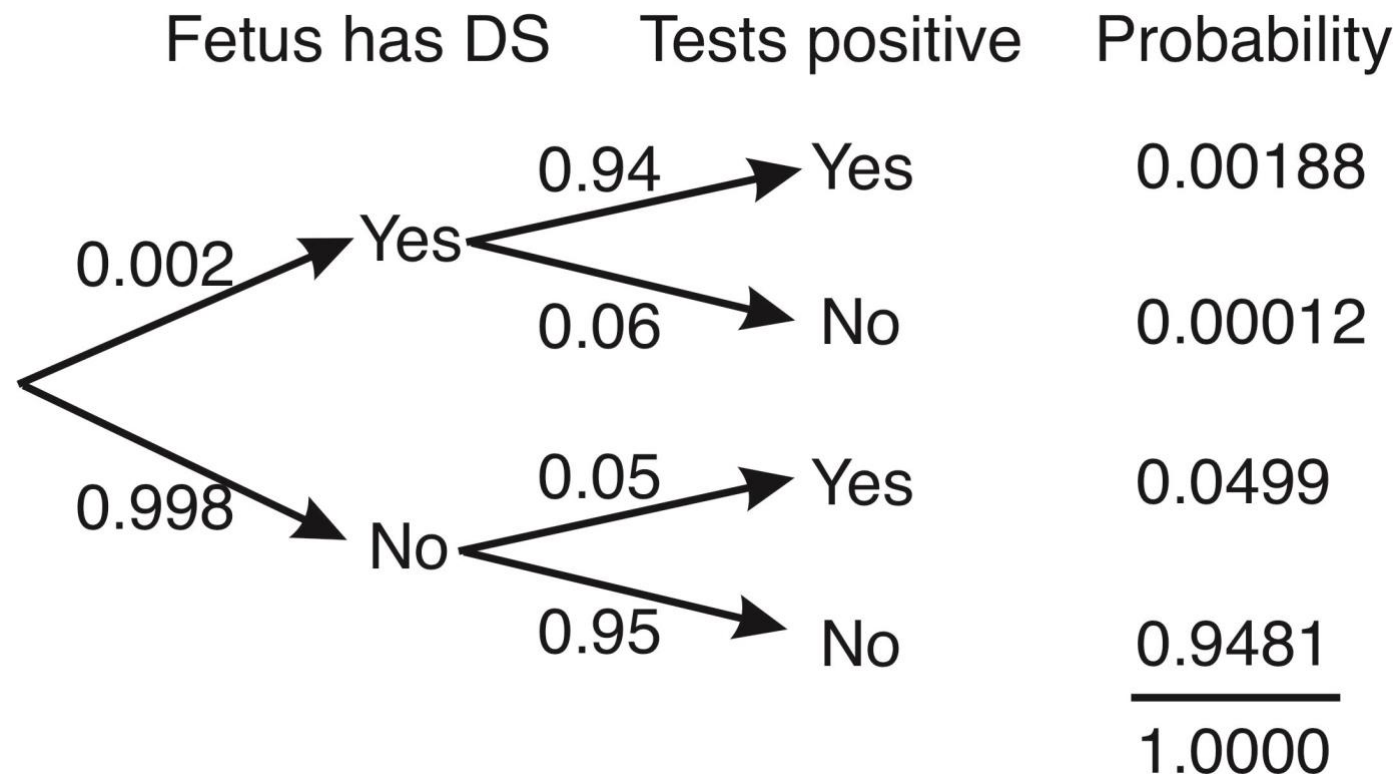Example: screening for Down syndrome (DS).

DS occurs in about 1 in 500 pregnancies. A "full integrated" combination test using 3 blood sera is widely used. It is cheap and risk-free. A newer DNA test is more accurate but not yet widely used.

Fetus has DS   Tests positive   Probability

Yes → Yes
Yes → No

No → Yes
No → No
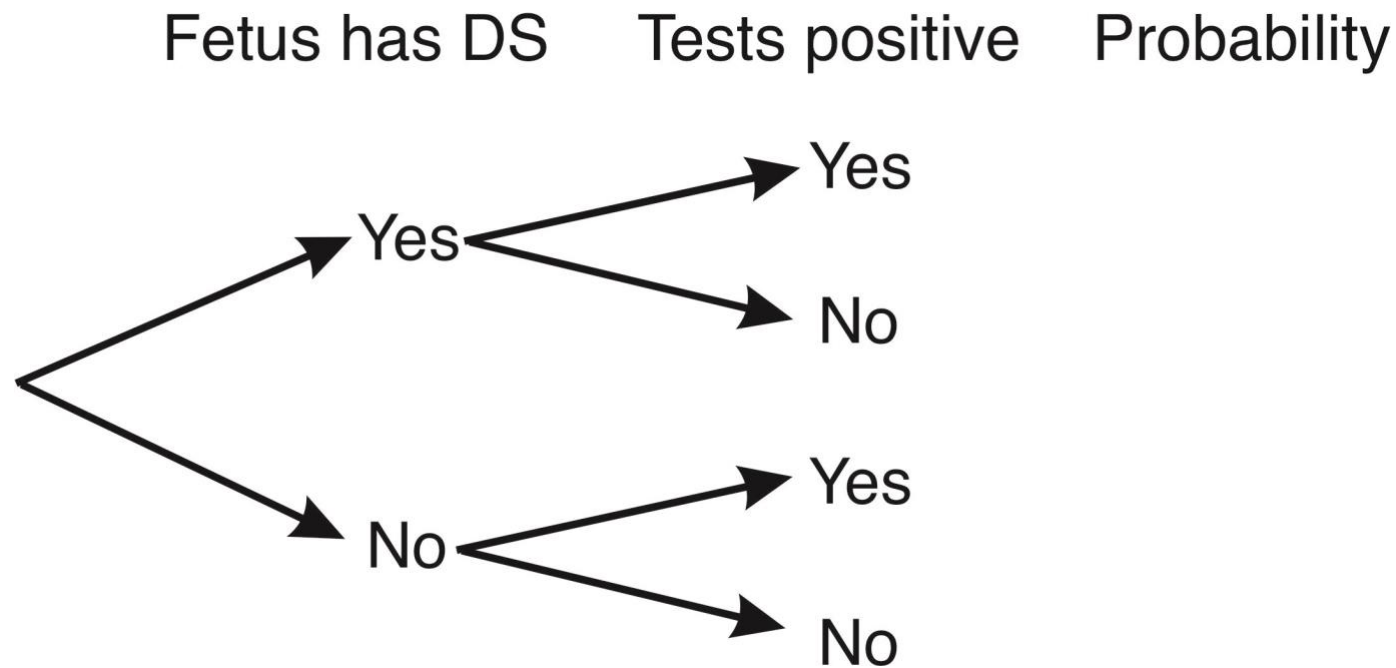
Conditional probability

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

The probability of a positive test result from the integrated test is 0.94, <u>given</u> that a fetus has DS. The probability of a positive result is 0.05, <u>given</u> that a fetus is not DS.
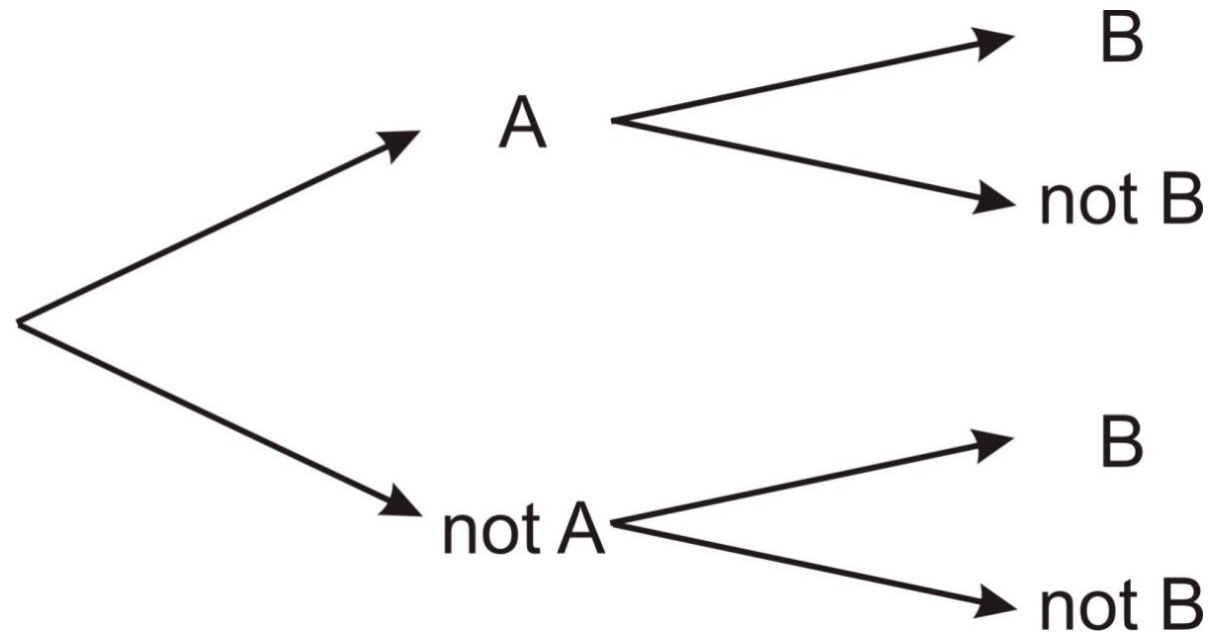
Fetus has DS	Tests positive	Probability
	0.94 → Yes	0.00188
0.002 → Yes		
	0.06 → No	0.00012
	0.05 → Yes	0.0499
0.998 → No		
	0.95 → No	0.9481
		1.0000

Conditional probability calculation

What is the probability that a fetus has DS given that the test is positive?
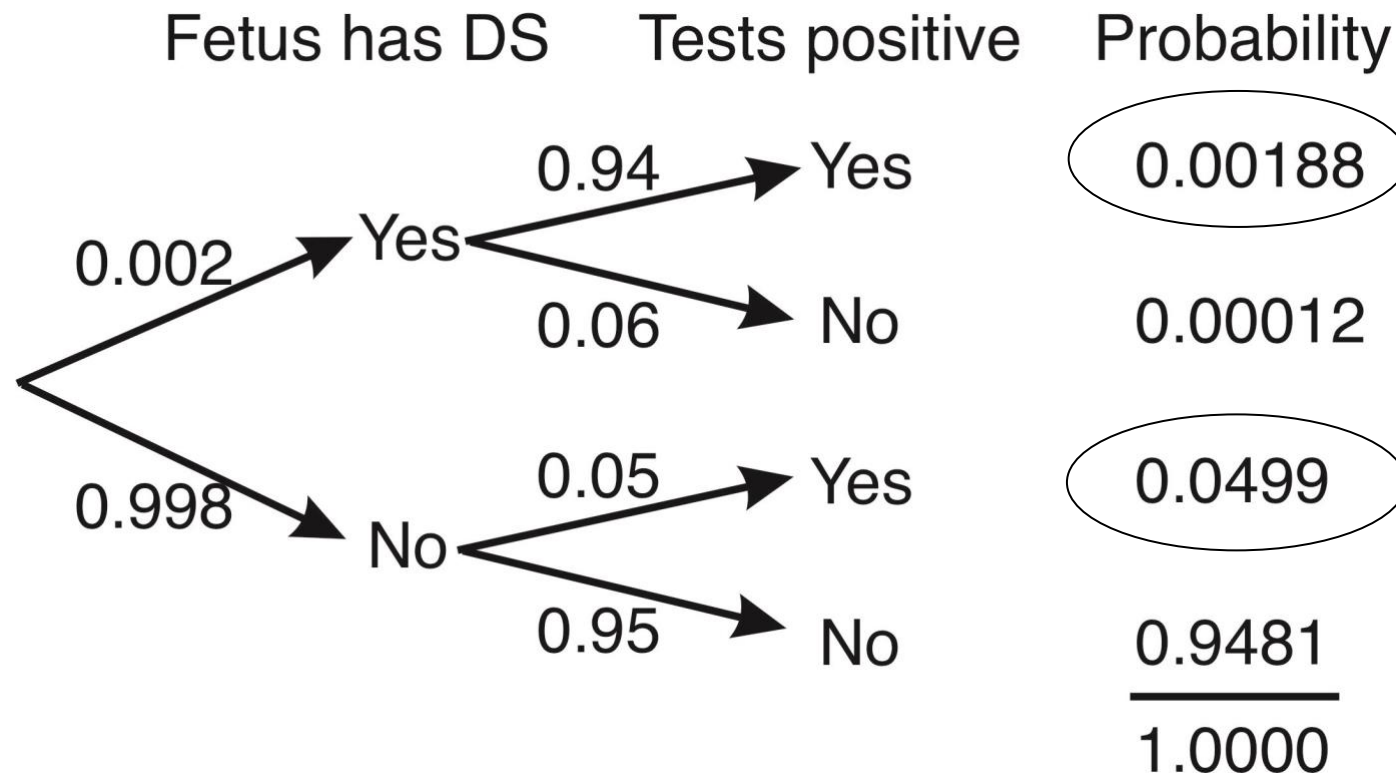
$$\Pr[DS \mid positive] = \frac{0.00188}{0.00188 + 0.0488} = 0.0363, \text{ less than } 4\%$$

**This calculation is formalized in Bayes' Theorem**

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

How Bayes' Theorem can be used

Hypothetical 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.

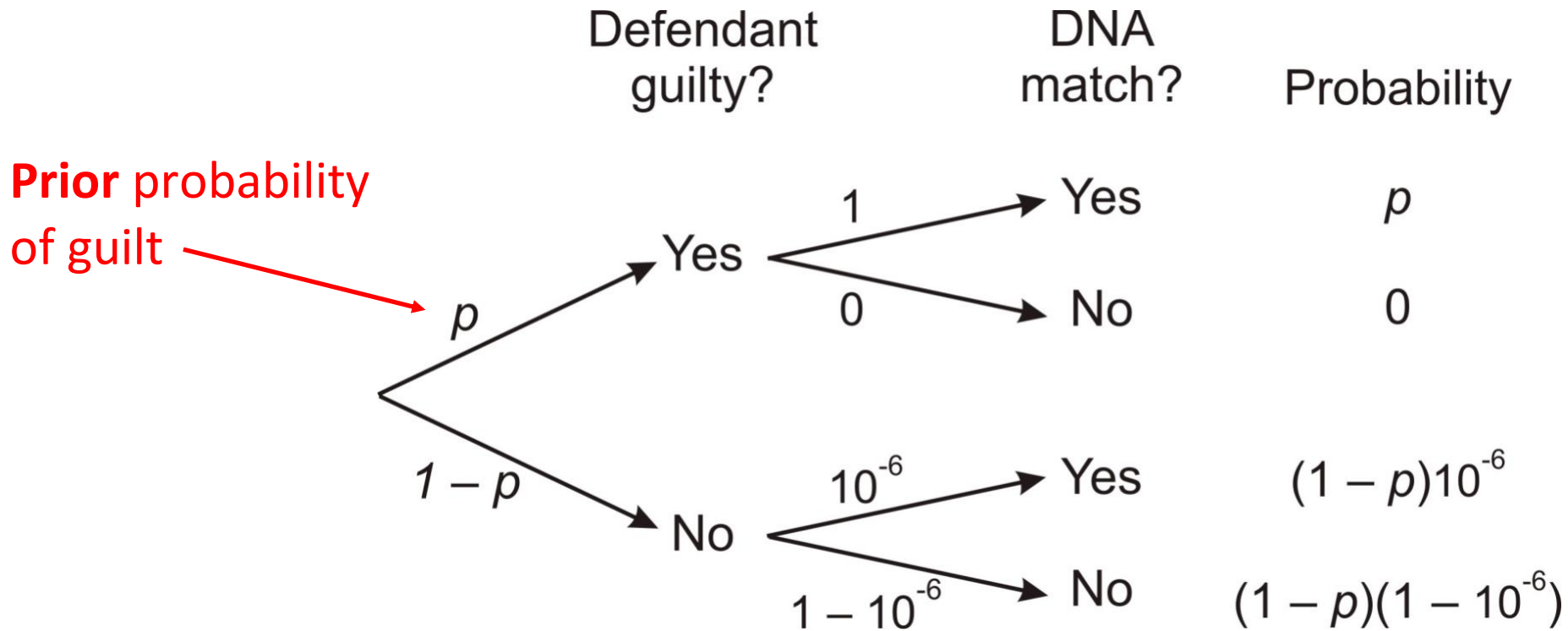What is the probability of guilt given a positive DNA match (assuming no contamination of samples)?
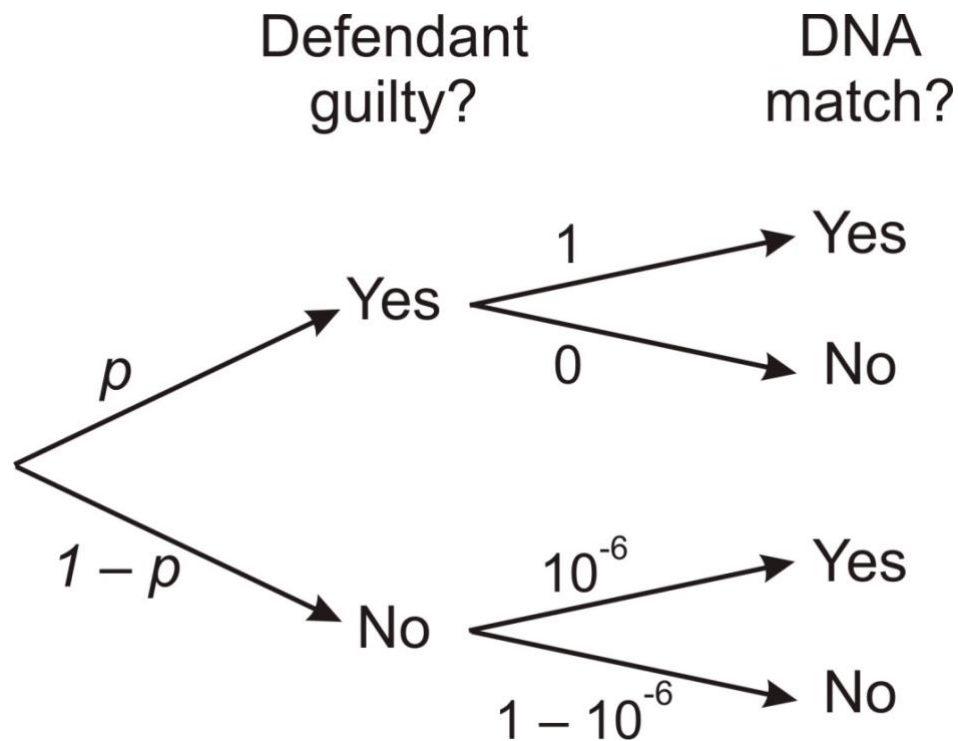
Bayesian inference in action

What is the probability of guilt given a positive DNA match?
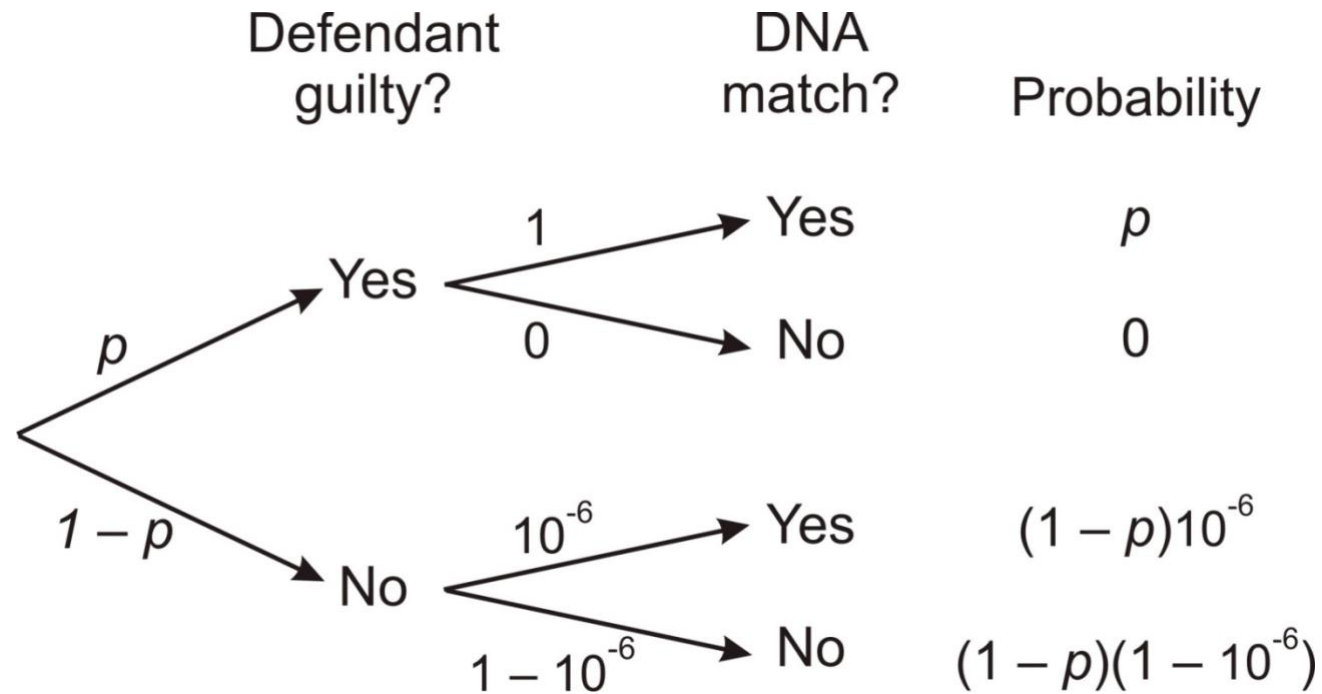
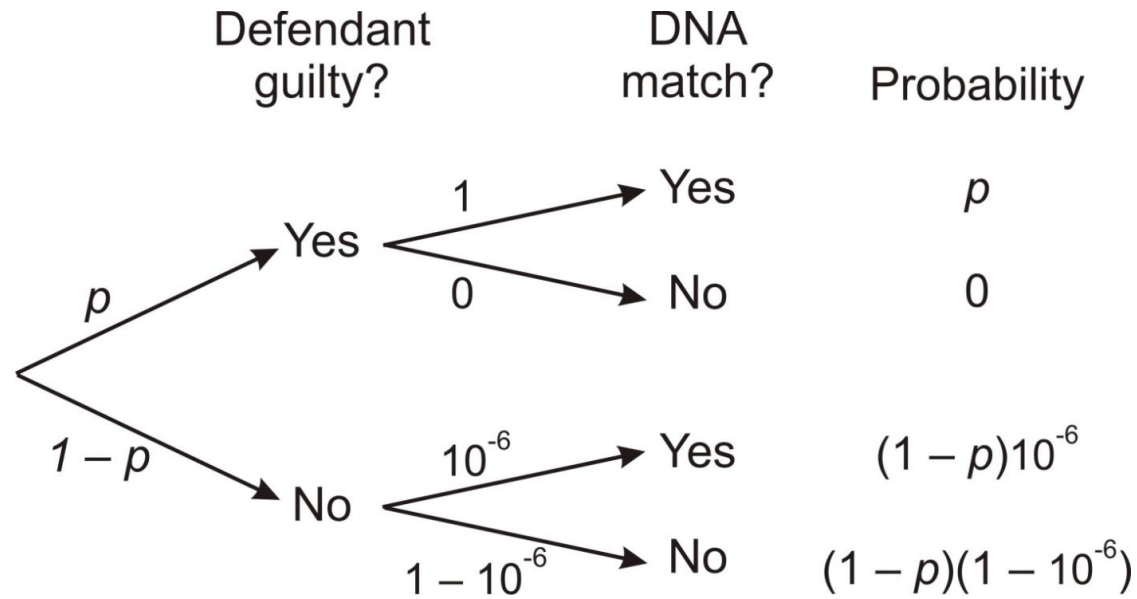$$\Pr[\text{guilt} \,|\, \text{match}] = \frac{1(p)}{1(p) + 10^{-6}(1-p)}$$

Prior and posterior probability

Prior probability of guilt →

Defendant guilty?		DNA match?	Probability

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

Posterior probability of guilt →

Bayesian inference in action

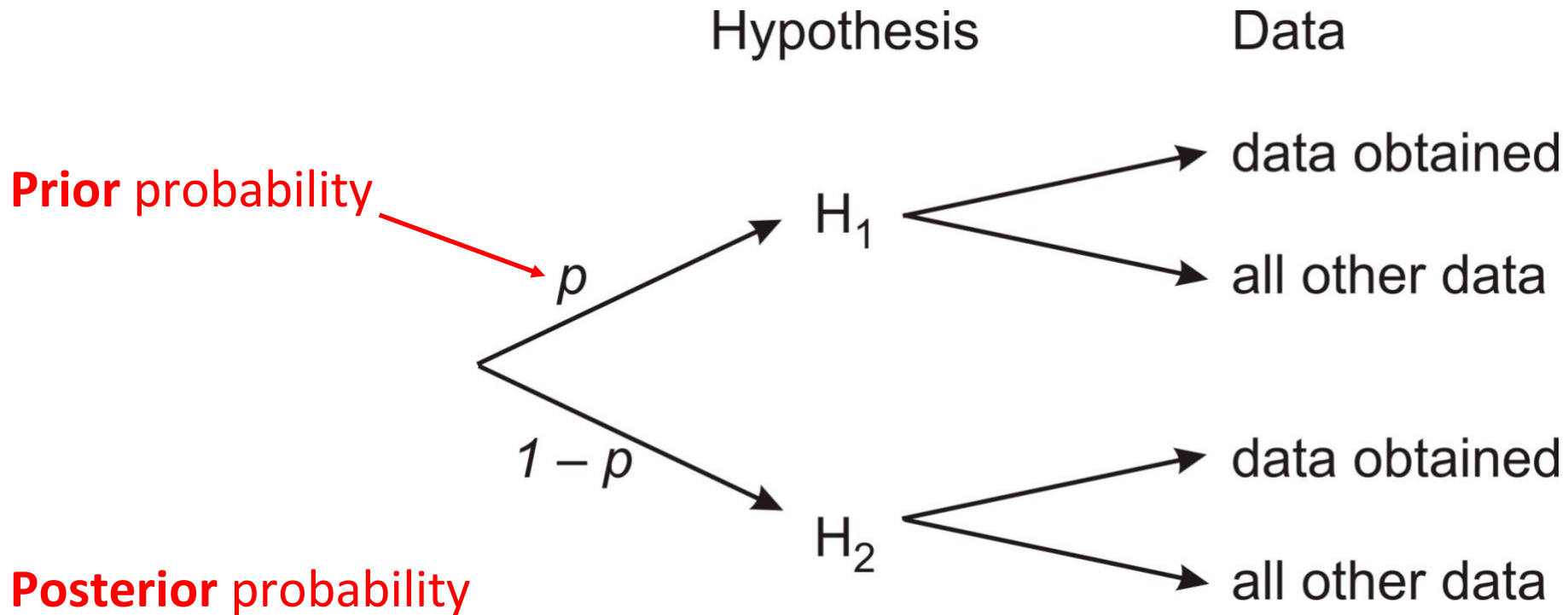$$\Pr[\text{guilt} \,|\, \text{match}] = \frac{1(p)}{1(p) + 10^{-6}(1 - p)}$$

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

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

So, is the defendant guilty or innocent?
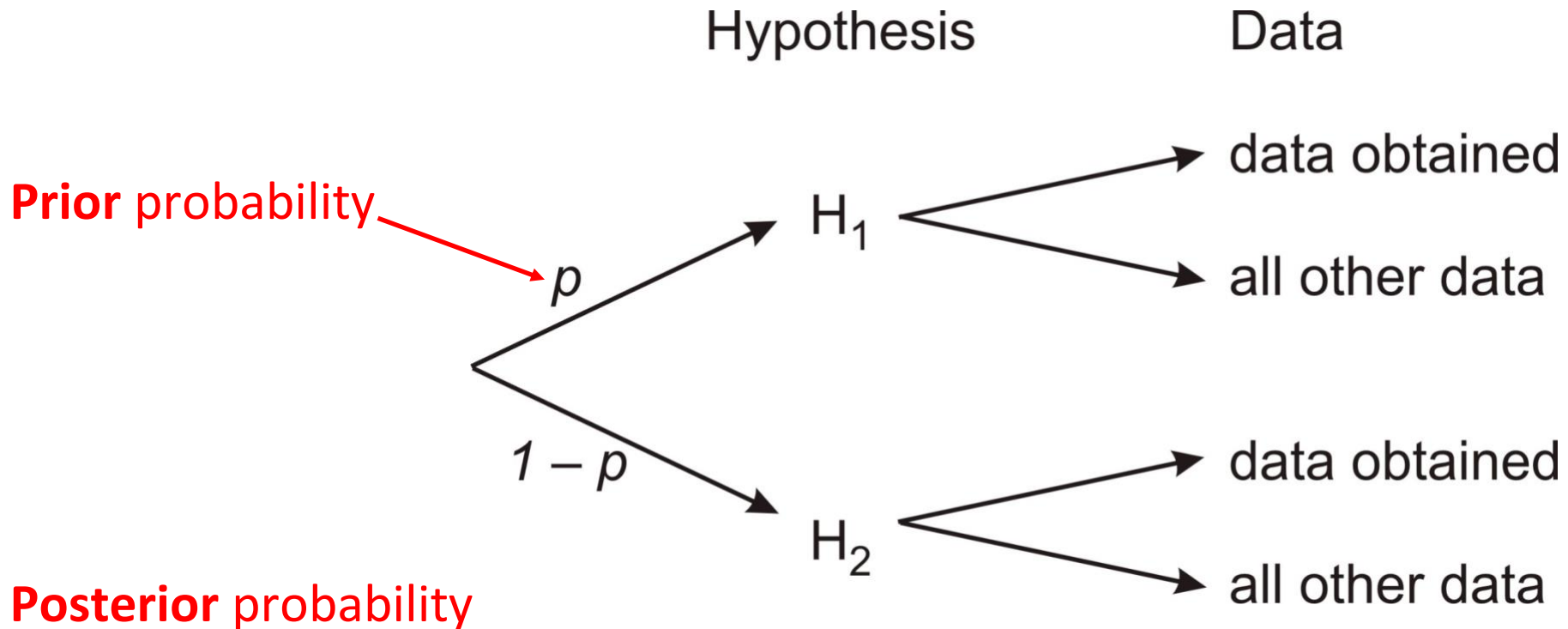
Bayesian inference with data

Prior probability

Posterior probability

Hypothesis · · · Data

H_1 → data obtained

H_1 → all other data

p

$1 - p$

H_2 → data obtained

H_2 → all other data

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

Bayesian inference goes beyond likelihood

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

Hypothesis   Data

Prior probability

                                → data obtained
                    H_1
                                → all other data
        p

        $1 - p$
                                → data obtained
                    H_2
                                → all other data

Posterior probability

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

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 hypothesis or parameter is treated as though it is a *random variable* that has a probability distribution, rather than a constant.

Here are several examples of how it works in practice.

Example 1 (discrete hypotheses): One species or two?

Data: Gill raker counts for 50 stickleback from a previously unsurveyed lake
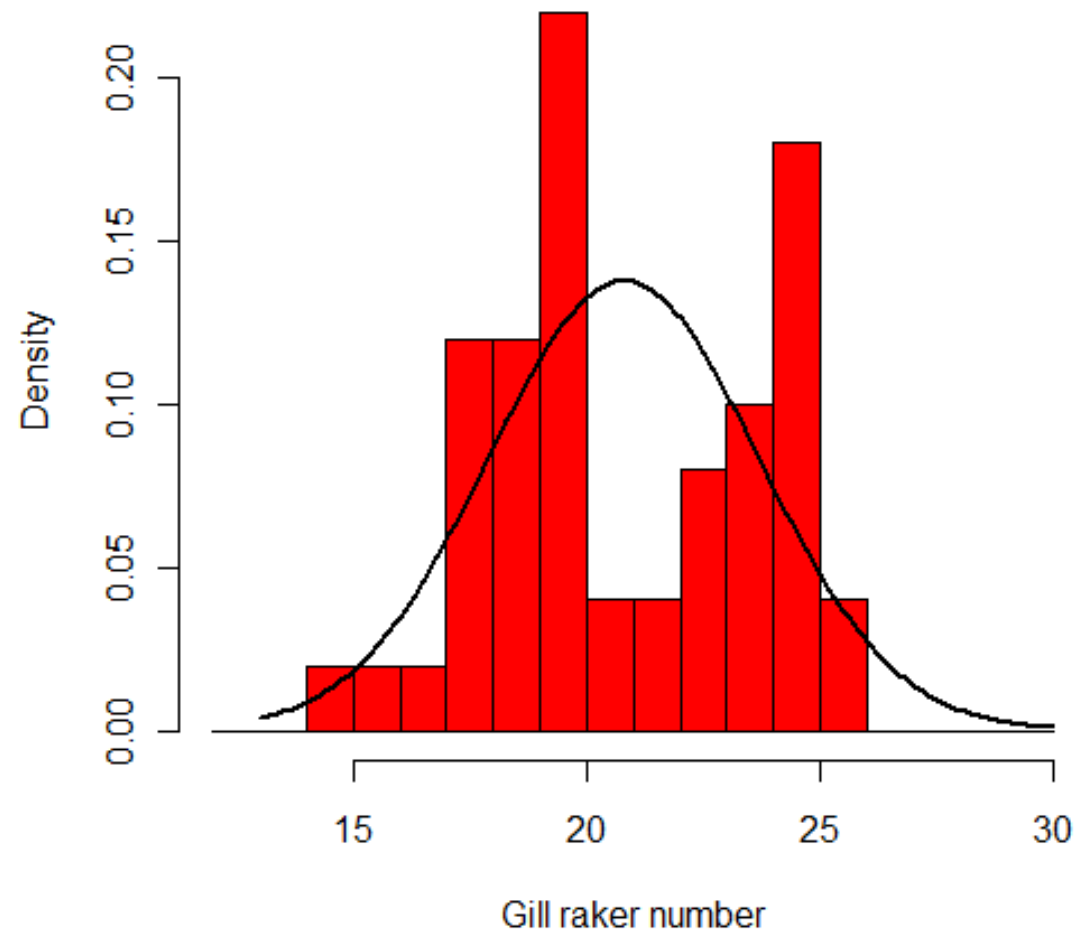
What is the probability that 2 species are present rather than 1?
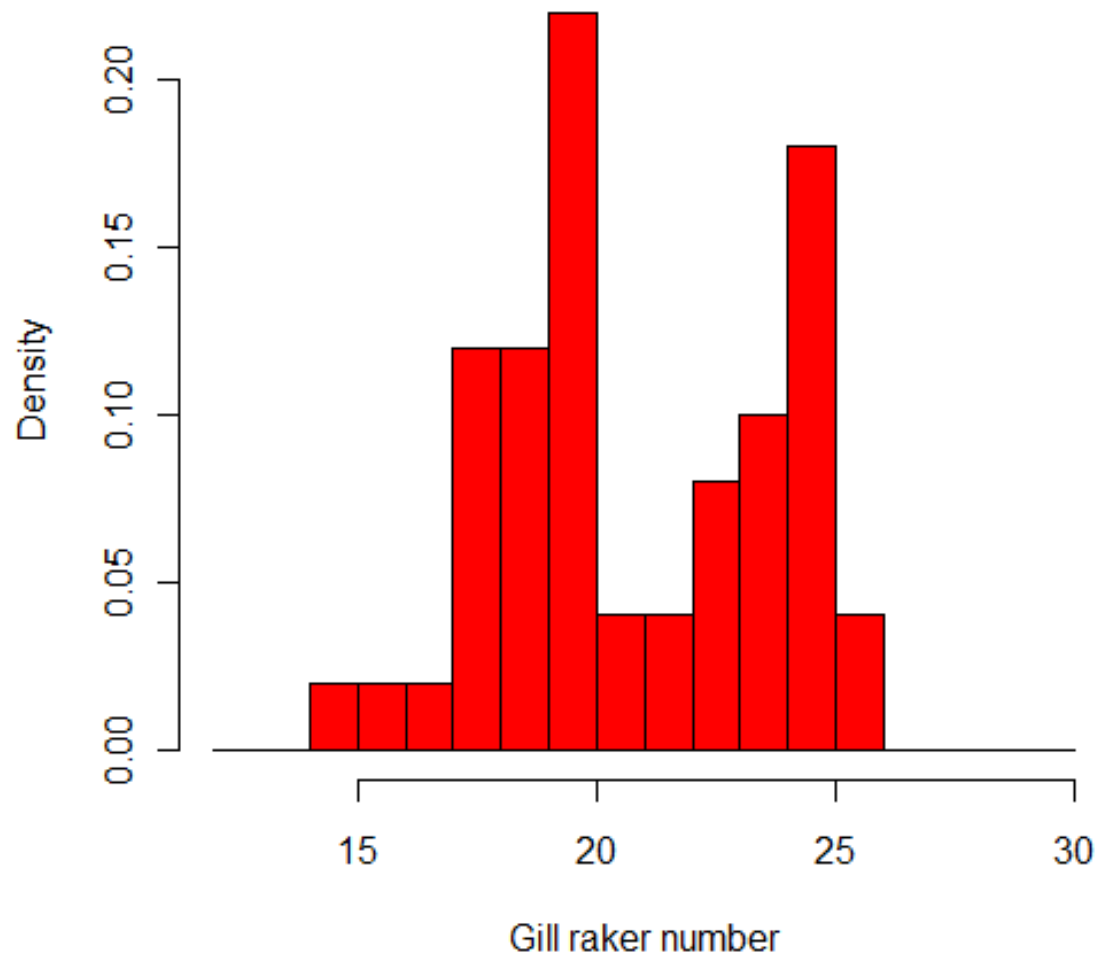
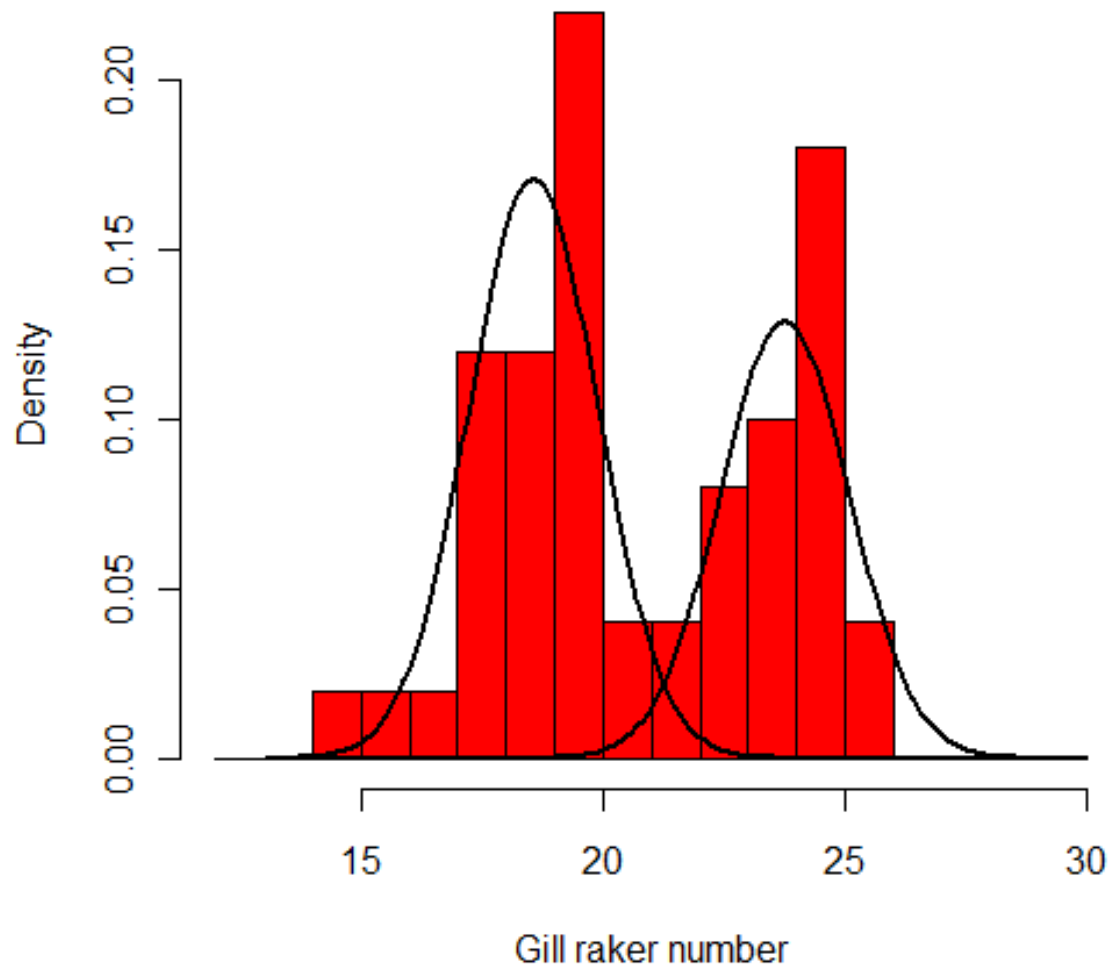## H₁: one species

Assume a normal distribution of measurements

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

Gill raker number

H₂: two species

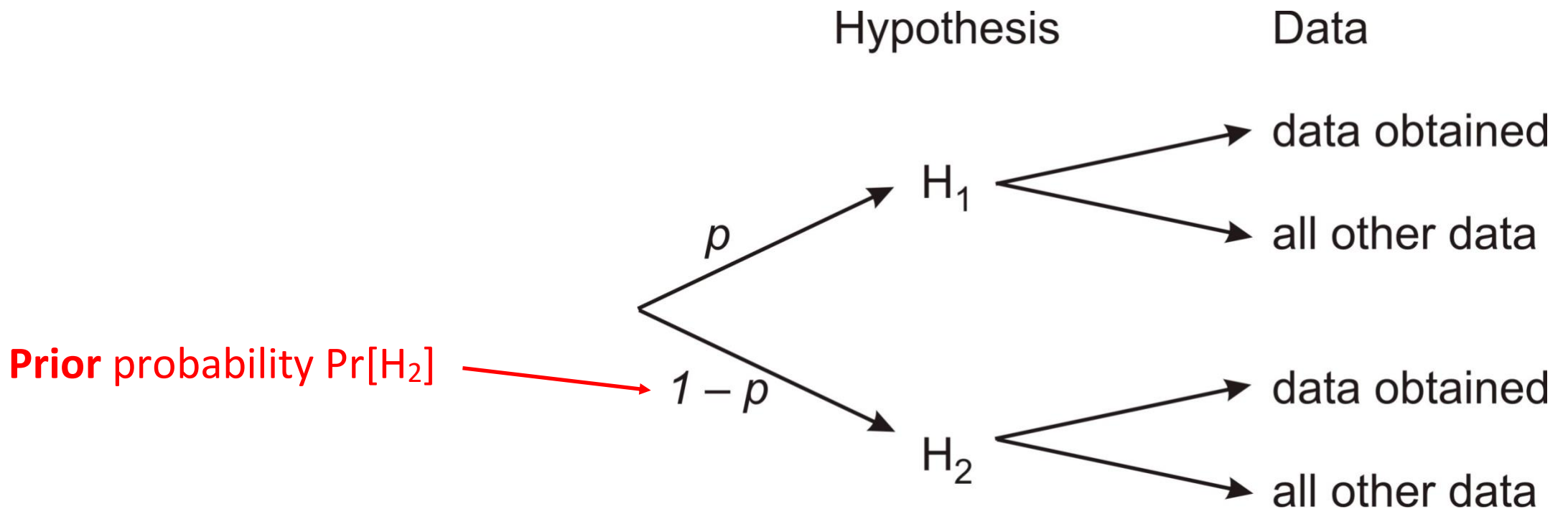Assume normal distributions with equal variance in both groups

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

Gill raker number

Posterior model probabilities

Plug the likelihoods into Bayes Theorem to calculate the posterior probability that two species are present (H_2) given the data

Posterior probability depends on the prior probability

Posterior model probabilities

Plug the likelihoods into Bayes Theorem to calculate the <u>posterior probabilities</u> of each hypothesis given the data.

Posterior probability depends on the prior probability

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

Prior probability $Pr[H_2]$	Posterior probability $Pr[H_2 \mid data]$
0.500	0.99
0.005	0.91
0.001	0.66

The strength of evidence for two species depends on the prior.

In general, *the prior becomes less influential the more data there are*

Example 2 (continuous variable): Bayesian estimation of a proportion

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

Goal: Estimate the proportion of males among the reproductive adults emerging from colonies.
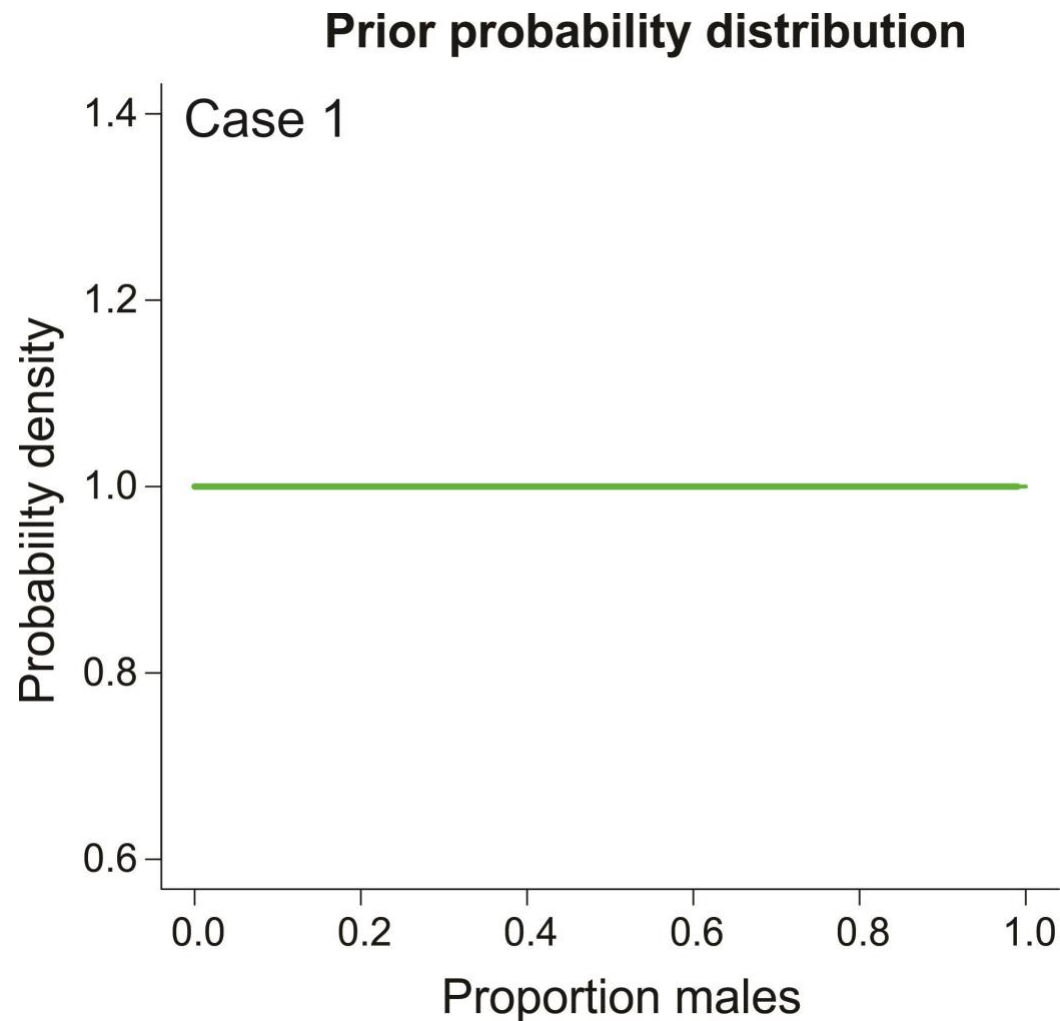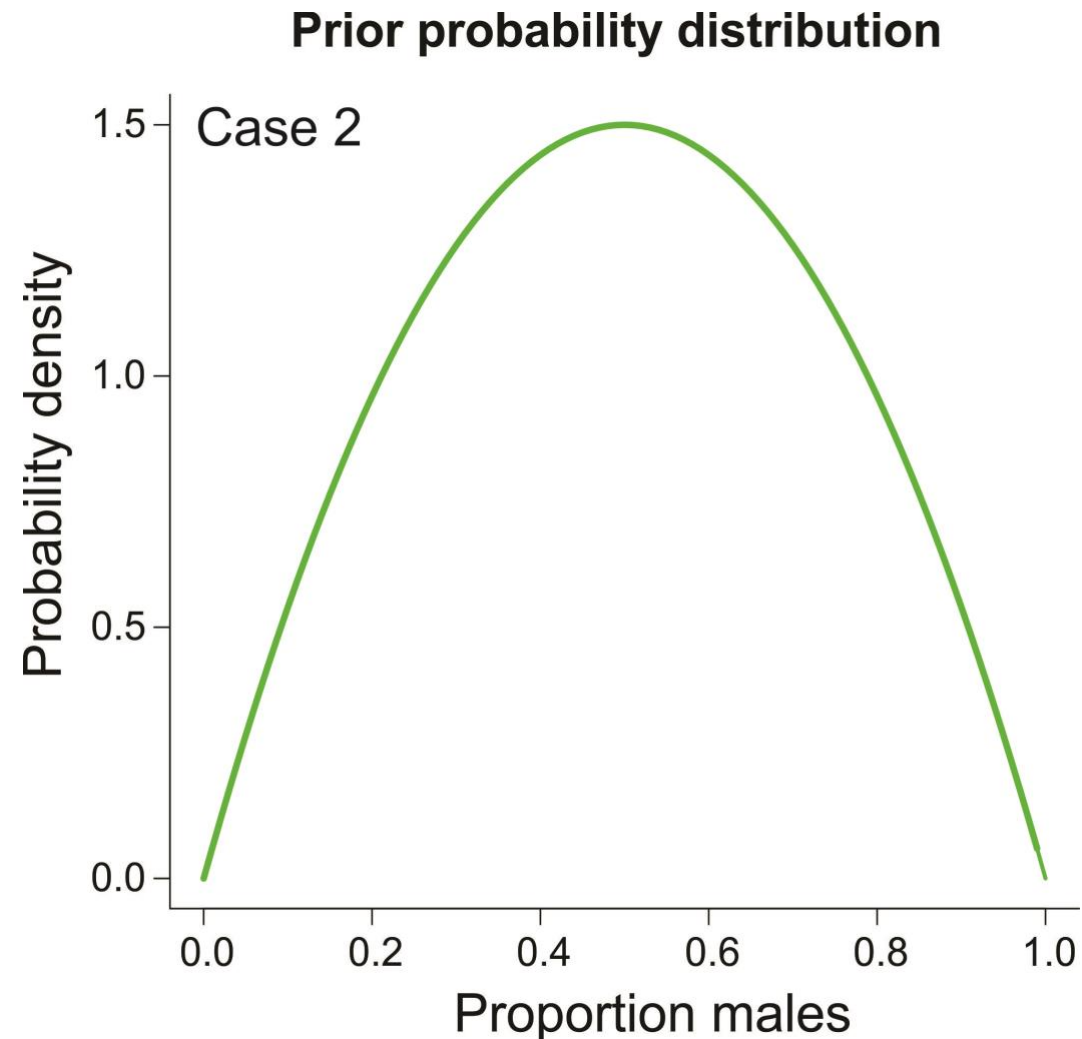
Bayesian estimation of a proportion

To begin, we need 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 incorporates previous knowledge.
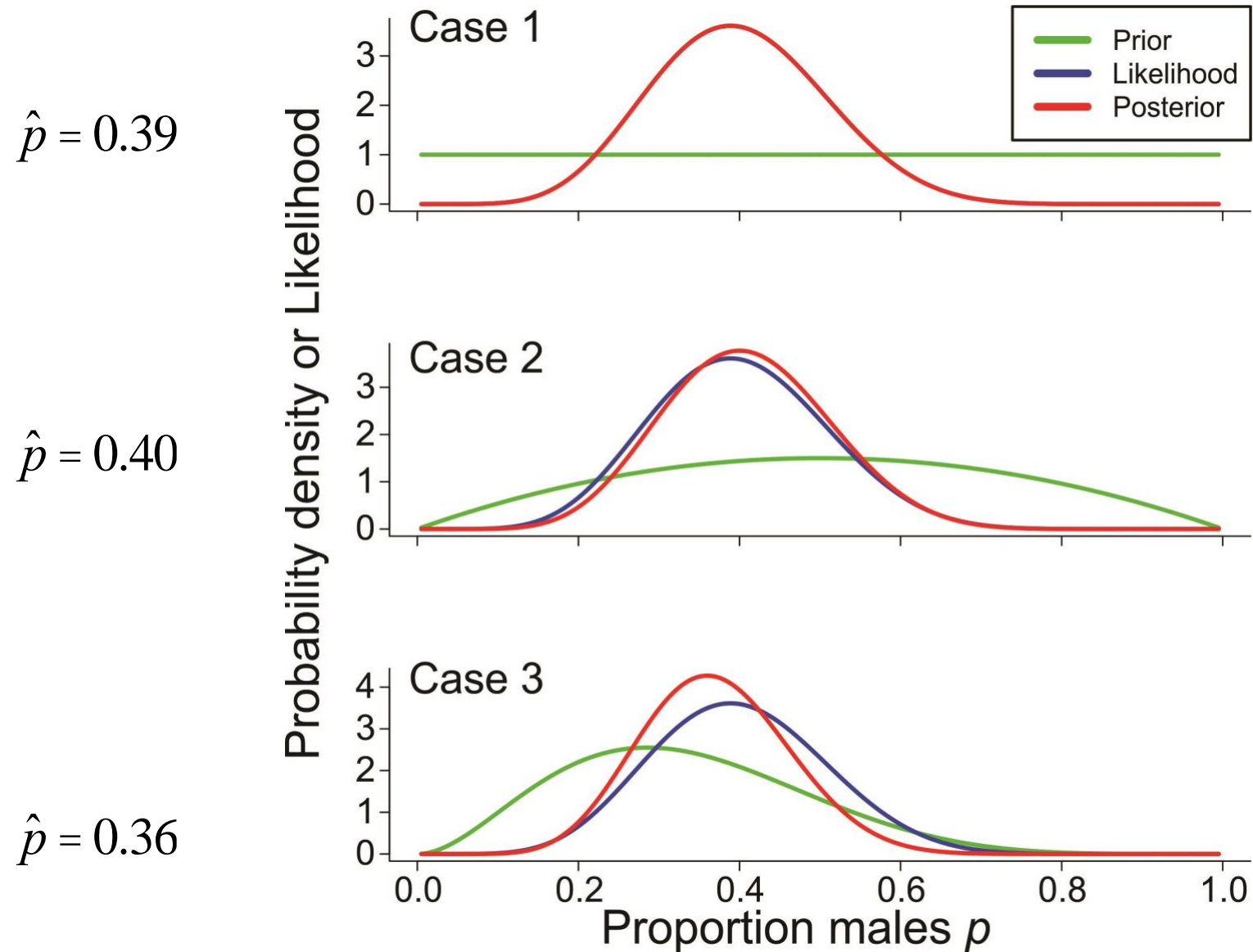

Prior probability distribution

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 incorporates this previous knowledge.

Prior probability distribution
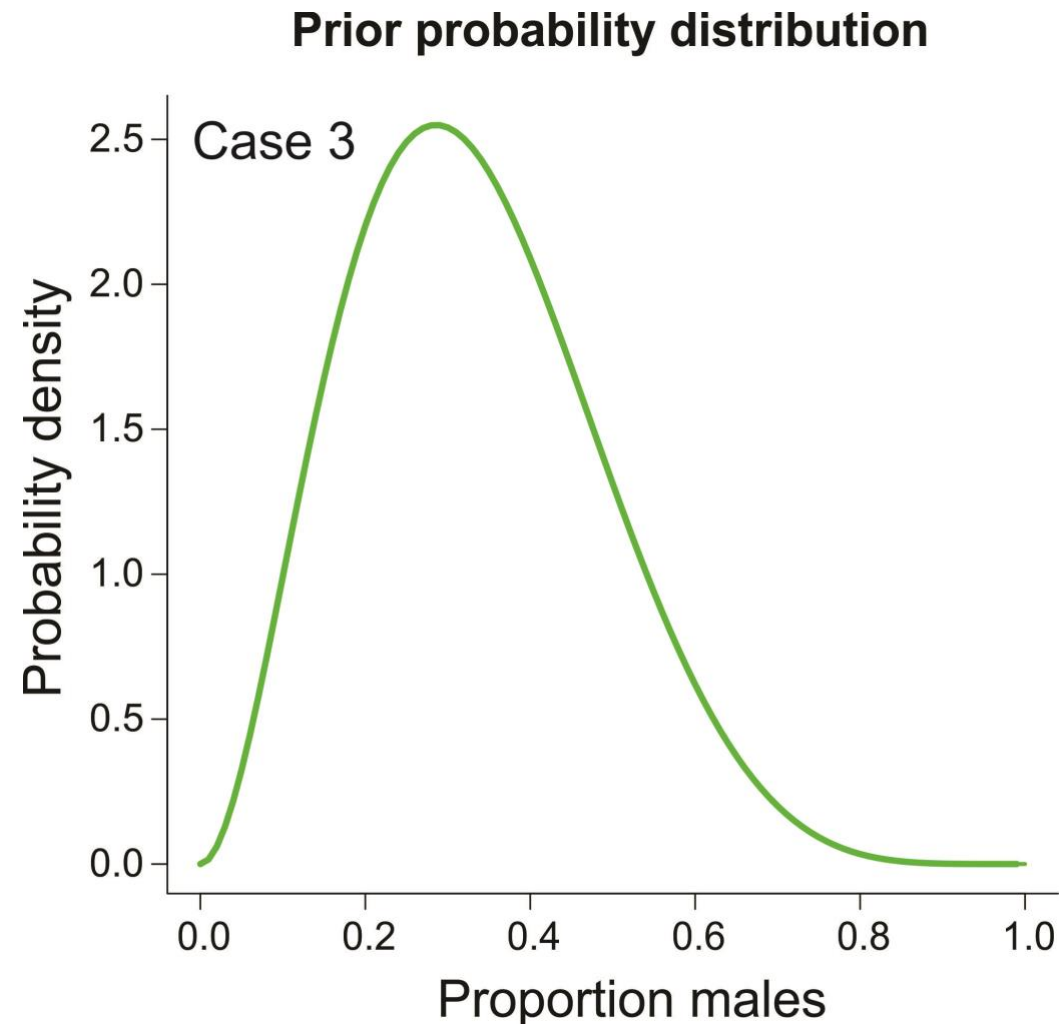
Bayesian estimation of a proportion

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

To resolve, we might all agree to use "noninformative" priors. But this stance prevents us from incorporating prior knowledge, 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 more explicitly on a survey of preexisting evidence.

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

Bayesian estimation of a proportion

95% credible interval for Case 1 (non-informative prior)
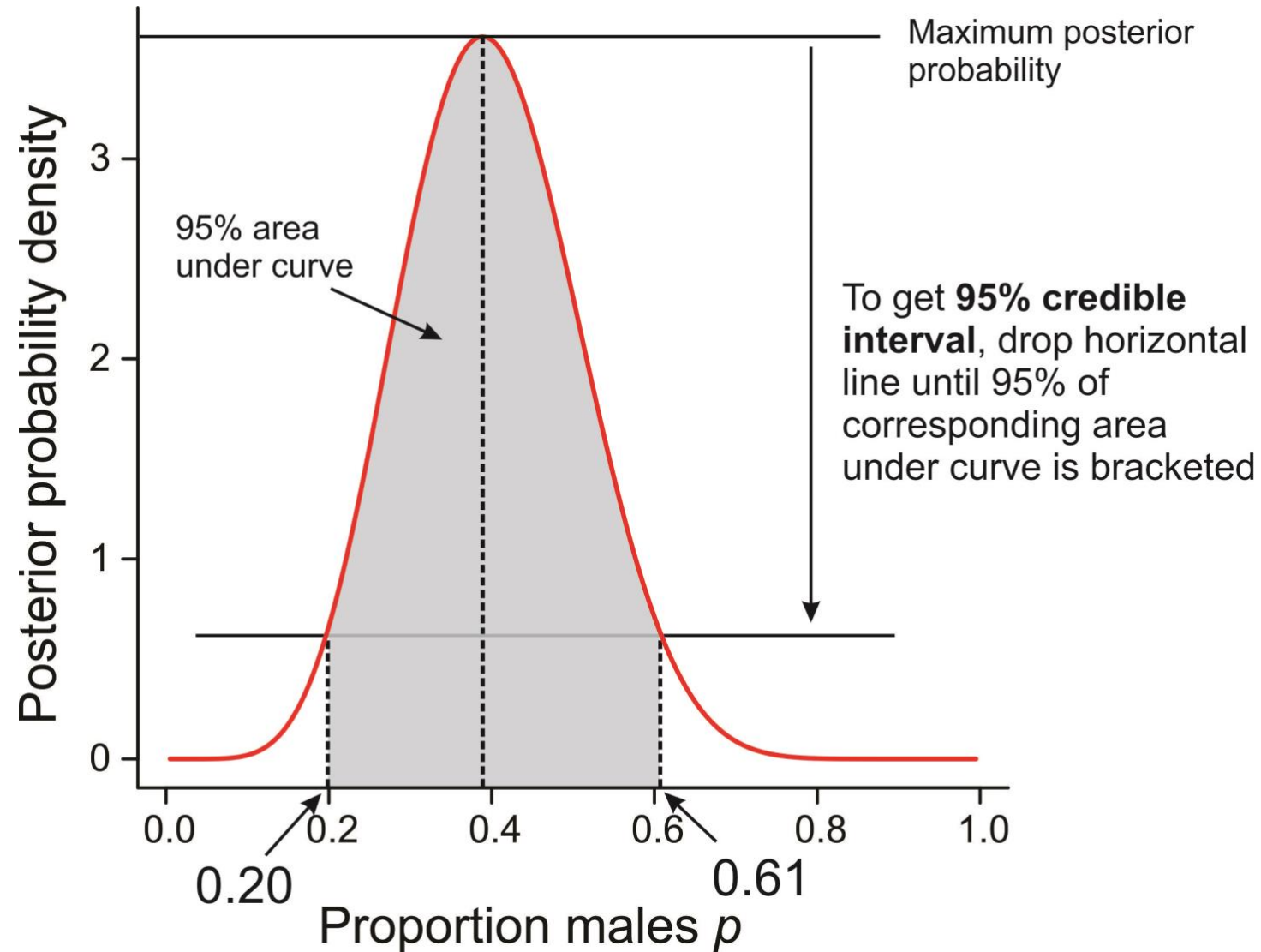
Bayesian estimation of a proportion

Interpretation of interval estimate

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

Frequentist 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.

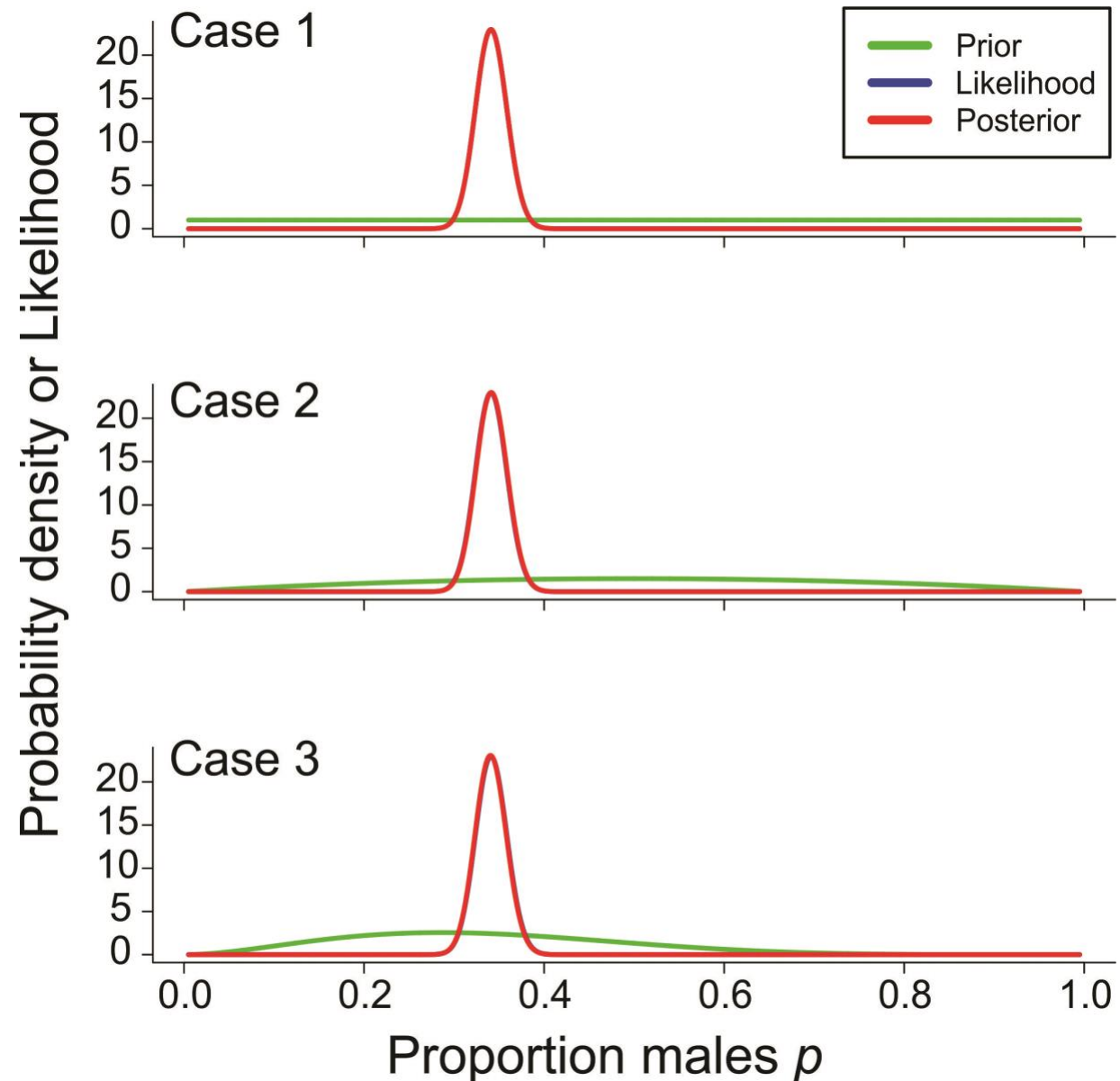95% <u>credible</u> interval: $0.20 < p < 0.61$

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 can be applied to hypothesis testing.
(ps: most Bayesians don't like hypotheses testing).

For example, when comparing means of two groups, we might consider a null and alternative hypothesis:

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_A: \mu_1 - \mu_2 \neq 0$$

Under standard <u>null hypothesis significance testing</u>, the data are compared only with the null hypothesis. The null hypothesis is the only hypothesis being tested with the data. The amount of support for the null and alternative hypotheses is never compared.

In contrast, the <u>Bayes factor</u> quantifies the amount of support for each hypothesis relative to the other hypothesis.

Bayesian hypothesis testing using the Bayes factor

Before seeing the data, the two hypotheses H_0 and H_A have prior probabilities $Pr[H_0]$ and $Pr[H_A]$. Note that $Pr[H_0] = 1 - Pr[H_A]$, since there are only two hypotheses. Hence, the ratio $Pr[H_A]/Pr[H_0]$ is called the prior odds of H_A. With a non-informative prior, $Pr[H_A]/Pr[H_0] = 1$.

Calculating the posterior probabilities involves multiplying the prior odds by a quantity called the **Bayes factor**. The transformation itself represents the evidence provided by the data.

$$\frac{Pr[H_A|data]}{Pr[H_0|data]} = \frac{Pr[data|H_A]}{Pr[data|H_0]} \times \frac{Pr[H_A]}{Pr[H_0]}.$$

$Pr[data|H_0]$ is the likelihood of H_0 and straightforward to calculate. Calculating $Pr[data|H_A]$ is a little more complicated because we have to integrate over the parameter space for $\mu_1 - \mu_2$. Let computer packages do this.

Bayesian hypothesis testing using the Bayes factor

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

A Bayes factor greater than 1 indicates that H_A has more support from the data than H_0. What value for the Bayes factor constitutes strong evidence for H_A?

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

Comparison of *P*-values from 855 *t*-tests in the published psychology literature with calculated Bayes factors (Wetzels et al. 2011).

Quantities obtained from the two approaches (*P*-value from *t*-test vs Bayes factor) are strongly correlated.
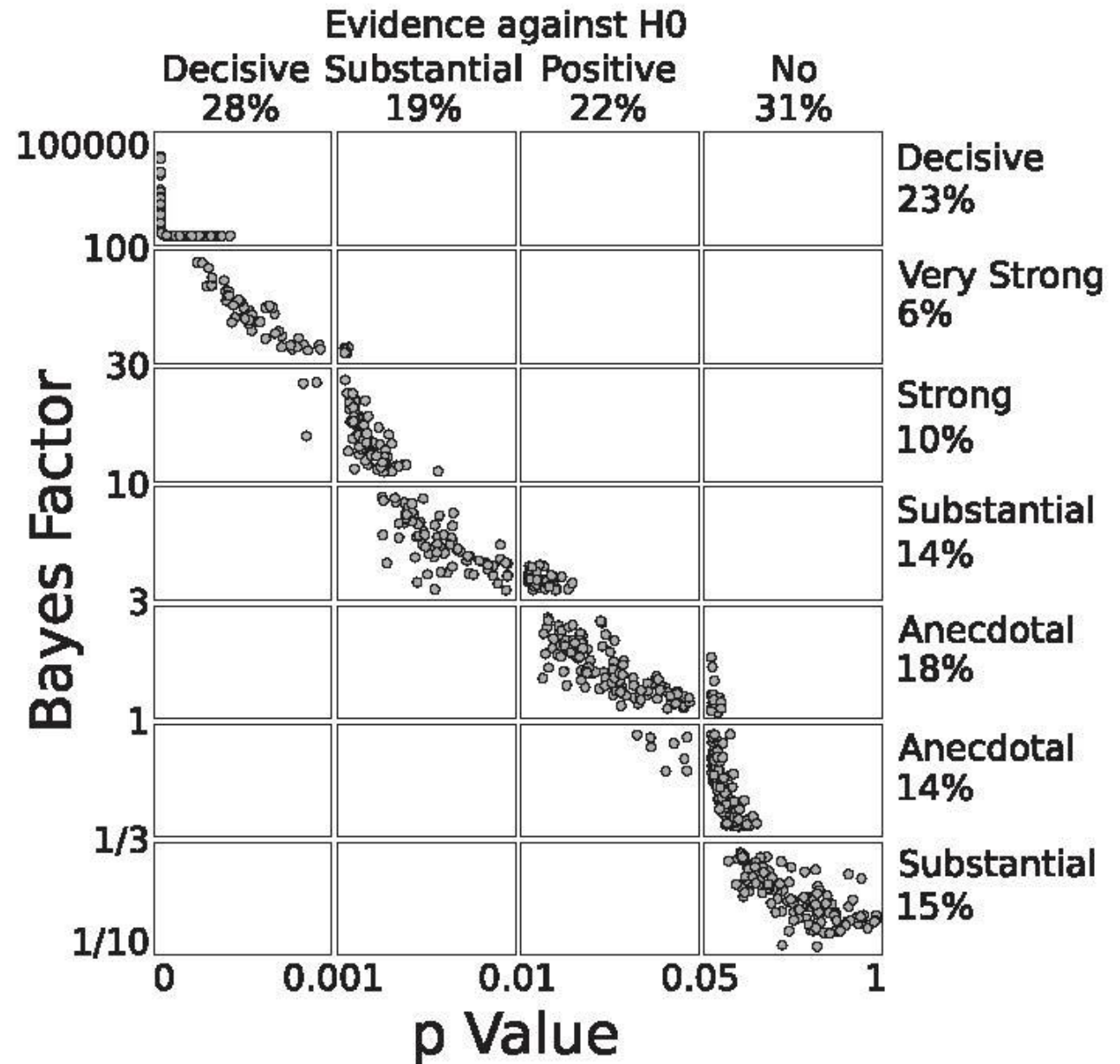
Fig. 3. The relationship between Bayes factor and *p* value. Points denote comparisons (855 in total). The scale of the axes is based on the decision categories, as given in Table 1.

Bayesian hypothesis testing

But notice how weak is the criterion *P* = 0.05 by the standard of the Bayes factor.
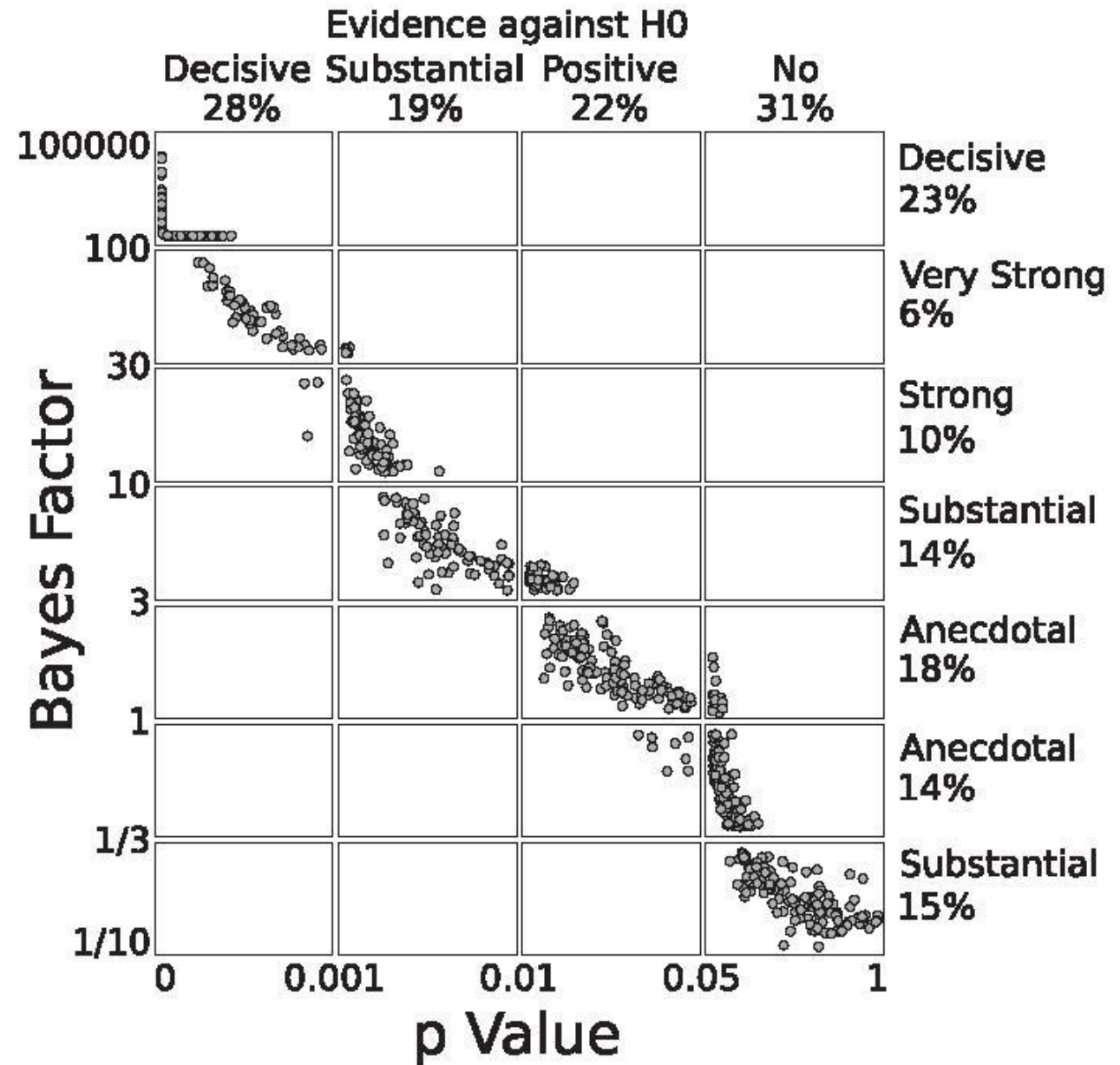
Fig. 3. The relationship between Bayes factor and *p* value. Points denote comparisons (855 in total). The scale of the axes is based on the decision categories, as given in Table 1.

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

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 as best compared to AIC.

$$\text{AIC} = -2\ln L(\text{model} \,|\, \text{data}) + 2k$$
$$\text{BIC} = -2\ln L(\text{model} \,|\, \text{data}) + k\log(n)$$

k is the number of parameters estimated in the model (including intercept and s^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

- R has many packages for Bayesian data analysis

Discussion paper for next week:

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

Download from "**handouts**" tab on course web site.

Presenters: Heather & Nick

Moderators: