BISC-869, Bayesian data analysis

March 29, 2022

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, *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?

In these examples there is no random trial, so no sampling error. *Information* is the source of uncertainty, not sampling error.

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.

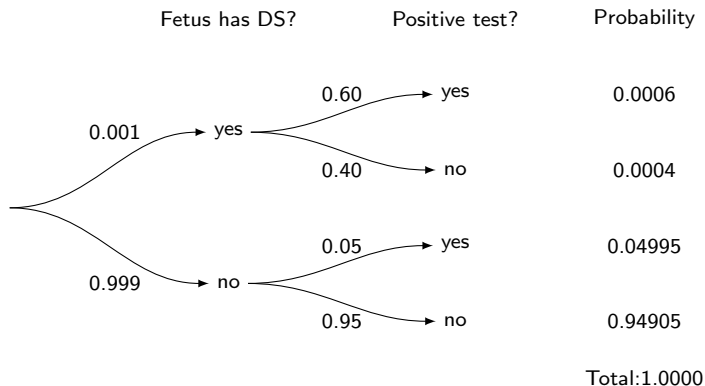**Bayesian methods are increasingly used in ecology and evolution**

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.

Brian Dennis, 1996, *Ecology*

<div align="center">Should we be using it?</div>
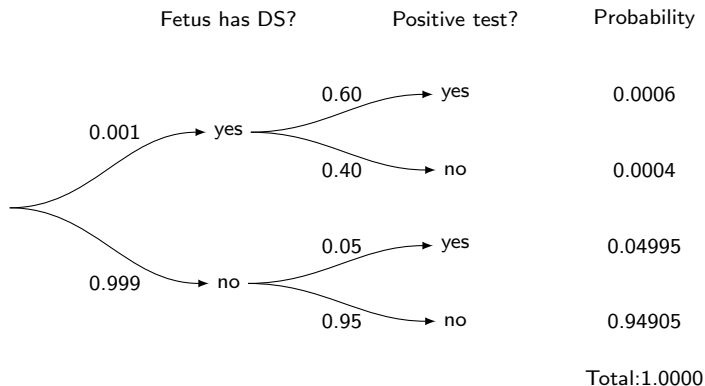
Bayes' Theorem itself is harmless

Down syndrome 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.

Fetus has DS?	Positive test?	Probability
	0.60 → yes	0.0006
0.001 → yes		
	0.40 → no	0.0004
	0.05 → yes	0.04995
0.999 → no		
	0.95 → no	0.94905

Total: 1.0000

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.
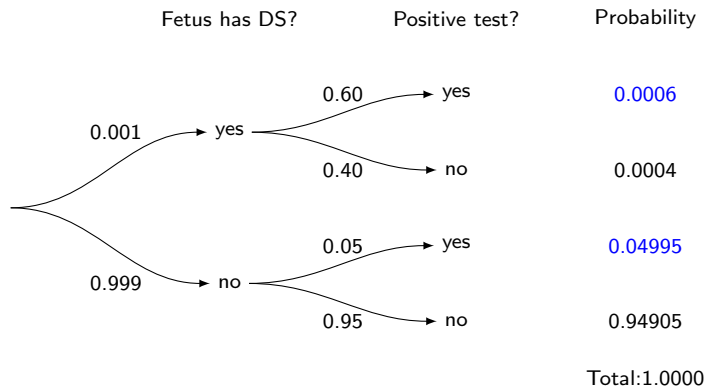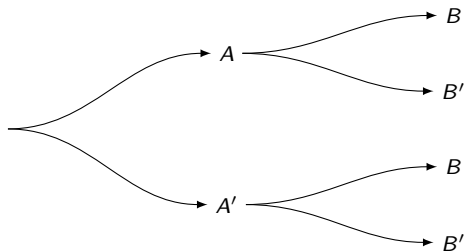
Fetus has DS?		Positive test?	Probability
	0.60 → yes		0.0006
0.001 → yes			
	0.40 → no		0.0004
	0.05 → yes		0.04995
0.999 → no			
	0.95 → no		0.94905

Total:1.0000

Conditional probability calculation

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

$$\Pr[\text{DS}|\text{positive}] = \frac{0.0006}{0.0006 + 0.04995} = 0.012$$

Fetus has DS?	Positive test?	Probability

Total: 1.0000
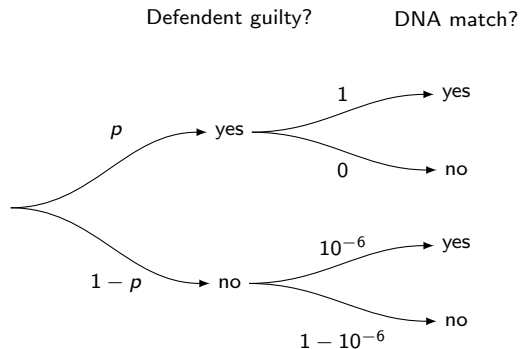
starting with

$$\Pr[A|B] \cdot \Pr[B] = \Pr[B|A] \cdot \Pr[A]$$

we get

$$\Pr[A|B] = \frac{\Pr[B|A] \cdot \Pr[A]}{\Pr[B]} = \frac{\Pr[B|A] \cdot \Pr[A]}{\Pr[B|A] \cdot \Pr[A] + \Pr[B|A'] \cdot \Pr[A']}$$

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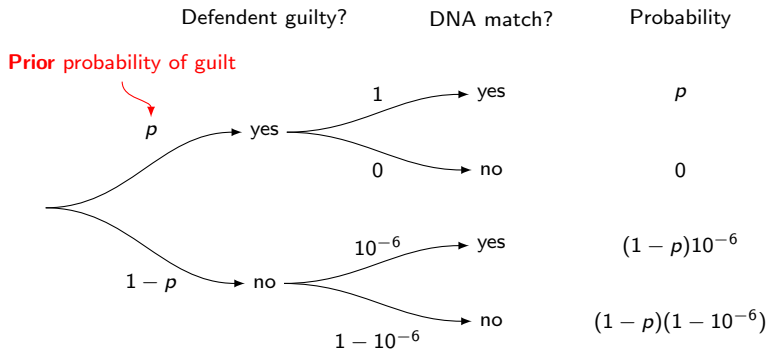
What is the probability of guilt given a positive DNA match (assuming no contamination of samples)?

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

$$\Pr[\text{guilt}|\text{DNA match}] = \frac{1 \cdot p}{1 \cdot p + 10^{-6} \cdot (1 - p)}$$
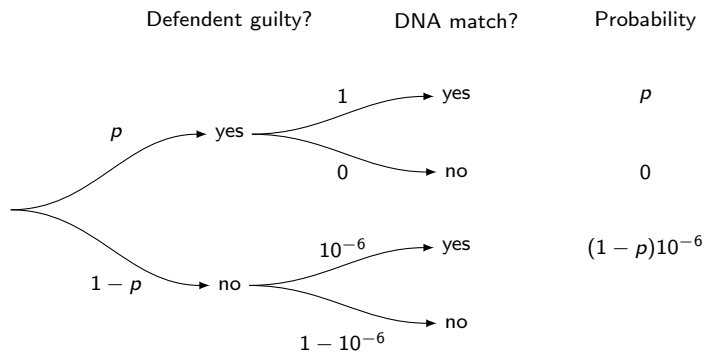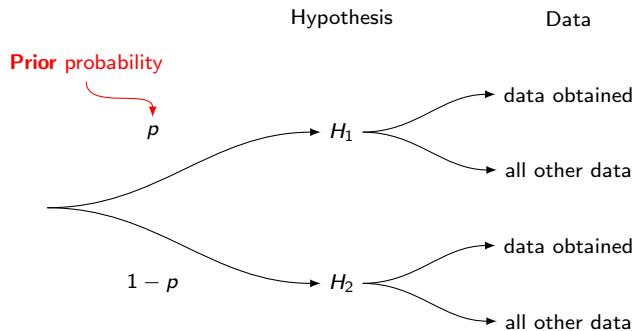
Posterior probability of guilt

	Defendant guilty?	DNA match?	Probability

Prior probability of guilt

p → yes

1 → yes p

0 → no 0

$1 - p$ → no

10^{-6} → yes $(1 - p)10^{-6}$

$1 - 10^{-6}$ → no $(1 - p)(1 - 10^{-6})$

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

if $p = 10^{-6}$, then $\Pr[\text{guilt}|\text{DNA match}] = 0.5$
if $p = 0.5$, then $\Pr[\text{guilt}|\text{DNA match}] = 0.999999$

So, is the defendant guilty or innocent?

Defendent guilty?	DNA match?	Probability

Defendent guilty?   DNA match?   Probability

p → yes
    1 → yes p
    0 → no   0

$1 - p$ → no
    10^{-6} → yes $(1 - p)10^{-6}$
    $1 - 10^{-6}$ → no

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

$\Pr[\text{data}|H_1]$ is the **likelihood** of H_1 given the data.

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.

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?
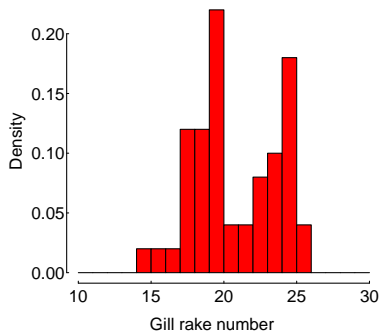
H_1: **one species**
Assume a normal distribution of measurements

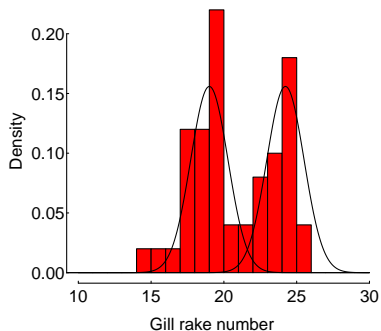$$\Pr[\text{data}|H_1] = \mathcal{L}[H_1|\text{data}] = e^{-124.06}$$

Note: $e^{-124.06}$ is the product of the probability of each outcome given the null.
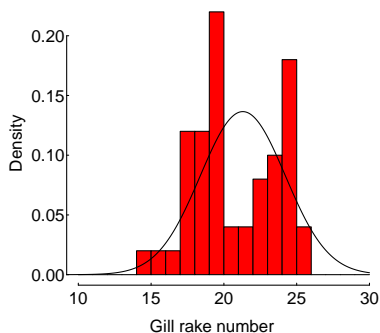
H_2: **two species**
Assume normal distributions with equal variance in both groups

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

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, $\Pr[H_2|\text{data}]$
0.500	0.99
0.005	0.91
0.001	0.66

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

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


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

Case 1: the "non-informative" prior: expression of total ignorance.
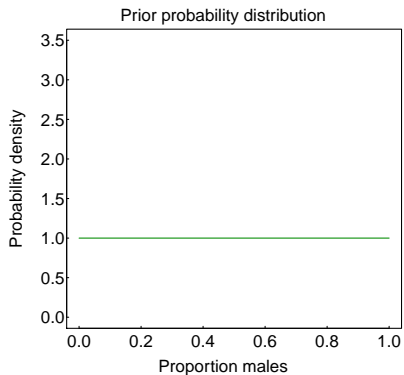
Case 2: Most species have a sex ratio close to 50:50, and this is predicted by simple sex-ratio theory. This prior probability distribution attempts to incorporate this previous information (this is really what priors are for).

Prior probability distribution
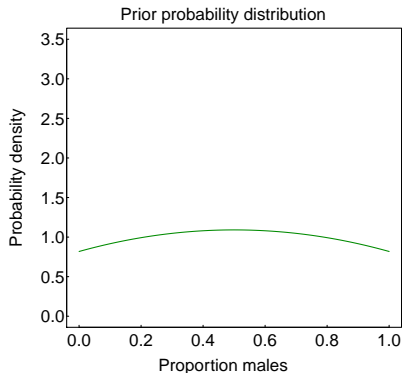
Case 3: Then again, female-biased sex ratios do exist in nature, more than male-biased sex ratios, especially in bees and other hymenoptera. The following prior attempts to incorporate this previous information.
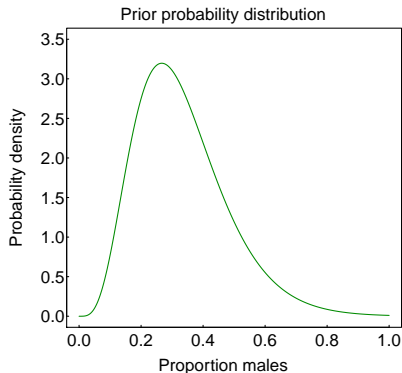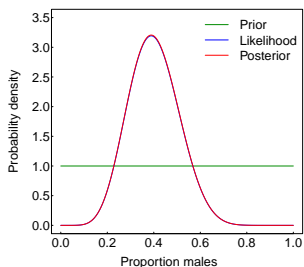
Data: from day 148 at nest S31: 7 males, 11 female. $\hat{p}_{\mathrm{MLE}} = 0.39$.
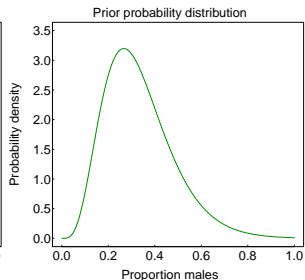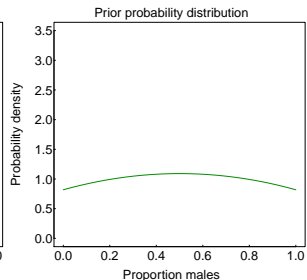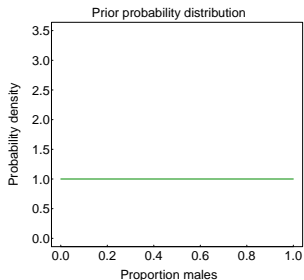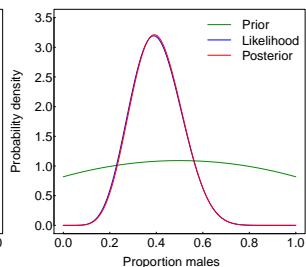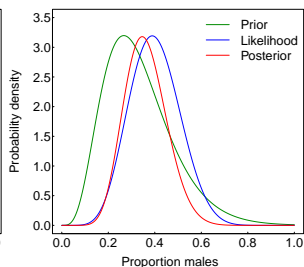
Data: from day 148 at nest S31: 7 males, 11 female. $\hat{p}_{\mathrm{MLE}} = 0.39$.

$\hat{p}_{\mathrm{MLE}} = 0.39$ $\qquad\qquad$ $\hat{p}_{\mathrm{MLE}} = 0.39$ $\qquad\qquad$ $\hat{p}_{\mathrm{MLE}} = 0.35$

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 "non-informative" 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 pre-existing evidence (lot of work).

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

95% credible interval
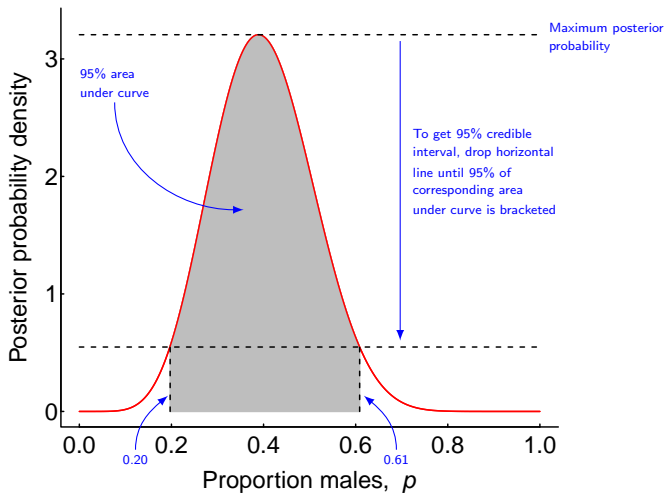
Interpretation of the interval estimates

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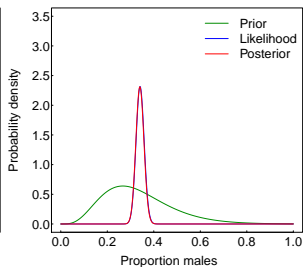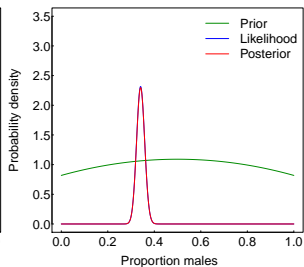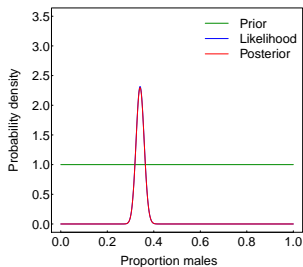
95% credible interval: $0.20 < p < 0.61$
(assuming Case 1, with non-informative prior).

Bayesian interpretation:
The probability is 0.95 that the population proportion lies between 0.20 and 0.61.

All the data: 253 males, 489 females $p_{\hat{\text{MLE}}} = 0.34$.

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

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.

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

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

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

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

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

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

$\Pr[\text{data}|H_A]$ 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).

If two hypotheses H_0 and H_A have the same prior probability, i.e., $\Pr[\text{data}|H_0] = \Pr[\text{data}|H_A]$, then the Bayes factor is just
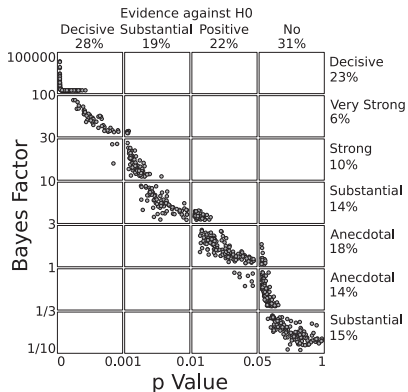
$$\text{Bayes factor} = \frac{\Pr[\text{data}|H_A]}{\Pr[\text{data}|H_0]}$$

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.

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.

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.

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

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

Requires a criterion to compare models, and 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.

$$\text{AIC} = -2\ln\mathcal{L}[\text{model}|\text{data}] + 2k$$

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

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

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.