A beginner's guide to Bayesian Statistics

$$p(c|r) = \frac{p(r|c)\,p(c)}{p(r)}$$

or

$$p(c|r) = \frac{p(r|c)\,p(c)}{\sum_{c*} p(r|c^*)\,p(c^*)}$$

Bayes

Thomas Bayes (1702-1761) was a mathematician and Presbyterian minister in England. His famous theorem was published posthumously in 1763, The simple rule has vast ramifications for statistical inference.

Bayes' successor, Pierre-Simon Laplace should really label this type of analysis, because it was Laplace who independently rediscovered and extensively developed the methods.

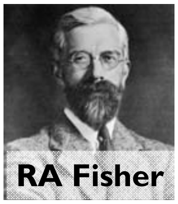
Bayes

There is another branch of statistics, called *frequentist*, which does not use Bayes' rule for inference and decisions. This approach is often identified Ronald Fisher ("F"-test).

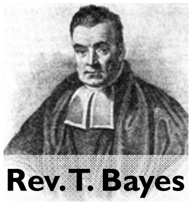
It is curious and re-assuring that the overwhelmingly dominant Fisherian approach of the 20th century is giving wayin the 21st century to a Bayesian approach that had its genesis in the 18th century.

Bayesian Statistics – NY Times

Some statisticians and scientists are optimistic that Bayesian methods can improve the reliability of research by allowing scientists to crosscheck work done with the more traditional or "classical" approach, known as frequentist statistics. The two methods approach the same problems from different angles.



RA Fisher



Rev. T. Bayes

Bayesian Statistics – NY Times

The essence of the frequentist technique is to apply probability to data. By contrast, Bayesian calculations go straight for the probability of the hypothesis, factoring in any other relevant information.

Scientists who have learned Bayesian statistics often marvel that it propels them through a different kind of scientific reasoning than they had experienced using classical methods.

Bayesian Statistics – NY Times

One downside of Bayesian statistics is that it requires prior information — and often scientists need to start with a guess or estimate.

Assigning numbers to subjective judgments is "like fingernails on a chalkboard," said physicist Kyle Cranmer, who helped develop a frequentist technique to identify the latest new subatomic particle — the Higgs boson.

Bayesian Statistics – NY Times

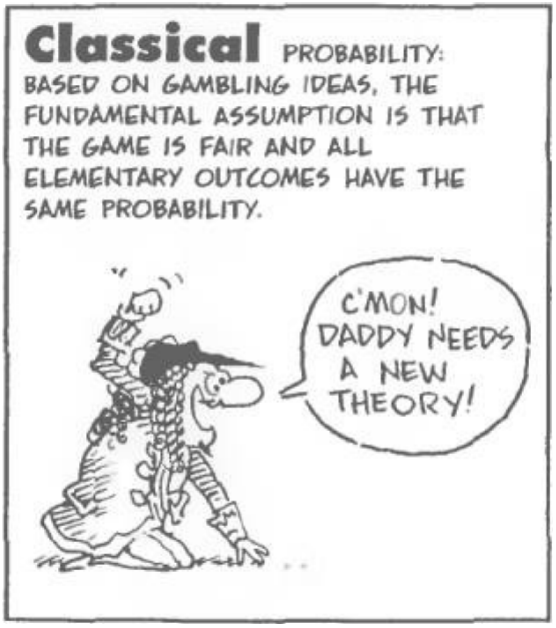
Critics of Bayesian Statististics say that the best cure for misleading findings is not Bayesian statistics, but good frequentist ones.

A psychologist found common statistical shenanigans in his field — logical leaps, unjustified conclusions, and various forms of unconscious cheating. He looked into Bayesian statistics and concluded that if people misused or misunderstood one system, they would do just as badly with the other.

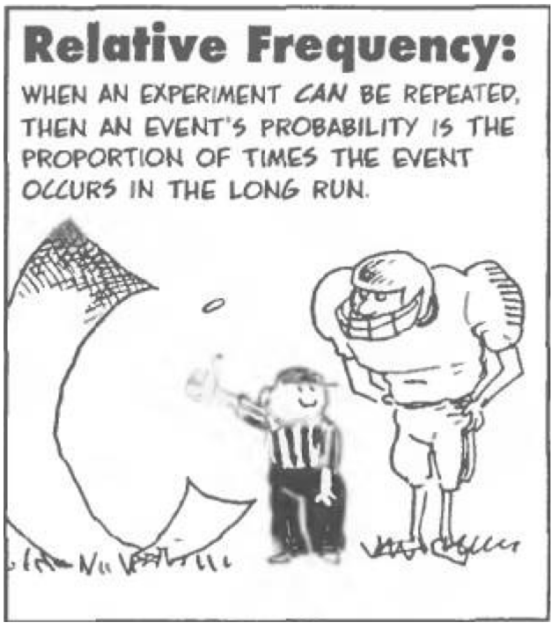
A beginner's guide to Bayesian Statistics

or

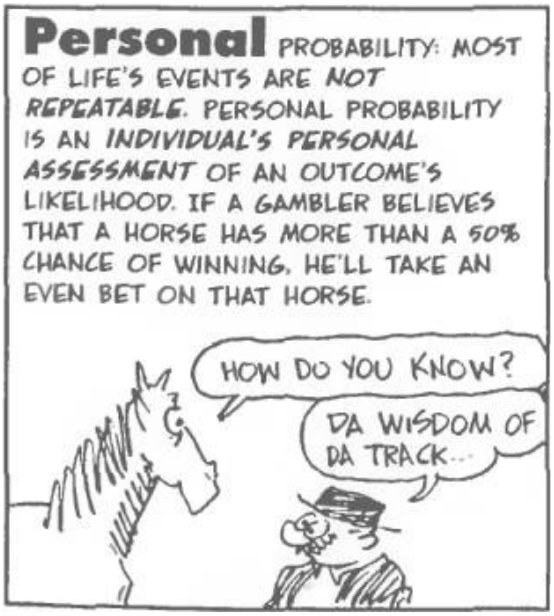
$$p(c|r) = \frac{p(r|c)\,p(c)}{\sum_{c^*} p(r|c^*)\,p(c^*)}$$



Classical PROBABILITY: BASED ON GAMBLING IDEAS, THE FUNDAMENTAL ASSUMPTION IS THAT THE GAME IS FAIR AND ALL ELEMENTARY OUTCOMES HAVE THE SAME PROBABILITY.



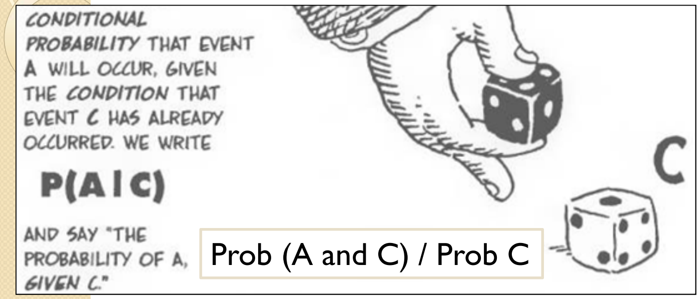
Relative Frequency: WHEN AN EXPERIMENT *CAN* BE REPEATED, THEN AN EVENT'S PROBABILITY IS THE PROPORTION OF TIMES THE EVENT OCCURS IN THE LONG RUN.



Personal PROBABILITY: MOST OF LIFE'S EVENTS ARE ***NOT REPEATABLE***. PERSONAL PROBABILITY IS AN ***INDIVIDUAL'S PERSONAL ASSESSMENT*** OF AN OUTCOME'S LIKELIHOOD. IF A GAMBLER BELIEVES THAT A HORSE HAS MORE THAN A 50% CHANCE OF WINNING, HE'LL TAKE AN EVEN BET ON THAT HORSE.

AN ***OBJECTIVIST*** USES EITHER THE CLASSICAL OR FREQUENCY DEFINITION OF PROBABILITY. A ***SUBJECTIVIST*** OR ***BAYESIAN*** APPLIES FORMAL LAWS OF CHANCE TO HIS OWN, OR YOUR, PERSONAL PROBABILITIES.

Conditional Probability



What is the probability of getting sum=7, Given that the white die is one?

$$\frac{1/36}{1/6} = 1/6$$

$$p(c|r) = \frac{p(r|c)\,p(c)}{p(r)}$$

$$= \frac{\dfrac{p(r \text{ and } c)}{p(c)}\,p(c)}{p(r)} = \frac{p(r \text{ and } c)}{p(r)}$$

Bayesian Probability Reasoning

A model of data specifies the probability of particular data values given the model's structure and parameter values.

In other words, a model specifies

***p*(data values | parameters values)**
along with the **prior, *p*(parameters values)**

We use Bayes' rule to convert that to what we really want to know, which is how strongly we should believe in the parameter values, given the data:

***p*(parameters values | data values)**

A beginner's guide to Bayesian Statistics

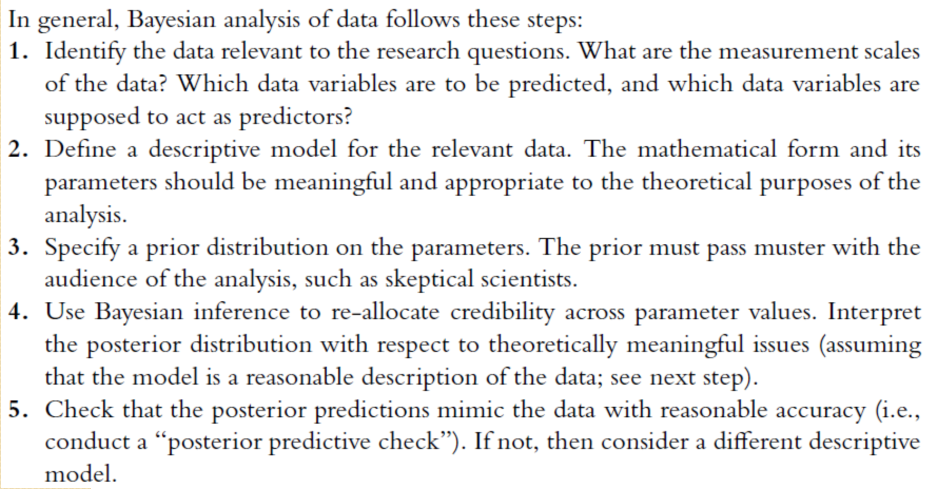
Bayesian Statistics

$$p(c|r) = \frac{p(r|c)\,p(c)}{p(r)}$$

or

$$p(c|r) = \frac{p(r|c)\,p(c)}{\sum_{c^*} p(r|c^*)\,p(c^*)}$$

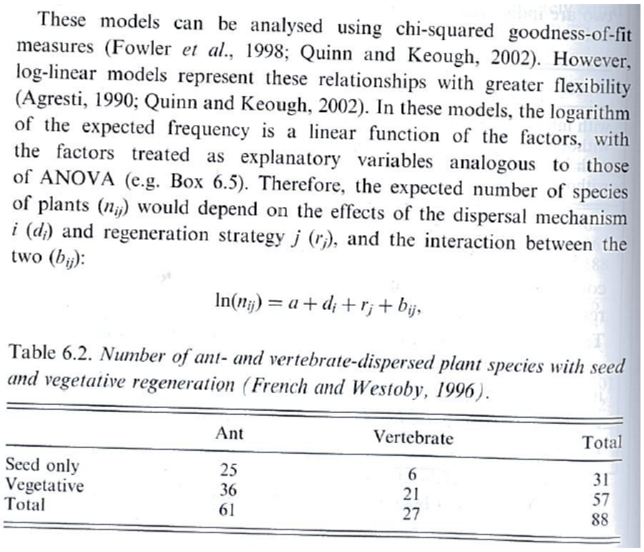
Steps to Bayesian Analysis



In general, Bayesian analysis of data follows these steps:

1. Identify the data relevant to the research questions. What are the measurement scales of the data? Which data variables are to be predicted, and which data variables are supposed to act as predictors?
2. Define a descriptive model for the relevant data. The mathematical form and its parameters should be meaningful and appropriate to the theoretical purposes of the analysis.
3. Specify a prior distribution on the parameters. The prior must pass muster with the audience of the analysis, such as skeptical scientists.
4. Use Bayesian inference to re-allocate credibility across parameter values. Interpret the posterior distribution with respect to theoretically meaningful issues (assuming that the model is a reasonable description of the data; see next step).
5. Check that the posterior predictions mimic the data with reasonable accuracy (i.e., conduct a "posterior predictive check"). If not, then consider a different descriptive model.

Example



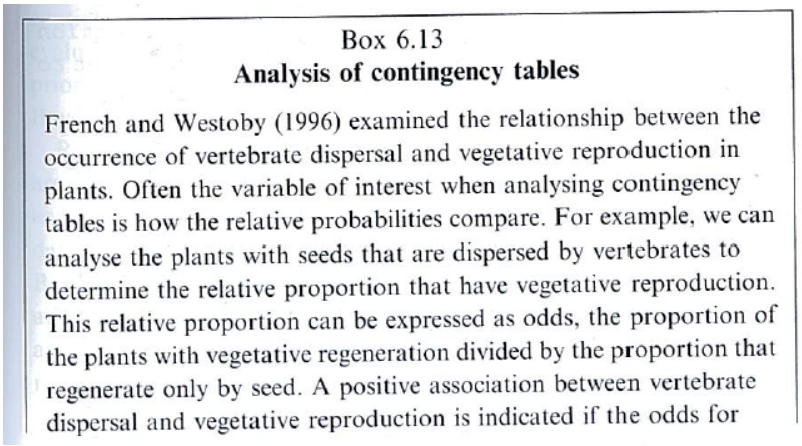
These models can be analysed using chi-squared goodness-of-fit measures (Fowler *et al.*, 1998; Quinn and Keough, 2002). However, log-linear models represent these relationships with greater flexibility (Agresti, 1990; Quinn and Keough, 2002). In these models, the logarithm of the expected frequency is a linear function of the factors, with the factors treated as explanatory variables analogous to those of ANOVA (e.g. Box 6.5). Therefore, the expected number of species of plants (n_{ij}) would depend on the effects of the dispersal mechanism i (d_i) and regeneration strategy j (r_j), and the interaction between the two (b_{ij}):

$$\ln(n_{ij}) = a + d_i + r_j + b_{ij},$$

Table 6.2. *Number of ant- and vertebrate-dispersed plant species with seed and vegetative regeneration (French and Westoby, 1996).*

	Ant	Vertebrate	Total
Seed only	25	6	31
Vegetative	36	21	57
Total	61	27	88

Example

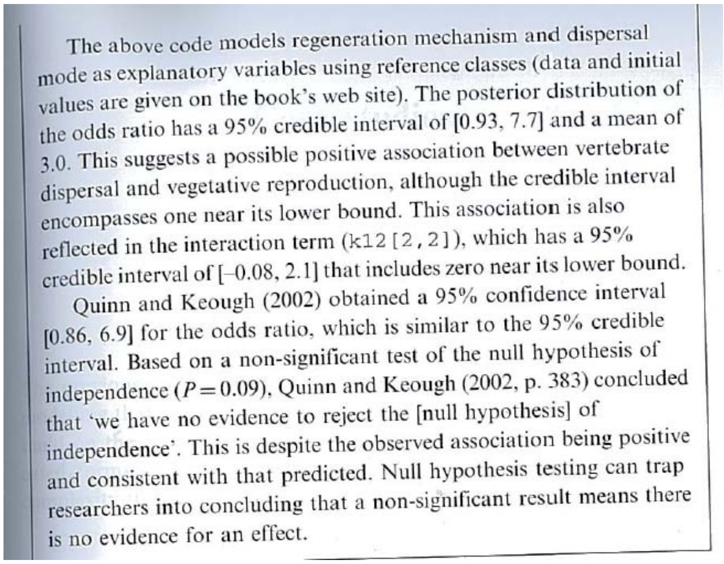


Box 6.13

Analysis of contingency tables

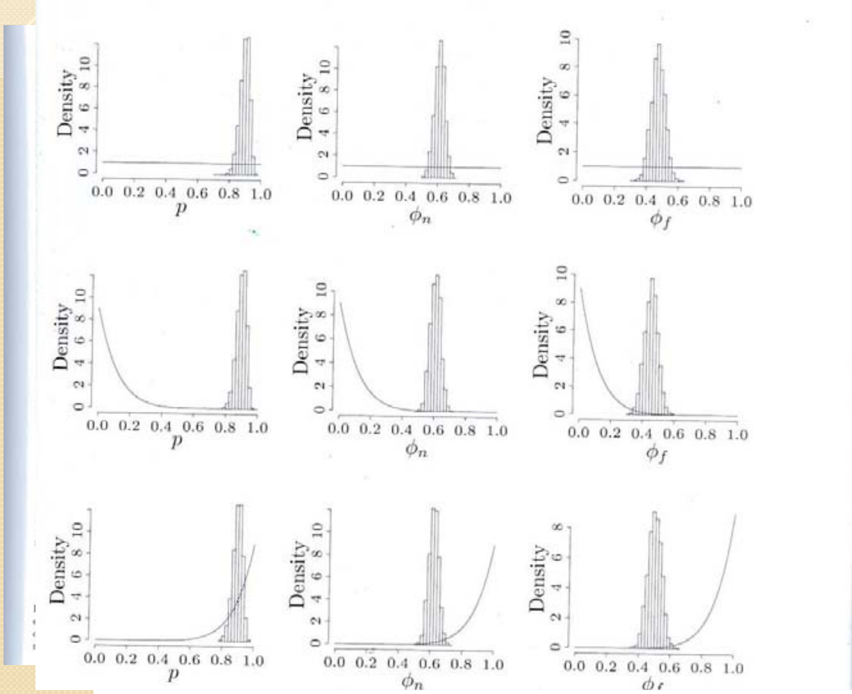
French and Westoby (1996) examined the relationship between the occurrence of vertebrate dispersal and vegetative reproduction in plants. Often the variable of interest when analysing contingency tables is how the relative probabilities compare. For example, we can analyse the plants with seeds that are dispersed by vertebrates to determine the relative proportion that have vegetative reproduction. This relative proportion can be expressed as odds, the proportion of the plants with vegetative regeneration divided by the proportion that regenerate only by seed. A positive association between vertebrate dispersal and vegetative reproduction is indicated if the odds for

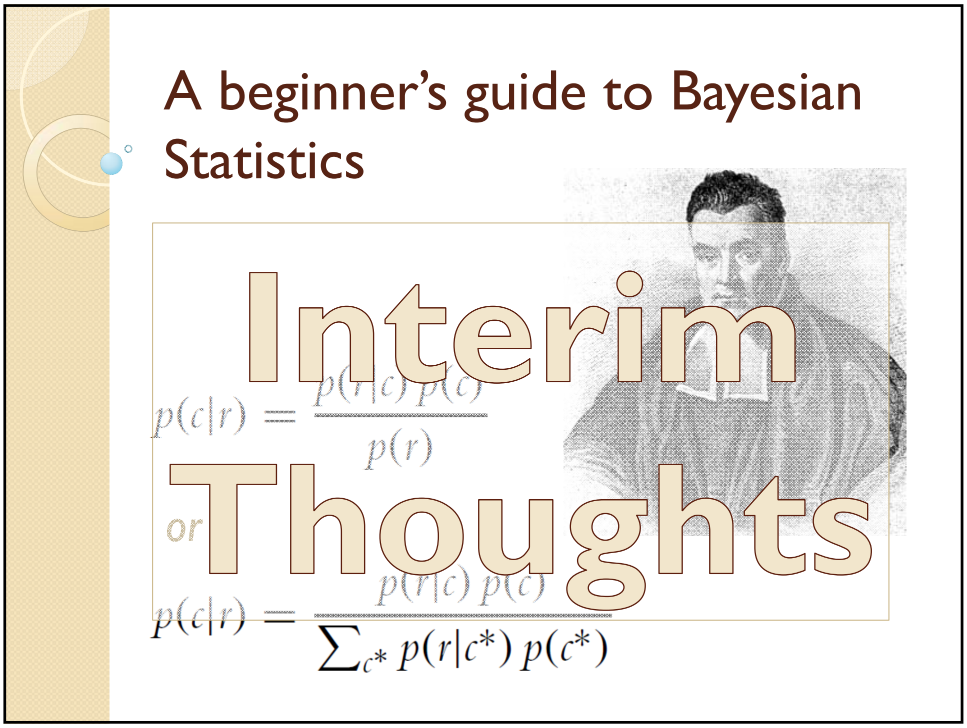
Example



The above code models regeneration mechanism and dispersal mode as explanatory variables using reference classes (data and initial values are given on the book's web site). The posterior distribution of the odds ratio has a 95% credible interval of [0.93, 7.7] and a mean of 3.0. This suggests a possible positive association between vertebrate dispersal and vegetative reproduction, although the credible interval encompasses one near its lower bound. This association is also reflected in the interaction term (`k12[2,2]`), which has a 95% credible interval of [–0.08, 2.1] that includes zero near its lower bound.

Quinn and Keough (2002) obtained a 95% confidence interval [0.86, 6.9] for the odds ratio, which is similar to the 95% credible interval. Based on a non-significant test of the null hypothesis of independence ($P = 0.09$), Quinn and Keough (2002, p. 383) concluded that 'we have no evidence to reject the [null hypothesis] of independence'. This is despite the observed association being positive and consistent with that predicted. Null hypothesis testing can trap researchers into concluding that a non-significant result means there is no evidence for an effect.





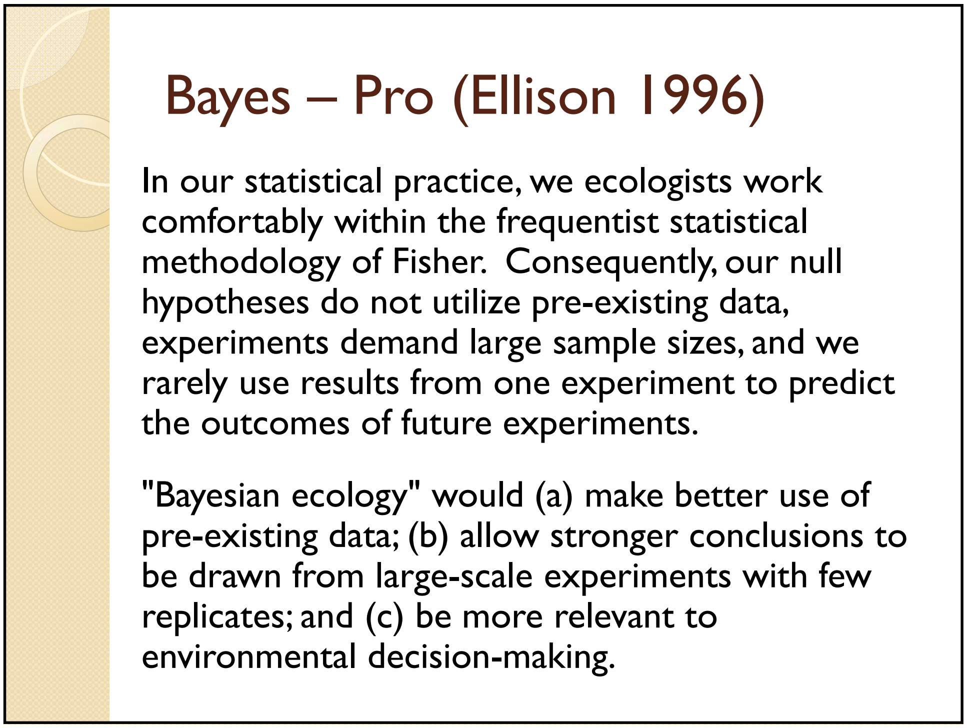
A beginner's guide to Bayesian Statistics

Interim Thoughts

$$p(c|r) = \frac{p(r|c)\,p(c)}{p(r)}$$

or

$$p(c|r) = \frac{p(r|c)\,p(c)}{\sum_{c^*} p(r|c^*)\,p(c^*)}$$



Bayes – Pro (Ellison 1996)

In our statistical practice, we ecologists work comfortably within the frequentist statistical methodology of Fisher. Consequently, our null hypotheses do not utilize pre-existing data, experiments demand large sample sizes, and we rarely use results from one experiment to predict the outcomes of future experiments.

"Bayesian ecology" would (a) make better use of pre-existing data; (b) allow stronger conclusions to be drawn from large-scale experiments with few replicates; and (c) be more relevant to environmental decision-making.

Bayes – Con (Dennis 1996)

Bayesian statistics involve substantial changes in the methods and philosophy of science. Before adopting Bayesian approaches, ecologists should consider carefully whether or not scientific understanding will be enhanced.

Frequentist statistical methods, while imperfect, have made an unquestioned contribution to scientific progress and are a workhorse of day-to-day research. Bayesian statistics, by contrast, have a largely untested track record.

Final Thoughts