Bayesian Statistics III

Matthieu Vignes

April 2024

Likelihood Ratio: how big is convincing?

We introduce the idea that drawing conclusions from likelihood ratios (or Bayes Factors) from fully specified models is context dependent. In particular, it involves weighing the information in the data against the relative (prior) plausibility of the models.

Overview

Recall that a fully specified model is one with no free parameters. So a fully specified model for X is simply a probability distribution $p(x|M)$. And, given observed data $X = x$, the LR for comparing two fully specified models, M_1 vs M_0, is defined as the ratio of the probabilities of the observed data under each model:

$$LR(M_1, M_0) := \frac{P(x|M_1)}{P(x|M_0)}.$$

For fully specified models, the likelihood ratio is also known as the *Bayes Factor* (BF), so we could also define the Bayes Factor for M_1 vs M_0 as:

$$BF(M_1, M_0) := \frac{p(x|M_1)}{p(x|M_0)}.$$

When comparing fully specified models the LR and BF are just two different names for the same thing.

In the previous example used to introduce LR, we considered the problem of determining whether an elephant tusk came from a savanna elephant or a forest elephant, based on examining DNA data. Specifically we computed the LR (or BF) comparing two models, M_S and M_F, where M_S denotes the model that the tusk came from a savanna elephant and M_F denotes the model that the tusk came from a forest elephant. In our example we found $LR = 1.8$, so the data favor M_S by a factor of 1.8. We commented that a factor of 1.8 is relatively modest, and not sufficient to convince that the tusk definitely came from a savanna elephant.

More generally, we would like to address the question: what value of the LR should we treat as "convincing" evidence for one model vs another?

It is crucial to recognise that the answer to this question has to be context dependent. In particular, the extent to which we should be "convinced" by a particular LR value has to depend on the relative plausibility of the models we are comparing. For example, in statistics there are many situations where we want to compare models that are not equally plausible. Suppose that model M_1 is much less plausible than M_0. Then, we must surely demand stronger evidence from the data (larger LR) to be "convinced" that it arose from model M_1 rather than M_0, than in contexts where M_1 and M_0 are equally plausible. Or when model M_1 is more plausible than model M_0.

In a disease example (continuous distributions), a medical screening test for the disease involves measuring the concentration (X) of a protein in the blood. In normal/healthy individuals X has a Gamma distribution with mean `1` and shape `2`. In diseased individuals the protein becomes elevated, and X has a Gamma distribution with mean `2` and shape `2`. If, for a particular patient, we observe $X = 4.02$, then the likelihood ratio for the

model that this patient is from the normal group (M_n) vs the model that the patient is from the diseased group (M_d) is `dgamma(4.02,scale=0.5,shape=2)/dgamma(4.02,scale=1,shape=2)` which is 0.0718. That is, the data favours this individual being diseased by a factor of approximately 14. The interpretation of the LR depends on the frequency of the disease in the population being screened. For example, suppose that only 0.5% of people screened actually have the disease. Then to outweigh that fact, we would have to demand a high LR to "convince" us that a particular person has the disease. In contrast, if 50% of people screened have the disease, then we might be convinced by a much smaller LR.

Calculations using Bayes theorem

The following calculation formalizes this intuition. Suppose we are presented with a series of observations $x_1, \ldots, x_n$, some of which are generated from model M_0, whilst the others are generated from model M_1. Let $Z_i \in \{0, 1\}$ denote whether x_i was generated from model M_0 or M_1, and let π_j denote the proportion of the observations that came from model M_j ($j = 0, 1$). So in the disease screening example, π_1 would be the proportion of screened individuals (prevalence) who have the disease. That is, $\pi_j = P(Zi = j)$. Bayes Theorem says that:

$$P(Z_i = 1|x_i) = P(x_i|Z_i = 1)P(Z_i = 1)/P(x_i). \text{ And also:}$$

$$P(x_i) = P(x_i|Z_i = 0)P(Z_i = 0) + P(x_i|Zi = 1)P(Z_i = 1).$$

Putting these together, substituting π_j for $P(Z_i = j)$, and dividing numerator and denominator by Pr(xi|Zi=0) yields:

$$P(Z_i = 1|x_i) = \pi_1 LR_i/(\pi_0 + \pi_1 LR_i),$$

where LR_i denotes the likelihood ratio for M_1 vs M_0 computed for observation x_i.

Numerical application

Suppose that only 0.5% of a screened population have a disease. Then a LR of 14 in favor of disease yields:

$$P(Z_i = 1|x_i) = 0.005 * 14/(0.995 + 0.005 * 14) \approx 6.6\%$$

In contrast, if 50% of screened people have the disease then the LR of 14 yields:

$$P(Z_i = 1|x_i) = 0.005 * 14/(0.995 + 0.005 * 14) \approx 93.3\%.$$

Thus in the first case, of patients with LR=14, on;ly a small proportion would actually have the disease. In the second case, of patients with LR=14, more than 90% would have the disease!

A useful formula

There is another way of laying out this kind of calculation, which may be slightly easier to interpret and remember, and also has the advantage of holding even when more than two models are under consideration. From Bayes theorem we have:

$$P(Z_i = 1|x_i) = P(x_i|Z_i = 1)P(Z_i = 1)/P(x_i), \text{ and}$$

$$P(Z_i = 0|x_i) = P(x_i|Z_i = 0)P(Z_i = 0)/P(x_i).$$

Taking the ratio of these gives:

$$P(Z_i = 1|x_i)/P(Z_i = 0|x_i) = \frac{P(x_i|Z_i = 1)P(Z_i = 1)}{P(x_i|Z_i = 0)P(Z_i = 0)}$$

This formula can be conveniently stated in words, using the notion of *odds*, as follows:

$$\text{Posterior Odds} = \text{Prior Odds} \times \text{LR (aka BF)},$$

as we have seen that the LR is sometimes referred to as the Bayes Factor (BF). Note that the "Odds" of an event E_1 vs an event E_2 means the ratio of their probabilities. So $P(Z_i = 1)/Pr(Z_i = 0)$ is the "Odds" of $Z_i = 1$ vs $Z_i = 0$. It is referred to as the *Prior Odds*, because it is the odds prior to seeing the data x. Similarly the *Posterior Odds* refers to the Odds of $Z_i = 1$ vs $Z_i = 0$ "posterior to" (after) seeing the data x.

Back to the numerical example

Suppose that only 0.5% of screened people have the disease. Then the prior odds for disease is 1/199. And a LR of 14 in favor of disease yields a posterior odds of 14/199 (or "14 to 199").

Suppose that 50% of screened people have the disease. Then the prior odds for disease is 1. And a LR of 14 in favor of disease yields a posterior odds of 14 (or "14 to 1").

In cases where there are only two possibilities, as here, then the posterior odds determines the posterior probabilities.

Exercise

a. Write a function to simulate data for the medical screening example above. The function should take as input the proportion of individuals who have the disease, and the number of individuals to simulate. It should output a table, one row for each individual, with two columns: the first column (x) is the protein concentration for that individual, the second column (z) is an indicator of disease status (`1`=disease, `0`=normal).

b. Write a function to compute the likelihood ratio in the medical screening example.

c. Use the above functions to answer the following question by simulation. Suppose we screen a population of individuals, 20% of which are diseased, and compute the LR for each individual screened. Among individuals with an LR "near" `c`, what proportion are truly diseased? Denoting this proportion $q_D(c)$, make a plot of $\log_{10}(c)$ (x axis) vs $q_D(c)$ (y axis), with `c` varying from `1/10` to `10`, say (so $\log_{10}(c)$ varies from `-1` to `1`.) Or maybe a wider range if you like, but the wider the range, the larger the simulation study you will need to get reliable results.

d. Use the computations introduced on this sheet to compute the theoretical value for $q_D(c)$, and plot these on the same graph as your simulation results to demonstrate that your simulations match the theory. It should provide a good agreement, provided your simulation is large enough.

e. Repeat the above, but in the case where only 2% of the screened population are diseased. Comment on differences.