

A PRIMER ON BAYESIAN APPROACHES TO HYPOTHESIS-TESTING

(AN OPTIONAL SUPPLEMENT TO THE BASIC LECTURE)

Overview

The Bayesian approach to testing hypotheses has two critical elements. First, a Bayesian approach stands the classical (by which I mean frequentist but I can't bring myself to type these "ist" designations) approach on its head. In the classical approach, our questions revolve around the probability of the data, given a specific hypothesis. In the Bayesian approach, our questions revolve around the probability of various hypotheses, given the data. You can see that the derivation of Bayesian methods will involve the same assumptions and manipulations of conditional probabilities that we discussed when we covered Bayesian estimation methods.

Technically, our Bayesian questions revolve around the likelihood of various hypotheses, given the data, which is not the same as the probabilities of various hypotheses; we will not delve into this distinction here - this is a heuristic exercise, after all. The point is that the classical approach is based on calculating probabilities of the data in hand under certain conditions (which are encompassed by the null and alternative hypotheses) and the Bayesian approach looks at probabilities of competing conditions (which are the hypotheses being compared) given the data in hand.

Second, a Bayesian approach integrates prior probabilities associated with competing conditions into the assessment of which condition is the most likely explanation for the data in hand. Put another way, a Bayesian approach evaluates the likelihoods of competing conditions by evaluating the change in the odds associated with the conditions, a change produced by assessing the data in hand. If the odds change sufficiently when the data are examined, then the scientist may alter his or her opinion about which competing condition is the most likely one. Now a dedicated Bayesian may take this process even further by weighing the change in the odds by the cost of mistakes in judgment. In other words, the dedicated Bayesian will weigh the new odds by the expected losses associated with an incorrect decision (accept one hypothesis when the other is true) and base the final decision on the combined evidence of the change in the odds and the possible costs associated with mistakes. Of course, one can integrate costs of mistakes into a classical approach, but a classical approach does not admit the concept of prior probabilities or include any criterion like assessing the change in the odds ratio suggested by analysis of the data in hand.

Perhaps the best way to appreciate a Bayesian approach and the methods therein is to walk through a relatively simple example. I have not presented all details in the example below and I am also presuming that you remember the basic elements of likelihood calculations from the lectures; in other words, this example is not intended as a self-contained tutorial.

Example

Situation. Scientists and policy makers in a fishery management agency are faced with a potential problem. Existing harvest rules for the species *Savory gustatorum* are based on two assumptions. First, the natural adult mortality rate is about 20% per year. Second, fishing under the rules imposes an annual adult mortality rate of 60%. Models of this fishery suggest that as long as annual adult survival is at least 20%, the population is sustainable and the harvesting is sustainable. But this year's annual survey of tagged adults suggests to some scientists in the agency that fishing mortality rates have increased, perhaps because more fishing licenses have been sold, because fisherpeople are getting better at finding the fish, because some individuals are not observing catch limits (fillets of this fish bring a nice price), or all of the above. The data in hand are that 60 of 80 tagged adults were landed (there is a reward for reporting the landing of a tagged fish). We assume that fish that aren't reported as landed survive the season (i.e.

natural mortality is negligible in the period between tagging and landing - this is just to make simple calculations for this pedagogic exercise). This is a higher fraction of tagged fish reported landed than in most previous years. The scientists are concerned because the models of the fishery indicate the emergence of real problems real quickly if less than 20% of the adults survive the season. These scientists think that fishing mortality rates are probably around 70% rather than 60%, which might create adult survival rates as low as 10% (if natural mortality rates remain at 20%) and so a change in rules for next season would be in order to prevent a problem from emerging. As often happens, this idea terrifies the policy people, who must develop new rules that would reduce fishing mortality and defend them before hostile audiences. So a rejection of the idea that fishing mortality is 60% is not to be made lightly. Let's take the data as they are and not worry about whether landings are under-reported, whether fishing and natural mortality are compensatory, or any other real-world complication.

A classical analysis. There is a simple classical analysis based on examining numbers of landings and survivors as binomially distributed random variables. If the rate of fishing mortality is 60%, then survival should be 40%. The expected number of survivors, that is, fish not landed, under a binomial model with probability of individual survival being 0.40 will be the product of the number of "trials" (number of tagged adults = 80) and the probability of individual survival ($p = 0.40$), in this case $np = 32$. When the number of trials is large, the number of "successes" converges on a t-distribution with expected value np and variance $np(1-p)$. We can then calculate a t-statistic as the observed number minus the expected number, divided by the square root of the variance. This yields a t-statistic value of -2.74 with 79 degrees of freedom. The probability of a result this extreme or more extreme under the null hypothesis is about 0.004 under a one-tailed test (alternative hypothesis is that mortality rates are larger than 0.60). So with this approach we would reject the null hypothesis in favor of a hypothesis that says that fishing mortality rates are higher than 60%.

A Bayesian approach. The Bayesian approach to the problem begins with specifying the hypotheses to be compared and the prior probabilities associated with each one. The goal of the Bayesian approach is to reach a decision, if possible, about which hypothesis is the more likely one, given the data in hand.

In our case, there are two hypotheses, that adult fishing mortality rates are 60% (which we will call **state A**) or that those rates are 70% (which we will call **state B**). The prior probability of state A being true is 0.80; this is based on the fact that in 8 of the last 10 years, the results of the annual tagging study would give an estimated fishing mortality rate of about 60%. The results of the last two years are suggesting a higher rate, perhaps 70%. The scientists doing the Bayesian analysis assign a prior probability of 0.40 to this state B because they have also noticed fewer juveniles and have found out that more licenses have been sold and think that these observations reinforce their sense that fishing mortality rates are higher. In our class notation then, $P^*(A) = 0.80$, $P^*(B) = 0.40$, and the prior odds ratio, $O^* = 2$.

The next step is to calculate the posterior odds ratio, O^{**} . Recall we calculate this as the product of the prior odds ratio and the likelihood ratio, $LR(AB)$, which is the ratio of the likelihood of the data, given state A, divided by the likelihood of the data, given state B. But before we do this, we need to clarify how we will decide which hypothesis we will accept as "more likely, given the data" and if we want to have a criterion for making the important decision about whether to reject state A in favor of state B. We need to face the decision issue because the management agency must take action if the fishing mortality appears to be rising so the outcome of our analysis is not to be taken lightly.

Of course, we could decide that we will change our adherence from state A to state B if the posterior odds ratio is different from the prior odds ratio. But given that there will be consequences to a decision, we might want to specify "how different" in some objective manner. This means that we might say that we'll turn to state B as the better choice if the magnitude of the posterior odds ratio is less than some value C because smaller values of O^{**} indicate a greater likelihood of state B. One way to proceed

is to set C as an odds ratio from a classical testing approach, namely, the ratio between the probability of rejecting the original hypothesis when it is really true and the probability of accepting it when it is indeed true, or $C = \alpha / (1 - \alpha)$ where α is the classical type I error probability.

With this criterion in mind, we calculate the likelihood of the data, that is, this year's tagging study results (80 fish tagged, 60 landed) under alternative binomial distributions (0.60 probability of landing or 0.70 probability of landing) because it is the ratio of these likelihoods that we want (that is, $LR(AB)$). Using the formula that $O'' = O' LR(AB)$, we find that $O'' = 0.06$. If we take $\alpha = 0.05$, then $C = 0.052$, and O'' is not quite small enough to change our minds in favor of state B. Of course, the reason for this is that the prior probabilities are very different and play a substantial role in determining the posterior odds ratio.

But no one at the agency is really happy with this result because it depends heavily on the prior probabilities. Some of the staff point out that if we had used diffuse (= uninformative) priors, the results would indicate that we should change our minds and accept state B. And even with the priors as stated, the pattern is right at the edge of our admittedly conservative criterion.

One approach is to model the time series of the data, that is, define state A as saying that fishing mortality rates have been constant for the last ten years and are the same for this year, and define state B as saying that the first 8 of the last ten years have experienced 60% fishing mortality rates and the most recent 2 years have experienced the higher rate and this higher rate accounts for this year's data. In this approach, we'd calculate the likelihoods of the longer series of data under each model and use that in our calculation of the posterior odds ratio. To do this, the scientists decide to use the data from the tagging studies in the odd numbered years to estimate the parameters of the binomial distributions. For state A, they use data from years 1, 3, 5, 7, and 9 to estimate a fishing mortality rate of 0.60 (which is how the data come out). For state B, they use data from years 1, 3, 5, and 7 to estimate the "old" fishing mortality rate (which comes out to 0.6) and the data from year 10 to estimate the "new" rate (which is 0.70). They then take data from years 2, 4, 6, and 8, along with this year's data, to examine each hypothesis; in this fashion they do not use the same data for parameter estimation and hypothesis testing. When the staff performs this analysis, the posterior odds ratio is actually slightly higher than 0.06; the minimal change is not surprising, given that the models differ only in the mortality rate for the last year of the data used in the test. We are in the same situation as before in terms of what to do.

But at this point, it's important to notice that there is an additional factor that must be taken into account, which is that the model for state A has a single parameter and the model for state B has two parameters. This difference could influence the "fit" of each model to the data (more parameters will always describe a given body of data better, but the improvement may not be meaningful if those additional parameters represent little more than random variation). We need to evaluate our competing models in this light. The typical method for doing so is to calculate a measure of "lack of fit" of model to data, a measure that is discounted by the number of parameters, and choose the model that minimizes this measure. Various measures are used for different types of problems; two popular measures are the Akaike Information Criterion (AIC) and the Bayesian Information Factor or Bayesian Information Criterion (BIC). The AIC, which has become very popular in ecology and evolutionary biology, is computed for a model of state A as

$$AIC(A) = -2 \log L(A) + 2p(A)$$

where $L(A)$ is the likelihood of the data under model A and $p(A)$ is the number of parameters embedded in model A. The BIC replaces the term $2p(A)$ with $p(A) + p(A) \log N$ where N = number of observations in the data. When we are comparing only two models, say A and B, we can calculate the change in AIC between them as

$$\Delta \text{AIC} = -2 \log \text{LR}(\text{AB}) + 2 [p(\text{A}) - p(\text{B})].$$

If $\Delta \text{AIC} > 0$, we would favor model B over model A (and vice-versa, of course). In our fisheries data, $\Delta \text{AIC} = 4.98$ so we would favor state B over state A even though there are more parameters in state B.

Yet all of these efforts may seem as frills that merely hide the fundamental dilemma that our Bayesian approach leaves us on the margin of a decision, and the agency must make a decision to rework the rules so as to lower next year's fishing mortality rate or do nothing and take a risk that the fishery will collapse.

A dedicated Bayesian will take one more step toward solving the problem. Let's consider the possible decisions and the costs of each decision:

Decision:	TRUE STATE:	
	A	B
Accept state A	0	u (A B)
Accept state B	u (B A)	0

where $u (A|B)$ is the cost of accepting state A when state B is true and $u (B|A)$ is the cost of accepting state B when state A is true. The u can be considered "loss functions" because they can, in principle, be not only scalars (i.e. a number representing an economic loss or time drain) but functions of parameters in the models or other complicated expressions. Now remember that, at least in principle, we could use our prior probabilities and likelihoods from the data to calculate posterior probabilities for each state, $P^{\wedge}(A)$ and $P^{\wedge}(B)$. Putting aside how we might do that for the moment, we could use these loss functions plus the posterior probabilities to calculate the expected loss under each decision. Let $E(u, A)$ be the expected loss if we accept state A and $E(u, B)$ be the expected loss if we accept state B. These can be calculated as

$$E(u, A) = 0 P^{\wedge}(A) + u (A|B) P^{\wedge}(B)$$

$$E(u, B) = u (B|A) P^{\wedge}(A) + 0 P^{\wedge}(B)$$

We can formulate a decision rule that incorporates the costs of wrong decisions using these values, which would be

if $E(u, A) < E(u, B)$, accept state A

if $E(u, A) > E(u, B)$, accept state B.

Note that this rule incorporates our analyses of the data through the effects of those analyses on the posterior probabilities and the weights those probabilities give to the loss functions.

There is a way to make this decision rule easier to use, especially for situations in which it is difficult to estimate explicit loss functions or in which only relative losses might be available. To begin, observe that, by dropping terms that multiply to zero, we can rewrite the expected loss equations as

$$E(u, A) = u (A|B) P^{\wedge}B$$

$$E(u, B) = u (B|A) P^{\wedge}A.$$

Remember that the posterior odds ratio is defined as $P^*(A) / P^*(B)$, although we often calculate it without ever explicitly calculating the posterior probabilities (i.e. we use $O^* = O^* LR(AB)$). Now we can rewrite the decision rule by substituting the posterior probabilities and loss functions

$$\text{if } u(A|B) P^*(B) < u(B|A) P^*(A), \text{ accept state A, otherwise accept state B.}$$

With some algebraic rearrangement, we can express the inequality as

$$\{u(A|B) / u(B|A)\} < \{P^*(A) / P^*(B)\}$$

and, remembering the definition of O^* , we can write

$$\{u(A|B) / u(B|A)\} < O^*$$

To simplify, let's define a "loss ratio" as $R = u(B|A) / u(A|B)$. We can think of R as the relative cost of mistakenly accepting state B compared to the cost of mistakenly accepting state A; if the loss functions are scalars, that is, simple numbers, then R becomes a simple relative cost. Now note that the inequality for our decision rule can be written as

$$1 / R < O^*$$

which allows us to develop the following decision rule:

$$\text{if } O^* R > 1, \text{ then accept state A}$$

$$\text{if } O^* R < 1, \text{ then accept state B.}$$

Now our agency economists estimate that the cost of mistakenly accepting state B, which includes wasting staff time and losing revenue from restricting fishing licenses, could be as high as \$100,000. The cost of mistakenly accepting state A, which involves doing nothing and allowing the fishery to collapse, which in turn means an economic loss to the region for at least five years, could run as high as \$1,000,000. Thus R is 0.10. The odds ratio was about 0.06, and the product is 0.006, which is much less than 1, and so given the expected losses from mistaken choices, the decision rule points clearly to accepting state B. Note that with an odds ratio of 0.06, the loss ratio would have to be enormously biased in the other direction to bring the product $O^* R$ above 1. If this fishery had no economic value to the region, or if there was a low economic loss that would occur only for a single year, the product might well take on a value that would indicate a choice to accept state A.