Archeologia e Calcolatori
13, 2002, 191-196

GETTING BAYESIAN IDEAS ACROSS TO A WIDE AUDIENCE

1. Introduction

I can understand how people may have reservations about Bayesian methods on the grounds that it can be difficult to find manageable mathematic expressions that do justice to one's prior knowledge and beliefs, and that in many cases the required computations may be excessively difficult. However, I cannot comprehend how it could not be *obvious* to everyone that the general Bayesian attitude – which I take to be that it is highly desirable to interpret new information in ways that explicitly and formally take account of prior knowledge and beliefs – is preferable to the attitude that prior knowledge and beliefs should have no place in the inferences drawn from any specific data set. This perhaps puts me at a disadvantage because if I could see how the latter attitude might make sense, then I could think of ways to frame arguments against it, rather than merely exclaiming that it's *obvious* that the Bayesian attitude makes more sense. As it is, I'm rather baffled by the fact that it doesn't seem to be obviously more reasonable to everyone, and I am at a loss to know how to construct arguments that will convince those who don't find the superiority of the Bayesian attitude self-evident. It is also disquieting that many statisticians who are vastly more experienced and talented than I am resist Bayesian approaches. However, I am heartened by the fact that other statisticians, equally or even more able than the first group, as far as I can tell, are strong advocates of Bayesian methods.

Beginning about 1962, for some years I taught introductory statistics courses in which I dealt with inference using the Neyman-Pearson approach, as best I understood it. This always seemed to me a rather austere and unnatural intellectual self-discipline, standing in much the same relationship to everyday reasoning as learning the moves and postures of classical ballet relates to just walking. Nevertheless, it seemed that this was the only way to do it, if one were to use statistical inference competently and correctly. It was probably always the most difficult and challenging part of the course for both me and students – far harder than teaching them how to plug numbers into formulas and compute a correlation coefficient or a chi-square statistic or the like – and a high proportion of students never really got it. Basically, the Neyman-Pearson approach to inference enables us to compute the probability of an observed sample statistic, *if it were the case that* the corresponding population parameter had some specified value, but it prohibits us from saying anything about the probability that the population parameter has any particular value, *given that* the sample statistic has a particular observed value.

This is supplemented with some rather murky (at least to students in an introductory course) considerations about the "power" of the test. In practice it was very difficult to get students to see how to make good use of the power concept, and virtually impossible to keep them from drawing conclusions about the probability of the population parameter, even though this was supposed to be a major "no-no." This made them "closet" Bayesians, but quite unskillful Bayesians.

I no longer recall when I first became aware of the Bayesian approach to inference, but I think that from the beginning it was like a great light dawning (like one of those light bulbs in the cartoons), and I was rather embarrassed that I had accepted the Neyman-Pearson doctrine as long as I had. So, as I say, it was unclear to me why everyone else exposed to Bayesian concepts should not become an instant convert. In the case of sophisticated statisticians, this is still unclear to me. For the relatively unsophisticated – and this includes all but a tiny handful of archaeologists – I suspect that a main source of resistance is that they never really understood the logic of Neyman-Pearson inference very well, but it provided the rituals they were taught to believe were founded on unchallengeable truths emanating from higher authorities (i.e., professional statisticians who really understood all those arcane symbols) and which could thus legitimize their conclusions. For this reason, they are uncomfortable about and suspicious of any alternative that they don't understand very well either, but which they have heard is controversial, and which asks them to unlearn a good deal of what (with considerable effort and pain) they have already managed to learn. Others (e.g. SHENNAN 1998) seem rather critical of formal statistical inference altogether. If I understand correctly, he advocates computing relatively uncontroversial descriptive statistics, then drawing conclusions intuitively. This is, indeed, often better than drawing unwarranted conclusions from poorly understood methods of formal inference. Particularly important is that it motivates us to try harder to obtain results whose patterning is so pronounced that formal tests of significance and formal confidence intervals are superfluous. Nevertheless, this approach fails to make good use of the possibilities of well-done formal inference in cases where the patterning is not so obvious.

I am not equipped to try to overcome resistances of trained statisticians to Bayesian attitudes. However, I can offer suggestions about how to get Bayesian ideas across to the much wider audience of relatively unsophisticated users of statistical methods. I suspect that an effective way to do this would be to write a truly introductory textbook that framed concepts of inference in Bayesian terms from the beginning. As *concepts* I do not think they are terribly difficult – in fact, I think they should be easier to grasp than the Neyman-Pearson concepts (or precepts). It is also desirable that such a book should emphasize estimation and confidence (or credibility) intervals more than hypothesis-testing.

Especially important is that it should give examples of Bayesian methods that are (a) applicable to classes of problems of real interest to nearly all archaeologists and (b) either require only very simple computations or can reliably and reasonably quickly get good results from widely-available computer packages that will run on standard desktops. Not surprisingly, I have drafted sections of such a book in preliminary form, although it remains to be seen whether I will ever follow through on this project. If I don't, someone else should. As an introductory book, much of it will necessarily cover topics such as descriptive statistics in ways not much different from the multitude of other introductory texts, and it will also have to deal with inferential statistics for which no easy Bayesian methods or nice computer programs yet exist.

2. Bayesian inference with an uninformative prior

I think that Bayesian inference with an uninformative prior should give students a logically coherent basis for what they tend to do, but incoherently, when they are trying to do classical inference. That is, rather than computing P(data|hypothesis) and then interpreting it as P(hypothesis|data), one recognizes that $P_{posterior}(H|D)$ is proportional to $L(D|H)P_{prior}(H)$, and, if one has good reason to think that $P_{prior}(H)$ is essentially flat over the relevant range of H, this becomes $P_{posterior}(H|D)$ is proportional to L(D|H). The student should be made aware that (a) you have to justify the assumption that $P_{prior}(H)$ is essentially flat, and (b) if you can justify a more informative distribution of $P_{prior}(H)$, and (c) this more informative distribution is mathematically tractable, you are missing an opportunity if you don't take advantage of this more informative prior.

3. Using an informative prior to improve estimates of true population proportions

One good example of a mathematically simple Bayesian method that very often applies to questions of archaeological interest is that of using the proportion of objects of a particular category in a collection to estimate the proportion of that category in the larger population represented by the sample. For example, if p is the observed proportion of type X is a given collection, what is a good estimate of π, the proportion of type X in the population represented by the sample, and what is an appropriate credibility interval for this estimate of π? In many cases, a beta distribution expresses our prior information quite well, and it is very easy to compute. Iversen (1984) goes into using a beta prior at some length, and he gives formulas for combining a beta distribution that captures one's prior information with one's sample data, to compute a posterior beta distribution that usually has a narrower credibility interval than either the prior distribution or a distribution computed from the sample data alone. These computations are quite easy. Nevertheless, his book, however elementary it may seem to statisticians, is far too advanced for most archaeologists. Much of what I think needs to be done in an introductory exposition is recasting what Iversen says in far more elementary terms and giving the mathematical and statistical background that he presupposes. It would also be desirable to give various examples of archaeological applications, such as those of ROBERTSON (1999, 2001).

One complication with the beta prior arises when one's prior estimate of p is quite different from the observed sample proportion, *p*, and both have fairly narrow credibility intervals. For example, suppose we are working in a region and time interval for which prior work has given us good reason to believe that type X constitutes 60 to 80% of the relevant sherd populations of nearly all sites in this region during this interval. A straightforward application of the methods described by Iversen would be to say that the mean of the prior probability distribution of π, μ_{prior}, is 0.7, and that it has a standard deviation (σ_{prior}) of 0.05. This means that the variance of the prior probability distribution will be $(0.05)^2$, or 0.0025. Plugging these values into equations given by IVERSEN (1984, 23) gives values for the constants a_{prior} and b_{prior} of 58.1 and 24.9, respectively[1]. That is, the prior beta distribution is the product of a normalizing coefficient (whose value need not be computed) times $\pi^{58.1-1}(1-\pi)^{24.9-1}$, or $\pi^{57.1}(1-\pi)^{23.9}$.

Suppose further that in a given collection of 200 sherds, which we have good reason to think reasonably approximates a simple random sample of the population of all sherds at a site within the given region and pertaining to the given time interval, only 60 of these sherds are of type X. That is, for this sample, $p = 0.30$. Using the binomial distribution, the probability that $p = 0.3$ when $\pi_{prior} = 0.7$ and the sample size (*n*) is 200, is $\pi^{60}(1-\pi)^{140}$, times another normalizing coefficient, whose value also need not be computed. Multiplying this probability by the prior probability function, we can simply add exponents of terms and get, for the posterior distribution, also a beta distribution, a different coefficient times $\pi^{57.1+60}(1-\pi)^{23.9+140}$, or $\pi^{117.1}(1-\pi)^{163.9}$. For this posterior distribution, $a_{post} = 118.1$ and $b_{post} = 164.9$. Plugging these back into Iversen's equations, we obtain $\mu_{post} = 0.417$, $\sigma^2_{post} = 0.000856$, and $\sigma_{post} = 0.0293$. That is, we are practically sure that, for this particular collection, the true value of p is somewhere between 0.36 and 0.47, most likely about 0.42.

Is this really sensible? The posterior probability distribution for p for this particular sample is quite different from both our prior distribution for p and the observed sample proportion, and there is almost no overlap in the credibility intervals. Should we be happy with this result? I found that students generally weren't, and they tended to give "wrong" interpretations of

[1] $a_{prior} = \mu_{prior}\{[\mu_{prior}(1-\mu_{prior})/\sigma^2_{prior}]-1\}$ and $b_{prior} = (1-\mu_{prior})\{[\mu_{prior}(1-\mu_{prior})/\sigma^2_{prior}]-1\}$.

cases like this example. Basically, what they were saying, although they didn't express it this way, was that they didn't actually think the above prior distribution applied after all. At first I thought they just hadn't gotten the Bayesian logic, but eventually I came to think they were acting more sensibly than I was. Something really does seem to be wrong here.

Tentatively, I think the problem is that a beta prior works pretty well as long as observed samples are small or observed proportions in large samples are not terribly different from μ_{prior} (or both, of course). But when the observed proportion in a large sample is more than about $2\sigma_{prior}$ away from μ_{prior}, the shapes of the tails of the beta distribution become very important, and they may do a very poor job of capturing our prior knowledge or beliefs. In such cases, some other distribution may work much better. We might, for example, begin with the idea that type X usually constitutes 60% to 80% of the relevant populations, but there is some small, but not tiny, probability that type X could be anywhere from 0% to 100% of a population. This might be expressed formally as the sum of two beta distributions, the one discussed above, plus another one that is flat over the interval from 0 to 1, with constants chosen to reflect one's prior relative belief about the proportion of populations for which p is probably between 0.6 and 0.8, relative to the proportion of populations in which we think p could be just about anything. We might, for example, believe that there is about a 95% probability that π is between 0.6 and 0.8, but we think the remaining 5% probability is evenly spread across the intervals 0 to 0.6 and 0.8 to 1. In effect, this would be like a beta distribution with heavy and flat tails. Since each of the two terms in this sum is a beta distribution, it seems we would get a posterior distribution that is also the sum of two beta distributions, so the computations should not be difficult. Obviously we have more difficulty in specifying our prior, since, in addition to μ_{prior} and σ_{prior}, we also have to have some basis for postulating a prior value for the ratio of the informative and uninformative beta distributions. That is, there are three parameters, rather than just two, that depend on prior information, and one's choice of a value for this third one may be particularly difficult to justify.

This is where my thinking now stands, for better or worse, on this complication about using beta priors to estimate proportions of categories of objects within populations. I hope I have made the issue clear, and I would greatly appreciate suggestions and comments from those more skilled in these matters than I am.

4. Good use of multiple radiocarbon dates and other information to estimate dates

Using multiple radiocarbon dates together with other information, such as stratigraphic sequences, is a prime example of an important problem that has become manageable by means of computer methods. Numerous papers by Buck and others (e.g., Buck *et al.* 1996) illustrate these methods. I do not have the competence to discuss them in any detail, and I am happy to leave such discussions to those actively developing these methods.

George L. Cowgill
Department of Anthropology
Arizona State University

REFERENCES

Buck C.E., Cavanagh W.G., Litton C.D. 1996, *Bayesian Approach to Interpreting Archaeological Data*, New York, Wiley.

Iversen G. 1984, *Bayesian Statistical Inference*, Beverly Hills, CA, Sage.

Robertson I.G. 1999, *Spatial and Multivariate Analysis, Random Sampling Error, and Analytical Noise: Empirical Bayesian Methods at Teotihuacan, Mexico*, «American Antiquity», 64, 137-152.

Robertson I.G. 2001, *Mapping the Social Landscape of an Early Urban Center: Socio-Spatial Variation in Teotihuacan*, Tempe, AZ (Doctoral Dissertation, Department of Anthropology, Arizona State University).

Shennan S. 1998, *Quantifying Archaeology*, (Second edition), Iowa City, University of Iowa Press.

ABSTRACT

A generally Bayesian attitude toward statistical inference seems to me so obviously superior to the “classical” Neyman-Pearson approach that it is difficult to comprehend why not everyone agrees. I believe that most non-statisticians learn classical procedures ritualistically but then interpret their results in naively Bayesian ways. It would be better if they became more sophisticated and knowing Bayesians. A truly introductory text on the logic of Bayesian inference, with some simple but useful applications, would probably help. Bayesian inference with an uninformative prior may yield the same results as classical inference, but with coherent rather than muddled logic. An example of a very useful but mathematically simple archaeological application of an informative prior is using prior information to improve estimates of true proportions of artifact categories in populations represented by small collections. However, a complication arises when the observed proportion in a fairly large sample is well outside the range considered at all likely for the relevant population, based on prior information. In this case, straightforward use of a *beta* prior distribution can yield results that seem unreasonable. Possibly our prior information is better represented by a modified *beta* distribution with “heavy” tails. Advice about this problem would be appreciated.