STUDENT PERCEPTIONS OF BAYESIAN STATISTICS

Gudmund R. Iversen, Swarthmore College, USA

The second half of this century has witnessed a very fruitful debate among statisticians about the relative merits of Bayesian and classical statistical inference. Neither side can claim a victory in this debate, since there is no way of proving that one approach is more correct than the other. But the debate has served the purpose of illuminating the strengths and weakness of each approach. The students we teach are to a very large extent exposed only to classical statistical inference. This is a choice made by their instructors, meaning all of us. This spring, in a small group of students studying both approaches I probed for their opinions of the two approaches. Not surprisingly, the Bayesian approach was well received by the students, even though they also had some misgivings about Bayesian statistics, at least early on.

INTRODUCTION

As you know, there are two kinds of statisticians in this world. There are the Bayesian statisticians and there are, for a lack of a better word, the classical statisticians. As the name indicates, the Bayesians do statistical inference using Bayes theorem, while the classical statisticians do statistical inference using confidence intervals and/or hypothesis testing. It is possible to compare them by saying that Bayesian statisticians are like Macintosh computer users while the classical statisticians are like Windows users. The reason Bayesian statisticians are like Macintosh users is that they know for sure they are right, so there simply is no reason to even discuss the issues. In spite of that, let us try some discussion anyway.

HISTORY

So that you can better evaluate my comments here today, let me give you a brief background describing my own exposure to Bayesian statistics. When I was a graduate student at the University of Michigan, in the US, in the early 1960s, I had the good fortune to work with the great Bayesian statistician Jimmy Savage. He had just come to Michigan from Chicago, and as you some of you may know, he had very bad eyesight. To lure him away from Chicago, the University of Michigan promised, among other things, that he could have a student to help him with many of his daily tasks such as picking up the mail, going to the library, etc.

I got that job. Not because I had any interest in Bayesian statistics at that time, much less because of any knowledge of Bayesian statistics on my part, but in typical graduate student fashion I needed the money. As it turned out, my year with Jimmy Savage became an wonderful, incredible experience. I helped him with his mail and his files and library errand and what not, but in between we talked at great length about Bayesian statistics and a good many other things, including neutrinos of all things. I still do not know why he gave so much of his time to this unknown graduate student, but I got to read his manuscripts and his professional correspondence and I learned a tremendous amount from him. My year with Savage became a mentor experience for me of the kind one can only hope to come away with as a graduate student, and I became rather hooked on Bayesian ideas. When I later transferred to another graduate school, I am certain it did not hurt me to have a letter of recommendation from him!

DISCUSSIONS

I do not see many discussions any more between the relative merits of the Bayesian and classical approaches to statistical inference. Since neither side can prove in any way that their approach to statistical inference is the better practice, it may also be that such discussions are not very fruitful. The main point I remember from those discussions is the advice given by classical statisticians to their Bayesian counterparts that the Bayesians should follow the example set by their great leader, the Reverend Bayes himself, and not have their papers published until three years after their deaths!

But there are discussions about what we should teach our students. Should we teach them Bayesian statistics or should we teach them classical statistical inference? Or maybe both? One such discussion took place about a year ago in a recent issue of *The American Statistician* across 33 pages with lead articles by Donald Berry, Jim Albert and David Moore, and discussions by Jeffrey Witmer, Thomas Short, Dennis Lindley, David Freedman and Richard Scheaffer. If you are interested in Bayesian statistics and/or statistics education, you will most likely recognize at least some of those names.

In their articles, Berry and Albert argue in favor of the teaching of Bayesian elementary statistics while Moore takes a more cautionary approach to the issue. Part of the papers touch on the amount of probability theory we should have in our introductory courses. Since David Moore is known to argue for less probability in these courses rather than more, it may not be too surprising that he argues against Bayesian statistics in introductory courses. The five paper discussants, Witmer, Short, Lindley, Freedman and Scheaffer, throw additional light on the issues that I will not take the time to get into here.

TEACHING CLASSICAL STATISTICS

All of us at this conference are excellent statistics teachers, I am sure. That is why we are at this conference in the first place. I know that I am one of the good teachers, because I can make even the most difficult and obscure statistical points understandable to my students, and they all leave college as better people because they took one or more statistics course from me.

So why is it that I got answers like these last fall when I asked on a test what the interpretation is of a confidence interval that goes from 23.6 mpg to 26.4 miles per gallon for the mean mileage of cars? Miles per gallon is the American way of measuring the mileage we get from a car. And I quote:

> "95% of the intervals would fall within the two values of the parameter." (At least the student knows there is a parameter involved and that the correct answer has something to do with having many different intervals from many different samples.)
>
> "95% of the intervals will lie in this interval." (At least the student knows the answer has something to do with constructing many intervals from many samples, but otherwise it is hard to know what to make of this answer.)
>
> "95% chance that the actual value would be contained within the confidence interval." (Do I have a closet Bayesian in the class?)
>
> "95% sure that μ is between 23.6 mpg and 26.4 mpg." (How can you get to be more Bayesian than this without ever having been exposed to Bayesian statistics? Bayesian statistics must have some sort of intuitive appeal.)
>
> "95% of the differences in the population parameter would be within the interval." (I still do not understand what that answer means.)
>
> "95% of our data would fall within this interval." (Quite often I get this answer, invented by the students, presumably from something I have said in class.)
>
> "95% of the values will fall in this range." (Another version of the same answer as above.)
>
> "95% of the cars in the population has mean between 23.6 and 26.4 mpg." (It is hard to imagine how a single car could at the same time have a mean.)
>
> "The true population value will lie between the parameters 95% of the time." (Even a diehard Bayesian would not go that far!)

I will save you from having to listen to any more quotes of wisdom from my students. Mind you, all of this is after I have very, very carefully laid out the correct answer to such a question in class several times before the test as well as promised there would be a confidence interval question on the test! In sheer desperation I have even dictated the correct answer in class the day before the test!

Aside from the fact that these answers show my failure as a teacher, I see the answers really more as cries in the wilderness from students who simply do not get it when it comes to classical statistical inference. Or, at least, they have not studied the material hard and long enough. This is specially so if they skip class and try to learn about confidence intervals from the notes of other students or on their own. This particular class was not a unique experience; at home I have a much longer collection of strange interpretations of confidence intervals. Also, my guess is that I am not very wrong if I say that many of you have had the same experiences I have had, both with confidence intervals and hypothesis testing. Significance levels and *p*-values rank right up there with confidence intervals as mysterious statistical concepts.

STUDENT OPINIONS

What has been missing in my comments up to this point is any statement from the students themselves about Bayesian versus classical statistics. Learned professors have been discussing and even fighting about which is the ‘right’ approach to statistical inference for a long time. They have also discussed to considerable lengths what approach to teach in their classes and their textbooks.

Maybe it is time to let the students, our customers, speak their minds on the issue of classical versus Bayesian statistical inference. With this talk in mind, I organized a small group of students this spring to read some of the Bayesian literature with me and then tell me what thoughts came to mind about the two approaches to statistical inference. These students had all taken at least one course in statistics prior to joining this seminar group. Clearly, in no way does this group represent any kind of random sample of students from Swarthmore College or from anywhere else, but there may still be things we can learn from their comments.

CLASSICAL INFERENCE

The group started with a review of classical statistical inference. Here are some of the things the students said about this approach. They found it hypocritical for classical statisticians to use the expression: “We are 95% confident that our confidence interval contains the unknown parameter.” As statisticians, we may all know the correct interpretation of such a statement, but the students found it all too easy to become what they called a closet Bayesian and interpret the interval as a Bayesian probability interval, even if they were not supposed to. In the end, we still do not know whether our particular confidence interval contains the unknown parameter or not, and we are not doing anyone a service by hiding our classical interpretation by such a statement.

Not surprisingly, the students objected to the arbitrariness of using 0.05 as a common significance level and 95% as a common confidence level. They were more approving of *p*-values and letting the reader decide the importance of a result. Also, the students bemoaned the fact that they hardly ever see any discussions of the statistical power of a test, after the test has been performed, in spite of the importance of this concept and how it had been stressed in their previous courses.

In short, the students all expressed opinions very much along the lines I expected them to have. I tried to stay neutral throughout these discussions, but from me and from other sources they knew there was an alternative view of statistical inference coming along very soon in the seminar. That could account for some of their opinions about classical statistical inference.

BAYESIAN INFERENCE

Surprisingly to me, the students did not see that it made much difference whether a parameter was considered in classical statistics as a fixed quantity or in Bayesian statistics as a quantity with its associated probability distribution. To many of us this distinction provides the fundamental difference between the two approaches, but that

point did not seem to have come across to the students after their early exposure to Bayesian statistics.

Not surprisingly, the role of the prior distribution played a large role in their discussions of Bayesian statistics. They worried that forcing the user to make use of a conjugate prior distribution because of mathematical convenience and elegance would limit the applicability of Bayesian statistics, be it a beta distribution for binomial data or a normal distribution for normal data. My guess is that, with further experience with Bayesian statistics, they would see that even conjugate priors allow for considerable flexibility, and with today's computing facilities conjugate priors are not as central as they once were, even though we all like to teach them to our students.

Furthermore, the students worried about the role of the prior distribution and its impact on the final results of the Bayesian analysis. However, I think they did realize that it would not be the end of the world if two people ended up with different posterior distributions because they had started with different prior distributions. This would be particularly so if the two prior distributions represented real differences in people's earlier knowledge. They liked how the prior distribution makes explicit the subjectivity that people always bring to any analysis. On the other hand, they argued that prior distributions could be hard to come up with, and not using Bayesian statistics avoids the subjectivity that comes from the use of prior distributions.

Also, not surprisingly, they liked the notion of Bayesian probability intervals as opposed to classical confidence intervals. This point reverts back to the difficulties many students have with confidence intervals, and it was a relief to my students that Bayesian statistics actually makes it possible to come up with intervals that have the "proper" interpretation. They also liked the idea of personal probabilities, such that it is possible to talk about probabilities of unique events and not limiting the probability concept to repeatable events.

I had expected unqualified enthusiasm for Bayesian statistics from my students, but that did not happen, at least not during the early part of the seminar. For example, they liked the fact that with classical statistics everyone analyzing the same data with the same methods would arrive at the same results. Since that is the case, the entire classical approach was seen as a way of communicating the actual information contained in the data without the influence of any subjective opinions on the part of the analyst.

Finally, because the classical and Bayesian approaches often lead to approximately the same numerical results, for example, a Bayesian probability interval based on a flat prior and a classical confidence interval are often numerically identical, they argued that the two approaches are really not so different after all. Granted, these students had had only a limited exposure to Bayesian statistics, but this way of thinking should strike terror in any statistician's heart. What this shows, I think, is that our customers are not all that moved by the differences between the two approaches the way we statisticians are. They did not say it, but to me I really heard the students say: "So there are differences, but as non-statisticians, who cares?"

This attitude was very surprising to me, but I figured that if I could have these students available to me over a longer period of time, then I could cure them of this opinion and make them think they way they really should think about these matters.

REFERENCES

Albert, J. (1997). Teaching Bayes' rule: A data-oriented approach. *The American Statistician*, *51*, 247-253. (Plus discussions by J. Witmer, T. H. Short, D. V. Lindley, D. A. Freedman, and R. L. Scheaffer, and reply).

Berry, D. A. (1997). Teaching elementary Bayesian statistics with real applications in science. *The American Statistician*, *51*, 241-246. (Plus discussions by J. Witmer, T. H. Short, D. V. Lindley, D. A. Freedman, and R. L. Scheaffer, and reply).

Moore, D. S. (1997). Bayes for beginners? Some reasons to hesitate. *The American Statistician*, *51*, 254-261. (Plus discussions by J. Witmer, T. H. Short, D. V. Lindley, D. A. Freedman, and R. L. Scheaffer, and reply).