TEACHING INTRODUCTORY STATISTICS FROM A BAYESIAN PERSPECTIVE

James Albert
Bowling Green State University
USA

This paper reviews the past and current interest in using Bayesian thinking to introduce statistical inference. Rationale for using a Bayesian approach is described and particular methods are described that make it easier to understand Bayes' rule. Several older and modern introductory statistics books are reviewed that use a Bayesian perspective. It is argued that a Bayesian perspective is very helpful in teaching a course in statistical literacy.

THE "STATISTICS FOR POETS" CLASS

Iversen (1992) describes three levels of an introductory statistics course taught at universities: the "mathematical statistics" class taught to students with three semesters of calculus, the "applied statistics" class focused on teaching statistical methods, and the lower-level introductory statistics class taught to students with only a high school mathematics background. As Iversen explains, there has a new trend at this lowest introductory level to teach statistics as a liberal art. The goal of this "statistics for poets" class is to communicate in a general way what statistics is all about. In this class, it is not necessary to teach particular inferential methods such as a t test. However, it is desirable to communicate a few fundamental inferential concepts, such as the distinction between samples and populations and that sample data provide an incomplete description of a population. The student should understand the correct interpretation of statistical "confidence" which underlies interval estimation and tests of hypotheses.

STUDENTS' CONCEPTIONS OF PROBABILITY

The traditional method of teaching statistical inference is based on the frequency notion of probability. Is this the natural interpretation of probability for students taking an introductory statistics class? Hawkins and Kapakia (1984) review research on children's conceptions of probability. Although the statistics students are older than the young children discussed in this article, many of the ideas expressed by Hawkins and Kapakia apply to these college students. The "classical" notion of probability based on the notion of equally likely outcomes doesn't provide a very good foundation for later work in probability where events are not equally likely. Although the frequency approach to teaching probability is helpful in the situation where students can perform random experiments, there are conceptual difficulties in distinguishing between the observed relative frequency and the actual probability of an outcome obtainable in an infinite sequence of experiments. Hawkins and Kapakia advocate teaching on the basis of intuitive or subjective probability. This approach is accessible to less mathematical sophisticated children, since it is based on comparisons of likelihoods, rather than specification of fractions. Also the condition of coherence can be helpful to explain that probabilities assigned by a student must satisfy particular rules. Hawkins and Kapakia believe that subjective probability is "closer to the intuition that they try to apply in formal probability situations." (p. 372). They think that frequentist and classical approaches should be blended together with subjective approaches – otherwise a focus on frequentist or classical notions "may well conflict with the children's expectations and intuitions" (p. 372).

Albert (2001) gave college students a short survey to learn about their interpretation of probability. For nine questions, the students were asked to provide probabilities and supporting explanation. Although the students were successful in specifying probabilities for stylized problems (say, balls in a box) with equally likely outcomes, they were less able to use frequency and subjective viewpoints to obtain probabilities. There was a strong tendency for these students to use the classical notion of probability even when it was inappropriate.

Steinberg and Von Harten (1982) also assert that a subjective approach allows the student to assign a probability to a wide range of situations. Moreover, Bayes' rule offers an opportunity for learning from experience. Falk and Konold (1992) also remark (p. 162) that "many people's sound intuition in learning from experience and revising their beliefs are consistent with Bayesian analysis." Although these authors recognize the incoherence in the students' specification of probabilities (Kahneman, Slovic, & Tversky, 1982; Konold, Pollatsek, Well, Lohmeier, & Lipson, 1993), they advocate capitalizing on their commonsense notions about subjective probability to establish students' confidence in their abilities to reason probabilistically.

Shaughnessy (1992) discusses the three views of probabilities and discusses which view should be taught in the grade schools. Shaughnessy (p. 468) "advocates a pragmatic approach which involves modeling several conceptions of probability. The model of probability that we employ in a particular situation should be determined by the task we are asking our students to investigate, and by the types of problems we wish to solve." From this viewpoint, one could regard the teaching of statistical inference in college as one particular stochastic challenge that we face as instructors and the Bayesian subjective view may be the interpretation of probability that facilitates the learning of this difficult topic.

WHY BAYES' IN TEACHING INFERENCE?

What are the advantages of a Bayesian viewpoint in the learning of statistical inference? Simply, Bayes' thinking is more intuitive than the frequentist probability viewpoint and better reflects the commonsense thinking about uncertainty that students have before taking the statistics class. Students have dealt with uncertainty in their lives and use words such as "likely", "possible", "rare", and "always" to reflect different degrees of uncertainty. Subjective probability is a way of assigning numbers, on a scale of 0 to 1, to these different degrees of uncertainty. Beliefs of a person can change as one obtains new information. As one obtains new information or data, a person's beliefs or subjective probabilities can change. Bayes' rule is the recipe for determining exactly how the probabilities change in the light of new evidence. The Bayesian paradigm reflects the scientific method of learning, where one has initial beliefs about the world, an experiment is observed, and the new beliefs blend one's previous opinion about the world with the information obtained in the experiment (Berry, 1997).

In a statistical estimation problem, one typically wants to be confident that a computed interval estimate contains the parameter of interest. In a testing problem, one is interested in the probability that a particular hypothesis is true. But, in the traditional frequentist viewpoint towards inference, one is confident only in the performance of the interval estimate or hypothesis test in repeated sampling. This is helpful in the situation where one is performing a large number of 95% confidence intervals or hypothesis tests of level .05, but doesn't help the applied statistician who is interested in making an inference based on a single dataset.

In contrast, Bayesian inferential statements are made conditional on the observed data. Since parameters are viewed random, it makes sense from a Bayesian perspective to say that the interval (.04, .54) covers the proportion p. It makes sense to talk about the probability that the null hypothesis is true. These are the types of inferential conclusions that applied scientists and students want to make. Actually, it is relatively common for a student in a traditional statistics class to make the incorrect statement that a proportion p falls in a 95% confidence interval with probability .95. (Actually, from a frequentist perspective, the probability that the interval covers p is either 0 or 1 – either the proportion is in the interval or it isn't.) On the other hand, it would be very unlikely for a student in a Bayesian class to use a frequency viewpoint to interpret a 95% Bayesian probability interval.

One criticism of the Bayesian approach in teaching is that the student is faced with the new problem of choosing a prior that reflects one beliefs about the parameter. However, in my experience, it seems that the prior construction is a useful way of clarifying the distinction between a parameter and a statistic. By thinking of alternative parameter values and their relative likelihoods, the student needs to have a clear notion of the parameter in the particular problem. In contrast, it is very easy for a student taught frequentist methods to confuse, say, the proportion in the entire proportion with the computed sample proportion.

INTRODUCING BAYES' RULE USING COUNTS

In this liberal arts statistics class, the big idea to teach is Bayes' rule. The strategy that we use is to develop some intuition for Bayes' rule using simple examples for two possible models or parameters. We present some methods in this section that seem helpful in communicating the big idea in this setting. Once the students understand Bayes' rule, then one can use the computer to automate Bayesian calculations for more sophisticated problems.

There has been interest among psychologists on people's concepts of probability and learning of probability concepts. One particular project focuses on Bayes' rule – what are effective ways of presenting tables of probabilities so that people can correctly state conditional probabilities? Gigerenzer and Hoffrage (1995) found that people generally are more successful in Bayes' rule calculations when presented with tables of counts rather than tables of probabilities. In our experience in teaching this introductory statistics class, we have also found that students seem to understand conditional probability statements better in a two-way table when the table has been presented as counts rather than probabilities.

Albert (1997) illustrates the use of a "Bayes' box" to perform inference about categorical models and proportions. The Bayes' box is a table where a set of hypothetical people are classified by a "model" and the "data" (positive or negative blood test result). This method is used, in probability form, by Antleman (1997) in doing Bayes calculations and Albert and Rossman (2001) use this table, in count form, as their basic method for introducing Bayes' rule.

Holt and Anderson (1996) illustrate another simple method of developing intuition for Bayes' box using counts rather than probabilities. To illustrate this method, suppose that one have two boxes. Box 1 has two white balls and one black ball; Box 2 has two black balls and one white ball. Suppose you choose a box at random and select one ball with replacement from the box – it turns out that this ball is white. What is the probability that you drew from Box 1? The student observes the balls in the two boxes. If both boxes are equally likely to be chosen, then all six balls are equally likely to be drawn. We are told that the ball drawn is white. We note that there are a total of 3 white balls, each likely to be chosen, and two of these white balls are in Box 1. So P(Box 1 | white ball drawn) = 2/3.

Now suppose that we draw a 2nd ball from the same box as we drew the 1st one. Again we draw a white – what is the new probability that we are drawing from Box 1? Currently we believe that P(Box 1) = 2/3 and P(Box 2) = 1/3, so our unknown box is twice as likely to be Box 1 than Box 2. We represent this knowledge by drawing two Box 1's and one Box 2. Each of the 9 balls shown in the diagram are equally likely to be chosen. We are again told that we have drawn a white. There are now 5 white balls that can be chosen and four come from Box 1. So P(Box 1 | 2nd white ball drawn) = 4/5.

It would be tedious to use this algorithm to perform Bayes' rule calculations for more interesting problems. However, in the construction of the multiple boxes, the student has to think about the current probabilities of the two boxes, and the student sees how the multiple white ball observations are resulting in higher probabilities given to the "BOX 1" model.

USING DISCRETE PRIORS

After Bayes' rule has been introduced for simple inferential problems, one is interested next in using Bayes' rule to learn about a population proportion or a population mean. But the use of continuous prior distributions to model beliefs about a proportion or a mean is much more difficult conceptually than the discrete approach outlined in Section 6. Discrete priors provide a useful bridge between the two-way Bayes' box calculations and the more realistic modeling using continuous priors. This approach is a common way of introducing Bayesian thinking for a proportion or a mean; see, for example, Albert (1995), Schmitt (1969), Blackwell (1969), and Phillips (1973).

To describe the use of discrete priors in a simple setting, suppose a student wishes to learn about the proportion p of students who smoke on-campus. Certainly she has some opinion about the location of this parameter, but it can be a difficult task to her to construct a prior that roughly matches her beliefs. To construct a discrete prior for p, she must (1) make a list of "plausible" values of the proportion and (2) assign relative likelihoods to these different proportion values. In our teaching, we give the student the equally-spaced proportion values 0, .1,

…, .9, 1 and focus our instruction on how to assign probabilities to these values. It is easier to think about relative likelihoods than probabilities and these likelihoods can be stated in terms of whole numbers. So the student can assign the whole number 10 to the most likely proportion value, 5 to the proportion values that are half as likely as the most likely value, and so on.

After the prior has been assigned and the data collected, the posterior calculations are straightforward. In our example, 40 students were surveyed and 9 smoked (so 31 do not smoke). The likelihood of this datum result is $L(p) = p^9 (1-p)^{31}$. A Javascript program has been written to accompany Albert and Rossman (2001) that automates the calculation of the posterior probabilities. The student enters his or her prior and the number of successes and failures in the sample. The program displays the likelihoods (in normalized form, where the maximum value is 10,000), the products of the prior probabilities and likelihoods, and the (normalized) posterior probabilities.

Since the posterior probabilities are computed easily using a computer, the instructor can focus the instruction on the use of the posterior probabilities for different inferences. A good estimate at the proportion is the value of p, here p = .3, with the highest posterior probability. The set of values {.2, .3} is a 95% probability set for the proportion of smokers on-campus – the probability that p is included in this set is approximately .95.

AUTHENTIC ASSESSMENT BY USE OF A STUDENT PROJECT

A good way for the student to learn the entire inferential process is by means of a sample survey project. Albert (2000) describes in detail the implementation of this type of project in the introductory statistics class. In this project, a group of students decides on two questions of interest, each having a yes/no response. The group takes a random sample of size 40 from the student undergraduate body with the goal of learning about each of the two proportions. In a project write-up, the student describes all aspects of the statistical investigation including:

- The background behind the questions that were asked. Why were the students interested in the answers to these questions?
- What did the students think they would find out? How are these beliefs reflected in the prior distribution for each proportion?
- How was the “random” sample chosen?
- Contrast the prior and posterior distributions. Were the sample results consistent with the students’ prior beliefs. If the results were inconsistent, how were the probabilities for p modified by the data?

There were several nice features of this particular survey project. First, the students got a good understanding of a parameter in the process of constructing the prior. The “proportion of all students that smoke” is better understood when the student thinks about the plausibility of alternative values of this proportion. Second, this project illustrates the scientific method in practice, where the student has initial opinions about a population, designs an experiment and collects data to learn about the population, and then modifies his/her opinions on the basis of the data.

SOME UNDERGRADUATE BAYESIAN BOOKS AND SOFTWARE

Following the work of Harold Jeffreys, Jimmie Savage, and Dennis Lindley, there was a strong interest in Bayesian inference in the 1960’s. A number of books appeared at this time that presented Bayesian thinking from an undergraduate viewpoint. The text by Schmitt (1969) is notable for presenting Bayesian thinking assuming only a modest high school mathematics background. Inference is performed for a proportion first using discrete priors, and the book focuses on computation rather than analytical work to obtain posterior distributions. Blackwell (1969) is a brief book that describes Bayesian thinking in very simple settings. Phillips (1973) is an excellent undergraduate text that focuses on Bayesian inference for problems in the social sciences.

Unfortunately, the books mentioned above are all out of print and there has been a recent revival of undergraduate Bayesian statistics texts that flows from the current interest Bayesian modeling in research and applications. Antleman (1997) presents Bayesian inference at an undergraduate level assuming some knowledge of calculus. Minitab commands and macros are used to illustrate the probability calculations. Berry (1995) presents Bayesian thinking at an introductory level assuming only a high school mathematics background. As in the earlier Bayesian books, Bayes' rule is illustrated first in Berry (1995) for simple inferential problems where the unknown model is categorical or the unknown proportion can take only a finite set of values. Another introductory Bayesian text, Albert and Rossman (2001), also focuses on the use of discrete priors to learn about means and proportions. Albert and Rossman's text is equally divided between data analysis and probability/inference topics. This book follows a "workshop" format, where the students learn the material by working on a number of directed activities. In a class that uses the Albert and Rossman text, the instructor does not lecture, but interacts with the students working on the activities in small groups.

One difficulty in teaching Bayesian inference is the lack of software to simply the computations of posterior distributions. Albert (1996) wrote a set of Minitab macros for introducing Bayesian inference. Although these macros were written in the older "exec" Minitab macro language, a toolbox of Bayesian Minitab macros is also available in the newer "local" style of macros. This set of introductory Bayes programs is also available in the MATLAB programming language. Tony O'Hagen's First Bayes package is well suited for introducing inference for proportion, mean, and regression problems. In this software, one can use the standard collection of conjugate distributions and their mixtures to model prior opinion. The First Bayes software makes it easy to graph and summarize the posterior and predictive distributions.

REFERENCES

Albert, J. (1996). *Bayesian Computation Using Minitab*. Belmont, CA: Duxbury Press.

Albert, J. (1995). Teaching inference about proportions using Bayes rule and discrete models. *Journal of Statistical Education, 3.* Available at: http://www.amstat.org/publications/jse/v3n3/albert.html

Albert, J. (1997). Teaching Bayes' rule: a data-oriented approach. *The American Statistician*, *51*, 247-253.

Albert, J. (2000). Using a sample survey project to assess the teaching of statistical inference, *Journal of Statistical Education, 8*. Available at: http://www.amstat.org/publications/jse/secure/v8n1/albert.cfm

Albert, J. (2002). College students' perception of probability. (Tech. Report 2002-12). Bowling Green State University, Department of Mathematics and Statistics.

Albert, J., & Rossman, A. (2001). *Workshop statistics: Discovery with data, a Bayesian approach*. Emeryville, CA: Key College.

Antleman, G. (1997). *Elementary Bayesian statistics*. Cheltenham: Edward Elgar Publishing.

Berry, D.A. (1995). *Basic statistics: A Bayesian perspective.* Wadsworth: Belmont.

Berry, D.A. (1997). Teaching elementary Bayesian statistics with real applications in science. *The American Statistician, 51*, 241-246.

Blackwell, D. (1969). *Basic Statistics*. New York: McGraw Hill.

Falk, R., & Konold, C. (1992). The psychology of learning probability. In F. Gordon and S. Gordon (Eds.), *Statistics for the Twenty-First century.* Washington DC: The Mathematical Association of America.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review, 102*, 684-704.

Hawkins, A.S., & Kapakia, R. (1984). Children's conceptions of probability – a psychological and pedagogical review. *Educational Studies in Mathematics, 15,* 349-377.

Holt, C.A., & Anderson, L. (1996). Classroom games: understanding Bayes' rule. *Journal of Economic Perspectives*, *10,* 179-187.

Iversen, G.R. (1992). Mathematics and statistics: An uneasy marriage. In F. Gordon and S. Gordon (Eds.), *Statistics for the Twenty-First Century*. Washington DC: The Mathematical Association of America.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education, 24*, 392-414.

Phillips, L.D. (1973). *Bayesian statistics for social scientists*. London: Nelson.

Schmitt, S. (1969). *Measuring uncertainty: An elementary introduction to Bayesian statistics*. Reading, MA: Addison-Wesley.

Shaughnessy, J.M. (1992). Research in probability and statistics: reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics and learning* (pp. 465-494). New York: Macmillan.

Steinberg, H., & Von Harten, G. (1982). Learning from experience – Bayes' theorem: a model for stochastic learning. *Proceedings of the First International Conference of Teaching Statistics*, Volume 2, (pp. 701-714). Sheffield, U.K.: Teaching Statistics Trust.