1

Bayes for Beginners? Some Pedagogical Questions

David S. Moore

Purdue University, West Lafayette, IN USA

S. Panchapakesan and N. Balakrishnan (eds.), *Advances in Statistical Decision Theory*, Birkhäuser, 1997, 3–17.

Abstract.

Ought we to base beginning instruction in statistics for general students on the Bayesian approach to inference? In the long run, this question will be settled by progress (or lack of progress) in persuading users of statistical methods to choose Bayesian methods. This paper is primarily concerned with the *pedagogical* challenges posed by Bayesian reasoning. It argues, based on research in psychology and education and a comparison of Bayesian and standard reasoning, that Bayesian inference is harder to convey to beginners than the already hard reasoning of standard inference.

Keywords and phrases: Bayesian inference, statistical education

1.1 Introduction

From *Advances in Statistical Decision Theory*, Birkhuser, Boston, 1997, 3-17.
It is a pleasure to dedicate this paper to Shanti Gupta, who offered me my first position and contributed in many ways to my statistical education.

Bayesian methods are among the more active areas of statistical research. Moreover, a glance at recent journals shows that researchers have made considerable progress in applying Bayesian ideas and methods to specific problems arising in statistical practice. It is therefore not surprising that some Bayesians have turned their attention to the nature of introductory courses in statistics for general students. These students come to us from many fields, with the goal of learning to read and perhaps carry out statistical studies in their own disciplines. Several recent Bayesian textbooks [e.g., Albert (1996a), Antleman

(1996), Berry (1996a)] are aimed at such students, and their authors have argued in favor of a Bayesian approach in teaching beginners [Albert (1995), (1996b), Berry (1996b)]. The arguments, put crudely to make the issues clear, are: (1) Bayesian methods are the only right methods, so we should teach them; (2) Bayesian inference is easier to understand than standard inference.

Although I do not accept argument (1), I have no wish to participate in the continuing debate about the "right" philosophy of inference. Inference from uncertain empirical data is a notoriously subtle issue; it is not surprising that thoughtful scholars disagree. I doubt that any of us will live to see a consensus about the reasoning of inference. Indeed, the eclectic approach favored by many practicing statisticians, who use Bayesian methods where appropriate but are unconvinced by universal claims, may well be a permanently justifiable response to the variety and complexity of statistical problems.

I do wish to dispute the second argument. I will give reasons why Bayesian reasoning is considerably more difficult to assimilate than the reasoning of standard inference (though, of course, neither is straightforward). I even doubt the common Bayesian claim that at least the *results* of Bayesian inference (expressed in terms of probabilities that refer to our conclusions in this one problem) are more comprehensible than standard results (expressed in terms of probabilities that refer to the methods we used). My viewpoint is that of a teacher concerned about issues of content and pedagogy in statistical education. I will point to research in education and psychology rather than in statistics.

I preface the paper's main argument by briefly stating in Section 2 what I consider to be decisive empirical and pragmatic reasons for not basing introductory courses for general students on Bayesian ideas at the present time. Those who think that the conditions I describe in Section 2 will change in the future can consider the rest of the paper as posing questions of pedagogy that we will have to face in the coming Bayesian era.

1.2 Unfinished business: The position of Bayesian methods

Let us attempt to be empirical. Here are two questions that bear on our decision about teaching Bayes to beginners:

- Are Bayesian methods widely used in statistical practice?
- Are there standard Bayesian methods for standard problem settings?

Several recent surveys suggest that Bayesian methods are little used in current statistical practice. Rustagi and Wright (1995) report the responses of all 103 statisticians employed in Department of Energy National Laboratories. Asked to choose from a long list "the three statistical techniques that have been most important to your work/research," only four mentioned "Bayesian methods." The top responses are not surprising: "Regression analysis" (63), "Basic statistical methods" (37), "Analysis of variance" (26), and "Design of experiments" (26). Bayesian methods tied for 17th/18th among 20 methods mentioned by more than one respondent. Turning to medical research, Emerson and Colditz (1992) catalog the statistical methods used in articles in the 1989 volume of the *New England Journal of Medicine*. Of 115 "Original Articles," 45 use t tests, 41 present contingency tables, 37 employ survival methods, and so on. Emerson and Colditz do not mention any use of Bayesian techniques.

The DOE respondents are professional statisticians working in a variety of applied areas. Medical research projects often engage trained statistical collaborators. These practitioners might be expected to employ more up-to-date methodology than those in fields that less often engage professional statisticians. Yet even here, Bayesian approaches have made very few inroads. The absence of Bayesian procedures in commercial statistical software systems is further evidence of lack of use. Although the statistical literature abounds in research papers developing and applying Bayesian ideas, these appear to be in the nature of demonstration pieces. I can find no empirical evidence of widespread use in actual practice.

Because Bayesian methods are relatively rarely used in practice, teaching them has an opportunity cost, depriving students of instruction about the standard methods that are in common use. It might be argued that we should teach the "right" methods, regardless of current practice. That is simply to invite students to go elsewhere, to turn to another of the many places on campus where basic statistics is taught. Aside from that pragmatic consideration, I believe that we have an obligation to meet the needs of our customers. Their need is to read and understand applications of statistics in their own disciplines. Those applications are not yet Bayesian.

Here is a second argument, also at least somewhat empirical. *Reading of current Bayesian literature strongly suggests that there are not yet standard Bayesian methods for standard problem settings.* There remains considerable disagreement among Bayesians on how to approach the inference settings usually considered in first courses.

Purists might insist on informative, subjective prior distributions—a position that says in effect that there can be no standard methods because there are no standard problems. Lindley (e.g., 1971) is a classical proponent of the subjective Bayesian position, but it is worth noting that arguments for the superiority of Bayesian reasoning usually start from this position. It is nonetheless now more common to concede that there is a need for standard methods to apply in standard settings. The usual approach is to use noninformative reference priors that are generated from the sampling distribution rather than from actual prior information. Not all Bayesians are comfortable with this triumph of pragmatic over normative thinking.

Let us accept the need for standard techniques, at least for beginners. Very well: which noninformative prior should we use? Berger, who favors this approach, admits (1985, p. 89) that "Perhaps the most embarrassing feature of noninformative priors, however, is simply that there are often so many of them." He offers *four* choices when θ is the probability of success in the binomial setting, and says, "All four possibilities are reasonable." Robert (1994, p. 119) presents an example due to Berger and Bernardo showing that simply reordering the parameters in the oneway ANOVA setting leads to four different reference priors.

Bayesian hypothesis testing in particular appears to be work in progress. One can find current research [e.g., Kass and Wasserman (1995)] starting from the premise that practically useful Bayesian tests remain an important open problem. The excellent survey of Kass and Raftery (1995) convinces me that Bayesian testing is not yet ready for prime time.

I believe that these negative answers to the two questions posed at the beginning of this section are relatively empirical and "objective." They are spelled out in much greater detail in Moore (1996). It seems to me that the weakest conclusion possible is that it is *premature* to make Bayesian methods the focus of basic methodological courses. I am not unalterably philosophically opposed to Bayesian statistics. I would support teaching Bayesian methods in introductory courses if they became standard methods widely accepted and widely used in practice. As Keynes once said, "When the facts change, I change my mind. What do you do, sir?" The rest of this paper therefore concerns issues that provide both secondary reasons to avoid Bayes-for-beginners at present and pedagogical challenges in a Bayesian future.

1.3 The standard choice

If we are to compare the accessibility of Bayesian reasoning with that of standard inference, it is wise to first state what the standard is. Our topic is first courses on statistical methodology for beginners who come from other disciplines and are not on the road to becoming professional statisticians. I will therefore give a partial outline of a "standard" elementary statistics course for this audience.

> **A. Data analysis**. We begin with tools and tactics for exploring data. For a single measured variable, the central idea is a *distribution* of values. We meet several tools for graphic display of distributions, and we learn to look for the shape, center, and spread of a distribution. We learn to describe center and spread numerically and to choose among competing descriptions.
>
> **B. Data production**. We now distinguish a sample from the underlying population, and statistics from parameters. We meet sta-

tistical designs for producing sample data for inference about the underlying population or process. Randomized comparative experiments and random sampling have in common the deliberate use of chance mechanisms in data production. We motivate this as avoiding bias, then study the consequences by asking "What would happen if we did this many times?" The answer is that the statistic would vary. The pattern of variation is given by a distribution, the *sampling distribution*. We can produce sampling distributions by simulation and examine their shape, center, and spread using the tools and ideas of data analysis.

C. Formal inference. We want to draw a conclusion about a parameter, a fixed number that describes the population. To do this, we use a statistic, calculated from a sample and subject to variation in repeated sampling from the same population. Standard inference acts as if the sample comes from a randomized data production design. We consistently ask the question "*What would happen if we did this many times?*" and look at sampling distributions for answers. One common type of conclusion is, "If we drew many samples, the interval calculated by this method would catch the true population mean μ in 95% of all samples." Another is, "If we drew many samples from a population for which $\mu = 60$ is true, only 1.2% of all such samples would produce an $\overline{x}$ as far from 60 as this one did. That unlikely occurrence suggests that μ is not 60."

Inductive inference is a formidable task. There is no simple path. But there may be relatively simpler and more complex paths. Consider these characteristics of the "standard" outline above:

- A parameter is a fixed unknown number that describes the population. It is different in nature from a statistic, which describes a sample and varies when we take repeated samples.

- Inference is integrated with data analysis through the idea of a distribution. The central idea of a sampling distribution can be presented via simulation and studied using the tools of data analysis.

- Probability ideas are motivated by the design of data production, which uses balanced chance mechanisms to avoid bias. The issue of sampling variability arises naturally, and leads naturally to the key question, "What would happen if we took many samples?"

- Probability has a single meaning that is concrete and empirical: "What would happen if we did this many times?" We can demonstrate how probability behaves by actually doing many trials of chance phenomena, starting with physical trials and moving to computer simulation.

- Inference consistently asks "What would happen if we did this many times?" Although we use probability language to answer this question, we require almost no formal probability theory. Answers are based on sampling distributions, a concrete representation of the results of repeated sampling.

- For more able students, study of simulation and bootstrapping is a natural extension of the "do it many times" reasoning of standard inference.

This simple outline of standard statistics can legitimately be criticized as lacking generality—standard inference is limited by acting *as if* we did proper randomized data production, for example. For beginners, however, it is clarity rather than generality that we seek.

1.4 Is Bayesian reasoning accessible?

I find that the reasoning of Bayesian inference, though purportedly more general, is considerably more opaque than the reasoning of standard inference.

- A **parameter** does describe the population, but it is a random quantity that has a distribution. In fact, it has two distributions, prior and posterior. So, for example, μ and $\overline{x}$ are both random. Yet μ is not "random" in the same sense that $\overline{x}$ is random, because a distribution for μ reflects our uncertainty, while the sampling distribution of $\overline{x}$ reflects the possibility of actually taking several samples.

- **Probability** no longer has the single empirical meaning, "What would happen if we did this many times?" Subjective probabilities are conceptually simple, but they are not empirical and don't lend themselves to simulation. Because we hesitate to describe the sampling model for the data given the parameter entirely in terms of subjective probability, we must explain several interpretations of "probability," and we commonly mix them in the same problem.

- The core reasoning of Bayesian inference depends not merely on probability but on **conditional probability**, a notoriously difficult idea. Beginners must move from the prior distribution of the parameter and the conditional distribution of the data given the parameter to the conditional distribution of the parameter given the data. Keeping track of what we are conditioning on at each step is the key to unlocking Bayesian reasoning. Thus when we consider the sampling distribution of $\overline{x}$, we think of μ as fixed because we are conditioning on μ. This allows us to deal with the practical observation that if we took another sample from the same population, we would no doubt obtain a different $\overline{x}$. Once we have sample data, we regard $\overline{x}$ as fixed and condition on its observed value in order to update the distribution of μ. This makes sense, of course, but in my experience even mathematics majors have difficulty keeping the logic straight.

Let me be clear that I am not questioning the coherence or persuasiveness of Bayesian reasoning, only its ease of access. In a future Bayesian era, we shall all have to face the task of helping students clear these hurdles. At present, they hinder an attempt to include a small dose of Bayes in a course that (for the reasons noted in Section 2) must concentrate on standard approaches when dealing with inference. In either case, we ought to recognize that they pose genuine barriers to students.

I find, to be blunt, that many Bayesian plans for teaching beginners deal with these basic conceptual difficulties by either over-sophistication or denial. Let me give an example of each.

Many Bayesians insist that they have only one kind of probability, namely, subjective probability. That strikes me as over-sophisticated. Physical and personal probabilities are conceptually quite different. There is a clear practical distinction between a prior distribution that expresses our uncertainty about the value of μ and the sampling distribution that expresses the fact that if we took another sample we would observe a different value of $\overline{x}$. Saying that "do it many times" is just one route to subjective probability is intellectually convincing to sophisticates but doesn't deal with the beginner's difficulty. What is more, the priors used in our examples are often "noninformative" priors that attempt to represent a state of ignorance. So some prior distributions represent prior knowledge, while others, mathematically similar in kind and appearing to give equally detailed descriptions of the possible values of the parameter, represent ignorance. There are sophisticated ways to explain that "ignorance" means ignorance relative to the information supplied by the data—but we are discussing the accessibility of core ideas to unsophisticated students.

Here is a passage in which Albert (1996b) seems to me to deal with the conceptual difficulties of conditional probabilities by denial:

> *Although one can be sympathetic with the difficulty of learning probability, it is unclear what is communicated about classical statistical inference if the student has only a modest knowledge of probability. If the student does not understand conditional probability, then how can she understand the computation of a p-value, which is a tail probability conditional on a particular hypothesis being true? What is the meaning of a sampling distribution if the student does not understand what model is being conditioned on?*

The concluding rhetorical question in this passage seems based on the view that all probabilities are conditional probabilities. That is over-sophisticated, but it also denies that there is a difference in difficulty between these questions:

A. Toss a balanced coin 10 times. What is the probability that exactly 4 heads come up?

B. Toss a balanced coin 10 times. Someone tells you that there were 2 or more heads. Given this information, what is the probability that exactly 4 heads come up?

That there is a "model being conditioned on" in Question A is language that a teacher not bound by precise Bayes-speak would avoid when addressing beginners. Question B, which involves *conditioning on an observed event*, is conceptually more complex than Question A, which does not. Speaking only of the formalities of probability, as opposed to its interpretation, students can grasp standard inference via probability at the level of Question A. Bayesian inference requires conditional probability of the kind needed for Question B. Why attempt to deny that the second path is harder?

Albert's point that P-values require conditioning displays a similar denial of the distinction between Questions A and B. To explain a P-value, we begin by saying, "Suppose for the sake of argument that the null hypothesis is true." To the Bayesian sophisticate, this is conditioning. To a teacher of beginners, however, it is like saying "Suppose that a coin is balanced." The supposition is the start of the reasoning we want our students to grasp: a result this extreme would be very unlikely to occur if H_0 were true; such a result is therefore evidence that H_0 is not true. This reasoning isn't easy (there is no easy road to inference), but it does not involve conditioning on an observed event. A grasp of P-values does not require, and is not much aided by, a systematic presentation of conditional probability. On the other hand, conditioning is so central to Bayesian reasoning that we must discuss it explicitly and very carefully. The distinction between Question A and Question B captures the distinction between what we must teach as background to standard (A) and Bayesian (B) inference.

There are numerous other complexities that the teacher of Bayesian methods must face (or choose to ignore). The use of default or reference priors opens a gap between Bayesian principle and Bayesian practice that is not easy to explain to beginners. The use of improper priors may sow puzzlement. The need to abandon what seemed satisfactory priors when we move from estimation to testing is annoying. And so on. I want, however, to concentrate on what appears to be the primary barrier to beginners' understanding of the core reasoning of Bayesian inference: the greater dependence on probability, especially conditional probability.

1.5 Probability and its discontents

Albert, in the passage I just cited, speaks for many teachers, Bayesian and non-Bayesian alike, when he says, "it is unclear what is communicated about classical statistical inference if the student has only a modest knowledge of probability." It is, unfortunately, a well-documented fact that the great majority of our beginning students *will*, despite our best efforts, have only a modest knowledge of probability.

Psychologists have been interested in our perception of randomness ever since the famous studies of Tversky and Kahneman [e.g., Tversky and Kahneman, 1983]. Bar-Hillel and Wagenaar (1993) offer a recent survey. Much recent work has criticized Tversky and Kahneman's attempt to discover "heuristics" that help explain why our intuition of randomness is so poor and our reasoning about chance events is so faulty. That our intuition about random behavior *is* gravely defective, however, is a demonstrated fact. To give only a single example, most people accept an incorrect "law of small numbers" [so named by Tversky and Kahneman (1971)] that asserts that even short runs of random outcomes should show the kind of regularity that the laws of probability describe. When short runs are markedly irregular, we tend to seek some causal explanation rather than accepting the results of chance variation. As Tversky and Kahneman (1983, p. 313) say in summary, "intuitive judgments of all relevant marginal, conjunctive, and conditional probabilities are not likely to be coherent, that is, to satisfy the constraints of probability theory."

That people's intuitions about chance behavior are systematically faulty has implications for statistical education. Our students do not come to us as empty vessels into which we pour knowledge. They combine what we tell them with their existing knowledge and conceptions to construct a new state of knowledge, a process that Bayesians should find natural. The psychologists inform us that our students' existing conceptions of chance behavior are systematically defective: they do not conform to the laws of probability or to the actually observed behavior of chance phenomena. At this point, researchers on teaching and learning become interested; the teaching and learning of statistics and probability has been a hot field in mathematics education research for more than a decade. Psychologists attempt to describe how people think. Education research looks at the effects of our intervention (teaching) on students' thinking. Results of this research are summarized in Garfield (1995), Garfield and Ahlgren (1988), Kapadia and Borovcnik (1991) and Shaughnessy (1992).

The consensus of education research is, if anything, more discouraging than the findings of the psychologists. Even detailed study of formal probability, so that students can solve many formally posed problems, does little to correct students' misconceptions and so does little to equip them to use probabilistic reasoning flexibly in settings that are new to them. Garfield and Ahlgren (1988) conclude that "teaching a conceptual grasp of probability still appears to be a very difficult task, fraught with ambiguity and illusion." Some researchers [see Shaughnessy, pp. 481–483] have been able to change the misconceptions of some (by no means all) students by activities in which students must write down their predictions about outcomes of a random apparatus, then actually carry out many repetitions and explicitly compare the experimental results with their predictions. Some of these same researchers go so far as to claim that "not only is traditional instruction insufficient, it may even have negative effects on students' understanding of stochastics." Shaughnessy (p. 484) also cites "strong evidence for the superiority of simulation methods over analytic methods in a course on probability." He stresses that changing ingrained misconceptions cannot be done quickly, but requires sustained efforts.

Research in education and psychology appears to confirm that conditional probabilities are particularly susceptible to misunderstanding. Garfield and Ahlgren (1988, p. 55) note that students find conditional probability confusing because "an important factor in misjudgment is misperception of the question being asked." Students find it very difficult to distinguish among $P(A|B)$, $P(A \text{ and } B)$, and $P(B|A)$ in plain-language settings. Shaughnessy (1992, pp. 473–476) discusses the difficulties associated with conditional probabilities at greater length. He agrees that "difficulties in selecting the event to be the conditioning event can lead to misconceptions of conditional probabilities." He also points to empirical studies suggesting that students may confuse conditioning with causality, are very reluctant to accept a "later" event as conditioning an "earlier" event, and are easily confused by apparently minor variations in the wording of conditional probability problems. The "Monty Hall problem" (that goat behind a door—a job for Bayes' theorem) and its kin remind us that it is conditional probability problems that so often give probability its air of infuriating unintuitive cleverness.

1.6 Barriers to Bayesian understanding

It appears that we must accept these facts as describing the environment in which we must teach: beginning students find probability difficult; they find conditional probability particularly difficult; there is as yet no known way to relieve their difficulties that does not involve extensive hands-on activity and/or simulation over an extended time period. I believe that our experience as teachers generally conforms to these findings of systematic study.

The unusual difficulty of probability ideas, and the inability of study of formal probability theory to clarify these ideas for students, argue against a mathematically-based approach in teaching beginners. *Mathematical understanding is not the only kind of understanding.* Mathematical language helps *us* to formulate, relate, and apply statistical ideas, but it does not help our students nearly as much as we imagine. Recognition of the futility of a formal approach has been one factor in moving beginning statistics courses away from the traditional probability-and-inference style toward a data analysis-data production-concepts of inference-methods of inference style that pays more attention to data. I believe that the findings I have cited also point to very substantial difficulties that stand in the way of effective Bayes-for-beginners instruction:

- If our intuition of chance is systematically incoherent, is it wise to rely on subjective probability as a central idea in a first statistics course?
- If the only known ways of changing misconceptions about chance behavior involve confronting misconceptions via physical chance devices and simulation—that is, by asking "What would happen if we did this many times?"—ought we not to make the answer to that question our primary interpretation of probability?
- If we want to help students see that the laws of probability describe chance outcomes *only* in the long run—the law of large numbers is true, but the law of small numbers is false—how can we avoid confusion if our central notion of probability applies to even one-time events?
- If teaching correct probability is so difficult and requires such intensive work, ought we not to follow Garfield and Ahlgren (1988) in asking "how useful ideas of statistical inference can be taught independently of technically correct probability?" Ought we not at least seek to minimize the number of probability ideas required, in order to leave time for data-oriented statistics?
- If conditional probability is known to be particularly difficult, should we not hesitate to make it a central facet of introductory statistics?

Statistical inference is not conceptually simple. Standard and Bayesian inference each require a hard idea—sampling distributions for standard inference, and conditional probability and updating via Bayes' rule for Bayesian inference. I claim—not only as a personal opinion, but as a reasonable conclusion from the research cited above—that the Bayesian idea is markedly more difficult for beginners to comprehend. Sampling distributions fit the "activity and simulation" mode recommended by most experts on learning, and the absence of conditioning makes them relatively accessible both conceptually and to simulation. Once students have grasped the "What would happen if we did this many times?" method, we can hope that they will also grasp the main ideas of standard inference. Bayes-for-beginners, on the other hand, must either shortchange the reasoning of inference or use two-way tables to very carefully introduce conditional probability and Bayes' theorem. Albert (1995) and Rossman and Short (1995) illustrate the care that is required.

Although I advocate a quite informal style in presenting statistical ideas to beginners, I recognize that different teachers prefer different levels of formality. I claim that at any level of formality,

> "State a prior distribution for the parameter and the conditional distribution of the data given the parameter. Update to obtain the conditional distribution of the parameter given the data."

is less accessible core reasoning than

> "What would happen if we did this many times?"

1.7 Are Bayesian conclusions clear?

We have seen that the need to understand probability, especially conditional probability, at a relatively profound level, is a barrier to understanding the reasoning of Bayesian inference. I believe that the same barrier stands in the way of understanding the results of inference. That is, Bayesian conclusions are perhaps not as clear to beginners as Bayesians claim.

It is certainly true, as Bayesians always point out, that users do not speak precisely in stating their understanding of the conclusions of standard inference. They often confuse probability statements about the *method* (standard inference) and probability statements about the *conclusion* (Bayesian inference). "I got this answer by a method that gives a correct answer 95% of the time" easily slides into "The probability that this answer is correct is 95%." If we regard this semantic confusion as important, we ought to ask whether the user of Bayes methods can explain without similar confusion what she means by "probability." The Bayesian conclusion is easy to state, but hard to explain. What is this "probability 95%"? Physical and personal probabilities are conceptually quite different, and the user of Bayes methods must be aware that probabilities are conditioned on different events at each stage. That a user gives a semantically correct Bayesian conclusion is not evidence that she understands that conclusion.

1.8 What spirit for elementary statistics instruction?

The continuing revolution in computing has changed the professional practice of statistics and our judgment of what constitutes interesting research in statistics. These changes are in turn changing the teaching of elementary statistics. We are moving from an over-emphasis on probability and formal inference toward a balanced presentation of data analysis, data production, and inference. See the report of the ASA/MAA joint curriculum committee [Cobb (1992)] for a clear statement of these trends. In particular, there is a consensus that introductions to statistics ought to involve constant interaction with real data in real problem settings. Real problem settings often have vaguely defined goals and require the exercise of judgment. The spirit of contemporary introductions to statistical practice is very different from the spirit of traditional "probability and statistics" courses.

Bayesian thinking fits uneasily with these trends. Exploratory data analysis allows the data to speak, and diagnostic procedures allow the data to criticize proposed models; Bayesians tend to say "no adhockery" and to start from models and structured outcomes rather than from data. Good designs for data production avoid disasters (voluntary response, confounding) and validate textbook models; many Bayesians question at least the role of randomization and sometimes the role of proper sample/experimental design in general. And the high opportunity cost of teaching conditional probability and Bayes' theorem in more-than-rote fashion forces cuts elsewhere.

Concentrating on Bayesian thinking is not in principle incompatible with data-oriented instruction. In practice, however, it is likely to turn elementary statistics courses back toward the older mode of concentrating on the parts of our discipline that can be reduced to mathematics. Avoiding this unfortunate retrogression is perhaps the most serious pedagogical challenge facing Bayes-for-beginners.

Acknowledgments

I am grateful to James Berger for discussions about Bayesian viewpoints, and to George Cobb and Paul Velleman for helpful comments on a first version of this paper.

References

Albert, J. A. (1995). Teaching inference about proportions using Bayes and discrete models, *Journal of Statistics Education* (electronic journal), **3**, no. 3.

Albert, J. A. (1996a). *Bayesian Computation Using Minitab*, Belmont, Ca.: Duxbury.

Albert, J. A. (1997). Teaching Bayes' rule: a data oriented approach. *The American Statistician*, to appear.

Antleman, G. (1996). *Elementary Bayesian Statistics*, Cheltenham, England: Edward Elgar Publishing.

Bar-Hillel, M. and Wagenaar, W. A. (1993). The perception of randomness, In *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues* (Eds., G. Keren and C. Lewis), pp. 369–393, Hillsdale, N.J.: Erlbaum Associates.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, New York: Springer.

Berry, D. A. (1996). *Statistics: A Bayesian Perspective*, Belmont, Ca.: Duxbury.

Berry, D. A. (1997). Teaching introductory Bayesian statistics with real applications in science: doing what comes naturally, *The American Statistician*, to appear.

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*, Reading, Ma.: Addison-Wesley.

Cobb, G. (1992). Teaching statistics, In *Heeding the Call for Change: Suggestions for Curricular Action* (Ed., L. A. Steen), pp. 3–43, Washington, D.C.: Mathematical Association of America.

Dickey, J. M. and Eaton, M. L. (1996). Review of J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*, *Journal of the American Statistical Association*, **91**, 906–907.

Emerson, J. D. and Colditz, G. A. (1992). Use of statistical analysis in the *New England Journal of Medicine*, In *Medical Uses of Statistics*, 2nd ed. (Eds., J. C. Bailar and F. Mosteller), pp. 45–57, Boston: NEJM Press.

Garfield, J. (1995). How students learn statistics. *International Statistical Review*, **63**, 25–34.

Garfield, J. and Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: implications for research, *Journal for Research in Mathematics Education*, **19**, 44–63.

Kapadia, R. and Borovcnik, M. (Eds.) (1991). *Chance Encounters: Probability in Education*, Dordrecht: Kluwer.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *Journal of the American Statistical Association*, **90**, 773–795.

Kass, R. E. and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion, *Journal of the American Statistical Association*, **90**, 928–934.

Lindley, D. V. (1971). *Bayesian Statistics: A Review*, Philadelphia: SIAM.

Moore, D. S. (1997). Bayes for beginners? Some reasons to hesitate, *The American Statistician*, to appear.

Robert, C. (1994). *The Bayesian Choice: A Decision-Theoretic Motivation*, New York: Springer.

Rossman, A. J. and Short, T. H. (1995). Conditional probability and education reform: Are they compatible? *Journal of Statistics Education* (electronic journal), **3**, number 2.

Rustagi, J. S. and Wright, T. (1995). Employers' contributions to the training of professional statisticians, *Bulletin of the International Statistical Institute, Proceedings of the 50th Session*, LVI, Book 1, 141–160.

Shaughnessy, J. M. (1992). Research in probability and statistics: reflections and directions, In *Handbook on Research on Mathematics Teaching and Learning* (Ed., D. A. Grouws), pp. 465–494, New York: Macmillan.

Tversky, A. and Kahneman, D. (1971). Belief in the law of small numbers, *Psychological Bulletin*, **76**, 105–110.

Tversky, A. and Kahneman, D. (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment, *Psychological Review*, **90**, 293–315.