Bayes for Beginners? Some Reasons to Hesitate

David S. Moore

The American Statistician, 51 (1997), 254–261

Is it reasonable to teach the ideas and methods of Bayesian inference in a first statistics course for general students? This paper argues that it is at best premature to do so. Surveys of the statistical methods actually in use suggest that Bayesian techniques are little used. Moreover, Bayesians have not yet agreed on standard approaches to standard problem settings. Bayesian reasoning requires a grasp of conditional probability, a concept confusing to beginners. Finally, an emphasis on Bayesian inference might well impede the trend toward experience with real data and a better balance between data analysis, data production, and inference in first statistics courses.

KEY WORDS: Bayesian methods; Statistical education.

David S. Moore is Shanti S. Gupta Distinguished Professor, Department of Statistics, Purdue University, West Lafayette, IN 47907–1399. 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.

1. INTRODUCTION

The issue I wish to consider is whether ideas and methods of Bayesian inference should be presented in a first statistics course for general students who must later read and perhaps employ statistics in their own disciplines.

This is a quite specific question. I agree that Bayesian methods are increasingly important and should form part of the training of professional statisticians. I also agree that a course for liberal arts students on, e.g., "risks and decisions" structured around subjective Bayesian ideas can be stimulating, particularly if it incorporates discussion of the conflict between probabilistic assessment of risks and societal response (e.g., Slovic, Fischhoff, and Lichtenstein (1982), Zeckhauser and Viscusi (1990)). These are quite different settings from that I have in mind, one more specialized and the other less constrained by student needs.

I do not wish to join the foundational debate over whether Bayesian inference is in some sense uniformly preferable to standard inference. That prejudges our question, forcing us to teach what is "right" regardless of customer needs or pedagogical barriers. Most statisticians remain eclectic, willing to employ Bayesian methods where appropriate but unconvinced by universalist claims. That being the case, I am unwilling to settle the content-of-instruction issue by appeals to one or another side in a debate among professionals.

There are, I think, some compelling reasons to hesitate to present Bayesian ideas in a first course in working statistics:

1. Bayesian methods are relatively rarely used in practice. Teaching them has an opportunity cost, depriving students of instruction about methods that are in common use.

2. It is unclear what Bayesian methods we should teach. Those who advocate them have not yet agreed on standard approaches to standard problems. And of course, lacking standard methods, we also lack standard software for implementing them.

3. A conceptual grasp of Bayesian methods rests on an understanding of conditional probability, a notoriously difficult idea. Although Bayesian conclusions are simple in form, the simplicity disappears when we ask "What do you mean by probability?" There is of course no easy path to understanding inference. Nonetheless, the fact that standard inference consistently asks "What would happen if we did this many times?" and answers by displaying a sampling distribution makes standard reasoning more accessible.

4. Inference is only part of statistics in practice. Data analysis and the design of data production were always important and have become more so in the past generation. They have recently begun to receive more adequate attention in instruction for beginners. For various reasons (emphasis on inference, ambiguity about the role of designed data production, dependence on conditional probability), Bayes-for-beginners tends to impede the trend toward greater emphasis on developing students' data sense. I like the trend and don't want to impede it.

The first two of these arguments are relatively "objective," resting on recent surveys of actual use of statistical methods in several fields (Section 2) and on browsing in Bayesian texts and research literature (Section 3). They are the primary focus of this paper. I also consider these arguments conclusive. It seems to me that the weakest conclusion possible is that it is *premature* to make Bayesian methods the focus of basic methodological courses. They haven't been sufficiently widely accepted by users, or even been sufficiently routinized by experts. That situation may change in the future. In the present, academic researchers ought not to impose a still-primitive version of how we think things should in principle be done.

Advocates of Bayes-for-beginners should ponder the similar intellectual case for basing our first courses on modern computer-intensive methods (bootstrapping, "nonparametric" fitting, robust inference). This is also an active area of research, and is also finding increasing application. Some statisticians find the empirical and exploratory spirit of these approaches more congenial than the decision orientation of Bayesian thinking. Yet it would also be premature to abandon the teaching of t procedures in favor of neural nets. We have an obligation to consider and to meet the needs of our customers. If we do not, they will simply go elsewhere.

The third argument enumerated above can be grounded in research in psychology and education about the difficulties attending probability concepts, but it is nonetheless more disputable. I briefly sketch the grounds for my opinion in Section 4; much greater detail appears in Moore (1997). My fourth point concerns the overall spirit and emphasis of our introductory courses rather than the species of formal inference that we teach. It is the point about which I feel most strongly. If, in a future Bayesian era, Bayes methods become standardized and widely used in routine statistical practice, I will happily teach them despite the pedagogical barriers noted in Section 4. I trust that even in such an era, statisticians will continue to insist that beginning students gain hands-on experience with data and understand core statistical ideas (such as the distinction between observation and experiment) that lie outside the domain of formal inference.

2. ARE BAYESIAN METHODS USED IN PRACTICE?

Our beginning students come to us from other fields of study. They come because their own fields employ statistical ideas and methods. Their first need is to be able to read literature in their own field. We ought to be attentive to our customers' expressed needs, rather than offer them what we imagine they ought to need. What statistics do knowledgeable practitioners in various fields apply? Let us search for data. There follows, in roughly descending order of statistical sophistication, a survey of recent empirical studies on the use of statistics in practice.

Professional statisticians employed in national laboratories may be thought to employ up-to-date and effective methodology. Rustagi and Wright (1995) carried out a census of statisticians employed in Department of Energy National Laboratories. Remarkably, they obtained responses from all 103 members of this population, 100 of whom hold a master's or doctorate degree. These statisticians work in a variety of applied fields, with some emphasis on physical sciences and environmental problems. Table 1 records their responses to a request to choose from a long list "the three statistical techniques that have been most important to your work/research." Only four of this sophisticated group mentioned Bayesian methods, although 37 reported training in these methods during their university careers. Techniques typical of those taught in a two-course sequence on statistical methods are most common. Other frequently mentioned topics (probability modeling, simulation, quality control, reliability) reflect the nature of the National Laboratories' work. When asked to choose the three most important techniques that were *not* part of their academic training, 19 named Bayesian methods—sixth place behind quality control, reliability, simulation, exploratory data analysis, and graphical display.

Medical research is a major and quite sophisticated consumer of statistical analyses. The surveys by Altman (1991) and Emerson and Colditz (1992) document the nature and growth of the use of statistics in medical journals in roughly the decade of the 1980s. In particular, Emerson and Colditz inventory the methods employed in the 115 "Original Articles" appearing in the *New England Journal of Medicine* in 1989, and Altman does the same for the 100 articles appearing in 1990. Table 2 presents some of the findings of Emerson and Colditz. Techniques and ideas from a standard first course predominate—Emerson and Colditz say (referring to a longer time period) that acquaintance with descriptive statistics, t procedures, and contingency tables would give "full access" to 73% of the articles. They identify increasing use of ANOVA, multiple regression, and survival analysis as notable recent trends. Altman points to meta-analysis, new techniques for design of clinical trials, and editorials in several medical journals encouraging more frequent use of confidence intervals. Neither paper contains any mention of Bayesian approaches. Some use could lie hidden in the "other methods" category of their tables, though Emerson and Colditz enumerate several of the "other methods" without using the word Bayes. Similar results appear in older surveys such as Hokanson, Luttman, and Weiss (1986).

In *psychiatry*, Everitt (1987) reprints a table from DeGroot and Mezzich (1985) that surveys 597 papers in the 1980 volumes of three major journals. Of these, 156 were surveys or contained no data. The remaining 441 use statistical methods to some degree. *One* paper among these 441 employed Bayesian methods. Everitt mentions more recent trends toward use of clustering, logistic regression, structural equation models, and Cox regression. Dunn et al. (1993), focusing on depression, point to some of the same innovations along with meta-analysis and an emphasis on controlled clinical trials. Neither they nor any of the discussants to Everitt's survey mention Bayesian approaches as either in use or promising. Generally similar conclusions appear in Hokanson et al. (1986). Nieminen (1995) compares the statistical methods employed in two types of papers in psychiatry (research on therapeutic communities and on psychiatric wards) across a large universe of international journals in the years 1987–1992. Comparison of means (t tests and ANOVA), cross-tabulations (often with the Pearson chi-square test), correlation, and "reliability analysis" (agreement of raters, intraclass correlation, etc.) are the most common techniques. Bayesian methods again go unmentioned.

Emulating Emerson and Colditz, Hammer and Buffington (1994) survey all articles published in 1992 in six *veterinary medicine* journals. About half contained statistics beyond simple numerical descriptions. The authors summarize: "Knowledge of 5 categories of statistical methods (ANOVA, t-tests, contingency tables, nonparametric tests, and simple linear regression) permitted access to 90% of the veterinary literature surveyed. These data may be useful when modifying the veterinary curriculum to reflect current statistical usage." Multiple regression, epidemiological methods, confidence intervals, and survival analysis fill out the authors' list. Bayesian methods are not mentioned.

In all the fields of work covered by these surveys, statistical methods taught in a standard first course go far toward giving access to the research literature. More important, the reasoning of tests and confidence intervals is essential to understanding research reports in all of these fields. The more specialized methods that are most often employed, and the newer techniques that are considered promising, vary with the area of application. As we move away from the areas that most often involve professional statisticians in their work, the statistical methods employed become more traditional. It would certainly be desirable to have comparable data for other fields, especially the social and behavioral sciences. Nonetheless, the available data suggest that *Bayesian methods are rarely used, relative to the methods of standard inference, in any field to which statistics is applied.*

Paul Velleman points out that the absence of Bayesian procedures in commercial statistical software is more global evidence of lack of use, as these packages respond quickly to customer demand. Velleman says, "Both the features and the advertising of software packages offer a good measure of what people who actually analyze data really want." The most recent new version announcement I have seen (as of July, 1996) is Minitab Release 11. Minitab claims to have added logistic regression, reliability/survival analysis, polynomial regression, gage R&R, and correspondence analysis. Several of these techniques appear in the lists I have cited. Minitab appears to find more demand for even gage R&R and correspondence analysis than for Bayesian procedures.

The fact that even sophisticated users rarely turn to Bayesian methods deserves attention. Statistical research journals are full of papers advancing and applying Bayesian ideas. Statisticians—especially Bayesians—therefore imagine that use of these ideas in practice is advancing rapidly. I can find no empirical evidence that this is true. The research papers are in the nature of demonstration pieces that suggest the possibilities of Bayesian analyses. These analyses have not yet passed the "Box test" that assesses the usefulness of a method by whether it is actually used.

This survey of empirical evidence concerning use in practice raises another issue beyond the question of why we should teach beginners an approach that appeals to us in principle but lies unused in practice. *Teaching Bayes to beginners has an opportunity cost.* If Bayesian ideas displace t procedures, contingency tables, regression, or ANOVA, they bar students from access to much literature in any field that applies statistics. If we manage to add Bayes to the list, perhaps by persuading students to elect further courses, we must ask whether the time might better be spent on logistic regression, simulation, survival analysis, or meta-analysis. Bayesian methods are not the only hot field in statistics research, and several others have already passed the Box test.

3. ARE THERE STANDARD BAYESIAN METHODS?

The content of a first course that aims to provide understanding and useful tools to students from other disciplines should, I think, consist mainly of standard material well-accepted by the profession and widely used in practice. The first section argued that Bayesian methods are not widely used. This section will suggest that even if we listen only to Bayesians, standard methods for standard problems are not yet agreed upon.

Much of the disagreement concerns the essential element distinguishing Bayesian from standard models, namely prior distributions for unknown parameters. There is a continuing tension in Bayesian circles between use of priors that try to reflect the actual partial knowledge of a decision-maker and "reference priors" that are automatically generated from the sampling distribution, taking no account of what partial knowledge may exist. (I adopt the terminology of Kass and Wasserman (1996) in calling priors generated by formal rules "reference priors.") The former class of priors are "informative;" reference priors are generally chosen to be "noninformative." The use of conjugate priors, which specify a parametric family of prior distributions on the basis of analytic convenience, but allow prior knowledge to choose the parameters of this family, lies between these extremes. Because knowledge of the mechanism that generates the unknown parameter is rarely complete enough to determine a prior distribution, informative priors are usually subjective. Simplifying a bit, we can imagine several Bayes-for-beginners approaches.

> **A. Purist Bayes** Relevant prior information is always available and should be expressed in an informative (usually subjective) prior distribution tailored to the problem at hand. This view is consistent, easy to explain, and in many settings intellectually attractive. Expositions of the advantages of Bayesian analysis emphasize the use of genuine prior information and the subjective interpretation of probability. Alas, there can then be no standard Bayesian analysis for standard problems, because every problem is potentially unique. We are left with a tool of great power in non-standard problems, but which is unlikely to ever be widely used in standard settings. Beginners come away with ideas but few usable tools.
>
> **B. Accessible Bayes** Emphasize the Bayesian Big Idea: express prior information in an informative prior distribution, use data to update this information to form a posterior distribution, base all inference on the posterior distribution. Restrict the settings considered to those in which beginners can implement the Big Idea, mainly discrete or conjugate priors. In a binomial problem, for example, the prior information just happens to be expressed by a beta distribution.

Emphasize estimation rather than testing, and rejoice that the one-way ANOVA setting is beyond the scope of the course. We can teach a beginner-friendly course that does provide usable tools for simple settings. However, the tools taught may not reflect actual Bayesian practice (let alone prevailing statistical practice). As Robert (1994, p. 98) notes, "the use of conjugate priors is strongly suspicious for most Bayesians since it is mainly justified on technical grounds rather than for fitting properly the available prior information."

C. Auto-Bayes In practice we do need standard methods for standard problems. We can get them by employing reference priors that are determined by the sampling distribution. Reference priors are generally noninformative. In effect, we first present beginners with an explanation of the role of prior information and perhaps even of the machinery for making use of it in the simplest cases. When we come to practical settings, however, we tell our students to ignore prior information. If our students are a bit sophisticated, we may explicitly argue (Box and Tiao, 1973, p. 2) that "In problems of scientific inference we would usually, were it possible, like the data 'to speak for themselves.' Consequently, it is usually appropriate to conduct the analysis as if a state of relative ignorance existed *a priori*."

There is no Bayesian consensus on the relative place of purist, accessible, and automated methods for dealing with specific standard settings. It was once common (e.g., Lindley 1971) to begin with axioms for coherent inference, show that these imply the existence of a subjective prior distribution, and insist that use of these subjective priors is essential to the Bayesian approach. Many Bayesians now criticize this purist stance (e.g., Berger 1985, pp. 198–199). Current opinion among Bayesians seems rather to favor some version of reference priors for common statistical settings. But Lindley (1971, pp. 71) is not alone in his criticism of "ready-made Bayesian analyses in which θ is just a parameter." Eaton and Dickey (1996, p. 906) exemplify the continuing debate over the role of "normative" arguments in Bayesian inference in commenting on "...the outstanding gaping hole in Bayesian theory as it exists today—the absence of a normative motivation for responding to new information by the use of probability conditioning, and hence the lack of any formal justification for statistical inference by means of Bayes' theorem."

For practical and pedagogical reasons, Bayes for beginners will almost surely employ some mix of conjugate and reference priors, focusing on procedures that are computationally feasible. This also reflects a feeling that teaching the Bayesian Big Idea is more important than teaching methods we would actually recommend in practice. Even assuming this, there remain many issues on which a Bayesian consensus has yet to emerge.

- There is no agreement as to which noninformative priors we should use and teach. "Perhaps the most embarrassing feature of noninformative priors, however, is simply that there are often so many of them." (Berger 1985, p. 89) Berger offers *four* choices when θ is the probability of success in the binomial setting, and says, "All four possibilities are reasonable." See Robert (1994, p. 119) for an example due to Berger and Bernardo showing that simply reordering the parameters in the oneway ANOVA setting leads to four different reference priors (all of them too messy for beginners to grasp, I might add).

- Noninformative prior distributions are generally improper when the parameter space is not compact. Shall we expose beginners to improper priors? Robert (1994, p. 112) notes that "...some statisticians object to the use of improper priors ...Such misgivings are not really justified since it is actually possible to work with improper priors, as long as we do not regard them as probability distributions." This may not seem promising material for a first course for general students. Berry (1996, p. 339) wisely lets pedagogical good sense prevail, saying only that the prior for a normal mean m is "flat over a substantial region of m-values." He also discusses normal (conjugate) priors for the mean m. He does not mention Jeffrey's advocacy of Cauchy priors.

- Yet other issues lurk beneath the surface, though we may choose to ignore them in a first course. Some types of noninformative priors depend on the choice of parametrization. We will hide this rather than admit that choosing between probability of success and odds ratio to parametrize a binomial setting changes our automated inference. Once we have made our choice of prior and obtained the posterior distribution, a loss function or utility function usually enables us to complete our inference. We may assume (silently) squared error and 0/1 loss functions for estimation and testing. Or we may simply give posterior distributions and comment informally on what actions they suggest.

Kass and Wasserman (1996) provide a thorough survey of proposals for the use of reference priors chosen by formal rules. To quote their preface, "We conclude that the problems raised by research on priors chosen by formal rules are serious and may not be dismissed lightly." Their exposition makes it clear that choice of reference priors is still work in progress.

The complexities of Bayesian *hypothesis testing* deserve separate mention. The gap between estimation and testing is wider for Bayesian than for standard inference. "In frequentist theory, estimation and testing are complementary, but in the Bayesian approach, the problems are completely different." (Kass and Raftery, 1995, p. 781) Bayesians must generally switch priors when moving from estimation to testing, because the continuous priors used for estimation problems put probability zero on a point null hypothesis.

The results of Bayes tests seem to be more sensitive to the choice of prior distribution than is the case for Bayes estimates. Again unlike the situation for estimation, the results of Bayes tests often diverge dramatically from the results of standard tests of the same hypotheses. A major strand of Bayesian literature demonstrates the divergence between P-values and posterior probabilities when the prior distribution places any point mass on H_0 and is diffuse on the space of alternatives. Berger and Delampady (1987) give many references and a good discussion. Interpretation of this conflict varies: Berger and Delampady (p. 319) feel that P-values as usually interpreted have a "hidden and extreme bias," but even some Bayesians (e.g., Robert, 1994, p. 189) can be found suggesting that many Bayes tests "are usually quite biased in favor of the alternative hypothesis." I rather like Dempster's (1971) suggestion that

Bayesian "predictive" concepts of probability are suitable for estimation, whereas frequentist "postdictive" probability better fits testing. Like many other variations of Bayesian thinking, Dempster's suggestion has not gained the general acceptance of Bayesians.

Even putting aside these issues, the choice of Bayesian testing methods is both quite complex and not at all settled. See Kass and Wasserman (1995) as a demonstration that Bayesian approaches to standard testing problems remain a research issue. These authors say in their preface, "For estimation problems, reference priors are often 'flat' ... but in testing such a prescription leads to serious difficulties. Thus an important problem is to define a reference Bayesian testing procedure that uses a proper prior ..." Robert (1994, p. 187) agrees with Kass and Wasserman that the current emphasis on reference priors leaves testing in an unsatisfactory state: "The recourse to noninformative prior distributions for testing hypotheses is rather limited, if not simply discouraged ..." Kass and Raftery (1995) provide a review of Bayesian hypothesis testing. They convince me that this topic is not yet ready for the general public.

It is, of course, possible to take definite positions on these issues. The difficulty is that no one set of positions is widely accepted by Bayesians. The most recent surveys of research on these issues remind us of their unsettled state. Even if we abandon the hope of contact with standard statistical usage, there is as yet no "Bayesian standard" to replace it.

4. IS BAYESIAN REASONING ACCESSIBLE?

The findings of the two previous sections give, I think, ample reason to avoid basing our first courses on Bayesian ideas. Let me now briefly raise an additional issue, based on pedagogical concerns. Bayesians generally argue that the *conclusions* of Bayesian inferences are clearer than those of standard inference. I suggest that the *reasoning* of Bayesian inference is considerably more difficult for naive beginners to grasp.

The chief barrier is, as usual in first statistics courses, probability. Both standard and Bayesian inference are based on probability, and both can be presented with greater or lesser degrees of formal probability. I believe that at any level of informality, Bayesian reasoning requires both a more complex notion of probability and more probabilistic machinery than standard inference.

The unusual difficulty of probabilistic ideas has been documented both by psychologists, who investigate how people think about chance, and by education researchers, who study the effects of our intervention (teaching) on students' thinking. The best known work by psychologists is that of Tversky and Kahneman (e.g., 1983); see Bar-Hillel and Wagenaar (1993) for a recent survey. Tversky and Kahneman (1983, p. 313) show that, "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." They also dispute Lindley's claim that coherent personal probabilities can be elicited: "we suspect that incoherence is more than skin deep."

Research bearing on the teaching and learning of statistics and probability is summarized in Garfield (1995), Garfield and Ahlgren (1988), Kapadia and Borovcnik (1991) and Shaughnessy (1992). This literature establishes: (1) Our intuition about chance is defective, and this is a major barrier to effective teaching of probability and statistics. (2) It has proved hard to correct students' misconceptions about probability; limited success has been reported only from quite extensive programs based on hands-on experience with chance experiments and simulation. (3) Conditional probability is a particularly confusing idea. Shaughnessy (1992, pp. 473–476) reviews at some length the difficulties that conditional probability poses for learners.

Against this background, consider these contrasts between standard and Bayesian inference, both restricted to the level of instruction for beginners.

In standard inference, a *parameter* (say, μ) is a fixed number that describes the population. It is different in kind from a *statistic* (say, $\overline{x}$) that describes a sample and is random because it varies in repeated sampling. In Bayesian inference, μ and $\overline{x}$ are both random; in fact, μ has two distributions, prior and posterior. Yet, to a beginner, μ and $\overline{x}$ are random in different senses, because the randomness of μ expresses our uncertainty, but the randomness of $\overline{x}$ reflects the possibility of repeated sampling.

Probability has a single meaning in standard inference. It answers the question "What would happen if we did this many times?" Probability is physical and empirical; we can explore it by hands-on repetition of chance phenomena and by simulation—the only means known to improve students' grasp of chance behavior. Bayesians must deal with subjective probability (hard to simulate), account for the fact that subjective probabilities often diverge from actual long-term patterns of chance phenomena, and explain the senses in which μ and $\overline{x}$ are both random.

Each style of inference involves a hard idea—there is no easy road to inference. Students of standard inference must grasp *sampling distributions*. We help them by simulations, by using the familiar tools of data analysis (histograms; look for shape, center, and spread), and by reminding them that inference is based on asking, "What would happen if we did this many times?" Bayesian inference requires *conditional probability and updating via Bayes' rule*. It is a reasonable conclusion from the psychological and educational research cited above that this is markedly harder than an approach that does not require conditioning on observed events and so can avoid conditional probability.

Bayesian conclusions are perhaps not as clear to beginners as is often claimed. It is true that users of standard inference often utter a probability statement about their conclusion ("The probability that this answer is correct is 95%") when standard inference requires a statement about the method ("I got this answer by a method that gives a correct answer 95% of the time"). 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." It is easy to say "probability 95%," but much harder to be clear about what this probability is conditioned on, or about whether it is a physical or subjective probability. A semantically correct interpretation of a Bayesian credible region may reflect no better understanding than a semantically incorrect interpretation of a standard confidence region.

A glance back at the comments on Bayesian hypothesis testing in Section 4 will suggest several other barriers to beginners' grasp of Bayesian reasoning. I understand and have some sympathy for Bayesian claims to employ a single coherent approach that works in very general settings. It is *a priori* unlikely that such a general and powerful method will also be simple. Adding data (years of teaching experience, considerable reading of Bayesian expositions, and the findings of research on such topics as our understanding of probability), I am *a posteriori* convinced that Bayesian reasoning is even harder than the already hard reasoning of standard inference. Those who wish to dispute this conclusion can find my reasons stated in much greater detail in Moore (1997).

5. WHAT DO WE WANT OUR STUDENTS TO LEARN?

Let me conclude with a less specific—but perhaps more important—reason to hesitate to base a first course on Bayesian ideas. The teaching of elementary statistics has only recently moved from an over-emphasis on the parts of our subject that can be reduced to mathematics (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 trends that I and many others think are healthy. Figure 1 is a summary of the committee's recommendations that has been approved by the ASA Board of Directors. It deserves serious consideration as a thoughtful statement by the statistical community of principles that should govern our first courses.

The last main point in Figure 1 concerns pedagogy (what we do to help students learn) rather than content (what we want our students to learn). "Active learning" is the byword of the effort to reform the teaching of the mathematical sciences in general. Although my topic here is content, it is fair to note that there is a synergy between active learning and content that focuses on data and scientific problems, thus requiring students to explore and interact with data in the setting of substantive problems.

What do we want our students to learn? In our more realistic moments, we recognize that *many students will not take away from our first courses any clear conceptual grasp of formal probability or of the more subtle varieties of inference.* I would place all flavors of hypothesis testing and all Bayesian reasoning in the "more subtle" category. Students will, if we provide the opportunity, take away a number of messages more valuable than a grasp of formal inference. Several such messages are alluded to in the first two points of the ASA/MAA committee report.

Here are some things I want my beginning students to learn. *First, look at your data*, starting with graphs and simple calculations. Look for overall patterns and for deviations such as outliers. Always ask what your data say in the context of the setting they describe. *Recognize the importance of data production.* Faulty data production, (e.g., voluntary response or confounding) can render data worthless in ways that no fancy analysis can rescue. Understand that an observed association does not imply causation, and that randomized comparative experiments are the gold standard for evidence of causation. *Randomness*, as exemplified by deliberate use of chance in designs for data production, produces regular patterns of long-run behavior described qualitatively by the law of large numbers and the central limit theorem. This regularity applies *only* in the long run. *The data take priority over any model* (such as a normal distribution or a linear relationship) used to analyze them. Analysis starts by looking at the data, and models and assumptions are judged by the data. "Data sense" might summarize my primary objectives for a first statistics course.

Real problem settings often have vaguely defined goals and require the exercise of judgment. They require us to ask about the data production design, to search for problem data points, to try several presentations of the data in an iterative search for the main features. We may, happily, find that all is in order and that a standard inference procedure does in fact help answer the substantive question. We may find that formal inference is unjustified or doesn't answer a question of real interest. Intelligent supporters of standard inference recognize that formal inference is not always appropriate. They also recognize that, even when appropriate, inference often plays a "confirmatory" role, confirming by calculation what examination of the data suggests. This understanding contributes to a willingness to reduce the traditional first-course emphasis on inference in favor of hands-on work to develop data sense.

There is, on the other hand, some tendency in Bayesian instruction to neglect data analysis and design of data production in favor of more attention to inference. No doubt this tendency reflects in part the opportunity cost of the need to explain the Bayes machinery. Successful presentation of conditional probability and Bayes' theorem will require careful work with two-way tables, as in Albert (1995) and Rossman and Short (1995). The decision-theoretic bent of many Bayesians, which is clearly reflected in the leading advanced texts, Berger (1985) and Robert (1994), also contributes to an over-emphasis on formal inference. Bayesian inference is not as well integrated with the design of data production and with data analysis as is standard inference. Some Bayesians deny the importance of randomization in data production, whereas standard inference sees randomization as validating standard sampling models. The spirit of data analysis (derived from John Tukey) is to minimize prior assumptions and allow the data to suggest models. This spirit fits uneasily with Bayesian emphasis on the importance of prior (prior *to the data*) distributions and clearly structured outcomes. Bayesian thinking seems to start with models rather than with data—"no adhockery" is a Bayesian dictum.

Instruction in Bayesian inference is not in principle incompatible with developing data sense. In practice, however, it is likely to turn elementary statistics courses back toward probabilistic formalism and to leave our beginning instruction less accessible, less in contact with practice, and less in contact with data. As George Cobb so nicely puts it, "Bayesian inference offers a way to make a *probability* course deal with statistics."

REFERENCES

1. Albert, J. A. (1995), "Teaching inference about proportions using Bayes and discrete models," *Journal of Statistics Education* (electronic journal), 3 , number 3.

2. Altman, D. G. (1991), "Statistics in medical journals: developments in the 1980s," *Statistics in Medicine*, 10, 1897–1913.

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

4. Berger, J. O., and Delampady, M. (1987), "Testing precise hypotheses," *Statistical Science*, 2, 317–352.

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

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

7. Cobb, G. (1992), "Teaching statistics," in *Heeding the Call for Change: Suggestions for Curricular Action*, L. A. Steen (ed.), Washington, DC: Mathematical Association of America, 3–43.

8. DeGroot, M. H., and Mezzich, J. E. (1985), "Psychiatric statistics," in *A Celebration of Statistics: The ISI Centenary Volume*, A. C. Atkinson and S. E. Fienberg (eds.), New York: Springer, 145–164.

9. Dempster, A. P. (1971), "Model searching and estimation in the logic of inference," in *Foundations of Statistical Inference*, V. P. Godambe and V. A. Sprott (eds.), Toronto: Holt, Rinehart, and Winston. 56–76.

10. 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.

11. Dunn, G., Hand, D. J., and Sham, P.C. (1993), "Statistics and the nature of depression," *Journal of the Royal Statistical Society*, A156, 63–87.

12. 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., J. C. Bailar and F. Mosteller (eds.), Boston: NEJM Press, 45–57.

13. Everitt, B. S. (1987), "Statistics in psychiatry," *Statistical Science*, 2, 107–134.

14. 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.

15. Hammer, A. S., and Buffington, C. A. (1994), "Survey of statistical methods used in the veterinary medical literature," *Journal of the American Veterinary Medical Association*, 205, 344–345.

16. Hokanson, J. A., Bryant, S. G., Gardner, R., Luttman, D. J., Guernsey, B. G., and Bienkowski, A. C. (1986), "Spectrum and frequency of use of statistical techniques in psychiatric journals," *American Journal of Psychiatry*, 143, 1118–1125.

17. Hokanson, J. A., Luttman, D. J., and Weiss, G. B. (1986), "Frequency and diversity of use of statistical techniques in oncology journals," *Cancer Treatment Reports*, 70, 589–594.

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

19. 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.

20. Kass, R. E., and Wasserman, L. (1996), "The selection of prior distributions by formal rules," *Journal of the American Statistical Association*, 91, 1343–1370.

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

22. Moore, D. S. (1997), "Bayes for beginners? Some pedagogical questions," in *Advances in Statistical Decision Theory and Methodology*, S. Panchapakesan and N. Balakrishnan (eds.), Boston: Birkhäuser, Boston, to appear.

23. Nieminen, P. (1995), "Statistical content of published therapeutic community research," *Therapeutic Communities*, 16, 239–251.

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

25. 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.

26. 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.

27. Slovic, P., Fischhoff, B., and Lichtenstein, S. (1982), "Facts vs. fears: Understanding perceived risk," In *Judgment under Uncertainty: Heuristics and Biases*, D. Kahneman, P. Slovic, and A. Tversky (eds.), New York: Cambridge University Press, 463–489.

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

29. Zeckhauser, R. J., and Viscusi, W. K. (1990), "Risk within reason," *Science*, 248, 559–564.

Table 1. Responses of 103 DOE Statisticians Asked to Name the Three Techniques Most Important in Their Work.

Statistical Technique	Top 3 Responses
Regression analysis	63
Basic statistical methods	37
Analysis of variance	26
Design of experiments	26
Probability modeling	22
Sampling, survey sampling	17
Simulation	16
Graphical display and data summary	12
Multivariate analysis	12
Quality control, acceptance sampling	12
Exploratory data analysis	11
Reliability, life data analysis	11
Nonlinear estimation	7
Biostatistics, bioassay	6
Nonparametric methods	5
Numerical analysis	5
Bayesian methods	4
Time series analysis	4
Categorical data analysis	3
Variance components	3
Ranking, paired comparisons	1
Other	5

Source: Rustagi and Wright (1995).

Table 2. Statistical Techniques Most Commonly Used in 115 *New England Journal of Medicine* Articles, 1989.

Statistical technique	Number of articles
t-tests	45
Contingency tables	41
Survival methods	37
Epidemiologic statistics	25
Nonparametric tests	24
Analysis of variance	23
Pearson correlation	22
Multiple regression	16
Multiway tables	11
Simple linear regression	10
Multiple comparisons	10
Adjustment and standardization	10

Source: Excerpted from Table 4 of Emerson and Colditz (1992).

Figure 1: Recommendations of the ASA/MAA Joint Curriculum Committee

The following three recommendations are intended to apply to any course whose goal is to introduce the nature of statistics to beginning students. For more advanced courses in the statistics curriculum, some of the recommendations would still apply; others would need to be modified.

1. **Emphasize the elements of statistical thinking:**
   (a) the need for data,
   (b) the importance of data production,
   (c) the omnipresence of variability,
   (d) the measuring and modeling of variability.
2. **Incorporate more data and concepts, fewer recipes and derivations. Wherever possible, automate computations and graphics**. An introductory course should:
   (a) rely heavily on *real* (not merely realistic) data,
   (b) emphasize *statistical* concepts, e.g., causation vs. association, experimental vs. observational and longitudinal vs. cross-sectional studies,
   (c) rely on computers rather than computational recipes,
   (d) treat formal derivations as secondary in importance.
3. **Foster active learning**, through the following alternatives to lecturing:
   (a) group problem solving and discussion,
   (b) laboratory exercises,
   (c) demonstrations based on class-generated data
   (d) written and oral presentations,
   (e) projects, either group or individual.