Commentary on Teaching Bayesian Statistics

David A. Freedman

1. INTRODUCTION

Should we teach Bayesian statistics in a general introductory course? I agree with Moore's answer: the subjective approach is best taught in later courses, after some probability has been developed; the beginning course should focus on standard frequentist techniques.

The subjective approach is elegant, powerful, and logically water-tight. It provides a rich set of metaphors, and a host of fascinating mathematical questions. However, there seem to be serious difficulties in applying the methods to real problems; see, for instance, LeCam (1977) or Freedman (1996). Such foundational issues are background to the present discussion of teaching.

At Berkeley, the introductory course does a lot on descriptive statistics—histograms and scatter plots, the average and the standard deviation, correlation and regression. Design issues are central: for instance, comparisons between observational studies and randomized controlled experiments. We emphasize the role of "box models" in statistical inference, and the advantages of probability samples over convenience samples. These are important topics, perhaps more important than any formal statistical calculations, and they are not easily reconciled with the Bayesian approach.

Would a Bayesian course be simpler to teach? On this score, Albert and Berry are too optimistic. For instance, students—and others—have a lot of trouble distinguishing P{A|B} from P{B|A}. Even the difference between P{A} and P{A|B} is problematic. There is a real psychological conflict between two facts about tossing a fair coin:

(i) the chance of 10 heads in a row is 1/1,024;

(ii) given that the first 9 tosses were heads, the chance of a head on the 10th toss is 1/2.

One consequence is the "gambler's fallacy"—a cognitive illusion whose power is demonstrated by Cohen (1981). Beginning students in our courses have a hard enough time with fractions, never mind these probabilistic subtleties. Pretest results from U.C. Berkeley may be of interest (Freedman, Pisani, Purves, 1997). Two items give the flavor: only half the students get the first question right, and one in six gets the second.

There are 100 million eligible voters in the United States. The Gallup poll interviews 5,000 of them. This amounts to 1 eligible voter out of every _____.

In the United States, 1 person out of every 500 is in the army, and 3 out of every 10,000 are army officers. What percentage of army personnel are officers, or can this be determined from the information given?

For many of these students, probability theory comes down to one question, "When do I add and when do I multiply?" To this audience, Bayes' theorem will not be an illumination.

Of course, significance testing present difficulties of its own, but students need to understand *P*-values because $P < .05$ has become a talisman of science. A parallel issue is canvassed by the three papers: To what extent are Bayesian methods used in practice? Such methods have a considerable presence in certain fields. In other areas, frequentist techniques dominate but practitioners have the idea that probabilities quantify uncertainty (including *P*-values). On balance, Moore is close to right: frequentist inference is the paradigm, and that is a powerful argument for the frequentist introductory course.

Can frequentist ideas be taught to beginners? Moore's comment (p.15) is on target: students come away with a reasonable grip on the ideas, if not the syntax. Test results from courses with a frequentist orientation are reported in Freedman, Pisani, and Purves (1997). These data span many years and show that much can be done, even for students whose technical skills are limited. For example, the average score was about 9/10 on the following question (Statistics 2 final exam, Berkeley, fall 1996; n=185).

On October 20, 1993, the *San Francisco Chronicle* reported a survey of top high school students in the U. S. According to the survey,

> "Cheating is pervasive. Nearly 80 percent admitted some dishonesty, such as copying someone's homework or cheating on an exam. The survey was sent last spring to 5,000 of the nearly 700,000 high achievers included in the 1993 edition of *Who's Who Among American High School Students*. The results were based on the 1,957 completed surveys that were returned. 'The survey does not pretend to be representative of all teenagers,' said *Who's Who* spokesman Andrew Weinstein. 'Students are listed in *Who's Who* if they are nominated by their teachers or guidance counselors. Ninety-eight percent of them go on to college.'"

Why isn't the survey representative of all teenagers? Is the survey representative of the nearly 700,000 high achievers included in the 1993 edition of *Who's Who Among American High School Students*?

---

2. IMPROPER PRIORS

Strictly speaking, Bayesian methods start with a proper prior. Some statisticians now use improper priors, while others use "priors" that are data-dependent. Such algorithms would be hard to defend on the usual Bayesian grounds, and frequentist operating characteristics need to be demonstrated. (Bayes' theorem is among other things a powerful heuristic engine: but few statistical algorithms are self-justifying, and even the best heuristics may lead us far astray.) These considerations are well beyond the scope of any conceivable introductory course.

3. ARGUMENTS FOR THE SUBJECTIVIST POSITION

Albert and Berry are quite pragmatic; other writers often suggest that frequentists are benighted. In the extreme, to be non-Bayesian is to be "incoherent" if not "irrational"; the introductory course should then be switched from an irrational paradigm to the rational alternative. Life is not that simple, and the subjectivist position is not compelling. For instance, one of their arguments runs as follows. If a statistician has to post odds on each of several events and cover bets on either side at any stakes, a clever bettor can make money whatever the state of nature may prove to be—unless a prior probability is used to set the odds. (The events have to form an algebra.) A "Dutch book" is a system of bets that makes money given any state of nature; at least in this context, "incoherence" just means the possibility of a Dutch book.

The theorem goes back to de Finetti; see Freedman and Purves (1969) for an extension. The conditions of the theorem, however, describe only a highly stylized version of applied statistics—after all, who will cover as many bets as the theorem requires? Therefore, the Dutch-book argument is too far removed from practice to have much force.

Another line of argument cites Edwards, Lindeman, and Savage (1963) to show that the prior does not really matter, being eventually swamped by the data. That is so in smooth, low-dimensional situations (and then frequentist methods will give essentially the same answers too). On the other hand, if the number of parameters is large, the situation is quite different; the latter case may be the relevant one. For reviews, see Diaconis and Freedman (1986, 1996).

Other arguments for axiomatics of the von Neumann-Morganstern-Savage type are almost purely normative; see Kreps (1988, p.4). These generally boil down to a truism: you should obey the axioms if, after careful consideration, you want to obey the axioms. Moreover, non-Bayesians turn out to be irrational only by a little semantic trick: according to the subjectivists' definition, "rational" behavior just means behavior that conforms to their axioms. On the descriptive side, of course, rational people generally do not behave like Bayesians—or frequentists, for that matter. See, for instance, Tversky and Kahneman (1986).

4. CONCLUSION

In the abstract, the Bayesian approach is tidier, and students are often impatient of practical detail. These are good arguments for a Bayesian introductory course. However, subjectivist ideas are inherently difficult, and the messy details may be the ones that matter. After much soul-searching, I opted long ago for a frequentist approach. I still think that was the right decision. The present papers make thoughtful contributions to an important issue, and they will stimulate many valuable conversations.

REFERENCES

Cohen, L. J. (1981), "Can Human Irrationality Be Experimentally Demonstrated" (with discussion), *Journal of behavioral and Brain Sciences*, 4, 317-70.

Diaconis, P. and Freedman, D. (1986), "On the Consistency of Bayes Estimates" (with discussion), *Annals of Statistics*, 14, 1-87.

Diaconis, P. and Freedman, D. (1996), *Consistency of Bayes Estimates for Nonparametric Regression: Normal Theory*, Technical report no. 414, Department of Statistics, University of California, Berkeley.

Edwards, W., Lindeman, H., and Savage, L. J. (1963), "Bayesian Statistical Inference for Psychological Research," *Psychological Review*, 70, 193-242.

Freedman, D. (1996), "Some Issues in the Foundations of Statistics" (with discussion), *Foundations of Science*, 1, 19-39. Polish Academy of Sciences, Warsaw. Reprinted by Kluwer, Dordrecht, The Netherlands.

Freedman, D. and Purves, R. (1969), "Bayes Method for Bookies," *Annals of Mathematical Statistics*, 40, 1177-86.

Freedman, D., Pisani, R., and Purves, R. (1997), *Instructors' Manual for Statistics*, 3rd ed. Department of Statistics, University of California, Berkeley; to be published by Norton, New York.

Kreps, D. (1988), *Notes on the Theory of Choice*, Boulder, Westview Press

Le Cam, L. (1977), "A Note on Metastatistics or 'An Essay Toward Stating a Problem in the Doctrine of Chances,'" *Synthese*, 36, 133-60.

Tversky, A. and Kahneman, D. (1986), "Rational choice and the framing of decisions," *Journal of Business*, 59, no. 4, part 2, S251-78.

Author's note

David A. Freedman is Professor of Statistics, University of California, Berkeley, California 94720