the bootstrap. For example, the first edition of *Numerical Recipes: the Art of Scientific Computing* by Press et al. (1986) sketches, in Section 14.5, a construction of bootstrap "pivotal" confidence limits for model parameters. The book cites as references two astrophysical papers published in 1976. In comparing these astrophysical papers with Efron (1979a) and with later bootstrap work, one sees again the historical role of the statistician in formulating, sharpening and developing a primitive new data-analytic idea. The bootstrap is not just a notion inflicted by theoretical statisticians upon reluctant data analysts. Also the reverse holds. Incidentally, the second edition of *Numerical Recipes* cites Efron.

Broadcasting bootstrap methods requires updating statistical education. Education goes beyond accessible software, mentioned in statement (f). Many undergraduate statistics texts fail to treat the Behrens–Fisher problem adequately, let alone developments of recent decades such as nonparametric regression, statistical graphics, generalized linear models or bootstrap. Why? I suggest the following: (a) Comprehension of modern statistical methods benefits from an actual need to analyze complex data. (b) Statistical theory relies on the mathematics of the twentieth century. (c) Using modern statistical methods, such as bootstrap, is computer-intensive. Meeting these three requirements is not so easy in large undergraduate classes. However, computing costs continue to drop as PC's become more powerful; students face a growing need to analyze the ambient information flood; and careful analysis of simple cases can develop statistical intuition. Meanwhile, MA-level courses can be effective in spreading modern statistical ideas to students in other fields. On bootstrap methods, we now have several trustworthy monographs.

Statements (d) and (e) flirt with double-think. The main thrust of bootstrap research, from 1979 onward, has been to understand what form of bootstrap works for what kind of statistical model. Young himself mentions the steady development of bootstrap techniques for time-series analysis. In preprints, this time-series research dates back to at least 1988. The work on squeezing better performance from bootstrap methods that is denigrated in assertion (d) resolved problems neglected according to statement (e), and these results are part of the ongoing research into diagnostics of bootstrap reliability. It is a noteworthy success that intuitive bootstrap critical values achieve the good small-sample performance of Welch's solution to the Behrens–Fisher problem or, more generally, of the Bartlett adjustment to likelihood ratio confidence sets and tests.

Statement (g) illustrates the numbing effect of familiar terminology. The word "nonparametric" is a blind description of what is actually a function-valued parameter. The word "likelihood" is equally a misnomer. Consider the three parameter lognormal model—smooth in the parameters and possessing finite Fisher information—whose likelihood function climbs to infinity at a most unlikely place. Bootstrap and empirical likelihood are complementary techniques rather than competitors. For instance, after empirical likelihood determines the shape of a confidence region, bootstrap provides a more accurate critical value for that region.

I conclude by mentioning two useful references not cited in Young's essay. The proceedings of the 1990 Trier conference (Jöckel, Rothe and Sendler, 1992) contain papers on random number generation and Monte Carlo tests as well as on bootstrap theory and applications. Janas (1993) surveys some of the earlier work on bootstrapping time series.

Comment

B. Efron

"My general feeling about bootstrapping is that I don't like it very much. It's easy for me to say that, because nowadays I don't have to do practical problems for a living."—Henry Daniels*, Statistical Science, August 1993.*

B. Efron is Professor of Statistics and Biostatistics, Statistics Department, Stanford University, Stanford, California 94305-4065.

In 1980 I gave a talk at Ann Arbor called "Six influential papers and what ever became of them." The six papers were classics of the postwar literature: Wilcoxon on rank tests, Huber on robust estimation, Robbins on empirical Bayes, James and Stein on shrinkage estimates, Cox on proportional hazards and Tukey on the jackknife variance estimate. The question raised in the talk, but not settled, was why two of these papers, Wilcoxon's and Cox's, seemed to leap into applied use, while the others com-