FRIEDMAN, J. H. and TIBSHIRANI, R. J. (1984). The monotone smoothing of scatter-plots. *Technometrics* **26** 243–250.
HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.
HASTIE, T. J. and TIBSHIRANI, R. J. (1985). Discussion of Peter Huber's "Projection Pursuit." *Ann. Statist.* **13** 502–508.
HINKLEY, D. V. (1978). Improving the jackknife with special reference to correlation estimation. *Biometrika* **65**, 13–22.
HYDE, J. (1980). *Survival Analysis with Incomplete Observations. Biostatistics Casebook.* Wiley, New York.
JAECKEL, L. (1972). The infinitesimal jackknife. Memorandum MM 72-1215-11. Bell Laboratories, Murray Hill, New Jersey.
JOHNSON, N. and KOTZ, S. (1970). *Continuous Univariate Distributions.* Houghton Mifflin, Boston, Vol. 2.
KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete samples. *J. Amer. Statist. Assoc.* **53** 457–481.
KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.
MALLOWS, C. (1974). On some topics in robustness. Memorandum, Bell Laboratories, Murray Hill, New Jersey.
MILLER, R. G. (1974). The jackknife—a review. *Biometrika* **61** 1–17.
MILLER, R. G. and HALPERN, J. (1982). Regression with censored data. *Biometrika* **69** 521–531.
RAO, C. R. (1973). *Linear Statistical Inference and Its Applications.* Wiley, New York.
SCHENKER, N. (1985). Qualms about bootstrap confidence intervals. *J. Amer. Statist. Assoc.* **80** 360–361.
SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.
THERNEAU, T. (1983). Variance reduction techniques for the bootstrap. Ph.D. thesis, Stanford University, Department of Statistics.
TIBSHIRANI, R. J. and HASTIE, T. J. (1984). Local likelihood estimation. Tech. Rep. Stanford Univ. Dept. Statist. **97**.
TUKEY, J. (1958). Bias and confidence in not quite large samples, abstract. *Ann. Math. Statist.* **29** 614.

Comment

J. A. Hartigan

Efron and Tibshirani are to be congratulated on a wide-ranging persuasive survey of the many uses of the boostrap technology. They are a bit cagey on what is or is not a bootstrap, but the description at the end of Section 4 seems to cover all the cases; some data y comes from an unknown probability distribution F; it is desired to estimate the distribution of some function $R(y, F)$ given F; and this is done by estimating the distribution of $R(y^*, \hat{F})$ given $\hat{F}$ where $\hat{F}$ is an estimate of F based on y, and y^* is sampled from the known $\hat{F}$.

There will be three problems in any application of the bootstrap: (1) how to choose the estimate $\hat{F}$? (2) how much sampling of y^* from $\hat{F}$? and (3) how close is the distribution of $R(y^*, \hat{F})$ given $\hat{F}$ to $R(y, F)$ given F?

Efron and Tibshirani suggest a variety of estimates $\hat{F}$ for simple random sampling, regression, and autoregression; their remarks about (3) are confined mainly to empirical demonstrations of the bootstrap in specific situations.

I have some general reservations about the bootstrap based on my experiences with subsampling techniques (Hartigan, 1969, 1975). Let $X_1, \ldots, X_n$ be a random sample from a distribution F, let F_n be the empirical distribution, and suppose that $t(F_n)$ is an estimate of some population parameter $t(F)$. The statistic $t(\hat{F}_n)$ is computed for several random subsamples (each observation appearing in the subsample with probability ½), and the set of $t(\hat{F}_n)$ values obtained is regarded as a sample from the posterior distribution of $t(F)$. For example, the standard deviation of the $t(\hat{F}_n)$ is an estimate of the standard error of $t(F_n)$ from $t(F)$; however, the procedure is not restricted to real valued t.

The procedure seems to work not too badly in getting at the first- and second-order behaviors of $t(F_n)$ when $t(F_n)$ is near normal, but it not effective in handling third-order behavior, bias, and skewness. Thus there is not much point in taking huge samples $t(\hat{F}_n)$ since the third-order behavior is not relevant; and if the procedure works only for $t(F_n)$ near normal, there are less fancy procedures for estimating standard error such as dividing the sample up into 10 subsamples of equal size and computing their standard deviation. (True, this introduces more bias than having random subsamples each containing about half the observations.) Indeed, even if $t(F_n)$ is not normal, we can obtain exact confidence intervals for the median of $t(F_{n/10})$ using the 10 subsamples. Even five subsamples will give a respectable idea of the standard error.

Transferring back to the bootstrap: (A) is the boot-

J. A. Hartigan is Eugene Higgins Professor of Statistics, Yale University, Box 2179 Yale Station, New Haven, CT 06520.