of the estimate (at normal means) is optimal for $W_i^l = 1 + \sqrt{n}\xi_i^l$.

For $n = 8$, obtain an efficient estimate from subsamples (1234), (1256), (1278), (1357), (1368), (1458), (1467); use as many as you need, and if $n > 8$ divide the sample into 8 groups as evenly as possible. I think it must be rare that the various approximations needed to connect the resampled computation to the computation of interest will be satisfied well enough to justify more than a few resamples. Perhaps this method might be called the *shoestring*.

ADDITIONAL REFERENCES

HARTIGAN, J. A. (1969). Using subsample values as typical values. *J. Amer. Statist. Assoc.* **64** 1303–1317.

HARTIGAN, J. A. (1975). Necessary and sufficient conditions for asymptotic joint normality of a statistic and its subsample values. *Ann. Statist.* **3** 573–580.

Rejoinder

B. Efron and R. Tibshirani

Professor Hartigan, who is one of the pioneers of resampling theory, raises the question of higher order accuracy. This question has bothered resamplers since the early days of the jackknife. Sections 7 and 8 of our paper show that the bootstrap can indeed achieve higher levels of accuracy, going the next step beyond simple estimates of standard error. The bootstrap confidence intervals we discuss are *not* of the crude (although useful) first-order form $\hat{\theta} \pm \hat{\sigma} z^{(\alpha)}$. They explicitly incorporate the higher order corrections about which Hartigan is legitimately concerned.

In particular the "z_0" term (7.8) is a correction for bias, and the acceleration constant "a," (7.16), is a correction for skewness. These correspond to Hartigan's $b(F)$ and $s_3(F)$, respectively. The reader who follows through Tables 5 and 7 will see these corrections in action. The fact that they produce highly accurate confidence intervals is no accident. The theory in Efron (1984a, 1984b) demonstrates higher order accuracy of the BC_a intervals in a wide class of situations. This demonstration does not yet apply to fully general problems, but current research indicates that it soon will. (The impressive higher order asymptotic results of Beran, Singh, Bickel, and Freedman, referred to in the paper, underpin these conclusions.)

It is worth mentioning that the bias and skewness corrections of the bootstrap confidence intervals are not of the simple "plug into an approximate pivotal" form suggested in Hartigan's remarks. The theory is phrased in a way which automatically corrects for arbitrary nonlinear transformations, even of the violent sort encountered in the correlation example of Table 5. In this sense the bootstrap theory does handle "non-normal situations."

Since this paper was written, research by several workers, including T. Hesterberg, R. Tibshirani, and T. DiCiccio, has substantially improved the computational outlook for bootstrap confidence intervals. It now appears possible that bootstrap sample sizes closer to $B = 100$ than $B = 1000$ may be sufficient for the task. However, these improvements are still in the process of development.

Professor Hartigan's last remarks, on the comparative efficiency of different resampling methods, need careful interpretation. There are two concepts of efficiency involved: the efficiency of the numerical algorithm in producing an estimate of variance, and the statistical efficiency of the estimate produced. There is no question that other resampling techniques, for example, the jackknife, can produce variance estimates more economically than does the bootstrap. We have argued, both by example and theory, that the bootstrap variance is generally more efficient as a statistical estimator of the unknown true variance.

This is not surprising given that methods like the jackknife are Taylor series approximations to the bootstrap (see Section 10). The simple idea in (2.3), substituting $\hat{F}$ for F, lies at the heart of all nonparametric estimates of accuracy. The bootstrap is the crudest of these methods in that it computes $\sigma(\hat{F})$ directly by Monte Carlo. For this reason it is also the method that involves the least amount of analytic approximation. It is perhaps surprising, and certainly gratifying, that a method based on such a simple form of inference is capable of producing quite accurate confidence intervals.

To say that the bootstrap is good, as we have been blatantly doing, doesn't imply that other methods are bad. Professor Hartigan's own work shows that for some problems, for example, forming a confidence interval for the center of a symmetric distribution, other methods are better. We hope that resampling methods in general will continue to be a lively research topic.