of the original error terms, e_i. The procedure suggested in Section 7 is an attempt to ensure that the error terms associated with x_i in the bootstrap samples capture some of the possible dependence on x_i. This approach leads to the "bootstrap" estimator for the covariance of $\tilde{\beta}$ being $V_{J(1)}$, and so this resampling method is as effective as the weighted jackknife in this regard. The potential advantage the bootstrap procedure has over the jackknife in general is in approximating the distribution of $(\hat{\beta} - \beta)$. In the case where the e_i's are independent and identically distributed the procedure suggested in Section 7 forces some arbitrary distribution on y_i^* through t_i^*, and the actual distribution of the e_i's is lost. Thus, the usefulness of this approach in generating confidence intervals by approximating the distribution of $(\hat{\beta} - \beta)$ is in question. Perhaps the jackknife–bootstrap hybrid is the answer to this problem and this model certainly deserves more investigation.

The bootstrap percentile method for calculating confidence intervals in regression has been investigated by Robinson (1985). He compared the bootstrap approach to the exact confidence intervals obtained by inverting permutation tests and suggested an adjustment to the bootstrap percentile method to improve its coverage probability.

The use of t-confidence intervals in the simulation study for the parameter $\theta = -\beta_1/2\beta_2$ should give moderate results with normal errors and β_2 away from 0. Weber and Welsh (1983) found that the standardised distribution of the jackknife statistic for θ can be very skewed and so one would not expect the symmetric t-confidence interval to give reasonable coverage in general. The adjusted percentile methods appear to be the appropriate way of obtaining nonparametric confidence intervals for θ.

REFERENCES

ROBINSON, J. (1985). Nonparametric confidence intervals in regression—the bootstrap and randomization methods. In *New Perspectives in Theoretical and Applied Statistics* (M. Puri, J. P. Vilaplana and W. Wertz, eds.). Wiley, New York.

WEBER, N. C. (1986). The jackknife and heteroskedasticity—consistent variance estimation for regression models. *Econom. Lett.* **20** 161–163.

WEBER, N. C. and WELSH, A. H. (1983). Jackknifing the general linear model. *Austral. J. Statist.* **25** 425–436.

WHITE, H. (1980). A heteroskedasticity—consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48** 817–838.

DEPARTMENT OF MATHEMATICAL STATISTICS
UNIVERSITY OF SYDNEY
SYDNEY, N.S.W. 2006
AUSTRALIA

H. P. WYNN AND S. M. OGBONMWAN

The City University and University of Benin

Along with commenting on this authoritative paper, we wish to make a plea for an approach to the computational problems of resampling and simulation

more generally. Our work divides into two parts: (1) accelerated simulation and (2) the use of "cheap" methods of inference based on a method closer in spirit to the inversion of permutation tests in rerandomization.

In simulation, and *discrete* simulation in particular, a basic approach is to sample from a set S of possible configurations, which may be termed the *reference set.* Thus, in the simple bootstrap S consists of all n^n possible with replacement samples obtained from a data sample $x = (x_1, \ldots, x_n)$. In a permutation test problem S would consist of all $n!$ permutations of x. In a sign test problem we may consider all 2^n configurations $(\pm x_1, \ldots, \pm x_n)$.

An idea mentioned, for example in Tukey et al. (1978), is to do complete enumeration over a restricted reference set $S' \subseteq S$, which has two important properties:

(1) S' has very much smaller cardinality than S.
(2) S' fills out S in a dense and uniform way so that any inference based on S' is close to that based on S.

We refer to S' as a simulation code or resampling code by analogy with coding theory. We may consider that a task of the simulation is to get close to the "true" value of the average of some statistic t:

$$\bar{t}_S = \frac{1}{N} \sum_{w \in S} t(x(w)),$$

where $N = \sharp(S)$. In approximating $\bar{t}_S$ by $\bar{t}_{S'} = 1/N' \Sigma_{w \in S'} t(x(w))$, where $N' = \sharp(S')$, we are led to a problem of numerical integration. There is a large literature on quadrature methods based on low discrepancy sequences: Halton–Hammersley sequences and so forth.

We can measure the discrepancy for the simple bootstrap as follows. First identify a bootstrap sample by the selected indices of the with-replacement sample: If, say, $n = 5$ and the sample chosen were $(x_1, x_1, x_2, x_2, x_5)$, we would refer to this as $(1, 1, 2, 2, 5)$. Each sample is thus in one-to-one correspondence with a point in the n^n hypercube grid of side n. If we suppress the data points we can consider the integration as being over this grid. Now S' will be a subset of this grid. The L^2 normed discrepancy of S' would be

$$\tau^2 = \frac{1}{n^n} \sum_{w \in S} g^2(w),$$

where $g(w)$, the local discrepancy, is

$$g(w) = F_{S'}(w) - F_S(w),$$

the difference between the "empirical" n-dimensional c.d.f. of S' and S. The latter is that for a uniform distribution on the grid. The *expected* value of τ^2 for the simple bootstrap sample with N' configurations independently uniformly chosen from S is

$$E(\tau^2) = \frac{1}{N'} \left\{ \left(\frac{1}{2}\right)^n \left(1 + \frac{1}{n}\right)^n - \left(\frac{1}{3}\right)^n \left(1 + \frac{3}{2n} + \frac{1}{2n^2}\right)^n \right\},$$

which tends to

$$\frac{1}{N'}\left\{\left(\frac{1}{2}\right)^n e - \left(\frac{1}{3}\right)^n e^{3/2}\right\} \quad \text{as } n \to \infty.$$

This can be used as a benchmark to compare the discrepancy of other deterministically chosen sets S', which can, incidentally, be explicitly computed also.

The "balanced bootstrap" independently discussed by Davison, Hinkley and Schechtman (1986) *always* has lower expected discrepancy. The balanced bootstrap is simple resampling with the added restriction that each data point x_i occurs equally often in the full set S'. In our notation this says that the one-dimensional discrepancies, considering the marginal distributions on each axis of the n^n grid, are zero.

The deterministic codes obtained in various ways by direct minimization of discrepancy or by judicious "space-filling" together with balance do even better. Using a version of Koksma's inequality, a bound, depending on discrepancy and the variation of t, can be placed on the deterministic error

$$|\varepsilon| = |\bar{t}_S - \bar{t}_{S'}|.$$

(See Niederreiter (1978) for a discussion of these bounds.) First-order balance decreases the bound.

Armed, hopefully, with a catalogue of, or easily complemented algorithm to find, the simulation codes S', we can do very large numbers of simulations in situations such as bootstrap inference or inversion of permutation tests. We should also mention the applications to more general numerical experimentation in areas such as computer aided design. But now to a cheap method of inference.

We believe that the key to methods of inference when simulation becomes easy is to let the configuration set S (or S') *depend* on the parameters θ of a model as follows. Let us consider a parameter dependent transformation of the data $y_\theta(x)$. For example, in a two-sample problem we may have

$$y_\theta = \left(x_1, \ldots, x_{n_1}, x_{n_1+1} - \theta, \ldots, x_n - \theta\right).$$

Let $T(y)$ be some statistic of interest. In the two-sample problem it could be

$$T(y) = \frac{1}{n_1}(y_1 + \cdots + y_{n_1}) - \frac{1}{n - n_1}(y_{n_1+1} + \cdots + y_n).$$

If we now compute all values $T(y_\theta^*)$ for all y_θ^* derived from our reference set S (or S') we will have an empirical c.d.f. for T say $\hat{F}_T(t|\theta)$. Smoothing the $T(y_\theta^*)$ values will produce a density $\hat{f}_T(t|\theta)$ that depends on θ. Then putting $t = t_0 = y_\theta$, the "observed" value, will yield a *likelihood* $\hat{L}(\theta)$. This likelihood is close in spirit to that used in deriving an inverted sign test or permutation test. The likelihood $\hat{L}(\theta)$ can be used to do "ordinary statistics"; for example, it can be combined with a prior $\pi(\theta)$ to obtain posterior confidence intervals.

The standard approach to simulation, and that adopted in Professor Wu's paper, is to obtain first an estimator $\hat{\theta}$ and then place confidence regions around this using variance estimation. Our approach, we repeat, is to simulate the values of a statistic for a range of selected θ values. Of course the choice of statistic is

critical, but so it is in nonparametric methods in which statisticians are quite happy to consider parameter dependent reference sets. Here is our recipe for straight line regression, referring to Professor Wu's paper.

First set (on the computer) the value of β. Then *evaluate* the

$$z_i = y_i - x_i^T\beta.$$

Bootstrap the z_i values (keeping the x_i values fixed). Regress each bootstrap z_i^* set back on the x_i values to obtain a $\hat{\beta}^*$ value for each bootstrap. Smooth the set of $\hat{\beta}^*$ to obtain $\hat{f}(\hat{\beta}|\beta)$. Note that $\hat{f}$ depends on the set value β. Put $\hat{\beta} = \hat{\beta}_0$ the value obtained from the original (unbootstrapped) z_i values and we have our generated likelihood $L(\beta)$. Here $\hat{\beta}$ plays the role of the statistic T.

REFERENCES

DAVISON, A., HINKLEY, D. and SCHECHTMAN, E. (1986). Efficient bootstrap simulation. *Biometrika* **73**. To appear.

NIEDERREITER, H. (1978). Quasi-Monte Carlo methods and pseudo-random numbers. *Bull. Amer. Math. Soc.* **84** 957–1041.

TUKEY, J. W., BRILLINGER, D. R. and JONES, L. V. (1978). *The Management of Weather Resources* **2**. Weather Modification Advisory Board, Statistical Task Force, U.S. GPO, Washington.

DEPARTMENT OF MATHEMATICS
UNIVERSITY STATISTICAL LABORAOTRY
CITY UNIVERSITY
NORTHAMPTON SQUARE
LONDON EC1V 0HB
ENGLAND

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BENIN
PMB 1154, EKENWAN ROAD
BENIN CITY
NIGERIA

REJOINDER

C. F. J. WU

University of Wisconsin-Madison

The overwhelming response to the paper reflects great interest and perhaps confusion in the bootstrap and the jackknife. The contributions of the discussants make the discussion informative, valuable and diverse. Their comments, even though I do not always agree with them, help clarify certain points, suggest new ideas and results, and in some cases prompt me to study the issues more carefully. Most of these comments can be grouped into five broad categories. My reply will concentrate on the major points of interest in each category.

Among the new ideas and results to which my response will not be directed, let me mention: robustification of resampled values (*Beran*), two interesting applications from genetics (*Felsenstein* and *Mitchell-Olds*), examples of inconsistency of bootstrap estimators (*Olshen* and *Srivastava*), use of weighted jackknife in variance components model (*Rao and Prasad*), results on the