(Felsenstein (1985)). A phylogeny is a tree showing the genealogy of a set of species. It has a shape and in some cases branch lengths as well. If we draw a subset of the characters that are used to estimate the tree and come up with a tree with (say) a shorter length of one of the branches, it may be the case that extrapolation using the factor in (4.4) will make the branch length negative, which is impossible. This indicates that a change in tree topology is called for, but there is no easy way of knowing what change. The problem is even more severe when the resampled estimate shows a change in topology—how is one to extrapolate it?

In the absence of any ability to extrapolate, one cannot use the ordinary jackknife or most of the variants Wu describes, but one can use the jackknife with subset size $(n + k - 1)/2$, or the bootstrap. In my 1985 paper, I mentioned the possibility that one could drop half the observations chosen at random instead of doing the bootstrap, but could not see what advantage this might have over the bootstrap. It would be nice to have some simulations comparing these two approaches. For now they both appear to be tenable approaches. In biology we have so little confidence in the independence of our data points and the validity of our assumptions that any method that comes within a factor of two of indicating a confidence set is probably good enough. The obsession statisticians have with correcting for biases of order $1/n$ must seem somewhat comical to most biologists, who only wish their problems were that small.

REFERENCE

FELSENSTEIN, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39** 783–791.

DEPARTMENT OF GENETICS SK-50
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195

D. A. FREEDMAN[1]

University of California, Berkeley

I comment first on that part of Wu's paper relating to the bootstrap. It seems worthwhile to restate the bootstrap idea in brief. Consider data X drawn from a stochastic model M, and a statistical functional $F(X, M)$ whose distribution is to be estimated. Let $\hat{M}$ be an estimator of M, computed from the data X. Let X^* be drawn from the model $\hat{M}$. The bootstrap consists in approximating the M-law of $F(X, M)$ by the $\hat{M}$-law of $F(X^*, \hat{M})$. In particular, if $\hat{M}$ is silly, the bootstrap cannot work: Like any statistical procedure, the bootstrap is model-dependent. (This is the statistics version of the *no free lunch* principle.)

If I understand it correctly, Wu observes that in a heteroscedastic regression model M, a bootstrap procedure based on a homoscedastic estimator $\hat{M}$ will be

---

[1] Research partially supported by National Science Foundation grant MCS83-01812.

inconsistent. Furthermore, a procedure based on a heteroscedastic estimator is in error by $O(1/n)$. I am sure both points are right, but do not quite see their force as arguments against the bootstrap. Probably Wu did not intend them as such.

For some discussion of the role of the model in the bootstrap, see Peters and Freedman (1984). For a mathematical demonstration that, in certain classes of heteroscedastic models, the bootstrap is asymptotically correct, see Freedman ((1981), Section 3; (1984), Section 3). For some simulation results in the same direction, see Daggett and Freedman (1985) and Freedman and Peters (1984a, b).

I turn now to the jackknife and consider only the simplest model, $Y_i = \alpha + \delta_i$. In this case, the "delete-one" jackknife estimate for α is the mean, and the jackknife standard error is just the familiar standard error for the mean. Since the regression model is balanced, there seems to be no difference between weighted and unweighted procedures.

I will construct independent δ_i's with expected value 0, but not identically distributed; their variances will approach 1, so the conditions (C) of Section 5 are satisfied. Still, the jackknife will not be such a good guide to the asymptotic distribution of the estimator. (Whether this example can be handled by the bootstrap depends on what information about the error structure is made available to the statistician: iid resampling would be a mistake.)

Suppose

$$\delta_i = 1/i \quad \text{with probability } 1 - 1/i^2,$$

$$\delta_i = 1/i - i \quad \text{with probability } 1/i^2.$$

Asymptotically,

$$\delta_1 + \cdots + \delta_n = \log n + U,$$

$$\delta_1^2 + \cdots + \delta_n^2 = V,$$

where U and V are random variables with a proper joint distribution. The sample mean is asymptotic to

$$(\log n + U)/n,$$

while the jackknife standard error is

$$\sqrt{V}/n$$

In particular, the jackknife misses the dominant bias term, $(\log n)/n$; and provides only erratic scaling for the U/n term.

Wu's Theorem 5 is not contradicted. The variance of the sample mean is $1/n + O(1/n^2)$, as is the expected value of the jackknife variance. However, the sample mean and jackknife variance have longish tails, so their asymptotic behavior is not well described by moments. (The formulas are a bit confusing, because it is standard deviation rather than variance that indicates scale.)

In this example, the sample mean is "super-efficient," but that is not essential, as I will now show. In essence, the jackknife estimates the scale of S by $\sqrt{V}$, where

$$S = (\delta_1 + \cdots + \delta_n)/\sqrt{n},$$
$$V = (\delta_1^2 + \cdots + \delta_n^2)/n.$$

Suppose the δ_i's are independent, with mean 0 and variance 1, but not identically distributed. Then $E(S) = 0$ and

$$\operatorname{Var} S = E(V) = 1,$$

the last display being the essence of Wu's theorem in our special case.

I will now sketch a construction to show that along some subsequences, S and V will converge, but $S/\sqrt{V}$ will be very large and positive, with probability about $1/e$; along other subsequences, $S/\sqrt{V}$ will be very large and negative, with the same probability; along still other subsequences, $S/\sqrt{V}$ tends to $N(0,1)$, only the last being desired by jackknife proponents.

Of course, S and V depend on n; the sequence of joint distributions is tight, and includes as limits all pairs of random variables S^* and V^* that can be represented in the following way. Let X_t be a process with stationary, independent increments, such that $E(X_t) = 0$ and $\operatorname{Var}(X_t) = t$. Then $S^* = X_1$ while V^* is the quadratic variation of X over the unit interval $[0,1]$.

In particular, the limits include all pairs S^* and V^* that can be represented in the form

$$S^* = W_1 + \cdots + W_N,$$
$$V^* = W_1^2 + \cdots + W_N^2,$$

where N is Poisson with parameter λ, the W_i's are independent and identically distributed with mean 0 and finite variance, independent of N, with $\lambda \operatorname{Var}(W_i) = 1$.

Now choose $\varepsilon > 0$ but small. Let

$$W_i = \varepsilon \quad \text{with probability } 1 - \varepsilon,$$
$$W_i = -(1 - \varepsilon) \quad \text{with probability } \varepsilon.$$

So $E(W_i) = 0$ and $\operatorname{Var}(W_i) = \varepsilon(1 - \varepsilon)$. Choose λ so $\lambda \operatorname{Var}(W_i) = 1$. Let N^+ and N^- be the number of positive and negative W_i's, respectively. These are independent Poisson variables, with parameters $\lambda(1 - \varepsilon) \approx 1/\varepsilon$ and $\lambda\varepsilon \approx 1$, respectively. Now

$$S^* = N^+\varepsilon - N^-(1 - \varepsilon),$$
$$V^* = N^+\varepsilon^2 + N^-(1 - \varepsilon)^2.$$

If $N^- = 0$, which has probability about $1/e$,

$$S^*/\sqrt{V^*} \approx \infty.$$

Otherwise,

$$S^*/\sqrt{V^*} \approx (1 - N^-)/\sqrt{N^-}.$$

Of course, the signs are easily reversed; the normal limit is the common one, and interweaving the three cases presents only the usual difficulties.

A finite-sample version of this disaster is even easier to arrange. For example, if $1 \le i \le n$, let

$$\delta_i = 0 \quad \text{with probability } 1 - \lambda/n,$$

$$\delta_i = \sqrt{n}\, W_i \quad \text{with probability } \lambda/n.$$

The moral seems clear: Second-order moment conditions or no, with skew long-tailed errors that change from observation to observation, the jackknife cannot be relied upon. On the other hand, preliminary calculations suggest that in our special case, with independence, in the presence of Lindeberg's condition both the jackknife and the bootstrap will perform adequately. Wu admits dependent errors, and this introduces further complications.

REFERENCES

DAGGETT, R. and FREEDMAN, D. (1985). Econometrics and the law: a case study in the proof of antitrust damages. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **1** 123–172. Wadsworth, Monterey, Calif.

FREEDMAN, D. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228.

FREEDMAN, D. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. *Ann. Statist.* **12** 827–842.

FREEDMAN, D. and PETERS, S. (1984a). Bootstrapping a regression equation: some empirical results. *J. Amer. Statist. Assoc.* **79** 97–106.

FREEDMAN, D. and PETERS, S. (1984b). Bootstrapping an econometric model: some empirical results. *J. Bus. Econ. Statist.* **2** 150–158.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

MALAY GHOSH[1]

University of Florida

Professor Wu is to be complimented for bringing out several important issues on jackknife, bootstrap and other resampling methods in regression analysis. Using a representation of the full-data least-squares estimator as a weighted average of corresponding least-squares estimators for appropriately chosen subsets, he has been able to motivate very successfully general-weighted jackknife in regression. I agree with the author that a jackknife that allows for the deletion of an arbitrary number of observations at a time is more flexible than the delete-one jackknife. However, I will be surprised if, for estimating nonsmooth functions such as the median, a delete-d jackknife estimator will necessarily rectify the deficiency of a delete-one jackknife estimator.

Although $v_{J(1)}$ enjoys the same robustness property of $v_{H(1)}$ when the errors are independent, but not identically distributed, and the design matrix satisfies

---

[1] Research supported by National Science Foundation grant DMS-8218091.