

and \hat{v} are of the form $\sum_{i=1}^n c_i r_i^2$, where the constants c_i depend on X but not on the σ_i^2 . For the situation in Section 10, the vectors $\mathbf{c} = (c_1, c_2, \dots, c_{12})$ are

$$v_{J(1)}: \quad (0.67, 0.28, 0.12, 0.04, 0.00, 0.00, 0.01, 0.06, 0.08, 0.06, 0.02, 0.46)$$

$$v_{H(1)}: \quad (0.52, 0.28, 0.13, 0.04, 0.01, 0.00, 0.02, 0.09, 0.09, 0.06, 0.02, 0.12)$$

$$\hat{v}: \quad (0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11, 0.11)$$

This raises several interesting questions, some of which are considered in the MINQUE literature:

- (i) For a given vector of variances $(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$, what is the best choice of the vector \mathbf{c} , say to minimize rms error?
- (ii) Given a set of possible variance vectors, is there a preferred general choice of \mathbf{c} ?
- (iii) Is there an adaptive way of selecting \mathbf{c} from the observed data, as suggested at the end of remark D?
- (iv) Wu's estimators $v_{J,r}$, $r > 1$, involve quadratic forms $\sum_i \sum_j c_i c_j r_i r_j$. Is there any real advantage to using the cross-terms $r_i r_j$, or does this just add to the variability of the estimator?

REFERENCES

- EFRON, B. and TIBSHIRANI, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other methods of statistical accuracy (with discussion). *Statist. Sci.* **1** 54–77.
- EFRON, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. SIAM, Philadelphia.
- REID, N. (1981). Estimating the median survival time. *Biometrika* **68** 601–609.

DEPARTMENT OF STATISTICS
SEQUOIA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

JOSEPH FELSENSTEIN

University of Washington

The emphasis the author places on using nonstandard subset sizes in jackknife procedures is important for at least one other reason. In some applications of resampling methods we are estimating an entity that lives in a space in which extrapolation is essentially impossible. Wu's equation (4.4) takes the estimate obtained from analysis of a sample and extrapolates its deviation from the overall estimate. However, if the space does not admit of extrapolation, then a choice of a subset size of $(n + k - 1)/2$ eliminates the extrapolation entirely. The value of $\tilde{\beta}_s$ is then the same as $\hat{\beta}_s$. By resampling and collecting a set of values of $\hat{\beta}_s$ we get variation that we take to be typical of the sampling variation in the estimate.

The example that brings this to mind is placing confidence intervals on phylogenies (evolutionary trees), to which I have applied a bootstrap technique

(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. FREEDMAN¹

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

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