involves inferring the variability in an unknown distribution from which your data were drawn by resampling from the

can be inferred by computing the variance of this collection of t^* values, and the confidence limits on the parameter

data. Suppose that you had data points $x_1, x_2, \ldots, x_n$, which you are willing to assume were drawn independently from the same distribution. From these, applying some method T of statistical es-

can be approximated by using the appropriate upper and/or lower percentiles of the observed distribution of the t^* values.

The justification for this resampling is that if the original sample size n is large

timation we obtain an estimate

each possible value of x will be repre-

$$t = T(x_1, x_2, \ldots, x_n) \qquad (1)$$

sented in the same proportion as in the underlying distribution, and resampling

site to be independently drawn from a distribution with 4^B possibilities whose

characters. I have recently discussed (Felsenstein, 1983a) some of the statistical

probabilities depends on the whole sequence, involved in choosing characters

be analyzed to obtain an estimate of the phylogeny. We then have r phylogenies. Each of these is a complicated multi-

they are both present in the true tree. But at least they cannot be contradictory: each being present on at least 95% of the

may also have branch lengths. Defining a confidence interval and summarizing it in a useable form is far from a simple matter.

In bootstrapping, confidence limits on a statistic are frequently constructed by the percentile method, which involves simply taking (for a 95% confidence in-

occur on at least one of the trees and must thus be either nested or disjoint.

The same argument has been used by Margush and McMorris (1981) to define "majority rule" consensus trees. These are trees composed of all those subsets that appear in a majority of a collection of trees. By the argument just given, these subsets must define a tree, since no two

TABLE 1. Fossil horse data of Camin and Sokal (1965). The states of each character are in a linear series. $-1, 0, 1, 2, \ldots,$ with the ancestral state being 0. The data are also shown in binary recoded form in which the nine multistate characters have been recoded into 20 binary factors. The first line of that table indicates the correspondence between the original and recoded characters. Bootstrap sampling of characters should be done before any recoding into binary factors.

Name				Characters						11112	22333	44566	77889
Mesohippus	0	0	0	0	0	0	0	0	0	00000	00000	00000	00000
Hypohippus	−1	3	3	0	0	0	0	0	1	00011	11111	00000	00001
Archaeohippus	1	0	0	0	0	0	0	1	1	10000	00000	00000	00101
Parahippus	1	1	1	1	0	0	0	0	1	10001	00100	10000	00001
Merychippus	2	2	2	2	1	1	1	2	1	11001	10110	11110	10111
Merych. secundus	2	2	2	2	1	−1	−1	2	1	11001	10110	11101	01111
Nannippus	2	2	1	2	1	1	1	2	1	11001	10100	11110	10111
Neohipparion	2	3	3	2	1	1	1	2	1	11001	11111	11110	10111
Calippus	2	2	1	2	1	−1	−1	2	1	11001	10100	11101	01111
Pliohippus	3	3	3	2	1	−1	−1	2	1	11101	11111	11101	01111

One difficulty in the interpretation of the result is that we may not have decided which subset of species interests us until after the bootstrap result is examined. This raises the "multiple tests" problem:

one would have to engage in an extrapolation to make their variance larger. The difficulty in envisaging a procedure like this is that the space of possible phylogenies does not lend itself readily to ex-

if we have 20 statistical tests, on average one should show significance at the 05%

trapolation: once a branch length has shrunk to zero it is not immediately ob-

dropped from the analysis. A weight of w means that the character is counted as if present w times, so that each change of state in the character is counted as if it were w changes of state.

This automatically accomplishes the duplication and deletion of characters without the necessity of recopying the data matrix. Different bootstrap samples could be fed into the programs by doing computer runs with different weights. The weights are generated by starting with

gram package PHYLIP, available free from me (see the Appendix below).

An Example

Table 1 shows the fossil horse data given by Camin and Sokal (1965 pp. 321–322) as a computational example. The full list of species and references for the original data are given by Camin and Sokal (1965). The data set has ten species and nine multistate characters. *Mesohippus* has been taken as the outgroup, as it

weights of zero for all characters. We then sample n characters at random with re-

was in Camin and Sokal's paper.

Figure 1 shows the results of running

placement (using a table of random numbers for example). Each time a character

a branch-and-bound program that finds all most parsimonious trees according to
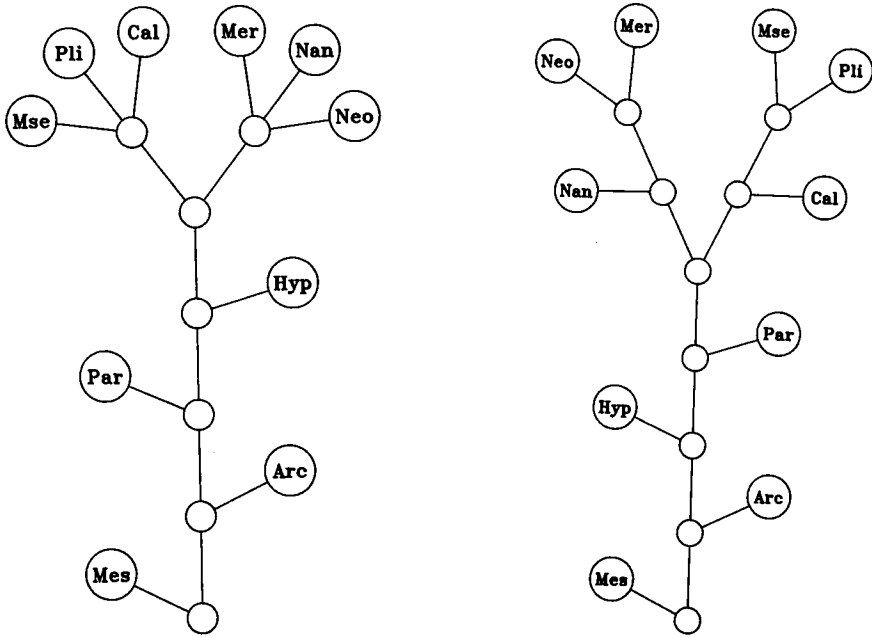
FIG. 1. All most parsimonious trees for the fossil horse data in Table 1 when phylogenies are evaluated by the Wagner parsimony criterion. There are ten most parsimonious trees in all. Nine of these can be generated by resolving each of the trifurcations in the left tree into all three possible bifurcations. The

Merychippus secundus.

characters is less than 0.05. This is easily

is that it provides a practical method, albeit a flawed one, for assessing the un-certainty inherent in such conclusions. I

———. 1982. The jackknife, the bootstrap, and other resampling plans. CBMS-NSF Regional Conference Series in Applied Mathematics No.