Estimation of Allocation Rates in a Cluster Analysis Context

K. E. Basford; G. J. McLachlan

Journal of the American Statistical Association, Vol. 80, No. 390. (Jun., 1985), pp. 286-293.

Estimation of Allocation Rates in a Cluster Analysis Context

K. E. BASFORD and G. J. McLACHLAN*

A sample of multivariate observations is assumed to be drawn from a mixture of a given number of underlying populations. The mixture likelihood approach to clustering is used to allocate each individual in the sample to its population of origin on the basis of the estimated posterior probabilities of population membership. Estimation of the correct allocation rate is considered for individual populations as well as for the overall mixture by averaging functions of the maximum of these posterior probabilities. The estimates of the correct allocation rates provide a means of assessing the performance of the mixture approach to clustering. The bootstrap technique is investigated for its effectiveness in reducing the bias of the estimates so obtained. Results are reported for three real data sets and a simulation study. It is demonstrated that the proposed estimates generally provide useful information on the unobservable allocation rates of the mixture approach. Encouraging results are obtained for the bootstrap method of bias correction applied to the estimates of the individual and overall allocation rates.

KEY WORDS: Allocation rate, correct; Mixture maximum likelihood method; Posterior probability estimate; Bias correction; Bootstrap method.

1. INTRODUCTION

Consider a sample of p-dimensional observations, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, where each may belong to one of several populations. Allocation of this sample is considered in a cluster analysis context in which it is not known from which population an observation comes. Under the so-called mixture maximum likelihood (ML) approach to clustering, the sample is assumed to have been drawn from a mixture of a specified number of populations in varying proportions. By adopting some parametric form for the density function in each underlying population, a likelihood can be formed in terms of the mixture density, and the unknown parameters can be estimated by the likelihood principle. An allocation rule based on the estimated posterior probabilities can then be found for assigning the observations to the various populations. The properties of the mixture approach have been considered by Day (1969), Wolfe (1970), Hosmer (1973), O'Neill (1978), Ganesalingam and McLachlan (1978, 1980a,b, 1981), Aitkin (1980), Mezzich and Solomon (1980), Aitkin et al. (1981), Symons (1981), Everitt and Hand (1981), and McLachlan (1982), among others. Hawkins et al. (1982) strongly supported the increasing emphasis on such an approach to clustering because it is model based. As they pointed out, most other methods of cluster analysis are based "on algorithms and criteria in the belief that intuitively reasonable criteria should produce good results over a wide range of possible [and generally unstated] models" (p. 353).

With any approach to clustering, there is the problem of assessing its performance. For the particular model under study, the superpopulation is assumed to be a mixture of a specified number of populations. With respect to a given population in the mixture, the correct allocation rate is defined to be the proportion of the observations from that population in the sample correctly allocated. In a cluster analysis context, the origin of each observation is unknown, and so the estimation of these proportions is a difficult task. This contrasts with the situation in discriminant analysis in which the origin of each observation is known. There, these proportions are observable and are referred to as the apparent allocation rates. In the discriminant analysis context, these rates are used to estimate the performance of the allocation rule when applied to subsequent observations of unknown origin.

There are few available results in the literature on the present objective of allocating an unclassified sample at hand and obtaining estimates of the correct allocation rates to assess the performance of the clustering method. A technique based on averaging appropriate functions of the maximum of the posterior probabilities was suggested by Ganesalingam and McLachlan (1980b) to estimate the overall error rate of the mixture approach. It worked well on a real data set—consisting of a mixture of two bivariate normal populations—reported in an unpublished Ph.D. dissertation by Ganesalingam. The present investigation examines the performance of such estimators not only for the overall allocation rate but also for the individual population allocation rates. In addition to the aforementioned bivariate example, two real data sets of more than two dimensions are considered, in which the number of underlying groups is greater than two. For these data sets the origin of each observation is known, and so the estimates of the allocation rates for the individual populations and the superpopulation can be compared with the true allocation rates. A simulation study is also conducted.

As the bias associated with the proposed method of estimation can be quite large, attention is focused on the problem of correcting the resulting estimates for bias by using the bootstrap technique. Some other methods of bias correction for this problem are briefly discussed.

2. MIXTURE APPROACH

Let Π denote the superpopulation, consisting of g populations $\Pi_1, \ldots, \Pi_g$ in some proportions $\pi_1, \ldots, \pi_g$; and let $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_g)'$. The density of an observation $\mathbf{x}$ in Π_i is given by $f_i(\mathbf{x}; \boldsymbol{\nu})$, where $\boldsymbol{\nu}$ denotes the vector of all unknown param-

---

* K. E. Basford is Biometrician, Department of Agriculture, and G. J. McLachlan is Associate Professor, Department of Mathematics, both at University of Queensland, St. Lucia 4067, Australia. The authors wish to thank the associate editor and the referees for their helpful comments.

Journal of the American Statistical Association
June 1985, Vol. 80, No. 390, Applications

eters in the population densities. Under the assumption that the sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$ has been drawn from Π, the likelihood is given by

$$L(\mathbf{x}_1, \ldots, \mathbf{x}_n; \boldsymbol{\nu}, \boldsymbol{\pi}) = \prod_{j=1}^{n} \left\{ \sum_{i=1}^{g} \pi_i f_i(\mathbf{x}_j; \boldsymbol{\nu}) \right\}. \quad (1)$$

Let $\hat{\boldsymbol{\phi}}$ be the estimate of $\boldsymbol{\phi} = (\boldsymbol{\nu}', \boldsymbol{\pi}')'$ obtained by likelihood estimation. Then each $\mathbf{x}_j$ $(j = 1, \ldots, n)$ can be allocated on the basis of the estimated posterior probabilities. The posterior probability that $\mathbf{x}_j$ belongs to population Π_r is given by

$$\theta_r(\mathbf{x}_j; \boldsymbol{\phi}) = \pi_r f_r(\mathbf{x}_j; \boldsymbol{\nu}) \Big/ \sum_{i=1}^{g} \pi_i f_i(\mathbf{x}_j; \boldsymbol{\nu}) \quad (2)$$

for $r = 1, \ldots, g$; and $\mathbf{x}_j$ is assigned to Π_r if

$$\theta_r(\mathbf{x}_j; \hat{\boldsymbol{\phi}}) > \theta_i(\mathbf{x}_j; \hat{\boldsymbol{\phi}}), \qquad i = 1, \ldots, g; \quad i \neq r. \quad (3)$$

For convenience $\theta_i(\mathbf{x}_j; \hat{\boldsymbol{\phi}})$ is denoted henceforth by $\hat{\theta}_{ij}$. If $\boldsymbol{\phi}$ were known, the allocation rule (3) would be the optimal or Bayes's rule (Anderson 1958), which maximizes the overall correct allocation rate.

The likelihood equation for $\boldsymbol{\phi}$, $\partial \log L / \partial \boldsymbol{\phi} = \mathbf{0}$, can be expressed as

$$\sum_{i=1}^{g} \sum_{j=1}^{n} \hat{\theta}_{ij} \frac{\partial}{\partial \hat{\boldsymbol{\nu}}} \log f_i(\mathbf{x}_j; \hat{\boldsymbol{\nu}}) = \mathbf{0} \quad (4)$$

and

$$\hat{\pi}_i = \sum_{j=1}^{n} \frac{\hat{\theta}_{ij}}{n}, \qquad i = 1, \ldots, g. \quad (5)$$

The computation is facilitated to a great extent by identifying these equations with the application of the EM algorithm of Dempster et al. (1977; see also Orchard and Woodbury 1972). For each $\mathbf{x}_j$ $(j = 1, \ldots, n)$, let the vector of indicator variables, $\boldsymbol{\gamma} = (\gamma_{1j}, \ldots, \gamma_{gj})'$, be defined by

$$\begin{aligned} \gamma_{ij} &= 1, & \mathbf{x}_j &\in \Pi_i, \\ &= 0, & \mathbf{x}_j &\notin \Pi_i; \end{aligned}$$

the expectation of γ_{ij} conditional on $\mathbf{x}_j$ is equal to θ_{ij}. Then it can be easily confirmed that (4) and (5) are obtained by differentiation of the expectation of the complete-data log-likelihood conditional on $\mathbf{x}_1, \ldots, \mathbf{x}_n$. This conditional expectation is effected here by replacing each indicator variable γ_{ij} by its expected value conditional on $\mathbf{x}_j$, θ_{ij}. Given some initial estimate of $\boldsymbol{\phi}$ in forming the $\hat{\theta}_{ij}$, a new estimate $\hat{\boldsymbol{\phi}}$ can be computed from (4) and (5), substituted back into the $\hat{\theta}_{ij}$ to yield a new estimate of $\boldsymbol{\phi}$, and so on. The process continues iteratively with the likelihood increasing at each step, until convergence for a sequence bounded above. An account of the convergence properties of the EM algorithm was given by Wu (1983). Generally, the convergence is slow but may be improved by using Aitken's acceleration process; see Louis (1982) for details of speeding up this algorithm.

For normal densities with unequal covariance matrices the likelihood is unbounded, since each data point gives rise to a singularity on the edge of the parameter space. It is known (Kiefer 1978), however, that there is a sequence of roots of the likelihood equation given by (4) and (5) that is consistent and asymptotically efficient for $\boldsymbol{\phi}$. With probability tending to 1, these roots correspond to local maxima of the likelihood. With mixture models the likelihood generally has multiple maxima, and so there is the problem of which root to choose. It does not necessarily follow that the estimator constructed by choosing the root that corresponds to the largest of the local maxima is consistent (Lehmann 1980). Ideally, we would like to have available a $\sqrt{n}$-consistent estimator of $\boldsymbol{\phi}$ with which to start the EM algorithm. This is possible with the data sets considered here, since we know the true origin of each observation, so we can use as our starting point the estimate of $\boldsymbol{\phi}$ formed using this knowledge. In practice, this information is unknown; so the EM algorithm must be repeated for several different sets of starting values of $\boldsymbol{\phi}$. (On the choice of suitable starting values, see Quandt and Ramsey 1978 and the subsequent discussion—Fowlkes 1979, Everitt and Hand 1981, and Basford and McLachlan 1984.)

The mixture approach is usually applied with the $f_i(\mathbf{x}; \boldsymbol{\nu})$ assumed to be multivariate normal densities with different mean vectors and perhaps different covariance matrices. This assumption can be checked to some extent by using the test for multivariate normality and homoscedasticity as described by Hawkins (1981). Since the true origin of each observation is unknown, the sample must be clustered first by using the mixture approach for multivariate normal densities with unequal covariance matrices. The test is then applied to these clusters as if they were the true groups with no misallocations. This is a rather crude approach, but according to Fatti et al. (1982), it appears to work fairly well. In the examples to be discussed, we were able to apply Hawkins's method directly, as we knew the true groups.

3. ESTIMATORS OF ALLOCATION RATES

Although the mixture approach may not give accurate estimates of the posterior probabilities, $\hat{\theta}_{ij}$, for the observations in the sample, it may still provide a satisfactory clustering of the data (Ganesalingam and McLachlan 1979). The case study of Hernandez-Avila (1979) suggests that the mixture approach applied, assuming multivariate normal densities, may well be reasonably robust from the cluster analysis viewpoint of separating samples in the presence of multimodality.

We wish to consider here, however, the situation in which n is sufficiently large, relative to the number of unknown parameters, for the $\hat{\theta}_{ij}$ to be regarded as reliable estimates of the posterior probabilities. In this situation, if the maximum of $\hat{\theta}_{ij}$ over $i = 1, \ldots, g$ is near to 1 for most of the observations $\mathbf{x}_j$, we can be confident that the mixture approach can cluster the sample at hand with a high degree of certainty. Conversely, if this maximum is generally well below 1, then the underlying populations are too close together for the sample to be clustered with any certainty. It would be informative, therefore, to have available a quick summary statistic for assessing the overall performance of the mixture approach in clustering the sample at hand. To this end, we consider the estimation of the correct allocation rates associated with an application of the mixture approach.

Let P_i denote the correct allocation rate with respect to population i ($i = 1, \ldots, g$) when the sample is allocated to $\Pi_1, \ldots, \Pi_g$ on the basis of rule (3); that is, $P_i = \#\{j : \mathbf{x}_j \in \Pi_i$ and $\mathbf{x}_j$ allocated to $\Pi_i\} / \#\{j : \mathbf{x}_j \in \Pi_i\}$. The overall correct allocation rate is

$$P = \sum_{i=1}^{g} \frac{n_i P_i}{n}, \tag{6}$$

where n_i is the number of observations from Π_i in the sample. This can be estimated by

$$T = \sum_{j=1}^{n} \max_r \frac{\hat{\theta}_{rj}}{n}. \tag{7}$$

Our reasoning for using T as an estimator of P follows from the ideas developed in a discriminant analysis context by Fukunaga and Kessel (1972, 1973), Lissack and Fu 1976, Glick (1978), Moore et al. (1976), and Schwemer and Dunn (1980). In the case of known $\boldsymbol{\phi}$, T [using θ_{rj} instead of $\hat{\theta}_{rj}$ in (7)] would be a consistent and unbiased estimator of the correct allocation rate for rule (3) applied to a subsequent observation randomly chosen from the superpopulation, and P tends in probability to this latter rate as $n \rightarrow \infty$.

It was decided to estimate the individual allocation rates P_i ($i = 1, \ldots, g$) by

$$T_i = \sum_{j=1}^{n} \max_r \hat{\theta}_{rj} \frac{I_i(\mathbf{x}_j)}{n\hat{\pi}_i}, \tag{8}$$

where

$$\begin{aligned} I_i(\mathbf{x}_j) &= 1, \quad \hat{\theta}_{ij} > \hat{\theta}_{rj}, \quad r = 1, \ldots, g, \quad r \neq i, \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{9}$$

It can be verified without difficulty that providing $\hat{\boldsymbol{\phi}}$ is consistent, $T - P$ and $T_i - P_i$ converge in probability to 0 as $n \rightarrow \infty$.

As explained in the introduction, some method of bias correction should be considered with applications of T and T_i ($i = 1, \ldots, g$), although it may involve considerable computation. Moreover, as explained by Efron (1982), even if correction for bias is not undertaken, it can still be of interest to compute the estimated bias of an estimator, along with an estimate of its standard deviation (SD) or root mean square error (RMSE). If the estimated bias is less than, say, $\frac{1}{4}$ of the estimated SD or RMSE, then bias is probably not a serious issue. Note, however, that as the quantities being estimated here are not parameters but random variables, the MSE does not equate simply to the variance plus the bias squared. Considering T_i as an estimator of P_i, we find that

$$\begin{aligned} \mathrm{MSE}(T_i) &= \mathrm{var}(T_i) + \{\mathrm{bias}(T_i)\}^2 \\ &\quad + \mathrm{var}(P_i) - 2\,\mathrm{cov}(T_i, P_i). \end{aligned}$$

4. BIAS CORRECTION OF THE ESTIMATED ALLOCATION RATES

We consider now the proposed bootstrap method of Efron (1979) to correct the estimates T and T_i ($i = 1, \ldots, g$) for bias (for a full account of the application of the bootstrap in a general context see also Efron 1982, 1983, and Efron and Gong 1983). It may be applied as follows:

Step 1. A new set of random variables, $(\tilde{\boldsymbol{\gamma}}_1', \tilde{\mathbf{x}}_1')', \ldots, (\tilde{\boldsymbol{\gamma}}_n', \tilde{\mathbf{x}}_n')'$, is generated from the distribution of $(\boldsymbol{\gamma}', \mathbf{x}')'$ with the estimates $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\nu}}$ used in place of the unknown $\boldsymbol{\pi}$ and $\boldsymbol{\nu}$. This can be achieved by generating for each j ($j = 1, \ldots, n$) a random variable that takes the values $1, \ldots, g$ with probabilities $\hat{\boldsymbol{\pi}}_1, \ldots, \hat{\boldsymbol{\pi}}_g$. If the generated value is equal to r, say, then $\tilde{\boldsymbol{\gamma}}_{ij} = 1$ for $i = r$ and $= 0$ otherwise. The observation $\tilde{\mathbf{x}}_j$ is then generated from the density $f_r(\mathbf{x}; \tilde{\boldsymbol{\nu}})$. This implies that $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n$ is a random sample from the mixture density

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{g} \hat{\pi}_i f_i(\mathbf{x}; \hat{\boldsymbol{\nu}}). \tag{10}$$

Note that unlike the indicator vectors $\boldsymbol{\gamma}_j$ associated with the original observations $\mathbf{x}_j$, the bootstrap indicators $\tilde{\boldsymbol{\gamma}}_j$ are known.

Step 2. The mixture approach is applied to the bootstrap sample $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n$ to produce new estimates of $\boldsymbol{\pi}$ and $\boldsymbol{\nu}$, say $\tilde{\boldsymbol{\pi}}$ and $\tilde{\boldsymbol{\nu}}$, knowledge of the $\tilde{\boldsymbol{\gamma}}_j$ is not used in this step.

Step 3. Each estimated allocation rate ($\tilde{T}_i$) for the clustering of the bootstrap sample ($\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n$) on the basis of (3), now using $\tilde{\boldsymbol{\pi}}$ and $\tilde{\boldsymbol{\nu}}$, is computed along with the true rate ($\tilde{P}_i$) as determined from the $\tilde{\boldsymbol{\gamma}}_j$.

Step 4. The expectation of the difference $d_i = \tilde{T}_i - \tilde{P}_i$ with respect to the bootstrap distribution given by (10) is referred to as the bootstrap bias, say b_i. It is approximated by $\bar{d}_i$, obtained by averaging d_i over K repeated realizations of samples from $\hat{f}(\mathbf{x})$; that is, $\bar{d}_i = \Sigma_{k=1}^{K} (d_{ik}/K)$, where $d_{ik} = \tilde{T}_{ik} - \tilde{P}_{ik}$ denotes the value of $\tilde{T}_i - \tilde{P}_i$ obtained on the kth bootstrap replication. The standard error of the Monte Carlo approximation, $\bar{d}_i$, to the bootstrap bias, b_i, is calculated as the positive square root of

$$\sum_{k=1}^{K} \frac{(d_{ik} - \bar{d}_i)^2}{K(K - 1)}.$$

The RMSE of T_i is estimated by the positive square root of

$$\sum_{k=1}^{K} \frac{(\tilde{T}_{ik} - \tilde{P}_{ik})^2}{K}.$$

Similarly, the bootstrap estimates of the bias and RMSE of T can be formed with respect to its estimation of the overall allocation rate, P, and we use b to denote the bootstrap bias and $\bar{d}$ its Monte Carlo approximation.

Step 5. The bias-corrected estimates of the individual and overall allocation rates are given by $T_i - \bar{d}_i$ ($i = 1, \ldots, g$) and $T - \bar{d}$, respectively.

This is the parametric version of the bootstrap method. One could consider a semiparametric version, using a different method of generating the bootstrap sample in step 1 but with steps 2–5 the same as previously outlined. In step 1, a sample of size n would be drawn with replacement from the observations $\mathbf{x}_1, \ldots, \mathbf{x}_n$, with each observation given equal weight $1/n$. The origin of each chosen observation $\mathbf{x}_j$ would be determined then in accordance with its estimated posterior probability, $\hat{\theta}_{ij}$, of belonging to the ith underlying population ($i = 1, \ldots, g$). That is, a random variable that takes the values $1, \ldots, g$ with probabilities $\hat{\theta}_{1j}, \ldots, \hat{\theta}_{gj}$, respectively, would be generated; the value of this generated random variable determines the (bootstrap) population of origin of the observation $\mathbf{x}_j$.

The number of bootstrap replications K used in forming the bootstrap estimates here was limited to 50 for economical considerations. In various other applications of the bootstrap, Efron (1979, 1981a,b) noted that the choice of replication number does not seem to be critical past 50 or 100. In addition, for some simulations performed, a components-of-variance argument (described in Sec. 7) indicated that $K = 50$ was adequate.

5. APPLICATION

The estimators proposed in Section 3 for assessing the performance of the mixture ML method of clustering are investigated at first by using three real data sets. As the true origin of each observation is known, the estimates of the allocation rates can be compared with the true allocation rates for each population and for the superpopulation. In each of these examples the estimates—and consequently the cluster—associated with population Π_i are taken to be those corresponding to $\bar{\mathbf{x}}_i$ and $\mathbf{S}_i$, and n_i/n (the usual estimates of $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$, and π_i when the origin of the data is known), as their initial values in the application of the EM algorithm. The results obtained by the parametric version of the bootstrap are discussed first. The corresponding results for the semiparametric version are considered at the end of this section.

Example 1

The data analyzed were taken from Habbema et al. (1974), in which in the context of genetic counseling, the question of discriminating between normal women and hemophilia A carriers was considered on the basis of the two variables, $x_1 = \log_{10}$ (AHF activity) and $x_2 = \log_{10}$ (AHF-like antigen). Reference data containing $n_1 = 30$ observations on known noncarriers and $n_2 = 45$ observations on known obligatory carriers were available. Let π_1 be the population of noncarriers, or normals, and π_2 be the population of carriers.

Application of Hawkins's test for multivariate normality and homoscedasticity (Hawkins 1981) suggests that the populations are normal but somewhat heteroscedastic. Hence the mixture method of clustering was applied to the data, assuming it to be a sample of observations of unknown origin taken from a mixture of two bivariate normal populations with unequal covariance matrices. The mixture approach misallocated 3 and 12 observations from π_1 and π_2, respectively, when used to cluster the sample of 75 observations into two populations, so $P_1 = .9$, $P_2 = .733$, and $P = .8$. The estimates of the allocation rates as given by T_1, T_2, and T, respectively, are shown in Table 1, along with the bootstrap estimates of their RMSE's and biases. The standard errors given are of the latter Monte Carlo approximations, $\bar{d}_i$ ($i = 1, \ldots, g$) and $\bar{d}$, to the bootstrap biases, b_i ($i = 1, \ldots, g$) and b. The bias-corrected estimates of the allocation rates obtained as a consequence are also displayed in Table 1. It can be seen from this table that the bootstrap method is effective in reducing the bias of not only the overall allocation rate but also each individual population rate.

Note that for this example, the mixing proportions π_1 and π_2 have been introduced solely for the purpose of forming the mixture ML to produce a clustering of the sample. They do not represent the prior probabilities of being a noncarrier and carrier, respectively, since each woman has her own prior probability of being a noncarrier given by her genetic chance of being normal, which can be assessed from her pedigree.

Example 2

We consider here the well-known set of *Iris* data published by Fisher (1936). It consists of four measurements on 50 plants from each of three species of Iris: *Iris setosa*, *Iris versicolor*, and *Iris virginica*, denoted here by populations Π_1, Π_2, and Π_3, respectively. Many clustering techniques have been applied to this data set (e.g., see Kendall 1966, Friedman and Rubin 1967, and Scott and Symons 1971). The *setosa* plants have

Table 1. Estimation of Correct Allocation Rates for Real Data Sets by Using Bootstrap Method of Bias Correction

			Parametric Version				*Semiparametric Version*			
Population	*True Rate*	*Estimated Rate*	*Estimate of RMSE*	*Estimate of Bias*	*SE*	*Corrected Estimate*	*Estimate of RMSE*	*Estimate of Bias*	*SE*	*Corrected Estimate*
Example 1: Habbema's Hemophilia A Data										
Π_1	.900	.938	.113	.052	.017	.886	.052	.009	.007	.929
Π_2	.733	.902	.193	.090	.024	.812	.133	.060	.017	.842
Π	.800	.921	.120	.076	.013	.845	.071	.035	.009	.883
Example 2: Fisher's Iris Data										
Π_1	1.000	1.000	.000	.000	.000	1.000	.000	.000	.000	1.000
Π_2	.900	.985	.035	.013	.005	.971	.053	.021	.007	.964
Π_3	1.000	.986	.031	.012	.004	.974	.024	.004	.003	.981
Π	.967	.990	.013	.008	.002	.982	.017	.008	.002	.982
Example 3: Pigeon Pea Data										
Π_1	1.000	.998	.019	.007	.002	.991	.025	.009	.003	.989
Π_2	.800	.979	.023	.009	.003	.970	.055	.028	.007	.951
Π_3	.967	.981	.037	.015	.005	.966	.033	.009	.005	.972
Π	.922	.986	.018	.011	.002	.975	.024	.014	.003	.973

been easily isolated, but there is generally overlap between the other two species. The mixture approach was applied as though the data were a sample of 150 observations of unknown origin. The population densities were taken to be multivariate normal with unequal covariance matrices, as Hawkins's test indicated normality with heteroscedasticity (see Fatti et al. 1982, p. 64). It is worth noting, however, that some other analyses (Small 1980 and Royston 1983) have cast doubt on the normality of this data set. The mixture approach misallocated five plants from Π_2 to Π_3, thus giving $P_1 = 1.0$, $P_2 = .9$, and $P_3 = 1.0$ with an overall allocation rate of $P = .967$. The results for this example are displayed in Table 1.

Concerning the individual population rates, every bootstrap sample gave an estimate of the allocation rate for Π_1 equal to the true rate of 1.0, so the estimates of the RMSE and bias were all 0. As noted earlier, this population is quite distinct from the other two. The individual rate for Π_2 and the overall rate are closer to the true rates as a consequence of bias correction, although the corrected estimates have only partially removed the overoptimism. For Π_3 the corrected estimate is further than the estimate, T_3, from the true rate, P_3.

Example 3

The data considered here are from a pigeon pea variety trial conducted at Redland Bay in Southeast Queensland in 1977. Two of the lines were from International Crops Research Institute for the Semi Arid Tropics (ICRISAT), Hyderabad, India; these were compared with a line developed in the Agriculture Department of the University of Queensland. Observations were made on plant height, seed weight, and pod number for each of 30 plants for each line. The three lines—Royes, ICP28, and ICP26—were denoted by Π_1, Π_2, and Π_3, respectively. The populations were taken to be multivariate normal with unequal covariance matrices because of significant heteroscedasticity shown in Hawkins's test for multivariate normality and homoscedasticity. According to this test, there is some doubt about normality. The mixture approach misallocated six plants from ICP28 as ICP26 and one plant from ICP26 as Royes. This gave allocation rates of $P_1 = 1.0$, $P_2 = .8$, and $P_3 = .967$ with $P = .922$.

It can be seen from Table 1 that with one exception, bias correction moved the inital estimates closer to the true rates. The corrected estimate of the rate for Π_3 is almost equal to the true rate. Otherwise, the improvement in the estimated rates is limited. The one instance in which bias correction moved the estimate away from the true rate (albeit by a small amount) is similar to the exception in Example 2. In each case the true rate is equal to unity, and so contrary to the general trend, the estimated rate is not optimistic.

We have also listed in Table 1, for Examples 1–3, the results corresponding to the semiparametric version of the bootstrap method. Contrasting these results with those of the parametric version, it can be seen that for Example 1 the latter is clearly preferable, but for Example 2 it does not give better estimates than the semiparametric version, and for Example 3 there is little to choose between the two versions. It has been noted for Examples 2 and 3 that there is some doubt about the normality of the data, and this might explain the lack of superiority of the parametric over the semiparametric version.

6. SIMULATION RESULTS

To examine the estimation of the allocation rates in more detail, a simulation study was undertaken in which samples were drawn from a mixture of equal proportions of two bivariate normal populations with equal covariance matrices ($\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$). Without loss of generality it was assumed that $\boldsymbol{\mu}_1 = (\Delta, 0)'$, $\boldsymbol{\mu}_2 = \mathbf{0}$, and $\boldsymbol{\Sigma}$ is the identity matrix, where $\Delta = \{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\}^{1/2}$ is the Mahalanobis distance between Π_1 and Π_2. Four different combinations of Δ and n were considered (cases A, B, C, and D), and 40 simulation trials were generated in each case except case C, where the number of trials was limited to 20 (see Table 2). An IMSL subroutine based on the inverse method was used to generate normal random variables from uniformly distributed deviates, which were produced by a multiplicative congruential generator of the form $X_{i+1} \equiv rX_i$ (modulo s), where $r = 7^5$ and $s = 2^{31} - 1$. For the simulation study, attention was confined to the parametric version of the bootstrap method of bias correction.

Before discussing the summarized results of the simulation study (in Table 5), we shall go through the first trial of case A, reporting relevant values for the overall allocation rate, then through several trials to illustrate how the summary table was obtained. In Table 3, we list the values of $\tilde{T}$ and $\tilde{P}$ obtained on the first five replications using the bootstrap method to correct T for bias as an estimate of P; for this first trial, $T = .854$ and $P = .767$. As explained in Section 4, the bootstrap estimate of the bias of T is given by the Monte Carlo approximation $\bar{d}$, obtained by averaging the values of $d = \tilde{T} - \tilde{P}$ from the $K = 50$ bootstrap replications performed. For this first trial, $\bar{d} = .115$, leading to $T - \bar{d} = .739$ as the corrected estimate of P. The RMSE of T was estimated by the positive square root of the average over the $K = 50$ bootstrap replications of d^2, yielding $(.0226)^{1/2} = .150$ as the estimate of the RMSE.

The preceding exercise was repeated for each of 40 simulation trials. To illustrate this process, we list in Table 4 the first five individual trial values for the overall allocation rate in case A. The averages and their standard errors in the columns headed P, T, $T - \bar{d}$, and Estimate of RMSE in Table 4 correspond to the entries for True Rate, Estimated Rate, Corrected Estimate, and Estimate of RMSE, respectively, for case A in Table 5. In this table RMSE and RMSE* refer to the RMSE's of the estimated rate and its bias-corrected version, respectively, assessed by the positive square root of the averages over the simulated values of $(T - P)^2$, respectively. In the context of estimating the error rate of a prediction rule, Efron (1983) discussed the "ideal constant" estimate, which, for the estimation of P, is given by $T - \theta$, where θ is the expectation of $T - P$. In our simulations we used $\bar{T} - \bar{P}$—the value of $T - P$ averaged over the number of trials performed—in place

Table 2. Design Aspects of Simulation Study

Case	Δ	n	Number of Trials
A	$\sqrt{2}$	60	40
B	2	60	40
C	2	120	20
D	3	60	40

Table 3. The First Five Bootstrap Replications and Summary Statistics for $K = 50$ Bootstrap Replications for the Overall Allocation Rate on Trial 1 in Case A, for Which $T = .854$ and $P = .767$

Bootstrap Replication	$\tilde{T}$	$\tilde{P}$	$d = \tilde{T} - \tilde{P}$	d^2
1	.771	.700	.071	.005
2	.962	.817	.145	.021
3	.888	.800	.088	.008
4	.907	.833	.074	.005
5	.922	.733	.189	.036
50 (average)			.115	.023

NOTE: The entries for the averages of d and d^2 over $K = 50$ replications give the bootstrap estimates of the bias of T and its MSE on this trial.

of the unknown θ. For example, for the overall allocation rate in case A, we see from Table 4 that $\bar{T} - \bar{P} = .84 - .74 = .10$. The RMSE of the ideal-constant estimated rate is denoted by $\text{RMSE}^{(\text{IC})}$ in Table 5, where correlation refers to the correlation between the uncorrected estimated rate minus the true rate and the bootstrap estimator of the expectation of this difference.

Note that since the estimated overall allocation rate, T, can be expressed as $T = \hat{\pi}_1 T_1 + \cdots + \hat{\pi}_g T_g$, it lies within the range of the estimates of the individual rates, T_i. This does not necessarily apply, however, to the rates after they have been averaged over the number of simulation trials performed in each case. For example, in case A, it can be seen that $\bar{T} = .84$ lies outside the range of the means of the estimates of the individual rates, given by $\bar{T}_1 = .839$ and $\bar{T}_2 = .802$.

The results reported in Table 5 for the estimation of the allocation rates are very encouraging not only for the overall but also for the individual rates. In each of the four cases, the bootstrap method of bias correction appreciably reduced the optimistic assessment provided in the first place by the uncorrected estimates. Moreover, bias correction of the estimated rates reduced the RMSE of each estimated rate in every case. Indeed, if we compared the RMSE of the corrected estimated rate, RMSE*, with that of the ideal constant estimated rate, $\text{RMSE}^{(\text{IC})}$, it can be seen that progress was generally made in reducing the MSE to this level. In all but one instance, the mean corrected estimated rate is greater than the corresponding true rate, suggesting that the optimism of the uncorrected estimate was not completely eliminated.

Table 4. The First Five Trials and Summary Statistics for 40 Trials for the Overall Allocation Rate in Case A

Trial	T	P	$\bar{d}$	$T - \bar{d}$	Estimate of RMSE
1	.854	.767	.115	.739	.150
2	.889	.700	.058	.831	.089
3	.786	.700	.131	.655	.173
4	.890	.733	.067	.823	.115
5	.953	.850	.041	.912	.065
40					
Average	.840	.740	.078	.762	.117
SE	.014	.010	.006	.019	.007

NOTE: $T - \bar{d}$ is the bias corrected estimate of P formed from the $K = 50$ bootstrap replications performed on each trial.

Table 5. Summary Results for the Simulation Trials in Cases A, B, C, and D

Case	Π_1	Π_2	Π
A			
True Rate	.764	.697	.740
SE	.024	.031	.010
Estimated Rate	.839	.802	.840
SE	.021	.026	.014
Corrected Estimate	.767	.731	.762
SE	.024	.028	.019
Estimate of RMSE	.137	.146	.117
SE	.008	.007	.007
RMSE	.146	.194	.154
RMSE*	.134	.193	.147
$\text{RMSE}^{(\text{IC})}$	.124	.163	.118
Correlation	−.163	−.590	−.625
B			
True Rate	.760	.814	.792
SE	.029	.027	.012
Estimated Rate	.874	.904	.906
SE	.020	.011	.010
Corrected Estimate	.828	.857	.860
SE	.022	.014	.013
Estimate of RMSE	.111	.106	.079
SE	.008	.007	.005
RMSE	.170	.168	.144
RMSE*	.141	.149	.117
$\text{RMSE}^{(\text{IC})}$	.125	.143	.089
Correlation	.177	.087	−.210
C			
True Rate	.809	.817	.815
SE	.025	.017	.009
Estimated Rate	.867	.870	.874
SE	.015	.009	.009
Corrected Estimate	.829	.834	.833
SE	.017	.011	.012
Estimate of RMSE	.083	.081	.066
SE	.006	.005	.005
RMSE	.100	.092	.082
RMSE*	.086	.083	.068
$\text{RMSE}^{(\text{IC})}$	.081	.075	.057
Correlation	−.134	−.282	−.466
D			
True Rate	.925	.900	.916
SE	.009	.015	.007
Estimated Rate	.944	.940	.945
SE	.005	.005	.004
Corrected Estimate	.920	.914	.920
SE	.008	.008	.007
Estimate of RMSE	.061	.071	.046
SE	.005	.005	.004
RMSE	.044	.093	.049
RMSE*	.043	.083	.043
$\text{RMSE}^{(\text{IC})}$	.039	.084	.040
Correlation	.077	.239	.061

NOTE: Entries for the true rate, estimated rate, corrected estimate, and estimate of RMSE are averages over the simulation trials. SE = standard error. The root mean square errors of the estimated rate, the bootstrap corrected rate, and the ideal constant estimated rate are denoted by RMSE, RMSE*, and $\text{RMSE}^{(\text{IC})}$ respectively. The correlation is between the (uncorrected) estimated rate minus the true rate and the estimated bias.

With the bootstrap method of bias correction of the apparent error rate in a discriminant analysis context, Efron (1982, 1983) noted a marked negative correlation between the apparent error rate minus the true rate and the bootstrap estimator of the bias, which appreciably increases the MSE of the corrected rate. The results in Table 5 for cases A and C indicate that in particular, for the overall allocation rate in the present problem, the estimated rate minus the true rate is negatively correlated with the bootstrap estimator of the bias.

7. DISCUSSION OF BOOTSTRAP ESTIMATES

The estimation of the allocation rates associated with the application of a cluster analysis method is a rather ambitious undertaking in the absence of data of known origin. Since the mixture ML approach to clustering is model based, however, it has been demonstrated that estimates of the allocation rates can be formulated under this model in terms of the relative sizes of the estimated posterior probabilities. In the three real examples presented, these estimates without correction for bias provided generally useful information on the allocation rates of the mixture approach to clustering. Correction of these estimates for bias should, however, be considered, since it was noted in each of the examples reported that the estimator T gave too optimistic an estimate of the overall true allocation rate. In addition, the estimates of the rates for the individual populations were optimistic in situations in which this was possible (i.e., in which the true rates were less than unity). A comparison of the estimated bias with the estimated RMSE for each estimator in these examples clearly shows that bias is a serious issue here; the ratio of the estimated bias to the estimated RMSE generally well exceeded the suggested threshold of $\frac{1}{4}$.

With respect to the fully parametric version of the bootstrap method applied to the three real data sets, bias correction nearly always improved the estimates of the various allocation rates. The two exceptions occurred in situations in which the true rate was unity. The computation required to implement the procedure would appear to be justified. The results for the last two examples were not as promising as for Example 1. This is not surprising, since Examples 2 and 3 were chosen not because they are favorable to accurate estimation of the allocation rates, but rather because they represent situations in which estimation of the allocation rates is just a realistic exercise in view of the size of the sample relative to the number of unknown parameters.

Favorable results were obtained in each of the four cases of the simulation study performed with the parametric version of the bootstrap method. They suggest that in such situations, estimation of the allocation rates and subsequent bias correction according to the bootstrap method is a very worthwhile exercise.

The bootstrap method can be used to estimate the RMSE's of the estimators of the individual and overall allocation rates (reported in Table 1 for the three examples considered and in Table 5 for the simulation study). Concerning the parametric version, in Example 1 the individual allocation rates were within one (estimated) RMSE of the uncorrected estimates; the overall rate was just over one RMSE below the estimated rate. For Examples 2 and 3, however, the true rates were within two RMSE's of the uncorrected estimates for only four of the eight rates. For the simulations there was fairly good agreement between the estimated and true RMSE's, except for case B, in which, for instance, the estimated RMSE was only 55% of the simulated RMSE of the estimator of the overall allocation rate. Note that Efron (1982) made various suggestions about how the bootstrap method can be used formally to give nonparametric confidence intervals. However, the usefulness of these ideas in the present context has yet to be investigated.

Concerning the number of bootstrap replications K for computing the Monte Carlo approximations to the proposed bootstrap estimates here, it can be seen from the results for Examples 1–3 in Table 1 that it would be preferable to choose $K > 50$. This is because the standard error of the Monte Carlo approximation ($\bar{d}_i$) to the bootstrap estimate of the bias of T_i (b_i) is not always very small relative to $\bar{d}_i$. Of course as $K \rightarrow \infty$, the standard error of $\bar{d}_i \rightarrow 0$, but there is no point in taking K to be any larger than is necessary to ensure that the standard error of $\bar{d}_i$ is small relative to the SD of b_i. Some indication about how large K should be to achieve this, at least in situations similar to the four cases in the simulation study, can be obtained by a components-of-variance analysis of the simulated data on which Table 5 is based. It shows that as $K \rightarrow \infty$ from $K = 50$, the trial-to-trial SD of b_1 would decrease from .034, .024, .014, and .020 to .029, .018, .009, and .018 in cases A, B, C, and D, respectively; similarly, for b_2 and b. Hence in these four cases, $K = 50$ would appear to be adequate because there would be only a moderate reduction in the SD of b_i or b for a larger value of K.

Unless n is very large, the bias of the estimate of $\boldsymbol{\phi} = (\boldsymbol{\pi}', \boldsymbol{\nu}')'$—$\hat{\boldsymbol{\phi}}$—will not be small; and this bias will manifest itself in the estimates of the posterior probabilities. We are currently investigating the use of bias-corrected versions of the parameter estimates in forming the estimated posterior probabilities, with a view to improving the allocation of the sample. Bias correction of $\hat{\boldsymbol{\phi}}$ by using the bootstrap method would involve term-by-term additive correction of a covariance matrix, which is fraught with danger. This can be avoided, however, by an appropriate reparameterization of the terms that occur in the expressions for the posterior probabilities.

8. OTHER METHODS OF BIAS CORRECTION

Other methods, besides the bootstrap, for estimating bias and in some instances variance include the jackknife, the delta method, and cross-validation. Efron (1982) gave an excellent account of the available methods. In particular he examined the relationships between the methods and identified those situations in which estimates produced according to some of the methods are similar or indeed the same. In the regression context Efron (1982) showed that the jackknife, cross-validation, and the bootstrap are closely connected in theory if not necessarily in their practical consequences.

These competitors of the bootstrap, however, appear not to be applicable in a straightforward manner to our clustering problem in which we wish to reduce the biases of the T_i and T as estimators of the unobservable allocation rates P_i and P. For instance, consider the application of the nonparametric jackknife procedure to this problem: It obviously cannot be applied to the estimates themselves. In considering the estimation of T, the jackknife produces an estimate by extrapolating T to its value at $n = \infty$, T_∞. We require, however, an unbiased estimator not of T_∞, but of $E(P)$, which may not be close to T_∞ for small n, although it does equal T_∞ in the limit. Moreover, there does not appear to be a useful role for the jackknife in providing approximations to the bootstrap expectations of the statistics $\tilde{T}_i$

$- \tilde{P}_i$ and $\tilde{T} - \tilde{P}$. The complication with the nonparametric jackknife in our problem is that the unknown quantities, $\boldsymbol{\gamma}_j$, on which P and the P_i depend are random variables and not parameters. We have seen that the bootstrap method overcomes this obstacle by adopting a parametric approach. For example, with the semiparametric version of the bootstrap, in which the bootstrap sample $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n$ is obtained nonparametrically, the corresponding bootstrap labels $\tilde{\boldsymbol{\gamma}}_1, \ldots, \tilde{\boldsymbol{\gamma}}_n$ are subsequently generated by using the parametric forms for the posterior probabilities of the $\tilde{\mathbf{x}}_j$. In any event, there appears little to be gained here by seeking a jackknife-type approximation. There would be no saving in computation, as we have seen that the parametric bootstrap expectations can be fairly well estimated by the Monte Carlo method, using, say, 50 bootstrap replications. Moreover, Efron (1982) demonstrated for various other problems, including estimation of the excess error rate in discriminant analysis, that the jackknife gives a more variable estimate than the Monte Carlo approximation.

Overall, it can be seen that the estimation problem considered here is far from straightforward. Nevertheless, it can be handled directly by either a parametric or semiparametric application of the bootstrap method.

[Received June 1983. Revised November 1984.]

REFERENCES

Aitkin, M. (1980), "Mixture Applications of the EM Algorithm in GLIM," *Proceedings in Computational Statistics*, Compstat 1980, Vienna, Austria: Physica Verlag, pp. 537–541.

Aitkin, M., Anderson, D., and Hinde, J. (1981), "Statistical Modelling of Data on Teaching Styles," *Journal of the Royal Statistical Society*, Ser. A, 144, 419–461.

Anderson, T. W. (1958), *An Introduction to Multivariate Statistical Analysis*, New York: John Wiley.

Basford, K. E., and McLachlan, G. J. (1984), "Efficient Likelihood Estimation With Mixture Models," unpublished manuscript, University of Queensland, Australia.

Day, N. E. (1969), "Estimating the Components of a Mixture of Two Normal Distributions," *Biometrika*, 56, 463–474.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, 7, 1–26.

——— (1981a), "Nonparametric Estimates of Standard Error: The Jackknife, the Bootstrap and Other Methods," *Biometrika*, 68, 589–599.

——— (1981b), "Nonparametric Standard Errors and Confidence Intervals," *Canadian Journal of Statistics*, 9, 139–172.

——— (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, Philadelphia: Society for Industrial and Applied Methematics.

——— (1983), "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation," *Journal of the American Statistical Association*, 78, 316–331.

Efron, B., and Gong, G. (1983), "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation," *The American Statistician*, 37, 36–48.

Everitt, B. S., and Hand, D. J. (1981), *Finite Mixture Distributions*, London: Chapman and Hall.

Fatti, L. P., Hawkins, D. M., and Raath, E. L. (1982), "Discriminant Analysis," in *Topics in Applied Multivariate Analysis*. ed. D. M. Hawkins, Cambridge, UK: Cambridge University Press, pp. 1–71.

Fisher, R. A. (1936), "Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, 7, 179–188.

Fowlkes, E. B. (1979), "Some Methods for Studying the Mixtures of Two Normal (Lognormal) Distributions," *Journal of the American Statistical Association*, 74, 561–575.

Friedman, H. P., and Rubin, J. (1967), "On Some Invariant Criterion for Grouping Data," *Journal of the American Statistical Association*, 62, 1159–1178.

Fukunaga, K., and Kessel, D. L. (1972), "Application of Optimum Error-Reject Functions," *IEEE Transactions on Information Theory*, 18, 814–817.

——— (1973), "Non-parametric Bayes Error Estimation Using Unclassified Samples," *IEEE Transactions on Information Theory*, 19, 434–440.

Ganesalingam, S., and McLachlan, G. J. (1978), "The Efficiency of a Linear Discriminant Function Based on Unclassified Initial Samples," *Biometrika*, 65, 658–662.

——— (1979), "Small Sample Results for a Linear Discriminant Function Estimated From a Mixture of Normal Populations," *Journal of Statistical Computation and Simulation*, 9, 151–158.

——— (1980a), "A Comparison of the Mixture and Classification Approaches to Cluster Analysis," *Communications in Statistics—Theory & Methods*, Ser. A, 9, 923–933.

——— (1980b), "Error Rate Estimation on the Bias of Posterior Probabilities," *Pattern Recognition*, 12, 405–413.

——— (1981), "Some Efficiency Results for the Estimation of the Mixing Proportion in a Mixture of Two Normal Distributions," *Biometrics*, 37, 23–34.

Glick, N. (1978), "Additive Estimators for Probabilities of Correct Classification," *Pattern Recognition*, 10, 211–222.

Habbema, J. D. F., Hermans, J., and Van den Broek, K. (1974). "A Stepwise Discriminant Analysis Program Using Density Estimation," *Proceedings in Computational Statistics*, Compstat 1974, Vienna, Austria: Physica Verlag, pp. 101–110.

Hawkins, D. M. (1981), "A New Test for Multivariate Normality and Homoscedasticity," *Technometrics*, 23, 105–110.

Hawkins, D. M., Muller, M. W., and ten Krooden, J. A. (1982), "Cluster Analysis," in *Topics in Applied Multivariate Analysis*, ed. D. M. Hawkins, Cambridge, UK: Cambridge University Press, pp. 303–356.

Hernandez-Avila, A. (1979), *Problems in Cluster Analysis*, unpublished Ph.D. dissertation, University of Oxford.

Hosmer, D. W. (1973), "On MLE of the Parameters of a Mixture of Two Normal Distributions When the Sample Size Is Small," *Communications in Statistics*, 1, 217–227.

Kendall, M. G. (1966), "Discrimination and Classification," in *Multivariate Analysis*, ed. P. R. Krishnaiah, New York: Academic Press, pp. 165–185.

Kiefer, N. M. (1978), "Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model," *Econometrika*, 46, 427–434.

Lehmann, E. L. (1980), "Efficient Likelihood Estimators," *The American Statistician*, 34, 233–235.

Lissack, T., and Fu, K. S. (1976), "Error Estimation in Pattern Recognition via L^α-Distance Between Posterior Density Functions," *IEEE Transactions on Information Theory*, 22, 34–45.

Louis, T. A. (1982), "Finding the Observed Information Matrix When Using the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 44, 226–233.

McLachlan, G. J. (1982). "The Classification and Mixture Maximum Likelihood Approaches to Cluster Analysis," in *Handbook of Statistics* (Vol. 2), eds. P. R. Krishnaiah and L. N. Kanal, Amsterdam: North-Holland, pp. 199–208.

Mezzich, J. E., and Solomon, H. (1980), *Taxonomy and Behavioral Science—Comparative Performance of Grouping Methods*, New York: Academic Press.

Moore, D. S., Whitsitt, S. J., and Landgrebe, D. S. (1976), "Variance Comparisons for Unbiased Estimators of Probability of Correct Classification," *IEEE Transactions on Information Theory*, 22, 102–105.

O'Neill, T. J. (1978), "Normal Discriminant With Unclassified Observations," *Journal of the American Statistical Association*, 73, 821–826.

Orchard, T., and Woodbury, M. A. (1972), "A Missing Information Principle: Theory and Applications," *Proceedings of the 6th Berkeley Symposium* (Vol. 1), Berkeley: University of California Press, pp. 697–715.

Quandt, R. E., and Ramsey, J. B. (1978), "Estimating Mixtures of Normal Distributions and Switching Regressions," *Journal of the American Statistical Association*, 73, 730–744.

Royston, J. P. (1983), "Some Techniques for Assessing Multivariate Normality Based on the Shapiro-Wilk W," *Applied Statistics*, 32, 121–133.

Schwemer, G. T., and Dunn, O. J. (1980), "Posterior Probability Estimators in Classification Simulations," *Communications in Statistics—Simulation and Computation*, Ser. B, 9, 133–140.

Scott, A. J., and Symons, M. J. (1971), "Clustering Methods Based on Likelihood Ratio Criteria," *Biometrics*, 27, 387–397.

Small, N. J. H. (1980), "Marginal Skewness and Kurtosis in Testing Multivariate Normality," *Applied Statistics*, 29, 85–87.

Symons, M. J. (1981), "Clustering Criteria and Multivariate Normal Mixtures," *Biometrics*, 37, 35–43.

Wolfe, J. H. (1970), "Comparative Cluster Analysis of Patterns of Vocational Interest," *Multivariate Behavioral Research*, 13, 33–44.

Wu, C. F. J. (1983), "On the Convergence Properties of the EM Algorithm," *Annals of Statistics*, 11, 95–103,