ASSESSMENT OF COMPOSITION OF STOCK MIXTURES

JEROME J. PELLA[1] AND TIMOTHY L. ROBERTSON[2]

ABSTRACT

Stocks of fish can occur in mixtures, and knowledge of the composition of such a mixture may be needed. An estimate of the proportion of the mixture arising from each stock potentially present as well as a measure of the precision of this estimate may suffice. To develop these estimators, we posit that the distributions of characters of individuals differ among the stocks and that rules have been developed by others with which some success in stock identification of individuals can be had. We require test samples of individuals from each stock included in the mixture with which to evaluate the rules; these samples must be other than the learning samples used to develop the rules. The rules are also applied to a sample from the mixture. Using the numbers of individuals in each test sample and sample of the mixture which are assigned to each stock, we can estimate the composition of the mixture and the precision of this estimation.

Approximations based on large samples underlie the estimation. Numerical studies provide some idea of the sample sizes required for the approximations to be satisfactory as well as of the behavior of the estimators as related to performance of rules and sample sizes.

We note that the roles of the learning and test samples from the segregated stocks may be interchanged, allowing a repetition of the procedure.

Stocks of fish frequently occur in mixtures. When these stocks are of the same species at the same life stage, the stock identity of an individual may be difficult or impossible to ascertain. Yet if the distributions of characters of individuals differ among stocks, some success may be had in identification of individuals in a mixture by use of discriminant analysis (e.g., Hill 1959; Fukuhara et al. 1962; Anas and Murai 1969; Parsons 1972; Cook and Lord 1978) or more simply by a verbal key (Konovalov 1975). In most important applications the correct identification of individuals is not of direct value. Rather the accurate determination of the proportions of the mixture belonging to each stock is desired.

Critical to accurate assessment of composition of a mixture are the rules of assignment of individuals to stocks. The rules applied to a vector of measurements on an individual assign the individual to one stock of those possible. Among rules, those with lowest error rates of assignments provide the most accurate assessments, of course. If individuals of known stocks, either those used in developing the rules or new individuals, are assigned to stocks using the rules, a measure of error rates is provided. Although some sense of the accuracy of the rules is obtained, this does not provide a satisfactory evaluation of possible errors in estimates of stock proportions from new mixtures.

Worlund and Fredin (1962) began to attack this problem. They developed an estimation procedure for stock proportions in a mixture of an arbitrary number of stocks. Further, under restrictive assumptions concerning knowledge of the accuracy of assignments, Worlund and Fredin developed an approximate variance expression for the estimates of stock proportions in the mixture when only two stocks composed the mixture. We extend their approach now, developing methodology to estimate stock proportions in mixtures of an arbitrary number of stocks as well as the variances of such estimates under less restrictive conditions.

BACKGROUND SITUATION AND SAMPLING THEORY

We assume K stocks are known to potentially occur in the mixture. Random samples of individuals are taken from each stock at a time when the stocks are completely segregated; these may be taken before or after the mixing. A random sample of individuals from the mixture is also taken.

[1]Northwest and Alaska Fisheries Center Auke Bay Laboratory, National Marine Fisheries Service, NOAA, P.O. Box 155, Auke Bay, AK 99821.

[2]Division of Fisheries, University of Alaska, Juneau, Box 1447, Juneau, AK 99802; present address: Alaska Department of Fish and Game, Division of Commercial Fisheries, 333 Raspberry Road, Anchorage, AK 99502.

Manuscript accepted November 1978.
FISHERY BULLETIN: VOL. 77, NO. 2, 1979.

The sample from a stock at time of segregation is partitioned into two subsamples, called the learning and test samples after Cook and Lord (1978). Learning samples from the K stocks are used to develop rules of assignment; the assumptions and methods used are arbitrary for our purpose. The realism of the assumptions forming the basis of the rules is not critical; performance of the rules and some knowledge of this performance is important. Performance of the rules is determined by their application to the test samples. The rules are also applied to the sample of the mixture. Using the numbers assigned to each of the stocks by application of rules to the test samples from the segregated stocks and the sample from the mixture, we can estimate the composition of the mixture and the precision of this estimation.

A caveat concerning situations in which the methodology is not appropriate is needed before we begin. What follows presumes the individuals of a stock in both the test sample and mixture sample are drawn from a common distribution of characters used in the rules. When the condition is violated, performance of the rules would differ impermissibly between test samples and that of the mixture. We must avoid characters on which a selection process occurs between the mixture and the separate stocks.

Test Sample Theory and Analysis

Once particular rules have been established from the learning samples (e.g., using discriminant analysis), individuals forming each stock in effect have been partitioned into K mutually exclusive groups corresponding to those assigned by the rules to one of each of the K stocks. We define ϕ_{kj} to be the proportion of the individuals comprising the kth stock which is assigned by the rules to the jth stock. Also we let t_{kj} be the number of individuals in the test sample from stock k assigned by the rules to stock j, and let $T'_k = (t_{k1}, t_{k2}, \ldots, t_{kK})$. Assuming the number of individuals in the test and learning samples is small as compared with the number of individuals composing the stock, the probability of the occurrence of vector T_k is, to a good approximation, given by the multinomial probability function, i.e.,[3]

$$P(T_k) = \binom{t_{k\cdot}}{t_{k1}\ t_{k2} \ldots t_{kK}} \phi_{k1}^{t_{k1}}\ \phi_{k2}^{t_{k2}} \ldots \phi_{kK}^{t_{kK}}. \qquad (1)$$

Because the probabilities ϕ_{kj} are usually unknown, we estimate them from T_k by the well-known maximum likelihood estimator

$$\hat{\Phi}'_k = (t_{k1}/t_{k\cdot}, t_{k2}/t_{k\cdot}, \ldots, t_{kK}/t_{k\cdot}) \qquad (2)$$

corresponding to the parameter vector $\Phi'_k = (\phi_{k1}, \phi_{k2}, \ldots, \phi_{kK})$. $\hat{\Phi}_k$ is unbiased and has the variance-covariance matrix

$$\Sigma\hat{\Phi}_k = \begin{bmatrix} \frac{\phi_{k1}(1-\phi_{k1})}{t_{k\cdot}} & -\frac{\phi_{k1}\phi_{k2}}{t_{k\cdot}} & \cdots & -\frac{\phi_{k1}\phi_{kK}}{t_{k\cdot}} \\ -\frac{\phi_{k2}\phi_{k1}}{t_{k\cdot}} & \frac{\phi_{k2}(1-\phi_{k2})}{t_{k\cdot}} & \cdots & -\frac{\phi_{k2}\phi_{kK}}{t_{k\cdot}} \\ & & \ddots & \\ & & & \frac{\phi_{kK}(1-\phi_{kK})}{t_{k\cdot}} \end{bmatrix}. \qquad (3)$$

Test samples from different stocks are statistically independent and covariance between elements of $\hat{\Phi}_k$ and $\hat{\Phi}_{k'}$, are zero for $k \neq k'$.

Mixed Sample Theory and Analysis

The mixture of stocks at the time of sampling is comprised of possibly as many as K stocks. Ignoring for the moment the actual stock composition of the mixture, our rules established from the learning samples partition the mixture into K mutually exclusive groups again corresponding to the K stocks to which individuals are assigned. We define λ_j to be the proportion of the individuals composing the entire mixture which would be assigned to the jth stock by the rules. Also we let m_j be the actual number of fish in the sample from the mixture which are assigned to stock j. If the size of the sample from the mixture is small compared with the number of fish composing the mixture, the probability of observing the vector $M' = (m_1, m_2, \ldots, m_K)$ is given by the multinomial probability function, i.e.,

[3]The dot notation implies summation over the subscript. Thus, $t_{k.} = \Sigma_{j=1}^{K} t_{kj}$ is the size of the test sample from the kth stock.

$$P(M) = \binom{m_.}{m_1\ m_2 \ldots m_K} \lambda_1^{m_1} \lambda_2^{m_2} \ldots \lambda_K^{m_K} . \quad (4)$$

We can estimate the probabilities λ_j from M by the maximum likelihood estimator

$$\hat{\Lambda}' = (m_1/m_., m_2/m_., \ldots, m_K/m_.) \quad (5)$$

corresponding to the parameter vector $\Lambda' = (\lambda_1, \lambda_2, \ldots, \lambda_K)$. $\hat{\Lambda}$ is unbiased and has the variance-covariance matrix

$$\Sigma_{\hat{\Lambda}} = \begin{bmatrix} \frac{\lambda_1(1-\lambda_1)}{m_.} & -\frac{\lambda_1\lambda_2}{m_.} & \cdots & -\frac{\lambda_1\lambda_K}{m_.} \\ -\frac{\lambda_2\lambda_1}{m_.} & \frac{\lambda_2(1-\lambda_2)}{m_.} & \cdots & -\frac{\lambda_2\lambda_K}{m_.} \\ & & \ddots & \\ & & & \frac{\lambda_K(1-\lambda_K)}{m_.} \end{bmatrix} . \quad (6)$$

$\hat{\Lambda}$ is a natural estimator of stock composition of the mixture. Unfortunately as we see next, its expected value, Λ, depends not only on stock composition, but also on the behavior of the rules.

Basic Relation Between Parameters of Test Samples and Those of the Sample from the Mixture

We know the mixture consists of individuals from at most K stocks. Let θ_k be the proportion of the individuals composing the entire mixture which are of the kth stock, where $0 \leq \theta_k \leq 1$ for all k and

$$\sum_{k=1}^{K} \theta_k = 1 .$$

The parameter vector $\Theta' = (\theta_1, \theta_2, \ldots, \theta_K)$ is unknown; its estimation is our objective. If the individuals of each of the stocks occurring in the mixture are a random sample from the character distribution of that stock, then the probability that a randomly sampled individual from the mixture is assigned to the jth stock, λ_j, is related to previously defined probabilities by the equation system

$$\lambda_j = \sum_{k=1}^{K} \theta_k \phi_{kj}, \quad j = 1, 2, \ldots, K. \quad (7)$$

The term, $\theta_k \phi_{kj}$, represents the probability a randomly sampled individual from the mixture is of stock k and assigned to stock j; summing over the K stocks gives the total probability the individual is assigned to stock j. This basic set of relationships can be expressed in matrix notation,

$$\Lambda = \Phi'\Theta \quad (8)$$

$$\text{where } \Phi = \begin{bmatrix} \Phi_1' \\ \Phi_2' \\ \vdots \\ \Phi_K' \end{bmatrix} = \begin{bmatrix} \phi_{11}\ \phi_{12} \ldots \phi_{1K} \\ \vdots \\ \phi_{K1}\ \phi_{K2} \ldots \phi_{KK} \end{bmatrix} .$$

ESTIMATION OF STOCK COMPOSITION OF MIXTURE

If $|\Phi| \neq 0$, we can solve Equation (8) for Θ,

$$\Theta = (\Phi')^{-1} \Lambda. \quad (9)$$

When the rules assign individuals from the stocks without error, $\Phi = I$, and $\Theta = \Lambda$. Then the natural estimator $\hat{\Lambda}$ is appropriate. But the rules will usually be imperfect, yet Equation (9) shows we can still solve for Θ without error provided Φ and Λ are known. Unfortunately neither Φ nor Λ is known in usual circumstances; however, we saw how to estimate them from the test and mixed samples using Equations (2) and (5). When Λ and Φ in Equation (8) are replaced by estimates from Equations (2) and (5), the problem of estimating Θ is a special case of estimation of the solution of a system of linear equations with random coefficients. Fuller[4] has provided several solutions for the general problem; these are applicable in the present case for large test and mixed samples. Later we indicate how large these samples must be.

[4]Fuller, W. A. 1970. Mimeographed class notes, Statistics 638, winter 1969-70. Iowa State Univ. Stat. Lab., 56 p., on file at the library of Northwest and Alaska Fisheries Center Auke Bay Laboratory, National Marine Fisheries Service, NOAA, P.O. Box 155, Auke Bay, AK 99821.

Fuller (see footnote 4) begins with the simple estimator

$$\hat{\Theta} = (\hat{\Phi}')^{-1}\ \hat{\Lambda} \qquad (10)$$

with the restriction that the event $|\hat{\Phi}| = 0$ must be impossible. The asymptotic variance-covariance matrix of $\hat{\Theta}$, $\Sigma_{\hat{\Theta}}$, is given by

$$\Sigma_{\hat{\Theta}} = (\Phi')^{-1}\ (\Sigma_{\hat{\Lambda}} + \Sigma_{\Phi'\Theta})\ \Phi^{-1} \qquad (11)$$

and $\Sigma_{\hat{\Lambda}}$ is defined by Equation (6).

We remark that variation in estimates of Θ arises additively from two sources: 1) sampling variation in estimation of the assignment composition of the mixture which is represented by $\Sigma_{\hat{\Lambda}}$; and 2) sampling variation in estimation of the probability of assignment matrix Φ which is represented by $\Sigma_{\Phi'\Theta}$. The diagonal elements of the $\Sigma_{\hat{\Theta}}$ are the variances of the elements of $\hat{\Theta}$; the square roots of these are the standard errors.

where
$$\Sigma_{\Phi'\Theta} = \begin{bmatrix} \sum_i \frac{\theta_i^2\phi_{i1}(1-\phi_{i1})}{l_{i.}} & -\sum_i \frac{\theta_i^2\phi_{i1}\phi_{i2}}{l_{i.}} & \cdots & -\sum_i \frac{\theta_i^2\phi_{i1}\phi_{iK}}{l_{i.}} \\ -\sum_i \frac{\theta_i^2\phi_{i2}\phi_{i1}}{l_{i.}} & \sum_i \frac{\theta_i^2\phi_{i2}(1-\phi_{i2})}{l_{i.}} & \cdots & -\sum_i \frac{\theta_i^2\phi_{i2}\phi_{iK}}{l_{i.}} \\ \vdots & \vdots & & \vdots \\ -\sum_i \frac{\theta_i^2\phi_{iK}\phi_{i1}}{l_{i.}} & -\sum_i \frac{\theta_i^2\phi_{iK}\phi_{i2}}{l_{i.}} & \cdots & \sum_i \frac{\theta_i^2\phi_{iK}(1-\phi_{iK})}{l_{i.}} \end{bmatrix}$$

Bias in estimation of Θ is approximately given by

$$B \cong (\Phi')^{-1}\ G\Theta, \qquad (12)$$

where

$$G = \begin{bmatrix} \frac{\phi^{11}\phi_{11}(1-\phi_{11})}{l_{1.}} - \sum_{k\neq 1}\frac{\phi^{k1}\phi_{11}\phi_{1k}}{l_{1.}} & \frac{\phi^{12}\phi_{21}(1-\phi_{21})}{l_{2.}} - \sum_{k\neq 1}\frac{\phi^{k2}\phi_{21}\phi_{2k}}{l_{2.}} & \cdots & \frac{\phi^{1K}\phi_{K1}(1-\phi_{K1})}{l_{K.}} - \sum_{k\neq 1}\frac{\phi^{kK}\phi_{K1}\phi_{Kk}}{l_{K.}} \\ \frac{\phi^{21}\phi_{12}(1-\phi_{12})}{l_{1.}} - \sum_{k\neq 2}\frac{\phi^{k1}\phi_{12}\phi_{1k}}{l_{1.}} & \frac{\phi^{22}\phi_{22}(1-\phi_{22})}{l_{2.}} - \sum_{k\neq 2}\frac{\phi^{k2}\phi_{22}\phi_{2k}}{l_{2.}} & \cdots & \frac{\phi^{2K}\phi_{K2}(1-\phi_{K2})}{l_{K.}} - \sum_{k\neq 2}\frac{\phi^{kK}\phi_{K2}\phi_{Kk}}{l_{K.}} \\ \vdots & \vdots & & \vdots \\ \frac{\phi^{K1}\phi_{1K}(1-\phi_{1K})}{l_{1.}} - \sum_{k\neq K}\frac{\phi^{k1}\phi_{1K}\phi_{1k}}{l_{1.}} & \frac{\phi^{K2}\phi_{2K}(1-\phi_{2K})}{l_{2.}} - \sum_{k\neq K}\frac{\phi^{k2}\phi_{2K}\phi_{2k}}{l_{2.}} & \cdots & \frac{\phi^{KK}\phi_{KK}(1-\phi_{KK})}{l_{K.}} - \sum_{k\neq K}\frac{\phi^{kK}\phi_{KK}\phi_{Kk}}{l_{K.}} \end{bmatrix}$$

and ϕ^{ij} is the element in the ith row and jth column of Φ^{-1}.

Improved estimators with smaller bias than $\hat{\Theta}$ can also be developed from Fuller's (see footnote 4) general results. These are the new estimators:

$$\tilde{\Theta} = (\hat{\Phi}' + \hat{G})^{-1}\,\hat{\Lambda} \text{ and} \tag{13}$$

$$\ddot{\Theta} = [I - (\hat{\Phi}')^{-1}\,\hat{G}\,]\,(\hat{\Phi}')^{-1}\,\hat{\Lambda}. \tag{14}$$

Here $\hat{G}$ is obtained by substituting the estimates for the unknown parameters of G. To the order of approximation provided by Fuller, these estimators have the same variance-covariance matrix as $\hat{\Theta}$. An internal estimate of this variance-covariance matrix $\Sigma_{\hat{\Theta}}$ can be obtained by substitution of observed values for parameters in Equation (11). We can substitute in Equation (11) for elements of Θ the corresponding elements of either $\hat{\Theta}$, $\tilde{\Theta}$, or $\ddot{\Theta}$. To distinguish between these possibilities, we label the internal estimators of $\Sigma_{\hat{\Theta}}$ as $\hat{\Sigma}_{\hat{\Theta}}$, $\hat{\Sigma}_{\tilde{\Theta}}$, or $\hat{\Sigma}_{\ddot{\Theta}}$, respectively. With the internal estimate of $\Sigma_{\hat{\Theta}}$, we can estimate not only Θ but also how precisely the estimation is accomplished.

To establish confidence intervals on the elements of Θ, we assume test and mixed samples are sufficiently large so that the estimators $\hat{\Theta}$, $\tilde{\Theta}$, or $\ddot{\Theta}$ are each approximately distributed as the multivariate normal with mean Θ and known variance-covariance matrix $\hat{\Sigma}_{\hat{\Theta}}$, $\hat{\Sigma}_{\tilde{\Theta}}$, or $\hat{\Sigma}_{\ddot{\Theta}}$, respectively. Then a 100(1 - α)% set of confidence intervals such that all the unknown elements of Θ are simultaneously covered by their respective intervals with a probability 1 - α is for the estimator $\hat{\Theta}$ (say) as follows (see Morrison 1967, section 4.4):

$$\begin{aligned}
\hat{\theta}_1 - (\hat{\sigma}_{11}{}^2\chi_{\alpha;K-1}{}^2)^{1/2} &\leqslant \theta_1 \leqslant \hat{\theta}_1 + (\hat{\sigma}_{11}{}^2\chi_{\alpha;K-1}{}^2)^{1/2}\\
\hat{\theta}_2 - (\hat{\sigma}_{22}{}^2\chi_{\alpha;K-1}{}^2)^{1/2} &\leqslant \theta_2 \leqslant \hat{\theta}_2 + (\hat{\sigma}_{22}{}^2\chi_{\alpha;K-1}{}^2)^{1/2}\\
&\;\vdots\\
\hat{\theta}_k - (\hat{\sigma}_{kk}{}^2\chi_{\alpha;K-1}{}^2)^{1/2} &\leqslant \theta_k \leqslant \hat{\theta}_k + (\hat{\sigma}_{kk}{}^2\chi_{\alpha;K-1}{}^2)^{1/2}\\
&\;\vdots\\
\hat{\theta}_K - (\hat{\sigma}_{KK}{}^2\chi_{\alpha;K-1}{}^2)^{1/2} &\leqslant \theta_K \leqslant \hat{\theta}_K + (\hat{\sigma}_{KK}{}^2\chi_{\alpha;K-1}{}^2)^{1/2}
\end{aligned} \tag{15}$$

where $\hat{\sigma}_{kk}{}^2$ is the element in the kth row and column of $\hat{\Sigma}_{\hat{\Theta}}$, and $\chi_{\alpha;K-1}{}^2$ is the value associated with a chi-square distribution with K-1 degrees of freedom such that 100α% of the distribution lies above it. When only two stocks occur in the mixture, this set of simultaneous intervals reduces to the familiar univariate normal approximation for setting confidence intervals:

$$\begin{aligned}
\hat{\theta}_1 - z_{\alpha/2}(\hat{\sigma}_{11}{}^2)^{1/2} < \theta_1 < \hat{\theta}_1 + z_{\alpha/2}(\hat{\sigma}_{11}{}^2)^{1/2}\\
\hat{\theta}_2 - z_{\alpha/2}(\hat{\sigma}_{22}{}^2)^{1/2} < \theta_2 < \hat{\theta}_2 + z_{\alpha/2}(\hat{\sigma}_{22}{}^2)^{1/2}
\end{aligned} \tag{16}$$

where $z_{\alpha/2}$ is the standardized normal deviate such that 100(α/2)% of the distribution lies below $-z_{\alpha/2}$ and 100(α/2)% lies above $z_{\alpha/2}$. These expressions are in terms of the estimator $\hat{\Theta}$; they apply as well to the other estimators when elements of $\tilde{\Theta}$ or $\ddot{\Theta}$ replace those of $\hat{\Theta}$ within them.

Worlund and Fredin (1962) developed the estimator $\hat{\Theta}$ in Equation (10). To translate their notation to ours, let

$$\begin{aligned}
\phi_{ij} &= P_{ij}\\
\theta_i &= F_i\\
\lambda_j &= \hat{R}_j
\end{aligned} \tag{17}$$

and permit the subscripts to take on letter values $a, b, c, \ldots$. In the special case when the mixture is comprised of only two stocks, they developed an asymptotic expression for the variance of $\hat{\theta}_1$ (the variance of $\hat{\theta}_2$ necessarily equals that of $\hat{\theta}_1$ since $\hat{\theta}_2 = 1 - \hat{\theta}_1$). In deriving the variance expression, they assumed Φ is known without error so that $\Sigma_{\hat{\Phi}'\Theta}$ is a null matrix; such is approximately true as $t_{1.}$ and $t_{2.}$ become large.

We consider two examples now to illustrate our notation and method in concrete terms. The first case restricts our general approach to the simplest situation of two stocks in the mixture; the second provides numerical computations for three stocks so that users may verify their understanding of the formulas.

Special Case of Two Stocks

We assume a set of rules based on learning samples from each stock has been developed which assigns each individual to one of the two stocks. Individuals in two test samples, size $t_{1.}$ from stock 1 and size $t_{2.}$ from stock 2, are assigned by the rules to either of the stocks. Of the $t_{1.}$ individuals, t_{11} are assigned to stock 1 and t_{12}, to stock 2. Of the $t_{2.}$ individuals, t_{21} are assigned to stock 1 and t_{22}, to stock 2. A sample from the mixture of size $m_.$ is assigned by the rules to the stocks—m_1 to stock 1 and m_2 to stock 2. Then

$$\hat{\Phi} = \begin{bmatrix} \hat{\phi}_{11} & \hat{\phi}_{12} \\ \hat{\phi}_{21} & \hat{\phi}_{22} \end{bmatrix} = \begin{bmatrix} t_{11}/t_{1\cdot} & t_{12}/t_{1\cdot} \\ t_{21}/t_{2\cdot} & t_{22}/t_{2\cdot} \end{bmatrix} \quad (18)$$

$$\hat{\Lambda} = \begin{bmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{bmatrix} = \begin{bmatrix} m_1/m_. \\ m_2/m_. \end{bmatrix} \quad (19)$$

$$(\hat{\Phi}')^{-1} = \begin{bmatrix} \hat{\phi}^{11} & \hat{\phi}^{21} \\ \hat{\phi}^{12} & \hat{\phi}^{22} \end{bmatrix} = \frac{1}{\hat{\phi}_{11}\hat{\phi}_{22} - \hat{\phi}_{12}\hat{\phi}_{21}} \begin{bmatrix} \hat{\phi}_{22} & -\hat{\phi}_{21} \\ -\hat{\phi}_{12} & \hat{\phi}_{11} \end{bmatrix} \quad (20)$$

$$\hat{\Theta} = (\hat{\Phi}')^{-1}\hat{\Lambda} = \begin{bmatrix} \dfrac{\hat{\lambda}_1 - \hat{\phi}_{21}}{\hat{\phi}_{11} - \hat{\phi}_{21}} \\ \dfrac{\hat{\phi}_{11} - \hat{\lambda}_1}{\hat{\phi}_{11} - \hat{\phi}_{21}} \end{bmatrix}. \quad (21)$$

The form of $\hat{\Theta}$ has been chosen to agree with the solution of Worlund and Fredin (1962); in developing this form, we used the facts for this special case that

$$\begin{aligned} \hat{\phi}_{22} &= 1 - \hat{\phi}_{21} \\ \hat{\phi}_{12} &= 1 - \hat{\phi}_{11} \\ \text{and } \hat{\lambda}_2 &= 1 - \hat{\lambda}_1 \end{aligned} \quad (22)$$

$$G = \begin{bmatrix} \dfrac{\phi^{11}\phi_{11}(1-\phi_{11}) - \phi^{21}\phi_{11}\phi_{12}}{t_{1\cdot}} & \dfrac{\phi^{12}\phi_{21}(1-\phi_{21}) - \phi^{22}\phi_{21}\phi_{22}}{t_{2\cdot}} \\ \dfrac{\phi^{21}\phi_{12}(1-\phi_{12}) - \phi^{11}\phi_{12}\phi_{11}}{t_{1\cdot}} & \dfrac{\phi^{22}\phi_{22}(1-\phi_{22}) - \phi^{12}\phi_{22}\phi_{21}}{t_{2\cdot}} \end{bmatrix} \quad (23)$$

$$\Sigma_{\hat{\Lambda}} = \frac{\lambda_1\lambda_2}{m_.} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad (24)$$

$$\Sigma_{\hat{\Phi}'\Theta} = \left(\frac{\theta_1^{\,2}\phi_{11}\phi_{12}}{t_{1\cdot}} + \frac{\theta_2^{\,2}\phi_{21}\phi_{22}}{t_{2\cdot}} \right) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \quad (25)$$

$\hat{G}$, $\hat{\Sigma}_{\hat{\Lambda}}$ and $\hat{\Sigma}_{\hat{\Phi}'\Theta}$ are obtained by substituting estimates for the corresponding parameters. Then

$$\hat{\Sigma}_{\hat{\Theta}} = (\hat{\Phi}')^{-1} (\hat{\Sigma}_{\hat{\Lambda}} + \hat{\Sigma}_{\hat{\Phi}'\Theta}) \hat{\Phi}^{-1} \quad (26)$$

provides internal estimates of the variance of $\hat{\theta}_1$ or $\hat{\theta}_2$ as well as their covariance.

Expressions for $\tilde{\Theta}$ and $\ddot{\Theta}$ follow directly from specialization of Equations (13) and (14) to two stocks; substitution of their elements into Equation (26) in place of those of $\hat{\Theta}$ provides $\hat{\Sigma}_{\tilde{\Theta}}$ and $\hat{\Sigma}_{\ddot{\Theta}}$, respectively.

Numerical Computations for Three Stocks

To illustrate the computations for a three-stock situation, we use the information reported by Cook and Lord (1978) regarding stock composition of high-seas mixtures of sockeye salmon, *Oncorhynchus nerka*. Their purpose was to estimate proportions of the mixture arising from each of three river systems — Egegik, Kvichak, and Naknek — of the Bristol Bay region of Alaska. Actually the application of our methods is inappropriate because Cook and Lord used individuals of test samples from the segregated stocks both to modify an original set of rules from the learning samples as well as to estimate Φ. Because our purpose is only to illustrate the computations, we will treat their observations as though the test samples had been used exclusively to estimate Φ. Using the test samples in developing the rules, as Cook and Lord did, should produce greater precision in estimation of composition of a mixture; the disadvantage at present is the inability to assess the precision of these enhanced estimates. In developing the variance-covariance matrix $\Sigma_{\hat{\Theta}}$, we assumed $\hat{\Phi}$ and $\hat{\Lambda}$ are statistically independent. Such is untrue if the test samples are used as by Cook and Lord both to develop the rules used to estimate Λ as well as to estimate Φ.

Test samples from the segregated stocks of the three rivers were assigned by the rules to these stocks (Table 1). Then the rules were applied to 101 fish caught on the high seas. Of these, 25 were assigned to Egegik, 22 to Kvichak, and 54 to Naknek. We identify Egegik, Kvichak, and Naknek as the first, second, and third streams in our subscript use. Computations using these data produce the following results:

$$\hat{\Phi} = \begin{bmatrix} 0.80000 & 0.08000 & 0.12000 \\ 0.04000 & 0.74000 & 0.22000 \\ 0.16667 & 0.20833 & 0.62500 \end{bmatrix}.$$

(Our $\hat{\Phi}$ is the transpose of $\hat{C}$ of Cook and Lord (1978).)

$$\hat{\Lambda} = \begin{bmatrix} 0.24752 \\ 0.21782 \\ 0.53465 \end{bmatrix}.$$

(Our $\hat{\Lambda}$ is the same statistic as R_u of Cook and Lord (1978); apparently they have numerical errors in their evaluation of R_u.)

TABLE 1.—Numbers of sockeye salmon in test samples from three Bristol Bay (Alaska) rivers—Egegik, Kvichak, and Naknek—assigned by rules to these rivers (source: Cook and Lord 1978).

Actual river	Assigned river		
	Egegik	Kvichak	Naknek
Egegik	40	4	6
Kvichak	2	37	11
Naknek	8	10	30

$$\hat{\Phi}^{-1} = \begin{bmatrix} 1.30019 & -0.07801 & -0.22218 \\ 0.03641 & 1.49782 & -0.53422 \\ -0.35885 & -0.47847 & 1.83732 \end{bmatrix}$$

$$\hat{G} = \begin{bmatrix} 0.00480 & -0.00086 & -0.00424 \\ -0.00154 & 0.00737 & -0.00666 \\ -0.00326 & -0.00651 & 0.01090 \end{bmatrix}$$

$$\hat{\Theta} = \begin{bmatrix} 0.138 \\ 0.051 \\ 0.811 \end{bmatrix} \quad \tilde{\Theta} = \begin{bmatrix} 0.145 \\ 0.062 \\ 0.793 \end{bmatrix} \quad \ddot{\Theta} = \begin{bmatrix} 0.145 \\ 0.062 \\ 0.793 \end{bmatrix}.$$

(Our $\hat{\Theta}$ is $\hat{U}$ of Cook and Lord (1978); their errors in evaluating R_u are responsible for the discrepancy with our estimate $\hat{\Theta}$.)

$$\hat{\Sigma}_{\hat{\Lambda}} = \begin{bmatrix} 0.00184 & -0.00053 & -0.00131 \\ & 0.00169 & -0.00115 \\ & & 0.00246 \end{bmatrix}$$

$$\hat{\Sigma}_{\hat{\Phi}'\Theta} = \begin{bmatrix} 0.00197 & -0.00050 & -0.00146 \\ & 0.00230 & -0.00180 \\ & & 0.00326 \end{bmatrix}$$

$$\hat{\Sigma}_{\hat{\Theta}} = \begin{bmatrix} 0.00975 & 0.00208 & -0.01184 \\ & 0.01454 & -0.01662 \\ & & 0.02846 \end{bmatrix}.$$

In computing $\hat{\Sigma}_{\hat{\Phi}'\Theta}$ and $\hat{\Sigma}_{\hat{\Theta}}$, $\hat{\Theta}$ is used as the estimate of Θ.

The 90% confidence set from Equation (15) using $\hat{\Theta}$ is as follows:

$$-0.074 \leq \theta_1 \leq 0.350$$
$$-0.208 \leq \theta_2 \leq 0.310$$
$$0.449 \leq \theta_3 \leq 1.173.$$

The elements of Θ must lie between 0 and 1; therefore, we can set the lower limits of the first two intervals to 0, and the upper limit of the third interval to 1. The actual composition was estimated by Cook and Lord (1978) from returning adults to Bristol Bay as

$$\Theta = \begin{bmatrix} 0.325 \\ 0.061 \\ 0.614 \end{bmatrix},$$

which falls within the intervals of the confidence set as would be expected. However, recall that the condition that test samples be used exclusively to evaluate the rules was violated; therefore, the confidence set is not valid. Further, Cook and Lord

(1978) have misgivings of probable occurrence of additional unaccounted stocks in the high-seas mixture.

BEHAVIOR OF ESTIMATORS AND ASYMPTOTIC FORMULAS

Of interest to investigators beginning studies of stock composition of mixtures is the behavior of our estimators as test and mixed sample sizes vary for fixed rules and the influence of rules on the estimators. Further, we remarked that our solution of the stock mixture problem assumes large test and mixed samples. Of concern is how large specifically the samples must be for the asymptotic expressions to be reasonably accurate. This examination will be restricted to the two-stock case which is general for our purpose in that any number of stocks can be partitioned into two groups; that is, we can evaluate the estimators for a particular stock when the remaining stocks are lumped into a second group after assignment by the rules to the individual stocks. Bias and variance for the particular stock would be unchanged then even if the stocks of the second group were treated severally.

We evaluate estimation behavior and asymptotic approximation for three choices of Φ representing rules of increasing accuracy:

$$\text{Case 1. } \Phi = \begin{bmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}$$

$$\text{Case 2. } \Phi = \begin{bmatrix} 0.75 & 0.25 \\ 0.10 & 0.90 \end{bmatrix}$$

$$\text{Case 3. } \Phi = \begin{bmatrix} 0.90 & 0.10 \\ 0.10 & 0.90 \end{bmatrix}$$

We let $\Theta' = (0.6, 0.4)$ for all three cases. Based on experience in identification of sockeye salmon in Bristol Bay, Alaska, using discriminant functions on scale features, the ranges of elements of Φ are realistic. The choice of Θ is arbitrary, of course.

Given Φ, Θ, and sample sizes $t_{1.}$, $t_{2.}$, and $m_.$, we can enumerate all possible sample points — t_{11}, t_{12}, t_{21}, t_{22}, m_1, and m_2 — as well as compute their probabilities of occurrence. In our evaluations, we always used equal test sample sizes. For each sample point, we can compute $\hat{\Theta}$, $\tilde{\Theta}$, $\ddot{\Theta}$, $\hat{\Sigma}_{\hat{\Theta}}$, $\hat{\Sigma}_{\tilde{\Theta}}$, and $\hat{\Sigma}_{\ddot{\Theta}}$. With these calculations for each point we can compute the mean and variance of each estimator by weighting its value at a sample point by the probability of that point.

Estimation of Θ by $\hat{\Theta}$, $\tilde{\Theta}$, or $\ddot{\Theta}$ requires the probability that $|\hat{\Phi}| = 0$ be zero; this condition is not met. If we supplement the procedure by assigning arbitrary values to the estimators when $|\hat{\Phi}| = 0$, means and variances of such modified estimators will approach the values we obtained by omission of such sample points. The probability that $|\hat{\Phi}| = 0$ rapidly decreases with increasing test sample sizes. For case 1 with test samples of 20, it is $<5 \times 10^{-4}$, and with test samples of 40, about 5×10^{-7}. The probability also decreases with improved identification of stocks. For case 3 with test samples of 20, the probability is $<4 \times 10^{-6}$. Weighting the arbitrary values of the estimators corresponding to such points by their probabilities makes their contributions to expectation computations negligible.

We found these numerical studies to be expensive, especially with large sample sizes. Therefore, we began omitting sample points whose probability was small even if $|\hat{\Phi}| \neq 0$. Criteria for omission of points are indicated in our tables; the justification is again their negligible contributions in expectation computations. Results will be discussed in terms of the first stock only.

We consider bias first. Bias of any estimator, $\hat{\theta}_1$, $\tilde{\theta}_1$, or $\ddot{\theta}_1$, is unaffected by changes in mixed sample size; however, bias decreases with increasing test sample size. For example, we computed biases for case 1 with three mixed sample sizes—20, 30, and 40 — at each of two choices of equal sized test samples — 20 and 30 (Table 2, lines 1 to 6). The occasional change in the last digit for biases at varying mixed sample sizes within fixed test sample sizes is probably caused by omission of improbable sample points in evaluation of expectations. Bias of $\hat{\theta}_1$ also is predicted by the asymptotic formula [Equation (12)] to vary only with test sample sizes, not mixed sample size (Table 2, last column).

Bias of $\hat{\theta}_1$ is of opposite sign from that of either $\tilde{\theta}_1$ or $\ddot{\theta}_1$ (Table 2, column $b_{\hat{\theta}_1}$ as compared with columns $b_{\tilde{\theta}_1}$ and $b_{\ddot{\theta}_1}$). Absolute value of bias of $\tilde{\theta}_1$ is less than that of either $\hat{\theta}_1$ or $\ddot{\theta}_1$. Generally, absolute value of bias of $\ddot{\theta}_1$ is also less than that of $\hat{\theta}_1$; the sole exception is case 1 with test samples of only 20.

Bias of $\tilde{\theta}_1$ or $\ddot{\theta}_1$ decreases with improved rules as we go from case 1 to case 2 to case 3, holding test and mixed sample sizes fixed. Biases computed for $\hat{\theta}_1$ decreased between case 1 and case 3 for which the Φ-matrices are both symmetric; however, for

TABLE 2.—Biases, ($b_{\hat{\theta}_1}$, $b_{\tilde{\theta}_1}$, and $b_{\ddot{\theta}_1}$), of estimators and asymptotic bias, b_1, from Equation (12) for indicated Φ-matrices, indicated test and mixed sample sizes, and $\Theta' = (0.6, 0.4)$.

Φ	Test sample sizes	Mixed sample size	$b_{\hat{\theta}_1}$	$b_{\tilde{\theta}_1}$	$b_{\ddot{\theta}_1}$	b_1
Case 1	20	[1]20	+0.01077	−0.00363	−0.01873	+0.00750
		[2]30	+ .01076	− .00362	− .01873	+ .00750
$\begin{bmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}$		[2]40	+ .01076	− .00364	− .01873	+ .00750
	30	[2]20	+ .00637	− .00123	− .00381	+ .00500
		[2]30	+ .00637	− .00124	− .00381	+ .00500
		[2]40	+ .00636	− .00124	− .00382	+ .00500
	40	[2]40	+ .00438	− .00063	− .00108	+ .00375
Case 2	20	[1]20	+ .01053	− .00090	− .00186	+ .00905
$\begin{bmatrix} 0.75 & 0.25 \\ 0.10 & 0.90 \end{bmatrix}$	30	[2]30	+ .00660	− .00037	− .00060	+ .00604
	40	[2]40	+ .00481	− .00022	− .00033	+ .00453
Case 3	20	[1]20	+ .00155	− .00014	− .00018	+ .00141
$\begin{bmatrix} 0.90 & 0.10 \\ 0.10 & 0.90 \end{bmatrix}$	30	[2]30	+ .00099	− .00007	− .00008	+ .00094

[1]Evaluated at all sample points except when $|\Phi| = 0$.
[2]Evaluated only at sample points for which probability of observing the outcomes of the test samples $>10^{-6}$ and $|\Phi| \neq 0$.

two of three combinations of test and mixed sample sizes, repeated under case 1 and case 2, bias increased between case 1 and case 2, the latter not having a symmetric Φ-matrix.

The predicted bias of $\hat{\theta}_1$ from the asymptotic formula [Equation (12)] agrees with actual bias of $\hat{\theta}_1$ reasonably well. The approximation obviously becomes more accurate as size of test samples increases or as rules improve.

Biases would appear negligible in comparison with magnitude of variances of the estimators next considered. Absolute value of bias in the situations evaluated represents at most 3.1% of the parameter value, $\theta_1 = 0.6$. Random errors in estimation are the main concern.

Variances of the estimators ($\hat{\Theta}$, $\tilde{\Theta}$, and $\ddot{\Theta}$) decrease as test samples become larger, agreeing in behavior with biases; in contrast to biases, variances also decrease as size of mixed samples increases. We computed variances under case 1 for the same test and mixed sample sizes described for bias evaluation (Table 3, lines 1 to 6). Although variance of any of the estimators ($\hat{\theta}_1$, $\tilde{\theta}_1$, and $\ddot{\theta}_1$) decreases with size of test or mixed samples, the rate decreases with size of either type when that of the other is fixed. For example, at test samples of

TABLE 3.—Variances ($\sigma^2_{\hat{\theta}_1}$, $\sigma^2_{\tilde{\theta}_1}$, and $\sigma^2_{\ddot{\theta}_1}$) of estimators and asymptotic variance (σ_{11}^2) from Equation (11) for indicated Φ-matrices, indicated test and mixed sample sizes, and $\Theta' = (0.6, 0.4)$.

Φ	Test sample sizes	Mixed sample size	$\sigma^2_{\hat{\theta}_1}$	$\sigma^2_{\tilde{\theta}_1}$	$\sigma^2_{\ddot{\theta}_1}$	σ_{11}^2
Case 1	20	[1]20	0.11232	0.06741	0.86891	0.06900
		[2]30	.08685	.05120	.68980	.05250
$\begin{bmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}$		[2]40	.07417	.04311	.60199	.04425
	30	[2]20	.07957	.06370	.23131	.06250
		[2]30	.05904	.04686	.17620	.04600
		[2]40	.04877	.03844	.14864	.03775
	40	[2]40	.04051	.03538	.03518	.03450
Case 2	20	[1]20	.04640	.04010	.04448	.03927
$\begin{bmatrix} 0.75 & 0.25 \\ 0.10 & 0.90 \end{bmatrix}$	30	[2]30	.02892	.02668	.02658	.02618
	40	[2]40	.02111	.01995	.01992	.01963
Case 3	20	[1]20	.02425	.02293	.02290	.02269
$\begin{bmatrix} 0.90 & 0.10 \\ 0.10 & 0.90 \end{bmatrix}$	30	[2]30	.01579	.01525	.01525	.01513

[1]Evaluated at all sample points except when $|\Phi| = 0$.
[2]Evaluated only at sample points for which probability of observing the outcomes of the test samples $>10^{-6}$ and $|\Phi| \neq 0$.

20, variance of $\tilde{\theta}_1$ decreases 24% when mixed samples increase from 20 to 30, but by only 16% when mixed samples increase further to 40. Similarly at a mixed sample of 40, variance of $\tilde{\theta}_1$ decreases by 11% and 8% as test samples increase from 20 to 30 and 30 to 40, respectively. The return to sampling effort of precision of estimation by increase in mixed sample size with test sample sizes fixed diminishes and is limited by test sample sizes. Return of precision to increase in test sample sizes is similarly related to and limited by mixed sample size.

Overriding both test and mixed samples in determining ultimate precision of estimation are the rules characterized by the Φ-matrix. As rules of assignment improve and the Φ-matrix approaches the identity matrix, precision of estimation at fixed test and mixed sample sizes increases.

In our evaluations, variance of $\tilde{\theta}_1$ is always less than that of $\hat{\theta}_1$. In this respect $\tilde{\theta}_1$ also enjoys considerable advantage over $\ddot{\theta}_1$ when rules are poor, case 1, and sample sizes are small. As test or mixed samples increase, the advantage diminishes until $\ddot{\theta}_1$ has the smaller variance. However, differences among variances of the three estimators ($\hat{\theta}_1$, $\tilde{\theta}_1$, and $\ddot{\theta}_1$) become negligible either as rules improve or samples sizes become large.

Predicted variance of the estimators of θ_1 from the asymptotic formula [Equation (11)] describes variance of $\tilde{\theta}_1$ remarkably well, even when rules are poor and sample sizes are small (Table 3, compare lines 1 to 7 of column $\sigma_{11}{}^2$ with column $\sigma_{\tilde{\theta}_1}{}^2$). With improved rules, variances of each of $\hat{\theta}_1$, $\tilde{\theta}_1$, and $\ddot{\theta}_1$ are well described by the asymptotic variance (Table 3, compare lines 8 to 12 of columns $\sigma_{\hat{\theta}_1}{}^2$, $\sigma_{\tilde{\theta}_1}{}^2$, and $\sigma_{\ddot{\theta}_1}{}^2$ with column $\sigma_{11}{}^2$).

Two evaluations concerning adequacy of internal variance estimation by $\hat{\Sigma}_{\hat{\Theta}}$, $\hat{\Sigma}_{\tilde{\Theta}}$, and $\hat{\Sigma}_{\ddot{\Theta}}$ conclude our numerical studies. Computations are heavy so the range of these studies is restricted. First, we computed the mean of the internal variance estimator $\hat{\sigma}_{\hat{\theta}_1}{}^2$ (i.e., of the element in the first row and first column of $\hat{\Sigma}_{\hat{\Theta}}$) for the cases of Φ and sample sizes used in the previous evaluations of bias and variance (Table 4). The mean of this internal variance estimator, $E(\hat{\sigma}_{\hat{\theta}_1}{}^2)$, generally exceeds the actual variance of $\hat{\theta}_1$, $\sigma_{\hat{\theta}_1}{}^2$. As rules improve with sample size fixed, percent bias changes from large positive values to small negative values.

Percent bias under case 1 decreases sharply with increase of test sample size, but increases slightly with increase of mixed sample size (Table 4, lines 1 to 7); conceivably omission of sample points in our evaluations underlies the slight increase with mixed sample size. Under any case of Φ, the internal estimator or variance of $\hat{\theta}_1$ becomes nearly unbiased at the largest sample sizes examined.

Our last computations are of the mean and variance of the internal variance estimators ($\hat{\sigma}_{\hat{\theta}_1}{}^2$, $\hat{\sigma}_{\tilde{\theta}_1}{}^2$, and $\hat{\sigma}_{\ddot{\theta}_1}{}^2$) (i.e., of the elements in the first row and column of $\hat{\Sigma}_{\hat{\Theta}}$, $\hat{\Sigma}_{\tilde{\Theta}}$, and $\hat{\Sigma}_{\ddot{\Theta}}$, respectively) for the three cases of Φ with test and mixed samples all of size 20. Also we determined the actual probability that 90% and 95% simultaneous confidence intervals from Equation (16) using either $\hat{\Theta}$, $\tilde{\Theta}$, or $\ddot{\Theta}$, each with its internal variance estimator, $\hat{\Sigma}_{\hat{\Theta}}$, $\hat{\Sigma}_{\tilde{\Theta}}$, or $\hat{\Sigma}_{\ddot{\Theta}}$, cover the actual composition vector $\Theta' = (0.6, 0.4)$ (Table 5).

Comparison of actual variances of the estimators ($\hat{\theta}_1$, $\tilde{\theta}_1$, and $\ddot{\theta}_1$) (Table 5, line 1) with the mean of the corresponding internal variance estimators (Table 5, line 2) shows the positive bias of each internal estimator diminishes as rules improve. Only the internal estimator of variance of $\hat{\theta}_1$ becomes negatively biased. Percent bias (Table 5, line 3) of each estimator decreases sharply with improvement of rules.

Variance of the internal variance estimators of $\hat{\theta}_1$ and $\ddot{\theta}_1$ are manyfold greater than that of $\tilde{\theta}_1$ under case 1 and case 2. With improved rules of case 3, all internal variance estimators have comparable variance.

Probabilities that simultaneous confidence intervals for each estimator ($\hat{\Theta}$, $\tilde{\Theta}$, and $\ddot{\Theta}$) cover the

TABLE 4.—Variance, $\sigma_{\hat{\theta}_1}{}^2$, of estimator, $\hat{\theta}_1$; mean, $E(\hat{\sigma}_{\hat{\theta}_1}{}^2)$, of internal variance estimator, $\hat{\sigma}_{\hat{\theta}_1}{}^2$; and percent bias for indicated Φ-matrices; for indicated test and mixed sample sizes; and $\Theta' = (0.6, 0.4)$.

Φ	Test sample sizes	Mixed sample size	$\sigma_{\hat{\theta}_1}{}^2$	$E(\hat{\sigma}_{\hat{\theta}_1}{}^2)$	Percent bias
Case 1	20	[1]20	0.11232	0.24894	122
		[2]30	.08685	.19693	127
$\begin{bmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}$		[2]40	.07417	.17096	130
	30	[2]20	.07957	.10171	28
		[2]30	.05904	.07707	31
		[2]40	.04877	.06450	32
	40	[2]40	.04051	.04289	5.9
Case 2	20	[1]20	.04640	.04894	5.5
$\begin{bmatrix} 0.75 & 0.25 \\ 0.10 & 0.90 \end{bmatrix}$	30	[2]30	.02892	.02931	1.3
	40	[2]40	.02111	.02125	0.7
Case 3	20	[1]20	.02425	.02394	−1.3
$\begin{bmatrix} 0.90 & 0.10 \\ 0.10 & 0.90 \end{bmatrix}$	30	[2]30	.01579	.01562	−1.1

[1]Evaluated at all sample points except when $|\hat{\Phi}| = 0$.
[2]Evaluated only at sample points for which probability of observing the outcomes of the test samples $>10^{-6}$ and $|\hat{\Phi}| \neq 0$.

TABLE 5.—Variances of the estimators ($\hat{\theta}_1$, $\tilde{\theta}_1$, and $\ddot{\theta}_1$), means of internal variance estimators, percent bias of internal variance estimators, variances of internal variance estimators, and probabilities of coverage of $\Theta' = (0.6, 0.4)$ by simultaneous 90 and 95% confidence intervals for three cases of Φ when test and mixed samples are all of size 20.[1]

Φ		$\begin{bmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}$			$\begin{bmatrix} 0.75 & 0.25 \\ 0.10 & 0.90 \end{bmatrix}$			$\begin{bmatrix} 0.90 & 0.10 \\ 0.10 & 0.90 \end{bmatrix}$	
Estimator	$\hat{\theta}_1$	$\tilde{\theta}_1$	$\ddot{\theta}_1$	$\hat{\theta}_1$	$\tilde{\theta}_1$	$\ddot{\theta}_1$	$\hat{\theta}_1$	$\tilde{\theta}_1$	$\ddot{\theta}_1$
Variance of estimator	0.11221	0.06740	0.86543	0.04634	0.04009	0.04259	0.02425	0.02293	0.02290
Mean of internal variance estimator	0.24824	0.12030	7.75453	0.04859	0.04683	0.07374	0.02394	0.02387	0.02386
Percent bias of internal variance estimator	121	78	796	4.9	16.8	73	-1.3	4.1	4.2
Variance of internal variance estimator	25.2471	0.10567	125.688	0.11388	0.00175	439.969	9.521×10^{-5}	9.166×10^{-5}	9.140×10^{-5}
Probability of coverage of Θ by 90% confidence intervals	0.933	0.949	0.950	0.906	0.917	0.917	0.890	0.897	0.897
Probability of coverage of Θ by 95% confidence intervals	0.976	0.981	0.981	0.956	0.960	0.960	0.947	0.950	0.950

[1]Evaluations only include sample points for which probability of observing the outcome of the test samples $>10^{-6}$ and $|\Phi| \neq 0$.

parameter vector $\Theta' = (0.6, 0.4)$ approach the intended levels of confidence as rules improve (Table 5, lines 5 and 6). For rules of case 1 or case 2, the level of confidence provided by any of the estimators exceeds that intended; such is preferable to the converse because the intervals provide at least the level of confidence the investigator intends. Our normality assumption used to construct confidence intervals will be better satisfied as mixed and test sample sizes increase. Apparently the internal variance estimators become less biased as test sample size increases. Therefore, we anticipate the level of confidence of intervals from any of the estimators will more closely approach the intended level as test sample size increases even when rules are poor.

Limited as these numerical studies are, they demonstrate that when sample sizes are small and rules are poor, $\tilde{\Theta}$ should be used to estimate composition of a mixture. We found then that $\tilde{\Theta}$ is least biased, has smallest variance, and its internal variance estimator itself has smallest variance. With larger sample sizes or good rules of assignment, the estimators $\hat{\Theta}$, $\tilde{\Theta}$, and $\ddot{\Theta}$ appear more nearly equivalent.

Decisions on sample sizes depend on desired precision and the rules characterized by Φ. The closer Φ is to an identity matrix or, equivalently, the better the identification of stocks, the fewer required individuals in test and mixed samples to achieve desired precision of composition estimation. With an accurate initial estimate of Φ from the learning samples, the corresponding asymptotic variance-covariance matrix at Equation (11) can be used to estimate sample sizes needed to achieve required precision. We recall that variance of $\tilde{\theta}_1$ is well described by the asymptotic variance-covariance matrix even when rules are poor and sample sizes are small, providing another reason for preferring $\tilde{\Theta}$ to $\hat{\Theta}$ or $\ddot{\Theta}$ in that circumstance.

AFTERWORD

Withholding individuals of samples from the separate stocks to form test samples must result in less effective rules than if the learning and test samples were pooled for rule formation. Although the practice is repaid in part by the ability to evaluate precision of composition assessment, the penalty at rule development can be further alleviated. Roles of the two samples from each of the separate stocks can be interchanged; either can be the learning or test sample. If each of the samples from the segregated stocks is partitioned into two approximately equal sized subsamples, two sets of rules can be formed; two estimates of Φ obtained; two estimates of Θ computed by any of $\hat{\Theta}$, $\tilde{\Theta}$, or $\ddot{\Theta}$; and two internal estimates of the variance-covariance matrices ($\hat{\Sigma}_{\hat{\Theta}}$, $\hat{\Sigma}_{\tilde{\Theta}}$, or $\hat{\Sigma}_{\ddot{\Theta}}$) calculated. The pairs of estimates are statistically dependent. Nonetheless, means of pairs of estimates of Θ and Σ_{Θ} have the same expectation and presumably greater precision than the individual members of the pairs. Exact evaluation of that enhanced precision for estimates of the composition vector Θ does not appear easy; however, use of the mean of internal estimates of the variance-covariance matrix in calculation of the confidence set Equation (15) provides an unknown but greater level of confidence than the indicated $100(1 - \alpha)\%$ value.

LITERATURE CITED

ANAS, R. E., AND S. MURAI.
1969. Use of scale characters and a discriminant function

for classifying sockeye salmon (*Oncorhynchus nerka*) by continent of origin. Int. North Pac. Fish. Comm., Bull. 26:157-192.

COOK, R. C., AND G. E. LORD.
1978. Identification of stocks of Bristol Bay sockeye salmon, *Oncorhynchus nerka*, by evaluating scale patterns with a polynomial discriminant method. Fish. Bull., U.S. 76:415-423.

FUKUHARA, F. M., S. MURAI, J. J. LALANNE, AND A. SRIBHIBHADH.
1962. Continental origin of red salmon as determined from morphological characters. Int. North Pac. Fish. Comm., Bull. 8:15-109.

HILL, D. R.
1959. Some uses of statistical analysis in classifying races of American shad (*Alosa sapidissima*). U.S. Fish Wildl. Serv., Fish. Bull. 59:268-286.

KONOVALOV, S. M
1975. Differentiation of local populations of sockeye salmon *Oncorhynchus nerka* (Walbaum). (Translated from Russ. by Leda V. Sagen.) Univ. Wash. Publ. Fish., New Ser., 6, 290 p.

MORRISON, D. F.
1967. Multivariate statistical methods. McGraw-Hill, N.Y., 338 p.

PARSONS, L. S.
1972. Use of meristic characters and a discriminant function for classifying spring- and autumn-spawning Atlantic herring. Int. Comm. Northwest Atl. Fish. Res. Bull. 9:5-9.

WORLUND, D. D., AND R. A. FREDIN.
1962. Differentiation of stocks. *In* N. J. Wilimovsky (editor), Symposium on pink salmon, p. 143-153. H.R. MacMillan Lectures in Fish., Univ. B.C., Vancouver, Can.