STOCK IDENTIFICATION BY MAXIMUM LIKELIHOOD METHOD:
DERIVATION OF VARIANCE ESTIMATOR

Kazuhiko Hiramatsu
Far Seas Fisheries Research Laboratory

Hirohisa Kishino
The Institute of Statistical Mathematics

Fisheries Agency of Japan

September 1989

THIS PAPER MAY BE CITED IN THE FOLLOWING MANNER:
Hiramatsu, Kazuhiko and Hirohisa Kishino. 1989. Stock Identification
by Maximum Likelihood Method: Derivation of Variance Estimator.
(Document submitted to the Annual Meeting of the International North
Fisheries Agency of Japan, Far Seas Fisheries Research Laboratory,
5-7-1, Orido, Shimizu, Japan 424
Abstract

Evaluating the precision of stock composition needs to consider two sources of sampling error, i.e., sampling of baseline data and that of mixed fishery data. Variance estimator including both sources is derived from the Fisher's information matrix.

1. Introduction

Maximum likelihood method has recently been used to resolve composition of a mixture of stock from genetic data and scale data (Fournier et al. 1984; Pella and Milner 1987; Millar 1987; Wood et al. 1987; Millar 1988). The estimations are based on the differences in the distribution of the measured characteristics among stocks. The probability density function of characteristics for the j-th stock, $g^j$, is estimated by maximizing the likelihood function for the baseline samples of the j-th stock. Let $\pi_j$ be the proportion of fish in the mixed samples that comes from the j-th stock. The maximum likelihood estimates for $\pi_j$'s are found by maximizing the following likelihood function:

$$\prod_j \sum_k \pi_j g^j(y_k),$$

where $y_k$ is the vector of characteristics of the k-th fish in the mixed samples. The precision of composition estimation has been evaluated by the following methods.

1. Asymptotic covariance matrix derived by Fisher's information matrix (Pella and Milner 1987)
2. Simulation (Fournier et al. 1984; Wood et al. 1987) or bootstrap resampling (Millar 1987)
3. Infenitecimal jackknife (Millar 1987)

Although both the baseline and mixed samples variability should be considered for evaluating the precision, method 1 and 3 took into
account only the mixed sample variability. On the other hand, method 2 can take into account both sources of variability, but the amount of computing is considerable.

In this paper, we expand the method 1 and derive the variance estimator taking account of both the baseline and mixed samples variability. The calculations are presented in a step by step manner.

2. Symbols and Definitions

\( m \): The number of stock groups,
\( n_j \): the number of the baseline samples in the j-th stock,
\( y_k^j \): characteristics of the kth fish in the j-th stock (vector),
\( g^j(y|\theta^j) \): the probability density function of characteristics for the j-th stock,
\( \theta^j \): parameters which specify the distribution of characteristics of j-th stock.
\( \theta^j = (\theta^j_1, \ldots, \theta^j_p)^t \), where \( t \) means transposed.

\( n \): The number of the mixed samples,
\( y_k \): characteristics of the k-th fish in the mixed samples (vector),
\( \pi_j \): the proportion of the mixed fishery that are of the j-th stock (estimated parameter), where \( \sum \pi_j = 1 \).
\( \pi = (\pi_1, \ldots, \pi_m) \).

The log-likelihood function for baseline data is:

\[
L^j(\theta^j | y_1^j, \ldots, y_n_j) = \sum_{k=1}^{n_j} \log \{ g^j(y_k^j | \theta^j) \}, \quad j=1, \ldots, m. \tag{1}
\]

The log-likelihood function for mixed data is:

\[
L(\pi, \theta | y_1, \ldots, y_n) = \sum_{k=1}^{n} \log \{ f(y_k | \pi, \theta) \}, \tag{2}
\]

where

\[
f(y | \pi, \theta) = \sum_{j=1}^{m} \pi_j g^j(y | \theta^j)
\]

is the probability density function of characteristics for the mixed
samples.

\[ \nabla_{\theta_i} \equiv \left( \frac{\partial}{\partial \theta_{1}^j}, \frac{\partial}{\partial \theta_{2}^j}, \ldots, \frac{\partial}{\partial \theta_{p}^j} \right)^t, \]

\[ \nabla_{\theta} \equiv \left( \nabla_{\theta_1}^t, \ldots, \nabla_{\theta_m}^t \right)^t, \]

\[ \nabla_{\pi} \equiv \left( \frac{\partial}{\partial \pi_1}, \ldots, \frac{\partial}{\partial \pi_{m-1}} \right)^t. \]

\[ i_{\pi \pi} \equiv E \left[ \nabla_{\pi} \log(f) \cdot \nabla_{\pi}^t \log(f) \right] = -E \left[ \nabla_{\pi} \nabla_{\pi}^t \log(f) \right], \]

\[ i_{\pi \theta} \equiv E \left[ \nabla_{\pi} \log(f) \cdot \nabla_{\theta}^t \log(f) \right] = -E \left[ \nabla_{\pi} \nabla_{\theta}^t \log(f) \right], \]

\[ i_{\theta \theta_i} \equiv E \left[ \nabla_{\theta_i} \log(g^j) \cdot \nabla_{\theta_i}^t \log(g^j) \right] = -E \left[ \nabla_{\theta_i} \nabla_{\theta_i}^t \log(g^j) \right], \]

\[ i_{\pi \pi} = n \cdot i_{\pi \pi}, \quad i_{\pi \theta} = n \cdot i_{\pi \theta}, \quad i_{\theta \theta_i} = n_j \cdot i_{\theta \theta_i}, \]

where \( E [ \cdot ] \) is an expectation and \( i \) is Fisher's information matrix.

Maximum likelihood estimates \( \hat{\theta} \) and \( \hat{\pi} \) are the value of \( \theta \) and \( \pi \) which satisfy the equations

\[ \nabla_{\theta_i} L^j (\theta^j) = 0 \quad j = 1, \ldots, m, \]

\[ \nabla_{\pi} L(\pi, \hat{\theta}) = 0, \quad \text{and} \quad \pi_m = 1 - \sum_{j} \pi_j. \]  

\[ \text{(3)} \]

\[ \text{(4)} \]

3. Derivation of Variance Estimator of \( \theta \)

For brevity, we omit the superscript \( j \) in this section. By the Taylor series expansion of \( \nabla_{\theta} L(\hat{\theta}) \) about \( \theta \),

\[ \nabla_{\theta} L(\hat{\theta}) \approx \nabla_{\theta} L(\theta) + \nabla_{\theta} \nabla_{\theta}^t L(\theta) \cdot (\hat{\theta} - \theta), \]

\[ \text{(5)} \]
where \( \hat{\theta} \) is maximum likelihood estimate and \( \theta \) is the true value. Since \( \nabla_{\theta} L(\hat{\theta}) = 0 \),

\[
\hat{\theta} - \theta \approx - \left( \nabla_{\theta} \nabla_{\theta}^t L(\theta) \right)^{-1} \cdot \nabla_{\theta} L(\theta),
\]

\[
\sqrt{n} (\hat{\theta} - \theta) \approx - \left( \frac{1}{n} \nabla_{\theta} \nabla_{\theta}^t L(\theta) \right)^{-1} \cdot \frac{1}{\sqrt{n}} \nabla_{\theta} L(\theta)
\]

\[
= \left( \frac{1}{n} \sum_{k=1}^{n} \nabla_{\theta} \nabla_{\theta}^t \log \{ g(y_k | \theta) \} \right)^{-1} \cdot \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \nabla_{\theta} \log \{ g(y_k | \theta) \}.
\]

Taking the limit of \( n \to \infty \),

\[
\frac{1}{n} \sum_{k=1}^{n} \nabla_{\theta} \nabla_{\theta}^t \log \{ g(y_k | \theta) \} \to E \left[ \nabla_{\theta} \nabla_{\theta}^t \log \{ g(Y | \theta) \} \right],
\]

(law of large numbers)

where \( Y \) represents the random variable following the distribution \( g(\cdot | \theta) \), and

\[
\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \nabla_{\theta} \log \{ g(y_k | \theta) \} \to N(0, U),
\]

(central limit theorem)

where \( N(0, U) \) is the normal distribution with mean 0 and variance \( U \).

Here,

\[
U = V \left[ \nabla_{\theta} \log (g) \right]
\]

\[
= E \left[ \nabla_{\theta} \log (g) \cdot \nabla_{\theta}^t \log (g) \right]
\]

\[
= i_{\theta \theta}.
\]
Hence,

\[ \sqrt{n} (\hat{\theta} - \theta) \sim i_{ee}^{-1} \cdot N(0, i_{ee}) \]

\[ \sim N(0, i_{ee}^{-1}), \text{ asymptotically}. \]

Finally, the asymptotic variance-covariance matrix of \( \hat{\theta} \) is obtained.

\[ \mathbb{V} [\hat{\theta}] = (n_i_{ee})^{-1} \]

\[ = i_{ee}^{-1}. \tag{6} \]

It can be estimated by \( \{-\nabla_{\theta} \nabla_{\theta}^t L(\hat{\theta})\}^{-1} \).

4. Derivation of Variance Estimator of \( \pi \)

By the Taylor series expansion of \( \nabla_{\theta} L(\hat{\pi}, \hat{\theta}) \) about \( \pi \) and \( \theta \),

\[ \nabla_{\pi} L(\hat{\pi}, \hat{\theta}) \approx \nabla_{\pi} L(\pi, \theta) + \nabla_{\pi} \nabla_{\theta}^t L(\pi, \theta) \cdot (\hat{\pi} - \pi) \]

\[ + \nabla_{\pi} \nabla_{\theta}^t L(\pi, \theta) \cdot (\hat{\theta} - \theta). \tag{7} \]

Since \( \nabla_{\pi} L(\hat{\pi}, \hat{\theta}) = 0 \),

\[ \sqrt{n} (\hat{\pi} - \pi) \approx - \left( \frac{1}{n} \nabla_{\pi} \nabla_{\theta}^t L(\pi, \theta) \right)^{-1} \]

\[ \times \left\{ \frac{1}{\sqrt{n}} \nabla_{\pi} L(\pi, \theta) + \frac{1}{\sqrt{n}} \nabla_{\pi} \nabla_{\theta}^t L(\pi, \theta) \cdot (\hat{\theta} - \theta) \right\}. \tag{8} \]

Taking the limit of \( n \to \infty \),

\[ \frac{1}{n} \nabla_{\pi} \nabla_{\theta}^t L(\pi, \theta) \to - i_{\pi \theta}. \]
\[
\frac{1}{\sqrt{n}} \nabla_n L(\pi, \theta) \rightarrow N(0, \pi_{\pi}).
\]

As for the baseline samples, we take the limit \( n_j \rightarrow \infty \) in the way \( n_j/n \rightarrow a_j > 0 \). First, we decompose as follows:

\[
\frac{1}{\sqrt{n}} \nabla_n \nabla^t \eta \cdot (\hat{\theta} - \theta)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{j=1}^{m} \nabla_n \nabla^t \eta_j \cdot L(\pi, \theta) \cdot (\hat{\theta}^j - \theta^j)
\]

\[
= \sum_{j=1}^{m} \frac{1}{\sqrt{a_j \cdot n}} \nabla_n \nabla^t \eta_j \cdot L(\pi, \theta) \cdot \sqrt{n_j} (\hat{\theta}^j - \theta^j).
\]

In this equation,

\[
\frac{1}{n} \nabla_n \nabla^t \eta \cdot L(\pi, \theta) \rightarrow -i_{\eta^t}.
\]

\[
\sqrt{n_j} (\hat{\theta}^j - \theta^j) \rightarrow N(0, i_{\eta^t}^{-1}).
\]

Hence,

\[
\sqrt{n} (\hat{\pi} - \pi) \rightarrow i_{\pi}^{-1} \cdot \left( w - \sum_{j=1}^{m} \frac{1}{\sqrt{a_j}} i_{\eta^t} u^j \right),
\]

where \( w \) and \( u^j \)'s follow the normal distributions with means 0 and variances \( i_{\pi}^{-1} \) and \( -i_{\eta^t}^{-1} i_{\eta^t} \)'s, respectively. Since the surveys of the mixed populations and the baseline populations are independent, \( w \) and \( u^j \)'s are independent among others. Therefore,
In particular, the asymptotic variance-covariance matrix of \( \pi \) is given by

\[
V[\hat{\pi}] = (n_i \pi_{\pi})^{-1} \cdot (n_i \pi_{\pi} + \sum_{j=1}^{n_i} (n_i \pi_{\pi}) \cdot (n_i \pi_{\theta})^{-1} \cdot (n_i \pi_{\pi})) \cdot (n_i \pi_{\pi})^{-1}
\]

\[
= (I_{\pi \pi})^{-1} \cdot I_{\pi \theta} \cdot \left[ \begin{array}{ccc} I_{\theta \theta} & \cdots & I_{\theta \theta} \\ \vdots & \ddots & \vdots \\ I_{\theta \theta} & \cdots & I_{\theta \theta} \end{array} \right]^{-1} \cdot (I_{\pi \pi})^{-1}.
\]  

The first term represents the contribution of the mixed samples variability and the second term is that of the baseline samples variability.

This result is similar with the covariance matrix of the classification approach (equation (11) in Pella and Robertson 1978). \( I_{\pi \pi} \) and \( I_{\pi \theta} \) are estimated by \( \nabla_{\pi} \nabla_{\pi}^t L(\hat{\pi}, \hat{\theta}) \) and \( \nabla_{\pi} \nabla_{\theta}^t L(\hat{\pi}, \hat{\theta}) \), respectively.

According to the definition of \( L \),

\[
\frac{\partial L}{\partial \pi_j} = \sum_{k=1}^{n} \frac{g^j(y_k | \theta^j) - g^m(y_k | \theta^m)}{f(y_k | \pi, \theta)}
\]

\[
\frac{\partial^2 L}{\partial \pi_j \partial \pi_j'} = \sum_{k=1}^{n} \frac{g^j(y_k | \theta^j) - g^m(y_k | \theta^m)}{f(y_k | \pi, \theta)}
\]

\[
\frac{\partial^2 L}{\partial \pi_j \partial \pi_j'} = \sum_{k=1}^{n} \frac{g^j(y_k | \theta^j) - g^m(y_k | \theta^m)}{f(y_k | \pi, \theta)}
\]
\[- \sum_{k=1}^{n} \frac{1}{\mathbb{P}(y_k|\pi, \theta)} \left( g^j(y_k|\theta^j) - g^m(y_k|\theta^m) \right) \times \left( g^{j'}(y_k|\theta^{j'}) - g^{m'}(y_k|\theta^{m'}) \right) \]

\[
\frac{\partial^2 L}{\partial \theta_s \partial \pi_j} = \frac{\partial}{\partial \theta_s} \left( \sum_{k=1}^{n} g^j(y_k|\theta^j) - g^m(y_k|\theta^m) \right) \mathbb{P}(y_k|\pi, \theta) \]

\[
= \sum_{k=1}^{n} \frac{1}{\mathbb{P}(y_k|\pi, \theta)} \left\{ f \left( \frac{\partial g^j}{\partial \theta_s} - \frac{\partial g^m}{\partial \theta_s} \right) - (g^j - g^m) \frac{\partial f}{\partial \theta_s} \right\} \]

\[
= \sum_{k=1}^{n} \frac{1}{\mathbb{P}(y_k|\pi, \theta)} \left\{ f \left( \frac{\partial g^j}{\partial \theta_s} - \frac{\partial g^m}{\partial \theta_s} \right) \right. \]
\[
\left. - (g^j - g^m) \left( \sum_{j=1}^{m-1} \pi_j \frac{\partial g^j}{\partial \theta_s} + (1 - \sum_{j=1}^{m-1} \pi_j) \frac{\partial g^m}{\partial \theta_s} \right) \right\},
\]

where \( \theta_s \) is an element of vector \( \theta \).

For the evaluation of \( V[\hat{\pi}] \), we need further to calculate \( \frac{\partial g^j}{\partial \theta} \) and \( \nabla_\theta \nabla_\theta^j L^j(\hat{\theta}) \) for the baseline likelihoods. Although they are usually obtained numerically, we will obtain the explicit formula for the multivariate normal distribution in Appendix. Explicit expression is also possible for the multinomial distributions.

Acknowledgments

This study was carried out under the ISM (Institute of Statical Mathematics) Cooperative Research Program (89-ISM-CRP-77).
References


Appendix: Multivariate Normal Distribution

The INPFC is now using the multivariate normal distribution model for the distribution of baseline data characteristics (Millar 1988). In this model, we can obtain an explicit expression of the variance of $\theta$ and $\pi$.

Now, $g^j$ is given by

$$g^j(y|\mu^j, \Omega) \sim \frac{1}{|\Omega|^{1/2}} \exp\{-\frac{1}{2} (y-\mu^j)^T \Omega^{-1} (y-\mu^j)\},$$

where $\mu^j = (\mu_{1j}, \ldots, \mu_{pj})^T$ and $\Omega$ are means and variance-covariance matrix of the distributions, respectively. We assume that variance matrix is common for all baseline populations. $|\Omega|$ represents determinant of matrix $\Omega$. In this case,

$$\theta = (\mu_1, \ldots, \mu_p, \mu_{11}, \Omega_{12}, \ldots, \Omega_{pp})^T,$$

where $\Omega_{st}$ is an element of matrix $\Omega$. Since $\Omega$ is symmetric matrix, the number of free parameters specifying $\Omega$ is $p(p+1)/2$. For simplicity, however, we deal here all elements of $\Omega$ as the free parameters.

The results are the same. The log-likelihood function for baseline data is given by

$$L'(\mu, \Omega | y) = \sum_{j=1}^m \sum_{k=1}^{n_j} \log\{g^j(y_k^j|\mu^j, \Omega)\}$$

$$= -\frac{1}{2} \sum_{j=1}^m n_j \log|\Omega| - \frac{1}{2} \sum_{j=1}^m \sum_{k=1}^{n_j} (y_k^j - \mu^j)^T \Omega^{-1} (y_k^j - \mu^j).$$

The maximum likelihood estimate of $\mu$ and $\Omega$ are as follows:

$$\hat{\mu}_j^i = \frac{1}{n_j} \sum_{k=1}^{n_j} y_{k}^j,$$

$$\hat{\Omega}_{st} = \frac{1}{n_j} \sum_{k=1}^{n_j} (y_{k}^j - \hat{\mu}_j^i)(y_{k}^j - \hat{\mu}_j^i)^T.$$
\[ \hat{\Omega}_{st} = \frac{1}{\sum n_j} \sum_{j=1}^{n_j} \sum_{k=1}^{n_j} (y_{kj} - \hat{\mu}_j)^t \cdot (y_{kj} - \hat{\mu}_j)_t. \]

In order to estimate the variance of \( \theta \) and \( \pi \), we should calculate

\[ \frac{\partial g^j}{\partial \mu_{qj}}, \quad \frac{\partial g^j}{\partial \Omega_{st}}, \quad \frac{\partial^2 L'}{\partial \mu_{rj} \partial \mu_{qj}}, \quad \frac{\partial^2 L'}{\partial \mu_{qj} \partial \Omega_{st}}, \quad \text{and} \quad \frac{\partial^2 L'}{\partial \Omega_{qr} \partial \Omega_{st}}. \]

To begin with,

\[ \frac{\partial g^j}{\partial \mu_{qj}} = \{ e_q \Omega^{-1} (y - \mu^j) \} g^j, \]

where \( e_q = (0, \ldots, 0, 1, 0, \ldots, 0)^t \),

the q-th element

\[ \frac{\partial g^j}{\partial \Omega_{st}} = -\frac{1}{2} \exp \left\{ -\frac{1}{2} (y - \mu^j)^t \Omega^{-1} (y - \mu^j) \right\} \cdot \left\{ \frac{1}{|\Omega|^{3/2}} \frac{\partial |\Omega|}{\partial \Omega_{st}} \right\} \]

\[ + \frac{1}{|\Omega|^{1/2}} (y - \mu^j)^t \frac{\partial \Omega^{-1}}{\partial \Omega_{st}} (y - \mu^j) \}

\[ = -\frac{1}{2} g^j \left\{ \frac{\Delta_{st}}{|\Omega|} - (y - \mu^j)^t \Omega^{-1} E_{st} \Omega^{-1} (y - \mu^j) \right\}. \]

Here, \( \Delta_{st} \) is the st-th cofactor of the matrix \( \Omega \) and \( E_{st} \) is defined as

\[ E_{st} = 1 \text{ in } (s, t) \text{ element,} \]
\[ = 0 \text{ elsewhere.} \]
\[
\begin{align*}
\frac{\partial L'}{\partial \mu^j} &= \frac{1}{2} \sum_{k=1}^{n_j} \sum_{u=1}^{p} \left\{ (\Omega^{-1})_{qu} + (\Omega^{-1})_{uq} \right\} (y_{kj} - \mu^j)u, \\
\frac{\partial^2 L'}{\partial \mu^j \partial \mu^j} &= -n_j (\Omega^{-1})_{rq}.
\end{align*}
\]

\[
\begin{align*}
\frac{\partial L'}{\partial \Omega_{st}} &= -\frac{1}{2} \sum_{j=1}^{m} n_j (\Omega^{-1})_{st} + \frac{1}{2} \sum_{j=1}^{m} \sum_{k=1}^{n_j} \sum_{u=1}^{p} \left\{ (\Omega^{-1})_{us} (y_{kj} - \mu^j)u \right\} \\
&\times \left\{ \sum_{u=1}^{p} (\Omega^{-1})_{tu} (y_{kj} - \mu^j)u \right\}.
\end{align*}
\]

\[
\begin{align*}
\frac{\partial^2 L'}{\partial \mu^j \partial \Omega_{st}} &= -\frac{1}{2} \sum_{k=1}^{n_j} \sum_{u=1}^{p} \left\{ (\Omega^{-1})_{qs} \right\} \\
&\times \left\{ \sum_{u=1}^{p} (\Omega^{-1})_{tu} (y_{kj} - \mu^j)u \right\}.
\end{align*}
\]

\[
\begin{align*}
\frac{\partial^2 L'}{\partial \Omega_{qr} \partial \Omega_{st}} &= -\frac{1}{2} \sum_{j=1}^{m} \sum_{k=1}^{n_j} \frac{\partial (\Omega^{-1})_{st}}{\partial \Omega_{qr}} + \frac{1}{2} \sum_{j=1}^{m} \sum_{k=1}^{n_j} \sum_{u=1}^{p} \frac{\partial (\Omega^{-1})_{us}}{\partial \Omega_{qr}} \\
&\times (y_{kj} - \mu^j)u \times \left\{ \sum_{v=1}^{p} (\Omega^{-1})_{tv} (y_{kj} - \mu^j)_{v} \right\}.
\end{align*}
\]
$$+ \left\{ \sum_{u=1}^{p} (\Omega^{-1})_{us}(y_{kj} - \mu_j) \right\} \cdot \left\{ \sum_{v=1}^{p} \frac{\partial (\Omega^{-1})_{tv}}{\partial \Omega_{qr}}(y_{kj} - \mu_j) \right\} \right]$$

$$= \frac{1}{2} \sum_{j=1}^{m} n_j (\Omega^{-1})_{sa}(\Omega^{-1})_{rt}$$

$$- \frac{1}{2} \sum_{j=1}^{m} \sum_{k=1}^{n_j} \sum_{u=1}^{p} \left\{ \sum_{v=1}^{p} (\Omega^{-1})_{us}(\Omega^{-1})_{rs}(y_{kj} - \mu_j) \right\} \left\{ \sum_{v=1}^{p} (\Omega^{-1})_{tv}(y_{kj} - \mu_j) \right\}$$

$$+ \left\{ \sum_{u=1}^{p} (\Omega^{-1})_{us}(y_{kj} - \mu_j) \right\} \left\{ \sum_{v=1}^{p} (\Omega^{-1})_{rv}(y_{kj} - \mu_j) \right\}$$

$$= \frac{1}{2} \sum_{j=1}^{m} n_j (\Omega^{-1})_{sa}(\Omega^{-1})_{rt}$$

$$- \frac{1}{2} \sum_{j=1}^{m} \sum_{k=1}^{n_j} \sum_{u=1}^{p} \left\{ (\Omega^{-1})_{rs}\sum_{v=1}^{p} (\Omega^{-1})_{us}(y_{kj} - \mu_j) \right\} \left\{ \sum_{v=1}^{p} (\Omega^{-1})_{rv}(y_{kj} - \mu_j) \right\}$$

$$+ (\Omega^{-1})_{ta}\sum_{v=1}^{p} (\Omega^{-1})_{us}(y_{kj} - \mu_j) \left\{ \sum_{v=1}^{p} (\Omega^{-1})_{rv}(y_{kj} - \mu_j) \right\}.$$