Comparing Mixture Estimates by Parametric Bootstrapping Likelihood Ratios

Joel H. REYNOLDS and William D. TEMPLIN

Wildlife managers and researchers often need to estimate the relative contributions of distinct populations in a mixture of organisms. Increasingly, there is interest in comparing these mixture contributions across space or time. Comparisons usually just check for overlap in the interval estimates for each population contribution from each mixture. This method inflates Type I error rates, has limited power due to its focus on marginal comparisons, and employs a fundamentally inappropriate measure of mixture difference. Given the difficulty of defining an appropriate measure of mixture difference, a powerful alternative is to compare mixtures using a likelihood ratio test. In applications where the standard asymptotic theory does not hold, the null reference distribution can be obtained through parametric bootstrapping. In addition to testing simple hypotheses, a likelihood ratio framework encourages modeling the change in mixture contributions as a function of covariates. The method is demonstrated with an analysis of potential sampling bias in the estimation of population contributions to the commercial sockeye salmon (Oncorhynchus nerka) fishery in Upper Cook Inlet, Alaska.

Key Words: Compositional data; Compositional difference; Discrete mixture analysis; Genetic stock identification; Mixed stock analysis; Mixture homogeneity.

1. INTRODUCTION

Mixed stock analysis (MSA) estimates the relative contributions of distinct populations in a mixture of organisms. MSA is an important tool in fisheries management and research (Begg, Friedland, and Pearce 1999; Shaklee, Beacham, Seeb, and White 1999), marine mammal research (Pella and Masuda 2001), and wildlife management and conservation (Pearce et al. 2000). The underlying approach has been applied to problems ranging from estimating the percentage of genes from source or parental populations (Planes and Doherty 1997) to field investigations of owl diets (Do and McLachlan 1984). Although methods for
MSA estimation have appeared in the fisheries literature for many years (e.g., Cassie 1954; Grant, Milner, Krasnowski, and Utter 1980; Pella and Milner 1987), and much longer in the statistics literature (see the extensive surveys in Blischke 1963; Redner and Walker 1984; Titterington, Smith, and Makov 1985; McLachlan and Peel 2000), new applications continually require methodological extensions.

There is growing interest in mixture homogeneity in space or time. Researchers assess differences in mixture estimates from independent samples by looking across the samples for overlap of the confidence intervals for a given population’s contribution (e.g., Wilmot, Kondzela, Guthrie, and Masuda 1998; McParland, Ferguson, and Liskauskas 1999; Seeb and Crane 1999; Shaklee et al. 1999; Ruzzante, Taggart, Lang, and Cook 2000). This approach suffers from both inflated Type I error rates due to multiple testing and inflated Type II error rates due to focusing on marginal summary statistics. The approach employs a fundamentally inappropriate measure of mixture difference that ignores the dependence among contribution estimates arising from their constrained sum (see Section 5).

MSA commonly uses maximum likelihood estimation; we extend this framework to a conditional likelihood ratio test of mixture homogeneity across independent samples, similar in flavor to McLachlan’s (1987) approach to testing the number of mixture components. Asymptotic theory, Monte Carlo simulation, or parametric bootstrapping provide approximate $P$ values. In addition to testing simple hypotheses, a likelihood ratio framework encourages explicit modeling of mixtures as functions of covariates. The method is implemented in the latest release of the freeware Statistical Package for Analyzing Mixtures [SPAM; version 3.6 (Reynolds 2001) available online at http://www.cf.adfg.state.ak.us/geninfo/research/genetics/software/spamnage.php].

Section 2 presents a motivating application from the sockeye salmon ($Oncorhynchus nerka$) commercial fishery in Upper Cook Inlet, Alaska, then Section 3 introduces the basic finite mixture model, derives the likelihood ratio test of $M$-sample mixture homogeneity, and presents three approaches to approximating the null reference distribution. Section 4 shows how parametric bootstrapping is used to test mixture homogeneity for the sockeye salmon data. Section 5 compares the performance of the likelihood ratio method and the confidence interval method both for the current application and in general. Marginal measures of “mixture difference” appropriate to compositional data are briefly discussed. The Appendix extends the finite mixture model to two-stage sampling.

2. APPLICATION: COMPARING SALMON HARVEST MIXTURES

The sockeye salmon fishery in Upper Cook Inlet, Alaska (Figure 1) is a major component of the local economy. Over the last ten years, the total annual value of commercial fisheries harvests in the region ranged from US$ 8.8 to $111.1 million, with sockeye salmon comprising 80%–97% of the annual value (Ruesch and Fox 1999). The fishing fleet is very efficient; the approximately 600 drift gillnet vessels can harvest as much as 70% of the available fish in a single 12-hour opening (Seeb et al. 2000).
Most sockeye salmon home with precision, returning from the ocean to their natal habitats to spawn and then die (Burgner 1991). Adults may spawn in many diverse environments (i.e., rivers, sloughs, lake shores), but survival of their offspring generally depends on the offspring successfully emigrating shortly after emergence to rear in a lake. This reliance on rearing lakes is characteristic of the majority of the species, and contrasts with the other Pacific salmonids (Burgner 1991). Over time, the low rates of straying (spawning in a

Figure 1. Upper Cook Inlet, Alaska. Numbers locate each baseline population listed in Table 1 (from Seeb et al. 2000). Commercial harvests occurred in the Central District.
Table 1. Baseline Populations and Associated Reporting Regions for the Mixture Analysis of Commercially Harvested Sockeye Salmon in Upper Cook Inlet, Alaska from Seeb et al. (2000)

<table>
<thead>
<tr>
<th>Region</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>West Cook Inlet</td>
<td>1–Chilligan R., 2–Crescent Lk., 3–Wolverine Ck., 4–McArthur R., 5–Packers Lk., 6–Coal Ck.</td>
</tr>
<tr>
<td>Northeast Cook Inlet</td>
<td>26–Daniels Lk., 27–Bishop Ck., 28–Swanson R.</td>
</tr>
</tbody>
</table>

NOTE: Numbers refer to labels in Figure 1. Ck.–Creek, Lk.–Lake, R.–River.

...location other than the natal habitat) and the demands of different spawning environments can lead to significant genetic, morphometric, and behavioral differences between sockeye populations within a relatively small geographic area (e.g., Woody, Olsen, Reynolds, and Bentzen 2000).

Maintaining genetic diversity and future productivity in the face of more immediate demands for economic returns by highly efficient fishers requires that fishery managers accurately identify the harvest contributions of the major Upper Cook Inlet sockeye salmon stocks. Sustainable management will be very difficult, if not unachievable, otherwise. There are 44 genetically distinct populations, or stocks, within the major sockeye salmon-producing areas of Upper Cook Inlet (Seeb et al. 2000). Overharvesting any of the stocks reduces both the region’s genetic diversity, through loss of unique combinations of genetic characters, and the fishery’s economic value, as lost stocks will not contribute to future harvests.

2.1 THE PROBLEM

The mixture of interest is the sockeye harvest in the Central District fishery during a 12-hour opening (Figure 1). Each boat delivers its catch to 1 of 11 processors. Traditionally, the harvest was sampled only at the largest processor, Wards Cove (Seeb et al. 2000). To ascertain whether this procedure produces biased mixture estimates, replicate samples from a second processor, Salamatof Seafoods, Inc., were collected on four openings during the 1997 and 1998 seasons. We wish to test the equality of the mixture estimates from the two processors.

2.2 THE DATA

Previous mixed stock analyses on Upper Cook Inlet sockeye have used scale patterns (Marshall et al. 1987), parasites (Waltemeyer, Tarbox, and Brannian 1993), and genetic
markers (Grant et al. 1980; Seeb et al. 2000). Of these characteristics, genetic markers provided the most accurate and precise mixture estimates (Seeb et al. 2000).

Samples of spawning salmon were collected from each of the 44 baseline populations (Figure 1, Table 1) (Seeb et al. 2000), with a target sample size of 100 individuals per population (Allendorf and Phelps 1981; Waples 1990). Allozyme electrophoresis was used to detect each individual’s genotype at 27 discriminating unlinked neutral loci [Seeb et al. (2000) detailed the tissue collection, loci, laboratory analyses, population genetic analyses, and results]. Geography (Figure 1) and genetic diversity (Figure 2), followed by extensive simulation analyses, were used to define six regions, or population aggregates, that were reliably identified in mixture estimation (Seeb et al. 2000). A region had to demonstrate a contribution estimate of \( \geq 90\% \), averaged over 500 bootstrap resamples, from simulated

![Figure 2. Neighbor-joining tree (Saitou and Nei 1987) of Cavalli-Sforza and Edwards chord length genetic distances (Cavalli-Sforza and Edwards 1967) between Upper Cook Inlet sockeye salmon populations (Table 1, Figure 1). Genetic distances are calculated from 27 unlinked allozyme electrophoresis loci (data in Seeb et al. 2000). Regional membership of each population is indicated in the right-hand column.](image-url)
mixture samples of 400 genotypes, where each mixture sample was drawn uniformly from the populations within the region of interest, and each genotype was parametrically bootstrapped from its population of origin’s allele frequency estimates. Six regions were thus identified (Figure 1, Figure 2): Kenai, Kasilof, West Cook Inlet, Susitna/Yentna, Knik, and Northeast Cook Inlet. Most sockeye salmon come from the first four regions (Tobias and Tarbox 1999), all of which contain major river drainages (Figure 1).

The mixture sample at each processor was obtained by two-stage sampling: boats were randomly sampled from the incoming sequence of deliveries, and a random sample of sockeye salmon were selected from each boat’s catch. Forty boats were sampled at Wards Cove at a rate of 10 fish per boat, for a target sample size of 400 fish per period. Salamatof Seafoods Inc., the smaller processor, serves a fleet of 20–30 boats. In 1997, 10 to 15 fish were sampled per boat depending on the number of boats returning, for a target sample size of 400 fish per period. In 1998, the goal was revised to 10 fish per boat for a target of 200 fish per period. Allozyme electrophoresis was used to detect each individual’s genotype at the 27 loci previously used to characterize the baseline populations [Seeb et al. (2000) detailed the genetic tissue collection, laboratory analysis, and results].

3. METHODS

3.1 THE FINITE MIXTURE MODEL

A friend goes into a candy store. Two large jars contain both strawberry and licorice candies, but in different proportions. She randomly grabs some candy from each well-mixed jar (the baseline populations), places it a single bag (the mixture), pays for it, walks out, and hands it to you. She tells you the original proportions in each jar, then says you may have some candy if you can tell her what portion of the mixture came from each jar. This is a finite mixture problem (only two jars contributed) with a discrete characteristic (flavor) and known source population characteristics (flavor proportions in each jar).

Mixture identifiability requires that the probability density functions of the characteristics (e.g., flavors) are linearly independent across the set of populations (e.g., jars) [Teicher (1963) and Redner and Walker (1984) provided a survey of general identifiability results; Lindsay (1995) provided a thorough discussion for mixtures of discrete characteristics]. Characteristics commonly used in fisheries include parasite assemblages (Moles and Jensen 2000; Urawa, Nagasawa, Margolis, and Moles 1998), scale patterns (Marshall et al. 1987), morphometrics and meristics (Fournier, Beacham, Ridell, and Busack 1984), artificial tags such as thermal marks, coded wire tags, or fin clips (Ihssen et al. 1981) and, increasingly, genetic markers (Seeb and Crane 1999; Ruzzante et al. 2000). The following presentation assumes a discrete characteristic, though this assumption is not essential.

Let $n$ items be randomly sampled from a mixture of $J$ populations. Let the $j$th population contribute an unknown proportion $\theta_j \geq 0$ to the mixture, $\sum \theta_j = 1; \theta = (\theta_1, \ldots, \theta_J)$. If the characteristic measured on the $i$th sample observation is denoted by $x_i$, then the probability of observing the sample $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$ is
Comparing Mixture Estimates

\[
\Pr(\mathbf{x} | \theta, \pi) = \prod_{i=1}^{n} \Pr(x_i | \theta, \pi) = \prod_{i=1}^{n} \left\{ \sum_{j=1}^{J} \theta_j \Pr(x_i | \pi_j) \right\},
\]

(3.1)

where \( \pi_j \) is a column vector of parameters specifying the probability density function for the characteristic in population \( j \), and \( \pi \) is the matrix \( [\pi_1| \cdots |\pi_J] \). For a discrete characteristic with \( k \) possible outcomes, \( \pi_j = (\pi_{j,1}, \ldots, \pi_{j,k}) \), \( \pi_{j,i} \geq 0, \sum \pi_{j,i} = 1 \), the vector of multinomial probabilities. The model, and its extension below, assumes that all potentially contributing populations are included in the set \{Pop. 1, Pop. 2, \ldots, Pop. \( J \)\} (see Smouse, Waples, and Tworek 1990). Multivariate characteristics are incorporated by expanding the \( x_i, \pi_j \), and \( \Pr(x_i | \pi_j) \) terms.

### 3.1.1 Estimation

With discrete characteristics, estimating the proportions \( \theta \) requires information on each contributing population’s parameters, \( \pi_j \) (Teicher 1963). Given the mixture sample, \( \mathbf{X} \), and a sample from each baseline population, \( \mathbf{Y}_j \), most fisheries applications fix the population-specific nuisance parameters at their estimates from each baseline population sample, \( \pi_j = \hat{\pi}_j(\mathbf{Y}_j) \). The conditional maximum likelihood problem \( L(\theta | \mathbf{X}, \hat{\pi}_j(\mathbf{Y})) \) (Millar 1987) is then solved using the expectation-maximization algorithm (EM, Dempster, Laird, and Rubin 1977). This avoids the difficulties with multiple optima in the unconditional maximum likelihood problem (Kiefer and Wolfowitz 1956). Assuming adequate baseline sample sizes, there is generally little information on \( \pi_j \) in \( \mathbf{X} \) relative to \( \mathbf{Y} \) (Milner, Teel, Utter, and Burley 1981). Sampling uncertainty in \( \hat{\theta} \) is accounted for by nonparametric bootstrap resampling from \( \mathbf{X} \) and parametric bootstrap resampling the \( \hat{\pi}_j \), producing bootstrap confidence intervals for each marginal contribution estimate, \( \hat{\theta}_j \) (ADF&G 2000). CMLE can produce biased estimates when contributing populations are missing from the baseline or, in the case of discrete characters, when \( \hat{\pi}_j \) gives zero probability to outcomes occurring in Population \( j \) but unobserved in its sample, \( \mathbf{Y}_j \); that is, sampling zeros (Smouse et al. 1990). For MSA using genetic characteristics, the contribution from a missing population can be estimated using the EM algorithm (Pella and Milner 1987; Smouse et al. 1990). Sampling zeros can be handled by adjusting the initial frequency estimates (e.g., Rannala and Mountain 1997), using the EM algorithm (Smouse et al. 1990) or shifting to a Bayesian analysis (Pella and Masuda 2001).

### 3.2 Extension to Two Mixture Samples

Let \( m \) index \( M \) independent simple random samples from possibly different mixtures of the same baseline populations, \( \theta_1, \theta_2, \ldots, \theta_M \). For example, \( m \) could index samples taken through time or space. Following the previous notation, the general mixture model for the sequence of samples, \( \{ \mathbf{X}_1 = \{x_{1,1}, \ldots, x_{1,n_1}\}, \ldots, \mathbf{X}_M = \{x_{M,1}, \ldots, x_{M,n_M}\} \} \), allowing each sample to come from a different mixture, \( \theta_1, \theta_2, \ldots, \theta_M \), is
This model assumes each populations’ characteristic parameters, $\pi_j$, remain constant with regard to the index $m$. We revisit this point in Section 5.

### 3.2.1 Estimation

The general $M$ mixture sample model, with samples coming from potentially different mixtures, is estimated by fitting each mixture independent of the others. The constrained $M$ sample model, with samples coming from a common mixture $\theta_0$, is estimated by combining the mixture samples into a single sample then fitting. Both cases follow from the likelihood under (3.2). Unconditional estimation is considered in Section 5.

### 3.3 Testing Mixture Equality

Assume independent simple random samples from $M$ potentially different mixtures, each mixture consisting of contributions from a common set of known baseline populations. A likelihood ratio test of equality of the $M$ mixture proportions, $H_0 : \pi_m = \pi_0$ for $m = 1, \ldots, M$, versus the general alternative, follows directly from (3.2):

\[
LR = \frac{L(\theta_1, \theta_2, \ldots, \theta_M | X_1, X_2, \ldots, X_M, \hat{\pi})}{L(\theta_0, \theta_0, \ldots, \theta_0 | X_1, X_2, \ldots, X_M, \hat{\pi})} = \prod_{m=1}^{M} \prod_{i=1}^{n_m} \left\{ \sum_{j=1}^{J} \theta_{m,j} \Pr(x_{m,i} | \hat{\pi}_j) \right\}.
\]

#### 3.3.1 Null Reference Distribution Method 1: Asymptotic Theory

Mixture homogeneity can be tested by comparing $-2 \times \ln(LR)$ to a $\chi^2$ with degrees of freedom $d = (J - 1) \times (M - 1)$, its asymptotic distribution under the null model (Stuart, Ord, and Arnold 1999). However, the asymptotic relationship becomes unreliable when any of the $\hat{\theta}_j$ nears the boundary of the parameter space (Stuart et al. 1999), which is often the case in fisheries application (Millar 1987). The family of asymptotic distributions for likelihood ratio tests on the boundary of the parameter space is known (Self and Liang 1987), but not easily employed.

#### 3.3.2 Null Reference Distribution Method 2: Monte Carlo Simulation ($\theta_0$ Known)

Monte Carlo simulation can approximate the null distribution if $\theta_0$ is known a priori
Let \( n_m \) be the number of observations in mixture sample \( m \), \( m = 1, \ldots, M \). Simulate \( R \) sets of \( n = \sum n_m \) observations from the single mixture model (3.1) by first drawing \( n \) observations from the multinomial with parameters \( \theta_0 \), giving the number of simulated observations originating from each source population. For each source population, draw that many observations from the population’s probability density function for the characteristics of interest, \( \hat{\pi}_j \). Take the full set of \( n \) simulated observations and fit (3.1), obtaining \( \hat{\theta}_0, \hat{\pi}_j \). Randomly partition the observations into \( M \) sets of size \( \{n_1, \ldots, n_M\} \) and fit (3.2), obtaining \( \{\hat{\theta}_{m,rr} : m = 1, \ldots, M\} \). The likelihood ratio, \( LR_{rr} \), is obtained from plugging the estimates into (3.3). The observed likelihood ratio, \( LR_{obs} \), is obtained by using the actual observations to fit (3.1) and (3.2), then plugging the estimates into (3.3). An approximate \( P \) value for the test is

\[
\frac{1 + \sum_r I(LR_{rr} \geq LR_{obs})}{1 + R}
\]

(Davison and Hinkley 1997), where the function \( I() \) is one when the argument is true and zero otherwise. \( R \) in the range 1,000–10,000 guarantees little loss of power due to finite simulation (Davison and Hinkley 1997, sec. 4.2.5).

### 3.3.3 Null Reference Distribution Method 3: Parametric Bootstrapping

(\( \theta_0 \) Unknown)

Usually \( \theta_0 \) is unknown prior to the analysis. One can estimate \( \theta_0 \) using the combined observations, then perform parametric bootstrapping (Davison and Hinkley 1997) to approximate the null reference distribution by following the steps outlined above using \( \hat{\theta}_0 \). In essence, this extends McLachlan’s (1987) approach for testing the number of components in a mixture to testing mixture homogeneity.

Uncertainty in the conditional values of the nuisance parameters, \( \hat{\pi}_j \), is incorporated in either simulation approach by parametric bootstrap resampling the baseline samples and re-estimating \( \hat{\pi}_j \) before simulating the null mixture sample during each iteration.

Significant mixture differences may lead to a model selection process exploring less-constrained null models. For example, we could fit a model where some of the \( M \) samples come from identical mixtures or where the \( M \) samples only differ in the contributions of a subset of the baseline populations. The software package SPAM (Reynolds 2001) currently allows the former investigation but not the latter.

### 3.4 Sockeye Application

Parametric bootstrapping was used to test the null hypothesis that the two processor samples came from the same mixture (\( R = 5,000 \) simulations). All mixture simulations and model fitting were done in SPAM (Reynolds 2001) using CMLE based on the EM algorithm and/or a conjugate gradient search algorithm [for algorithm implementation details see Pella, Masuda, and Nelson (1996)]; final analysis of the simulation results was conducted.
in S-Plus 2000 (Insightful, Inc., Seattle, WA). The $\hat{\pi}_j$, parameterizing population $j$’s allele frequency distribution for each of 27 genetic markers, were parametrically bootstrapped before generating each null mixture simulation to account for baseline uncertainty in the null reference distribution. The mixture samples are from two-stage sampling. Extending (3.2) to this sampling design (see the Appendix) results in a likelihood ratio identical to (3.3), so details of simulating the null reference distribution remain as given above.

### 4. RESULTS

The likelihood ratio test showed evidence that processors sampled different harvest mixtures for one of the four opening dates (Table 2). Ninety percent bootstrap confidence intervals were calculated for each opening to compare with other published testing procedures (Figure 3) and for a posteriori insight when mixture estimates were found to differ (Section 5). Mixture homogeneity and its associated likelihood ratio test focus on baseline populations, but we followed the general practice of presenting and publishing region interval estimates.

Two sets of Efron’s percentile intervals were calculated using 1,000 resamples (Davison and Hinkley 1997).

1. For $\sum_{j \text{in Region } h} \hat{\theta}_{\text{processor } A,j}$, the processor-specific estimates of each region’s total contribution (Table 2, Figure 3); published assessments of mixture homogeneity check whether each region’s intervals overlap across mixtures (e.g., processors).

2. For the difference in processor-specific estimates of each region’s total contribution,

$$\sum_{j \text{ in Region } h} \left( \hat{\theta}_{\text{processor } A,j} - \hat{\theta}_{\text{processor } B,j} \right)$$

(Table 2), a “natural” extension of approach 1. Both approaches ignore the dependence among region estimates and therefore should not be relied upon in practice (see Section 5).

In the July 14, 1997, opening, the boat from which each fish was sampled was not recorded, making it impossible to replicate the two-stage sampling in the bootstrap confidence interval calculations. All interval estimates (Table 2, Figure 3) therefore assume simple random sampling and hence may underestimate the true variance. Interval calculations incorporated parametric resampling of $\hat{\pi}_j$ and nonparametric resampling of each mixture sample.

### 5. DISCUSSION

#### 5.1 Upper Cook Inlet Sockeye Salmon

Processor-specific mixture differences may arise when spatial heterogeneity in the harvestable Central District sockeye salmon mixture occurs with clustering, during harvest, of boats delivering to a specific processor. If such clustering regularly occurs, then the current harvest-sampling plan may need to be revised. One solution would be a weighted
<table>
<thead>
<tr>
<th>Opening date</th>
<th>Processor</th>
<th>N</th>
<th>West Cook Inlet</th>
<th>Southeast Yellowtail</th>
<th>Knik</th>
<th>Northeast Cook Inlet</th>
<th>Kennicott</th>
<th>Kukakof</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 July 1987</td>
<td>WC</td>
<td>369</td>
<td>0.00</td>
<td>0.16</td>
<td>0.02</td>
<td>0.02</td>
<td>0.79</td>
<td>0.03</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Sal.</td>
<td>361</td>
<td>0.06</td>
<td>0.05</td>
<td>0.02</td>
<td>0.00</td>
<td>0.84</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.01, 0.12)</td>
<td>(0.12, 0.22)</td>
<td>(0.07, 0.22)</td>
<td>(0.01, 0.01)</td>
<td>(0.73, 0.97)</td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WC-Sal</td>
<td></td>
<td>(0.06, 0.06)</td>
<td>(0.06, 0.06)</td>
<td>(0.01, 0.01)</td>
<td>(0.06, 0.06)</td>
<td>(0.75, 0.91)</td>
<td>(0.06)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>90% CI</td>
<td></td>
<td>(−0.12, 0.02)</td>
<td>(−0.18, 0.01)</td>
<td>(−0.01, 0.01)</td>
<td>(−0.16, 0.07)</td>
<td>(−0.08, 0.07)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 July 1987</td>
<td>WC</td>
<td>389</td>
<td>0.02</td>
<td>0.07</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
<td>0.02</td>
<td>0.171</td>
</tr>
<tr>
<td></td>
<td>Sal.</td>
<td>364</td>
<td>0.00</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.85</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.01, 0.13)</td>
<td>(0.01, 0.12)</td>
<td>(0.01, 0.01)</td>
<td>(0.02, 0.02)</td>
<td>(0.76, 0.92)</td>
<td>(0.06)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WC-Sal</td>
<td></td>
<td>(0.06, 0.06)</td>
<td>(0.06, 0.06)</td>
<td>(0.02, 0.02)</td>
<td>(0.06, 0.06)</td>
<td>(0.07, 0.07)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>90% CI</td>
<td></td>
<td>(−0.03, 0.05)</td>
<td>(−0.08, 0.07)</td>
<td>(−0.02, 0.02)</td>
<td>(−0.01, 0.01)</td>
<td>(−0.07, 0.07)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 July 1988</td>
<td>WC</td>
<td>364</td>
<td>0.04</td>
<td>0.31</td>
<td>0.08</td>
<td>0.01</td>
<td>0.26</td>
<td>0.19</td>
<td>0.280</td>
</tr>
<tr>
<td></td>
<td>Sal.</td>
<td>159</td>
<td>0.01</td>
<td>0.25</td>
<td>0.01</td>
<td>0.01</td>
<td>0.51</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.21, 0.40)</td>
<td>(0.21, 0.40)</td>
<td>(0.26, 0.45)</td>
<td>(0.26, 0.45)</td>
<td>(0.28, 0.50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>WC-Sal</td>
<td></td>
<td>(−0.18, 0.09)</td>
<td>(−0.10, 0.24)</td>
<td>(−0.08, 0.14)</td>
<td>(−0.04, 0.02)</td>
<td>(−0.06, 0.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>90% CI</td>
<td></td>
<td>(−0.18, 0.09)</td>
<td>(−0.10, 0.24)</td>
<td>(−0.08, 0.14)</td>
<td>(−0.04, 0.02)</td>
<td>(−0.06, 0.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 July 1988</td>
<td>WC</td>
<td>369</td>
<td>0.05</td>
<td>0.57</td>
<td>0.04</td>
<td>0.00</td>
<td>0.63</td>
<td>0.01</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td>Sal.</td>
<td>197</td>
<td>0.01</td>
<td>0.27</td>
<td>0.09</td>
<td>0.00</td>
<td>0.69</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.11, 0.30)</td>
<td>(0.22, 0.43)</td>
<td>(0.24, 0.43)</td>
<td>(0.24, 0.43)</td>
<td>(0.26, 0.43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>WC-Sal</td>
<td></td>
<td>(−0.07, 0.15)</td>
<td>(−0.11, 0.22)</td>
<td>(−0.16, 0.05)</td>
<td>(−0.02, 0.01)</td>
<td>(−0.17, 0.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>90% CI</td>
<td></td>
<td>(−0.07, 0.15)</td>
<td>(−0.11, 0.22)</td>
<td>(−0.16, 0.05)</td>
<td>(−0.02, 0.01)</td>
<td>(−0.17, 0.17)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Processor sampled: WC—Wards Cove, Sal.—Sealsetof Seafoods, Inc. Conditional maximum likelihood mixture estimates are shown for each sample of size N, where contributions from each of the 44 baseline populations were estimated then summed into the management regions for display (see Figure 1 and Table 1). Ninety percent bootstrap confidence intervals (Miller’s percentile method; 1,000 replications, Davidson and Hinckley 1997) are given for both the processor-specific region contribution estimates (listed below the contribution point estimates, see also Figure 8) and the marginal difference in processor-specific region contribution estimates (labelled "90% CI WC-Sal"). Although commonly used to assess mixture homogeneity (Do the processor-specific intervals overlap? Do the marginal difference intervals contain zero?), both approaches have low power to detect mixture differences as they ignore the dependence among region contributions. Approximate P values are from parametric bootstrapping the likelihood ratio test of mixture homogeneity using N = 5,000 replications. Note that the likelihood ratio tests the baseline population contributions, not the region contributions.
average of mixture estimates across processors, with weight proportional to each processor’s portion of the total harvest.

5.2 Method Comparison

Mixture homogeneity should not be assessed using marginal confidence intervals. Checking for overlapping marginal confidence intervals both inflates Type I error rates, due to the simultaneous inference across $M$ mixture samples, and inflates Type II error rates due to the use of marginal (region-specific) measures of mixture difference. The Type I inflation could be partially accounted for by using $(1 - \alpha)^{(1/M)}$ 100% level intervals (Hsu 1996), but this would not assuage the inflation due to repeating the “overlap check” across $J - 1$ sets of intervals.

Fundamentally, these marginal comparisons entail a loss of power and employ an inappropriate measure of mixture difference. Both deficiencies stem from ignoring the intercomponent dependencies. Mixture estimates live in the simplex, $\hat{\theta}_i \geq 0$, $\sum \hat{\theta}_i = 1$; changing one component’s contribution necessitates changing another’s. This strong dependence guaran-
Figure 4. Three-component projections of the processor-specific nonparametric bootstrap mixture estimates for the July 14, 1997, collections. The processor-specific mixture estimates differed at three of six regions (Table 2), so results are displayed for the four-component mixture (West Cook Inlet, Susitna/Yentna, Kenai, All Others). Rather than display the tetrahedral sample space, all four three-component projections are shown. Each projection was obtained by dropping a component and renormalizing—in effect, shining a light from the dropped vertex of the tetrahedron and marking the shadows cast on the far wall; the walls were then laid flat. Wards Cove resamples (left), Salamatof Seafoods, Inc. resamples (right). The closer a point is to a vertex in a ternary diagram, the greater the contribution of that component to the three-component mixture. Ternary diagrams compress distances between mixtures that fall near the boundaries (Billheimer et al. 2001), visually masking substantive mixture differences. Even so, the Wards Cove sample mixture clearly has less Kenai and West Cook Inlet contributions and more Susitna/Yentna contributions than the Salamatof Seafoods, Inc. sample mixture (left vs. right figures). This difference underlies the significant likelihood ratio test result (Table 2).

COMPARING MIXTURE ESTIMATES

Figure 4. Three-component projections of the processor-specific nonparametric bootstrap mixture estimates for the July 14, 1997, collections. The processor-specific mixture estimates differed at three of six regions (Table 2), so results are displayed for the four-component mixture (West Cook Inlet, Susitna/Yentna, Kenai, All Others). Rather than display the tetrahedral sample space, all four three-component projections are shown. Each projection was obtained by dropping a component and renormalizing—in effect, shining a light from the dropped vertex of the tetrahedron and marking the shadows cast on the far wall; the walls were then laid flat. Wards Cove resamples (left), Salamatof Seafoods, Inc. resamples (right). The closer a point is to a vertex in a ternary diagram, the greater the contribution of that component to the three-component mixture. Ternary diagrams compress distances between mixtures that fall near the boundaries (Billheimer et al. 2001), visually masking substantive mixture differences. Even so, the Wards Cove sample mixture clearly has less Kenai and West Cook Inlet contributions and more Susitna/Yentna contributions than the Salamatof Seafoods, Inc. sample mixture (left vs. right figures). This difference underlies the significant likelihood ratio test result (Table 2).

tees a loss of power when assessing marginal rather than joint summaries. The marginal focus of both confidence interval approaches prevents either method from detecting the mixture difference on the July 14, 1997, sampling event (Table 2, Figures 3, 4).

The implicit marginal measure of mixture difference, \( \Delta(\theta_A, \theta_B) = \{\theta_{A,i} - \theta_{B,i}\} \), assumes an additive structure that ignores intercomponent dependence (Aitchison 1992). Thus, while the mixture difference on the July 14, 1997, sampling event appears driven by simultaneous shifts in the contributions from West Cook Inlet, Susitna/Yentna, and Kenai regions (Table 2, Figure 4), the marginal difference confidence intervals ("90% CI WC – Sal", Table 2) only suggest a possible shift in the Susitna/Yentna contribution.

Appropriate measures of mixture difference, ones that capture the intercomponent dependence, are an ongoing research topic in compositional data analysis (e.g., Aitchison 1982, 1986, 1992; Billheimer, Guttorp, and Fagan 2001). Both the additive and the centered log-ratio transformations of Aitchison (1982, 1986) lead to a metric on the simplex (Aitchison 1992). Billheimer et al. (2001) have extended this to a regression framework in conjunction with the logistic normal distribution (Aitchison 1982, 1986). Unfortunately, this alternative approach to testing mixture homogeneity requires strictly positive \( \hat{\theta_i} \), severely limiting practical implementation. Further, the metric can be difficult to interpret (Billheimer et al. 2001).
A more subtle criticism of the interval overlap method is that it assesses mixture equality not at the baseline population level but at the region aggregate level, potentially obscuring relevant population level heterogeneity.

Researchers must use caution investigating mixture homogeneity and interpreting marginal contribution interval estimates. The likelihood ratio approach provides a test of mixture homogeneity that accounts for the inter-component dependencies while offering better control of Type I and Type II error rates. It can be adapted to a test of regional aggregates, if desired. More importantly, the approach may be extended to a paradigm for developing models of mixture variation across time or space.

5.3 Model Extensions

The \( M \)-mixture model can be extended to allow the baseline population parameters, \( \pi_j \), to change with the mixture index \( m \). This would require population-specific learning samples from each occasion of change.

6. CONCLUSIONS

Mixed stock analysis has a long history in fisheries and wildlife management. Advances in genetic marker technology have magnified the method’s importance by simplifying the collection and analysis of field samples, thus encouraging the development of extensive baseline population databases and the sampling of mixtures through space or time. However, the common methods of testing mixture homogeneity in space or time are fraught with statistical deficiencies. Although confidence interval methods may provide some insight, researchers must use caution interpreting these marginal measures of difference as they ignore the intercomponent dependence and therefore suffer low power. The likelihood ratio test presented here provides a statistically sound method for such tests while researchers await the development of more direct, and readily interpretable, measures of mixture difference.

The likelihood ratio approach can be used to develop more refined models of mixture variation, providing greater insight into wild populations subject to research and management. Such efforts can provide insight into the adequacy of mixture sampling protocols (illustrated here), investigation of marine migration patterns (Seeb and Crane 1999), and temporal and spatial stability of scientifically or economically important mixtures (Ruzzante et al. 2000).

APPENDIX: TWO-STAGE SAMPLING \( M \)-MIXTURE MODEL

Mixture samples in the sockeye harvest application were obtained by two-stage sampling. Model (3.2) is easily extended to this case. Let \( k \) index the sequence of \( K_m \) primary sampling units randomly selected from the \( m \)th of \( M \) independent mixtures. Let \( i_k \) index the sequence of \( n_{mk} \) secondary sampling units randomly selected from the \( k \)th primary unit.
Comparing Mixture Estimates

from the $m$th mixture. The possibly multivariate characteristic observed on the secondary sampling unit $i_k$ in the $m$th mixture is denoted $x_{m,i_k}$. Following the text, $\theta_{m,j} \geq 0$ is the unknown proportion of the $m$th mixture contributed by population $j$ (out of $J$ contributing populations), $\sum_j \theta_{m,j} = 1$ for each $m$, and $\hat{\theta}_j$ is the vector of parameter estimates specifying the characteristic distributions in population $j$. The resulting likelihood ratio for testing homogeneity of the $M$ mixtures is:

$$
\frac{\prod_{m=1}^{M} \left( \prod_{k=1}^{n_{m,k}} \sum_{j=1}^{J} \theta_{m,j} \Pr(x_{m,i_k} | \hat{\theta}_j) \right)}{\prod_{m=1}^{M} \left( \prod_{i=1}^{\Sigma n_{m,y}} \sum_{j=1}^{J} \theta_{0,j} \Pr(x_{m,i} | \hat{\theta}_j) \right)}
$$

Each mixture is assumed homogeneous across its associated primary sampling units, so the likelihood ratio under two-stage sampling (A.1) reduces to that for simple random sampling (3.3).

**ACKNOWLEDGMENTS**

Sincere thanks to the following colleagues for helpful discussions and/or reviews of earlier drafts: Eric Anderson, Ed Debevec, David Evans, Rich Hinrichsen, and two anonymous reviewers. One reviewer’s comment indirectly led to a greater awareness of the problem of defining a distance metric for compositional data; sincere thanks to Dean Billheimer for conversations expanding on that theme.

The genetics data for the population baselines and the Wards Cove mixtures are available in Seeb et al. (2000). The genetics data from the Salamatof Seafoods, Inc. mixtures are available by contacting Dr. Lisa Seeb at the Gene Conservation Laboratory: lisa.seeb@fishgame.state.ak.us.

Contribution PP-204 of the Alaska Department of Fish and Game, Commercial Fisheries Division, Juneau, Alaska, USA.

[Received November 2000. Revised March 2003.]

**REFERENCES**


J. H. REYNOLDS AND W. D. TEMPLIN


C
OMPARING
MIXTURE
ESTIMATES
73


