Optimal sampling design for using the age-length key to estimate age composition of a fish population

Han-Lin Lai
Washington State Department of Fisheries
1111 Washington Street SE, P.O. Box 43144
Olympia, Washington 98504-3144

An age-length key (ALK) is the traditional method for estimating age composition of an application of fish population. Tanaka (1953) showed that the ALK method is a double-sampling technique (Cochran 1977). The first stage uses simple random sampling to collect a very large, but relatively inexpensive, length sample. The second stage subsamples a small number of fish from the first-stage length samples for relatively costly age determination (age subsample). Age subsampling in the second stage can be taken in two fundamental ways, one in which age subsamples are randomly taken from the entire first-stage length samples (random-age subsampling; Kimura 1977) and another in which age subsamples are taken from each stratified length-stratum (stratified-age subsampling; Tanaka 1953). In stratified-age subsampling, fixed-age subsampling (a constant number of age specimens is taken at each length-stratum) and proportional age subsampling (age specimens is proportional to the random length-frequency) are the most popular (Ketchen 1949). However, a general stratified-age subsampling (number of age specimens varied at each length-stratum) can also be used. Stratified-age subsampling is the major focus of this paper. The similarities in results obtained from proportional and random-age subsamplings are given in the Discussion.

This paper was motivated by two articles, Lai (1987) and Jinn et al. (1987), using optimal sampling designs to estimate age composition of a fish population using ALK. The two articles differ significantly. Lai (1987) was based on the classic double-sampling technique and derived the optimal allocation of length samples and age subsamples using Kimura's Vartot (Kimura 1977) and the Cauchy-Schwartz inequality (Kendall & Stuart 1977) for fixed-age subsampling and proportional-age subsampling. Lai (1987) incorrectly used random-age subsampling for proportional-age subsampling. In contrast, Jinn et al. (1987) used a Bayesian approach to estimate age composition, variance, and covariance for a general stratified-age subsampling. They used the iterative method of Roa & Ghangurde (1972) to obtain the optimal allocation of length samples and age subsamples for each length-stratum, with a set of known per-unit costs for ageing a fish in each stratum.

The length-based optimal sampling design of Jinn et al. (1987) has advantages. Because the age of a fish can be expressed as a function of its length (e.g., a von Bertalanffy growth relationship) and because older fish are more difficult to age, the per-unit cost of ageing a fish can be used to estimate the difficulty of ageing older (larger) fish. In addition, the covariance components are important statistics because the sum of all age proportions equals 1, indicating that the estimates of age composition are not mutually independent. The disadvantage of the Bayesian approach of Jinn et al. (1987) is that their method is mathematically complicated and does not provide explicit expressions for the optimal allocations of length and age samples for an ALK, and thus requires substantial computing effort.

The purpose of this paper is to derive a length-based optimal sampling design for an ALK using a classic double-sampling technique. The covariance of age composition also is derived using the method of Kimura (1977). This paper also provides answers to the question many fishery scientists have asked me: What is the explicit solution to a length-based optimal sampling design for an ALK? A discussion on the general applicability of ALK in the sampling program with complexity of fishery-time-areal stratification and tows/trips clusterization is also provided.

Methods

I use the following notation for an ALK with a general stratified-age subsampling:

- N = total number of length samples;
- N_i = number of fish in the ith length-stratum, i=1,...,L;
- l_i = proportion of fish in the ith length-stratum, ($\hat{l}_i = N_i/N$);
- n_i = number of age sub-samples randomly taken from the ith length-stratum;
- n_{ij} = number of fish from n_i assigned to the jth age-class;
- q_{ij} = proportion of fish in the ith length-stratum that fall into the jth age-class ($\hat{q}_{ij} = n_{ij}/n_i$);
- A = number of age-classes;
- L = number of length-strata;
- p_j = proportion of population in the jth age-class;

Manuscript accepted 10 December 1992.
Fishery Bulletin, U.S. 92:382-388 (1993).

$Var(p_j)$ = variance of p_j;
$Cov(p_j,p_k)$ = covariance between p_j and p_k.
A caret denotes the estimate of each variable.

The unbiased estimate of p_j (Tanaka 1953) from an ALK is

$$\hat{p}_j = \sum_{i=1}^{L} \hat{l}_i \hat{q}_{ij} . \quad (1)$$

The variance of $\hat{p}_j$ has been derived by Tanaka (1953) and Kimura (1977); however, the approximate form of variance is more frequently used (Kutkuhn 1963, Southward 1963, Doubleday & Rivard 1983, Lai 1987):

$$Var(\hat{p}_j) \approx \sum_{i=1}^{L}\left[\frac{\hat{l}_i^2 \hat{q}_{ij}(1-\hat{q}_{ij})}{n_i} + \frac{\hat{l}_i(\hat{q}_{ij}-\hat{p}_j)^2}{N}\right]. \quad (2)$$

The terms in the right-hand side of Eq. 2 represent the portion of the total variance due to variation within length-strata and that due to variation between strata, respectively.

The covariance of $\hat{p}_j$ and $\hat{p}_k$ is derived using the method of Kimura (1977):

$$Cov(\hat{p}_j,\hat{p}_k) = \sum_{i=1}^{L}\frac{-\hat{l}_i(1-\hat{l}_i)\hat{q}_{ij}\hat{q}_{ik}}{Nn_i} + \sum_{i=1}^{L}\frac{-\hat{l}_i^2\hat{q}_{ij}\hat{q}_{ik}}{n_i} + \sum_{i=1}^{L}\frac{\hat{l}_i\hat{q}_{ij}\hat{q}_{ik}}{N} - \frac{\hat{p}_j\hat{p}_k}{N}$$
$$\approx \sum_{i=1}^{L}\frac{-\hat{l}_i^2\hat{q}_{ij}\hat{q}_{ik}}{n_i} + \sum_{i=1}^{L}\frac{\hat{l}_i\hat{q}_{ij}\hat{q}_{ik}}{N} - \frac{\hat{p}_j\hat{p}_k}{N}. \quad (3)$$

The approximate form of covariance omits the first term because this term is small compared with the sum of the other two terms. A quadratic loss function (Jinn et al. 1987) is used to infer the precision of the estimated age composition $\hat{\mathbf{p}}'=(\hat{p}_1,\hat{p}_2,\ldots,\hat{p}_A)$:

$$\mathcal{L}(\hat{\mathbf{p}},\mathbf{p}) = (\hat{\mathbf{p}}-\mathbf{p})'\mathbf{W}(\hat{\mathbf{p}}-\mathbf{p})$$
$$= \sum_{j=1}^{A} w_{jj} Var(\hat{p}_j) + \sum_{j\neq k}^{A} w_{jk} Cov(\hat{p}_j,\hat{p}_k)$$
$$= E\left[\sum_{j=1}^{A} w_{jj}(\hat{p}_j-p_j)^2\right] + E\left[\sum_{j\neq k}^{A} w_{jk}(\hat{p}_j-p_j)(\hat{p}_k-p_k)\right]. \quad (4)$$

The loss function presented in Eq. 4 is identical to Kimura's Vartot provided that **W** is an identity matrix, i.e., $w_{jj}=1$ and $w_{jk}=0$ for $j\neq k$.

Substituting Eq. 2 and the approximate form of Eq. 3 into Eq. 4 and collecting terms, we obtain:

$$\mathcal{L} = \sum_{i=1}^{L}\left(\frac{a_i-u_i}{n_i}\right) + \frac{b+v-m}{N}, \quad (5)$$

where $a_i = \sum_{j=1}^{A} w_{jj}\hat{l}_i^2\hat{q}_{ij}(1-\hat{q}_{ij})$,

$u_i = \sum_{j\neq 1}^{A} w_{jk}\hat{l}_i^2\hat{q}_{ij}\hat{q}_{ik}$,

$b = \sum_{i=1}^{L}\sum_{j=1}^{A} w_{jj}\hat{l}_i(\hat{q}_{ij}-\hat{p}_{ij})^2$,

$v = \sum_{i=1}^{L}\sum_{j\neq k}^{A} w_{jk}\hat{l}_i\hat{q}_{ij}\hat{q}_{ik}$.

$m = \sum_{j\neq k}^{A} w_{jk}\hat{p}_j\hat{p}_k$,

and where a_i, u_i, b, m, and v are all positive.

A linear cost function is used for the optimal sampling design:

$$C = c_1 N + \sum_{i=1}^{L} c_{2i} n_i, \quad (6)$$

where C is total cost, c_1 is per-unit cost of collecting a random length sample, and c_{2i} is per-unit cost for ageing a fish in the ith length-stratum.

Survey designs generally are based on two constraints: (i) a fixed total cost, i.e., minimize the loss function in Eq. 4 at a fixed cost; or (ii) a desired precision of the estimators, i.e., minimize the total cost at a given level of the loss function. Therefore, the problem for optimal allocation becomes one of determining the optimal set of N^* and n_i^*'s which minimizes L at a given total cost or which minimizes total cost at a desired precision level of L (N^* and n_i^* are the optimal sample sizes of length and age samples, respectively). Kendall & Stuart (1977, Sect. 39.20) and Cochran (1977, Sect. 5.5) show that choosing the optimal set of N^* and n_i^*'s to minimize L for a fixed C or to minimize C for a fixed L are both equivalent to minimizing the product of L and C:

$$\mathcal{L}C = \left[\sum_{i=1}^{L}\frac{a_i-u_i}{n_i} + \frac{b+v-m}{N}\right]\cdot\left[c_1 N + \sum_{i=1}^{L} c_{2i} n_i\right]. \quad (7)$$

Applying the Cauchy-Schwarz inequality to Eq. 7, the product LC is

$$\mathcal{L}C = \left[\sum_{i=1}^{L}\left(\sqrt{\frac{a_i-u_i}{n_i}}\right)^2 + \left(\sqrt{\frac{b+v-m}{N}}\right)^2\right]\left[(\sqrt{c_1 N})^2 + \sum_{i=1}^{L}(\sqrt{c_{2i} n_i})^2\right]$$
$$\geq \left[\sum_{i=1}^{L}\sqrt{c_{2i}(a_i-u_i)} + \sqrt{b+v-m}\right]^2.$$

Kendall & Stuart (1977) showed that the minimum value of the product LC occurs when

$$\frac{\sqrt{c_{21}n_1}}{\sqrt{\frac{a_1-u_1}{n_1}}} = \ldots = \frac{\sqrt{c_{2i}n_i}}{\sqrt{\frac{a_i-u_i}{n_i}}} = \ldots = \frac{\sqrt{c_{2L}n_L}}{\sqrt{\frac{a_L-u_L}{n_L}}} = \frac{\sqrt{c_1 N}}{\sqrt{\frac{b+v-m}{N}}} = \text{constant}>0 \quad (8)$$

Use the terms of the equality between the ith and the (L+1)th terms and rearrange the variables to obtain $r_i^*=n_i^*/N^*$, the optimal subsampling ratio between age subsamples and length samples in the ith length stratum. The solution of r_i^* is:

$$r_i^* = \sqrt{\frac{c_1(a_i - u_i)}{c_{2i}(b+v-m)}}. \quad (9)$$

For a survey design subject to a given total cost C, r_i^* is the optimal subsampling ratio required to reach the minimum loss function value (min. $\mathcal{L}$). Thus, substitute Eq. 9 into Eq. 6 and solve for the optimal set of N^* and n_i^*:

$$N^* = \frac{C}{c_1 + \sum_{i=1}^{L} c_{2i} r_i^*},$$

$$n_i^* = r_i^* N^*,$$

$$\min. \mathcal{L} = \sum_{i=1}^{L}\left(\frac{a_i - u_i}{n_i^*}\right) + \frac{b+v-m}{N^*}. \quad (10)$$

For a survey design subject to a desired precision level of $\mathcal{L}$, r_i^* is the optimal subsampling ratio to reach the minimum total cost (min. C). The optimal set of N^* and n_i^* can be obtained by substituting Eq. 9 into Eq. 5:

$$N^* = \frac{1}{\mathcal{L}}\left[\sum_{i=1}^{L}\left(\frac{a_i - u_i}{r_i^*}\right) + b+v-m\right],$$

$$n_i^* = r_i^* N^*, \quad (11)$$

$$\min. C = c_1 N^* + \sum_{i=1}^{L} c_{2i} n_i^*.$$

Similarly, the above derivation can be extended to the traditional fixed- and proportional-age subsampling schemes. For these age subsampling schemes, the per-unit cost for ageing a fish is not length-specified (i.e., $c_{2i} = c_2$ for all i's). The loss and cost functions in Eq. 5 and 6 are modified according to the definition of the two age subsampling schemes: (1) $n_i = n/L$ for fixed-age subsampling, and (2) $n_i = n\, l_i$ for proportional-age subsampling, where $n = \Sigma n_i$. The loss function for fixed-age subsampling is

$$\mathcal{L} = \frac{\sum_{i=1}^{L} L(a_i - u_i)}{n} + \frac{b+v-m}{N} \quad (12)$$

and that for a proportional-age subsampling is

$$\mathcal{L} = \frac{\sum_{i=1}^{L} L(a_i - u_i)/\hat{l}_i}{n} + \frac{b+v-m}{N}. \quad (13)$$

The cost function for both age subsampling schemes is:

$$C = c_1 N^* + c_2 n \quad (14)$$

Using the Cauchy-Schwarz inequality, the optimal subsampling ratio (r^*) for either minimizing $\mathcal{L}$ at a fixed total cost C or minimizing total cost C at a desired precision level of $\mathcal{L}$ for a fixed-age subsampling is:

$$r^* = \frac{n^*}{N^*} = \sqrt{\frac{c_1 \sum_{i=1}^{L} L(a_i - u_i)}{c_2(b+v-m)}} \quad (15)$$

The optimal set of N^* and n^* and min. $\mathcal{L}$ subject to fixed cost is:

$$N^* = c_1 + c_2 r^*,$$

$$n^* = r^* N^*, \quad (16)$$

$$\min. \mathcal{L} = \frac{\sum_{i=1}^{L} L(a_i - u_i)}{n} + \frac{b+v-m}{N}.$$

and the optimal set of N^* and n^* and min. C subject to a desired precision level of $\mathcal{L}$ is

$$N^* = \frac{1}{\mathcal{L}}\left[\frac{\sum_{i=1}^{L}(a_i - u_i)}{r^*} + b+v-m\right],$$

$$n^* = r^* N^*, \quad (17)$$

$$\min. C = c_1 N^* + c_2 n^*.$$

For proportional-age subsampling, the optimal subsampling ratio (r^*) for either minimizing $\mathcal{L}$ at a given total cost C, or minimizing total cost C at a desired precision level of $\mathcal{L}$, is:

$$r^* = \sqrt{\frac{c_1 \sum_{i=1}^{L} (a_i - u_i)/\hat{l}_i}{c_2(b+v-m)}} \quad (18)$$

The optimal set of N^* and n^* and min. $\mathcal{L}$ subject to a given total cost C is

$$N^* = c_1 + c_2 r^*,$$

$$n^* = r^* N^*, \quad (19)$$

$$\min. \mathcal{L} = \frac{\sum_{i=1}^{L}(a_i - u_i)/\hat{l}_i}{n^*} + \frac{b+v-m}{N^*}.$$

and the optimal set of N^* and n^* and min. C subject to a desired precision level of $\mathcal{L}$ is:

$$n^* = \frac{1}{\mathcal{L}}\left[\frac{\sum_{i=1}^{L}(a_i - u_i)/\hat{l}_i}{r^*} + (b+v-m)\right],$$

$$n^* = r^* N^*, \tag{20}$$

$$\min.\ C = c_1 N^* + c_2 n^* .$$

The solutions given in Eq. 16–20 are similar to that of Lai (1987), provided that the matrix **W** is an identity matrix.

Example

The lemon sole (= English sole *Pleuronectes vetulus*) example of Jinn et al. (1987) is used to illustrate the length-based optimal sampling design. The ALK and per-unit costs are summarized in Table 1. The total cost is C=\$229.15, and the per-unit cost for collecting a length of fish is c_1=\$0.15. For simplicity without loss of generality, consider the special case: $w_{jk}=0$ for $j \neq k$. Three different sets of w_{jj} are used to reflect different aspects of interest:

Table 1

Age-length key and length-frequency distribution of male lemon sole (= English sole, *Pleuronectes vetulus*) collected from Strait of Georgia, British Columbia. Original dataset is from Ricker (1975:68), and per-unit cost for ageing is from Jinn et al. (1987).

Length stratum (cm)	Number of age sub-samples	Age						Number of length samples	Per-unit cost of ageing
		4	5	6	7	8	9		
27	6	5	1					6	1.0
28	9	3	4	2				9	1.2
29	10	4	4	1	1			30	1.4
30	10	1	5	4				51	1.6
31	10		8	2				54	1.8
32	10	1	7	1	1			48	2.0
33	10	1	3	3	2	1		41	2.2
34	10		2	6	1	1		27	2.4
35	10		1	4	3		2	13	2.6
36	6			1	3	2		6	2.8
37	3			1	1	1		3	3.0
38	1					1		1	3.2
Age proportion		0.12	0.48	0.26	0.09	0.04	0.01		
Variance ($\times 10^{-3}$)		1.09	2.98	2.46	0.88	0.33	0.04		

Case 1 {1,1,1,1,1,1}: equal interest in estimating all $\hat{p}_i$'s;

Case 2 {10,30,30,10,1,1}: increase precision of four major age-classes with larger Var($\hat{p}_i$);

Case 3 {1,1,1,1,10,60}: interest in older but rare age-classes.

The results obtained from length-based design are compared with those from fixed- and proportional-age subsampling schemes. The average per-unit cost for ageing a fish (c_2) is calculated as the weighted mean of c_{2i}, which is c_2=\$1.96. The optimal set of {N^*, n_i^*} and min. $\mathcal{L}$ subject to the given total cost of \$229.15 are computed using Eq. 9 and 10 for length-based age subsampling, Eq. 15 and 16 for fixed-age subsampling, and Eq. 18 and 19 for proportional-age subsampling.

Precision improves substantially when length-based age subsampling rather than fixed-age subsampling is used in all three cases (Table 2). In the first two cases, however, precision improves marginally by using length-based age rather than proportional-age subsamplings. When rare and older fish (ages 8 and 9, Case 3) are of interest, precision is substantially improved by using length-based age instead of proportional-age subsampling. This is due to the fact that proportional-age subsampling is not designed to increase age subsamples from length-strata consisting of older age-classes.

The optimal set of {N^*, n_i^*} and min. C subject to a desired precision level of $\mathcal{L}^*$=0.01 are computed from Eq. 9 and 11 for length-based age subsampling, Eq. 15 and 17 for fixed-age subsampling, and Eq. 18 and 20 for proportional-age subsampling. Tables 2 and 3 show similar trends. The cost efficiency of length-based age subsampling is superior to fixed-age subsampling in all cases. However, cost efficiency is only marginally

Table 2

Optimal sample sizes of length and age samples and min. $\mathcal{L}$ ($\mathcal{L}$: loss function) subject to fixed total cost, C=$229.15, for three different age subsampling schemes. Age-length key dataset is listed in Table 1. Per-unit cost of observing a length sample, c_l=$0.15. Case 1 $\{w_i\}$=$\{1,1,1,1,1,1\}$; Case 2 $\{w_i\}$=$\{10,30,30,10,1,1\}$; and Case 3 $\{w_i\}$=$\{1,1,1,1,10,60\}$.

Length (cm)		Case 1	Case 2	Case 3
27		2	2	2
28		4	5	3
29		14	13	10
30		21	23	15
31		15	18	11
32		16	16	12
33		17	16	18
34		9	9	11
35		5	4	13
36		2	1	3
37		1	1	1
38		1	1	1
Length-based sampling design	n^*	107	109	100
	N^*	183	182	204
	min. $\mathcal{L}$	0.0058	0.1312	0.0110
Improvement of precision (%)[1] vs.				
Fixed-age subsampling		36.26	39.15	28.57
Proportional-age subsampling		4.92	4.23	19.12
Fixed-age subsampling	n^*	106	106	104
	N^*	146	141	172
	min. $\mathcal{L}$	0.0091	0.2156	0.0154
Proportional-age subsampling	n^*	103	103	103
	N^*	179	177	183
	min. $\mathcal{L}$	0.0061	0.1370	0.0136

[1] % = percent difference of min. $\mathcal{L}$ between length-based sampling design and fixed or proportional-age subsampling.

Table 3

Optimal sample sizes of length and age samples and minimum total cost (min. C) subject to a desired precision level of loss function, $\mathcal{L}$=0.01 for the three different age subsampling schemes. Age-length key dataset is listed in Table 1. Per-unit cost of observing a length sample, c_l=$0.15. $\{w_i\}$ for each case are the same as that in Table 2.

Length (cm)		Case 1	Case 2	Case 3
27		1	3	2
28		3	6	4
29		8	17	11
30		12	30	17
31		9	23	12
32		9	21	13
33		10	21	20
34		5	12	12
35		3	5	14
36		1	2	3
37		1	1	2
38		1	1	1
Length-based sampling design	n^*	63	142	111
	N^*	106	236	223
	min. C	135.98	300.6	257.25
Cost efficiency (%)[1] vs.				
Fixed-age subsampling		35.14	39.09	27.06
Proportional-age subsampling		1.08	3.68	17.33
Fixed-age subsampling	n^*	97	229	160
	N^*	133	304	265
	min. C	209.66	493.48	352.68
Proportional-age subsampling	n^*	62	141	140
	N^*	108	242	249
	min. C	137.46	312.07	311.16

[1] % = percent difference of min. C between length-based sampling design and fixed- or proportional-age subsampling.

different between length-based age and proportional age subsamplings in Cases 1 and 2. In Case 3, the cost efficiency of length-based age subsampling increases subtantially over proportional-age subsampling.

Discussion

To draw a general conclusion, many different sets of w_{ij}'s and full matrices of **W** also were investigated. The results from these additional analyses were similar to that of Tables 2 and 3. In general, the length-based age subsampling is superior to either fixed- or proportional-age subsampling. However, precision improvement and cost efficiency depend on the weights placed on particular age-classes. Total cost will change in accord with the different weights and desired precision. A higher total cost should be allowed for cases where sampling is designed to improve the precision of highly variable estimates, usually young and old age-classes.

A larger budget will increase precision of the estimates, especially for highly variable age-classes; however, as Lai (1987) showed, there is a point of diminishing returns as the budget increases (Fig. 1). For the examples used in this paper, precision improvement is marginal when total cost (C) increased beyond $40 for Cases 1 and 3, and beyond $120 for Case 2. Kimura (1989) showed that satisfactory results from cohort analysis can be obtained at low sampling levels (i.e., total cost) provided that the representativeness of the samples can be maintained.

It is difficult to compare the methods of Jinn et al. (1987) with those of this paper because the Bayesian approach and classic sampling techniques are derived from different theoretical backgrounds. Nontheless, the results obtained from this paper are similar to that of

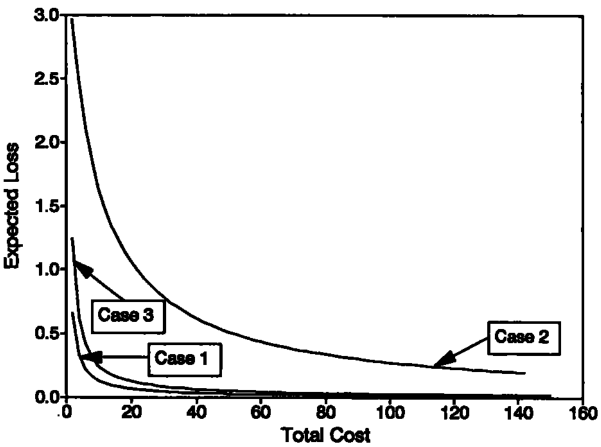

Figure 1
Relationship between minimum loss function value and total cost for the lemon sole (= English sole *Pleuronectes vetulus*) example. The relationship can be obtained from substituting N^* and n_i^* in Eq. 10 into the third equation in Eq. 10.

Jinn et al. (1987) although the values of N^* and n_i^* are different. An advantage of the classic double-sampling technique is the explicit solutions of the optimal set of N^* and n_i^*, which reduces computational effort.

In this paper, the optimal sampling design is for stratified-age subsampling. For random-age subsampling in which the number of age subsamples (n) is randomly taken from the entire length sample of size N, the estimated variance and covariance (Kimura 1977) are

$$Var(\hat{p}_j) = \sum_{i=1}^{L} \left[\frac{\hat{l}_i \hat{q}_{ij} (1-\hat{q}_{ij})}{n} + \frac{\hat{l}_i (\hat{q}_{ij}-\hat{p}_j)^2}{N} \right]$$

and

$$Cov(\hat{p}_j, \hat{p}_k) = \sum_{i=1}^{L} \frac{-\hat{l}_i \hat{q}_{ij} \hat{q}_{ik}}{n} + \sum_{i=1}^{L} \frac{\hat{l}_i \hat{q}_{ij} \hat{q}_{ik}}{N} - \frac{p_j \hat{p}_k}{N}.$$

These two equations are similar to that of proportional-age subsampling in which $n_i = n\hat{l}_i$ is substituted into Eq. 2 and 3. However, the estimated variance and covariance for proportional-age subsampling are approximate, and those for random-age subsampling are not. The similarities of random- and proportional-age subsamplings can be anticipated because N is a random sample from a population so that $E(\hat{l}_i)=E(N_i/N)=l_i$, and n is randomly taken from N so that $E(n_i/n)=N_i/N$. This indicates that the size of each n_i will be approximately proportional to l_i, i.e., $n_i/n \approx N_i/N \approx \hat{l}_i$ and $n_i \approx n \cdot l_i$ (Kutkuhn 1963).

An ALK requires a large random sample of fish from which a length-stratified subsample is collected for ageing. Most fishery data are collected either from surveys in which fish from different tows are sampled or from commercial catches in which fish from different vessel-trips are sampled. Pooling data over such clusters is necessary because of the cost of data gathering. In addition to cluster sampling, fisheries data are frequently stratified into time-area, fisheries (or gears), and sex strata (Kimura 1989). The question is how to make the optimal sampling design of ALK generally applicable. To address this, the following factors must be considered: (1) Need of stratification, (2) ALK sampling within stratum, and (3) combined-strata estimation.

Westrheim & Ricker (1978) showed the need for stratification. An ALK obtained from a population at a time-interval should not be universally applied to length-frequency datasets from other populations or other time-intervals if growth and survival rates are different among the populations and time-intervals. Therefore, the factors that may result in differences in growth and survival rates should be evaluated, and stratification should account for these factors.

Current sampling programs (e.g., Doubleday & Rivard 1983, Quinn et al. 1983, Kimura 1989) adopted the strategy in which length-frequency data collected from clustered sampling units (e.g., tows or vessel-trips) within a stratum are pooled, from which a length-stratified subsample is collected for ageing. Southward (1963) evaluated an old method (Southward 1963:12) in which a set of length and age data is collected from each landing of a vessel-trip. Because this old method is not developed from a probability sampling design, the within-vessel variability in fish lengths is assumed to be less than between-vessel variability. Southward (1963) showed that this assumption is not valid and the estimated variances of age composition from this old method are so large that little confidence can be placed in it.

The length-frequency data pooled over clusters should be a representative sampling of that stratum. Therefore, the weighting factor of each sample should be included in the pooling. Ignoring the weighting factor will bias the estimated age composition (Kimura 1989). Quinn et al. (1983) described a sampling-rate method in which a fixed proportion of halibut were sampled from landings >1000 lbs for length data, and then age data were subsampled from the pooled length samples from these landings. All length samples are self-weighted and can be pooled directly.

Quinn et al. (1983) evaluated the methods of combined-strata estimation and found that the "project-and-add" method (total catch-at-age is estimated for each stratum and then the estimates are added over strata; see Quinn et al. 1983) produces unbiased estimators if all strata are sampled. The project-and-add method uses the concept of a stratified, random sampling technique (Cochran 1977). Therefore, Cochran's rules (Cochran 1977:98) of optimal allocation for stratified random sampling can be applied. In a given stratum, take larger length and age samples if (1) the stratum is larger, (2) the stratum is more variable internally, and (3) sampling is cheaper in the stratum. The first two rules are the basis of the Neyman allocation (Cochran 1977:99). The sampling-rate method proposed by Quinn et al. (1983) built upon the first rule and can easily incorporate other rules in the sampling program.

It is clear that the number of strata and variability (vartot or loss function) of each stratum should be evaluated first for designing ALK sampling. Then, the total cost is allocated into various strata according to Cochran's rules. Once the total cost for each stratum is determined, an optimal sampling design for ALK can be applied. The sampling rate method of Quinn et al. (1983) can be used to collect the optimal length sample size from clusters. After pooling the length samples, a length-stratified subsample is collected for ageing.

Acknowledgments

I thank D.K. Kimura for his help with derivation of the covariance. Thanks to M. Sigler for his constructive comments on an earlier draft. This paper is partially supported by the U.S. Agency for International Development, Fisheries Stock Assessment CRSP (DAN-4146-G-SS-5071-00).

Citations

Cochran, W.G.
1977 Sampling techniques, 3rd ed. John Wiley, NY.

Doubleday, W. G., & D. Rivard (editors)
1983 Sampling commercial catches of marine fish and invertebrates. Can. Spec. Publ. Fish. Aquat. Sci. 66, 290 p.

Jinn, J. H., J. Sedransk, & P. Smith
1987 Optimum two-phase stratified sampling for estimation of the age composition of a fish population. Biometrics 43:343–353.

Kendall, M., & A. Stuart
1977 The advanced theory of statistics, vol. 3, 3rd ed. Griffin & Co., London, 585 p.

Ketchen, K. S.
1949 Stratified subsampling for determining age distributions. Trans. Am. Fish. Soc. 79:205–212.

Kimura, D. K.
1977 Statistical assessment of age-length key. J. Fish. Res. Board Can. 31:317–324.
1989 Variability in estimating catch-in-numbers-at-age and its impact on cohort analysis. *In* Beamish, J. R., & G. A. McFarlane (eds.), Effects of ocean variability on recruitment and an evaluation of parameters used in stock assessment models, p.57–66. Can. Spec. Publ. Fish. Aquat. Sci. 108.

Kutkuhn, J. H.
1963 Estimating absolute age composition of California salmon landings. Calif. Fish Game Fish. Bull. 120, 29 p.

Lai, H. L.
1987 Optimum allocation for estimation age composition using age-length key. Fish. Bull., U.S. 85:179–185.

Quinn, T. J. II, E. A. Best, L. Bijsterveld, & I. R. McGregor
1983 Port sampling for age composition of Pacific halibut landings. *In* Doubleday, W. G., & D. Rivard (eds.), Sampling commercial catches of marine fish and invertebrates, p. 194–205. Can. Spec. Publ. Fish. Aquat. Sci. 66.

Rao, J. N. K., & P. D. Ghangurde
1972 Bayesian optimization in sampling finite populations. J. Am. Stat. Assoc. 67:439–443.

Ricker, W. E.
1975 Computation and interpretation of biological statistics of fish populations. Fish. Res. Board Can. Bull. 191, 382 p.

Southward, G. M.
1963 Sampling landings of halibut for age composition. Int. Pac. Halibut Comm. Sci. Rep. 58, 31 p.

Tanaka, S.
1953 Precision of age-composition of fish estimated by double sampling method using the length for stratification. Bull. Jpn. Soc. Sci. Fish. 19:657–670.

Westrheim, S. J., & W. E. Ricker
1978 Bias in using an age-length key to estimate age-frequency distributions. J. Fish. Res. Board Can. 35:184–189.