OPTIMUM ALLOCATION FOR ESTIMATING AGE COMPOSITION USING AGE-LENGTH KEY

HAN-LIN LAI[1]

ABSTRACT

A new optimum allocation method for age-length keys (ALK) was developed by applying Kimura's Vartot, an error index of estimated age composition. The method is applied to Pacific cod, sablefish, and walleye pollock. At the present working capacity, the total of 10,000 minutes (about 70 working days) will approximate the most effective cost to estimate age composition for the three species. Increasing costs beyond this level will show no more gain.

Age-length keys (ALK) are widely used for estimating age compositions in fisheries. The theory of ALK is based on a double sampling technique with stratification (Tanaka 1953). The first stage involves a simple random sampling for a relatively large size, less costly length sample. The second stage involves a stratified random sampling for a smaller size, more costly age subsample from each length stratum. Following the approximation of Kutkuhn (1963) and Southward (1963), the proportion of fish at the ith age class (p_i) and variance of p_i are estimated as

$$p_i = \sum_{j=1}^{L} l_j q_{ij} \tag{1}$$

$$\mathrm{Var}(p_i) = \sum_{j=1}^{L} \left[\frac{l_j^2 q_{ij} (1 - q_{ij})}{n_j} + \frac{l_j (q_{ij} - p_i)^2}{N} \right] \tag{2}$$

where l_j is the proportion of fish that fall into the jth length stratum,
N is total length sample size,
n_j is the size of age subsample in the jth length stratum,
q_{ij} is the proportion of n_j fish classified into the ith age class,
A is the number of age classes, and
L is the number of length strata.

Kimura (1977) defined Vartot as the sum of all variances of the p_i:

$$\mathrm{Vartot} = \sum_{i=1}^{A} \mathrm{Var}(p_i) = E\left[\sum_{i=1}^{A} (\hat{p}_i - p_i)^2 \right] \tag{3}$$

which is an error index for assessing precision of the ALK. Furthermore, Vartot is the expectation of the squared distance between the estimated age composition $\hat{P}' = (\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_A)$ and the true age composition of the population $P' = (p_1, p_2, \ldots, p_A)$. Then, $D = \sqrt{\mathrm{Vartot}}$ can be interpreted as a kind of average Euclidean distance between $\hat{P}$ and P in an A-dimensional space. Kimura (1983) indicated that D can be viewed as the percent error of the estimated accumulated age proportion (i.e., the percent error of $\Sigma p_i = 1$).

This paper derives a new method for the optimum allocation of ALK, applying the properties of Vartot and Cauchy-Schwarz inequality (Kendall and Stuart 1977). The optimum sizes of length sample and age subsample are determined so that either Vartot is minimized subject to a fixed total cost or the total cost is minimized subject to a desired level of Vartot. Although this method is basically derived for the problem that all age classes are of equal interest, it can be modified by adding weighting factors to the ages which are important to population dynamics. This method was applied to Pacific cod, *Gadus macrocephalus*, from the Washington coast; sablefish, *Anoplopoma fimbria*, from the Gulf of Alaska; and walleye pollock, *Theragra chalcogramma*, from the eastern Bering Sea.

METHODS

Two subsampling schemes related to ALK are frequently used by fisheries biologists: 1) fixed age sub-

[1]School of Fisheries, WH-10, University of Washington, Seattle, WA; present address: Center for Quantitative Science, HR-20, University of Washington, Seattle, WA 98195.

Manuscript accepted February 1987.
FISHERY BULLETIN: VOL. 85, NO. 2, 1987.

sampling, in which the size of age subsample in all length strata is constant (i.e., $n_j = n/L$, where $n = \Sigma n_j$ is total age subsample size), and 2) random age subsampling, in which the size of the age subsample in each length stratum is proportional to the length sample size for all length strata (i.e., $n_j = nl_j$). Thus applying Equations (2) and (3), Vartot for a fixed age subsample (Appendix A) is

$$\text{Vartot} = \frac{a_1}{n} + \frac{a_2}{N} \tag{4}$$

and Vartot of a random age subsample is

$$\text{Vartot} = \frac{b_1}{n} + \frac{b_2}{N} \tag{5}$$

$$\text{where } a_1 = \sum_{i=1}^{A} \sum_{j=1}^{L} [L\, l_j^2\, q_{ij}\,(1 - q_{ij})]$$

$$a_2 = b_2 = \sum_{i=1}^{A} \sum_{j=1}^{L} [l_j\,(q_{ij} - p_i)^2]$$

$$b_1 = \sum_{i=1}^{A} \sum_{j=1}^{L} [l_j\, q_{ij}\,(1 - q_{ij})].$$

It should be noted that a_1, a_2, b_1, and b_2 are all positive values.

The total cost of a survey can be expressed as some function of the costs of the two sampling stages. The commonly used form of the cost function is a simple linear equation relating the unit cost of observing the length and the age of a fish to N and n (Tanaka 1953; Kutkuhn 1963; Southward 1963):

$$C = c_1 N + c_2 n \tag{6}$$

where C is the total cost, c_1 is the unit cost of observing the length of a fish, and c_2 is the unit cost of determining the age of a fish. For optimum allocation, the problem becomes one of determining the values of N and n, which will provide an estimated age composition with a minimum Vartot subject to a given total cost C.

Cauchy-Schwarz inequality, which is frequently used in sampling theory (Cochran 1977; Kendall and Stuart 1977; Schweigert and Sibert 1983), is applied to find optimum set of N^* and n^* for an ALK. The set of N^* and n^* is obtained when they minimize the product of D^2 (= Vartot) and C.

For a fixed age subsample, using Equations (4) and (6) and the Cauchy-Schwarz inequality (Appendix B), we obtain

$$r^* = (n/N)^* = \sqrt{a_1 c_1 / a_2 c_2}. \tag{7}$$

This quantity is the optimum subsampling ratio (r^*) required to reach the minimum (min.) Vartot subject to the cost function given in Equation (6). Therefore, N^* and n^* are dependent on Equation (6):

$$N^* = C/(c_1 + c_2 r^*) \tag{8}$$

$$n^* = r^* N^* \tag{9}$$

$$\text{min. Vartot} = \frac{a_1}{n^*} + \frac{a_2}{N^*}. \tag{10}$$

Similarly, the optimum allocation of a random age subsample can be obtained using Equations (5) and (6) and the Cauchy-Schwarz inequality. The solutions of r^*, N^*, n^*, and min. Vartot are

$$r^* = (n/N)^* = \sqrt{b_1 c_1 / b_2 c_2} \tag{11}$$

$$N^* = C/(c_1 + c_2 r^*) \tag{12}$$

$$n^* = r^* N^* \tag{13}$$

$$\text{min. Vartot} = \frac{b_1}{n^*} + \frac{b_2}{N^*}. \tag{14}$$

Generally, survey designs are based on two constraints (Schweigert and Sibert 1983). The first, as derived previously, relates to obtaining the precision of p_i, viz., to minimize Vartot at a fixed total cost. The second determines the total cost required to achieve a given precision, that is, to minimize total cost (min. C) at a desired level of Vartot. In the latter problem, r^* in Equations (7) and (11) will also minimize the product of D^2C irrespective of the value of C and D^2, i.e., it will minimize D^2 for fixed C or C for fixed D^2 (Kendall and Stuart 1977, Section 39.20). The solutions of n^* and N^* are now dependent on the desired level of Vartot.

For a fixed age subsample, substituting r^* of Equation (7) into Equation (4), we obtain

$$N^* = (a_1/r^* + a_2)/D^2 \tag{15}$$

$$n^* = r^* N^* \tag{16}$$

$$\text{min. } C = c_1 N^* + c_2 n^* \tag{17}$$

For a random age subsample, N^* is obtained by substituting Equation (11) into Equation (5)

$$N^* = (b_1/r^* + b_2)/D^2 \tag{18}$$

$$n^* = r^*N^* \tag{19}$$

$$\min.\ C = c_1N^* + c_2n^*. \tag{20}$$

Because $0 < r \leqslant 1$, Equations (7) and (11) indicate that

$c_1/c_2 \leqslant a_2/a_1$ for fixed age subsample and

$c_1/c_2 < b_2/b_1$ for random age subsample

must be held for the optimum allocation. When equality holds, $r = 1$ and then ALK degenerates to a simple random sampling for age samples.

EXAMPLES

Three sets of ALK data are used for the example: Pacific cod, aged by the scale method, from the Washington coast (Kimura 1977); sablefish, aged by the otolith method, from the Gulf of Alaska; and walleye pollock, aged by the otolith method, from the eastern Bering Sea (Lai 1985). The parameters of Vartot, a's and b's, are calculated and summarized in Table 1. Accurate cost estimates are difficult to determine; therefore, time measurements required for observing a length and determining an age of fish are used for c_1 and c_2. The total cost C is thus the total time required to build an ALK. The measurements of c_1 and c_2 for the three species (Table 1) are primarily based on the author's experience.

Given a total of 120 working days and 6 working hours per day to a fisheries scientist, the optimum allocation is summarized in Table 2. Under this budget, a random age subsample can provide higher precision than a fixed age subsample (improved 10%, 15%, and 18% respectively for Pacific cod, sablefish, and walleye pollock). Also, for this budget, the error of estimated cumulated age proportion is less than 2.5% for the three species using either fixed or random age subsamples.

Using Equations (7), (15), (16), and (17) for a fixed age subsample and Equations (11), (18), (19), and (20) for a random age subsample, costs under various desired D are minimized for the three species (Table 3). At the same level of D, a fixed age subsample requires much larger sample sizes of length and age than a random age subsample does for the three species. The greatest benefit of using random age subsample is that it drastically reduces the total cost required to obtain the same level of D. The total cost is reduced by 35%, 45%, and 55% respectively for Pacific cod, sablefish, and walleye pollock for any given D when a random instead of a fixed age subsample is used.

Figure 1 shows the relationships of D and total cost. Whether a fixed or a random age subsample is used for the three species, it is obvious that D decreases rapidly until C = 10,000 minutes, which is nearly 70 working days. A point of diminishing

TABLE 1.—Parameters of Vartot and costs for optimum allocation.

	Pacific cod	Sablefish	Walleye pollock
a_1	0.4495	1.2222	1.5501
a_2	0.2660	0.1717	0.1217
b_1	0.2766	0.6235	0.6634
b_2	0.2660	0.1717	0.1217
[1] C	43,200	43,200	43,200
[1] c_1	0.5	0.5	0.5
[1] c_2	15	15	15

[1] In minutes, assuming 120 working days for each species and 6 working hours per day.

TABLE 2.—Optimum allocation of minimizing Vartot for Pacific cod, sablefish and walleye pollock. ($D = \sqrt{\text{Vartot}}$.)

	Pacific cod		Sablefish		Walleye pollock	
	fixed	random	fixed	random	fixed	random
N^*	10,641	13,121	5,534	7,555	5,093	7,551
n^*	2,525	2,443	2,695	2,629	4,064	3,943
D^*	0.014	0.013	0.022	0.019	0.020	0.017

TABLE 3.—Optimum allocation minimizing total cost for various desired precision level ($D = \sqrt{\text{Vartot}}$). Parameters and c_1 and c_2 are listed in Table 1.

		Pacific cod		Sablefish		Walleye pollock	
D		fixed	random	fixed	random	fixed	random
0.05	N^*	865	701	1,073	786	826	557
	n^*	206	130	523	273	659	291
	C^*	3,508	2,307	8,372	4,492	7,002	3,186
0.04	N^*	1,351	1,095	1,676	1,228	1,290	870
	n^*	321	204	817	427	1,030	454
	C^*	5,482	3,605	13,080	7,019	11,940	4,978
0.03	N^*	2,401	1,947	2,979	2,182	2,310	1,547
	n^*	570	362	1,451	759	1,831	808
	C^*	9,745	6,410	23,254	12,478	19,450	8,850
0.02	N^*	5,401	4,380	6,702	4,910	5,160	3,480
	n^*	1,282	815	3,265	1,708	4,118	1,817
	C^*	21,924	14,422	52,320	28,076	43,761	19,912
0.01	N^*	21,604	17,520	26,809	19,640	20,640	13,922
	n^*	5,121	3,262	13,059	6,832	16,473	7,269
	C^*	87,699	57,687	209,279	112,305	175,043	79,650

C^*: in minutes.

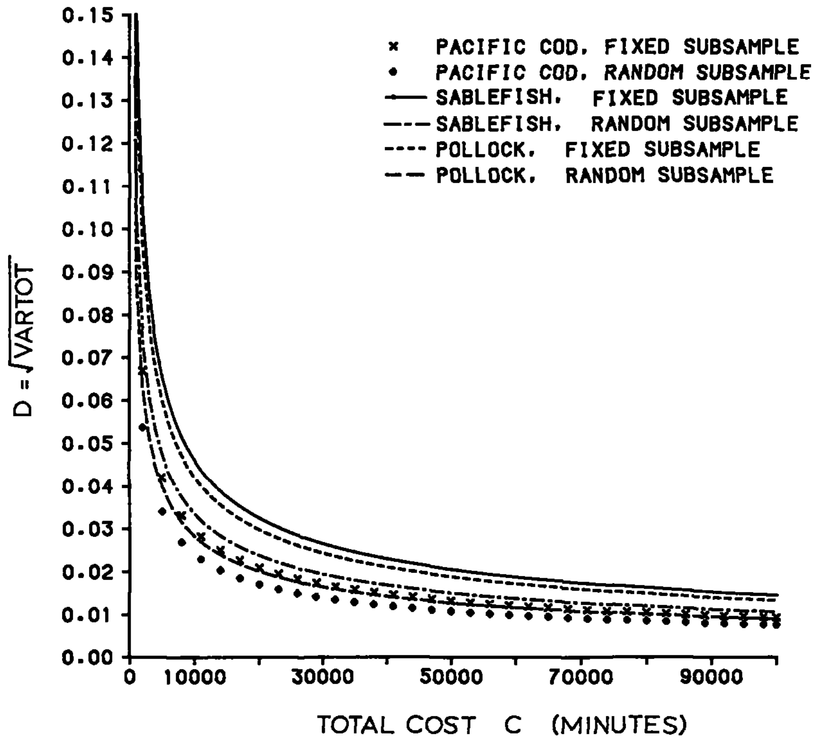

FIGURE 1.—The relationship of D (= $\sqrt{\text{Vartot}}$) and total cost for Pacific cod, sablefish, and walleye pollock, using fixed or random age subsample.

returns is reached beyond this total cost and the curves become flatter for C greater than 10,000. These results indicate that setting a precision at D = 0.02, 0.025, and 0.03 respectively for Pacific cod, walleye pollock, and sablefish using random age subsamples would represent a reasonable compromise between cost and precision. Increasing total cost beyond this level will show no more gains from the ALK.

DISCUSSION

It is obvious that the random subsampling scheme is superior to the fixed subsampling scheme. However, it is more important to realize that there is a cap on total cost for ALK. This cap represents the most effective budget for ALK. Vartot of estimated age composition will not decrease significantly for a greater budget. For the three species, total cost of 10,000 minutes (about 70 working days) is the upper limit. This indicates that approximately 2,000 length observations and 800 random age subsamples for sablefish, 2,500 and 1,200 for walleye pollock, and 3,000 and 600 for Pacific cod represent the best compromise between cost and precision of estimates ($\sqrt{\text{Vartot}}$ = 0.03, 0.025, and 0.02 for the three species respectively).

Although it can be argued that minimizing Vartot may not be sufficient for optimum sampling design for all age classes, it is necessary to consider that some age classes are rare in the commercial catch and are therefore difficult to sample precisely. However, these age classes do not generally represent significant contributions to biomass, and it therefore seems reasonable to concentrate on the major age classes. If these rare age classes are important to population dynamics, the optimum allocation can be addressed as a multiple minima. The objective function can be rewritten as

$$M(N,n) = \sum w_i \,\text{Var}(p_i), \qquad \text{for } i = 1, 2, \ldots, A$$

where w_i is weighting factor. A larger w_i must be given to those age classes which are of interest, whereas the mathematical expressions of optimum allocation are the same as Equations (7) to (14), and subject to the same cost function (Equation (6)), except that a's and b's are weighted by w_i. In fact, minimizing Vartot is a special case of minimizing $M(N,n)$ = Vartot for all $w_i = 1$.

Another argument may relate to the possibility that the cost function may be more complicated so that traveling and overhead costs can be taken into account. In such cases, cost function may become a nonlinear form, and the explicit expressions of n^* and N^* cannot be obtained. However, the technique of nonlinear programming can be applied to find numerical solutions of n^* and N^*. In general, it is to

$$\text{minimize } M(N,n) = \sum w_i \operatorname{Var}(p_i)$$

$$\text{subject to } C = c(N,n)$$

where C is total cost and $c(N,n)$ is cost function. Many optimization programs can be employed from popular computer software packages for main frame computers. Bunday (1984) provided several BASIC programs for constrained optimization, which may be useful in personal computers. It should be noted, however, that the sufficient and necessary conditions of this constrained minimum must be proved. Theoretically, there is a unique minimum if objective function is convex and constraint function is concave (Bunday 1984).

ACKNOWLEDGMENTS

This study was funded by the Northwest and Alaska Fisheries Center, NMFS, Seattle, WA. I am deeply indebted to Donald R. Gunderson, Vincent F. Gallucci, and Loh-Lee Low for their guidance and Pat Sullivan for his comments. Many thanks to the referees for their helpful comments, and particularly for rewording a more accurate description of Vartot.

LITERATURE CITED

BUNDAY, B. D.
1984. Basic optimisation methods. Edwards Arnold, Lond., 128 p.

COCHRAN, W. G.
1977. Sampling techniques. 3d ed. John Wiley & Sons, Inc., N.Y., 428 p.

KENDALL, M., AND A. STUART.
1977. The advanced theory of statistics. Vol. 3. 3d ed. Griffin & Co., Lond., 585 p.

KIMURA, D. K.
1977. Statistical assessment of age-length key. J. Fish. Res. Board Can. 31:317-324.
1983. Determing appropriate sampling levels for estimating age distributions for trawl landings. Wash. Dep. Fish., Tech. Rep. 80, 18 p.

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

LAI, H. L.
1985. Evaluation and validation of age determination of sablefish, pollock, Pacific cod and yellowfin sole; optimum sampling design using age-length key; and implications of aging variability in pollock. Ph.D. Thesis, Univ. Washington, Seattle, 426 p.

SCHWEIGERT, J. F., AND J. R. SIBERT.
1983. Optimizing survey design for determining age structure of fish stocks: an example from British Columbia Pacific herring (*Clupea harengus pallasi*). Can. J. Fish. Aquat. Sci. 40:588-597.

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.

APPENDIX A. Derivation of Vartot

For a fixed age subsampling, substituting $n_j = n/L$ into Equation (2), the $\mathrm{Var}(p_i)$ is

$$\mathrm{Var}(p_i) = \sum_{j=1}^{L} \left[\frac{L\, l_j^2\, q_{ij}\,(1 - q_{ij})}{n} + \frac{l_j\,(q_{ij} - p_i)^2}{N} \right]. \tag{A.1}$$

Applying Equation (3),

$$\mathrm{Vartot} = \sum_{i=1}^{A} \mathrm{Var}(p_i)$$

$$= \sum_{i=1}^{A} \sum_{j=1}^{L} \left[\frac{L\, l_j^2\, q_{ij}\,(1 - q_{ij})}{n} + \frac{l_j\,(q_{ij} - p_i)^2}{N} \right]$$

$$= \frac{\sum_{i=1}^{A} \sum_{j=1}^{L} [L\, l_j^2\, q_{ij}\,(1 - q_{ij})]}{n} + \frac{\sum_{i=1}^{A} \sum_{j=1}^{L} [l_j\,(q_{ij} - p_i)^2]}{N}$$

$$= \frac{a_1}{n} + \frac{a_2}{N}$$

which is Equation (4).

For a random age subsampling, substituting $n_j = n\, l_j$ into Equation (2) and applying Equation (3), the Vartot is

$$\mathrm{Vartot} = \sum_{i=1}^{A} \sum_{j=1}^{L} \left[\frac{l_j\, q_{ij}\,(1 - q_{ij})}{n} + \frac{l_j\,(q_{ij} - p_i)^2}{N} \right]$$

$$= \frac{\sum_{i=1}^{A} \sum_{j=1}^{L} [l_j\, q_{ij}\,(1 - q_{ij})]}{n} + \frac{\sum_{i=1}^{A} \sum_{j=1}^{L} [l_j\,(q_{ij} - p_i)^2]}{N}$$

$$= \frac{b_1}{n} + \frac{b_2}{N}$$

which is Equation (5).

APPENDIX B. Derivation of Equation (7)

The Cauchy-Schwarz inequality (Cochran 1977, p. 97) is

$$(\sum_h A_h^2)(\sum_h B_h^2) - (\sum_h A_h B_h)^2 = \sum_i \sum_{i>j} (A_i B_j - A_j B_i)^2 \geq 0. \quad \text{(B.1)}$$

Therefore,

$$(\sum_h A_h^2)(\sum_h B_h^2) \geq (\sum_h A_h B_h)^2.$$

For a fixed age subsample, let

$$A_1 = \sqrt{a_1/n};\ A_2 = \sqrt{a_2/N};\ B_1 = \sqrt{c_2 n};\ \text{and}\ B_2 = \sqrt{c_1 N}.$$

Applying Equation (B.1), the product of D^2 and C is

$$D^2 C = \left(\frac{a_1}{n} + \frac{a_2}{N}\right)(c_2 n + c_1 N) \geq \left(\sqrt{a_1 c_2} + \sqrt{a_2 c_1}\right)^2. \quad \text{(B.2)}$$

The product D^2C will be minimized, provided that the equality of Equation (B.2) holds. Setting equality of Equation (B.2) and expanding both sides, we find the solution is

$$r^* = (n/N)^* = \sqrt{a_1 c_1 / a_2 c_2} \quad \text{(B.3)}$$

which is Equation (7).

For a random age subsample, let

$$A_1 = \sqrt{b_1/n};\ A_2 = \sqrt{b_2/N};\ B_1 = \sqrt{c_2 n};\ \text{and}\ B_2 = \sqrt{c_1 N}.$$

The reader can derive Equation (11) by the procedures identical to Equations (B.2) and (B.3).