

Analysis of a Mixed Normal Distribution by Means of the First Difference of the Logarithmic Frequency Distribution

Yoshiharu MATSUMIYA*¹, Masao OKA*², Kazuhiko HIRAMATSU*³
and Kenji ASANO*¹

*1 Faculty of Bioresources, Mie University

*2 Faculty of Fisheries, Nagasaki University (the deceased)

*3 National Research Institute of Far Seas Fisheries

Abstract

A simple method which utilizes the fact that the differential coefficients of a parabola give a straight line is proposed for analysing the polymodal length distribution into two or more individual normal distributions. A differential value of $\ln Y_i$ ($= y_i$, logarithmic frequency of length classes l_i), $\frac{dy_i}{dl_i}$ can be approximately given by

$$\frac{dy_i}{dl_i} = \frac{1}{2} \left(\frac{y_i - y_{i-1}}{l_i - l_{i-1}} + \frac{y_{i+1} - y_i}{l_{i+1} - l_i} \right)$$

where $\ln Y = \ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(l - m)^2}{2\sigma^2}$ (for a normal distribution the mean is m and standard deviation σ).

When several straight lines sloping downward to the right can be fitted to the first difference of the logarithmic frequency distribution, the intersections of these straight lines and $dy/dl = 0$ represent mean length of several normal distribution. The method above is applied to the saury and the porgy length distribution.

Key words: mixed normal distribution, logarithmic polymodal frequency, application to saury and porgy

Attempts have been made since 1920's to use a method in which a polymodal length distribution is assumed to have been derived from a mixed normal distribution, to analyse into two or more individual normal distributions (Buchanan-Wollaston and Hodgson¹), Cassie²), Tanaka³). This method is available when age determination is difficult, and is a fundamental one for estimating the age of an organism having no distinct age characters at all, from its length distribution data.

The use of a method of fitting a parabola to the logarithms of a frequency distribution has been advocated for a long time. In the present work, we propose a simple method which utilizes the fact that the differential coefficients of a parabola give a straight line.

Method

Assuming that length l has a normal distribution in each of the individual groups of a frequency distribution; and that for a given group the mean is m , standard deviation σ , and the total number of individuals N ; then the length distribution expected in that group becomes

$$f(l) = \frac{N}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(l-m)^2}{2\sigma^2} \right\}. \quad (1)$$

When the frequency values of length classes $l_1, l_2, l_3, \dots, l_i$ are given in terms of $Y_1, Y_2, Y_3, \dots, Y_i$, respectively, we get

$$\ln Y = \ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(l-m)^2}{2\sigma^2}. \quad (2)$$

If $y = \ln Y$, we get the following expression which describes a parabola

$$y = -pl^2 + ql - r. \quad (3)$$

Here, if we differentiate both sides of the above expression by l , then we get

$$dy/dl = -2pl + q, \quad (4)$$

and if $dy/dl = R$, R becomes a linear function with respect to l .

If a differential value of the logarithmic frequency of l_i , i.e. $\ln Y_i = y_i$, is expressed terms of R_i , the values of R_i can be approximately given by

$$R_i = \frac{1}{2} \left(\frac{y_i - y_{i-1}}{l_i - l_{i-1}} + \frac{y_{i+1} - y_i}{l_{i+1} - l_i} \right). \quad (5)$$

If intervals of length classes are taken as l unit, R_i can be obtained by

$$R_i = \frac{1}{2} (y_{i+1} - y_{i-1}). \quad (6)$$

As described above, if three or more values of R_i obtained for each of the individual length classes can be plotted on an identical straight line, then these length frequencies can be assumed to be points on an identical normal distribution curve. When several straight lines sloping downward to the right can be fitted to the first difference of the logarithmic frequency distribution, then the intersections of these straight lines and $R = 0$ represent mean length of the groups, and m and σ are given by

$$m = q/2p, \quad \sigma = \sqrt{1/2p}. \quad (7)$$

Application

Figure 1 shows an example in which our method has been applied to a body length distribution of the saury (*Cololabis saira*) caught with lift net in the coastal waters of Japan. The distribution could be clearly separated into three groups having means and standard deviations in the order of ascending magnitude: 21.8, 26.1, 30.4 for mean; and 1.26, 1.73, 1.58 for standard deviation (unit; cm).

Another applied example is a fork length distribution of the porgy (*Taius tumifrons*) caught with trawl net in the East China Sea (Tanaka³⁾, Table 2). As shown in Table 1, the distribution could be analysed into five groups, or to be exact, year-classes. Table 1 also summarizes the results obtained by methods other than the

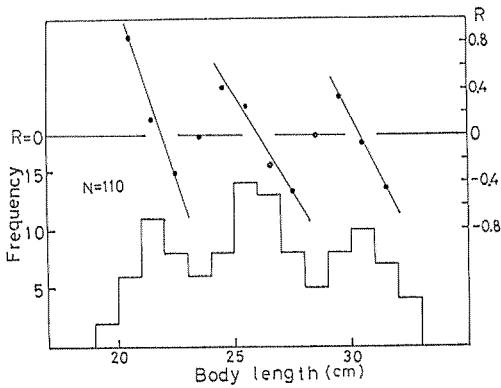


Fig. 1. Length distribution of the saury, and differential calculus of logarithmic frequency.

Table 1. Comparison of the results obtained by five methods A: the authors' method. B: Buchanan-Wollaston's method. C: Casie's method. D: Tanaka's method. E: Akamine's method.

Item	Method	Group (Age)				
		I	II	III	IV	V
Mean (cm)	A	11.0	15.4	19.9	23.2	26.1
	B	11.05	15.32	19.85	23.58	26.82
	C	11.02	15.33	19.85	23.46	26.92
	D	10.99	15.26	19.84	23.50	26.82
	E	11.00	15.29	19.71	23.45	27.26
Standard deviation (cm)	A	0.95	1.26	1.64	1.89	1.80
	B	0.844	1.161	1.412	1.212	1.443
	C	0.76	1.15	1.32	1.29	1.54
	D	0.8	1.2	1.4	1.2	1.4
	E	0.873	1.138	1.416	1.637	1.162

authors's. Some of the data in the Table have been quoted from Tanaka³⁾. It can be said from the Table that a satisfactory agreement has been obtained between the results of the authors's and the other methods.

Discussion

Our method has some disadvantages: the use of differences involves a danger of magnifying the errors in the frequency distribution. Moreover, it would be difficult to expect full precision when there are few, only two in particular, points available for determining the position of the straight lines. It is desirable to draw the straight lines freehand giving special attention to the points in the central part.

Bhattacharya⁴⁾ proposes a modification of our technique and provides several techniques for calculating the proportions of individuals (fish) present in each of the Gaussian components.

Recently, methods such as non-linear optimization, have been proposed and put to practical applications, in which a mixed normal distribution is analysed into its individual groups by means of computers (Hasselblad⁵⁾, Akamine⁶⁾, Kimura and Chikuni⁷⁾). The method in this paper is a graphical one which may nowadays belong to the category of classical techniques, but it proved to be less troublesome and, moreover, to give rather satisfactory results. Also, it can readily be put into practical use by means of a personal computer graphics technique.

Acknowledgements

We wish to express our profound gratitude to Dr. David Gilbert, Fisheries Research Centre, MAF New Zealand, for reviewing the manuscript and giving us the important advice.

References

- 1) BUCHANAN-WOLLASTON, H. J. and W. C. HODGSON. A new method of treating frequency curves in fishery statistics, with some results. *J. Cons. Explor. Mer*, 4: 207– 225 (1929).
- 2) CASSIE, R. M. Some uses of probability paper in the analysis of size frequency distributions. *Aust. J. Mar. Freshw. Res.*, 5: 513– 522 (1954).
- 3) TANAKA, S. A method of analysing a polymodal frequency distribution and its application to the length distribution of the porgy, *Taius tumifrons* (T. and S.). *J. Fish. Res. Board Canada*, 19: 1143– 1159 (1962).
- 4) BHATTACHARYA, C. G. A simple method of resolution of a distribution into Gaussian components. *Biometrics*, 23: 115– 135 (1967).
- 5) HASSELBLAD, V. Estimation of parameters for a mixture of normal distributions. *Technometrics*, 8: 431– 444 (1966).
- 6) AKAMINE, T. Comparison of algorithms of several methods for estimating parameters of a mixture of normal distributions. *Bull. Jap. Sea Reg. Fish. Res. Lab.*, 37: 259– 277 (1987).
- 7) KIMURA, D. K. and S. CHIKUNI. Mixtures of empirical distributions: An interative application of the age-length key. *Biometrics*, 43: 23– 35 (1987).

対数頻度の 1 次階差による複合正規分布の分解

松宮義晴・岡 正雄・平松一彦・浅野謙治

放物線の微係数が直線になることを利用して, polymodal な体長組成データを複数の正規分布に分解する簡便な方法を提案する。

体長階級 l_i の対数頻度 y_i の微分値は近似的に

$$\frac{dy_i}{dl_i} = \frac{1}{2} \left(\frac{y_i - y_{i-1}}{l_i - l_{i-1}} + \frac{y_{i+1} - y_i}{l_{i+1} - l_i} \right)$$

として表わされる。ここで $y = \ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(l-m)^2}{2\sigma^2}$ (平均 m , 標準偏差 σ の正規分布) とする。対数頻度分布の 1 次階差に幾本かの右さがりの直線をあてはめられた場合, これらの直線と $dy/dl = 0$ との交点, それぞれの正規分布の平均体長を表現することになる。

上記の方法をサンマとキダイの体長組成の具体例へ適用した。