日本海ブロック試験研究集録
水産庁　日本海区水産研究所
第 11 号, 23 ~ 30 , 1987

# 体長組成の分解と成長曲線に関する歴史的考察

赤　嶺　達　郎
(日本海区水産研究所)

体長組成データを多重正規分布とみなして，個々の分布に分解する試みは1920年代より行われてきた。1955年以前については田中（1956）に，それ以降については松崎ほか（1983）に概略が述べられている。また成長曲線がバータランフィの式にＳ字カーブをいくつか加えた階段状の形になることは古くから知られていたが，この拡張式について本格的に検討がなされたのは1970年代になってからである。ここではこの２つのテーマについての最近20年間の主要な論文について論評を加える。ただし，筆者の英語力と根性の不足により十分に理解できてない部分があり，全く誤まった評価を下した危険性がある。疑問を持たれた方は原文を直接読まれることをお勧めする。

## 1．多重正規分布のパラメータ推定

(1)　Hasselblad (1966)

このテーマについて最重要の論文である。目的関数は最尤法を採用し，反復法を開発し（彼はこれを一般化最急降下法と呼んだが，最急降下法ではなく狭義の反復法である），実用的なプログラムNORM-SEP (Normal Separator Program) を供給した。またニュートン法（初期値は反復法より与えた）と比較し，所用時間は大差ないがニュートン法の方が少し精度が良いことも確認している。特殊例として$\sigma$が$\mu$に比例する場合および$\sigma$が一定の場合についても反復式を導いている。さらに正規分布が２つの場合について実験を行い，共分散行列（ヘシアン行列の逆行列）の対角成分の大きさを調べて，$\mu$の差が$\sigma$の２倍以上でないと精度が悪くなることも明らかにした。

つまりこのテーマのほとんどの部分を解決している。彼の手法は従来の手法とは全く異なる画期的なもので実用面からみても全く申し分ない。Cohen（1966）のコメントはこのテーマ一般についての注意事項のようなものである。彼の成功は大型計算機の導入によるところが大きい。Marquardt が1963年に既にマルカール法を発表しているが，この時期にこれだけのことをやったのは驚嘆すべきことである。

日本でこの論文が十分に評価されてきたとは言い難い（嶋津1980，松崎ほか1983，田中1985等ではかなり評価されているが)。それは反復法の原理が理解しにくかったためと思われる（反復法の原理についてはたとえば伊理1981を参照のこと。また一松1982によれば反復 iteration はなんらかの意味で収束する列を得るときに限って使い，単なる繰り返し recursion とは区別した方がよいとのことである)。彼の反復法では$\mu$と$\sigma$については正規分布（$n=1$）の自然な拡張となっているが，$P$についてはかなり巧妙である。以下に赤嶺（1987）の記号に従って検討を加えてみる。

まず$\sum P_i=1$より

$$P_n=1-\sum_{i=1}^{n-1}P_i$$

として$P_n$をのぞく。$F$：ヒストグラム，$g=\sum P_i N_i$，$N$：正規分布，$Y=-\ln L$，$L$：尤度関数として

$$\frac{\partial Y}{\partial P_i}=0 \text{ より } \sum_x \frac{F}{g}N_i=\sum_x \frac{F}{g}N_n$$

を得る。このままでは$P_i$が消えてしまっているので両辺に$P_i$をかける。

$$P_i\sum_x \frac{F}{g}N_i=P_i\sum_x \frac{F}{g}N_n \qquad \text{(A)}$$

これを変形して

$$P_i^{\text{new}}=\left[P_i\sum_x \frac{F}{g}N_i\Big/\sum_x \frac{F}{g}N_n\right]^{\text{old}} \qquad \text{(B)}$$

などとしても収束しない。実は(A)は$i=1\sim n-1$における連立式となっている。

$$P_1\sum_x \frac{F}{g}N_1=P_1\sum_x \frac{F}{g}N_n$$

$$\wr \qquad\qquad \wr$$

$$P_{n-1}\sum_x \frac{F}{g}N_{n-1}=P_{n-1}\sum_x \frac{F}{g}N_n$$

$i=n$の場合は恒等式である。$i=1\sim n$について右辺と左辺を別々にすべて加えると，

$$\sum_{i=1}^{n}\text{左辺}=\sum_x \frac{F}{g}g=\sum_x F=T \text{（総データ数）}$$

$$\sum_{i=1}^{n}\text{右辺}=\sum_x \frac{F}{g}N_n$$

となる。そこで(B)の分母を$T$でおきかえて

$$P_i^{\text{new}}=\left[P_i\sum_x \frac{F}{g}N_i\right]^{\text{old}}\Big/T$$

とするとうまく収束する。

連立方程式の解法において右辺と左辺を別々にすべて加えるというテクニックはよく用いられるが，彼がどうしてこの変形に到達したのか不明である。おそらく試行錯誤の末に得たと思われるが，何らかの（幾可学的？）考察によるものかもしれない。彼の反復法はいわゆる“名人芸”である。多重確率分布のようなタイプはこの反復法で適用可能であろう。しかし，最小２乗法等には適用はむずかしいようである。

(2) Cohen (1966)

Cohen は Hasselblad の論文に寄せて以下のコメントをした。

① サンプル数は大きく。

② ヒストグラムの巾は小さく。

③　正規分布の数は既知。

④　大型計算機が必要。

①については当然であるが，正規分布が十分に離れているような条件の良いデータについては，それほど大きくなくてもよい。②についてはヒストグラムを用いずに生データから直接推定するのが最良で，HASSELBLAD の反復法でも対応可能である。線型近似による誤差伝播則より推定値の誤差（分散）はヒストグラムの巾の2乗に比例することが示される。③については当然の前提であるが，不明の場合にはAIC（赤池の情報量規準）によって判定するのが妥当であろう。④については現在ではパソコンで十分である。

以上のようにCOHENの指摘はこのテーマについての注意事項のようなもので（大半は現在ではほとんど意味を持たない），HASSELBLAD の論文を批判するものではなく，むしろ賞賛しているようである。

### (3)　北野（1978）

HASSELBLAD の反復法で$\sigma$が一定の場合の特殊例について，セイコーS500のアセンブラでプログラム化した。言語が特殊だったため残念ながら普及しなかった。

### (4)　MACDONALD and PITCHER（1979）

この論文では目的関数としてカルバック情報量（これは計算不能）の近似より$\chi^2$最小化法を採用している。カルバック情報量は“あてはまり”の尺度であるから，これの近似式であるAICは“モデルの選択”規準に用いられる。確率分布のパラメータ推定については最尤法がベストである。これはヒストグラムの巾を小さくしていった場合，最尤法の解は精度が向上するが（計算量は増加する），最小2乗法系の手法は適用不能になることから簡単に理解されよう。最尤法については既にHASSELBLADが解決している。

最適化法としてシンプレックス法を採用している。これは直接探索法 direct search だから精度はかなり悪い。またすべてのパラメータを同時に動かすことができないので部分ごとに行なっている。この論文の意義を見つけることは困難なように思う。

### (5)　嶋津（1980）

大型計算機のサブルーチンを使用して初めてこのテーマにマルカール法を適用した。ただし目的関数は最小2乗法と$\chi^2$最小化法であった。この方面の研究（および資源解析学全般）の進むべき方向を示した意義は大きい。

### (6)　赤嶺（1982）

パソコンを用いてガウス・ザイデル法で解いた。目的関数は最小2乗法であり，メモリーを最小限に押さえたため計算効率は悪い。収束に時間がかかるのが欠点であるが，このテーマの入門としては適当であったと思う。

もともとこのプログラムは田中（1956）の方法の改良として生まれたものである。田中の方法は当時の他手法と異なり，反復修正というアイデアを含んでいてガウス・ザイデル法的であり歴史的に重要である。

(7) **松崎ほか (1983)**

外国ではNORMSEPの改良版であるENORMSEP (Expanded N〜) が用いられている。これはHASSELBLADの方法にHARDINGの方法で初期値を与えるものである。これに改良を加えて移殖したもので，FORTRAN Ⅳで書かれている。プロッターで作図も行なう。

初期値推定は非線型最適化法では重要なテーマであるが，プログラムによる完全自動化はまだ問題があるように思う。局所的最小値やAICとのからみがあるからである。なお，HASSELBLADの反復法自体はHARDINGの方法とは無関係のものである。

(8) **赤嶺 (1984)**

ガウス・ザイデル法をマルカール法に改良したことにより，収束が速くなり実用的となった。ただし，やり直しループで $H$ と $g$ の値を保存していないというケアレスミスがあった。このループは実質上収束判定としてしか機能していなかったため，致命的なミスとならず発見が遅れてしまった。

(9) **赤嶺 (1985)**

前報のプログラムのミスを修正し，目的関数は最尤法と最小2乗法系4種について比較した。最尤法がベストであるが，端で分布が切断されているようなデータについては最小2乗法や $\chi^2$ 最小化法の方がよい（数値積分や数値微分が高速かつ高精度で行えれば最尤法でも適用可能となる)。またキダイのデータに適用し，解の近傍の共分散行列を求めた。これより $\mu$ は比較的安定しているが，$P$ と $\sigma$ は相関が高く，隣の群と負の相関を示して不安定であることが明らかとなった。

(10) **AKAMINE (1987)**

マルカール法とHASSELBLADの反復法，ガウス・ザイデル法とその変法を最尤法で比較した。マルカール法とHASSELBLADの反復法の実用性が再確認された（マルカール法はニュートン法の拡張であるが，ニュートン法と反復法の比較はHASSELBLADが既に行っている)。

マルカール法はHASSELBLADの反復法より現代的であるが，それは以下の理由による。

① 解の近傍で2次収束に近づくので，後半の収束が速く精度が高い。

② 原理が単純で理解しやすくプログラムも作りやすい。そのわりによく働く。

③ 収束判定を自動的にやってくれる。$\lambda$ を連続10回大きくしても $\Delta Y < 0$ とならない場合を収束と判定しているが，これは使用した計算機（言語）とアルゴリズムの精度の限界で，これ以上計算してもムダということである。伊理 (1981) によればパラメータの修正量の大きさで判定する方法は好ましくない。

④ もともと最小2乗法用に開発された手法であるが，ニュートン法の自然な拡張なので一般の最適化法にも適用できる（一般の最適化法にはDAVIDONの方法が良いとされているが，両手法の比較は今後の課題である)。

## 2．成長曲線の拡張とパラメータ推定

(1) **PITCHER and MACDONALD (1973)**

バータランフィの拡張式として2種類提示している。最初はバータランフィの式を切断して直線で接

続するモデルである。(生物学的域値を考えると)実際の成長はこれに近いものかもしれないが，パラメータ推定ではスイッチを制御するパラメータについての数値微分が必要となる。システムダイナミックス等のモデルで使用するには良いかもしれない。

もう1つのモデルは，

$$l = l_\infty [1 - \exp \{-C \sin \frac{2\pi}{52} (t - t_1) - K(t - t_0)\}]$$

である。$t$ の単位を週としたため52で割っている。この式の導き方は示されていない。$C$ や $t_0$ のパラメータの性質があいまいなような印象を受ける。最適化法は計算機による直接探索法となっているが，具体的には書かれていない。

(2) PAURY and DAVID (1981)

使用している式は(1)の2番目の式と同等のものであるが，$t$ の単位は年である。$C$と$t_0$の性質のあいまいさは同様である。バータランフィの式よりもリチャードの式に近い。導き方はこれも示されていない。(1)とは独立に導かれた可能性がある。拡張式自体は高校数学レベルで導くことができるので，オリジナリティにこだわる必要はないと思う。

最適化法は多重正規分布のヒストグラムから直接に求めるもので，プログラム名はELEFAN Ⅰ (Electronic Length Frequency Analysis) という。この手法は詳しく解説されているが，あまり精度がよくないようである。

(3) AKAMINE (1986a)

拡張式は

$$l' = a(l_\infty - l)\, f(t),\quad f(t+1) = f(t)$$

$$l = l_\infty (1 - \exp [-K \{F(t) - F(t_0)\}]),\quad F'(t) = f(t)$$

である。たとえば

$$f(t) = \begin{cases} 1: & n + t_1 \leqq t < n + t_2 \\ 0: & n + t_2 \leqq t < (n+1) + t_1 \end{cases}$$

$n$：整数

とすれば(1)の最初のモデルとなる。ここで採用したのは

$$f(t) = \frac{1+a}{2} + \frac{1-a}{2} \cos 2\pi (t - t_1)$$

$$F(t) = \frac{1+a}{2} t + \frac{1-a}{4\pi} \sin 2\pi (t - t_1)$$

で，$t$ の単位は年である。これは(1)の2番目および(2)のモデルと同等であるが，$t = t_0$ で $l = 0$，$a < 0$ でマイナス成長が現われるのでパラメータの性質がわかりやすい。

最適化法はマルカール法で，目的関数は重みつき最小2乗法なので実用的である。別タイプの拡張式との比較，もとの式との比較，定差図法との比較，$l'$の極値，ロジスティックおよびゴンパーツの式の拡張も行なっている。リチャードの式（赤嶺1986b）についてのマルカール法によるパラメータ推定プ

ログラムを準備中である。

## 3．境界領域的なもの

(1) SCHNUTE and FOURNIER (1980)

MACDONALD and PITCHER (1979) の発展型で成長曲線と組み合わせて論じている。したがって同じ欠点を受け継いでいる。

彼らはバータランフィの式の取扱いについて，３つのパラメータ ($l_\infty$, $K$, $t_0$) のかわりに ($l$, $L$, $k$) を推薦している。$l$ は最も若い年齢群の平均体長，$L$ は最高令のそれ，$K=-\ln k$ である。この方が多重正規分布から推定しやすくパラメータも安定するというのがその理由である。

赤嶺 (1986a) ではバータランフィの式の区間推定も行なっているが (p92の Fig. 3)，$K$ と $t_0$ には正の相関が，これらと $l_\infty$ には負の相関が認められる。したがって $l$ と $L$ の付近は曲線が最も変化しにくい部分と考えられる。逆に $l_\infty$ や $t_0$ は最も変化する部分である。$k$ が少し動くだけで $K$ が大きく変化することは自明である。つまり ($l$, $L$, $k$) はパラメータ推定において ($l_\infty$, $K$, $t_0$) よりも高い精度が要求されるわけである。したがってパラメータを ($l$, $L$, $k$) に変更するメリットはほとんどない。

(2) SCHNUTE (1981)

微分方程式の形式で成長モデルを提示している。これはリチャードの式，$n$ 次式，指数式を含んでいるが，成長式として重要なのはリチャードの式だけである。データ解析には微分型は不適当で積分型に変換する必要がある。したがってリチャードの式（積分型）だけを扱えば十分ではないかという気がする。最適化法は例によってシンプレックス法である。

(3) ASANO and TANAKA (1985)

多重正規分布の $\mu$ だけをヒストグラムの凹凸から判定しようとするプログラムである。実際の体長組成データは条件が悪くうまく分解できないことが多いため，このような方向の方が実用的な気もする。

(4) 田中・田中 (1986)

成長を加味したベイズ型モデルにより多重正規分布を分解しようとする方法で，最適化法は DAVIDON 法を採用している。現代的なアプローチなので成果が期待される。

## 参考文献

赤嶺　達郎 (1982)　Polymodal な度数分布を正規分布へ分解する BASIC プログラム．日水研報 (33)，163－166.

―――― (1984)　MARQUARDT 法による Polymodal な度数分布を正規分布へ分解する BASIC プログラム．日水研報 (34)，53－60.

―――― (1985)　Polymodal な度数分布を正規分布へ分解する BASIC プログラムの検討．日水研報 (35)，129－159.

AKAMINE, T. (1986a)　Expansion of growth curves using a periodic function and BASIC programs by

MARQUARDT's method. 日水研報 (36), 77－107.

赤嶺 達郎 (1986b) リチャードの成長式. 日水研連絡ニュース (338), 2－4.

AKAMINE, T. (1987) Comparison of algorithms of several methods for estimating parameters of a mixture of normal distributions. 日水研報 (37), 259－277.

ASANO, K. and S. TANAKA (1985) A simple method of analysing the polymodal frequency distributions of fish egg dismeters and its application to the ovarian eggs of Japanese common mackerel. *Bull. Fac. Fish., Mie Univ.*, (12), 13－27.

COHEN, A. C. (1966) Discussion of "Estimation of parameters for a mixture of normal distributions" by Victor Hasselblad. *Technometrics*, 8 (3), 445－446.

HASSELBLAD, V. (1966) Estimation of parameters for a mixture of normal distributions. *Technometrics*, 8 (3), 431－444.

一松 信 (1982) 数値解析. 朝倉書店, 東京, pp. 163.

伊理 正夫 (1981) 数値計算. 朝倉書店, 東京, pp. 173.

北野 裕 (1978) 複合正規分布の分解. 漁業資源解析のための電子計算機プログラム集, 水産庁, 19－30.

MACDONALD, P. D. M. and T. J. PITCHER (1979) Age-groups from size-frequency data : a versatile and efficient method of analyzing distribution mixtures. *J. Fish. Res. Board Can.*, (36), 987－1001.

松崎憲四郎・松崎加奈恵・小川 数也 (1983) 年令群クラス分けプログラムとその応用. 日本ベントス研究会誌 (25), 26－32.

PAURY, D. and N. DAVID (1981) ELEFAN Ⅰ, a BASIC program for the objective extraction of growth parameters from length-frequency data. *Meeresforsch*, (28), 205－211.

PITCHER, T. J. and P. D. M. MACDONALD (1973) Two models for seasonal growth in fishes. *J. Appl. Ecol.*, (10), 599－606.

SCHNUTE, J. (1981) A versatile growth model with statistically stable parameters. *Can. J. Fish. Aquat. Sci.*, (38), 1128－1140.

——— and D. FOURNIER (1980) A new approach to length-frequency analysis : growth structure. *Can. J. Fish. Aquat. Sci.*, (37), 1337－1351.

嶋津 靖彦 (1980) 体長組成から年齢組成を推定する一方法. 昭和54年度漁業資源研究会議西日本底魚部会会議報告, 36－48.

田中 栄次・田中 昌一 (1986) 成長を加味した体長組成の複合正規分布への分解方法について. 昭和61年度日本水産学会秋期大会講演要旨集, 13－13.

田中 昌一 (1956) Polymodalな度数分布の一つの取扱方及びそのキダイ体長組成解析への応用. 東海水研報 (14), 1－13.

——— (1985) 水産資源学総論. 恒星社厚生閣, 東京, pp. 381.

[質 疑 応 答]

池田（日水研）　体長組成をいくつかの正規分布に分解した際，$\chi^2$－検定によってもとの資料とのfitnessを吟味することは意味がないのか。

赤嶺　ヒストグラムの作り方（幅および起点）は何通りもあるからfitness自体はあまり意味がない。確率分布のパラメータ推定では尤度のほうが基本的である。