

混合正規分布とEMアルゴリズム

誌名	水産海洋研究
ISSN	09161562
著者	赤嶺, 達郎
巻/号	69巻3号
掲載ページ	p. 174-183
発行年月	2005年8月

寄稿

混合正規分布とEMアルゴリズム

赤嶺達郎

Mixture of Normal Distributions and EM algorithm

Tatsuro AKAMINE[†]

EM algorithm for estimating category proportions is involved in Hasselblad (1966) method for estimating a mixture of normal distributions. Although the convergence of EM algorithm is proved mathematically, the sufficient condition and the necessary conditions are obtained for convergence of Hasselblad method. When proportion converges into 0 which is outside of its range, strange phenomena occur. In this paper, Hasselblad method based on the strict iteration method for a mixture of normal distributions, mathematical proof of EM algorithm by using Kullback-Leibler information quantity, comparison with general maximization methods, numerical examples of bad convergence, and the special case of a mixture of normal distributions which is difficult to apply EM algorithm are explained.

Key words: EM algorithm, Hasselblad method, mixture of normal distributions, proportion

複数の年級群を含む魚類の体長組成データは多峰型の分布を示すことが多く、これを混合正規分布とみなして各正規分布の混合率、平均、および分散を推定する試みが1940年代頃から行われてきた(田中, 1985)。当初は正規確率紙などを用いて曲線を当てはめる手法が中心であったが、Hasselblad (1966)以降は計算機を用いた手法が主体となった。体長組成データの年齢分解における基本的な解析方法は以下の4つに大別できる。

- (a) 各正規分布の混合率、平均、および分散をすべて推定する(混合率は既知)。
- (b) 成長曲線などが得られている場合、それらから各正規分布の平均と分散を推定し、これを用いて体長組成データから混合率だけを推定する。
- (c) 一部の標本を抽出して年齢査定を行い、各年級群の体長分布データを得る。これを用いて体長組成データから混合率だけを推定する。
- (d) 各正規分布の平均が成長曲線などにしたがる場合、そのパラメータを混合率および分散とともに推定する。

パラメータ推定の立場からみると(a)がもっとも一般的で、(b)と(c)は(a)のパラメータの一部を外部から与えて固定したものに相当する。いくつかの重要魚種では体長組成

が団塊状となってしまうため、(a)ではうまく推定できない場合がある。そのため(a)において平均と標準偏差に線型関係を導入するなどの制約条件を付けたモデルが従来から工夫されてきた。(d)もそのような工夫の一種である。混合率が未知の場合には適当ないくつかの混合数を仮定して、それぞれにおいてパラメータ推定を行った結果を比較検討することが多い。(b)は(a)で推定できない魚種にも適用可能であり、(c)の一種とみなすこともできる。(c)は混合正規分布ではなく一般的な混合分布となるが、遺伝子頻度データを用いて混合率だけを推定するモデル(岸野, 1999; 北田, 2001)と数学的に同等である。

これらすべてがHasselblad (1966)の開発した反復法(以下、Hasselblad法)によって解決されたわけで、その収束原理についても1970年代にEMアルゴリズムであることが明らかにされた。一方、Akamine (1998)は一般的な最適化法であるヤコビ法の近似としてHasselblad法が導けることを示した。さらに赤嶺(2001)では混合率の推定において収束するための十分条件を提示したが、EMアルゴリズムによって常に収束が保証されているのであるから、このような条件がつくのは奇妙な印象を受ける。

この小論では混合率の推定においては、EMアルゴリズムだけでは収束に関する議論として不十分であること、および混合率の一部が0に収束する場合は見かけの収束であることを示す。これに関連して混合正規分布におけるHasselblad法の導出、EMアルゴリズムによる収束原理の説明、一般的な最適化法との比較および数値例による収束の検討、分散に特別な条件をつけた場合の計算方法などについて

2005年4月20日受付、2005年6月7日受理

独立行政法人水産総合研究センター中央水産研究所
National Research Institute of Fisheries Science, Fisheries Research
Agency, 2-12-4 Fukuura, Yokohama, 236-8648, Japan

[†] akabe@affrc.go.jp

て解説する。

混合正規分布モデル

最初に一般的な混合分布を以下のように定義する。

$$G(x) = \sum_{i=1}^n p_i g_i(x), \quad \sum_{i=1}^n p_i = 1, \quad \sum_x g_i(x) = 1 \quad (1)$$

ここで $G(x)$ は混合分布, $p_i > 0$ は混合率, $g_i(x) > 0$ はそれぞれの確率分布 (連続分布の場合には確率密度関数), n は混合数 (既知), x についての Σ は $x = x_1, \dots, x_m$ における総和で連続分布の場合には積分となる。上述の体長組成解析の(c)では $g_i(x)$ が年級 i の体長組成データであり, $G(x)$ は推定された全体の体長組成である。体長組成解析の(b)も各正規分布はデータとみなせるので, これと同じモデルとなる。体長組成解析の(a)では各確率分布は正規分布:

$$g_i(x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2} \frac{(x-\mu_i)^2}{\sigma_i^2}\right) \quad (2)$$

となる。この場合は正式には推定するパラメータも含めて $g_i(x) = g_i(x; \mu_i, \sigma_i^2)$ および $G(x) = G(x; \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ と表すべきであるが, 数式が煩雑となるため(1)式と同じ書き方とする。各正規分布のパラメータも同時に推定する場合には, 常に正式な数式を念頭に置いて読んでほしい。なお確率分布に関する合成関数的なものも混合分布と呼ぶことがあるので注意する必要がある (竹内, 1989)。

最尤法

1950年代までの混合正規分布におけるパラメータ推定方法はグラフを用いたり変数変換した上での当てはめで, 最小2乗法的なものであった。Hasselblad (1966) は計算機を用いて, 初めて最尤法をこの分布に適用した。観測された体長組成データとして $h^*(x)$ を頻度, $h(x)$ をその相対頻度とする,

$$h^*(x) = Hh(x), \quad \sum_x h(x) = 1 \quad (3)$$

ここで H は総頻度である。これより混合正規分布の対数尤度は

$$\log L = \sum_x h^*(x) \log G(x) = HY, \quad Y = \sum_x h(x) \log G(x) \quad (4)$$

となるから, 目的関数 $Y < 0$ を最大にするパラメータの値を求めればよい。これは非線型モデルなので反復法でないと解くことができない。パラメータは p_i, μ_i, σ_i^2 の3種類であるが, p_i については(1)式の制限 $\sum p_i = 1$ があるため本質的に $(n-1)$ 個なので, 合計で $(3n-1)$ 個のパラメータを推定する問題である。

Hasselblad法

最尤法では Y の最大値を求めるのだから, 各パラメータで微分して0とおけばよい。最初に p_i について微分してみる。(1)式の制限があるためラグランジュの未定乗数法 (赤嶺, 1997) を用いる。

$$Y^* = Y - \lambda \left(\sum_i p_i - 1 \right) \quad (5)$$

とおいて, p_i で微分して0とおくと

$$\sum_x \frac{h(x)g_i(x)}{G(x)} = \lambda \quad (6)$$

を得る。ここで未定乗数 λ の値が問題となるが,

$$\lambda = \lambda \sum_i p_i = \sum_x \sum_i \frac{h(x)p_i g_i(x)}{G(x)} = \sum_x h(x) = 1 \quad (7)$$

であるから, 最終的に解が満たす n 個の非線型方程式:

$$\sum_x \frac{h(x)g_i(x)}{G(x)} = 1 \quad (i = 1, \dots, n) \quad (8)$$

を得る。この方程式を数値的に解くためHasselbladはこの両辺に p_i を掛けて,

$$p_i^{(t+1)} = \left(p_i \sum_x \frac{h(x)g_i(x)}{G(x)} \right)^{(t)} \quad (9)$$

という反復法を導いた。ここで各変数の右肩の (t) は t 回目の反復計算の値を意味している。解の近傍では(8)式が成立するため, 収束が非常に遅くなるのが分かる。この(9)式は岸野(1999)の(7.22)式および北田(2001)の(11.5)式と同一の式で, 推定方法として広く用いられており, EMアルゴリズムとして解説されている。以降では数式が煩雑となるため, $h = h(x), g_i = g_i(x),$ および $G = G(x)$ と略記することにする。

平均 μ_i については Y を微分して0とおくと,

$$\sum_x \frac{hg_i(x-\mu_i)}{G} = 0 \quad (10)$$

という非線型方程式を得る。これよりHasselbladは

$$\mu_i^{(t+1)} = \left(\frac{\sum_x \frac{hg_i x}{G}}{\sum_x \frac{hg_i}{G}} \right)^{(t)} \quad (11)$$

という反復式を導いた。同様に分散 σ_i^2 についても,

$$\sum_x \frac{hg_i \{\sigma_i^2 - (x-\mu_i)^2\}}{G} = 0 \quad (12)$$

という非線型方程式を得るから,

$$\sigma_i^{2(t+1)} = \left(\frac{\sum_x \frac{hg_i(x-\mu_i)^2}{G}}{\sum_x \frac{hg_i}{G}} \right)^{(t)} \quad (13)$$

という反復式を導いた。これらの反復法はBASICプログラムや表計算ソフトのワークシートが供給されており、広く活用されている (Akamine, 1987; 相澤・滝口, 1999; 五利江, 2002)。しかし収束する理由が説明できないため、数値解析上はそれについての議論が必要となる。それがEMアルゴリズムによる解釈やヤコビ法からの近似である。

EM アルゴリズム

1) 反復式

Hasselblad法が収束する理由について最初にEMアルゴリズムを用いて説明したのはOrchard and Woodbury (1972)である。以後、何人かが同様の証明を行っている (Kimura and Chikuni, 1987; Millar, 1987)。EMアルゴリズムはもともとランダムに欠測データが発生する場合に、欠測データを推定するEステップと、欠測データを含む完全データを用いてパラメータの最適化を行うMステップを交互に行う反復法として開発された。その後、欠測データだけでなく潜在変数についても、同様に有効であることが判明し広く普及した。その際に混合正規分布のパラメータ推定におけるHasselblad法の成功が大きく寄与したようである (宮川, 1987)。

最初に潜在データ $z_i(x)$ として x におけるカテゴリー i の全体に対する相対頻度を設定して、Eステップで推定してみよう。潜在パラメータ $q_i(x)$ として x においてカテゴリー i に属する確率を設定すると、これは「ベイズの定理」から事後確率として

$$q_i(x) = \frac{p_i g_i(x)}{\sum_i p_i g_i(x)} = \frac{p_i g_i(x)}{G(x)} \quad (14)$$

と推定される。したがって潜在データとして期待値を採用すれば、

$$z_i(x) = h(x) q_i(x) = \frac{h(x) p_i g_i(x)}{G(x)} \quad (15)$$

と推定される。これらは自明な関係式であるが、重要なのはこのEステップは定義式だけなので、実際の計算量は0ということである。なお岸野 (1999) および北田 (2001) では本論文と記号が異なっているが、 $\theta_j = p_j$ 、 $p_i^{(0)} = h(i)$ 、 $p_i^{(j)} = g_j(i)$ 、 $n_i^{(0)} = Hh(i)$ 、 $n_i^{(j)} = Hz_j(i)$ と書き換えて、 $i \rightarrow x$ 、 $j \rightarrow i$ と置換すれば一致する。前者の(7.27)式、後者の(11.4)式が(15)式である。

次にMステップにおける実際のパラメータの最適化を

考える。

$$\sum_x \sum_i z_i(x) = 1, \quad \sum_x \sum_i p_i g_i(x) = 1 \quad (16)$$

であるから、潜在データ $z_i(x)$ について対数尤度を考えると、

$$\begin{aligned} Q &= \sum_x \sum_i z_i(x) \log(p_i g_i) \\ &= \sum_x \sum_i z_i(x) \log p_i + \sum_x \sum_i z_i(x) \log g_i(x) \end{aligned} \quad (17)$$

となる。EMアルゴリズムではMステップでこの Q 関数を最大化する。そうすると後述のように対数尤度 Y もMステップで増加することが証明される。

混合率 p_i についてEMアルゴリズムを適用してみよう。(1)式の制限があるから、ラグランジュの未定乗数法を適用して

$$Q^* = Q - \lambda \left(\sum_i p_i - 1 \right) \quad (18)$$

とおいて、 p_i で微分して0とおくと

$$\sum_x \frac{z_i(x)}{p_i} = \lambda \quad (19)$$

を得る。これより

$$\lambda = \lambda \sum_i p_i = \sum_x \sum_i z_i(x) = 1 \quad (20)$$

となるから、結局、

$$p_i^{(t+1)} = \sum_x z_i(x)^{(t)} \quad (21)$$

という反復式を得る。これはHasselblad法の(9)式と同一である。Hasselblad法では解を満たす等式を変形して反復式を導いたが、EMアルゴリズムでは Q 関数を最大化することによって反復式を自然に導いている。なお岸野(1999)の(7.26)式、北田(2001)の(11.3)式が(21)式である。

平均 μ_i については Q を微分して0とおくと

$$\sum_x z_i(x)(x - \mu_i) = 0 \quad (22)$$

となるから、

$$\mu_i^{(t+1)} = \frac{\sum_x z_i(x)^{(t)} x}{\sum_x z_i(x)^{(t)}} \quad (23)$$

という反復式を得る。これもHasselblad法の(11)式と同一である。分散 σ_i^2 についても Q を微分して0とおくと

$$\sum_x z_i(x) \{ \sigma_i^2 - (x - \mu_i)^2 \} = 0 \quad (24)$$

となるから、

$$\sigma_i^{2(t+1)} = \frac{\sum_x z_i(x)^{(t)} (x - \mu_i^{(t+1)})^2}{\sum_x z_i(x)^{(t)}} \quad (25)$$

という反復式を得る. この式はHasselblad法の(13)式と比較すると, 右辺の μ_i の右肩の (t) が $(t+1)$ に変更されているが, 反復法においては大同小異である.

Hasselblad法では反復式が収束するかどうか数学的な保証は何もなかったが, EMアルゴリズムでは毎回のMステップで Q 関数の最大化が行われ, 対数尤度が常に増加することが証明されている(この節の最後を参照). したがって対数尤度の上限値が存在する限り, EMアルゴリズムは対数尤度の極大値に収束するわけである.

2) KL情報量

EMアルゴリズムの収束を証明する前に, 証明で使用するKL(カルバック・ライブラー)情報量を簡単に解説する. KL情報量は

$$\text{KL}(p, q) = \int p(x) \log \frac{p(x)}{q(x)} dx \geq 0 \quad (26)$$

で定義される. ここで $p(x)$ と $q(x)$ は確率分布である. $p(x) \equiv q(x)$ のとき等号が成立する. 坂元ほか(1983)では $\log x \leq x-1$ を用いて離散系の場合の

$$\sum_x p(x) \log p(x) \geq \sum_x p(x) \log q(x) \quad (27)$$

を証明しているが, ここではラグランジュの未定乗数法を用いて証明する. $p(x)$ と $q(x)$ は確率分布だから

$$\sum_x p(x) = 1, \quad \sum_x q(x) = 1 \quad (28)$$

である. ここで未定乗数 λ を用いて

$$E = \sum_x p(x) \log q(x) - \lambda \left(\sum_x q(x) - 1 \right) \quad (29)$$

の最大値を求める. $q(x)$ で微分して0とおくと

$$\frac{p(x)}{q(x)} = \lambda \quad (30)$$

となる. これより

$$\lambda = \lambda \sum_x q(x) = \sum_x p(x) = 1 \quad (31)$$

であるから,

$$p(x) = q(x) \quad (32)$$

を得る. これがすべての x で成立するとき E は最大となるから, (27)式は証明された. なお混合分布の対数尤度(4)についても,

$$Y = \sum_x h(x) \log G(x) \leq \sum_x h(x) \log h(x) \quad (33)$$

であることが証明された.

3) 収束の証明

Eステップを潜在変数の推定, Mステップを実際のパラメータの推定と解釈すると, EMアルゴリズムは基本的な最適手法であるガウス・ザイデル法(以下, GS法)と似たアイデアの手法となるため, きわめて頑健な手法となることが期待される. 収束の一般的な証明は既にいくつかの教科書に載っている(McLachlan and Krishnan, 1997; 岸野, 1999; 渡辺・山口, 2000; McLachlan and Peel, 2000; 岩崎, 2002). Jensenの不等式を用いることが多いが, ここでは上田(2003)にあるKL情報量を用いた一般的な証明を混合分布に適用してみよう. これによって一般的な証明の理解が容易になると思われる. 潜在データ $z_i(x)$ と潜在パラメータ $q_i(x)$ を(14)および(15)式と同様に定義する. まず次のKL情報量を考える.

$$\text{KL}(\xi_i, q_i) = \sum_i \xi_i \log \frac{\xi_i}{q_i} \quad (34)$$

ここで ξ_i は q_i と少し異なる値で,

$$\sum_{i=1}^n \xi_i = 1, \quad \xi_i > 0 \quad (35)$$

と定義する. 定義式(14)を代入すると

$$\text{KL}(\xi_i, q_i) = \sum_i \xi_i \log \xi_i - \sum_i \xi_i \log(p_i g_i) + \sum_i \xi_i \log G \quad (36)$$

となる. 最後の項は $\log G$ だから

$$\log G = \sum_i \xi_i \log(p_i g_i) - \sum_i \xi_i \log \xi_i + \text{KL}(\xi_i, q_i) \quad (37)$$

となる. この式に l 回目の古い値として $\xi_i = q_i^{(l)}$ を代入し, $h(x)$ をかけ x について和をとると

$$Y = \sum_x h(x) \log G(x) = \Theta + \sum_x h(x) \text{KL}(q_i^{(l)}, q_i) \quad (38)$$

を得る. ここで

$$\begin{aligned} \Theta &= \sum_x \sum_i z_i(x)^{(l)} \log(p_i g_i) - \sum_x \sum_i z_i(x)^{(l)} \log q_i^{(l)} \\ &= Q - \sum_x \sum_i z_i(x)^{(l)} \log q_i^{(l)} \end{aligned} \quad (39)$$

である. $(t+1)$ 回目のMステップにおいて Q を最大化すると, このとき(39)式の第2項は変化しないため, Θ も最大化される. したがって(38)式において $q_i^{(l)}$ は古い値, q_i は新しい値となるから, KL情報量は正の値となって増加する. 以上より対数尤度 Y は必ず増加することが証明された.

最適化法との比較

前節の証明によりEMアルゴリズムにおける反復過程では常に尤度が上昇することが分かった。しかし一般の最適化法においても尤度が常に上昇するようにパラメータ値を修正しているが、収束が非常に遅かったり、途中で止まってしまう場合がある。またすべてのモデルに対してEMアルゴリズムが適用できるわけではない。したがって一般の最適化法と比較して、収束の速さや収束域の広さを検討することは重要である。

最初に最急降下法について述べる。この方法は傾斜法の基本的な手法であり、Hasselblad自身もHasselblad法を一般化最急降下法と解釈していた。パラメータを θ 、パラメータ空間における目的関数 Y の法線ベクトルを \mathbf{g} とする、つまり

$$\mathbf{g} = \left(\frac{\partial Y}{\partial \theta_i} \right) \quad (40)$$

であるが、 $-\mathbf{g}$ を最急降下ベクトルと呼ぶ。最急降下法は

$$\Delta\theta = -k\mathbf{g} \quad (41)$$

と表すことができ ($\Delta\theta$ はパラメータの修正量、 k は修正因子)、最小値を求める手法であるが、 Y の最大化は($-Y$)の最小化なので最尤法にも適用できる。この手法でも各ステップで対数尤度は必ず上昇するが、一般に解には収束しない。それは「鳥かご現象」という一部のパラメータだけが動いて、他のパラメータがほとんど変化しない現象が発生するからである。これは修正ベクトルが部分空間内に閉じこめられている現象である (Fig. 1)。Hasselblad法では鳥かご現象は発生しないが、それはヤコビ法の近似と見なせるからである (Akamine, 1998; 赤嶺, 1999)。

収束は遅いが確実に解に収束する方法としてGS法がある。これは1変数のニュートン法をすべてのパラメータについて順番に行う方法で、単純な方法で計算量も多いが、

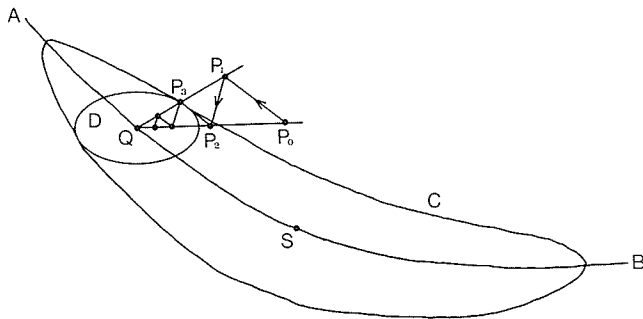


Figure 1. Image of the cage phenomenon of the steepest descent method in 3 dimension. AB, C, D, P, Q, and S indicate the axis of the banana shape, contour of the objective function whose shape looks like banana, subspace, trajectory of iteration, convergence point in subspace, and true solution, respectively.

確実な方法である。ここで「順番に行う」ところが重要である。最初、ガウスは各ステップで修正量が最大のパラメータを選択して修正していく方法を考えたが、鳥かご現象と同じような状況に陥ってしまう。すべてのパラメータを同等に扱うことが重要である。EMアルゴリズムは潜在パラメータと実際のパラメータを交互に最適化していると解釈できるので、原理的にGS法に似ている。

GS法を単純化して、すべてのパラメータで同時にニュートン法を行う手法がヤコビ法である。ヘッセ行列を

$$\mathbf{H} = \left(\frac{\partial^2 Y}{\partial \theta_i \partial \theta_j} \right) \quad (42)$$

と定義し、対角成分だけそのままにして他の成分をすべて0とおいた行列を、

$$\mathbf{D} = \text{diag}\mathbf{H} \quad (43)$$

と表す。これを用いるとヤコビ法は

$$\mathbf{D}\Delta\theta = -\mathbf{g} \quad (44)$$

と表せる。Hasselblad法はヤコビ法の近似として導けるが、実際 Akamine(1987)の Table 4において両者は同程度の収束を示している (Fig. 2)。Hasselblad法は解の近傍で収束が遅いが、ヤコビ法は解の近傍での収束が悪く、解に到達しない場合も多い。

この欠点を克服したのがマルカール法で、

$$(\mathbf{H} + \lambda\mathbf{D})\Delta\theta = -\mathbf{g} \quad (45)$$

と表すことができる。 λ は修正因子で、最初 $\lambda=1$ として、収束するに従って小さくしていく。 $\lambda=0$ のとき多変数のニュートン法となる。多変数のニュートン法は解の近傍では2乗収束するため、急速に収束するが、収束域が非常に

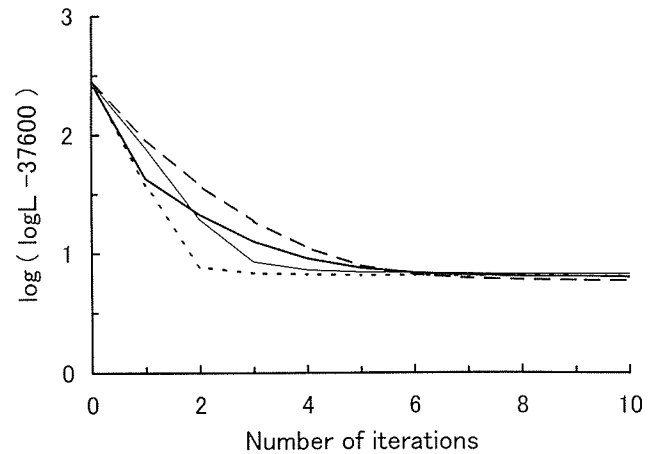


Figure 2. Convergence of estimating methods for the body length data of porgy (Table 4 of Akamine (1987)). Thick line, thin line, broken line, and dotted line indicate Hasselblad method, Jacobi method, Marquardt method, and Gauss-Seidel method, respectively.

狭いため、そのままでは使い物にならない。マルカール法は λ の値につれてヤコビ法から多変数のニュートン法に自然に移行するため、実用的かつ高精度な手法である。ただし計算量が多く、プログラムの作成も簡単ではない。

これらの一般的な最適化法と比較してHasselblad法は以下のような長所を持つ。

- (a) 収束域が広く確実に収束する。これは最大の長所であって、いくら精度が高くても収束域の狭い手法は使い物にならない。
- (b) プログラミングが容易で、表計算ソフトでも簡単に作成できる。制限条件のついたモデルへの拡張も容易である。
- (c) 計算量が小さいため、反復回数が多くても全体の計算時間は少ない。

短所としては、

- (a) 解の近傍での収束が遅い。

が挙げられるが、現在の計算機的能力からするとほとんど問題にならない。したがって混合率および混合正規分布のパラメータ推定について、Hasselblad法はほぼ満足のいくレベルに達しているといえる。

収束の検討

ここから混合正規分布においてうまく収束しない場合を検討する。最初の例として対数尤度の上限値が存在しない、つまり ∞ となる例が知られている。任意の x において、任意の正規分布の分散の値を0に近づけていくと、この正規分布はディラックのデルタ関数(超関数)に近づき、対数尤度は ∞ となる。この場合は数学的に収束しない。しかしこれは分散の初期値を0に近い値に設定した場合に起きる現象なので、通常の体長組成解析などでは発生しない。

実際にしばしば起きる現象として問題なのは、分散の値がどんどん大きくなる場合である。魚類の体長組成を分解する場合には、右端の方に飛び離れた点が生じやすく、このような現象がしばしば起きる。そのようなデータを除去したり、正規分布の数を減らすなどの対策が必要である。

混合率だけを推定する場合には上記2つの問題は生じない。しかも $0 < p_i < 1$ なのでパラメータ値は発散しない。しかし赤嶺(2001)は「縮小写像の原理」(伊理, 1981)を適用して以下の判定式を導いた。(9)式を $p_i^{(t+1)} = y_i(p_1^{(t)}, \dots, p_n^{(t)})$ と書き直すと、

$$\begin{aligned} \sum_i |dp_i^{(t+1)}| &= \sum_i \left| \sum_k \frac{\partial y_i}{\partial p_k} dp_k \right|^{(t)} \leq \sum_i \sum_k \left| \frac{\partial y_i}{\partial p_k} \right|^{(t)} |dp_k^{(t)}| \\ &\leq \max \left(\sum_i \left| \frac{\partial y_i}{\partial p_k} \right|^{(t)} \right) \sum_k |dp_k^{(t)}| = |\mathbf{J}^{(t)}| \sum_i |dp_i^{(t)}| \end{aligned} \quad (46)$$

ここで $|\mathbf{J}^{(t)}|$ は行列 $\mathbf{J}^{(t)}$ のノルムである。このとき

$$\frac{\partial y_i}{\partial p_k} = - \sum_x \frac{hp_i g_i g_k}{G^2} \quad (i \neq k) \quad (47)$$

$$\frac{\partial y_k}{\partial p_k} = \sum_x \frac{hg_k(G - p_k g_k)}{G^2} = \sum_x \sum_{i \neq k} \frac{hg_k p_i g_i}{G^2} = \Psi_k \quad (48)$$

であるから、最終的に次の判定式を得る ($i \neq k$ についての Σ は $i=k$ を除くすべての i における総和)。

$$|\mathbf{J}^{(t)}| = \max \left(\sum_i \left| \frac{\partial y_i}{\partial p_k} \right|^{(t)} \right) = 2 \max \Psi_k^{(t)} \leq L < 1 \quad (49)$$

ここで L はリプシッツ定数と呼ばれるもので、数列の収束判定に用いられるものである。上式が成立するときHasselblad法は収束するが、これは十分条件なので、実際はこの条件を満たさなくても収束する。なお必要条件は(各 p_i が発散しなければよいので)。

$$\frac{\partial y_k}{\partial p_k} = \Psi_k \leq 1 \quad (k=1, \dots, n) \quad (50)$$

である。

混合群の数が必要以上に多いと、いくつかの群で $p_i \rightarrow 0$ となるが、 $p_i > 0$ がモデルの前提条件なので、通常の正常な収束とは異なる。Hasselblad法では初期値として $p_i > 0$ を与えると、以降の p_i は絶対に0にならないから数学的には永久に収束しないが、数値上はどんどん0に近づく。これを「見かけの収束」と呼ぶことにする。混合分布では分布の個数に制限がないから、このような問題が生じるのである。この問題は $p_i = 1/b_i$ と変換すると理解しやすい。このとき $1 < b_i < \infty$ であり、

$$G(x) = \sum_{i=1}^n \frac{g_i(x)}{b_i}, \quad \sum_{i=1}^n \frac{1}{b_i} = 1 \quad (51)$$

となる。これは(1)式と数学的に同一のモデルである。反復式はEMアルゴリズムより

$$b_i^{(t+1)} = \left(\frac{b_i}{\Omega_i} \right)^{(t)}, \quad \Omega_i = \sum_x \frac{hg_i}{G} \quad (52)$$

となる。 $p_i \rightarrow 0$ のとき $b_i \rightarrow \infty$ となるから、これは発散である。 $b_i^{(t+1)} = \eta_i(b_1^{(t)}, \dots, b_n^{(t)})$ として、(47)式と(48)式を書き直すと、

$$\frac{\partial \eta_i}{\partial b_k} = - \frac{\sum_x \frac{hb_i g_i g_k}{G^2 b_k^2}}{\Omega_i^2} \quad (i \neq k) \quad (53)$$

$$\frac{\partial \eta_k}{\partial b_k} = \frac{\sum_x \frac{hg_k}{G^2} \left(G - \frac{g_k}{b_k} \right)}{\Omega_k^2} = \Phi_k \quad (54)$$

Table 1. Body length distribution of porgy (Akamine, 1987).

Class mark	Frequency	Class mark	Frequency	Class mark	Frequency
7.5	7	15.5	1439	23.5	310
8.5	79	16.5	921	24.5	228
9.5	509	17.5	448	25.5	168
10.5	2240	18.5	512	26.5	140
11.5	2341	19.5	719	27.5	114
12.5	623	20.5	673	28.5	64
13.5	476	21.5	445	29.5	22
14.5	1230	22.5	341	30.5	0

Table 2. Example 1 of Hasselblad method with Table 1 data to estimate parameter p . Each symbol is explained in the text.

No. of normal distribution		1	2	3	4	5	6	7
Initial value	p	0.05	0.3	0.2	0.08	0.02	0.15	0.2
	μ	11.001	15.29	19.704	23.445	27.262	10.75	11.25
	σ	0.8732	1.1381	1.4139	1.6455	1.1603	1	1
After 10000 iterations	p	0.384	0.306	0.178	0.083	0.022	0.026	6.6×10^{-8}
	Ω	1	1	1	1	1	1	0.99884
	Ψ	0.081	0.072	0.183	0.291	0.258	0.925	0.99884
	Φ	0.081	0.072	0.183	0.291	0.258	0.925	1.00116

となる。正常な収束は $p_i \rightarrow a > 0$ であるから、 $\Omega_i \rightarrow 1$ 、および $\Psi_i = \Phi_i < 1$ となる。しかし見かけの収束は $p_i \rightarrow 0$ ($p_i \rightarrow \infty$) であるから、 $\Psi_i \rightarrow \Omega_i < 1$ 、 $\Phi_i \rightarrow 1/\Omega_i > 1$ 、したがって $\Psi_i = 1/\Phi_i < 1$ となる。ここで $\Omega_i < 1$ となっているから、正常に収束していないことが分かる。

数値計算例 1

Akamine (1987) で使用したキダイの体長組成データ (Table 1) について、表計算ソフト (相澤・滝口, 1999; 五利江, 2002) を用いて混合率の推定を行った。最初に正規分布の数を 5 つとしてすべてのパラメータを推定したところ、Akamine (1987) と同一の解に収束した ($\log L = -37605.843$)。そこで得られた平均 μ と標準偏差 σ の値を固定して、以下では混合率 p だけの推定を行った。平均と標準偏差を固定した 2 つの正規分布を新たに 6、7 番目として加えた場合の結果を Table 2 に示す。 p_7 はほとんど 0 に収束したが、10000 回反復しても値が安定しなかった。このとき確かに $\Omega_7 = \Psi_7 = 0.99884$ 、 $\Phi_7 = 1.00116$ なので、互いに逆数となっている。それ以外では $\Omega = 1$ 、 $\Psi = \Phi < 1$ であるが、 $\Psi_6 > 0.5$ となっているため十分条件 (49) を満たしていない。余分な 2 つの正規分布を削除すると、この十分条件が満たされて急速かつ正常に収束する。7 番目の正規分布だけを削除した場合は、正常に収束するが、必要条件しか満たさない

め収束が遅い。

正規分布を 2 つ追加したときの最大対数尤度は $\log L = -37602.714$ となっていて、正規分布が 5 つの場合のよりも大きくなっているが、これは当然である。赤池の情報量規準:

$$AIC = -2 \times \text{最大対数尤度} + 2 \times \text{パラメータ数}$$

で判定すると、正規分布を 2 つ追加したことによって最大対数尤度は 3.129 増加したが、パラメータ数も 2 増加して 7 となっている。6 番目の正規分布だけを追加した場合には最大対数尤度は同一の $\log L = -37602.714$ となるが、パラメータ数は 6 である。AIC では値の小さいモデルが有効と判断されるため、6 番目の正規分布だけを追加したモデルが有効と判断される。しかしこれはあくまでも数値上の話なので、6 番目に追加した人工的な正規分布が生物学的に有効かどうかは別問題である。

数値計算例 2

次に左側半分正規分布を 5 つ詰め込んだ例を Table 3 に示す。この場合は見かけ上すぐに収束するが、正常に収束しているのは 1 番目と 5 番目だけで、残りの 3 つはすべて $\Omega = \Psi = 1/\Phi < 0.5$ となっている。右側半分には分布を当てはめていないため、対数尤度は $\log L = -124475.22$ と非常

Table 3. Example 2 of Hasselblad method with Table 1 data to estimate parameter p . All μ are fixed in small number. Each symbol is explained in the text.

No. of normal distribution		1	2	3	4	5
Initial value	p	0.2	0.2	0.2	0.2	0.2
	μ	14	14.5	15	15.5	16
	σ	1	1	1	1	1
Solution	p	0.548	3.4×10^{-204}	2.5×10^{-250}	2.8×10^{-215}	0.452
	Ω	1	0.473	0.400	0.455	1
	Ψ	0.078	0.473	0.400	0.455	0.095
	Φ	0.078	2.112	2.502	2.199	0.095

Table 4. Example 3 of Hasselblad method with Table 1 data to estimate parameter p . All μ are fixed in large number. Each symbol is explained in the text.

No. of normal distribution		1	2	3	4	5
Initial value	p	0.2	0.2	0.2	0.2	0.2
	μ	35	36	37	38	39
	σ	1	1	1	1	1
After 3 iterations	p	1	2.2×10^{-15}	1.5×10^{-25}	2.6×10^{-36}	3.7×10^{-48}
	Ω	1	1.3×10^{-5}	5.3×10^{-9}	1.4×10^{-12}	1.6×10^{-16}
	Ψ	0	1.3×10^{-5}	5.3×10^{-9}	1.4×10^{-12}	1.6×10^{-16}
	Φ	0	76679	1.9×10^8	7.3×10^{11}	6.5×10^{15}

に小さくなっていて、統計学的なパラメータ推定としてほとんど意味がない。

数値計算例3

さらに極端な例として正規分布5つを右側の離れた点においた例をTable 4に示す。この場合はわずか3回でほとんど収束する。1番目だけ $p=\Omega=1$, $\Psi=\Phi=0$ となるが、残りはすべて $p=0$, $\Omega=\Psi=1/\Phi=0$ となる。不適当な当てはめなので対数尤度は $\log L = -2898038$ と非常に小さくなっていて、統計学的には意味がない。

条件つき最適化

Hasselblad (1966)は混合正規分布において分散が一定、および標準偏差が平均に比例する場合についても反復式を導いている。これらはEMアルゴリズムでも同様に導くことができる。しかしこれらを一般化した $\sigma = \alpha\mu + \beta$ という条件下では、どちらも有効な反復式を導くことができない。

1) 分散が同一の場合

この場合は(17)式の Q 関数において各正規分布の分散を同一の σ^2 とおき、 σ^2 で微分して0とおけば

$$\sum_x \sum_i z_i(x) \{ \sigma^2 - (x - \mu_i)^2 \} = 0 \quad (55)$$

を得るから、

$$\sigma^{2(t+1)} = \sum_x \sum_i z_i(x)^{(t)} (x - \mu_i^{(t+1)})^2 \quad (56)$$

という反復式を得る。右辺で μ_i の右肩がHasselblad法と異なっているのは(25)式と同様である。

2) $\sigma = \alpha\mu$ の場合

この場合も(17)式の Q 関数において各正規分布に $\sigma_i = \alpha\mu_i$ を代入して、 α^2 で微分して0とおけば

$$\sum_x \sum_i z_i(x) \left\{ \alpha^2 - \frac{(x - \mu_i)^2}{\mu_i^2} \right\} = 0 \quad (57)$$

という方程式を得るから、

$$\alpha^{2(t+1)} = \sum_x \sum_i z_i(x)^{(t)} \frac{(x - \mu_i^{(t+1)})^2}{\mu_i^{(t+1)2}} \quad (58)$$

という反復式を得る。同時に Q 関数を μ_i で微分して0とおくと

$$\sum_x z_i(x) \{x(x-\mu_i) - \alpha^2 \mu_i^2\} = 0 \quad (59)$$

という2次方程式を得る. Hasselblad (1966)と同様にして

$$\mu_i^{(r+1)} = \frac{-B + \sqrt{B^2 + 4AC}}{2A} > 0 \quad (60)$$

$$A = \alpha^{2(r+1)} \sum_x z_i(x)^{(r)} > 0$$

$$B = \sum_x z_i(x)^{(r)} x$$

$$C = \sum_x z_i(x)^{(r)} x^2 > 0$$

とすれば正根を得る. (58)および(60)式では右辺の α と μ がどちらも新しい値となっているため, このままでは解くことができない. しかしHasselblad法ではどちらも古い値となっているので, 実用上はどちらかを古い値にすればよい. つまりこの場合には Q 関数の最大化自体に反復法を導入していることになる. なお赤嶺(1999)はニュートン法を用いる次式を提示したが, (60)式とほとんど同じ収束を示す.

$$\mu_i^{(r+1)} = \frac{A\mu_i^2 + C}{2A\mu_i + B} \quad (61)$$

3) $\sigma = \alpha\mu + \beta$ の場合

これは上記2つのモデルを一般化したものである. この場合に Q 関数をそれぞれ α , β , μ_i で微分して0とおくと, 以下の3式を得る.

$$\frac{\partial Q}{\partial \alpha} = \sum_x \sum_i z_i(x) \mu_i \left\{ \frac{(x-\mu_i)^2}{(\alpha\mu_i + \beta)^3} - \frac{1}{\alpha\mu_i + \beta} \right\} = 0 \quad (62)$$

$$\frac{\partial Q}{\partial \beta} = \sum_x \sum_i z_i(x) \left\{ \frac{(x-\mu_i)^2}{(\alpha\mu_i + \beta)^3} - \frac{1}{\alpha\mu_i + \beta} \right\} = 0 \quad (63)$$

$$\frac{\partial Q}{\partial \mu_i} = \sum_x z_i(x) \left\{ \frac{\alpha(x-\mu_i)^2}{(\alpha\mu_i + \beta)^3} + \frac{x-\mu_i}{(\alpha\mu_i + \beta)^2} - \frac{\alpha}{\alpha\mu_i + \beta} \right\} = 0 \quad (64)$$

最後の(64)式は μ_i についての2次方程式となるので(59)式

と同じ方法で解くことができるが, 上の2式は α と β についての簡単な式とならない. したがってHasselblad法およびEMアルゴリズムの適用は困難である. 実用上は表計算ソフトなどの最適化法を用いれば簡単に解が求まる. 初期値が大雑把すぎて収束しないときは, $\sigma = \alpha\mu$ の場合の解などを初期値として用いればよい.

数値計算例4

Table 1のデータについて $\sigma = \alpha\mu + \beta$ を仮定した混合正規分布モデルのパラメータ推定を行った. 何の制約も仮定しないで推定した μ と σ の値をTable 2の1~5番目に示しているが, 1~4番目の μ と σ について回帰直線を求めると $\sigma = 0.0621\mu + 0.1895$ となる. 5番目の σ はこの直線にのらないが, このデータは本来のキダイのデータ(田中, 1956; 赤嶺, 1985a)から右端の飛び離れた5個体のデータを削除したものである. それらを追加すればこの直線にのると思われる.

この仮定のもとではEMアルゴリズムは適用できないため, 表計算ソフトの最適化法(東海, 1997)を用いた. 通常の大雑把な初期値では収束しなかったため, ヤコビ法で何回か反復計算した値を初期値として使用した. 精度は良

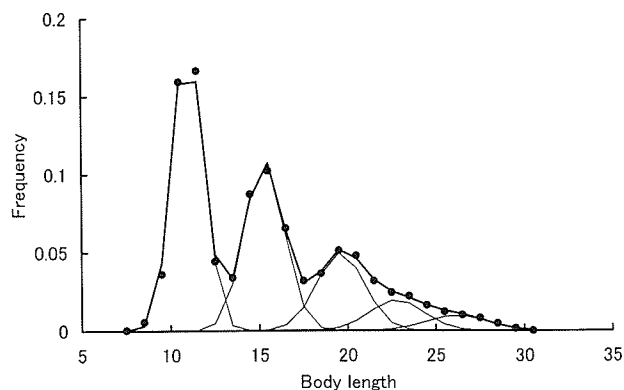


Figure 3. Example of estimation for the mixture of normal distributions under the condition $\sigma = \alpha\mu + \beta$ (Table 5). Closed circles, thick line, and thin lines indicate data distribution, mixture of normal distributions, and each normal distribution, respectively.

Table 5. Example 4 of Hasselblad method with Table 1 data to estimate all parameters. This is the special case in which $\sigma = \alpha\mu + \beta$, and both parameter is estimated as $\alpha = 0.05514$, $\beta = 0.27262$.

No. of normal distribution		1	2	3	4	5
Solution	p	0.4115	0.3043	0.1686	0.0743	0.0412
	μ	11.005	15.293	19.600	22.780	26.203
	σ	0.8794	1.1159	1.3534	1.5287	1.7175

くないが、解を Table 5 と Fig. 3 に示す。パラメータ数は 11 で $\log L = -37613.203$ となっている。何の制約もない場合はパラメータ数 14 で $\log L = -37605.843$ であったから、パラメータ数は 3 減少し、最大対数尤度は 7.36 減少している。AIC の値で判断する限りでは $\sigma = \alpha\mu + \beta$ という仮定は有効でなかったが、削除したデータを追加すれば有効となる可能性がある。

まとめ

混合正規分布のパラメータ推定における EM アルゴリズムを検討して以下の結論を得た。

- (a) 混合率の推定における EM アルゴリズムは Hasselblad 法と同一である。最初の発見者名を尊重して Hasselblad 法と呼ぶべきである。
- (b) EM アルゴリズムでは目的関数 Y の収束は保証されているが、パラメータの収束は保証されていない。混合率の推定において正常な収束では解において $\Omega_i = 1$ が成立するが、 $p_i \rightarrow 0$ の場合には $\Omega_i < 1$ となっているため「見かけの収束」と呼ぶべきである。この場合は余分な混合群を削除すると正常に収束する。
- (c) 混合正規分布のパラメータ推定における EM アルゴリズムは、分散の推定において Hasselblad 法と若干異なる。とりわけ $\sigma = \alpha\mu$ という条件を付けた場合は、そのままでは解けないため、Hasselblad 法と同様の反復式に修正する必要がある。
- (d) 混合正規分布のパラメータ推定において $\sigma = \alpha\mu + \beta$ という条件を付けた場合は、EM アルゴリズムおよび Hasselblad 法の適用は困難である。

Hasselblad 法はアルゴリズムが単純なため広く用いられており、実用性の高い手法である。混合率が 0 に漸近したり、分散がどんどん大きくなるような場合は余分な混合群を削除すれば正常に収束する。アルゴリズムの問題ではなく、データおよび計算上の仮定（混合群の数など）の問題である。体長組成解析における注意点は目合選択性のため体長の小さい部分のデータが欠けている場合、最尤法に基づく Hasselblad 法は適用困難となる点で、このような場合は最小 2 乗法を適用した方がよい（赤嶺, 1985b）。EM アルゴリズムは様々な統計モデルに適用可能なので、今後はさらに広い範囲で活用されるだろう。ただしモデルによっては極端に収束が遅い事例もあるので注意が必要である（宮川, 1987）。

謝辞

混合正規分布のパラメータ推定および EM アルゴリズムの検討に関しては、故加藤史彦氏、故石岡清英氏、および故松宮義晴教授のご協力とご助力によるところが大きかつ

た。また元広島県水産試験場長の猪子嘉生氏、中央水産研究所の西田宏主任研究官には貴重なご助言をいただいた。記して深謝の意を表します。

引用文献

- 相澤 康・滝口直之 (1999) MS-excel を用いたサイズ度数分布から年齢組成を推定する方法の検討。水産海洋研究, **63**, 205-214.
- 赤嶺達郎 (1985a) 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.
- 赤嶺達郎 (1985b) 数値計算プログラムの作り方と使い方の問題点。日本海ブロック試験研究集録, **6**, 77-85.
- Akamine, T. (1987) Comparison of algorithms for estimating parameters of a mixture of normal distributions. Bull. Japan Sea Reg. Fish. Res. Lab., **37**, 259-277.
- 赤嶺達郎 (1997) 成長・生残モデルにおける最適制御理論。中央水研報, **10**, 135-167.
- Akamine, T. (1998) Convergence of Hasselblad method for estimating parameters of a mixture of probability distributions. Fish. Sci., **64**, 997-998.
- 赤嶺達郎 (1999) 混合正規分布のパラメータ推定における Hasselblad 法の収束。中央水研報, **14**, 49-58.
- 赤嶺達郎 (2001) VPA における近似式と反復法の数学的検討。中央水研報, **16**, 1-16.
- 五利江重昭 (2002) MS-excel を用いた混合正規分布のパラメータ推定。水産増殖, **50**, 243-249.
- Hasselblad, V. (1966) Estimation of parameters for a mixture of normal distributions. Technometrics, **8**, 431-444.
- 伊理正夫 (1981) 数値計算。朝倉書店、東京。173 pp.
- 岩崎 学 (2002) 不完全データの統計解析。エコノミスト社、東京、348 pp.
- Kimura, D. K. and S. Chikuni (1987) Mixture of empirical distributions: an iterative application of the age-length key. Biometrics, **43**, 23-35.
- 岸野洋久 (1999) 生のデータを料理する。日本評論社、東京、227 pp.
- 北田修一 (2001) 栽培漁業と統計モデル分析。共立出版、東京、335 pp.
- McLachlan, G. J. and T. Krishnan (1997) The EM algorithm and extensions. John Wiley & Sons, New York, 274 pp.
- McLachlan, G. J. and D. Peel (2000) Finite Mixture Models. John Wiley & Sons, New York, 419 pp.
- Millar, R. B. (1987) Maximum likelihood estimation of mixed stock fishery composition. Can. J. Fish. Aquat. Sci., **44**, 583-590.
- 宮川雅巳 (1987) EM アルゴリズムとその周辺。応用統計学, **16**, 1-12.
- Orchard, T. and M. A. Woodbury (1972) A missing information principle: theory and applications. Proc. 6th Berkeley Symp. on Math. Statist. Prob., **1**, 697-715.
- 坂本慶行・石黒真木夫・北川源四郎 (1983) 情報量統計学。共立出版、東京、236 pp.
- 竹内 啓編 (1989) 統計学事典。東洋経済新報社、東京、1185 pp.
- 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.
- 田中昌一 (1985) 水産資源学総論。恒星社厚生閣、東京、381 pp.
- 東海 正 (1997) MS-excel のソルバーによる曳網の網目選択性 Logistic 式パラメータの最尤推定。水産海洋研究, **61**, 288-298.
- 上田修巧 (2003) EM 法。(ワシ他「計算統計 I 確率計算の新しい手法」)。岩波書店、東京、157-167.
- 渡辺美智子・山口和範編 (2000) EM アルゴリズムと不完全データの諸問題。多賀出版、東京、258 pp.