

混合ベータ分布のテスト得点分布への当てはめ

前川 眞一*

要 約

本研究においては、テストの得点分布を連続量であると見なし、それにベータ分布の混合を当てはめる方法について考察した。また、有限混合分布一般の当てはめに用いられる EM アルゴリズムに関して概説し、数値計算における初期値の選択法に関して提案を行った。

1 はじめに

テスト得点の分布は、基本的には離散分布である。したがって、任意の $p, 0 \leq p \leq 1$ に対して、パーセンタイルが一意に定まるとは限らない。このことは、要約統計量としてパーセンタイルを求める際にはあまり問題にはならないが、等百分位法を用いたテストの等化に際しては問題である。よく知られているように、分布関数 F を持つテスト 1 の得点 X_1 を、分布関数 G を持つテスト 2 の得点 X_2 に等化する場合、等化された得点 $X_{2,1}$ は

$$X_{2,1} = G^{-1}(F(X_1))$$

という形になる。したがって、元来離散分布であるテスト 2 の得点分布から、この分布関数の逆関数を如何に定めるかによって、等化の結果が異なってくるからである。

通常、この問題は、得点分布を何らかの方法で平滑化することによって対処されている場合が多い。得点分布の等化の方法としては、単純に、度数分布の移動平均を取る方法や、核関数を用いる方法、直接的に分布関数を平滑化する方法等、様々なものが考えられるが、その詳細は、例えば Braun, et.al, 1982 や Kolen & Brennan, 1995 にゆずる。

他方、離散分布に、パラメトリックな連続分布

を当てはめることにより、この問題を解決することも出来る。例えば、最も単純な場合、各分布に正規分布を当てはめれば、上記の等百分位法はいわゆる線形等化法に帰着する。ただ、観測されるテスト得点の分布は、明らかに定義域が 0 点と満点の間に限定されているし、また、形状的にも正規分布とは見なせないものも多く、同じパラメトリックな分布でも、より豊かな表現力を持つものを用いた方が望ましい場合が多い。

本稿では、このような視点から、テスト得点の分布に混合ベータ分布を当てはめることを試みる。ベータ分布を選択した理由としては、テスト得点の取る範囲が 0 点から満点までの間に限定されていること、ならびに、その形状がかなり複雑になる場合が多いこと、また、有限な定義域を持つ分布の中では、ベータ分布が豊かな表現力を持つこと等が挙げられる。

次節では、まず、有限混合分布一般に利用可能なパラメタの推定法に関して概説し、その後ベータ分布に特化した部分について述べる。また、初期値の選択方法に関しても説明を加える。そして、最後に、試験得点のデータに本方法を適用した例を示す。

* 大学入試センター研究開発部 試験作成支援研究部門

〒153-8501 目黒区駒場 2-19-23 mayekawa@hum.titech.ac.jp

2002 年 1 月 18 日 受理

2 有限混合分布の当てはめ

2.1 記 法

有限混合分布 (finite mixture distributions) に関しては, Everitt (1981), McLachlan & Basford (1988) や Titterton et al. (1985) 等に詳しいが, 本稿では以下のような記法を用いる.

1. データ

素データは, 一般的には, 総観測値数を N とし,

$$\mathbf{x}_i, i = 1, 2, \dots, N, \quad (1)$$

という形で表現される. しかし, 後の説明の利便性を考え, これを m 個のカテゴリの度数の形で要約した度数分布,

$$(\mathbf{x}_i, c_i), i = 1, 2, \dots, m, \text{ and } \sum_{i=1}^m c_i = N \quad (2)$$

ただし, $\mathbf{x}_i$ は観測された第 i 番目のカテゴリの値, c_i はその度数, を考える. すると, カテゴリの値に重複を許し, 素データの全ての値をもって各カテゴリを定義すれば, 素データの形は

$$c_i = 1, i = 1, 2, \dots, N, \quad (3)$$

と置いた場合の度数分布形式の特殊形となる.

2. 密度関数と分布関数

成分数が q 個の混合分布の密度関数は以下のように表現することが出来る.

$$f(\mathbf{x} | \Xi, \rho) = \sum_{k=1}^q f_k(\mathbf{x} | \xi_k) \rho_k, \quad \sum_{k=1}^q \rho_k = 1, \quad (4)$$

ただし, $f_k(\mathbf{x} | \xi_k), k = 1, 2, \dots, q$ は, 混合されるべき密度関数で, それぞれのパラメタを $\xi_k, k = 1, 2, \dots, q$ とし, それらをまとめて Ξ と書く. また, $\rho_k, k = 1, 2, \dots, q$ は各成分の混合比率であり, これもパラメタとなる. 同様に, 分布関数は, 各成分の分布関数を F_k とすれば,

$$F(\mathbf{x} | \Xi, \rho) = \sum_{k=1}^q F_k(\mathbf{x} | \xi_k) \rho_k \quad (5)$$

与えられる. なお, 通常は f_k として同一分布族 $f_k(\mathbf{x} | \xi_k) = f(\mathbf{x} | \xi_k), k = 1, 2, \dots, q$ を考えることが多い.

なお, 混合分布の平均と分散は, 各成分の平均と分散を μ_k, σ_k^2 とすれば,

$$\mathcal{E}(X) = \sum_{k=1}^q \rho_k \mu_k \quad (6)$$

$$\text{var}(X) = \sum_{k=1}^q \rho_k (\sigma_k^2 + \mu_k^2) - (\mathcal{E}(X))^2 \quad (7)$$

となる.

3. 尤度

データ $(\mathbf{x}_i, c_i), i = 1, 2, \dots, m$ が与えられた時のパラメタの尤度は

$$l = \prod_{i=1}^m f(\mathbf{x}_i | \Xi, \rho)^{c_i} \quad (8)$$

$$= \prod_{i=1}^m \left(\sum_{k=1}^q f_k(\mathbf{x}_i | \xi_k) \rho_k \right)^{c_i} \quad (9)$$

である.

2.2 最尤法によるパラメタの推定

上記の尤度 l を最大とするようなパラメタおよび混合比率の値が最尤解となるが, 混合分布のパラメタの推定の場合, EM アルゴリズム (Dempster, et.al., 1977) を用いた解法が一般的である. なお, EM アルゴリズムでは E ステップで対数完全データ尤度の欠測値に関する期待値を取り, M ステップにおいてその期待対数完全データ尤度を最大にするようなパラメタの値を求める, という過程を繰り返していくことにより, l の最大化を図ることになる.

1. 欠測データ

欠測データとして, 各観測値の第 k 番目の成分への所属の有無を表すインディケータ

$$z_{ik} = \{z_{ik}\}, i = 1, 2, \dots, m, k = 1, 2, \dots, q \quad (10)$$

$$z_{ik} = \begin{cases} 0 & \text{観測値 } i \text{ は} \\ & \text{第 } k \text{ 成分には属さない.} \\ 1 & \text{観測値 } i \text{ は} \\ & \text{第 } k \text{ 成分に属する.} \end{cases},$$

$$\sum_{k=1}^q z_{ik} = 1, i = 1, 2, \dots, m \quad (11)$$

を考える.

2. 完全データ

完全データは

$$(\mathbf{x}_i, \mathbf{z}_i), i = 1, 2, \dots, m \quad (12)$$

であり, その (同時) 密度関数は

$$f(\mathbf{x}_i, \mathbf{z}_i | \Xi, \rho) = \prod_{k=1}^q (f_k(\mathbf{x}_i | \xi_k) \rho_k)^{z_{ik}} \quad (13)$$

である. 従って, 完全データ尤度は

$$L = \prod_{i=1}^m \left(\prod_{k=1}^q (f_k(\mathbf{x}_i | \xi_k) \rho_k)^{z_{ik}} \right)^{c_i}, \quad (14)$$

その対数 (対数完全データ尤度) は

$$\begin{aligned} \ln L &= \sum_{i=1}^m c_i \left(\sum_{k=1}^q (z_{ik} \ln f_k(\mathbf{x}_i | \xi_k) + z_{ik} \ln \rho_k) \right) \\ &= \sum_{k=1}^q \left(\sum_{i=1}^m c_i z_{ik} \ln f_k(\mathbf{x}_i | \xi_k) \right) \\ &\quad + \sum_{k=1}^q \left(\sum_{i=1}^m c_i z_{ik} \ln \rho_k \right) \end{aligned} \quad (15)$$

となる.

3. 欠測データの条件付き分布

$\mathbf{z}_i$ の $\mathbf{x}_i$ を与えたときの条件付き密度関数は

$$f(\mathbf{z}_i | \mathbf{x}_i, \Xi, \rho) = \prod_{k=1}^q (h_{ik})^{z_{ik}} \quad (17)$$

ただし,

$$h_{ik} = \frac{f_k(\mathbf{x}_i | \xi_k) \rho_k}{\sum_{j=1}^q (f_j(\mathbf{x}_i | \xi_j) \rho_j)} \quad (18)$$

である. 従って, z_{ik} の条件付き期待値は h_{ik}

となる.

4. 対数完全データ尤度の期待値

$\mathbf{x}_i$ を与えたときの $\mathbf{z}_i$ の条件付き分布に関する対数完全データ尤度の期待値は

$$\begin{aligned} \mathcal{E} \ln L &= \sum_{k=1}^q \left(\sum_{i=1}^m c_i h_{ik} \ln f_k(\mathbf{x}_i | \xi_k) \right) \\ &\quad + \sum_{k=1}^q \left(\sum_{i=1}^m c_i h_{ik} \ln \rho_k \right) \end{aligned} \quad (19)$$

となる.

5. EM アルゴリズム

従って, 混合分布のパラメタの最尤解を求めるための EM アルゴリズムは以下ようになる.

(a) E step

現在のパラメタの値を Ξ^0, ρ^0 として,

$$\begin{aligned} h_{ik}^0 &= h_{ik}(\Xi^0, \rho^0) \\ &= \frac{f_k(\mathbf{x}_i | \xi_k^0) \rho_k^0}{\sum_{j=1}^q (f_j(\mathbf{x}_i | \xi_j^0) \rho_j^0)} \end{aligned} \quad (20)$$

を計算する.

その値を用いて

$$\begin{aligned} \mathcal{E} \ln L(\Xi, \rho) &= \sum_{k=1}^q \left(\sum_{i=1}^m c_i h_{ik}^0 \ln f_k(\mathbf{x}_i | \xi_k) \right) \\ &\quad + \sum_{k=1}^q \left(\sum_{i=1}^m c_i h_{ik}^0 \ln \rho_k \right) \end{aligned} \quad (21)$$

を計算する.

(b) M step

Ξ^0, ρ^0 から, 従って h_{ik}^0 を既知として $\mathcal{E} \ln L(\Xi, \rho)$ を最大とするようなパラメタ Ξ, ρ の値を求め, それらを次の Ξ^0, ρ^0 とする.

なお, $\mathcal{E} \ln L(\Xi, \rho)$ の形から, 混合比率と各成分のパラメタは独立に取り扱うことが可能であり, 混合比率に関しては,

$$\rho_k = \frac{\sum_{i=1}^m c_i h_{ik}^0}{\sum_{j=1}^q \sum_{i=1}^m c_i h_{ij}^0}, k = 1, 2, \dots, q \quad (22)$$

が $\mathcal{E} \ln L(\Xi, \rho)$ を最大にすることは明かである. また, $\mathcal{E} \ln L(\Xi, \rho)$ を最大にする Ξ に関しては, 各成分のパラメタ ξ_k が互いに無関

係の場合には、各成分ごとに、密度関数 f_k を持つデータが

$$(x_i, c_i h_{ik}^0), i = 1, 2, \dots, m \quad (23)$$

という形の度数分布形式のデータとして与えられたときの最尤解として求めることができる。

例えば、正規分布の混合の場合には、各成分のパラメタを $\xi_k = \{\mu_k, \Sigma_k\}, k = 1, 2, \dots, q$ とすれば、

$$\mu_k = \frac{1}{\sum_{i=1}^m c_i h_{ik}^0} \sum_{i=1}^m c_i h_{ik}^0 x_i \quad (24)$$

$$\Sigma_k = \frac{1}{\sum_{i=1}^m c_i h_{ik}^0} \sum_{i=1}^m c_i h_{ik}^0 x_i x_i' - \mu_k \mu_k' \quad (25)$$

となるが、これは Day (1969) が示したものに他ならない。

2.3 最尤解のための初期値

上記の最尤解を用いてパラメタ Ξ と ρ を推定するためにはその初期値が必要となるが、なるべく最尤解に近いものを用いることが望ましい。ここでは、データとして与えられる相対累積度数分布を、パラメタが既知の分布の分布関数の重み付き和で近似することにより初期値を求める方法について考察する。(前川, 1999b を参照のこと。)なお、以下の方法において、密度関数の代わりに分布関数を用いた理由は、範囲が定まっている分布関数の方が数値的に安定していると思われるためである。相対度数分布と密度関数を用いても同様の方法が利用できるはずである。また、以下の議論では、同一分布族からの混合を仮定するが、そうでない場合も同様に扱うことが出来る。

いま、パラメタ ξ_k が 2 個の成分 $\xi_k = (a_k, b_k)$ (例えば、正規分布の場合は平均と分散、ベータ分布の場合は α と β) で構成されているとして、

$a_r, b_r, r = 1, 2, \dots, R$ を初期値として用いるパラメタの値の候補であるとする。例えば、ベータ分布の場合、 a と b の値として 1.001, 3, 5, 7, 9, 11 の 6 個の値を用いれば、その全ての組み合わせとして $\xi_1^* = (a_1 = 1.001, b_1 = 1.001)$, $\xi_2^* = (a_2 = 1.001, b_2 = 3)$ 等の $R = 36$ の候補が存在する。

これらの分布関数を

$$F_r(x) = \int_{-\infty}^x f(y | \xi_r^*) dy, \quad (26)$$

それをデータとして与えられている点 $x_i, i = 1, 2, \dots, m$, で評価したものを $F_r(x_i)$ とする。また、データとして与えられる相対累積度数分布を

$$G(x_i) = \sum_{j=1}^i c_j, i = 1, 2, \dots, m \quad (27)$$

とする。本節で述べる初期値を求める方法は、

$$G(x_i) = \sum_{r=1}^R F_r(x_i) \beta_r + \varepsilon_i, i = 1, 2, \dots, m \quad (28)$$

という回帰分析モデル (相対累積度数分布が基準変数、混合成分の候補として与えられる分布関数が説明変数群) を考え、この残差の 2 乗和を最小とするような q 個の説明変数群を選択するというものである。そのためには回帰分析における変数選択の技法を用いればよいが、目的が最尤解の算出のための初期値の導出であるため、完全に最小の RSS を与えるサブセットを求める必要はない。ここでは、以下に示す maxr2 法 (SAS Institute, Inc., 1985, p765) を多変量に拡張したものが適当であると考えられる。

maxr2 法には変数増加法 (前進選択型) と変数減少法 (後進選択型) の 2 者が考えられるが、変数増加法の概略は以下のようなものである。IN を $F_1, F_2, \dots, F_R$ という説明変数のうち、現在モデルに入っている $r < q$ 個の変数のリスト、OUT をそれ以外の $R - r$ 個の変数のリストとする。IN の初期値は空であり、OUT の初期値は $\{1, 2, \dots, R\}$ である。まず、OUT の中から、RSS を最も減少させるような変数をひとつ選び IN の要素に追加する。次に、現在の IN の要素のそれぞれに関して、OUT の要素のひとつと交換することにより RSS が減少するか否かを調べ、もしも減少するのであれば、交換する。この手続きを繰り返していき、全ての比較を尽くした時点で $r + 1$ の変数群を決定する。

この方法で選ばれる変数は IN と OUT の交換の順序に依存するが、all-possible-regression と比較して計算回数が格段に少なく、最適解に近いものを与えてくれることが知られている。

変数増加法による maxr2 法

IN is the list of the variables currently included in the regression.
OUT is the list of the variables currently not included in the regression.

nextin is the function which enter the best variable in OUT given IN.
replacer is the function which swaps the one in IN and the one in OUT.
q is the total number of variables to be selected.

```

free IN;      * list of variables currently in the model;
OUT=1:nx;     * list of variables currently not in the model;

do i=1 to q;

    * find the one which best reduces rss;
    run nextin( bestid1, Rss, C, OUT, locy, eps, 0<>(print-2) );
    print i "Variable" bestid1 "entered." Rss;
    if bestid1 ^= 0 then do;
        IN=IN||bestid1;
        * for each variable already entered,
          find a candidate not entered which replaces it;
        * This is order dependent.;

        do j=1 to ncol(IN);

            INj=IN[j];
            run replaceR( bestid2, bestRss, C, INj, OUT, locy, eps, 0<>(print-3) );
            * print i j "replace" bestid2 bestrss;
            if bestid2 ^= 0 & bestid2 ^= INj then do;
                * swap INj and bestid2: take INj out, put bestid2 in;
                if ncol(IN) > 0 then IN[loc(IN=INj)]=bestid2;
                OUT[loc(OUT=bestid2)]=INj;
                C=sweep(C, Inj||bestid2);
                Rss=sum(vecdiag(C[locy,locy]));
                if print >= 3 then
                    print "Variable" INj "replaced by variable" bestid2 Rss;
            end;
        end;
        if print >= 2 then do;
            print "Best Model Found:" i Rss;
            print IN;
            Vn=shape(varnames[IN],1);
            print Vn;
        end;
        rssall=rssall//Rss;
        * sort IN for display purpose;
        rank=rank(IN);
        dum=IN; dum[rank]=IN';
        INall[i,1:ncol(dum)]=dum;
    end;
    else goto done;

end; * of q loop;

done: *;
```

この方法を混合分布のパラメタの初期値に利用する場合、混合比率パラメタ ρ の初期値は回帰係数 β として与えられるが、この方法ではそれが必ずしも正で、和が 1 になるとは限らない。したがって何らかの工夫をする必要がある。ひとつは、上記のアルゴリズムの中で RSS を評価する際に、この制約を入れることが考えられる。これは、

$$\beta_j = \frac{\gamma_j^2}{\sum_k \gamma_k^2} \quad (29)$$

という条件の下で回帰を行えばよい。より単純には、混合比率、すなわち回帰係数は一定という制約の元に変数選択を行い、最後に制約条件を満たす混合比率を計算するという方法があるが、今回はこの方法を用いた。

3 混合ベータ分布の場合

3.1 テスト得点の範囲の変換

通常のテスト得点は、素点の範囲が区間 $[min, max]$ に限定されている。例えば、100 点満点のテストの場合、 $[0, 100]$ である。このような形をしたデータに混合ベータ分布を当てはめるためには、何らかの形でテスト得点の範囲を区間 $[0, 1]$ に変換しなければならない。そのための方法としては、例えば、以下のような方法がある。

各テスト得点 x は 1 の幅を持つ連続区間 $(x - 1/2, x + 1/2)$ の代表値であると考ええる。すなわち、 min は区間 $(min - 1/2, min + 1/2)$ を代表する値、 max は区間 $(max - 1/2, max + 1/2)$ を代表する値となる。従って、素点の範囲の長さは

$$R = max + 1/2 - (min - 1/2) = max - min + 1 \quad (30)$$

となるため、それを区間 $(0, 1)$ に入れるためには

$$y = (x - min + 1/2) / R \quad (31)$$

という変換を施せばよい。例えば、4 点満点のテスト ($min = 0, max = 4$) の得点は

x	0	1	2	3	4
y	0.1	0.3	0.5	0.7	0.9

と変換される。(100 点満点のテストの場合は末尾

の図表を参照されたい。) また、この逆変換は

$$x = Ry + min - 1/2 \quad (32)$$

となる。

3.2 最尤解

混合ベータ分布のパラメタは、 $\xi_k = \{a_k, b_k, \rho_k\}$, $k = 1, 2, \dots, q$ であるが、ここに、 a_k, b_k は第 k 成分を成すベータ分布のパラメタである。

前節に示したように、EM アルゴリズムを用いて最尤解を求める場合には、パラメタ間に制約がなければ、各成分ごとに独立に M ステップを行うことが出来る。したがって、本節では、度数分布の形でデータが与えられた場合の(単純な)ベータ分布の最尤解を求める方法を示すが、実際の混合分布の計算では、M ステップにおいて、以下の a, b を a_k, b_k に、また、 $d_i, i = 1, 2, \dots, m$ を (23) 式に示す $c_i h_{ik}^0, i = 1, 2, \dots, m, k = 1, 2, \dots, q$ に置き換えて q 回繰り返す必要がある。

1. モデル

$$x_i \sim Be(a, b), \text{ iid}, i = 1, 2, \dots, m \quad (33)$$

密度関数

$$f(x_i | a, b) = \frac{1}{B(a, b)} x_i^{a-1} (1 - x_i)^{b-1} \quad (34)$$

ただし、 $\Gamma(a)$ をガンマ関数として、

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad (35)$$

2. 尤度

データ

$$(x_i, d_i), i = 1, 2, \dots, m, \sum_{i=1}^m d_i = n \quad (36)$$

が与えられたときの対数尤度は

$$\ln L = \sum_{i=1}^m d_i ((a-1) \ln x_i + (b-1) \ln (1-x_i) - \ln B(a, b)) \quad (37)$$

3. 最尤方程式

以下の微係数を用いて Newton-Raphson 法等

を用いて解を求める.

1 次微分

$$\begin{aligned} \frac{\partial \ln L}{\partial a} &= \sum_{i=1}^m (d_i \ln x_i) - n(\psi(a) - \psi(a+b)) \end{aligned} \quad (38)$$

$$\begin{aligned} \frac{\partial \ln L}{\partial b} &= \sum_{i=1}^m (d_i \ln (1 - x_i)) - n(\psi(b) - \psi(a+b)) \end{aligned} \quad (39)$$

ただし,

$$\psi(a) = \frac{d \ln \Gamma(a)}{da} \quad (40)$$

2 次微分

$$\frac{\partial^2 \ln L}{\partial a^2} = -n(\psi'(a) - \psi'(a+b)) \quad (41)$$

$$\frac{\partial^2 \ln L}{\partial b^2} = -n(\psi'(b) - \psi'(a+b)) \quad (42)$$

$$\frac{\partial^2 \ln L}{\partial a \partial b} = n\psi'(a+b) \quad (43)$$

ただし,

$$\psi'(a) = \frac{d\psi(a)}{da} \quad (44)$$

4. 数値計算

上記の最尤方程式の解は解析的に求めることが出来ない. したがって, M ステップにおいても繰り返し計算が必要となる. しかし, EM アルゴリズム全体の収束という観点からは, その繰り返しの回数は少数で十分であることが知られている.

ψ 関数の計算には a が大きい場合には Abramowitz & Stegun (1972) の 6.3.18 を, a が小さい場合には 6.3.5 の漸化式を利用すればよい. また, ψ' 関数の計算は上式を微分したものを利用すればよい.

次節に数値例を示すが, 計算に用いたプログラムにおいては Newton-Raphson 法と Marquardt の方法を併用している. また, 繰り返し計算を加速するため, Ramsay (1975) の方法 (Mayekawa, 1995 を参照) を用いている. なお, 計算機プログラム (SAS/IML) は著者より入手可能である.

4 試験得点データへの適用例

一般に, 試験得点は, 試験に含まれる項目の正誤反応に配点と呼ばれる重みを掛け合わせて計算されるため, その分布は 0 点から満点までの有限区間における離散分布となる. 従って, 受験生の反応には依存しないシステマティックな欠測の部分が存在する 경우가多い. たとえば, 20 項目からなる 100 点満点のテスト得点の場合, すべての配点が等しく 5 点であれば, テスト得点は 0, 5, 10, ... 95, 100 という値しかとらない. また, 配点が異なる場合でも, 最低の配点がたとえば 2 点であれば, 1 点や 99 点という得点は存在しない.

また, 経験的に, 試験得点の分布は様々な形 (平均点付近で対称, 裾が長い, 双峰等) をとることが知られている. 以下の例では, 試験得点の分布を代表するようなデータをいくつか選び, それに本稿で提案したベータ分布の混合を当てはめた例を示す.

分布 1 は, 双峰で尖度の小さなものであり, 分布 2 は, 対称性が大きく正規分布に近いものである. また, 分布 3 は, 高得点部分に度数が多く, 下方に長く裾を引く分布である. なお, 分布 3 に関しては, その高得点部分において 3 点刻みで度数が多くなっているが, これは, 大部分の配点が 3 点であることに由来している.

各分布に 3 成分のベータ分布を当てはめた例を本稿の最後に図表として示す. ただし, 分布 1 は, その計算に用いたコンピュータの出力をすべて掲載しているが, 残りの二つの分布に関しては, データと当てはめられた密度関数のプロットのみを示してある. また, 各分布ごとに, 推定されたパラメータを表 1 に, 素点と当てはめられた混合ベータ分布の平均と標準偏差を表 2 に示した.

表 2 とグラフで見る限り, 当てはまりは 3 成分でかなりよく, 平均点の差も 1 点未満に収まっている. 特に, 分布 1 に関してはその双峰分布としての特徴がよく表され, また, 分布 3 の高得点部分では, システマティックな欠測の部分をうまく補間するような形で混合ベータ分布が当てはめられているのがわかる. なお, 単純な形状をしている分布 2 に関しては, 2 成分で十分であることがわかる.

表 1 混合ベータ分布のパラメタの推定値

分布 1	成分 1	成分 2	成分 3
A	3.1116	4.4537	8.3436
B	1.0089	2.1916	13.9391
RHO	0.1307	0.6335	0.2358
分布 2	成分 1	成分 2	成分 3
A	2.2650	5.7506	9.4810
B	1.1467	3.7230	11.6927
RHO	0.0118	0.6729	0.3153
分布 3	成分 1	成分 2	成分 3
A	3.7802	25.0123	8.3308
B	1.0010	3.6146	3.3576
RHO	0.2553	0.2042	0.5405

表 2 平均、標準偏差

分布 1	平均	標準偏差
混合ベータ分布	0.6115	0.2086
100 点満点に換算	61.7685	21.0742
観測値	61.2631	21.0809
分布 2	平均	標準偏差
混合ベータ分布	0.5574	0.1581
100 点満点に換算	56.3045	15.9752
観測値	55.8005	15.9647
分布 3	平均	標準偏差
混合ベータ分布	0.7655	0.1443
100 点満点に換算	77.3157	14.5838
観測値	76.8490	14.5544

5 終わりに

本稿では、テスト得点の分布に混合ベータ分布を当てはめることを試みた結果、前節に示すように、混合ベータ分布は、成分の数が 3 つ程度で十分に豊かな表現力を持ち、実際のテスト得点の分布によく当てはまることを示した。したがって、テストの得点分布のパラメトリックな表現が望ましい場合には、本方法はきわめて有用であると考えられる。また、前節に示したような、欠損度数を多く持つテスト得点の分布を、視覚的に要約するためにも、有用であると思われる。なお、分布の視覚的要約という観点からは、パラメトリックとノンパラメトリックの中間的方法として、得点分布に制約を課した多項分布を当てはめる方法 (Mayekawa, 1996) を利用することも考えられるが、その場合

は、多項分布の確率がテスト得点のスプライン関数として表現出来ることとなる。ただ、この場合、得られる分布はあくまで多項分布であり、その取り扱いが、混合分布に比して容易ではない。

また、本方法とその他の平滑化の方法との比較に関しては、今後、研究の余地があると思われる。特に、 N が小さい場合の頑強性に関しては、興味深い。

最後に、実際に等百分位法を行う場合には、混合ベータ分布の分布関数の逆関数を求める必要が生じる。このためには、(5) 式 で与えられる分布関数を用いて、方程式

$$F(x) = p \quad (45)$$

を数値的に解くことも考えられるが、分布関数の線形逆補間を行うので十分であろう。

参考文献

- Abramowitz, M. & Stegun, I.A. (1972) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, Inc., New York.
- Braun, H.I. & Holland, P.W. (1982) Observed-score test equating: A mathematical analysis of some ETS equating procedures. In Holland, P.W. & Rubin, D.B. Eds. *Test Equating*. Academic Press, Inc. New York.
- Everitt, B.S. (1981) *Finite mixture distributions*. London: Chapman and Hall.
- Day, N.E. (1969) Estimating the Components of a Mixture Normal Distributions. *Biometrika* vol.56, 3, 463-474.
- Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm (with Discussion), *Journal of the Royal Statistical Society Series B*, **39**, 1-18.
- Kolen M.J. & Brennan, R.L. (1995) *Test Equating. Methods and Practices*. New York, N.Y.: Springer-Verlag New York, Inc.
- McLachlan, G.J., & Basford, K.E. (1988) *Mixture models: inference and applications to clustering*. New York, NY: Marcel Dekker, Inc.
- Mayekawa, Shin-ichi. (1996) Maximum likelihood estimation of the cell probabilities under linear constraints. *Behaviormetrika*, Vol.23, No.1,

- 111–128
- Mayekawa, Shin-ichi. (1995) An efficient algorithm for feed-forward neural network regression analysis. *Behaviormetrika*. Vol.22, No.1, 67–90
- Ramsay, J.O. (1975) Solving implicit equations in psychometric data analysis. *Psychometrika*, **40**, 337–360.
- SAS Institute Inc. (1985) SAS User's Guide: Statistics, Version 5 Edition. Cary, NC: SAS Institute Inc.
- Titterington, D.Dm, Smith, A.F.M., & Makov, U.E. (1985) *Statistical analysis of finite mixture distributions*. Chichester: John Wiley & Sons Ltd.
- 前川真一 (1999b) ニューラルネットワークを用いた回帰分析の初期値の導出 日本行動計量学会 第27 回大会発表

図表（プログラムの出力）

ML Estimation of Mixture Beta Parameters

TITLE

Original: 10B2

RANGE		SMALLO1	N	R	NROWX	NPARAM	PRINT	ABMIN	ABMAX
0	100	0.00001	38450	3	101	8	2	1.001	30

INIT	SEED	LMAX	EPS	EPSG	RELAX	RELAXR	EPSNEG	RESTORE	FIRSTR
1	92384	1000	1E-8	0.0001	10	1	0.1	2	5

OUTH

B2H

	INIT	
Initial Estimates	1	(初期値)

INIT	LMLHMBP	ABR	Component1	Component2	Component3
1	8005.1623	A	2.000	3.000	4.000
		B	1.001	2.000	3.000
		RHO	0.309	0.444	0.247

(繰り返し	尤度	改善分	改善率	変化量		Marquardt			Relax)
LLLL	LMLHMB	LMLHIMP	LMLHIMPR	MAXADAB	MAXADR	LAMBDA1			_AREL_
1	8111.934	106.77162	0.0131623	0.3174825	0.0038174	1E-7	1E-6	1E-6	0
LLLL	LMLHMB	LMLHIMP	LMLHIMPR	MAXADAB	MAXADR	LAMBDA1			_AREL_
2	8143.8731	31.939178	0.0039219	0.1792075	0.0002755	1E-10	1E-8	1E-9	0
LLLL	LMLHMB	LMLHIMP	LMLHIMPR	MAXADAB	MAXADR	LAMBDA1			_AREL_
3	8160.9168	17.043674	0.0020885	0.0920893	0.0004511	1E-11	1E-10	1E-11	0
LLLL	LMLHMB	LMLHIMP	LMLHIMPR	MAXADAB	MAXADR	LAMBDA1			_AREL_
4	8172.8735	11.956709	0.001463	0.0690327	0.0006565	1E-11	1E-11	1E-11	0
LLLL	LMLHMB	LMLHIMP	LMLHIMPR	MAXADAB	MAXADR	LAMBDA1			_AREL_
5	8183.1136	10.240056	0.0012514	0.0555776	0.0007777	1E-11	1E-11	1E-11	0
LLLL	LMLHMB	LMLHIMP	LMLHIMPR	MAXADAB	MAXADR	LAMBDA1			_AREL_
6	8192.9052	9.7916626	0.0011951	0.0621665	0.0008621	1E-11	1E-11	1E-11	0
LLLL	LMLHMB	LMLHIMP	LMLHIMPR	MAXADAB	MAXADR	LAMBDA1			_AREL_
7	8202.816	9.9107374	0.0012082	0.0740204	0.0009297	1E-11	1E-11	1E-11	-3.411693

途中省略

LLLL	LMLHMB	LMLHIMP	LMLHIMPR	MAXADAB	MAXADR	LAMBDA1				_AREL_
464	8873.1814	0.003689	4.1575E-7	0.0018653	0.000093	1E-11	1E-11	1E-11	-0.00007	
465	8873.1817	0.0003046	3.4332E-8	0.0011598	0.0000175	1E-11	1E-11	1E-11	-0.00007	
466	8873.1818	0.00013	1.4648E-8	0.0006935	0.0000134	1E-11	1E-11	1E-11	-0.00007	
467	8873.1819	0.0000825	9.3015E-9	0.0004021	0.0000129	1E-11	1E-11	1E-11	-1.409691	

TITLE

Original:Beta2

Iteration Terminated.

LLLL	LMLHMB	LMLHIMP	LMLHIMPR	MAXADAB	MAXADR
467	8873.1819	0.0000825	9.3015E-9	0.0004021	0.0000129

INIT	LMAX	EPS	EPSG	RELAX	RELAXR	EPSNEG	RESTORE	FIRSTR
1	1000	1E-8	0.0001	10	1	0.1	2	5

N	R	NPARAM	LMLHMB	AIC
38450	3	8	8873.1819	-17730.36

ABMIN	ABMAX
1.001	30

(パラメタの推定値)

Estimated Parameters

ABR Component1 Component2 Component3

A	3.1116	4.4537	8.3436
B	1.0089	2.1916	13.9391
RHO	0.1307	0.6335	0.2358

(ベータ分布の統計量)

MU	SIGMA2	SIGMA
0.6115695	0.0435374	0.2086563

(範囲を変換した場合の統計量)

MUX	SIGMA2X	SIGMAX
61.768524	444.12538	21.074282

(観測された統計量)

XBAR	S2	S
61.263121	444.4057	21.080932

Original:Beta2

Estimated Density and Distribution Functions

(得点 XORG	中点 X9	度数 FREQ	相対度数 EPDF	正規化密度 PDF1	確率密度 PDF	相対累積度数 ECDF	分布関数) CDF
0	0.0050	1	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0149	0	0.0000	0.0000	0.0001	0.0000	0.0000
2	0.0248	0	0.0000	0.0000	0.0002	0.0000	0.0000
3	0.0347	0	0.0000	0.0000	0.0005	0.0000	0.0000
4	0.0446	0	0.0000	0.0000	0.0010	0.0000	0.0000
5	0.0545	1	0.0000	0.0000	0.0018	0.0001	0.0000
6	0.0644	1	0.0000	0.0000	0.0031	0.0001	0.0001
7	0.0743	2	0.0001	0.0000	0.0050	0.0001	0.0001
8	0.0842	6	0.0002	0.0001	0.0079	0.0003	0.0002
9	0.0941	0	0.0000	0.0001	0.0121	0.0003	0.0002
10	0.1040	4	0.0001	0.0002	0.0182	0.0004	0.0004
11	0.1139	5	0.0001	0.0003	0.0268	0.0005	0.0006
12	0.1238	13	0.0003	0.0004	0.0383	0.0009	0.0009
13	0.1337	16	0.0004	0.0005	0.0537	0.0013	0.0014
14	0.1436	21	0.0005	0.0007	0.0733	0.0018	0.0020
15	0.1535	29	0.0008	0.0010	0.0979	0.0026	0.0029
16	0.1634	49	0.0013	0.0013	0.1280	0.0038	0.0040
17	0.1733	55	0.0014	0.0016	0.1638	0.0053	0.0054
18	0.1832	90	0.0023	0.0020	0.2057	0.0076	0.0072
19	0.1931	89	0.0023	0.0025	0.2537	0.0099	0.0095
20	0.2030	124	0.0032	0.0030	0.3076	0.0132	0.0123
21	0.2129	142	0.0037	0.0036	0.3670	0.0169	0.0156
22	0.2228	174	0.0045	0.0043	0.4313	0.0214	0.0196
23	0.2327	191	0.0050	0.0049	0.4999	0.0263	0.0242
24	0.2426	247	0.0064	0.0057	0.5717	0.0328	0.0295
25	0.2525	239	0.0062	0.0064	0.6459	0.0390	0.0355
26	0.2624	273	0.0071	0.0071	0.7212	0.0461	0.0423
27	0.2723	302	0.0079	0.0079	0.7965	0.0539	0.0498
28	0.2822	352	0.0092	0.0086	0.8706	0.0631	0.0580
29	0.2921	376	0.0098	0.0093	0.9423	0.0729	0.0670
30	0.3020	359	0.0093	0.0100	1.0107	0.0822	0.0767
31	0.3119	410	0.0107	0.0106	1.0747	0.0929	0.0870
32	0.3218	469	0.0122	0.0112	1.1334	0.1051	0.0979
33	0.3317	424	0.0110	0.0117	1.1863	0.1161	0.1094
34	0.3416	500	0.0130	0.0122	1.2328	0.1291	0.1214
35	0.3515	470	0.0122	0.0126	1.2727	0.1413	0.1338
36	0.3614	516	0.0134	0.0129	1.3057	0.1547	0.1466
37	0.3713	498	0.0130	0.0132	1.3319	0.1677	0.1597
38	0.3812	513	0.0133	0.0134	1.3516	0.1810	0.1729
39	0.3911	505	0.0131	0.0135	1.3652	0.1942	0.1864
40	0.4010	497	0.0129	0.0136	1.3731	0.2071	0.2000
41	0.4109	498	0.0130	0.0136	1.3761	0.2201	0.2136
42	0.4208	517	0.0134	0.0136	1.3747	0.2335	0.2272
43	0.4307	480	0.0125	0.0136	1.3699	0.2460	0.2408
44	0.4406	496	0.0129	0.0135	1.3623	0.2589	0.2543
45	0.4505	518	0.0135	0.0134	1.3529	0.2724	0.2678
46	0.4604	504	0.0131	0.0133	1.3425	0.2855	0.2811

47	0.4703	503	0.0131	0.0132	1.3317	0.2985	0.2943
48	0.4802	522	0.0136	0.0131	1.3213	0.3121	0.3075
49	0.4901	533	0.0139	0.0130	1.3119	0.3260	0.3205
50	0.5000	474	0.0123	0.0129	1.3042	0.3383	0.3335
51	0.5099	509	0.0132	0.0129	1.2984	0.3515	0.3463
52	0.5198	524	0.0136	0.0128	1.2951	0.3652	0.3592
53	0.5297	486	0.0126	0.0128	1.2943	0.3778	0.3720
54	0.5396	504	0.0131	0.0128	1.2964	0.3909	0.3848
55	0.5495	553	0.0144	0.0129	1.3013	0.4053	0.3977
56	0.5594	451	0.0117	0.0130	1.3091	0.4170	0.4106
57	0.5693	500	0.0130	0.0131	1.3196	0.4300	0.4236
58	0.5792	562	0.0146	0.0132	1.3326	0.4447	0.4367
59	0.5891	531	0.0138	0.0133	1.3480	0.4585	0.4500
60	0.5990	534	0.0139	0.0135	1.3654	0.4724	0.4634
61	0.6089	520	0.0135	0.0137	1.3845	0.4859	0.4770
62	0.6188	553	0.0144	0.0139	1.4050	0.5003	0.4909
63	0.6287	565	0.0147	0.0141	1.4264	0.5150	0.5049
64	0.6386	613	0.0159	0.0143	1.4485	0.5309	0.5191
65	0.6485	525	0.0137	0.0146	1.4708	0.5446	0.5336
66	0.6584	559	0.0145	0.0148	1.4929	0.5591	0.5482
67	0.6683	576	0.0150	0.0150	1.5145	0.5741	0.5631
68	0.6782	560	0.0146	0.0152	1.5353	0.5886	0.5782
69	0.6881	570	0.0148	0.0154	1.5548	0.6035	0.5935
70	0.6980	605	0.0157	0.0156	1.5727	0.6192	0.6090
71	0.7079	572	0.0149	0.0157	1.5887	0.6341	0.6246
72	0.7178	582	0.0151	0.0159	1.6026	0.6492	0.6404
73	0.7277	613	0.0159	0.0160	1.6141	0.6651	0.6564
74	0.7376	614	0.0160	0.0161	1.6228	0.6811	0.6724
75	0.7475	628	0.0163	0.0161	1.6286	0.6975	0.6885
76	0.7574	619	0.0161	0.0161	1.6312	0.7136	0.7046
77	0.7673	582	0.0151	0.0161	1.6304	0.7287	0.7208
78	0.7772	578	0.0150	0.0161	1.6261	0.7437	0.7369
79	0.7871	601	0.0156	0.0160	1.6179	0.7593	0.7530
80	0.7970	552	0.0144	0.0159	1.6058	0.7737	0.7689
81	0.8069	591	0.0154	0.0157	1.5896	0.7891	0.7848
82	0.8168	583	0.0152	0.0155	1.5690	0.8042	0.8004
83	0.8267	626	0.0163	0.0153	1.5440	0.8205	0.8158
84	0.8366	622	0.0162	0.0150	1.5145	0.8367	0.8310
85	0.8465	534	0.0139	0.0147	1.4802	0.8506	0.8458
86	0.8564	576	0.0150	0.0143	1.4411	0.8656	0.8603
87	0.8663	536	0.0139	0.0138	1.3971	0.8795	0.8743
88	0.8762	530	0.0138	0.0133	1.3481	0.8933	0.8879
89	0.8861	457	0.0119	0.0128	1.2941	0.9052	0.9010
90	0.8960	472	0.0123	0.0122	1.2350	0.9175	0.9135
91	0.9059	489	0.0127	0.0116	1.1709	0.9302	0.9254
92	0.9158	440	0.0114	0.0109	1.1019	0.9416	0.9367
93	0.9257	491	0.0128	0.0102	1.0280	0.9544	0.9472
94	0.9356	303	0.0079	0.0094	0.9496	0.9623	0.9570
95	0.9455	373	0.0097	0.0086	0.8670	0.9720	0.9660
96	0.9554	279	0.0073	0.0077	0.7808	0.9792	0.9742
97	0.9653	215	0.0056	0.0068	0.6916	0.9848	0.9815
98	0.9752	331	0.0086	0.0059	0.6008	0.9934	0.9879
99	0.9851	0	0.0000	0.0051	0.5106	0.9934	0.9934
100	0.9950	253	0.0066	0.0042	0.4252	1.0000	0.9980

分布 1 (観測された得点分布 * と当てはめられた混合ベータ分布 + ならびに混合分布の各成分)

TITLE	R	1	2	3	N	LMLHMB	AIC
Original:Beta2	3	A 3.112	4.454	8.344	38450	8873.18191370	-17730.36382739
		B 1.009	2.192	13.939			
		R 0.131	0.634	0.236			
Observed(*) and Expected(+) Frequency Distributions					Weighted Component Densities		
0	1				0.0		0.0
1	0				0.0	1	0.0
2	0				0.1	2	0.0
3	0				0.2	3	0.2
4	0				0.4	4	1.4
5	1				0.7	5	5.2
6	1				1.2	6	15.4
7	2				1.9	7	38.4
8	6				3.0	8	83.9
9	0				4.6	9	164.9
10	4				6.9	10	298.3
11	5				10.2	11	503.9
12	13 +				14.6	12	803.7
13	16 +				20.4	13 3	1221.0
14	21 **				27.9	14 3	1778.3
15	29 ***				37.3	15 3	2496.7
16	49 ****				48.7	16 2 3	3393.5
17	55 ****				62.4	17 2 3	4481.4
18	90 *****				78.3	18 2 3	5766.8
19	89 *****				96.6	19 2 3	7249.3
20	124 *****				117.1	20 2 3	8921.1
21	142 *****				139.7	21 2 3	10766.1
22	174 *****				164.2	22 2 3	12762.1
23	191 *****				190.3	23 2 3	14879.1
24	247 *****				217.7	24 2 3	17080.1
25	239 *****				245.9	25 2 3	19326.1
26	273 *****				274.5	26 2 3	21572.1
27	302 *****				303.2	27 1 2 3	23774.1
28	352 *****				331.4	28 1 2 3	25884.1
29	376 *****				358.7	29 1 2 3	27859.1
30	359 *****				384.8	30 1 2 3	29658.1
31	410 *****				409.1	31 1 2 3	31243.1
32	469 *****				431.5	32 1 2 3	32582.1
33	424 *****				451.6	33 1 2 3	33650.1
34	500 *****				469.3	34 1 2 3	34429.1
35	470 *****				484.5	35 1 2 3	34907.1
36	516 *****				497.0	36 1 2 3	35079.1
37	498 *****				507.0	37 1 2 3	34950.1
38	513 *****				514.5	38 1 2 3	34529.1
39	505 *****				519.7	39 1 2 3	33832.1
40	497 *****				522.7	40 1 2 3	32880.1
41	498 *****				523.8	41 1 2 3	31701.1
42	517 *****				523.3	42 1 2 3	30322.1
43	480 *****				521.5	43 1 2 3	28777.1
44	496 *****				518.6	44 1 2 3	27098.1
45	518 *****				515.0	45 1 2 3	25320.1
46	504 *****				511.0	46 1 2 3	23476.1
47	503 *****				506.9	47 1 2 3	21599.1
48	522 *****				503.0	48 1 2 3	19717.1
49	533 *****				499.4	49 1 2 3	17859.1
50	474 *****				496.5	50 1 2 3	16049.1
51	509 *****				494.3	51 1 2 3	14309.1
52	524 *****				493.0	52 1 2 3	12655.1
53	486 *****				492.7	53 1 2 3	11101.1
54	504 *****				493.5	54 1 2 3	9657.2
55	553 *****				496.4	55 1 2 3	8330.6
56	451 *****				498.3	56 1 2 3	7124.4
57	500 *****				502.3	57 1 2 3	6039.2
58	562 *****				507.3	58 1 2 3	5073.2
59	531 *****				513.1	59 1 2 3	4222.2
60	534 *****				519.8	60 1 2 3	3480.4
61	520 *****				527.0	61 1 2 3	2840.7
62	553 *****				534.8	62 1 2 3	2294.9
63	565 *****				543.0	63 1 2 3	1854.4
64	613 *****				551.4	64 1 2 3	1450.3
65	525 *****				559.9	65 1 2 3	1133.5
66	559 *****				568.3	66 1 2 3	875.3
67	576 *****				576.6	67 1 2 3	667.5
68	560 *****				584.4	68 1 2 3	502.4
69	570 *****				591.9	69 1 2 3	373.0
70	605 *****				598.7	70 1 2 3	272.9
71	572 *****				604.8	71 1 2 3	196.6
72	582 *****				610.1	72 1 2 3	139.3
73	613 *****				614.4	73 1 2 3	97.1
74	614 *****				617.8	74 1 2 3	66.4
75	628 *****				620.0	75 1 2 3	44.5
76	619 *****				621.0	76 1 2 3	29.2
77	582 *****				620.7	77 1 2 3	18.7
78	578 *****				619.0	78 1 2 3	11.7
79	601 *****				615.9	79 1 2 3	7.1
80	552 *****				611.3	80 1 2 3	4.2
81	591 *****				605.1	81 1 2 3	2.4
82	583 *****				597.3	82 1 2 3	1.3
83	626 *****				587.8	83 1 2 3	0.7
84	622 *****				576.5	84 1 2 3	0.4
85	534 *****				563.5	85 1 2 3	0.2
86	576 *****				548.6	86 1 2 3	0.1
87	536 *****				531.8	87 1 2 3	0.0
88	530 *****				513.2	88 1 2 3	0.0
89	457 *****				492.6	89 1 2 3	0.0
90	472 *****				470.1	90 1 2 3	0.0
91	489 *****				445.7	91 1 2 3	0.0
92	440 *****				419.5	92 1 2 3	0.0
93	491 *****				391.4	93 1 2 3	0.0
94	303 *****				361.5	94 1 2 3	0.0
95	373 *****				330.1	95 1 2 3	0.0
96	279 *****				297.2	96 1 2 3	0.0
97	215 *****				263.3	97 1 2 3	0.0
98	331 *****				228.7	98 1 2 3	0.0
99	0 *****				194.4	99 1 2 3	0.0
100	253 *****				161.9	100 1 2 3	0.0
DOMAIN FREQ1 ONE CHAR = 12 , ? = OUT OF RANGE, " = MULTI					FREQ2	DOMAIN FREQ1 ONE CHAR = 1E3 , ? = OUT OF RANGE, " = MULTI	FREQ2

分布 2 (観測された得点分布 * と当てはめられた混合ベータ分布 + ならびに混合分布の各成分)

TITLE	R	1	2	3	N	LMLHMB	AIC
Original:Beta4	3	A 2.265	5.751	9.481	58953	26052.56665555	-52089.13331109
		B 1.147	3.723	11.693			
		R 0.012	0.673	0.315			
Observed(+) and Expected(+) Frequency Distributions					Weighted Component Densities		
0	1			0.0	0		0.0
1	0			0.1	1		0.0
2	0			0.2	2		0.0
3	0			0.3	3		0.0
4	0			0.4	4		0.1
5	2			0.6	5		0.4
6	1			0.9	6		1.4
7	0			1.2	7		4.1
8	0			1.7	8		10.6
9	1			2.5	9		24.1
10	4			3.7	10		50.2
11	6			5.3	11		96.4
12	5			7.5	12		173.3
13	11			10.7	13		294.8
14	19			14.9	14		477.9
15	8			20.4	15		743.0
16	30 +			27.7	16		1113.4
17	39 +			36.9	17	2	1614.7
18	52 +*			48.5	18	"	2274.1
19	68 +*			62.7	19	"	3119.5
20	66 +**			80.0	20	"	4177.9
21	124 +**			100.7	21	23	5474.3
22	132 +**			125.0	22	23	7030.7
23	145 +**			153.1	23	23	8863.9
24	183 +**			185.3	24	2 3	10985.
25	191 +**			221.6	25	2 3	13398.
26	272 +**			262.0	26	2 3	16100.
27	293 +**			306.5	27	2 3	19076.
28	340 +**			354.7	28	2 3	22307.
29	409 +**			406.6	29	2 3	25761.
30	457 +**			461.6	30	2 3	29399.
31	529 +**			519.3	31	2 3	33176.
32	578 +**			579.2	32	2 3	37037.
33	661 +**			640.6	33	2 3	40922.
34	688 +**			702.9	34	2 3	44770.
35	793 +**			765.4	35	2 3	48514.
36	833 +**			827.3	36	2 3	52089.
37	864 +**			889.1	37	2 3	55429.
38	941 +**			946.8	38	2 3	58473.
39	1042 +**			1003.0	39	2 3	61165.
40	1069 +**			1056.0	40	2 3	63454.
41	1131 +**			1105.3	41	2 3	65298.
42	1117 +**			1150.3	42	2 3	66664.
43	1197 +**			1190.8	43	2 3	67528.
44	1283 +**			1226.4	44	2 3	67877.
45	1291 +**			1257.1	45	2 3	67710.
46	1279 +**			1282.6	46	2 3	67033.
47	1293 +**			1303.0	47	23	65866.
48	1264 +**			1318.4	48	3 2	64236.
49	1313 +**			1329.0	49	3 2	62179.
50	1315 +**			1335.0	50	3 2	59738.
51	1288 +**			1336.7	51	3 2	56963.
52	1357 +**			1334.5	52	3 2	53908.
53	1333 +**			1328.8	53	3 2	50630.
54	1279 +**			1319.8	54	3 2	47185.
55	1284 +**			1308.1	55	3 2	43634.
56	1292 +**			1293.9	56	3 2	40031.
57	1294 +**			1277.7	57	3 2	36431.
58	1242 +**			1259.8	58	3 2	32883.
59	1196 +**			1240.4	59	3 2	29433.
60	1218 +**			1219.9	60	3 2	26119.
61	1239 +**			1198.3	61	3 2	222974.
62	1189 +**			1175.8	62	3 2	20025.
63	1180 +**			1152.5	63	3 2	217291.
64	1133 +**			1128.4	64	3 2	214786.
65	1105 +**			1103.6	65	3 2	212518.
66	1134 +**			1078.0	66	3 2	210487.
67	1066 +**			1051.6	67	3 2	28690.1
68	1075 +**			1024.2	68	3 2	7119.7
69	1022 +**			995.7	69	3 2	5763.8
70	1025 +**			966.0	70	3 2	4608.0
71	959 +**			935.0	71	3 2	3635.7
72	874 +**			902.7	72	3 2	2828.8
73	858 +**			868.9	73	3 2	2168.6
74	776 +**			833.7	74	3 2	1636.6
75	780 +**			796.9	75	3 2	1214.6
76	768 +**			758.7	76	3 2	885.4
77	688 +**			719.0	77	3 2	635.0
78	716 +**			678.0	78	3 2	443.3
79	583 +**			635.9	79	3 2	303.6
80	564 +**			592.8	80	3 2	202.8
81	515 +**			549.0	81	3 2	131.9
82	493 +**			504.7	82	3 2	83.3
83	486 +**			460.3	83	3 2	50.9
84	417 +**			416.1	84	3 2	30.0
85	320 +**			372.4	85	3 2	17.0
86	370 +**			329.6	86	3 2	9.2
87	229 +**			288.1	87	3 2	4.7
88	244 +**			248.4	88	3 2	2.3
89	272 +**			210.7	89	3 2	1.0
90	118 +**			175.5	90	3 2	0.4
91	200 +**			143.2	91	3 2	0.2
92	151 +**			114.1	92	3 2	0.1
93	48 +**			88.4	93	3 2	0.0
94	92 +**			66.3	94	3 2	0.0
95	60 +**			48.1	95	3 2	0.0
96	7 +**			33.7	96	3 2	0.0
97	54 +**			23.1	97	3 2	0.0
98	6 +**			15.9	98	3 2	0.0
99	0 +**			11.6	99	3 2	0.0
100	14 +**			8.9	100	3 2	0.0
DOMAIN FREQ1	ONE CHAR = 26 , ? = OUT OF RANGE, " = MULTI			FREQ2	DOMAIN FREQ1	ONE CHAR = 2E3 , ? = OUT OF RANGE, " = MULTI	FREQ2

分布 3 (観測された得点分布 * と当てはめられた混合ベータ分布 + ならびに混合分布の各成分)

TITLE	R	1	2	3	N	LMLHMB	AIC
Original:Beta6	3	A 3.780	25.012	8.331	35318	21605.95157612	-43195.90315223
		B 1.001	3.615	3.358			
		R 0.255	0.204	0.540			
Observed(*) and Expected(+) Frequency Distributions					Weighted Component Densities		
0 3				0.0	0		0.0
1 0				0.0	1		0.0
2 0				0.0	2		0.0
3 0				0.0	3		0.0
4 0				0.1	4		0.0
5 0				0.1	5		0.0
6 0				0.2	6		0.0
7 0				0.2	7		0.1
8 0				0.3	8		0.1
9 0				0.5	9		0.3
10 1				0.6	10		0.6
11 0				0.8	11		1.2
12 1				1.0	12		2.1
13 0				1.3	13		3.5
14 0				1.6	14		5.8
15 1				1.9	15		9.2
16 2				2.3	16		14.2
17 2				2.8	17		21.3
18 3				3.3	18		31.1
19 2				3.9	19		44.4
20 2				4.6	20		62.2
21 6				5.4	21		85.7
22 3				6.3	22		116.0
23 9				7.4	23		154.8
24 7				8.6	24		203.8
25 7				10.0	25		264.9
26 11				11.6	26		340.4
27 12				13.4	27		432.5
28 11				15.4	28		544.2
29 11				17.8	29 1		678.1
30 13				20.4	30 1		837.5
31 21				23.4	31 1		1025.9
32 22 +				26.8	32 "		1246.7
33 33 +				30.6	33 "		1503.8
34 38 +				34.9	34 "		1801.1
35 37 +				39.7	35 "		2142.9
36 34 +				45.0	36 13		2533.3
37 54 ++				51.0	37 13		2976.7
38 63 ++				57.6	38 13		3477.6
39 71 ++				64.9	39 13		4040.2
40 81 ***				72.9	40 1 3		4668.9
41 99 ***				81.7	41 1 3		5367.9
42 96 ***				91.3	42 1 3		6141.2
43 93 ***				101.7	43 1 3		6992.7
44 140 *****				113.1	44 1 3		7925.8
45 135 *****				125.4	45 1 3		8943.5
46 143 *****				138.6	46 1 3		10049.
47 180 *****				152.8	47 1 3		11243.
48 171 *****				168.0	48 1 3		12528.
49 192 *****				184.2	49 1 3		13905.
50 191 *****				201.4	50 1 3		15374.
51 208 *****				219.6	51 1 3		16933.
52 222 *****				238.8	52 1 3		18681.
53 259 *****				259.0	53 1 3		20314.
54 257 *****				280.0	54 1 3		22130.
55 300 *****				302.0	55 1 3		24022.
56 351 *****				324.7	56 1 3		25984.
57 304 ***** +				348.3	57 1 3		28008.
58 387 *****				372.5	58 1 3		30086.
59 372 *****				397.2	59 1 3		32207.
60 389 ***** +				422.5	60 1 3		34359.
61 484 *****				448.2	61 1 3		36529.
62 449 *****				474.2	62 1 3		38702.
63 485 *****				500.3	63 1 3		40862.
64 582 *****				526.5	64 1 3		42993.
65 529 *****				552.7	65 1 3		45075.
66 571 *****				578.7	66 1 3		47089.
67 639 *****				604.4	67 1 3		49014.
68 607 *****				629.8	68 2 1		50829.
69 682 *****				654.8	69 2 1		52511.
70 660 *****				679.3	70 2 1		54038.
71 676 *****				703.3	71 2 1		55387.
72 760 *****				726.9	72 2 1		56535.
73 709 *****				749.9	73 2 1		57459.
74 728 *****				772.5	74 2 1		58138.
75 874 *****				794.9	75 2 1		58549.
76 740 ***** +				816.9	76 2 1		58675.
77 898 *****				838.9	77 2 1		58497.
78 912 *****				860.9	78 2 1		58000.
79 781 ***** +				882.9	79 2 1		57171.
80 1014 *****				904.9	80 21		56002.
81 932 *****				927.0	81 12		54489.
82 887 ***** +				948.8	82 1 2		52630.
83 1163 *****				970.0	83 1 2		50430.
84 919 ***** +				990.0	84 1 2		47901.
85 955 ***** +				1007.9	85 1 2		45060.
86 1196 *****				1022.5	86 1 2		41930.
87 680 ***** +				1032.2	87 1 2		38543.
88 1091 *****				1035.3	88 1 3		34940.
89 1297 *****				1029.6	89 1 3		31169.
90 541 ***** +				1012.8	90 1 3		27287.
91 1247 *****				982.7	91 3 1		23360.
92 1161 *****				937.4	92 3 1		19462.
93 307 ***** +				875.9	93 3 1		15677.
94 1361 *****				798.3	94 3 1		12095.
95 641 ***** +				707.2	95 3 1		8812.1
96 79 ***** +				607.5	96 3 1		5926.2
97 1171 *****				507.7	97 3 2		3534.1
98 271 ***** +				419.9	98 3 2		1722.8
99 0 ***** +				358.2	99 2 1		556.4
100 589 *****				334.3	100 2 1		44.9
DOMAIN FREQ1	ONE CHAR = 26	?, ? = OUT OF RANGE, " = MULTI	FREQ2	DOMAIN FREQ1	ONE CHAR = 1E3	?, ? = OUT OF RANGE, " = MULTI	FREQ2

Fitting the Mixture of Beta Distributions to the Test Score Distributions

MAYEKAWA Shin-ichi*

Abstract

In this paper, a method to fit the finite mixture of beta distributions using the EM algorithm is presented. In addition, a general method to fit a finite mixture distribution using the EM algorithm is reviewed and a method to find a rational initial parameter estimates is discussed.

As an example, the smoothing of the score distributions of the National Center Test is presented.

Key words: finite mixture distributions, beta distributions, EM algorithm, The National Center Examinations

* Department of Applied Statistics and Measurement, Research Division, The National Center for University Entrance Examinations, 2-19-23 Komaba, Meguro-ku, Tokyo 153-8501 Japan
mayekawa@hum.titech.ac.jp