

研究種目：若手研究（B）

研究期間：2007～2008

課題番号：19700259

研究課題名（和文） 混合分布モデルの一致推定量の研究

研究課題名（英文） Consistency of maximum likelihood type estimators for mixture models

研究代表者

田中 研太郎（TANAKA KENTARO）

東京工業大学・大学院社会理工学研究科・助教

研究者番号：00376948

研究成果の概要：

罰則付き最尤推定量が強一貫性を持つことを厳密に証明することができた。この結果により、Hathaway(1985) や McLachlan and Peel(2000) などによって言及された未解決問題を肯定的に解決することに成功した。また、罰則付き最尤推定アルゴリズムを、EMアルゴリズムを基に導出し、数値実験で適切に動くことを確かめた。これにより、従来の最尤推定では破綻してしまう可能性があるパラメータの推定において、その危険性を克服した推定が可能になった。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	900,000	0	900,000
2008年度	700,000	210,000	910,000
年度			
年度			
年度			
総計	1,600,000	210,000	1,810,000

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：統計的推測

1. 研究開始当初の背景

混合分布モデルとは、いくつかの確率分布を重みを付けて足し合わせて得られる確率分布モデルのことであり、画像認識・音声認識・遺伝情報解析および神経回路網といった情報科学のさまざまな分野に応用されている。一方、そのような分野でよく使われる混合正規分布において、パラメータの非識別性や、尤度関数の非有界性により最尤推定が不可能であるという問題点があることも

知られており、それらの数理的な解明が求められている。

混合分布モデルにおける理論的な研究で、本研究と関連があるとして次のものがあげられる。Redner(1981) は、Wald(1949)の結果を利用して混合分布モデルで母数空間をコンパクトなものに制限すれば最尤推定量の一致性が成り立つことを示している。Hathaway(1985) は、混合正規分布において標準偏差の比に制限を付ければ、最尤推定量の一致性が成り立つことを示している。Chen and Calbfeisch(1996) は、成分数に対する一致推定量を重みに対

する罰則付き尤度を考えることで構成している。また、ベイズ的なアプローチとしては、Richardson and Green(1997) による reversible jump MCMC 法 や、混合正規分布の事後分布の収束のオーダーを扱った Ghosal and Van der Vaart(2006) などがある。

2. 研究の目的

尤度を最大化するパラメータを選ぶという最尤推定の手法は、通常であれば、非常によい性質を持つため、統計的推測の場面においてよく使われる。しかし、混合正規分布などを含む混合位置尺度分布モデルにおいては、ある成分の位置母数をサンプルの1つの点と等しくとり、その同じ成分の分散を0に近づけると、無限に大きい尤度を得られてしまう。このことから、混合位置尺度分布モデルにおいては、最尤推定が、そのままでは意味をなさなくなるという問題点がある。

本研究では、混合位置尺度分布モデルにおいて尤度が発散してしまうという問題点を、パラメータ空間に制限をつけた最尤推定や、罰則付きの最尤推定などを扱うことによって回避することを考える。そして、その推定量の一致性などの漸近的な性質を明らかにするものである。また、本研究で生み出される推測方法を多数のシミュレーションを通して比較することにより、実際の応用の場面で有効な推測方法を確立することを目的としている。

3. 研究の方法

(1) 混合位置尺度分布において、尺度母数に対する罰則付き尤度を考えたとき、罰則付き最尤推定量の強一致性を厳密に証明した。また、尺度母数の比に対して罰則を入れる方法を考え、同様の性質が成り立つことを証明した。これらの証明においては、制約付きの最尤推定量における強一致性の証明のアイデアを用いた。確率的な評価においては、「丘本の不等式」を用いている。

(2) 罰則付き最尤推定のアルゴリズムを、EM アルゴリズムを基に導出した。アルゴリズムの妥当性は、数値実験で検証した。

(3) 罰則付き最尤推定のアルゴリズムにおいて現れる正則化パラメータを、クロスバリデーションによって選択する方法を与えた。そのアルゴリズムの妥当性を、数値実験によって検証した。

4. 研究成果

(1) 混合位置尺度分布において、「尺度母数」に対する罰則付き尤度を考えたとき、罰則付き最尤推定量の強一致性を厳密に証明した。

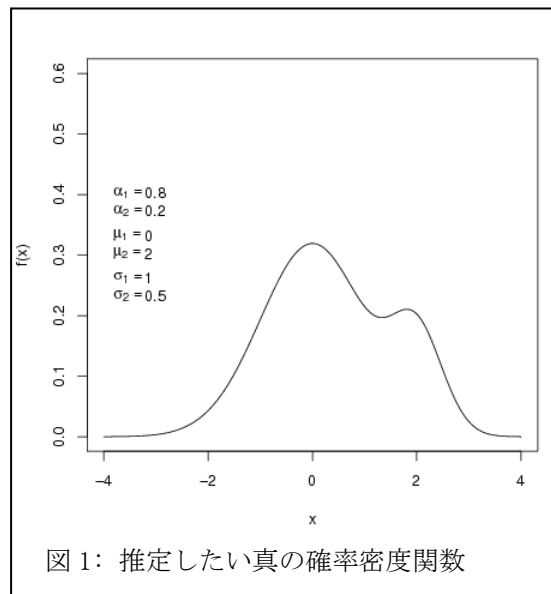
(2) 混合位置尺度分布において、「尺度母数の比」の最小値に対する罰則付き尤度を考えたとき、罰則付き最尤推定量の強一致性を厳密に証明した。

(3) (1), (2)の罰則付き最尤推定量の罰則を、標本サイズの増加とともに、強一致性を保ったまま緩和できることを証明した。

(4) 混合位置尺度分布において、「尺度母数の比」に対する「制限」を付けたときの尤度を考えたとき、その制約付き最尤推定量が強一致性を持つことを厳密に証明した。この結果により、Hathaway(1985) や Mclachlan and Peel(2000) などによって言及された未解決問題をより広いクラスのモデルに対して肯定的に解決することに成功した。

(5) 制約付き最尤推定と罰則付き最尤推定を求める反復アルゴリズムを、EM アルゴリズムを基に導出した。また、数値実験を通して、そのアルゴリズムが尤度の発散を回避して、妥当な推定を与えることを確かめた。

いま、図1の密度関数を持つ2成分混合正規分布から100個のデータが発生していると



この混合正規分布に対して、2成分混合正規分布モデル

$$\mathcal{C}_1 \phi\left(\frac{x - \mu_1}{\sigma_1}\right) + \mathcal{C}_2 \phi\left(\frac{x - \mu_2}{\sigma_2}\right)$$

のパラメータ(各成分の重み、平均、分散)を最尤推定量を求める通常のEMアルゴリズムで推定すると、図2のように、異常に高いピークを持った密度関数が推定されてしまう。このピークがあるせいで尤度が無限大に近づいていくため、このEMアルゴリズムの反復をさらに繰り返すと、計算不能に陥ってしまう。従って、真の密度関数を推定することができない。

一方、制約付き最尤推定を行うために変更したEMアルゴリズムを使ってデータから推定した密度関数は、図3のようになった。通常のEMアルゴリズムにおいて異常に高いピークが生じてしまう現象が回避されて、真の密度関数に近い関数の形になっている。また、罰則付き最尤推定を行うようにEMアルゴリズムを変更したアルゴリズムを使ってデータから密度関数を推定すると、図4のようになった。やはり、ピークを持ってしまう現象が回避され、良好な推定結果が出ていることが確認できる。

(6) 制約付き最尤推定や罰則付き最尤推定においては、正則化パラメータが出てくるため、これらについても適切に選択しなくてはいけない。そこで、クロスバリデーションの手法を用いることにより、正則化パラメータを選択する方法についても定式化した。また、これらの手法が実用的であるかどうかを、数値実験で確かめた。

図3と図4の推定結果を得るときにも、クロスバリデーションを使って得られた正則化パラメータを用いており、よって、この数値例からも、クロスバリデーションによる、適切な正則化パラメータが選択できることが確認できる。

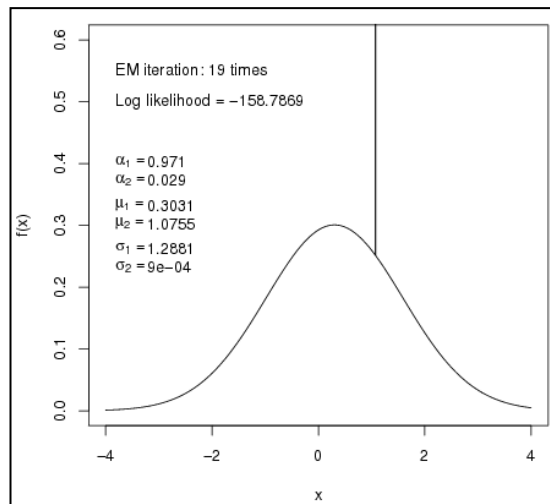


図2：通常のEMアルゴリズムにより推定された密度関数

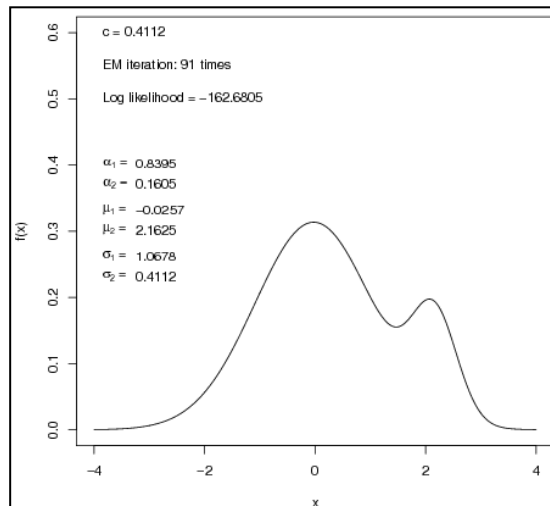


図3：制限付きのEMアルゴリズムにより推定された密度関数

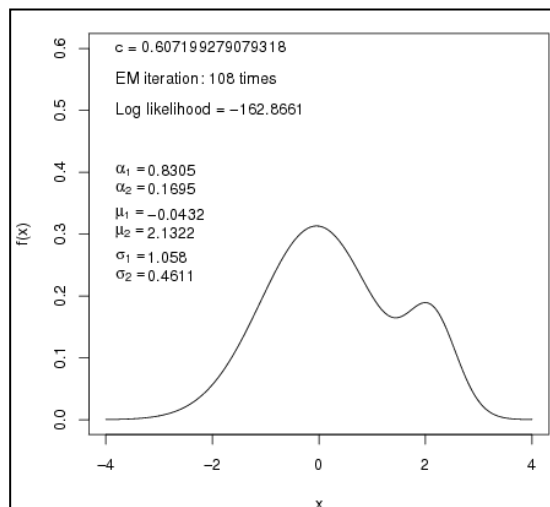


図4：罰則付きのEMアルゴリズムにより推定された密度関数

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 3件)

(1) Kentaro Tanaka. (2008).

“Strong Consistency of the Maximum Likelihood Estimator for Finite Mixtures of Location-Scale Distributions When Penalty is Imposed on the Ratios of the Scale Parameters”, *Scandinavian Journal of Statistics*, **36**, pp.171-184, 査読有り.

(2) 宮川雅巳, 田中研太郎. (2007).

“2値信号と対応のある量的反応における等SN比性の検定”. *品質*, **37**, pp.49-57, 査読有り.

(3) 宮川雅巳, 田中研太郎, 岩澤智之, 中西寛子. (2007). “マハラノビス・タグチ・システムにおける実際の誤判別率”. *品質*, **37**, pp.101-106. 査読有り.

[学会発表] (計 2件)

(1) 田中研太郎. “グラフィカルモデリングと時系列データ”, 統計関連学会連合大会, 2008年9月8日, 慶應義塾大学矢上キャンパス.

(2) 田中研太郎. “尺度母数の比を制限した混合分布における最尤推定量の一致性”, 統計関連学会連合大会, 2007年9月6日, 神戸大学.

6. 研究組織

(1) 研究代表者

田中 研太郎 (TANAKA KENTARO)

東京工業大学・大学院社会理工学研究科・助教

研究者番号: 00376948

(2) 研究分担者

—

(3) 連携研究者

—