

統計科学のフロンティア 7

特異モデルの統計学

統計科学のフロンティア 7

甘利俊一 竹内啓 竹村彰通 伊庭幸人 編

特異モデルの統計学

未解決問題への新しい視点

福水健次 栗木哲
竹内啓 赤平昌文

岩波書店

編集にあたって

非正則モデルの意義

数理科学の教科書には、主として統計的データ解析における標準的な問題が扱われている。

連続的なデータについては、それらが正規分布に従うものと仮定して、その平均や分散についての検定や推定を行う問題である。そこで標準的な、たとえば t 検定の手法などとともに(中級以上の教科書であれば)そのよさ(最適性)についても示されている。それをやや一般化したものが線形回帰モデルであり、それについて最小 2 乗推定法が説明され、また誤差項が正規分布であることを仮定して、最小 2 乗推定量の最適性が説明され、また係数に関する仮説に対する t 検定や F 検定が示されている。

また離散データについては、2 項分布、あるいはポアソン分布を前提にして、推定や検定について説明されているであろう。

これらは数理統計学の問題の中で、いわば最も簡単な場合といえるが、多くの現実の場において実際に現われる問題であって、その手法は有効に使われている。数理統計学の理論は、これらの手法の概念(推定量、検定の棄却域、信頼区間等)とその性質(不偏性その他)そのよさ(最小分散不偏推定量、一括最強力不偏検定)等を説明して、手法のもつ意味を明らかにしているのである。

このような標準的な問題についてのモデルは 3 つの方向に拡張されている。

第一は指数型分布といわれるもので、観測値の組をベクトル \mathbf{X} で表わすとき、その確率密度が未知母数ベクトル $\boldsymbol{\theta}$ をふくんで

$$p(\mathbf{x}, \boldsymbol{\theta}) = h(\mathbf{x}) \exp \sum_{j=1}^m t_j(\mathbf{x}) \theta_j(\boldsymbol{\theta})$$

という形に表わされる場合である。このようなモデルについては未知母数に関する推定、検定の手法の最適性が標準的な場合と同様に証明される。

第二は、さらに一般的に観測値 X_1, \dots, X_n が互いに独立に、また未知母

数ベクトル θ をふくむ密度関数 $f(x, \theta)$ をもつ分布に従うというモデルであり、この場合は推定や検定の手法について一般に最適性は証明できないが、 θ の関数としての f についてなめらかさの条件を仮定とすると、 n が大きいときほぼ最適な性質(漸近最適性)をもつ手法が存在すること、とくに最尤推定量や尤度比検定が漸近最適性をもつことが証明されている。

第三はデータの従う確率分布が、有限個のパラメータをふくむ密度関数によって表現されることを仮定しない場合である。たとえば X_1, \dots, X_n が互いに独立に連続な密度をもつ分布に従うことだけで仮定するのである。このような仮定はノンパラメトリックモデルとよばれる。その場合にもたとえば X の分布の中央値について仮説検定や区間推定を行うことができる。ノンパラメトリックモデルに関する統計的手法、とくにノンパラメトリック検定については古くからの研究があり、多くの手法が提案してきた。

ところで上記の第二の場合、すなわち密度関数が一定の条件(正則条件とよばれる)を満たす場合は、正則な場合とよばれるが、そこでもしそれが満たされない場合には、どのようなことがおこるかが問題となる。それが非正則な場合(non-regular case)の問題である。

このことを考えるには 2 つの理由がある。

ひとつは、正則条件は、たとえば最尤推定量の漸近的最適性を証明するために必要とされるものであり、いわば数学的理論構成の便宜上導入されたものであるから、それが本当に必要なものか、つまりそれが成り立たなければ定理がもはや成立しなくなるのかをチェックする必要があるということである。実は正則条件はいくつかの部分から成り立っているのであり、どの部分が成立しなければ、推測方式の性質についての結論がどのように変わるのでをチェックすることが必要になる。

次に、いろいろな複雑な現実の問題において、実際に正則条件が成立しない場合がおこることである。そのような場合には、標準的な手法が適用できなかったり、あるいはそれから誤った結論に導かれたり、あるいはその効率が悪かったりすることがおこり得る。たとえば最尤推定量がもはや漸近的最適性をもたないことがある。そのような場合には他のよりよい手法を探さなければならない。このような問題は、ある意味では非正則な場

合の特殊な場合と見なすことができるが、現実のデータに対応するモデルに即して具体的に考えなければならない。この巻ではとくにこのようなモデルを一般的な非正則モデルとは区別して、特異モデル(singular model)とよぶことにする。

非正則な場合の統計的推測理論については比較的古くから、散発的にいろいろなことが発見されてきたが、それらはまだあまり一般的な知識となっていないように思われる。これに対して特異モデルについては、最近この巻の執筆者をふくむ一部の研究者の精力的な研究によって、まとまった理論的成果が得られている。

そこでこの巻では特異モデルを中心的なケースとして、それに補論として一般の非正則モデルについての理論の概略について述べることにした。

(竹内 啓)

目 次

編集にあたって

特異モデルの統計学	福水健次・栗木哲	1
補論 A 非正則な場合の推測理論	竹内啓	231
補論 B 非正則モデルの情報損失	赤平昌文	245
索 引	271	

裝丁 蟻名優子

特異モデルの統計学

福水健次・栗木哲

目 次

1 特異点をもつ統計モデル	4
1.1 特異モデルの典型例——識別不能性と境界をもつパラメータ空間	4
1.2 特異モデル——接ベクトル集合の特異な構造	21
1.3 特異モデルのさまざまな例	37
1.4 最尤推定量の数値解法	53
2 パラメータ制約モデルの漸近論	56
2.1 有限混合モデルと遺伝連鎖解析	56
2.2 ランダム係数回帰モデルとプロファイル解析	60
2.3 最尤推定と尤度比検定の漸近論	64
2.4 尤度比の極限分布の例	75
2.5 定理の証明	82
3 チューブ法——正規確率場の幾何学	87
3.1 はじめに	87
3.2 チューブの体積と正規確率場の最大値分布	88
3.3 チューブ体積公式	92
3.4 チューブ座標と臨界半径	100
3.5 チューブ法による極限分布近似の例	106
3.6 オイラー標数法	109
3.7 チューブ法の歴史と文献	118
4 凸多面錐をパラメータ空間とするモデル	121
4.1 順序制約と単調回帰モデル	121
4.2 単調回帰モデルにもとづく統計推測	123
4.3 凸多面錐を対立仮説とする検定	135
4.4 同時信頼区間の構成	143
5 無限次元の特異モデル	146
5.1 無限次元空間の中の接錐	146
5.2 ガウス過程による尤度比の解析	160
5.3 尤度比の発散とそのオーダー	173
6 その他の話題	193
6.1 対数尤度関数の大域的性質	193
6.2 罰則付き最尤法	198
6.3 特異モデルにおける Bayes 推定	203
付 錄	214
1 確率論からの準備	214
2 多様体についての必要事項	216
3 オイラー標数	223
参考文献	226



本稿は「特異モデル」にまつわる統計的な諸問題を論じている。特異モデルという用語から、特殊なモデルの特殊な話題を想像する読者もいるかもしれない。しかし、本稿で扱う特異モデルの例のほとんどは、統計科学や情報科学において頻繁に用いられる一般的なモデルである。これらのモデルは、特異性に起因する複雑で興味深い挙動を示すが、それらは従来、各モデル固有の問題として議論されてきた。本稿の目的は、それらの現象を特異モデルという共通の観点から見直すことである。

特異モデルは、通常の統計的漸近理論が対象とする「正則モデル」に対峙する概念と考えられる。これは従来「非正則モデル」と呼ばれていた。しかしながら、通常の漸近理論が成立するためにはいくつもの正則条件が必要で、それがどう破られるかに依存して、生じる現象は異なってくる。本稿で扱うのは、統計モデルが滑らかでない点を有するという、ある特定のクラスの非正則モデルである。その点を強調するため、特異モデルという新しい語を用意した。なお、非正則モデル一般に関する解説は補論を参照していただきたい。

本稿は全 6 章から構成されている。まず 1 章で特異モデルの全体像を概観する。特異モデルは有限次元の接錐をもつ場合と無限次元の接錐をもつ場合がある。2 章から 4 章は前者の有限次元の場合を扱い、5 章は後者の無限次元の場合を扱う。モデルの特異性は、統計推測のさまざまな側面に影響を及ぼすが、本稿では話題を主に尤度にもとづく議論に限定し、それ以外の話題は最後の 6 章で簡単に扱うこととした。中心的な話題である統計的漸近理論に興味のある読者は、1 章前半(1.1 節, 1.2 節), 2 章, 5 章, 6 章後半(6.2 節, 6.3 節)を中心に読み進めてほしい。有限次元の特異モデルでは、尤度比検定などに関し、より進んだ議論が可能な場合がある。それらのうちで重要なものを 3 章, 4 章で論じた。なお執筆は、主として 1 章, 5 章, 6 章を福水が、2 章, 3 章, 4 章を栗木が担当した。

1 特異点をもつ統計モデル

特異モデルの代表的な例は、本章 1.1 節で述べる、識別不能なパラメータをもつ場合とパラメータ空間に端点が存在する場合である。これらには有限混合モデルや隠れマルコフモデル、制約つき推定問題、分散成分モデルといった応用上重要なモデルが含まれる。本章では、識別不能性や境界をもつモデルの何が特異なのかを説明した後、1.2 節で統計モデルの特異点を一般的に定義する。さらに、1.3 節でさまざまな統計モデルに現われる特異点の例を概観し、1.4 節で数値的問題との関係について触れる。

1.1 特異モデルの典型例—— 識別不能性と境界をもつパラメータ空間

特異モデルの一般的な定義を与える前に、本節ではその典型例である識別不能性をもつモデルとパラメータ空間に境界をもつモデルについて具体例を通して説明する。特異モデルの何が特異であるかを理解するためには、特異でない場合、すなわち正則な場合の統計的推測に関する理論を知っている必要がある。そこでまず、(a) 項で本稿の主たる議論の対象である最尤推定量の統計的挙動について基本的事項を復習する。パラメトリック推定や検定、漸近理論についてさらに詳しく知りたい読者は、たとえば稻垣(2003)や van der Vaart(1998)などを見ていただきたい。後者は漸近理論の本格的教科書である。

(a) 正則モデルにおける最尤推定と推定量の挙動

観測されたサンプル X_1, \dots, X_n を用いて、さまざまな量の間の関係を導いたり推論を行ったりするためには、その背後にある確率分布を推測することが重要である。そのためには、サンプルを発生させた「真の」分布

を含んでいると想定される確率分布の族を用意し、与えられたサンプルにもっともよく当てはまる分布をその中から選ぶという方法がよく用いられる。候補となる確率分布の族が有限個のパラメータによって定められているとき、これをパラメトリックモデルとよぶ。

より正確には、測度空間 $(\mathfrak{X}, \mathfrak{B}, \nu)$ 上の確率密度関数の族を統計モデルといい、統計モデル \mathcal{S} が、 \mathbb{R}^m の部分集合 Θ をパラメータ空間として $\mathcal{S} = \{f_\theta \mid \theta \in \Theta\}$ と表わせるとき、 \mathcal{S} をパラメトリックモデル(parametric model)という。パラメトリックモデルに対するパラメータの与え方は一通りとは限らない。統計モデル \mathcal{S} に特定のパラメータを与えて $\mathcal{S} = \{f_\theta \mid \theta \in \Theta\}$ の形に表わすことをパラメトリゼーションという。本稿では以降、統計モデルというとパラメトリックモデルを指すことにする。統計モデルに属する確率密度関数に対して、 $x \in \mathfrak{X}$ を明示する際には $f(x|\theta) = f_\theta(x)$ の記法もよく用いる。サンプル X_1, \dots, X_n が、 $f_{\theta_0}(\theta_0 \in \Theta)$ により定まる確率分布から発生した n 個の独立なサンプルであると仮定して、真のパラメータ θ_0 を推定するパラメトリック推定の問題を考えよう。たとえば、真の分布が、分散共分散行列が単位行列で平均ベクトルが未知の m 次元正規分布だとわかっているとき、

$$f(x|\theta) = \frac{1}{(2\pi)^{m/2}} \exp\left(-\frac{1}{2}\|x - \theta\|^2\right) \quad (1)$$

で定義される正規分布の平均値モデル $\{f(x|\theta) \mid \theta \in \mathbb{R}^m\}$ を使って、独立なサンプル X_1, \dots, X_n から真の平均値 θ_0 を推定する問題は、そのひとつの例である。実際に遭遇する現実世界の問題においては、設定したモデルが真の確率分布を完全に含むと仮定するのは理想的すぎるが、このような理想化によって統計的現象の数理的把握が容易になり、現実に役に立つ方法論も考えやすくなる。もちろん、それらの理論が実際の現象にどれくらい適合するかを検証することはきわめて重要であるが、本稿ではモデルが真の分布を含むという前提のもとで理論を開いていく。

与えられたサンプルからパラメータを推定する一般的な手法として最尤推定(maximum likelihood estimation)がある。最尤推定は

$$\prod_{i=1}^n f(X_i|\theta)$$

により定義される尤度(likelihood), あるいはその対数をとった対数尤度(log likelihood)

$$\sum_{i=1}^n \log f(X_i|\theta)$$

の最大値を達成するパラメータ θ を推定量として用いる。対数関数は狭義単調増加なので同じものを定める。これを最尤推定量(maximum likelihood estimator, MLE)とよび $\hat{\theta}_n$ で表わす。最尤推定量はサンプルに依存するので, $\hat{\theta}_n(X_1, \dots, X_n)$ と書くのがより正確であるが, 通常 $\hat{\theta}_n$ と略記し, また n も省略して $\hat{\theta}$ と書くことが多い。最尤推定量はサンプルに依存する確率変数であり, その確率分布の性質を調べることは統計的推測理論の中心的課題のひとつである。本稿の議論の中心も最尤推定量とそれに関連する統計量の統計的性質にある。

最尤推定量の分布を調べるには大きく分けて 2 種類のアプローチがある。そのひとつは, サンプルを発生させている分布に強い仮定をおいて最尤推定量の分布を詳しく調べる方法である。典型的な仮定はサンプルが正規分布に従うとするものである。ここで例として, 先に述べた正規分布の平均値モデルの最尤推定量の分布を求めてみよう。この場合の対数尤度は

$$-\frac{1}{2} \sum_{i=1}^n \|X_i - \theta\|^2 - \frac{nm}{2} \log(2\pi)$$

であるので, 第 1 項を最大化することにより, 最尤推定量は

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

で与えられ, サンプルによる標本平均値に一致する。各 X_i が, 平均ベクトルが θ_0 , 分散共分散行列が単位行列 I_m の m 次元正規分布 $N_m(\theta_0, I_m)$ に従うことから, その線形和で与えられる $\hat{\theta}_n$ も正規分布に従い, その分布は $N_m\left(\theta_0, \frac{1}{n}I_m\right)$ であることがわかる。いいかえると, 任意の n に対して

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \sim N_m(0, I_m)$$

が成立する。ここで記号 \sim は左の確率変数が右の確率分布に従うことを表

わす. この例のように, 任意のサンプル数 n について推定量の正確な分布を求めようとする立場を精密標本理論(exact sample theory)という. 一般に分布の正規性を仮定すると推定量に関する性質を詳しく調べられるが, その場合でもパラメータ θ のとりうる範囲に制約が加わると問題は突然難しくなる. 本節(c)項で例のひとつを示し, 2章, 4章で詳しい議論を行う.

最尤推定量の精密な分布を求めるることは困難な場合が多いため, サンプル数 n が非常に大きいという仮定のもとで, 最尤推定量の統計的性質を一般的に導こうとするのが第2のアプローチであり, 統計的漸近理論(statistical asymptotic theory)とよばれる. 統計モデル $\{f(x|\theta) \mid \theta \in \Theta\}$ があり, サンプル X_1, \dots, X_n が真のパラメータ θ_0 で定まる確率分布に従う独立同分布確率変数であるとき, ある種の正則条件のもと, 最尤推定量 $\hat{\theta}_n$ は真のパラメータ θ_0 に確率収束^{*1}する. この性質を最尤推定量の一致性(consistency)という. さらにいくつかの正則条件を仮定すると, θ の最尤推定量 $\hat{\theta}_n$ は

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{} N_m(0, I^{-1}(\theta_0)) \quad (n \rightarrow \infty) \quad (2)$$

と法則収束することが知られている. ここで $m \times m$ 行列 $I(\theta)$ の (a, b) 要素は

$$I(\theta)_{ab} = \int \frac{\partial \log f(x \mid \theta)}{\partial \theta^a} \frac{\partial \log f(x \mid \theta)}{\partial \theta^b} f(x \mid \theta) d\nu(x) \quad (1 \leq a, b \leq m)$$

で与えられ, **Fisher 情報行列**(Fisher information matrix)とよばれる. 式(2)は, \sqrt{n} のスケール変換のあと最尤推定量の分布が真のパラメータを中心とした正規分布に収束することを示しており, この性質を最尤推定量の漸近正規性(asymptotic normality)という. とくに, その正規分布のばらつきの程度, すなわち推定の良し悪しは Fisher 情報行列の逆行列で与えられる. したがって, Fisher 情報行列 $I(\theta)$ はパラメータ θ の推定しやすさを表わしていると考えてもよい.

実は, 推定量がある条件を満たす場合, 極限分布の分散共分散行列は, 半正定値性による半順序の意味で Fisher 情報行列の逆行列を下回らないことが知られている. これは補論 A でも述べられている Cramér-Rao の不

*1 概収束, 確率収束, 法則収束といった確率変数の収束に関しては付録 1 にまとめた.

等式^{*2}の漸近版である。一般に、推定量 T_n があって、任意の θ に対して $\sqrt{n}(T_n - \theta)$ が f_θ のもとで $n \rightarrow \infty$ のとき L_θ という分布に法則収束し、 L_θ に従う確率変数の分散共分散行列が $I^{-1}(\theta)$ に等しいとき、この推定量 T_n は漸近有効(asymptotically efficient)であるという。式(2)は最尤推定量が漸近有効であることを示しており、ある意味での最適性を保証している。

式(2)を見ると、漸近正規性が成り立つためにはいくつかの条件が必要であることに気づく。まず Fisher 情報行列の定義から、確率密度関数は真のパラメータ θ_0 でパラメータに関して微分可能でなければならない。収束先の正規分布の分散共分散行列が Fisher 情報行列の逆行列で与えられている以上それは可逆でなければならない。また、 θ_0 の近傍で正規分布を考えるために、 θ_0 はパラメータ空間の内点になければならない。

漸近正規性に関する厳密な議論は 2 章で改めて述べるが、正則な場合の漸近正規性をあまり厳密でない形で導出しておこう。ここでは表記を簡単にするためパラメータの次元を 1 次元とするが、多次元の場合への拡張も容易である。まず、積分と微分の交換が可能となる条件を課しておくと、

$$\begin{aligned} E_\theta \left[\frac{\partial \log f(X|\theta)}{\partial \theta} \right] &= 0, \\ E_\theta \left[\frac{\partial^2 \log f(X|\theta)}{\partial \theta \partial \theta} \right] + E_\theta \left[\frac{\partial \log f(X|\theta)}{\partial \theta} \frac{\partial \log f(X|\theta)}{\partial \theta} \right] &= 0 \quad (3) \end{aligned}$$

が成り立つ。最尤推定量は対数尤度関数の最大値であるので、それがパラメータ空間の内点にあるならば、次の尤度方程式を満足する。

$$\sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(X_i | \hat{\theta}_n) = 0$$

この左辺を θ_0 を中心に Taylor 展開すると、

$$\begin{aligned} \sum_{i=1}^n \frac{\partial \log f(X_i | \theta_0)}{\partial \theta} + \sum_{i=1}^n \frac{\partial^2 \log f(X_i | \theta_0)}{\partial \theta \partial \theta} (\hat{\theta}_n - \theta_0) \\ + \frac{1}{2} \sum_{i=1}^n \frac{\partial^3 \log f(X_i | \tilde{\theta})}{\partial \theta \partial \theta \partial \theta} (\hat{\theta}_n - \theta_0)^2 = 0 \end{aligned}$$

*2 Cramér-Rao の不等式については稻垣(2003)などを参照。

を得る。ここで $\tilde{\theta}$ は θ_0 と $\hat{\theta}_n$ を両端にもつ閉区間内の点である。これより

$$\begin{aligned} & \sqrt{n}(\hat{\theta}_n - \theta_0) \\ &= \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial \log f(X_i | \theta_0)}{\partial \theta}}{-\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \log f(X_i | \theta_0)}{\partial \theta \partial \theta} - \frac{1}{2n} \sum_{i=1}^n \frac{\partial^3 \log f(X_i | \tilde{\theta})}{\partial \theta \partial \theta \partial \theta} (\hat{\theta}_n - \theta_0)} \end{aligned}$$

と書ける。式(3)を使うと、中心極限定理^{*3}により分子は $N(0, I(\theta_0))$ に法則収束する。また、分母の 3 階微分の項が 2 階微分の項に比べて無視しうるほど小さいことを仮定しておくと、大数の法則により分母は $-E_{\theta_0} \left[\frac{\partial^2}{\partial \theta \partial \theta} \log f(X_i | \theta_0) \right]$ という値に確率収束するが、式(3)よりこれは $I(\theta_0)$ に一致する。したがって Slutsky の補題により $\sqrt{n}(\hat{\theta}_n - \theta_0)$ は $N(0, I(\theta_0)^{-1})$ に法則収束することがわかる。

本稿で議論する特異モデルでは、先の証明に必要な仮定の多くが成り立たない。ある場合には真のパラメータ θ_0 がパラメータ空間の境界上にあったり、他の例では Fisher 情報行列が逆行列をもたなかったりする。そのような場合には先述の証明がそのままでは適用できないばかりか、そもそも漸近分布が正規でなかったり、 \sqrt{n} のスケール変換では収束しなかったりするという一見異常な現象が見られる。それらに共通の見方を与えようとするのが本稿の目的である。

本稿では最尤推定量の漸近的挙動を調べる際、主に最大対数尤度比(maximum log likelihood ratio)を通して考える。これは、真のパラメータが θ_0 であるとき

$$L_n(\theta) = \sum_{i=1}^n \log \frac{f(X_i | \theta)}{f(X_i | \theta_0)}$$

の記法のもと

$$\sup_{\theta \in \Theta} L_n(\theta)$$

により定義される統計量である。最尤推定量 $\hat{\theta}_n$ が存在する場合には

*3 中心極限定理、大数の法則、Slutsky の補題などに関しては付録 1 を参照。

$$\sup_{\theta \in \Theta} L_n(\theta) = L_n(\hat{\theta}_n)$$

が成り立つ。本稿では誤解のない限り、最大対数尤度比を単に尤度比とよぶこととする。

尤度比は、サンプル X_1, \dots, X_n を生成するモデルのパラメータ θ が θ_0 であるか否かの統計的検定(尤度比検定, likelihood ratio test)のための統計量(尤度比検定統計量)として用いることができる。すなわち、棄却点とよばれる定数 c に対し

$L_n(\hat{\theta}_n) < c$ のとき仮説 $\theta = \theta_0$ を受容し、

$L_n(\hat{\theta}_n) \geq c$ のとき仮説 $\theta = \theta_0$ を棄却する($\theta \neq \theta_0$ と判断する)

とする統計的推測である。検定の対象の命題 $\theta = \theta_0$ は帰無仮説とよばれ、またその否定 $\theta \neq \theta_0$ は対立仮説とよばれる。棄却点 c は、あらかじめ与えられた定数 $\alpha(0 < \alpha < 1)$ に対し、

$$\Pr(L_n(\hat{\theta}_n) \geq c | \theta = \theta_0) = \alpha$$

を満たすように α の関数 $c = c_\alpha$ として決めておく。 α は、 $\theta = \theta_0$ が真であるときに $\theta \neq \theta_0$ であると誤判断する確率であり、第1種の誤り、あるいは検定のサイズとよばれる。棄却点 c を決めるためには、 $\theta = \theta_0$ であるときの $L_n(\hat{\theta}_n)$ の分布が必要となる。この分布は、帰無仮説のもとでの分布という意味で、帰無分布とよばれることがある。本稿の主要な目標のひとつは、特異モデルにおける尤度比検定統計量の帰無分布がどのようなものであるかを見ていくことである。

最尤推定量の漸近正規性と類似の正則条件のもと、漸近正規性と $L_n(\theta)$ の Taylor 展開を用いると、

$$\begin{aligned} L_n(\hat{\theta}_n) \\ = \frac{1}{2} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial \log f(X_i | \theta_0)}{\partial \theta} \right)^T I(\theta_0)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial \log f(X_i | \theta_0)}{\partial \theta} \right) \\ + o_p(1) \end{aligned}$$

と表わされることがわかる^{*4}。これにより

*4 $o_p(1)$ や $O_p(1)$ については付録 1 を参照。

$$2L_n(\hat{\theta}_n) \implies \chi_m^2 \quad (n \rightarrow \infty)$$

という法則収束が示される。ここで χ_m^2 は自由度 m のカイ 2 乗分布であり、 m はパラメータ θ の次元に一致する。 χ_m^2 は密度関数

$$g_m(x) = \frac{1}{2^{m/2}\Gamma(m/2)}x^{\frac{m}{2}-1}e^{-\frac{1}{2}x} \quad (x \geq 0, \Gamma(t) \text{ はガンマ関数})$$

をもつ確率分布で、標準正規分布に従う独立な m 個の確率変数の 2 乗和の分布としてよく知られている。カイ 2 乗分布への法則収束は、漸近正規性と同様、正則な漸近理論の基本的事実である。この事実から正則条件を満たす場合には、尤度比検定の棄却点 c_α を、漸近近似の意味で自由度 m のカイ 2 乗分布の上側 $100\alpha\%$ 点とおくことができる。

本稿が尤度比に主眼を置いているのは、次のような理由による。モデル $\mathcal{S} = \{f(x|\theta) \mid \theta \in \Theta\}$ に対し、 f_0 を真の密度関数とするとき、

$$\sup_{\theta \in \Theta} L_n(\theta) = \sup_{f \in \mathcal{S}} \sum_{i=1}^n \log \frac{f(X_i)}{f_0(X_i)}$$

と書き直せば明らかなように、定義されている確率密度関数族が同一である限り、尤度比はパラメトリゼーションに依存しない量である。したがって尤度比を考察することは、パラメトリゼーションに依存しない性質、すなわち「モデル」の性質を調べていることになる^{*5}。

以上の観点に立って本稿では最尤推定量に関する漸近理論を主に論じるが、もっと一般の推定量に関しても同様の漸近正規性や漸近カイ 2 乗性が導かれることが知られている(van der Vaart, 1998, 5 章)。しかしながら本稿では特異点に関する現象をなるべく簡単に知ってもらうために、6 章を除いては最尤推定量を中心にして議論する。

最尤推定量の漸近正規性は、AIC(Akaike information criterion)や MDL(minimum description length)といったモデル選択規準にも深く関わっている。確率分布の差異をはかるために、密度関数 $f_0(x)$ から $f_\theta(x)$ への

*5 もちろん問題によってはパラメータに物理的な意味づけがあり、パラメータの挙動そのものが知りたい場合もある。しかし特異モデルの場合には、パラメータの統計的挙動はパラメトリゼーションに依存してきわめて複雑な様相を呈し、現状では一般的な結果を述べるのは難しい。

Kullback-Leibler(KL)ダイバージェンス(KL ダイバージェンス)

$$K(f_0 \parallel f_\theta) = \int f_0(x) \log \frac{f_0(x)}{f_\theta(x)} d\nu$$

がよく用いられる。KL ダイバージェンスはつねに非負で、2つの確率分布が等しいときに限り 0 となる。 f_0 に関する期待値をサンプルで置き換えたものを $\hat{K}_n(\theta)$ とおくと、尤度比との間に

$$\hat{K}_n(\theta) = -\frac{1}{n} \sum_{i=1}^n \log \frac{f_\theta(X_i)}{f_0(X_i)} = -\frac{1}{n} L_n(\theta)$$

なる関係をもつので、尤度最大化は KL ダイバージェンス最小化を目的としていると考えてもよい。

いま、 f_0 からのサンプル $X_n = (X_1, \dots, X_n)$ にもとづく推定量 $\hat{\theta}$ の良さを、真の密度関数 f_0 からの KL ダイバージェンスによってはかり、そのサンプルによる期待値

$$E_{X_n}[K(f_0 \parallel f_{\hat{\theta}})]$$

を予測誤差とよぶことにしよう。ここで、モデルが正則条件を満たし、 $\hat{\theta}$ がその最尤推定量であるとき、漸近正規性と Taylor 展開を用いると、

$$E_{X_n}[K(f_0 \parallel f_{\hat{\theta}})] - E_{X_n}[\hat{K}_n(\hat{\theta})] = \frac{m}{n} + o(n^{-1})$$

であることが示される。この事実は AIC の理論的基礎づけとなっている。しかしながら、本稿で論じるような特異モデルでは、最尤推定量の漸近正規性が成立しておらず、AIC をそのまま用いることは理論的に妥当性をもっていない。また、MDL 規準の導出もやはり漸近正規性を用いており、それが成り立たないモデルに適用することはできない。

(b) パラメータの識別不能性——有限混合モデル

識別不能なパラメータをもつ統計モデルの例として有限混合モデルについて説明する。有限混合モデル(finite mixture model)は、パラメータ a をもった確率密度関数 $p(x|a)$ に対して、

$$f(x|\theta) = \sum_{k=1}^K c_k p(x|a_k)$$