

# ロバスト推定法とデータ解析への応用

## 1. 最尤法

物理現象を説明するモデルには、多くの場合未知パラメータが含まれていて、測定値からそれを決定しなければならない。決定されたパラメータによって、さらに別の現象に関する予言を与えることができ、それを実験で検証することにより、一歩一歩理解を深めていく。しかし測定値には誤差がつきものであるから、誤差の影響を最小限に押さえることが、よいデータ解析の基本条件である。すなわち、測定値  $y_j (j=1 \sim n)$  は、未知のパラメータ  $x_i (i=1 \sim m)$  の関数に誤差  $\varepsilon_j$  を加えたものとして、

$$y_j = f_j(x_1, \dots, x_m) + \varepsilon_j \quad (1.1)$$

の形であらわされる。 $f_j$  は既知の関数であるが、一般に  $x_i$  の 1 次関数とは限らない。 $y_j$  からどのように  $x_i$  を求めたらよいか、これが本稿の主題である。

誤差  $\varepsilon_j$  が従う分布がわかっている場合には、さいゆうほう最尤法 (maximum likelihood method) とよばれる方法を用いるのが普通である。最尤法は、データ数が充分大きい場合に偏りをもたない (漸近的不偏) だけでなく、有限個のデータに対しても一般に良い性質をもっている<sup>1)2)</sup>。

パラメータ  $\mathbf{x}$  (以下便宜のためベクトル記法を用いる) が与えられたとき、測定値  $y_j$  が出現する確率密度を  $p_j(y_j | \mathbf{x})$  とおけば、 $n$  回の独立な測定値  $\mathbf{y}$  が出現する同時確率密度は、

$$L(\mathbf{y} | \mathbf{x}) = \prod_{j=1}^n p_j(y_j | \mathbf{x}) \quad (1.2)$$

と書くことができる。この量は本来確率変数である測定値  $\mathbf{y}$  の関数であるが、 $\mathbf{y}$  を固定してパラメ

ータ  $\mathbf{x}$  の関数とみたとき、尤度 (likelihood) とよび、パラメータの値に対する「もっともらしさ」をあらわすものとする。最尤法とは、尤度を最大にするパラメータを求める方法である。

測定値の誤差  $\varepsilon_j$  が、分散  $\sigma_j^2$  の正規分布 (Gauss 分布) に従うならば、

$$p_j(y_j | \mathbf{x}) = \frac{1}{\sqrt{2\pi} \sigma_j} \exp\left[-\frac{(y_j - f_j(\mathbf{x}))^2}{2\sigma_j^2}\right] \quad (1.3)$$

であるから、最尤法は結局、

$$M(\mathbf{x}) = \sum_{j=1}^n \left[ \frac{y_j - f_j(\mathbf{x})}{\sigma_j} \right]^2 \quad (1.4)$$

を最小にする  $\mathbf{x}$  を求めることに帰着する。これを最小 2 乗法という。最小 2 乗法とは、誤差が正規分布に従う場合の最尤法である。(1.4) 式の  $M$  は、 $\chi^2$  分布に従い、 $\chi^2$  検定に用いられるので、カイ 2 乗とよばれることも多い。

## 2. 現実のデータ

データ解析の教科書の中には、測定誤差が正規分布に従うことを、頭から前提しているものが多い。しかしこの前提の根拠は薄弱である。中心極限定理による説明はあまり意味をもたない。正規分布は、いわば「完全に管理された一種の理想状態を定式化したもの」<sup>3)</sup> とみるべきであろう。実際、曲線のあてはめ (curve fitting) などを 1 度でも試みたことのある人なら、確率的にはまずほとんどあらわれるはずのない、 $3\sigma$  も  $4\sigma$  も離れた点が意外に出てきて、当惑した経験をもっているであろう。

このようなデータに対し、不用意に最小 2 乗法を適用するとどうなるであろうか。当然のことながら、これらの離れた点は  $M$  に大きく寄与し、

$M$  すなわちカイ 2 乗の値は非常に大きくなる。そればかりでなく, fit そのものがそれに引きづられて乱れてしまう。

なぜ実際のデータは正規分布から外れるのであろうか。一つの理由は分解能である。測定の誤差には測定器の分解能が重なっていることが多く, 分解能はしばしば Cauchy 分布 (Lorentzian ともいう),

$$p(x) = \frac{\Gamma}{\pi} \cdot \frac{1}{(x-x_0)^2 + \Gamma^2} \quad (2.1)$$

で近似される。ここで  $x_0$  は分布の中心,  $\Gamma$  は半値幅である。Cauchy 分布は正規分布よりはるかに長いすそ (tail) をもち, 割に大きな値の誤差がしばしば出現する。もちろん, 技術的によく管理された測定においては, Cauchy 分布ほどすその長い分布は現実的ではないが, 一つの極端な例として考えることができる。もし Cauchy 分布よりさらにすその長い分布が実際にあらわれたら, 測定の過程に何か不適切な操作がなかったかどうか検討すべきであろう。

この他, なんらかの理由による系統誤差もありうるし, 考えている模型の不完全さに由来する誤差もある。現実のデータには, 測定上の不備による誤った測定値や, パンチミスの可能性さえ排除できない。

### 3. ロバスト推定法

このように, 現実の問題では種々の誤差が混っているばかりか, 誤差分布の形が正確には知られていないことが多い。このような場合にどうしたらよいであろうか。

一つの方法は, データから分布の形を推定することである。つまり分布の形(とくにすその長さ)を適当な形でパラメトライズし, データを見てからもっとも適当な分布を選び, それから推定を行なうこのような方法を一般に適応型 (adaptive) 推定という。たとえば, Roos らは, 素粒子データを“平均”するのに, 自由度  $N$  の Student 分布  $S_N(x)$  を仮定して最尤法を用いることを提案して

いる<sup>5)</sup>。Student 分布は,  $N=1$  では Cauchy 分布となり,  $N \rightarrow \infty$  の極限では正規分布に帰着するので, 両者の内挿公式とみることもしできる。かれらは素粒子データ数百例を分析して, その誤差分布が,  $N=10$  の Student 分布で近似できることを見出した。

しかしこのような方法はデータ数が充分大きくなければ適応できないし, 計算量も莫大になるのであまり実用的ではない。これに対し, 誤差分布の形を厳密に考えずに, あらかじめ仮定した形から少しずれた場合にも, その影響をできるだけ受けにくい推定法という考え方がある。これを適応型推定に対して単純 (ロバスト) 推定とよぶ。たとえば, よく管理された誤差分布  $p_0(x)$ ——普通は正規分布を考える——に, 測定器の操作ミス, 誤ったデータ処理, 模型の不完全さなどに由来する粗大誤差  $q(x)$  を小さな確率  $\alpha$  で取り入れたものの, すなわち,

$$p(x) = (1-\alpha) p_0(x) + \alpha q(x)$$

を考える。これは Huber のモデルといわれている<sup>6)</sup>。このとき  $q(x)$  の存在の影響をあまり受けにくい推定法を, ロバスト (robust) であるという。

ロバスト推定法は, 大きくわけて 3 種のタイプがあり,  $M$  推定 (maximum-likelihood-type estimation),  $L$  推定,  $R$  推定とよばれる<sup>6)</sup>。ここでは  $M$  推定についてだけ考える。 $M$  推定とは, その名のとおりの最尤法を一般化したものである。尤度を,

$$L(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^n p_j(y_j - f_j(\mathbf{x})) \quad (3.2)$$

とおけば, 最尤法は,

$$\begin{aligned} -\log L(\mathbf{y}|\mathbf{x}) &= \sum_{j=1}^n -\log p_j(y_j - f_j(\mathbf{x})) \\ &= \min \end{aligned} \quad (3.3)$$

とあらわされる。 $M$  推定法とは, これを一般化して,

$$\sum_{j=1}^n \phi_j(y_j - f_j(\mathbf{x})) = \min \quad (3.4)$$

となる  $\mathbf{x}$  を求める方法であり,  $\phi$  の取り方によ

り種々の方法が考えられる。最小2乗法は、

$$\phi_j(y_j - f_j(\mathbf{x})) = (y_j - f_j(\mathbf{x}))^2 / \sigma_j^2 \quad (3.5)$$

と取ったことに対応するから、規格化された残差  $v_j = (y_j - f_j(\mathbf{x})) / \sigma_j$  が正規分布  $p_0(x)$  からみて適当な範囲に入っている場合には、 $v$  の関数として共通に、

$$\phi(v_j) \approx v_j^2 \quad (3.6)$$

であることが望ましい。

式(3.4)を直接解くには、線形の問題でも多変数関数の最適化プログラムが必要であり、計算は容易でない。そこで、最小2乗法の  $\phi$  との比を、

$$w(v) = \phi(v) / v^2 \quad (3.7)$$

とおけば、(3.6)より、 $|v|$ の小さい領域では  $w \approx 1$  である。ロバストにするには大きな  $v$  に対し、 $w$  を小さくとればよい。すると(3.4)は、

$$\tilde{M}(\mathbf{x}) = \sum_{j=1}^n (y_j - f_j(\mathbf{x}))^2 / \sigma_j^2 \cdot w(v_j) = \min \quad (3.8)$$

とあらわせ、(1.4)式で  $\sigma^{-2}$  を  $w \cdot \sigma^{-2}$  に置きかえたものと形式的に同じである。もちろん  $w$  は  $v$  の関数であるが、大部分のデータに対しては  $w \approx 1$  であることが期待されるので、iterative に解くことができる。すなわち  $w$  を、1回前の  $v_j = y_j - f_j(\mathbf{x})$  の値に対する  $w(v_j)$  に固定して(3.8)を最小2乗法プログラムで解き、その解にもとづいて  $w$  を修正しもう1度解く。これを何回か繰り返し、 $w$  の値が変化しなくなったところで止める。この  $w$  のことを、「調節重み」(adjustable weight) とよぶことにする。

$M$  推定法では、一部の測定値の重みを落としてあてはめるので、理想に近いデータに対しては最小2乗法より効率が少し落ちる(つまり決定したパラメータの分散が大きくなる)。また調節重みの修正という別の繰り返し加わるので計算時間はそれだけ増大する。しかしこれらは、粗大誤差や非正規誤差が万一含まれていた場合に備えての保険料のようなものであろう。

#### 4. SALSにおけるロバスト推定法

著者らは、自然科学におけるデータ解析のための、最小2乗法汎用標準プログラム SALS (Statistical Analysis with Least Square fitting の略)の開発を、東大大型計算機センターおよび科学研究費丘本班の協力のもとに進めている<sup>6)</sup>。SALS にはいくつかの特徴があるが、その一つは上に述べたロバスト推定法の考えを大幅に取り入れたことである。

SALS で採用したロバスト推定法は、 $M$  推定法の中で、Biweight 推定法および Huber の推定法とよばれるものである<sup>6)</sup>。

Huber の推定法では、 $w(v)$  として、

$$w(v) = \begin{cases} 1 & |v| \leq c \\ c/|v| & |v| > c \end{cases} \quad (4.1)$$

とおく。 $|v| \leq c$  では通常最小2乗法と同様に振舞い、 $|v| \geq c$  では絶対値の和を最小にする方向に働く。Huber は、モデル(3.1)に対し、この  $w$  の取り方が最適であることを議論している。

一方 Biweight 法では調節重みを、

$$w(v) = \begin{cases} [1 - (v/c)^2]^2 & |v| \leq c \\ 0 & |v| > c \end{cases} \quad (4.2)$$

とする。これは図1のように、 $|v|$ が小さい時は1に近く、 $|v|$ が大きい時は0となり、中間では連続的に変化している。

ここで問題になるのは  $c$  の取り方であるが、SALS では  $c$  を固定せず、 $|v|$  の median を規準としてその5~10倍をとることにしている。これは fit の途中段階では  $c$  を大きめにとり、残差全体が減少するにつれて  $c$  を小さく取るようになっていく。

#### 5. ロバスト推定法の使用例

ここで、ロバスト推定法の実例として、2変数の非線形モデルに対し、SALS の重み調節機能を用いたシミュレーションの結果を伊藤氏の報

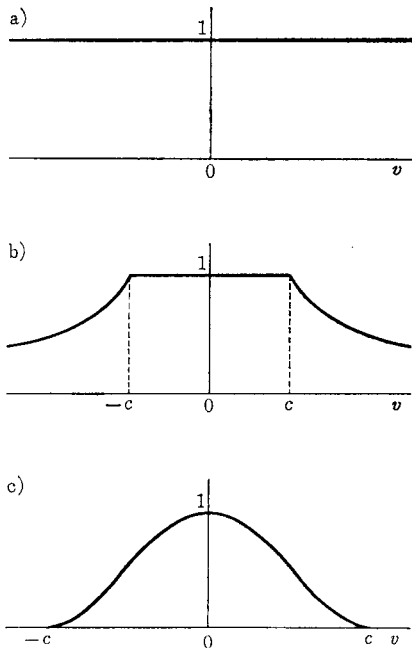


図1 調節重み  $w$ . a) 最小2乗法 b) Huber 法  
c) Biweight 法

告<sup>8)</sup>から引用したい。これは SALS の簡略版である SALS-MINI システム<sup>9)</sup>によって実行したものである。

モデル関数としては、sine 関数、

$$f_j(x_1, x_2) = x_1 \sin(2\pi x_2 q_j) \quad (5.1)$$

を取り<sup>2)</sup>、パラメータの真値としては  $x_1=2$ ,  $x_2=0.8$  と仮定し、初期値  $x_1=1.5$ ,  $x_2=0.75$  から計算した(図2)。横軸(制御変数)  $q_j$  は  $0 \leq q \leq 5$  の間の一様乱数によって与え、測定データ  $y_j$  は、真値に乱数誤差を加えてつくった。測定誤差  $\sigma_j$  は、0.2 (一定) と仮定した。

非線形最小2乗法としては、SALS-MINI システムの Gauss-Newton 法を用いた。線形部分の解法は Householder 法である。

### (1) 正規誤差の場合

乱数として、平均0、標準偏差0.2の正規乱数を用いて測定データをつくった。データ数30のデータを10組つくり計算した。パラメータ最確値の分散と偏り、および収束に要した、重み調節ならびに Gauss-Newton 法のサイクル数は表1のと

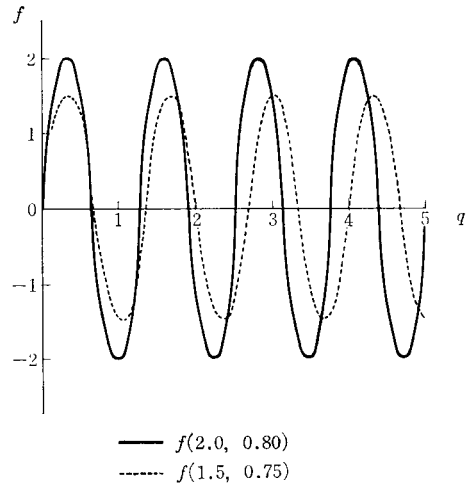


図2 真値(実線)と出発値(点線)におけるモデル関数

おりである。分散と偏りは、最小2乗法、Huber の推定法、Biweight 法で有意の差がない。

### (2) Cauchy 誤差の場合

乱数として、平均0、半値幅0.2のCauchy乱数を用いて測定データをつくった。表2で見ると、分散、偏りとも Biweight 法が抜群に近い。Huber 法は、最小2乗法と Biweight 法の中間である。図3に、最小2乗法と Biweight 法を用いた場合の残差  $y_j - f_j(\hat{x})$  のプロットを示

表1 正規誤差の場合

(a) パラメータの最確値と偏り

	$x_1$		$x_2$		$\chi^2$		$\tilde{\chi}^2$	
	LS	BIW	LS	BIW	LS	BIW	BIW	
10組の平均値	2.02	2.02	0.8002	0.8002	29	30	26	
10組の標準偏差	0.06	0.06	0.0011	0.0012	7	7	6	
真値からの偏り	0.02	0.02	0.0002	0.0002	(1)	(2)		

LS: 最小2乗法 BIW: Biweight 法

HUB: Huber 法 (この場合 LS と同じなので表から省いた)  $\tilde{\chi}^2$  は、(3.8)式の  $\tilde{M}$  のことである。

(b) 収束に要したサイクル数(上記第1組の例)

	重み調節 サイクル	Gauss-Newton 法 のサイクル	合計
L S	1	3	3
BIW	3	3 + 1 + 1	5

表 2 Cauchy 誤差の場合

(a) パラメータの最確値と偏り

	$x_1$			$x_2$		
	LS	HUB	BIW	LS	HUB	BIW
10組の平均値	2.08	1.98	1.99	0.784	0.790	0.798
10組の標準偏差	0.76	0.58	0.08	0.027	0.011	0.003
真値からの偏り	0.08	0.02	0.01	0.016	0.010	0.002

(b) 収束に要するサイクル数 (第1組の例)

	重み調節 サイクル数	Gauss-Newton のサイクル数	合計
LS	1	6	6
HUB	4	6 + 3 + 3 + 1	13
BIW	4	4 + 3 + 2 + 1	10

す。Biweight 法では 2~3 個のデータに対し、調節重み  $w$  の値が 0 となっている。

### (3) 異常値の影響

パンチミスなどで起こりうる異常に飛び離れた測定値 (outlier) を含む場合を数例試みた。20 個の正規乱数測定値 ((1) と同様) に、 $f_{21}=100$ ,

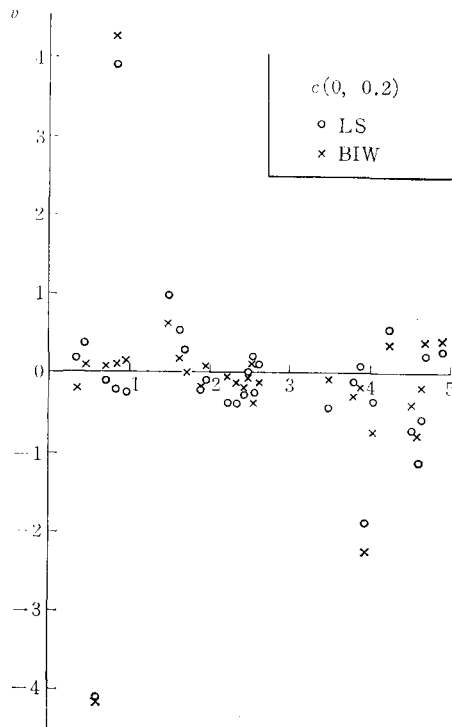


図 3 Cauchy 誤差の場合の残差分布。(本図では  $v_j = y_j - f_j(\hat{x})$  で、(3.7) の  $v$  とは異なる。)

$f_{22}=1000$  という測定値を加え、計算を行なった。いうまでもなく最小 2 乗法は、outlier に引かれて発散したが、Biweight 法では直ちに  $w_{21}=w_{22}=0$  となって収束し抵抗力は抜群であった。Huber 法では重み調節 4 サイクルの後、 $w_{21}=0.041$ ,  $w_{22}=0.004$  となり、パラメータの最確値としては  $\hat{x}_1=1.92$ ,  $\hat{x}_2=0.808$  を得た。かなり抵抗力はあるものの、 $w(v)$  の減少の仕方がゆるやかなのでどうしても影響は残る。

このテストでは Biweight 法が総合的にすぐれた性質をもつことがわかった。このような数値実験はいろいろ条件を変えて種々の問題に対してなされるべきである。

## 6. 問題点

ロバスト推定法の最大の問題点は、理論的な模型に多大の信頼をおいていることである。模型の予言から誤差の数倍以上離れている測定値はほとんど無視してしまうが、もしかしたらその点こそ重要な意味をもっているかもしれない。使い方を誤ると、データの特徴を示す大切な部分 (たとえばピーク) の重みを 0 とおいて、残りに fit してしまうかもしれない。ロバスト推定法の結果は、そのままのみにすべきものではなく、「診断」のための資料を与えるものと考えるべきである。データ、模型、初期値に対する十分な吟味が大切である。

## 文 献

- [1] W. T. Eadie et al.: *Statistical Methods in Experimental physics*, North Holland Pub., 1971.
- [2] S. Brandt: *Statistical and Computational Methods in Data Analysis*, 邦訳: 吉城, 高橋, 小柳訳「データ解析の方法」みすず書房, 1976.
- [3] 三浦良造: 「ロバスト推定法」シンポジウム「自然科学のためのデータ解析」報告集 1976, p. 90.
- [4] 中川徹, 朽津耕三: 「分光データ処理のための数学的手法, あてはめ法——その理想と現実」分光研究 24(1975), 109, 165.

- [5] M. Roos et al. : *Physica Finica*, **10** (1975)  
21. Particle Data Group : *Rev. mod. Phys.* **48**  
(1976), S 1.
- [6] 中川徹 : 「最小 2 乗法標準プログラムの開発」東  
大大型計算機センターニュース, **8**(1976), No. 5, 68;  
No. 6, 89.
- [7] 小柳義夫 : 「最小 2 乗法における新しい手法」応  
用物理, **46** (1977), 55.
- [8] 伊藤徹三 : 「SALS における重み調節使用例」丘  
本班研究会 “統計的データ解析と統計プログラムパッ  
ケージ” (1978).  
本節のデータはすべて伊藤氏より提供されたもので  
ある.
- [9] 小柳義夫 : 「最小 2 乗法標準プログラム (SALS)  
の開発」シンポジウム “自然科学のためのデータ解  
析” 報告集 (1976), p. 129.

おやなぎ・よしお 1943年生  
1966年 東京大学理学部物理学科卒  
1971年 同大学院博士課程修了  
現 在 高エネルギー物理学研究所勤務  
専 攻 素粒子論

## 支部ニュース

### 九州支部

1年後に九州地区での春季大会を控え、そろそろ準備にとりかからねばと思っていますが、大会に関する皆さまへの協力依頼は次回ということにしまして、今回は52年度下期の支部活動状況を報告させていただきます。

- 講演会：製鉄所の経営企画における管理技法の適用について(52.10.18)新日鉄 亀沢善一郎氏
- 研究会：多段工程、複数製品のバッチ製造ラインにおける最適設備能力、最適ロットサイズ決定の近似解法(52.11.15)

三菱化成 長田 博氏

- 講演会：目視検査の構造分析とその最適化  
新日鉄 岡本久人氏

最適選択の諸方式：Play the Winner Rule につ  
いて(53.1.17) 熊本女子大 城島邦行氏

- 研究会：連続モデルを利用した待行列問題の近似解法(53.2.14) 九大 S. K. ビスワス氏
- 運営委員会：支部総会議案、春季大会の運営方法等に関する審議(53.2.22) (支部事務局 迫)