

講座——物理学周辺の確率統計——

ロバスト推定法とデータ解析

小 柳 義 夫

「足して二で割る」のは日本人のお家芸である。同じ量を数回測定したとき、我々は何気なく平均をとる。いったい平均には何の根拠があるのだろうか。一つだけ飛び離れた値があった場合にどうするか。勝手に一つのデータを除いて平均してもよいだろうか。それとも主観的判断を避けるためにあくまで全体の平均を取るべきか。

このような問題に答えるのがロバスト推定法という考え方である。本稿ではロバスト推定法の意味について議論するとともに、ロバスト推定法を組み込んだ「最小二乗法標準プログラム SALS」を紹介する。

§1. はじめに

私が統計学なるものを初めて学んだのは大学の教養課程で、経済学の先生からであった。社会科学の科目の中では比較的なじみやすく、まあそのうちに必要になるだろうと思って取ったことを覚えている。何やら分布とか何やら検定とかが次々に出て来て、正直なところあまりおもしろくなかった。

多くの物理屋にとっても事情は似たり寄ったりであろう。何かを測定したり推定したりすれば必ず「誤差」が問題になるし、90%信頼度で上限がいくらなどと日常的に議論しているわけであるから、統計学は物理屋と無縁ではない。従って社会科学向きとはニュアンスの違った、物理屋向けの確率統計^{1,2)}があってもよいであろう。「サイコロ、ルーレット、くじ引き…」が第1ページに出てくる統計学はもう御免である。私はこのシリーズに期待している。

§2. 最小二乗法

世界方程式というようなものがある。第一原理から森羅万象を説明してくれるなら物事は簡単であるが—ただし物理屋は全員失業することになる—物理現象を説明するには、多くの未知パラメータを含む何らかの「モデル(模型)」を用いざるを得ない。これらのパラメータは、第一原理から予言されることはまれ

で、多くの場合、実験データから決定しなければならない。ひとたびパラメータの値が決定されると、さらに別な現象に関する予言を与えることができ、次にはそれを実験で検証する。実証科学である物理学は、こうして一步一步自然の理解を深めていく。

しかし測定値には誤差がつきものであるから、求めたパラメータの値にも誤差が影響を与える。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。

データを解析するには、モデルの概念を明確にすることが重要である。モデルという場合には、単に研究の対象とする物理的なモデルだけではなく、誤差のモデル、つまりどのような誤差が含まれているかということまで考慮に入れなければならない。すなわち、データの数値が得られるまでのあらゆる過程を総合的に考察することが、よいデータ解析の条件である。

ここで一組の測定値 $y_i (i=1 \sim m)$ を考えよう。これは、未知のパラメータ $x_j (j=1 \sim n)$ を含むモデル関数に、誤差 ε_i が加わったものと考えることができる。すなわち

$$y_i = f_i(x_1, \dots, x_n) + \varepsilon_i. \quad (2.1)$$

f_i は既知の関数であり、 x_j に対し線形の場合も、非線形の場合もある。 y_i からどのように x_j を求めたらよいか、モデル f_i は正しくデータ y_i を記述しているか、これが本稿の課題である。

誤差 ε_i がどのような分布に従っているかが既知の

場合は最尤法 (maximum likelihood method) とよばれる方法を用いるのが普通である。

以下ベクトル記法を用いることにする。パラメータ x が与えられたとき、測定値 y_i が出現する確率密度を $p_i(y_i|x)$ とおけば、 m 回の独立な測定値 y が出現する同時確率密度は、

$$L(y|x) = \prod_{i=1}^m p_i(y_i|x) \quad (2.2)$$

と書くことができる。これは一種の事後的確率であり、測定結果が知れてから、 y がその値をとる確率がいくらであったかを示すものである。 y は確率変数 (ある分布に従ってランダムな値を取る変数) であるから、その関数である L も確率変数である。他方見方を変えて、 y をある特定な測定値に固定し、 L をパラメータ x の関数と見たとき、これを尤度 (likelihood) と呼ぶ。尤度は、測定値が与えられたときの、パラメータの値に関する「もっともらしさ」を表すと考えられる。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 (最尤法の情報エントロピー最大化原理としての裏付けは、文献 3 参照)。

測定値の誤差 ε_i が、分散 σ_i^2 の正規分布 (ガウス分布) に従うならば、

$$p_i(y_i|x) = \frac{1}{\sqrt{2\pi} \sigma_i} \exp \left[-\frac{(y_i - f_i(x))^2}{2\sigma_i^2} \right] \quad (2.3)$$

であるから、最尤法は結局

$$M(x) = \sum_{i=1}^m \left[\frac{y_i - f_i(x)}{\sigma_i} \right]^2 \quad (2.4)$$

を最小にする x を求めることに帰着する。これを最小二乗法という。つまり、最小二乗法とは、誤差が正規分布に従うときの最尤法である。 $M(x)$ の最小値は χ^2 分布に従い、 χ^2 検定に用いられるので、カイ二乗推定法と呼ばれることも多い。

ある量 x を、同一精度で直接測定したデータを $y_1, \dots, y_m$ とすると、

$$f(x) = x$$

であるから、最小二乗法は推定値 $\hat{x}$ として、

$$\hat{x} = \sum_{i=1}^m y_i / m$$

を与える。すなわち、平均とは誤差が正規分布に従う場合の、同精度直接測定に対する最尤推定値である。

§3. 測定誤差の分布

前節で述べたことからすれば、最小二乗法という統

計的手法は、すでに確立された手法のように思われるかも知れない。しかしそれは必ずしも正しくない。最も問題になるのは、誤差が正規分布に従うという仮定である。

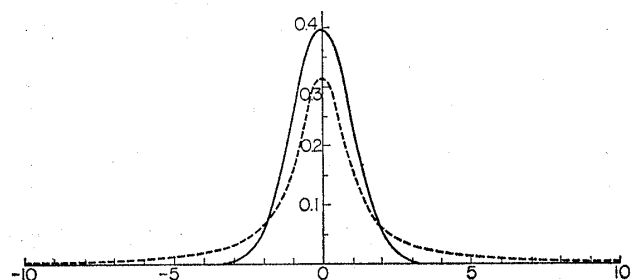
正規分布は何か特別の意味を持つものであろうか。しばしば正規分布は中心極限定理によって説明されている。すなわち、小さな誤差が充分多く重ね合わさると、各々の誤差がどんな分布をしていても、全体は正規分布に近づくというわけである。具体的なモデルとしては、誤差に関する Laplace のモデル等があげられる。しかしこの前提が成立するのは、モデル $f_i(x)$ が充分正しく現象を記述し、かつ測定上の不備や系統誤差、データ作成時のまちがいなどが一切含まれていない場合である。データ解析の途中においては、正しいモデルが見出されているとは限らず、測定値にも種類の不備があることが当然予想される。

つまり、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」とみるべきである。正規分布の特徴は、非常にすそが短いことである。たとえば $|\varepsilon_i| > 3\sigma_i$ の誤差が出現する確率はわずか 0.3% である。現実の測定値の分布としては、正規分布よりすそその長い分布を想定するのが妥当であろう。

すその長い分布の極限と考えられるのが、コーシー (Cauchy) 分布である。物理屋には、ローレンツ分布、またはブライト・ウィグナー型分布といった方がなじみやすいかも知れない。これは

$$p(x) = \frac{\gamma}{\pi} \frac{1}{(x - x_0)^2 + \gamma^2}$$

という分布関数をもつ。ここで x_0 は分布の中心、 γ は半値幅 (片側) である。 $p(x)$ は $x \rightarrow \infty$ で x^{-2} というゆっくりした速さで減少する (第 1 図)。コーシー分布のすそがいかに長いかは、平均を求める積分



第 1 図 正規分布 (実線) とコーシー分布 (破線)。

$$\int_{-\infty}^{\infty} xp(x)dx$$

さえ収束しないことからわかる。

一般に、技術的によく管理された測定においては、コーシー分布ほどすその長い誤差分布はありえず、コーシー型もしくはこれ以上にすその長い分布が現れた場合は、測定の過程をチェックし直すことが重要である。

§4. 最小二乗法の欠点

このようなデータに対し、不用意に最小二乗法を適用するとどうなるであろうか。すその長い分布に従う誤差には、当然大きな値も数多く出現する。コーシー分布では、半値幅 γ の10倍以上の誤差の現れる確率は、実に $(2/\pi) \tan^{-1}(1/10) \approx 6\%$ もある。

例として、§2で述べた等精度直接測定に対する最小二乗推定である単純算術平均を考えよう。算術平均では、すべての測定値は推定値に対し平等に $1/m$ の寄与を持っているから、一つでも飛び離れた測定点があれば、その値に比例して推定値が大きくなってしまふ。実際、誤差がコーシー分布に従う測定値の平均は、 m がいくら大きくなっても真値に収束せず、かえって発散することが知られている。¹⁾

同じことが一般の最小二乗問題についても言える。例えば、データ中に飛び離れた測定値が一つ含まれていたとする。このデータに対して最小二乗法を適用するとどうなるか。残念なことに、飛び離れた値ほどパラメータの決定に大きく影響し、あてはめ全体がこの1点のために大きく引きずられることになる。後の実例では、それがはっきりと現れている。

パラメータの決定の良さを表すのに、求めたパラメータの分散を目安とすることができる。つまり、測定誤差がある分布に従うとき、ある方法でパラメータを決定すると、真値からどの程度ずれるかを示すものである。統計学の教えるところによれば、Cramér-Rao の情報不等式により、推定したパラメータの分散には下限がある。すなわち、どんな方法を用いても、分散をある値より小さくすることはできない。この下限を実限する方法が最もよいパラメータの決定法である。誤差が正規分布に従う場合は、最小二乗法が最小の分散を持つ推定法である。

先程の等精度直接測定の問題を考える。平均値の他

にも推定の方法は色々ある。その一例が、中央値 (median) である。これはデータを大きさの順に並べたときの中央のデータ値 (偶数なら中央の2個の平均) のことである。中央値は、その定義からわかるように、2個や3個飛び離れた値があってもほとんど影響を受けない。その反面、データの数値としては中央の1~2個のみを用い、他は順序づけのためにしか用いていないので、データの情報を十分に生かしていない面がある。例えば、正規誤差に従う測定値から中央値を用いて推定したときの推定値の分散は、平均値の分散に比べ、(データ数が充分大きい極限で)、 $\pi/2 = 1.57$ 倍大きいことが知られている。¹⁾ 1よりは大きいだが、思ったほど大きいわけではない。他方、誤差がコーシー分布に従う場合には、中央値の分散は、最小の分散 (これはコーシー分布を仮定した最尤法によって実現される) に比べて、わずか1.22倍大きいだけである。コーシー分布に対し“平均”が発散してしまうのとは雲泥の差といわなければならない。このように、分布の形が正確には知られていない場合にも有効な推定法を、ロバスト推定法と呼ぶ。中央値は平均値よりもロバストであるといえることができる。

§5. ロバスト推定法

現実のデータ解析においては、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、データと比較して最も適当な分布を選び、それから推定を行う。このような方法を一般に適応型 (adaptive) 推定という。Particle Data Group の Roos らは、素粒子データを“平均”するのに、自由度 N のスチューデント分布 $S_N(t)$ を仮定し、最尤法により“平均値”と N とを同時に決定する方法を提案している。スチューデント分布は、 $N=1$ ではコーシー分布となり、 $N \rightarrow \infty$ 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 $N=10$ のスチューデント分布で近似できることを見出した。⁴⁾

このような方法はデータ数が充分大きくなければ用

いられないし、最尤法を直接解くには一般の多変数最小化法が必要であり、計算量も莫大になるのであまり実用的でない。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、その影響ができるだけ小さくなるような推定法が、(単純)ロバスト推定法である。

ロバスト推定法には、大きくわけて3種のタイプがあり、M推定、L推定、R推定と呼ばれる。⁶⁾ここでは物理などのデータ解析に関連の深いM推定についてだけ触れる。M推定のMはmaximum likelihoodの意で、その名の通り最尤法を一般化したものである。尤度を

$$L(y|x) = \prod_{i=1}^m p_i(y_i - f_i(x)) \quad (5.1)$$

とおけば、最尤法は

$$-\log L(y|x) = \sum_{i=1}^m -\log p_i(y_i - f_i(x)) = \min \quad (5.2)$$

と表される。M推定法とは、これを一般化して

$$\sum_{i=1}^m \phi_i(y_i - f_i(x)) = \min \quad (5.3)$$

となる x を求める方法である。 ϕ の取り方によって種々の方法が考えられる。最小二乗法は、

$$\phi_i(y_i - f_i(x)) = (y_i - f_i(x))^2 / \sigma_i^2 \quad (5.4)$$

と取ったことに対応している。等精度直接測定の場合に

$$\phi_i(y_i - \bar{y}) = |y_i - \bar{y}| / \sigma \quad (5.5)$$

と取れば、 ϕ_i の和を最小にする $\bar{y}$ は y_i の中央値である(証明は読者に任せる)。

我々の場合、大部分のデータは正規分布に従うことを期待しているので、規格化された残差 $v_i = (y_i - f_i(x)) / \sigma_i$ が正規分布からみて適当な範囲に入っている場合には最小二乗法と同じ形、すなわち v の関数として

$$\phi(v_i) \approx v_i^2 \quad (5.6)$$

であることが望ましい。最小二乗法の欠点は飛び離れた値の影響が大きすぎる点にあるので、 v が大きくなるにつれて、 $\phi(v)$ を v^2 より小さくすればよい。 ϕ と v^2 との比を

$$w(v) = \phi(v) / v^2 \quad (5.7)$$

とおけば、 $|v|$ の小さい領域では $w \approx 1$ であり、 $|v|$ の大きなところでは w を0に近づければよい。

式(5.3)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、計算は容易でないが、大部分のデータに対して w はほぼ一定(≈ 1)であることが期待できるので、反復法によって解くことができる。すなわち、(2.4)のかわりに

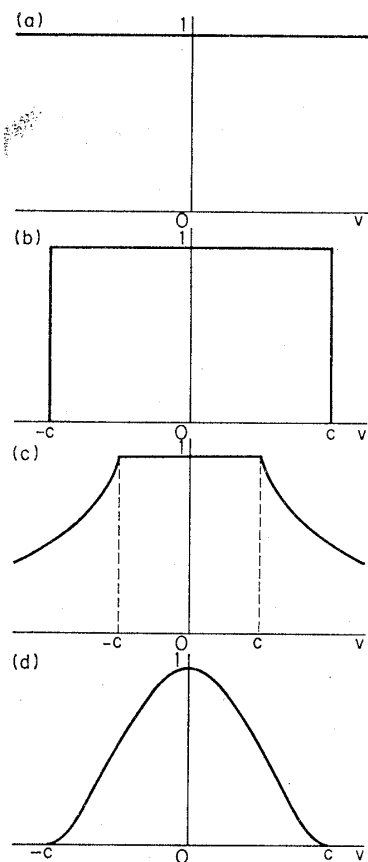
$$M(x) = \sum_{i=1}^m [(y_i - f_i(x)) / \sigma_i]^2 w(v_i) \quad (5.8)$$

を最小化すればよいわけであるが、 $w(v_i)$ としては1回前の v_i の値に対する $w(v_i)$ に固定して(5.8)を最小二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打ち切る。この w のことを「調節重み」と呼ぶことにする。

§6. 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、最小二乗法標準プログラムシステム SALS (statistical analysis with least square fitting) 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きた。⁶⁾ 1979年4月には、SALS第2版が完成し、5月からは東大センターのライブラリー・プログラムとして登録された他、⁷⁾ 各地のセンターでも使われ始めている。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。ロバスト推定法については、統計学の専門家の間にも種々の議論があり、その評価も定まったわけではないが、後に示すように有効に働く場合も多いので、種々の問題に対して経験を重ねて行きたい。

我々の基本的な考え方は、ロバスト推定法をデータ、モデル、初期値等の「診断」の道具として使うということである。ロバスト推定法は、モデルの予言値から大きく飛び離れた測定値を無視する方向に働くが、本当に無視してよいかどうかはユーザが判定すべきものである。ロバスト推定法は、これこれの測定点について、測定値かモデルかに異常があることを警告するが、それを判断するのは人間である。パンチミス、測定器の故障、調整不良、誤った処理の可能性もあるし、模型の不完全さ、未知の新現象かも知れない。もちろん単なる統計的ゆらぎのこともある。いずれにせよ、ロバスト推定法の結果はそのまよう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べきである。



第2図 調節重み。(a) 単純最小二乗法, (b) カット付最小二乗法, (c) Huberの方法, (d) biweight法。

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の推定法とよばれているものである。この他に、最小二乗法の変形として、規格化残差が一定以上大きな測定値の重みを0とする「カット付最小二乗法」を用いることもできる。⁷⁾

biweight法では $w(v)$ として、

$$w(v) = \begin{cases} [1 - (v/c)^2]^2 & |v| \leq c \\ 0 & |v| > c \end{cases} \quad (6.1)$$

とおく。これは第2図のように、 $|v|$ 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。

Huber法は、

$$w(v) = \begin{cases} 1 & |v| \leq c \\ c/|v| & |v| > c \end{cases} \quad (6.2)$$

とする。 $|v| \leq c$ では通常の最小二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。

biweight法と異なり、大きな $|v|$ に対し、調節重み w は、小さくはなるが決して0にはならない。

カット付最小二乗法は、

$$w(v) = \begin{cases} 1 & |v| \leq c \\ 0 & |v| > c \end{cases} \quad (6.3)$$

ととったものに対応している。

ここで問題となるのは c の取り方であるが、SALSでは c を固定せず、 $|v_i|$ の中央値を規準にして、biweight法とHuber法では10~100倍、カット付最小二乗法では100~1,000倍を取っている（ユーザが指定することもできる）。これは、あてはめの途中段階では c を大きめにとり、残差の大半が減少するにつれて c を小さく取るようになっていく。 $|v_i|$ の平均ではなくて中央値を取ったのも、一種のロバストネスを考慮したためである。

v_i が標準偏差 σ の正規分布に従う場合、 $|v_i|$ の中央値は 0.674σ であるから、その10倍とすると $c=6.74\sigma$ であり、この中には99.999%以上のデータが含まれる。biweight法でも、97%以上のデータの調節重みは0.8以上である。

§7. 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、5変数の非線形モデルに対しSALSを用いたシミュレーションの結果を示す。モデル関数は

$$f(q) = \frac{h\Gamma^2}{(q-q_0)^2 + \Gamma^2} + a_0 + a_1 q \quad (7.1)$$

で、1次のバックグラウンドの上に、ロレンツ型のピークが乗っている。ただし(括弧内は真値)

$$q_0 \quad \text{ピークの位置} \quad (2.5)$$

$$\Gamma \quad \text{半値幅(片側)} \quad (1.8)$$

$$h \quad \text{ピークの高さ} \quad (5.0)$$

$$a_0 \quad \text{バックグラウンドの係数} \quad (3.0)$$

$$a_1 \quad \text{バックグラウンドの係数} \quad (0.1)$$

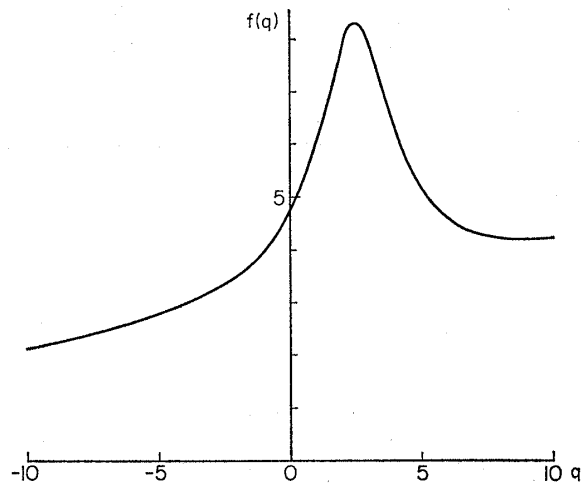
で、 $q=-10 \sim 10$ である(第3図)。

測定値は上記の q の区間に間隔0.04で51点を取り、真値を用いて計算した理論値に、正規乱数誤差およびコーシー乱数誤差を加えて、それぞれ10組作成した。

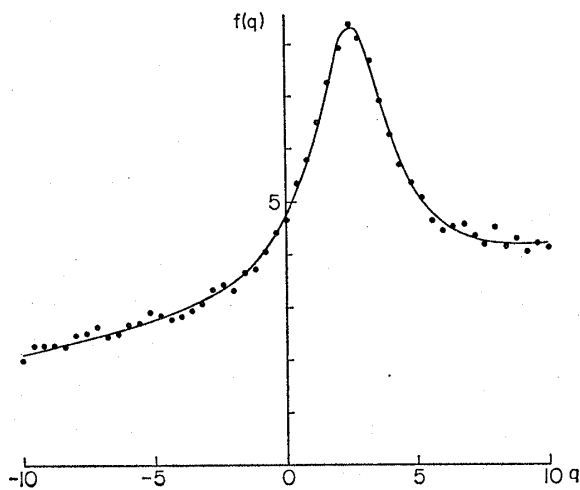
(i) 正規誤差の場合

標準偏差0.08の正規乱数を誤差として用いた。第1組のデータを第4図に示す。推定法としては単純最小二乗法とbiweight法とを用いた。第4図には、この

データに対する、あてはめたパラメターの値による理論値も書き入れてあるが、単純最小二乗法と biweight 法とはほとんど一致している。



第3図 真のモデル関数.

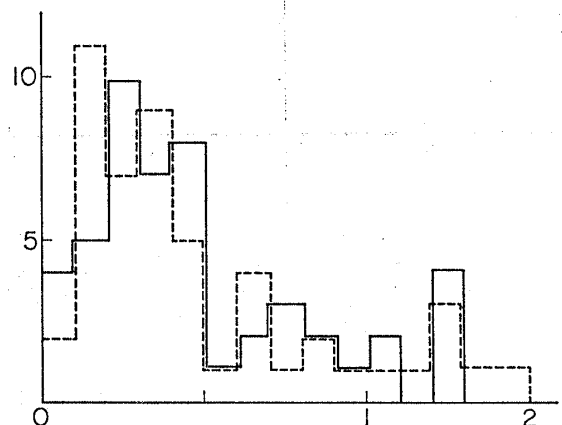


第4図 正規誤差データと、求めたパラメターによるモデル関数 (最小二乗法と biweight 法とは一致している).

第1表に、各パラメターについて、推定値の真値からのずれを、10組のデータについて二乗平均したものを示した。biweight 法でも、最小二乗法とほとんど同じ精度で推定できていることがわかる。ちなみに、最終段階においては、調節重みの合計は49~50(データ数は51)で、各重みはほとんど1に近い。正規誤差の場合、誤差の評価を2倍程度まちがえても、これによるパラメター推定値の分散への影響は高々 1/8 程度である(つまり偏差へは 1/16)から、この結果は当然である。

求めたパラメターの真値からのずれをヒストグラムの形で第5図に示す。ただし、ずれは、各パラメターの推定誤差でスケールしてある。これによっても、両者の推定法がほぼ同一の精度を与えていることがわかる。

反面ロバスト推定法は、重み調節の反復という余分の手続きが必要なので、計算時間が多く必要である。問題や初期値によっても異なるが、この場合では1組



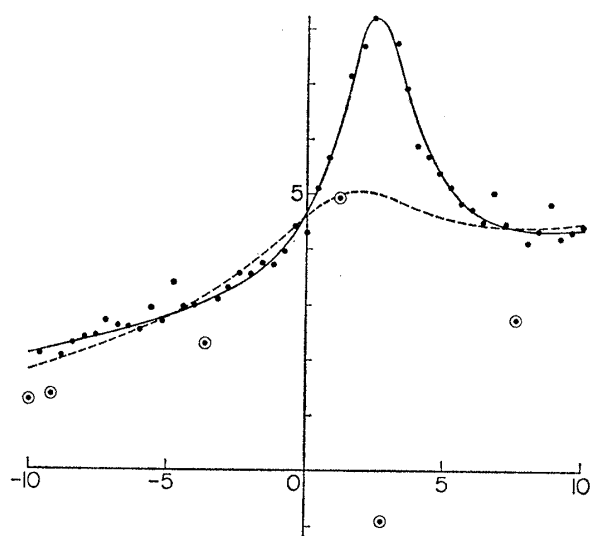
第5図 正規誤差データ10組から求めたパラメターの真値からのずれのヒストグラム (実線は biweight 法, 破線は最小二乗法). 横軸は各パラメターごとにスケールしてある.

第1表 正規誤差に対するパラメター推定値の真値からのずれ(二乗平均と最大値).

パラメター	真 値	ずれの平均	ずれの最大値	ずれの平均	ずれの最大値
ピーク位置 q_0	2.5	0.017	0.036	0.017	0.037
半 値 幅 Γ	1.8	0.014	0.027	0.016	0.029
ピークの高さ h	5.0	0.035	0.085	0.034	0.083
定 数 項 a_0	3.0	0.014	0.027	0.014	0.025
1 次 係 数 a_1	0.1	0.0025	0.0046	0.0025	0.0044
単純最小二乗法				biweight 法	

第2表 コーシー誤差に対するパラメータ推定値の真値からのずれ(二乗平均と最大値).

パラメータ	真 値	ずれの平均	ずれの最大値	ずれの平均	ずれの最大値
ピーク位置 q_0	2.5	0.444	1.11	0.019	0.037
半 値 幅 Γ	1.8	0.785	2.06	0.051	0.090
ピークの高さ h	5.0	1.121	3.04	0.064	0.120
定 数 項 a_0	3.0	1.088	3.39	0.038	0.070
1 次 係 数 a_1	0.1	0.1546	0.485	0.0042	0.0071
単純最小二乗法				biweight 法	

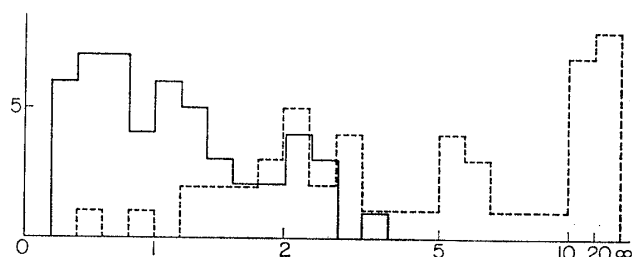


第6図 コーシー誤差データと, 求めたパラメータによるモデル関数(実線は biweight 法, 破線は最小二乗法). 丸で囲んだデータは, biweight 法で重みが 0 となった点.

のデータを解析するのに, 最小二乗法では平均0.47秒, biweight 法では平均0.77秒を要している(東大大型計算機センタ HITAC 8800 の CPU 時間による).

(ii) コーシー誤差の場合

乱数として, 半値幅(片側) 0.08 のコーシー乱数を用いて測定データをつくった. 第1組のデータおよび最小二乗法および biweight 法によってあてはめたグラフを第6図に示す. 測定値のうち丸で囲んだデータは, biweight 法で重みが 0 となったデータ点である. このグラフからもわかるように, biweight 法がほぼ良好な結果を与えているのに対し, 単純最小二乗法は完全に外れている. biweight 法なら重み 0 となるような飛び離れたデータ点に引きずられて, ピークの部分ではほとんどの点にも合っていない. これに対し,



第7図 コーシー誤差データ10組から求めたパラメータの真値からのずれのヒストグラム(実線は biweight 法, 破線は最小二乗法). 横軸は各パラメータごとに, 第5図と同じスケールを用いた.

biweight 法では, 大部分の点の近くを通る良好なあてはめになっている.

正規誤差の場合と同様, 各パラメータの推定値の真値からのずれの二乗平均を第2表に, また各パラメータの真値からのずれのヒストグラムを第7図に示す. スケールは正規誤差の場合と同じものを用いた. 横軸の絶対値は意味はないが, ロバスト推定法の有効性が一目瞭然であろう.

はじめにも述べたように, 実際にはコーシー分布ほどすその長い誤差は現れず, もし現れたら測定に何かの誤りがあったと考えてチェックを行うべきである. 測定の悪さを, 解析法で救うことはできない. しかし以上の数値実験の結果は, 一つの極端な場合として, 考慮に値するであろう.

§8. ロバスト推定法の問題点

ロバスト推定法の最大の問題点は, モデルおよびパラメータ初期値に多大の信頼をおいていることである. とくに biweight 法は, モデルの予言から誤差の数倍以上離れている点の重みを 0 としてしまうが, もしか

したらその点こそ重要な意味を持っているかも知れない。また今の正規誤差の例でも、ピークの高さ h の初期値を半分の 2.5 にとると、ピークのデータの重みが小さくなり、 h がなかなか増大しない。ロバスト推定法は、初期値の偏見に左右されやすい。

しかし既に述べたように、ロバスト推定法はデータ解析の手段であり、データとモデルを診断する道具である。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い。重みが小さくなったら、その点に関する測定過程やモデルを充分吟味すべきである。

飛び離れた測定値が存在していると、単純最小二乗法は、どの測定値にも平等に合わないモデルを与えるが、ロバスト推定法は、とにかくも何割かの測定値を説明するモデルを与えてくれる。この結果を解釈するのは人間である。ロバスト推定法は、診断重視のデータ解析法である。

文 献

1) W. T. Eadie, *et al.*: *Statistical Methods in*

Experimental Physics (North-Holland Pub., 1971). データ解析にたずさわる物理屋にぜひ一読を勧めたい。邦訳がないのが残念である。

- 2) S. Brandt: *Statistical and Computational Methods in Data Analysis*. 邦訳: 吉城 肇, 高橋秀知, 小柳義夫訳: データ解析の方法 (みすず書房, 1976). 上記の本よりやや伝統的スタイルであるが、物理屋向きに書かれている。
- 3) 赤池弘次: 数理科学 No. 153 (1976) 5.
- 4) M. Roos, *et al.*: *Phys. Fenn* **10** (1975) 21. Particle Data Group: *Rev. Mod. Phys.* **48** (1976) S1; *Phys. Lett.* **75B** (1978) i.
- 5) 三浦良造: 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——ロバスト推定法。
- 6) 中川 徹, 小柳義夫, 戸川隼人: 東京大学大型計算機センターニュース **11** (1979) No. 4, p. 70——最小二乗法標準プログラムの開発 (4). ここには詳しい報文目録が含まれている。
- 7) 中川 徹, 小柳義夫: 最小二乗法標準プログラム SALS, 第2版, 利用の手引き, 改訂2版, 第1部基礎篇. 中川 徹, 小柳義夫, 戸川隼人: 同上, 第2部制御・解法篇 (東京大学大型計算機センター, 1979).

BUTSURI

Vol. 34, No. 10

Reviews

Solid Grains in Interstellar Space.....	Syuzo Isobe	830
Long-Time Tails of Spin Correlation Functions and Magnetic Resonance.....	Kazukiyo Nagata	839
Pion Condensation in High Density Nuclear Matter.....	Hiroshi Toki and Yasuhiko Futami	848
Unsteady Stagnation-Point Flow.....	Yaso'o Matunobu	855
A New Type of Nuclear Interactions, Centauro.....	Yoichi Fujimoto and Masanobu Tamada	865
Sound Velocity of Mercury in Metal-Nonmetal Transition Region near Liquid-Gas Critical Point.....	Kenjiro Suzuki and Masaaki Inutake	869
Fine Structure in the Melting of DNA.....	Hideki Tachibana and Sizue Ueno	872
Robust Estimation and Data Analysis.....	Yoshio Oyanagi	885

Instruments and Techniques

Low Light Level Image Sensors and Observing Technique.....	Yuji Kiuchi	878
--	-------------	-----

Published by the Physical Society of Japan

Room No. 211, Kikai-Shinko Building
3-5-8, Shiba-Koen, Minato-ku, Tokyo 105, Japan