

Chebyshev 基準による回帰直線 の幾何学的導出方法

尾 崎 雄一郎

1. はじめに

回帰直線を求める方法として最小自乗法がよく知られているが、これ以外に偏差の絶対値の和を最小にする最小絶対値法や偏差の絶対値の最大値を最小にする Chebyshev 基準 (ミニマックス法) が知られている。Kelley [3] は一変数の多項式で表される Chebyshev 基準による回帰曲線を求める問題を、また Wagner [7] は最小絶対値法と Chebyshev 基準による回帰平面を求める問題をリニアール・プログラミングの問題に変換する方法を示した。Appa and Smith [1] は最小絶対値法と Chebyshev 基準による回帰平面が示す若干の特徴をリニアール・プログラミングを利用して明らかにした。Farebrother [2] はこれら 2 つの方法と関連する過去の研究結果を主としてリニアール・プログラミングの観点から歴史的に考察するとともに、Chebyshev 基準による回帰直線を図を用いて解く 1 つの方法を提案した。しかし、この図による方法は間接的で、計算に時間がかかる。尾崎 [5] は Chebyshev 基準による回帰直線が示す多くの特徴をリニアール・プログラミングを利用して証明し、尾崎 [6] は原点を通る Chebyshev 基準による回帰直線を一般的に求める計算方法を見つけ、これに基づいてこの回帰直線がもつ幾つかの特徴を明らかにした。

Chebyshev 基準による回帰直線をリニアール・プログラミングを用いて求める方法は一般的で、理論的に極めて優れている。しかし、この回帰直線をリニアール・プログラミングを用いて実際に求めようとすると、たとえ観測値の数がそれほど多くなくても、変換によって制約式と変数の数が多くなり、解くのに極めて時間がかかる。この論文において Chebyshev 基準による回帰直線を図を利用して求める効率的で、一般的な方法を提案するとともに、この回帰直線がもつ幾つかの特徴を明らかにする。この論文は尾崎 [5] で示した特徴を利用して尾崎 [4] の方法を発展させたものである。

2. Chebyshev 基準による回帰直線とリニアール・プログラミング

独立変数を x 、従属変数を y とし、 m 個の観測値

$$(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$$

が与えられているものとする。これらの観測値に対して偏差の絶対値の最大値を最小にする直線

$$(1) \quad y = ax + b$$

の傾き a と定数項 b を推定する問題を考える。観測値の数 m は統計的な推定が意味をもつために $m \geq 3$ であるとする。このとき、 i 番目の観測値と直線 (1) との偏差 e_i を

$$(2) \quad e_i = y_i - (ax_i + b) \quad (i = 1, \dots, m)$$

と表す。Chebyshev 基準による回帰直線の係数は偏差の L_∞ ノルムである

$$\max_{i=1,\dots,m} |e_i|$$

を最小にするという問題の最適解としてえられる。いま、

$$(3) \quad z = \max_{i=1,\dots,m} |e_i|$$

とすると、(2)と(3)より

$$(4) \quad z \geq |e_i| = |y_i - ax_i - b| \geq 0 \quad (i=1, \dots, m)$$

という式が成立する。(4)より

$$z \geq y_i - ax_i - b, \quad z \geq -y_i + ax_i + b, \quad z \geq 0 \quad (i=1, \dots, m)$$

という不等式をえるから、Chebyshev 基準による回帰直線の傾き a と定数項 b は、

$$(5) \quad z$$

を次の制約条件の下で最小にする

$$(6) \quad ax_i + b + z \geq y_i, \quad -ax_i - b + z \geq -y_i, \quad z \geq 0 \quad (i=1, \dots, m)$$

というリニアール・プログラミングの問題の最適解としてえられる。ここで、 a , b , z が変数であり、 a と b には符号の制約がない。なお、傾き a が非負であるとか、定数項 b がゼロである、あるいは直線 (1) が特定の点を通るといような条件を (6) に付け加えることもできる。リニアール・プログラミングの問題 (5)–(6) には常に最適解が存在する(尾崎[5], 定理1を参照)から、Chebyshev 基準による回帰直線 (1) を必ず求めることができる。

3. Chebyshev 基準による回帰直線の図による導出方法

本節で Chebyshev 基準による回帰直線を図を用いて求める方法を例を用いて説明する。

例題 1. 第1図の点 A, O, B, C, D で示される 5 個の観測値

$$A(-1, 2), O(0, 0), B(2, 5), C(4, 2), D(5, 6)$$

が与えられているものとする。

最初に、これらの 2 つの点を結んだときに、残りのすべての点がこの直線によってできる同じ半平面 (あるいはこの直線上) にあるような直線だけを考える。第1図の例では点 O と C を結んでできる直線 α に関して残りの点 A, B, D はすべて同じ側にあるので、直線 α はこの条件を満たすものの 1 つである。次に、直線上にない残りの点の中から直線 α との偏差の絶対値が最大となるものを見つけ、その最大値を記録する。いまの例では直線 α に対して偏差の絶対値が最大となるのは点 B であり、このときの最大値を t_1 とすると、 $t_1=4$ である。同様に、点 B と D を結んでできる直線 β を取り上げる。残りの点 A, O, C はすべて直線 β に関して同じ側にあり、直線 β からの偏差の絶対値が最大となるのは点 O で、その最大値を t_2 とすると、 $t_2=13/3$ である。また、点 A と B を結んでできる直線 γ を考える。残りの点 O, C, D はすべて直線 γ の片側の平面にあり、直線 γ から偏差の絶対値が最大となるのは点 C で、その最大値は $t_3=5$ になる。さらに、点 A と O を結んでできる直線からの偏差の絶対値が最大となるのは点 D で、その最大値は $t_4=16$ であり、同様に点 C と D を結んでできる直線からの偏差の絶対値が最大となるのは点 A で、その最大値は $t_5=20$ である (第1図には、点 A と O を結んだ直線と、点 C と D を結んだ直線は繁雑さを避けるために描いてな

である。

4. 計算の簡単化

いま述べた方法の中で $\min_s t_s$ を求めるステップは次の定理を利用することによってもっと簡単にできる。

Chebyshev 基準による回帰直線からの偏差が最大である観測値は 3 個存在し、これらの観測値の 1 つたとえば (x_2, y_2) はこの回帰直線の上方(下方)に、2 つの観測値たとえば (x_1, y_1) と (x_3, y_3) は下方(上方)に位置し、上方(下方)に位置する観測値の x 座標 x_2 は下方(上方)に位置する 2 つの観測値の x 座標 x_1 と x_3 (以下簡単化のために $x_1 < x_3$ と仮定する)による閉区間内にある、すなわち

$$x_1 \leq x_2 \leq x_3$$

が成立する(尾崎[5], 定理 11 を参照)。さらに、 x_2 が x_1 と x_3 による开区間内にあるとき、すなわち

$$x_1 < x_2 < x_3$$

が成立するときには、Chebyshev 基準による回帰直線は一意である(尾崎[5], 定理 5 と定理 7 を参照)。

この定理により、2 つの点を結んでできる直線のうち残りのすべての点が同一の半平面上にあるような直線に対して偏差の絶対値が最大となる点、たとえば (x_2, y_2) の x 座標が、この直線を決定している 2 つの点、たとえば $x_1 < x_3$ である (x_1, y_1) と (x_3, y_3) の x 座標の間にある、すなわち $x_1 < x_2 < x_3$ であるならば、もはや適当な 2 点を結ぶ他の直線とそこからの偏差の絶対値が最大となる点を見つけ、その最大値を調べるという作業を続ける必要はない。これらの 3 点を使って前節で述べた方法にしたがって回帰直線を求めればよい。ただし、直線を決定している 2 点の x 座標のいずれかと偏差の絶対値が最大である点の x 座標が一致する場合には、多重解の可能性があるので、それを知るために偏差の絶対値が最大となっている点を通る他の直線を調べる必要がある。いまの例では、2 点 $O(0, 0)$ と $C(4, 2)$ を結ぶ直線 α から偏差の絶対値が最大であるのは点 $B(2, 5)$ であり、点 B の x 座標の 2 が点 O と C の x 座標による开区間の $0 < x < 4$ 内にあるから、Chebyshev 基準による回帰直線は一意である。したがって、他の 2 点を通る直線を調べることなく、直ちにこれら 3 点を用いて回帰直線を求めればよい。

さらに、 $g \in M_1 = \{i \mid \min y_i\}$, $k \in M_2 = \{i \mid \max y_i\}$ である観測値 (x_g, y_g) と (x_k, y_k) が各々 M_1 と M_2 の唯一つの要素であるとき、求める回帰直線の傾き $\tilde{a}$ は、 $x_k > x_g$ であるならば、 $\tilde{a} > 0$ であり、 $x_k < x_g$ であるならば、 $\tilde{a} < 0$ であり、 $x_k = x_g$ であるならば、 $\tilde{a} \geq 0$ のいずれも成立するという定理(尾崎[5], 定理 12 と定理 15 を参照)を利用すると、Chebyshev 基準による回帰直線を効率的に求めることができる。

5. Chebyshev 基準による回帰直線の特徴

定数項をもつ最小自乗法による回帰直線がデータの平均を通ることはよく知られている。例題 1

の Chebyshev 基準による回帰直線 (7) も 5 個の観測値の平均 $(2, 3)$ を通る. しかし, いまのデータから観測値 $D(5, 6)$ を取り除くと, 新しい平均は $(5/4, 9/4)$ となるが, 回帰直線は (7) のまま不変で, 新しい平均を通らない. したがって, Chebyshev 基準による回帰直線は最小自乗法による回帰直線と異なって一般にデータの平均を通らない.

次に, Chebyshev 基準による回帰直線は最小自乗法による回帰直線と異なって, 必ずしも一意ではないことを示そう.

例題 2. 第 2 図の点 A, B, C で示される 3 個の観測値

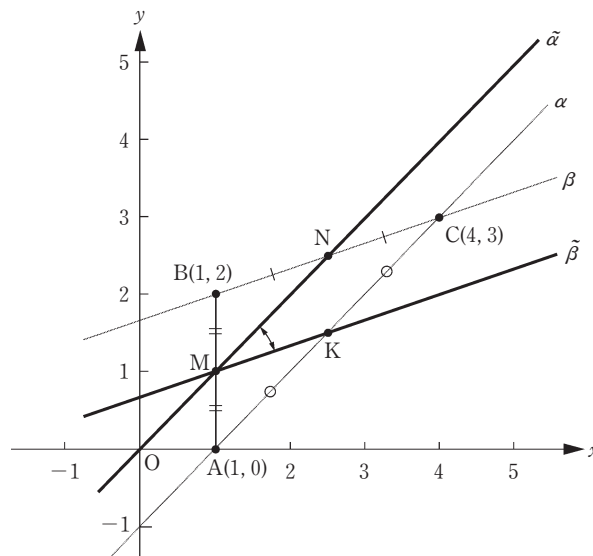
$$A(1, 0), B(1, 2), C(4, 3)$$

が与えられているものとする.

この場合, 点 A と C を結ぶ直線 α から偏差の絶対値が最大になるのは点 B で, このときの偏差の絶対値の最大値は 2 である. また, 点 B と C を結ぶ直線 β から点 A までの偏差の絶対値の最大値も 2 となって, ともに最小である. 直線 α を決定する点 A と C を底辺とし, 点 B を頂点とする $\triangle BAC$ の辺 BA の中点 M と辺 BC の中点 N を結んでできる直線 $\tilde{\alpha}$ と, 直線 β を決定する点 B と C を底辺とし, 点 A を頂点とする $\triangle ABC$ の辺 AB の中点 M と辺 AC の中点 K とを結んでできる直線 $\tilde{\beta}$ はともに求める回帰直線である. さらに, 点 M を通り傾きがこれら 2 つの直線の間にあるもの, すなわち

$$y = \tilde{\alpha}x + 1 - \tilde{\alpha}, \quad 1/3 \leq \tilde{\alpha} \leq 1, \quad \tilde{z} = 1$$

もすべて Chebyshev 基準による回帰直線である. したがって, この場合 Chebyshev 基準による回帰直線は一意ではない.



第 2 図

例題 3. 第 3 図の点 A, B, C, D で示される 4 個の観測値

$$A(1, 1), B(1, 3), C(4, 2), D(4, 4)$$

が与えられているものとする。

先に説明した方法により、この場合 Chebyshev 基準による回帰直線は

$$y = \frac{1}{3}x + \frac{5}{3}, \quad z=1$$

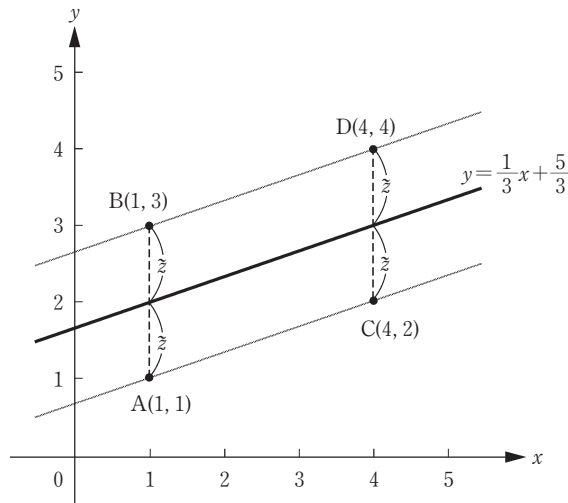
であり、これは一意である。Chebyshev 基準による回帰直線は、例題 1 が示すように

$$(10) \quad x_i \neq x_j \quad (i, j=1, \dots, m; i \neq j)$$

が成立するならば、一意であり、したがって(10)は Chebyshev 基準による回帰直線が一意となるための十分条件である(尾崎[5], 定理 5 を参照)。これと同じことであるが、例題 2 が示すように Chebyshev 基準による回帰直線が多重解であるならば、

$$(11) \quad x_i = x_j \quad (i \neq j)$$

が成立する 2 つの観測値 (x_i, y_i) と (x_j, y_j) が存在する。したがって、(11) は Chebyshev 基準による回帰直線が多重解となるための必要条件である(尾崎[5], 定理 6 を参照)。しかし、この例題 3 が示すように、(10) は Chebyshev 基準による回帰直線が一意となるための必要条件ではなく、また (11) はこの回帰直線が多重解となるための十分条件でもない。



第3図

例題 4. 第 4 図の点 A から E で示されている 5 個の観測値

$$A(1, 1), B(2, 2), C(3, 5), D(4, 4), E(5, 5)$$

が与えられているものとする。

この場合、偏差の L_1 ノルムを最小にする最小絶対値法による回帰直線は

$$y = x \quad (L_1)$$

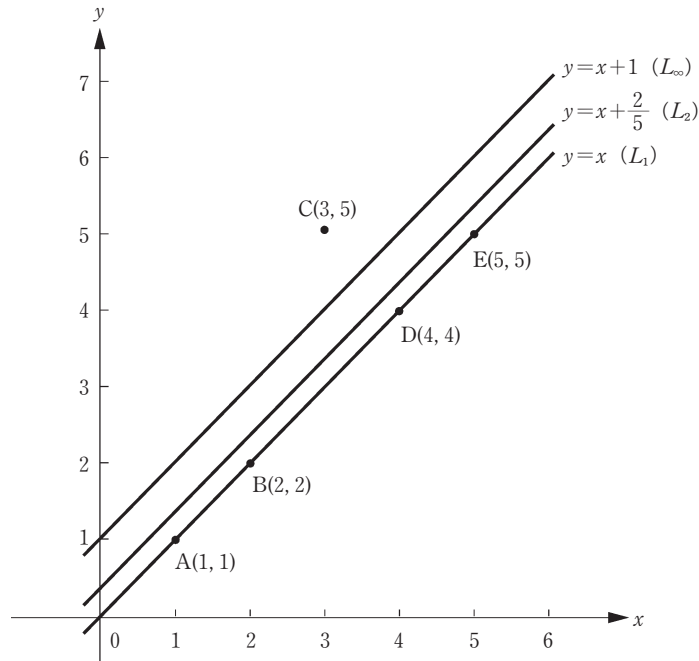
であり、偏差の L_2 ノルムを最小にする最小自乗法による回帰直線は

$$y = x + \frac{2}{5} \quad (L_2)$$

であり、Chebyshev 基準による回帰直線は

$$y = x + 1 \quad (L_{\infty})$$

となる。このように最小絶対値法による回帰直線は例外的な観測値に影響を受けにくい。最小自乗法と Chebyshev 基準による回帰直線は例外的な観測値によって大きく影響を受けることがある。



第 4 図

6. おわりに

Chebyshev 基準による回帰直線を求めるためにリニア・プログラミングの問題(5)–(6),あるいはこの双対問題をシンプレックス法か双対シンプレックス法を用いて解こうとすると,たとえ観測値の数がそれほど多くなくても非常に時間がかかる。しかし,この論文で説明した図を利用する方法を用いれば,観測値の数が多くても非常に短い時間で解くことができる。

Chebyshev 基準による回帰直線はデータの中の 3 個の観測値だけに依存し,他の多数の観測値にある範囲内で全く依存しないし,必ずしも一意ではなく,そのうえ例外的な観測値に影響されることがある。しかし,正常なデータから求めた Chebyshev 基準による回帰直線から偏差の絶対値の最大値の最小値に等しい幅で上下 2 本の平行な 2 直線を引き,新しい観測値が追加されたとき,この観測値がこれらの平行な 2 直線によってできる帯の中に入れば,この観測値を正常な値と判定し,この帯の中に入らなければ,異常な値と判定するという一種の品質管理にこの回帰直線は利用できる。

参考文献

- [1] Appa, G., and Smith, C., “On L_1 and Chebyshev Estimation,” *Mathematical Programming*, Vol. 5 (1973), pp. 73-87.
- [2] Farebrother, R. W., “The Historical Development of the L_1 and L_∞ Estimation Procedures,” in Y. Dodge (Ed.), *Statistical Data Analysis Based on the L_1 -Norm and Related Methods*. Amsterdam : North-Holland, 1987, pp. 37-63.
- [3] Kelley, J. E., Jr., “An Application of Linear Programming to Curve Fitting,” *Journal of the Society for Industrial and Applied Mathematics*, Vol. 6 (1958), pp. 15-22.
- [4] 尾崎雄一郎, 「Chebyshev 基準による回帰直線の幾何学的導出方法」, 『第 11 回中部支部研究発表会アブストラクト集』, 日本オペレーションズ・リサーチ学会, 1984 年, pp. 22-27.
- [5] 尾崎雄一郎, 「Chebyshev 基準による回帰直線の特徴づけ定理」, 『名城論叢』, 第 2 巻, 第 4 号 (2002 年), pp. 21-33.
- [6] 尾崎雄一郎, 「Chebyshev 基準による原点を通る回帰直線とリニアール・プログラミング」, 『名城論叢』, 第 3 巻, 第 1 号 (2002 年), pp. 1-10.
- [7] Wagner, H. M., “Linear Programming Techniques for Regression Analysis,” *Journal of the American Statistical Association*, Vol. 54 (1959), pp. 206-212.