Ⅱ−20 環境情報の計測誤差に伴う近似式パラメータの安定性について

山下清吾[1]　　草間晴幸[2]
Seigo Yamashita　　Haruyuki Kusama

【抄録】計測データをもとに近似曲線で推定式を導く手法は今なお幅広く用いられている。様々な数学処理ソフトで近似曲線の係数は容易に求まる。しかし、求めた係数の信頼性は、計測誤差に依存し、数%のデータの違いで変動する場合もある。殊に、自然環境データは、サンプリング誤差がつきもので、BOD等の水質指標については２０%以上のケースも少なくない。 通常の曲線近似計算過程での行列演算では、Condition Numberが大きくなり、求解に不安定さを伴う。本研究では、3次放物線での近似過程で通常の解放によるものと、独立変数にチェビシェフ関数を組み込んだものとを比較して解の安定性を検証した。データのばらつきが１５%を越えた場合に後者の優位性を認めた。

1. はじめに

土木工学の諸分野において、実験、観測、計測データをもとに経験式を導き、ある種の物理量やパラメータ等を推定することは、幅広く用いられており、重要な工学的手法として今後も利用されていくことには疑問の余地はない。 最新の研究論文を見ても、森田ら[1)]の海水ＣＯＤと透明度関係式や、海野ら[2)]が示したリン平衡濃度と飽和吸着量の関係を表す推定式など、使用例は少なくない。さらに、水質や大気中の成分測定など、自然環境データ計測が関係することから、土木工学分野のなかでも、殊に、多くの環境情報を扱う研究報告の中に回帰的手法を用いる事例を挙げることができる。

これらの経験式導入では、最少自乗法に代表される手法を用いて、線形、指数、累乗関数が直接あるいは、対数を介在させることによる間接手法などで近似させることになる。 しかし、自然環境データの場合、鋼材やコンクリートのような人工的な管理下で製造される材料とは異なり、測定機会が異なれば、数パーセントから数十パーセントの計測値のばらつきは、ある程度避けられないのが現実である。このことを考慮に入れて脇屋[3)]は溶存酸素量や電気伝導度などの水質指標測定に際して、河川水理条件の違いにはよるが、信頼できるデータ獲得のためには、１０個から２０個に及ぶサンプル数が必要とも報告している。また、澤里[4)]の土壌水分測定器に関する報告でも、飽和土の比較的高い土壌においては、乾燥した土壌よりも測定精度が低くなる傾向を示すなど、測定機器を因とするデータ解析のための入力データ無視できないばらつきが存在する。

以上のことから、環境情報に関わるデータをもとに解析する場合には、被測定領域の条件に起因するものや、測定機器精度に由来するものなどの相当量のばらつきが常に存在することを念頭に入れねばならない。

2. 目的

近年では、表計算ソフトはもとより様々な数学ソフトの普及により、近似直線・曲線は、比較的容易に求めることができる。 しかし、筆者の調べた範囲では、世界的に多くのユーザーを持つ有名ソフトでも、入力データを直接、最少自乗法で回帰計算し、係数をそのまま算出している。 殊に、3 次放物線を含む多項式関数（Polynomial

1： 山下清吾　豊田工業高等専門学校環境都市工学科　助教授
2： 草間晴幸　名古屋市立大学大学院芸術工学研究科　教授

Function) では、独立変数の2乗項、3乗項など、比較的「鈍い」項の係数の大幅な変化で入力データの変化に対応し、ベストフィットを求める過程で、不安定なパラメータを計上することになる。ここで言う「不安定」とは、入力データの僅かな違い、ばらつきで、経験式で最も重要な、曲線の曲率を表す係数や切片が大きく変化することである。

本研究では、3 次放物線での近似過程で通常の解放によるものと、独立変数にチェビシェフ関数を組み込んだものとを比較して解の安定性の差異を検証し、所謂、生のデータを扱う場合の経験式導入での危険性と、その対策のひとつを述べることを目的とする。

3. 多項関数の最小自乗法による近似

多項関数の最小自乗近似 (Polynomial Least Squares Approximation) は以下の式で表される。

$$f(x)=a_1\varphi_1(x)+a_2\varphi_2(x)+....+a_m\varphi_m(x) \quad (1)$$

ここに、$a_1, a_2, ..., a_m$ ：係数

$\varphi_1, \varphi_2, ..., \varphi_m$ ：任意の関数

上記の任意関数に、各々１， X, X^2, X^3 が適用されて、2次、3次の放物線関数となる。今、j 個の2変数、例えばX(j)とY(j)、があり、近似式による誤差の2乗の合計を目的関数 G とすると、

$$G(a_1,a_2...a_m)=\sum_{j=1}^{n}[a_1\varphi_1(x_j)+a_2\varphi_2(x_j)+...+a_m\varphi_m(x_j)-y_j]^2 \quad (2)$$

となる。上式を係数 $a_1, a_2, ..., a_m$ について、目的関数 G をm個偏微分し、ゼロと置けば、左辺がm x mの行列となり、(右辺は 1 x m のベクトル) 一般式が式（3）で表される連立方程式を解くだけとなる。

$$\sum_{k=1}^{m}a_k\left[\sum_{j=1}^{n}\varphi_k(x_j)\varphi_i(x_j)\right]=\sum_{j=1}^{n}y_i\varphi_i(x_j) \quad (3)$$

ただし、 $i = 1, 2, ... m$

上式（3）のφ_k項に以下に表されるチェビシェフ関数を挿入する。

$$T_n(x)=\cos[n\cos^{-1}x], \quad -1\le x\le 1 \quad (4)$$

また、$\theta=\cos^{-1}(x)$, $(0\le\theta\le\pi)$ とおけば、

$$T_n(x)=\cos(n\theta) \quad (5)$$

となり、簡便なかたちとなる。 例えば、n=2であれば、

$$T_2(x)=\cos(2\theta)=2\cos^2(\theta)-1=2x^2-1$$

となる。 3次放物線を近似式にする場合は、式(1)中の $\varphi_1, \varphi_2, \varphi_3, \varphi_4$ の4項が対象となり、直接的には、各々に以下の（6）に示す関数を挿入して鋭敏化させることで、入力値のばらつきによる係数の変化を最小限に抑えることができる。

$$\begin{aligned}\varphi_1(x)&=1\\ \varphi_2(x)&=2x-1\\ \varphi_3(x)&=8x^2-8x+1\\ \varphi_4(x)&=32x^3-48x^2+1\end{aligned} \quad (6)$$

線形システム Ax = B において、以下に示す Condition Number が Bベクトルの比較的小さな変化に対する鋭敏さの程度をはかる尺度となることは、Atkinson[4] による著書をはじめとする幾多の数値解析テキストに示されている。

$$cond(A)=\|A\|\cdot\|A^{-1}\| \quad (7)$$

ここに、$\|A\|$ はNorm

Condition Number が小さな状態を 良好な状態 (Well-Conditioned) と呼ぶが, 鋭敏な関数の挿入により、良好な状態で近似式を導くのである。

4. 近似検定

表 1 に示す 21 組の 2 変数の値を近似検定用入力データとして Atkinson[4] が近似例として用いたものを使った。これは、図 1.でもわかるように典型的な 3 次放物線で近似できるよう用意されたものである。

表1. 近似検定用データ

#	Yi	Vi	#	Yi	Vi
1	0.000	0.486	12	0.550	1.102
2	0.050	0.866	13	0.600	1.099
3	0.100	0.944	14	0.650	1.017
4	0.150	1.144	15	0.700	1.111
5	0.200	1.103	16	0.750	1.117
6	0.250	1.202	17	0.800	1.152
7	0.300	1.166	18	0.850	1.265
8	0.350	1.191	19	0.900	1.380
9	0.400	1.124	20	0.950	1.575
10	0.450	1.095	21	1.000	1.857
11	0.500	1.122	–	–	–

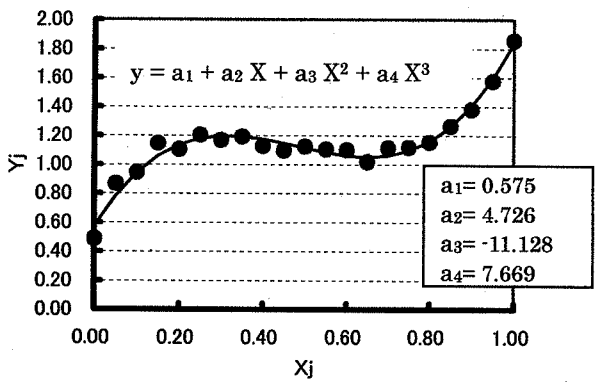

図 1. 近似曲線（ﾁｪﾋﾞｼｪﾌ関数挿入なし）

表 1. のオリジナルデータの各々の数値が最大で n%のばらつきを持つデータを 5 %きざみで 20%までの 4 ケースについて乱数発生させて後、3 次放物線による近似をチェビシェフ関数挿入無しと、挿入有りとのケースで比較した。オリジナルと併せて 5 通りの入力データがあり、近似法に挿入無しと、挿入有りがあるため、係数 a の解が 10 通りある。ここに図で示すのは紙面の都合もあり、最大 10%ばらつきケースのみを図 2. に示す。

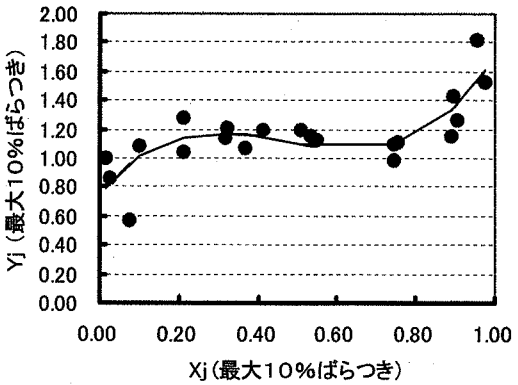

図 2. 最大 10%ばらつき近似

10 通りの異なるケースについて Fortran プログラムを作成し、シミュレーションを行った。計算結果である係数 a の変化の様子を以下の表 2. に、係数 a の変動率（挿入無しをベースとする）を表 3. に示す。

表 2. 入力値誤差に対応する放物線係数 a の値

最大誤差	チェビシェフ関数挿入	a1	a2	a3	a4
0%	無	0.575	4.726	−11.128	7.669
	有	0.575	5.054	−11.128	7.669
5%	無	0.571	4.311	−9.759	6.668
	有	0.571	4.523	−9.759	6.668
10%	無	0.743	3.369	−8.191	5.789
	有	0.743	3.801	−8.191	5.789
15%	無	0.677	3.439	−7.642	5.363
	有	0.677	3.792	−7.642	5.363
20%	無	0.687	3.577	−7.550	4.826
	有	0.507	3.831	−7.898	5.149

表 3. 各最大誤差に対応する係数 a の変動率

最大誤差	a1	a2	a3	a4
0%	0.00	6.94	0.00	0.00
5%	0.00	4.93	0.00	0.00
10%	0.00	12.80	0.00	0.00
15%	0.00	10.24	0.00	0.00
20%	−26.17	7.10	4.61	6.69

式（7）で紹介したCondition Numberを5つの入力用データについて求めた結果を次の表4．に示す。いずれも5桁を超える数値であり、挿入なしでは、求解が不安定である。

表4．Condition Number

最大誤差	0%	5%	10%	15%	20%
C.N.	22000	22548	25874	50170	16035

5．係数の信頼度について

前頁の表3．で示した変動率は、係数aの信頼度を表すものと考えることができる。小さいほど安定した係数と言える。最大ばらつき(誤差)が15%までは、a2以外の係数に変化はない。(少なくとも、表で用いた精度の範囲内では計上されない。)しかし、係数a2はXの1次項にかかるものであり、オリジナルデータ（0%）や5%程度の入力値のばらつきに、5%もの変動をみせている。これは、導かれる経験式に要求される精度により、無視できたり、重要な変化とみなされたりするであろうが、チェビシェフ関数挿入による近似式中の係数が安定しているのであるから、いわゆる「生」のままの解析結果を用いる理由はない。

表3でもわかるとおり、最大誤差が大きいケースで、係数の変動率が小さいケースがある。これは、シミュレーションに用いた、ばらつきを持つデータが、最大変動をコントロールした乱数発生数値によるものであり、分散の大小とは直接関係がないことによる。用いたデータはあくまで、各々の最大誤差での1つの現出である。したがって、わずかな変動は、別の発生させたデータで行えば、別の変動率を呈するである訳だが、3次放物曲線による近似においては、最大誤差15%までは係数a2によって、挿入無しの関数のもつ非鋭敏性を補完している結果となった。

さらに、最大ばらつきが20%のケースでは、すべての係数に変動がみられた。　係数a2だけに関しては、その結果、変動率が減少することとなったが、これは全体とのバランスによるものであり、すべての係数への信頼度は低くなる。この結果も、データ自体が1つの現出したものなので、異なったデータセットでは、係数相互の大小は、ばらつくものの、複数回のシミュレーションの結果、15%を越えて20%程度になると、すべての係数が変動していく傾向を認めた。

6．まとめ

3次放物線での近似過程で通常の解放によるものと、独立変数にチェビシェフ関数を組み込んだものとを比較して解の安定性の差異を検証し、以下の結果を得た。

（1）　入力値誤差に対応する3次放物線係数aの値は、1次項の係数が、真の値とみなしたデータから少ないばらつきを持つデータでも5%程度の変動をみせた。

（2）　真の値とみなしたデータから最大で各数値が10%の計測誤差を超える場合、1次項の係数のみではあるが10%以上の変動をみせた。

（3）　真の値とみなしたデータから最大で各数値が15%の計測誤差を超え、20%程度になる場合、4つの係数すべてが変動した。

（4）　水質データや大気データには、20%内外のサンプリング誤差がある場合が多いので、3次放物曲線による近似はチェビシェフ関数挿入近似などを用いて解の安定した方法が望ましい。

参照文献

1) 森田健二，竹下彰：アマモ場分布限界水深の予測評価法、土木学会論文集，No.741，pp.49-56, 2003.

2) 海野修司，岡本正美，永渕正夫：浄化汚泥をもちいたリン除去技術，土木学会論文集，No.741, pp.111-121, 2003.

3) 脇屋佳代：水質調査におけるサンプリング誤差に関する研究，豊田工業高等専門学校卒業研究論文，2002.

4) 澤里恵美：誘電式土壌水分計の測定精度に関する研究，豊田工業高等専門学校卒業研究論文，2002.

5) Atkinson, K. :Elementary Numerical Analysis, John Wiley & Sons, Inc., 1985.