

〔論 説〕

## 順序制約情報に基づくミカエリス・メンテン式の パラメータの推定、パートⅡ

孫 宏 傑

### 〔要 旨〕

本稿は前稿<sup>6</sup>のミカエリス・メンテン式のパラメータの推定アルゴリズムを改良し、より収束のはやい推定が実現できた。また、最小二乗誤差が最小という基準で推定結果が得られた。これらの結果から順序制約情報に基づくアプローチが有意義だと結論できる。

### 1. はじめに

ミカエリス・メンテン式<sup>2</sup> (MM 式) は酵素反応速度理論の基本的な公式である。酵素反応速度  $v$  は式

$$v = \frac{V_{max}s}{K_m + s} \quad (1)$$

で表れる。ここで、 $s$  は基質濃度、 $V_{max}$  は基質濃度が無限大のときの反応速度である。また、 $K_m$  は  $v = V_{max}/2$  を与える基質濃度である。MM 式は  $s \ll K_m$  のとき、 $v \approx (V_{max}/K_m)s$ 、すなわち、 $v$  が  $s$  と比例して変化する。また、 $s \gg K_m$  のとき、 $v \approx V_{max}$ 、すなわち、 $v$  が飽和状態になり、 $s$  の効用がなくなる。MM 式のパラメータの推定問題とは実験データ  $(s_i, v_i)$ 、 $v_i = E(v_i) + e_i$ 、 $Var(v_i) = \sigma^2$ 、 $i = 1, \dots, n$  から  $V_{max}$  と  $K_m$  を推定することである。現在、 $V_{max}$  と  $K_m$  の推定がおもに非線形回帰で行うが、それ以外の方法がいくつか提案された。そのうちの一つ、Direct Linear Plot 法<sup>3,4</sup> (DLP 法) がある。DLP 法は実験データを  $(s, v)$  空間の代わりに  $(K_m, V_{max})$  空間に変換し、変換後の空間でパラメータの推定を行う。変換は式

$$V_{max} = \frac{K_m + s}{s} \times v = \frac{v}{s} \times K_m + v \quad (2)$$

で行い、 $n$  組のデータを  $n$  本の直線に変換し、最大  $n \times (n-1)/2$  個の交点を生成する。理論上、もしデータに誤差がない場合、 $n$  本の直線が一点に収束し (図 1)、この収束点が真の  $V_{max}$  と  $K_m$  となる。しかし、実際のデータに誤差があるため、一点に収束することはまれである (図 2)。

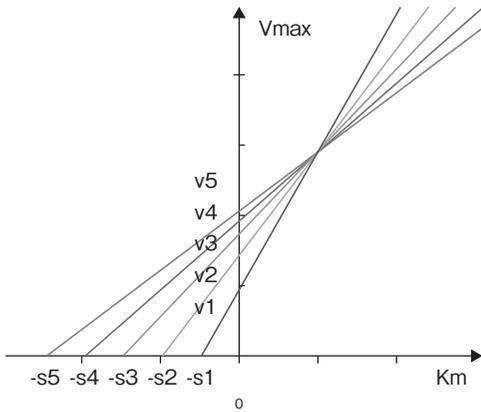


図 1. 理想状況

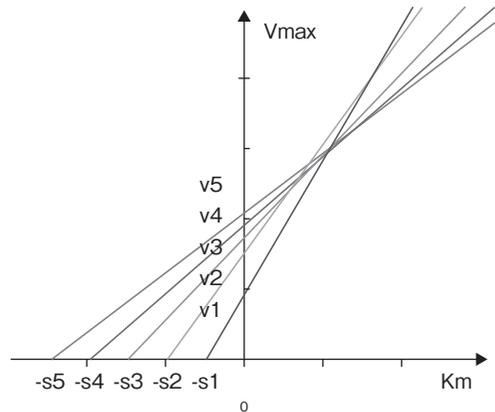


図 2. 現実状況

DLP 法は  $V_{max}$  と  $K_m$  の両軸上の最大  $n \times (n-1)/2$  個の交点のメディアンを推定値とし、そのバラツキを分散とする。直観的に DLP 法がシンプルで分かりやすい。しかし、 $(K_m, V_{max})$  空間上、時々、マイナス  $V_{max}$  とマイナス  $K_m$  が発生し、物理現象として解釈できない問題がある。また、データ数  $n$  が大きくなると  $n \times (n-1)/2$  個の交点の範囲も大きくなり、すなわち、分散が大きくなるという致命的欠陥がある。

DLP 法は  $V_{max}$  と  $K_m$  の解を  $(K_m, V_{max})$  空間で求める発想がよい。変換前と変換後に確率構造が同じというよい性質がある。しかし、解の結果がよくない。実際、 $(K_m, V_{max})$  空間での各直線に順序関係があり、この順序情報を利用できれば、よりよい  $V_{max}$  と  $K_m$  の推定が考えられる。この観点に基づき、前稿では、直線の順序関係に基づく  $(K_m, V_{max})$  の推定アルゴリズムを提案した。提案したアルゴリズムは DLP 法より真の  $V_{max}$  と  $K_m$  に近いが、しかし、 $V_{max}$  と  $K_m$  の特性がよくわからない。また、 $V_{max}$  と  $K_m$  への収束速度が遅い。本稿は前稿のアルゴリズムの問題点を解決するため、アルゴリズムに修正を加えた。ここで、その結果を報告する。

本稿の第 2 節で順序制約情報に基づく DLP 法の修正を理論面から説明し、修正アルゴリズムを提案する。第 3 節で、基質濃度のデータに重複がない場合に対し、修正アルゴリズムの振る舞いを説明する。第 4 節で、基質濃度のデータに重複がある場合を扱う。理論上、どちらの場合も非線形回帰と同じ最適解が得られるが、実際非常に近い数値解となった。計算精度が原因ではないかと考えられる。おわりに、順序制約情報に基づくミカエリス・メンテン式のパラメータの推定の意義と計算上の注意点についてまとめる。

## 2. 順序制約情報の利用

順序制約情報の利用は基本的に、図2の現実状況を図1の理想状況に修正し、その過程で順序情報を利用することである。実際、その修正は、 $v_1, v_2, \dots, v_n$  に対して行い、修正の結果、全ラインが一つの点で交差する。

まず、表示法を明記する。 $K_m$  軸上の最大  $n \times (n-1)/2$  個の交点を  $k_1, k_2, \dots, k_{n \times (n-1)/2}$  と表し、その大小関係を整理し、結果を  $k_{[1]} \leq k_{[2]} \leq \dots \leq k_{[n \times (n-1)/2]}$  とする。また、式(2)の直線  $V_{max}$  が  $K_m$  軸上の  $K_m$  点において、平均と分散は以下のとおりである。

$$E(V_{max}) = \frac{K_m + s}{s} \times E(v) \quad (3)$$

$$Var(V_{max}) = \left(\frac{K_m + s}{s}\right)^2 Var(v) = \left(\frac{K_m + s}{s}\right)^2 \sigma^2 \quad (4)$$

次に、図3の交点  $k_i$  と  $k_j$  について考える。真の  $k_m$  が  $k_j$  より右のほうにあると仮定する。 $k_i$  は2つの直線  $l_q$  と  $l_t$  の交点で、 $k_i$  の左側においては  $l_q$  が  $l_t$  の上方に、右側においては  $l_q$  が  $l_t$  の下方にある。一方、 $k_j$  においては、 $l_q$  が  $k_j$  までに  $l_t$  の上方にあるべきで、現在の下方にあるのは順序の違反であり、修正を行う必要がある。

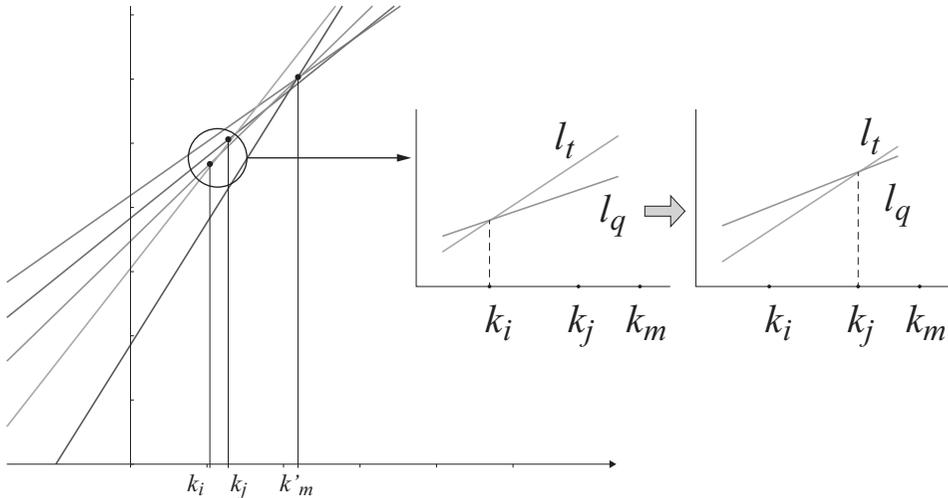


図3. 順序違反の修正

修正は順序制約情報に基づいて行う。 $k_j$  において、 $l_q$  と  $l_t$  の  $V_{max}$  値は記号  $V_{max,q}$  と  $V_{max,t}$  で表すと、

$$\begin{aligned} V_{max,q} &= \frac{k_j + S_q}{S_q} v_q = \frac{v_q}{S_q} \times k_j + v_q \\ V_{max,t} &= \frac{k_j + S_t}{S_t} v_t = \frac{v_t}{S_t} \times k_j + v_t \end{aligned} \quad (5)$$

である。また、分散は

$$\begin{aligned} Var(V_{max,q}) &= \left( \frac{k_j + S_q}{S_q} \right)^2 \sigma^2 \\ Var(V_{max,t}) &= \left( \frac{k_j + S_t}{S_t} \right)^2 \sigma^2 \end{aligned} \quad (6)$$

である。 $k_j$ における修正後の  $V_{max}$  値は記号  $V'_{max}$  で表し、 $V_{max,q}$  と  $V_{max,t}$  の重み付け平均で与える。

$$\begin{aligned} V'_{max} &= w_q \times V_{max,q} + w_t \times V_{max,t} \\ w_q &= \frac{Var(V_{max,t})^{-1}}{Var(V_{max,q})^{-1} + Var(V_{max,t})^{-1}} \\ w_t &= \frac{Var(V_{max,q})^{-1}}{Var(V_{max,q})^{-1} + Var(V_{max,t})^{-1}} \end{aligned} \quad (7)$$

$V'_{max}$ を求めた後、次に  $v_q$  と  $v_t$  を修正する。

$$\begin{aligned} v'_q &= \frac{S_q}{k_j + S_q} V'_{max} \\ v'_t &= \frac{S_t}{k_j + S_t} V'_{max} \end{aligned} \quad (8)$$

これで、一つの修正がおわる。修正の結果、 $v'_q \geq v_q$ 、 $v'_t \leq v_t$  となる。

同じように、右側に違反があれば、同様の修正を行う。

前稿で提案したアルゴリズムが上記の考えに基づき、左右から交互的に順序の違反を隣の交点までに修正しながら、 $V_{max}$  と  $K_m$  の解を求めた。このアルゴリズムが収束するが、収束したときの  $V_{max}$  と  $K_m$  の特性がよくわからない。原因は当然、左右から交互的に順序の違反を修正することに起因する。

## 定義

一組の  $(K'_m, V'_{max})$  に対し、平均二乗誤差 (mean squared error, MSE) が次式で定義する。

$$MSE(K'_m, V'_{max}) = \frac{1}{n-2} \sum_{i=1}^n \left( v_i - \frac{V'_{max} \times S_i}{K'_m + S_i} \right)^2 \quad (9)$$

交点の数が最大で  $n \times (n-1)/2$  個があるため、MSE の個数も最大で  $n \times (n-1)/2$  個ある。

この中から最小の MSE に対応する  $K'_m$  を記憶しておく。 $|k_{[1]} - K'_m|$  と  $|k_{[n \times (n-1)/2]} - K'_m|$  を比較し、最も遠いほうから  $K'_m$  へと修正していく。

修正アルゴリズム

1. 実験データから  $(K_m, V_{max})$  空間上の  $n \times (n-1)/2$  個の交点を求める。計算式は下記のとおりである。

$$V_{ij} = \frac{s_i - s_j}{\frac{s_i}{v_i} - \frac{s_j}{v_j}}, \quad K_{ij} = \frac{v_i - v_j}{\frac{v_i}{s_i} - \frac{v_j}{s_j}}, \quad i, j = 1, 2, \dots, n, i \neq j$$

2.  $n \times (n-1)/2$  組の  $(K_m, V_{max})$  のうち、マイナス  $K_m$  の組を除いて、個々の  $(K_m, V_{max})$  に対し、MSE を求める。その中から、最小の MSE に対応する  $k'_m$  を記憶しておく。
3.  $K_m$  軸上の  $n \times (n-1)/2$  個の交点  $k_1, k_2, \dots, k_{n \times (n-1)/2}$  から、 $k_{[1]} \leq k_{[2]} \leq \dots \leq k_{[n \times (n-1)/2]}$  を求める。 $|k_{[1]} - k_{[n \times (n-1)/2]}| \leq e$  の場合、アルゴリズムを終了する。 $e$  は指定精度である。
4.  $|k_{[1]} - k'_m|$  と  $|k_{[n \times (n-1)/2]} - k'_m|$  を比較し、大きい方から  $k'_m$  へ修正する。修正は式(5), (6), (7), (8)に基づいて行う。
5. 1. に戻り、再度計算を行う。

この修正アルゴリズムは以下の特徴がある。

- I. マイナス  $V_{max}$  とマイナス  $K_m$  の解釈ができる。マイナスになったのは単にデータに誤差があるため、修正で正しい値に戻る。
- II. 修正を行うときの進行方向が交互ではなく常に最も離れたところから  $k'_m$  へと変更したため、収束ははやくなった。
- III.  $k'_m$  は修正のたびに再度計算されるため、常に MSE の最も小さいところをキープしている。このため、修正アルゴリズムが終わったところ、平均二乗誤差最小の意味で  $V_{max}$  と  $K_m$  が求められた。

### 3. 基質濃度のデータに重複がない場合

基質濃度のデータに重複がある場合、修正アルゴリズムの 1. の計算ができなくなるため、特別な対処が必要。ここで、まず重複データのないケースを考える。表 1 は R.J. Ritchie and T. Prvan<sup>5</sup> の論文にあるデータで、修正アルゴリズムの収束特徴とパラメータの推測値を説明するために引用した。上記の論文に、真の  $K_m$  値が 0.560~0.766 の間にあると記載された。

基質濃度 $s$	反応速度 $v$
0.138	0.148
0.220	0.171
0.291	0.234
0.560	0.324
0.766	0.390
1.460	0.493

表 1. 反応速度と基質濃度の関係, R.J. Ritchie and T. Prvan

図 4 から図 9 までは、修正アルゴリズムで表 1 のデータを処理したときの状況を表している。図の横軸は  $K_m$  で、縦軸は  $V_{max}$  である。図 4 にマイナス  $K_m$  とマイナス  $V_{max}$  の交点があり（表 1 のデータ (0.220, 0.171), (0.291, 0.234) に対応）、順序違反のため、 $k_m=0.6007$  に交点するように修正を加えた。修正の結果は図 5 であった。図 6 は 2 回目の修正後の状況であった。図 7, 8, 9 は修正アルゴリズムの収束状況を表している。最終的に解は  $K_m=0.5939$ ,  $V_{max}=0.6890$  であった。

R.J. Ritchie and T. Prvan<sup>5</sup>が非線形回帰で解を求めた結果、 $K_m=0.597$ ,  $V_{max}=0.690$  であった。

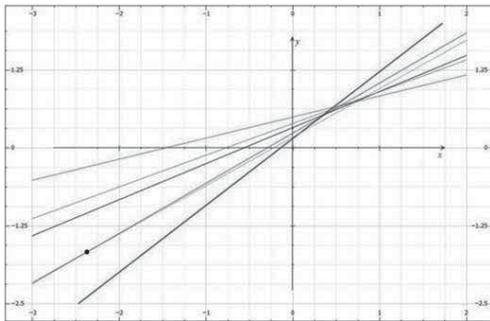


図 4 初期状況, マイナス交点がある

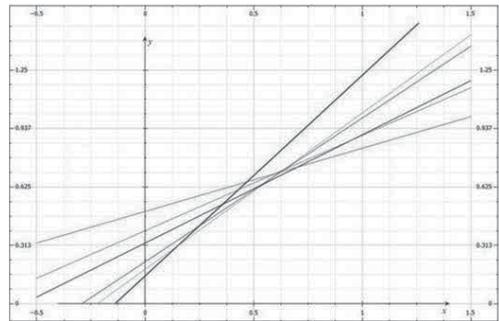


図 5 1 回目の修正後の状況

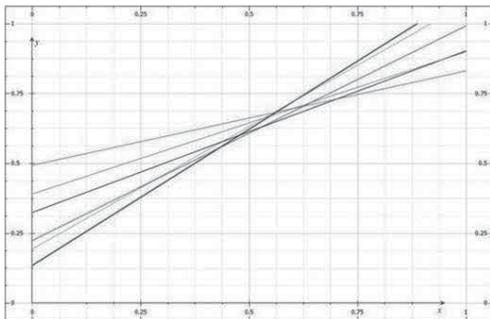


図 6 2 回目の修正後の状況

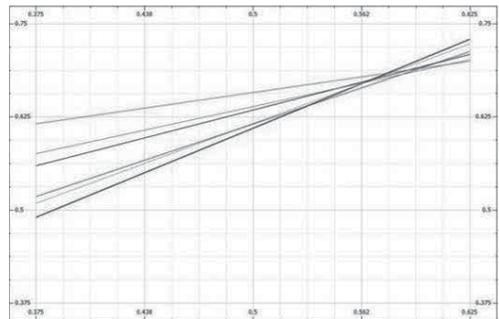


図 7 10 回目の修正後の状況

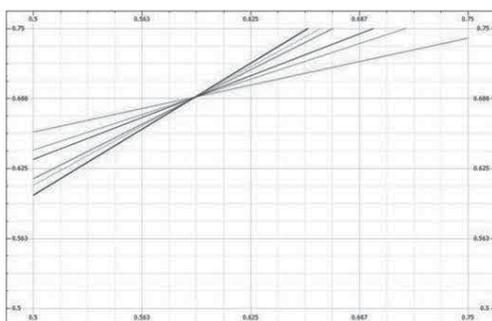


図 8 30回目の修正後の状況

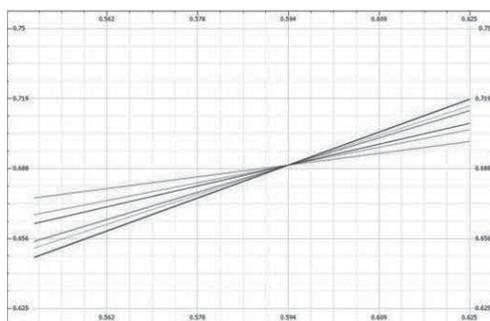


図 9 50回目の修正後の状況

修正アルゴリズムの解と少し乖離があった。この乖離の意味を検討する必要があるが、平均二乗誤差最小という意味で、修正アルゴリズムの解が意義あるといえる。

#### 4. 基質濃度のデータに重複がある場合

基質濃度のデータに重複がある場合、修正アルゴリズムの1. の  $V_{ij}$  と  $K_{ij}$  の計算式の分母が計算の途中にゼロになるため、計算ができなくなる。一方、修正アルゴリズムが線形理論に基づいているため、同じ基質濃度に対し反応速度の平均で代替して推定結果に影響がない。もちろん、重複データの影響が反応速度の分散に算入すればよい。重複した分、分散が小さくなり、計算式(6), (7)の中に重みが大きくなる。

表 2 は D.M. Bates (1988) の著書 (pp. 269) の中、Puromycin 実験の基質濃度と反応速度

基質濃度 $s$	反応速度 $v$
0.02	76
0.02	47
0.06	97
0.06	107
0.11	123
0.11	139
0.22	159
0.22	152
0.56	191
0.56	201
1.10	207
1.10	200

表 2. Puromycin 実験データ, D.M. Bates (1988), pp. 269

に関するデータである。基質濃度のデータに重複がある。このデータセットの  $K_m$  と  $V_{max}$  を求める。

R 言語の非線形回帰関数 `nls()` で表 2 のデータを処理すると、 $K_m = 6.412e-02$ 、 $V_{max} = 2.127e+02$  が得られた。一方、修正アルゴリズムの方は  $K_m = 0.06410$ 、 $V_{max} = 212.6868$  という結果であった。やはり、少し乖離があった。

## 5. おわりに

本稿は順序制約情報の利用という観点からミカエリス・メンテン式のパラメータの推定問題を研究してきた。提案した修正アルゴリズムは最初二乗誤差という基準で理論上最適な解が得られる。しかし、2つの例で示したように、非線形回帰の結果と若干の乖離がある。原因が不明だが、計算精度が原因ではないかと考えられる。いちおう、修正アルゴリズムが倍精度でコーディングしたが、原因を究明するには更なる調べが必要である。

ミカエリス・メンテン式のパラメータの意味がよく図形を用いて説明される。Lineweaver-Burk plot, Hanes-Woolf plot, Direct Linear Plot がある。図形解が解析学上どんな特性があるかよくわからない。これらに対し、本稿のアプローチは新たな解釈となり、非線形回帰の解の解釈にも繋がる。

本稿のアプローチが大量の計算が必要で、非線形回帰より手間がかかる。しかし、順序制約情報の有用性が確認された例としては有意義だと結論できる。

## 参 考 文 献

1. R.E. Barlow, D.J. Bartholomew, J.M. Bremner and H.D. Brunk (1972), *Statistical Inference under Order Restrictions*, John Wiley & Sons, New York
2. D.M. Bates (1988), *Nonlinear Regression Analysis and Its Applications*, 33-36, 269. John Wiley & Sons, New York
3. A. Cornish-Bowden (1981), *Robust Estimation in Enzyme Kinetics*, in *Kinetic Data Analysis: Design and Analysis of Enzyme and Pharmacokinetic Experiments*, edited by L. Endrenyi, Plenum Press, New York and London, 105-119.
4. R. Eisenthal and A. Cornish-Bowden (1974), *The Direct Linear Plot*, *Biochem. J.* 139, 715-720
5. R.J. Ritchie and T. Prvan (1996), Current Statistical Methods for Estimating the  $K_m$  and  $V_{max}$  of Michaelis-Menten Kinetics, *Biochemical Education*, 24(4), 196-206
6. 孫 宏傑「順序制約情報に基づくミカエリス・メンテン式のパラメータの推定について」, 九州産業大学「経営学論集」Vol. 27, No. 4, pp. 67-72, 2017年3月