

生化学反応系のためのベイズ的システム同定法

吉本潤一郎^{†,††} 銅谷 賢治^{†,††}

† 沖縄科学技術大学院大学先行研究事業 〒904-2234 沖縄県うるま市州崎 12-22

†† 奈良先端科学技術大学院大学 〒630-0192 奈良県生駒市高山町 8916-5

E-mail: †{jun-y.doya}@oist.jp

あらまし 未知の細胞内分子機構に関する仮説モデルを検証する上で、計算機シミュレーションは不可欠なツールである。しかしながら、シミュレーションの挙動を規定する力学パラメータには未知のものが多いため、利用可能な実験データからこれらのパラメータを高い信頼性で自動的に同定できる手法の開発が期待されている。本報告では、このパラメータ同定問題を取り上げ、ベイズ推定法に基づく解法を提案する。提案手法によれば、実験データを最も良く再現するパラメータの推定だけでなく、従来手法ではあまり対象とされていないパラメータ推定量の信頼区間の表現や類似したシミュレーション挙動を示すパラメータ空間の可視化という問題まで含めて統一的に扱うことができる。そして、提案手法を3つのベンチマーク問題への適用し、その有用性を示す。

キーワード シグナル伝達, 生化学反応方程式, システム同定, ベイズ推定, Metropolis-Hastings 法

Bayesian system identification method for biochemical reaction systems

Junichiro YOSHIMOTO^{†,††} and Kenji DOYA^{†,††}

† Initial Research Project, Okinawa Institute of Science and Technology
12-22 Suzaki, Uruma, Okinawa 904-2234, JAPAN

†† Nara Institute of Science and Technology 8916-5 Takayama-cho, Ikoma, Nara 630-0192, JAPAN
E-mail: †{jun-y.doya}@oist.jp

Abstract In a computational approach to life science, computer simulation is essential for verifying hypothetical kinetic models of intracellular molecular mechanisms. However, kinetic parameters used in simulation are often unidentified or unreliable. It is highly desired to develop a method that can automatically estimate the parameters from accessible experimental measurements with high confidence. In this report, we address this estimation problem and present a Bayesian method for solving it. Our contribution is to unify the following three issues in the framework of Bayesian inference: 1) identifying unknown parameters to reproduce experimental measurements; 2) representing confidence intervals of estimated parameters; and 3) visualizing the parameter space in which the kinetic behaviors are similar to experimental measurements. Existing methods rarely dealt with the last two issues. The effectiveness of our method is demonstrated in three benchmark problems.

Key words signal transduction, biochemical kinetics, system identification, Bayesian inference, Metropolis-Hastings method

1. はじめに

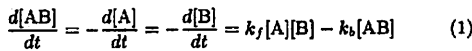
代謝、細胞分化や神経細胞のシナプス可塑性など生命現象を司る分子機構の解明は生命科学における重要な課題の一つである。近年、その研究アプローチの一つとして、細胞内分子反応系の仮説モデルを構築し、計算機シミュレーションによってその妥当性を検証する数値的研究が、生理実験的研究と並行して行われるようになってきた [1]~[3]。このような数値的研究では、シミュレーションの動的特性を決める反応速度係数や

酵素反応係数などのパラメータ設定が仮説の妥当性を根拠付ける鍵となる。多くの場合、これらのパラメータは過去の生理実験に関する文献報告に基づいて設定されるが、ハンドチューニングやランダムサーチによって説明したい現象と一致するようにパラメータ設定されることも少なくない。これは、対応する生理実験がいまだに実施されていなかったり、実験条件によって異なる値が報告されていたりするためである。このような背景から、数値的な意味で合理性が保証されたパラメータ同定法の開発が始められてきている [4]~[6]。

現在のところ、進化アルゴリズムに基づく手法が最も精度良くパラメータを推定できるということがベンチマーク結果から報告されている[4]、利用可能な実験データの情報量が少ない状況で数多くの未知パラメータを点推定しようとする、データの出現の仕方に関して大きな分散を持つことになる。したがって、同定されたパラメータの信頼性を評価することも重要な問題の一つである。また、モデルの構造が複雑になるにたがって実験データを良く再現できる未知パラメータは一点ではなく、ある部分空間を構成するようになるため、そのような部分空間の可視化は同定された動力学特性を調べる上で有用な情報になると考えられる。そこで本研究では、以上を統一的に扱える枠組みとして、ベイズ推定法に基づく生化学反応系のパラメータ同定法を提案し、その有効性を3つのベンチマーク問題への適用例とともに示す。

2. 問題設定

本研究では、空間的に均質な生化学反応の動力学系を対象とする。このような系では各素過程が質量作用の法則にしたがうことを仮定し、その挙動は一階常微分方程式としてモデル化される[7]。例えば、図1(a)のように2つの物質AとBが結合し、物質ABが合成される素過程は以下で表される。



ここで、記号[X]は物質Xの濃度を表している。 k_f (k_b)は、図中左から右への合成(右から左への分解)の速さを表す反応係数である。また、基質Sが酵素Eを仲介して生成物Pへと変化する酵素反応をモデル化する場合、図1(b)のように2段階の合成/分解過程に分解して表現されることもあるが、その近似として以下のMichaelis-Menten方程式が用いられることも多い[7]。



ここで、 k_{cat} は $k_{cat}[E]$ が最大反応速度となるような係数であり、 K_M はMichaelis係数と呼ばれる反応係数である。系全体としての挙動はこれらの素過程の和として与えられるので、対象とする生化学反応系は以下の一階常微分方程式モデルとして一般化できる。

$$\dot{x}(t) = f(x(t), u(t), q_1), \quad t \in [T_0, T_f] \quad (3a)$$

$$x(T_0) = q_0 \quad (3b)$$

ここで、 $\dot{x} \equiv dx/dt$ はベクトル x の時刻 t に関する微分(速度)を表しており、 T_0 と T_f ($T_0 < T_f$)は、それぞれ、着目する生化学反応の初期時刻と終了時刻である。 $x \equiv (x_1, \dots, x_D)'$ は系を構成する D 個の反応物の濃度をベクトルとして表現したものであり、系の状態ベクトルである。プライム記号(')は転置を表している。 q_1 は系の挙動を特徴付ける M 次元反応係数ベクトルである。例えば、図1で示される単一の合成反応のみから構成される系の場合、状態ベクトルおよび反応係数ベクトルは、それぞれ、 $x = ([A], [B], [AB])'$ および $q_1 = (k_f, k_b)'$ で与えら

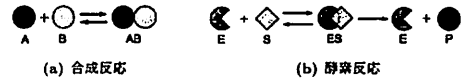


図1 化学反応素過程の例

れる。また、多くの生化学実験では、薬品や酵素の注入など外部から操作可能な条件が存在するため、それらを外部入力ベクトル $u(t)$ として表している。このような外部入力ベクトル $u(t)$ を含む例は、後の4.で紹介する。 $f(\cdot, u(t), q_1): x(t) \mapsto \dot{x}(t)$ は、(1)式や(2)式のように、系を構成する全ての素過程の構造、外部入力ベクトル $u(t)$ 、および、反応係数ベクトル q_1 によって一意に決定される関数である。 q_0 は系の初期条件を表す D 次元ベクトルである。

本研究では、ある生化学実験から得られた反応物の濃度時系列データから(3)式のモデルに基づいて系の振舞いを同定する逆問題を扱う。これは以下のように定式化できる。系を構成する D 個の反応物のうち実験中に計測可能なものの集合を $I_D \subset \{1, \dots, D\}$ で表し、 N 個の時点 $\{t_n \in [T_0, T_f] | n = 1, \dots, N\}$ において反応物の濃度計測が実施されるものとする。そして、時点 t_n における i 番目($i \in I_D$)の反応物の濃度測定値を $y_i(t_n)$ で表すものとする。すなわち、理想的な計測条件下では $y_i(t_n) = x_i(t_n)$ であり、 $x_i(t_n)$ は微分方程式(3)の解である。しかしながら、実際には誤差のない正確な測定は困難であることから、測定誤差 $\epsilon_i(t_n)$ で表し、観測過程が以下で与えられるものと仮定する。

$$y_i(t_n) = x_i(t_n) + \epsilon_i(t_n), \quad n = 1, \dots, N, \quad i \in I_D \quad (4)$$

逆問題の目的は、 $U \equiv \{u(t) | t \in [T_0, T_f]\}$ のもとで測定時系列データ $Y \equiv \{y_i(t_n) | i \in I_D; n = 1, \dots, N\}$ をよりよく再現するパラメータ $q \equiv (q'_0, q'_1)'$ を推定することである。より狭義には、パラメータ q の推定に加えて全ての素過程の構造、すなわち、 f の関数族を同定することまでを含めて逆問題というが、本研究では全ての素過程の構造は既知であることを仮定し、パラメータ q の推定問題のみに注目する。

3. ベイズ推定法によるシステム同定

3.1 ベイズ推定としての定式化

(3)式と(4)式で表現される動力学系の観測過程は、システム工学でいうところの状態空間モデルの一種とみなすことができる。一般の状態空間モデルでは、扱いの容易さから測定誤差の統計的性質を標準偏差 σ_i の定常白色ガウスノイズ、すなわち、 $\epsilon_i(t_n) \sim N(\epsilon_i(t_n); 0, \sigma_i)$ と仮定することが多い^(注1)。また、測定誤差が大きさがその時点での状態量に比例する状況では、 $\epsilon_i(t_n) \sim N(\epsilon_i(t_n); 0, \sigma_i x_i(t_n))$ と仮定することもある。一方で、生化学反応実験では、状態量 $x_i(t_n)$ およびその測定値 $y_i(t_n)$ が常に非負の値を取るという制約があり、実空間中のいたるところで正の確率密度を持つガウス分布によって測定誤差 $\epsilon_i(t_n)$ をモデル化することは、この制約を反することとな

(注1) : $N(x; \mu, \sigma) \equiv \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$

り望ましくないと考えられる。本研究では、このような非負性を考慮して、真の状態 $x_i(t_n)$ が与えられたときの濃度測定値 $y_i(t_n)$ の確率分布が以下のガンマ分布にしたがうことを仮定する。

$$p(y_i(t_n)|x_i(t_n); \gamma_i) = \text{Ga}(y_i(t_n); \gamma_i, x_i(t_n)/\gamma_i) \quad (5)$$

ここで、 $\text{Ga}(x; a, b) \equiv \frac{1}{\Gamma(a)} x^{a-1} e^{-bx}$ であり、 $\Gamma(x) \equiv \int_0^\infty t^{x-1} e^{-t} dt$ はガンマ関数である。この確率分布にしたがえば、 $y_i(t_n)$ が負となる確率が常に 0 となることが保証され、非負性の制約条件が自然に導入できる。また、 γ_i は分布の広がり制御するパラメータであり、その値が十分に大きな場合には、 $\text{Ga}(y_i(t_n); \gamma_i, x_i(t_n)/\gamma_i) \approx N(y_i, x_i(t_n), x_i(t_n)/\sqrt{\gamma_i})$ という近似が成り立つ。すなわち、測定誤差の統計的性質を近似的に $\epsilon_i(t_n) \sim N(\epsilon_i(t_n); 0, x_i(t_n)/\sqrt{\gamma_i})$ で与えていることに対応し、 $1/\sqrt{\gamma_i}$ はノイズの強度を表すパラメータとみなすこともできる。多くの場合、ノイズ強度を実験前に予測することは困難なことから、以下では、 $\gamma \equiv (\gamma_i)_{i \in I_D}$ も未知パラメータの一部とみなし、推定対象となるパラメータの全集合を $\theta \equiv (q', \gamma')$ で表す。そして、 θ をベイズ推定の枠組みで同定することを考える。

(5) 式において添字 (i, n) に関する条件付き独立性を仮定すると、外部入力時系列 U のもとで未知パラメータ θ が与えられた時に、測定時系列データ Y が得られる条件付き確率密度は以下で与えられる。

$$p(Y|\theta; U) = \prod_{n=1}^N \prod_{i \in I_D} \text{Ga}(y_i(t_n); \gamma_i, \gamma_i x_i(t_n; \theta, U)) \quad (6)$$

ここで、 $x_i(t_n; \theta, U)$ は、 θ と U が与えられたときの微分方程式 (3) の解である。さらに、未知パラメータ θ に関する事前知識を事前分布 $p(\theta)$ として反映することができれば、ベイズの定理より以下の関係を導くことができる。

$$p(\theta|Y; U) = \frac{p(Y|\theta; U)p(\theta)}{\int p(Y|\theta; U)p(\theta)d\theta} \quad (7)$$

ここで、 $p(\theta|Y; U)$ は未知パラメータ θ に関する事後分布であり、測定時系列データ Y から未知パラメータ θ が真であると帰納される尤もらしさを確率分布として表現したものである。従来よりパラメータ推定問題によく適用される最尤推定法や最小二乗法などの誤差関数最小化手法[3], [4] では、未知パラメータを点推定するため、測定値がノイズを含んでいる場合には、パラメータ推定量もおおよそデータ量の反比例する分散を持つことになる。これは、短い時間間隔での計測が困難なため、観測時点数 N が小さくなりがちな生化学反応実験データに適用する際には同定パラメータの信頼性という点で問題となる。一方で、ベイズ推定法では、後で述べるように事後分布を利用したパラメータの信頼区間を定量的に評価することが可能である。また、パラメータを最大事後確率 (MAP) 推定量 $\theta_{MAP} \equiv \text{argmax}_\theta p(\theta|Y; U)$ として点推定することも可能である。これらの特徴が、生化学反応実験データに適用する上でのベイズ推定法の大きな利点である。

事前分布の与え方については、いくつかの方法が考えられる。パラメータベクトル q の j 番目の要素を q_j で表すものとし、 q_j の取り得る値には有限区間 $[q_{\min}, q_{\max}]$ に制約されていることが既知の場合には、 $q_i \in [q_{\min}, q_{\max}]$ に対して $p(q_j) = 1/(q_{\max} - q_{\min})$ を満たす一様事前分布を与えればよい。また、 q_j が非負であるという制約しか事前に分からない場合^(注2)でも、非負の q_i に対して $p(q_i) = (\text{定数})$ 、かつ、負の q_i に対して $p(q_i) = 0$ を満たすような事前分布を仮定し、形式上の事後分布を求めることは可能である。しかしながら、この事前分布は特異分布 (improper distribution) となるため、(7) 式の分母で与えられる周辺尤度を求めたい場合には不都合が生じる。本研究では、この問題を避けるために以下の事前分布を与えるものとする。

$$p(q_j) = \frac{1}{2\xi q_j} \exp\left\{-\left|\frac{\log q_j}{\xi}\right|\right\}, \quad j = 1, \dots, D+M$$

ここで、 ξ はある与えられた定数である。また、パラメータ γ に関しては、ノイズの逆分散に対応することから、その値をベイズ推定する時によく用いられるガンマ事前分布 $p(\gamma_i) = \text{Ga}(\gamma_i; \alpha, \beta)$ 、 $i \in I_D$ を与えるものとする[8]。ここで、 α と β はいずれもある与えられた定数である。そして、未知パラメータ θ の全空間に対する事前分布は以下で与えられるものとする。

$$p(\theta) = \left(\prod_{j=1}^{D+M} p(q_j)\right) \left(\prod_{i \in I_D} p(\gamma_i)\right) \quad (8)$$

この事前分布の設定が最適であるという理論的根拠はないが、実際上うまく働くことを 4. の応用例を通して示す。

3.2 実装法

本研究では、実装の容易さから、Metropolis-Hastings (MH) 法[9], [10] 基づいて (7) 式を近似計算する。MH 法の基本原理は以下の通りである。 h 回目の反復での確率的サンプル値を $\theta^{(h)}$ で表すものとし、 $h+1$ 回目の反復手続きを考える。まず、遷移カーネルと呼ばれる $\theta^{(h)}$ のみに依存した条件付き確率分布 $K(\cdot|\theta^{(h)})$ を用いて新たな確率的サンプル候補 $\tilde{\theta}^{(h+1)} \sim K(\tilde{\theta}^{(h+1)}|\theta^{(h)})$ で生成する。ここで、遷移カーネル K は、 $K(\tilde{\theta}^{(h+1)}|\theta^{(h)}) = K(\theta^{(h)}|\tilde{\theta}^{(h+1)})$ という対称性を満たしているものとする。もし、 $p(\tilde{\theta}^{(h+1)}|Y; U) > p(\theta^{(h)}|Y; U)$ を満たしていれば、 $\tilde{\theta}^{(h+1)}$ は常に受理される。この条件が満たされない場合には、 $\frac{p(\tilde{\theta}^{(h+1)}|Y; U)}{p(\theta^{(h)}|Y; U)}$ の確率で、 $\tilde{\theta}^{(h+1)}$ が受理される。以上で $\tilde{\theta}^{(h+1)}$ が受理された場合には、 $\theta^{(h+1)} = \tilde{\theta}^{(h+1)}$ となるが、受理されなかった場合は、 $\theta^{(h+1)} = \theta^{(h)}$ となる。この手続きを無限回繰り返すことができれば、受理されたサンプルの集合 $\{\theta^{(h)}\}$ は真の事後分布 $p(\theta|Y; U)$ にしたがって分布することが理論的に保証されている。その一方で、有限個のサンプル集合 $\{\theta^{(h)}\}$ によって表現される経験分布が真の事後分布 $p(\theta|Y; U)$ に近づく収束速度は遷移カーネル $K(\cdot|\cdot)$ の設計に大きく依存することが知られている[11], [12]。これは、系の構造

(注2)：生化学反応方程式の性質上、パラメータベクトル q の全ての要素は常に非負の値でなければならない。

手続き 1 (修正 MH 法)

```

1 保持するサンプル候補数  $L$  ( $L > D + M + |I_D|$ ) と反復回数  $H$  を適切に設定
2 初期サンプル候補集合  $\{\bar{\theta}^{(l)} | l = 1, \dots, L\}$  を適切に初期化
3 for  $h = 1, \dots, H$ 
4   for  $l = 1, \dots, L$ 
5     ▷ 新しいサンプル候補の生成
6      $l_1$  と  $l_2$  を集合  $\{1, \dots, L\}$  からランダム選択
7     区間  $[0, 1)$  の一様分布  $U(0, 1)$  から  $\eta$  をランダム選択
8      $\bar{\theta}^{(l)} := \bar{\theta}^{(l_1)} + \eta (\bar{\theta}^{(l_2)} - \bar{\theta}^{(l_1)})$ 
9     ▷ サンプル候補の受理・棄却の判定
10     $r$  を  $U(0, 1)$  からランダム選択
11    if  $r < \min \left\{ 1, \frac{p(\bar{\theta}^{(l)})Y; U)}{p(\bar{\theta}^{(l_1)})Y; U} \right\}$ , then
12       $\bar{\theta}^{(l)} := \bar{\theta}^{(l)}$  ▷  $\bar{\theta}^{(l)}$  を受理
13    else if ▷  $\bar{\theta}^{(l)}$  を棄却し、 $\bar{\theta}^{(l)}$  をそのまま保持
14  end if
15 end for
16 end for

```

図 2 事後分布近似のための Metropolis-Hastings 法

によって関数 f の形が大きく変化する生化学反応系のパラメータ推定問題への適用を考えると大きな問題となる。

この問題点を軽減するために、元の MH 法を修正し、適応的に遷移カーネルを構成できる図 2 の手続きを提案する。この手続きが元の MH 法と大きく異なる点は確率的サンプル候補を生成するステップ 5-7 である。これらのステップでは、保持されているサンプル候補集合 $\{\bar{\theta}^{(l)} | l = 1, \dots, L\}$ が局所的な事後分布の形状にしたがって分布し、新たなサンプル候補がこの分布に基づいて生成されるサンプリング方法となっている。このため、元の MH 法が苦手としている事後分布において次元間に強い相関がある状況でも効率よいサンプリングが実行できると期待できる。また、元の MH 法と同様に、ステップ 9 の条件判定では、2 点における事後確率密度を計算する必要があるが、これらの間の比のみを評価すればよいので、(7) 式の分母の積分計算は不要となる。また、(6) 式の $x_i(t_n; \theta, U)$ の評価も、 θ の実現値として $\bar{\theta}^{(l)}$ が与えられるため、任意の常微分方程式数値計算法を用いて求めることが可能である。現在のところ、 $H \rightarrow \infty$ の極限において、受理されたサンプル集合 $\{\bar{\theta}^{(hL+l)} | h = 1, \dots, H; l = 1, \dots, L\}$ が真の事後分布 $p(\theta|Y, U)$ にしたがって分布するという理論的保証はないため、このサンプリング手続きは発見的手法ではあるが、その手続きが Leapfrog 法 [12] や adaptive direction sampling 法 [13] と類似していることから、いずれ理論的な保証も与えられるものと考えられる。

さらに、図 2 の手続きで得られたサンプル集合 $\{\bar{\theta}^{(h)} | h = 1, \dots, LH\}$ から、事後分布 (7) 式の連続性を復元するために、本研究では、以下のカーネル密度推定法 [14] で与えられる $\tilde{p}(\theta|Y; U)$ を近似事後分布とする。

表 1 推定対象となる合成反応系の設定

開始・終了時刻	$T_0 = 0, T_f = 5$
測定時点	$\{t_n = 0.1(n-1) n = 1, \dots, 51\}$
状態ベクトル	$\omega \equiv ([A], [B], [AB])'$
外部入力ベクトル	$u \equiv 0$
測定可能インデックス	$I_D \equiv \{2, 3\}$
反応係数ベクトル	$q_1 \equiv (k_f, k_b)'$
初期状態ベクトル	$q_0 \equiv ([A](0), [B](0), [AB](0))'$
測定ノイズ	$\epsilon_i(t_n) \sim N(\epsilon_i(t_n); 0, \sigma_i x_i(t_n))$
未知パラメータの真値	$q_1 = (5.0, 0.1)'$ $q_0 = (0.4, 0.3, 0.01)'$ $\sigma_i = 0.1$ for all $i \in I_D$

$$\tilde{p}(\theta|Y; U) \approx \frac{1}{LH} \prod_{h=1}^{LH} \mathcal{N}_{\dim(\theta)}(\theta; \theta^{(h)}, V) \quad (9a)$$

$$V \equiv \left(\frac{4}{(\dim(\theta) + 2)LH} \right)^{\frac{2}{\dim(\theta)+4}} \cdot \text{Cov}(\{\theta^{(h)}\}) \quad (9b)$$

ここで、 $\dim(\theta) \equiv D + M + |I_D|$ は未知パラメータ θ の次元数を表しており、 $\mathcal{N}_p(x; \mu, \Sigma)$ は、平均ベクトル μ と共分散行列 Σ の p 次元正規分布にしたがう確率変数 x の確率密度関数である。 $\text{Cov}(\{\theta^{(h)}\})$ は、サンプル集合 $\{\theta^{(h)} | h = 1, \dots, LH\}$ のサンプル共分散行列である。(9) 式で表現される確率密度は混合ガウス分布となるために、事後分布に対する周辺化演算や次元間の共分散計算が容易に実装できるという利点がある。

4. 応用例

本節では、提案手法を 3 つのベンチマーク問題に適用し、その有用性を示す。なお、全ての問題に共通して、事前分布を規定する定数は $\alpha = 10^{-4}$ 、 $\beta = 1$ 、および、 $\xi = 1$ に設定するものとした。また、図 2 の手続きにおける反復回数 H は、受理される総サンプル数が $LH = 10^7$ 個となるように設定し、保持するサンプル候補数 L は、 $\dim(\theta)$ の約 2 から 3 倍程度に設定した。初期サンプル候補集合 $\{\theta^{(h)}\}$ の各要素は、区間 $[0.99 \times 10^{-10}, 1.01 \times 10^{-10}]$ の一様乱数を用いて初期化した。

4.1 部分観測性を持つ合成反応系

まず、比較的小規模な問題例として図 1(a) で示される単一の合成反応からなる系のパラメータ推定問題に提案手法を適用した。ただし、ここでは、全ての反応物の測定が困難な実験状況を仮定し、物質 A が測定不可能であるものとした。したがって、(1) 式で与えられる関数 f の構造は既知であるものの、未知変量に関して双線形となり単純な線形回帰法で未知パラメータを同定することが困難な問題になっている。そして、(3) 式と (4) 式に現れる各変数とパラメータを表 1 のように設定し、模擬的な時系列データ $Y \equiv \{y_i(t_n) | i \in I_D, n = 1, \dots, 51\}$ を用意した。

この時系列情報 Y に対して、図 2 と (9) 式の計算手続きを適用し、近似事後分布 $\tilde{p}(\theta|Y; U)$ を得た。なお、図 2 において保持するサンプル候補数は $L = 20$ に設定した。図 3 は、利用可能な時系列情報 Y 、および、近似事後分布 $\tilde{p}(\theta|Y; U)$ を各未知パラメータ次元に関して周辺化したものを示している。た

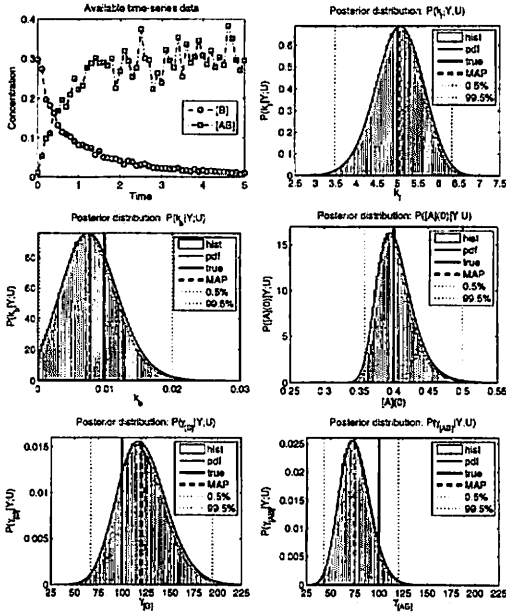


図3 合成反応系同定問題において利用可能な測定時系列データ（左上図）と獲得された近似事後分布（残りの図）

だし、紙面の都合とその推定が容易なことから、周辺化事後分布 $\hat{p}([B](0)|Y;U)$ および $\hat{p}([AB](0)|Y;U)$ に関する図は割愛した。図3において、ヒストグラムはMH法で受理されたサンプル集合 $\{\theta^{(h)}|h=1, \dots, LH\}$ の分布を示したものであり、曲線は(9)式で与えられる近似事後分布を示している。太線で示される実線と破線は、それぞれ、未知パラメータの真値とサンプル集合 $\{\theta^{(h)}|h=1, \dots, LH\}$ から求められるMAP推定量 θ_{MAP} を表している。また、細い点線で示される区間は、推定量の99%信頼区間としてサンプル集合 $\{\theta^{(h)}|h=1, \dots, LH\}$ の0.5%および99.5%百分位点を示したものである。図3に示されるように、MAP推定量 θ_{MAP} は真値とよく一致しており、また、信頼区間も真値を挟むように決定されていることが分かる。2つのパラメータ $\gamma_{[B]}$ と $\gamma_{[AB]}$ に関する信頼区間はかなり広く決定されているように見えるが、 $\sigma_i = 1/\sqrt{\gamma_i}$ の関係に注意すれば、物質B(AB)の測定ノイズ強度 σ に対する信頼区間を $[0.07, 0.12]$ ($[0.09, 0.16]$) と決定していることと等価であり、十分に実用的な信頼区間の設定であるといえる。以上のことから、提案手法はパラメータ推定法として望ましい基本性質を持っていることが分かる。

さらに、提案手法によれば、類似した動的挙動を示すパラメータ空間を可視化することも可能である。図4は、3つのパラメータ $\{k_f, k_b, [A](0)\}$ に注目し、それらの2変量同時事後分布を示したものである。ここで、白丸で示された点はパラメータの真値を示している。この問題では、確率密度の高い領域は等方円状に広がるのではなく、ある曲面に縮退した形で広がっていることが分かる。これは、類似した時系列を復元するため

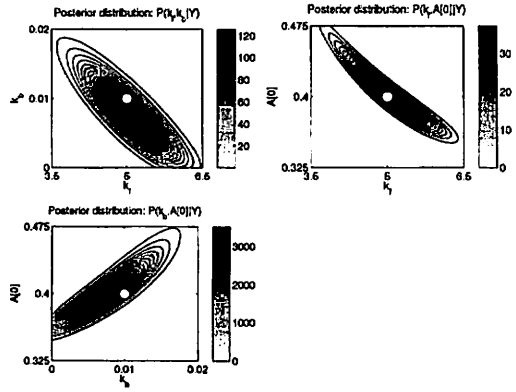


図4 3つの未知パラメータに関する2変量同時事後分布（合成反応系同定問題）

には、未知パラメータの次元間である拘束条件を満たす必要があることを意味している。特に、 $(k_f, A[0])$ に関する同時事後分布の図では、確率密度の高い領域が双曲線上に現れ、(1)式における双線形構造の拘束が事後分布として反映されていることが分かる。

4.2 バイオリアクタ

次に、酵素反応を含む系としてKremlingらがバイオリアクタシステムを想定して開発したベンチマーク問題[5]に提案手法を適用した。文献[5]では、事前に想定される2種類の反応構造からいずれが尤もらしいかを選択するモデル判別問題も扱っているが、ここではその構造は既知として、反応系の挙動が以下の微分方程式にしたがうものと仮定した。

$$d[B]/dt = (\mu - q_{in})[B] \quad (10a)$$

$$d[S]/dt = q_{in}(c_{in} - [S]) - r_1 M_w[B] \quad (10b)$$

$$d[M1]/dt = r_1 - r_2 - \mu[M1] \quad (10c)$$

$$d[M2]/dt = r_2 - r_3 - \mu[M2] \quad (10d)$$

$$d[M3]/dt = r_{syn} - \mu[M3] \quad (10e)$$

$$r_1 = \frac{r_{1,max}[S]}{K_s + [S]}; \quad r_2 = \frac{k_2[M3][M1]}{K_{M1} + [M1]}; \quad r_3 = \frac{k_{1,max}[M2]}{K_{M2} + [M2]}$$

$$\mu = Y_{x/s} \cdot r_1; \quad r_{syn} = \frac{k_{syn,max} \cdot K_{IB}}{K_{IB} + [M2]}$$

ここで、6つの反応係数パラメータ $r_{1,max} = 2.4 \times 10^4$, $K_s = 0.4437$, $K_{M1} = 12.2$, $r_{3,max} = 3.0 \times 10^6$, $K_{M2} = 10.0$, $M_w = 3.423 \times 10^{-4}$ は文献[5]と同様に既知であるものとし、4つの反応係数 $q_i \equiv (Y_{x/s}, K_2, k_{syn,max}, K_{IB})'$ を推定するものとした。 $u \equiv (c_{in}, q_{in})'$ はバイオリアクタへ入力される基質の濃度と流量を表す変数であり、この問題における外部入力ベクトルとなる。状態変数ベクトルは $x \equiv ([B], [S], [M1], [M2], [M3])'$ であり、全ての反応物の濃度が測定可能であるものと仮定した。測定時系列データ Y は、Kremlingらが公開しているウェブページ^(注3)からデフォルトの設定で生成されるものを利用した。

(注3) : <http://sysbio.ist.uni-stuttgart.de/projects/benchmark/>

表2 バイオリアクタ問題に対する推定結果

反応係数	真値	MAP	99%信頼区間
$Y_{x/s}$ ($\times 10^{-6}$)	7.00	6.95	[6.79, 7.11]
K_2 ($\times 10^{+6}$)	6.00	5.98	[5.56, 6.42]
k_{synmax} ($\times 10^{-2}$)	1.68	1.17	[0.89, 1.63]
K_{IB} ($\times 10^{-2}$)	1.00	1.95	[0.99, 4.56]

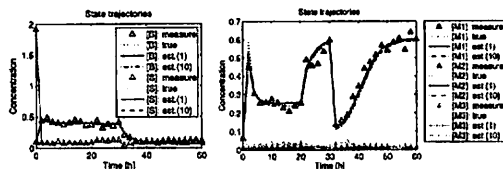


図5 推定値から再現される状態ベクトル軌道 (バイオリアクタ問題)

表2は、提案手法によって求められた反応係数ベクトル q_1 に関するMAP推定値および99%信頼区間を示したものである。 $Y_{x/s}$ および K_2 に関しては、非常に正確に推定できていることが分かる。一方で、 k_{synmax} および K_{IB} に関する推定値は、信頼区間から真値が外れるなど比較的大きな誤差が見える。この原因を調べるために、真の状態ベクトル軌道とMAP推定値 θ_{MAP} によって再現される状態ベクトルの軌道を比較した。図5はその比較を示したものであり、実線(点線)は θ_{MAP} によって再現される状態ベクトル軌道(真の状態ベクトル軌道)を示している。また、記号“ Δ ”は測定時系列データ Y の値を示している。この図から、再現される状態ベクトル軌道は真のものと同様に類似しており、測定値との誤差も非常に小さいことが分かる。したがって、測定時系列が持つ情報量自体が正確なパラメータ推定を行うためには不十分なものであったと考えられる。実際に、全く同条件の実験が10回繰り返されることを仮定し、その全ての測定データから推定されるパラメータによって状態ベクトル軌道を再現すると、図5の破線で示されるものとなり、真の状態ベクトル軌道とほぼ正確に一致していることが分かる。

4.3 代謝パスウェイモデル

未知パラメータ数の次元が比較的大きな問題として、Molesらがパラメータ推定法の性能評価に用いた代謝パスウェイモデルのパラメータ推定問題[4]に提案手法を適用した。想定する反応構造や利用可能な観測時系列データ Y は、文献[4]と同じものを用いて、未知とされる36次元の反応係数ベクトル q_1 の推定精度を調べた。図6はその結果を示したものである。ここで、横軸は反応係数ベクトル q_1 の要素番号を表しており、縦軸は相対誤差^(注4)を示したものである。また、“SRES”は文献[4]において進化アルゴリズムの一種を用いて得られた最良の推定値を示しており、“Bayesian”は我々の提案手法から得られたMAP推定値を示している。特に最大誤差に注目すると、提案手法は従来手法に比べて大きく改善されており、全体的に見ても提案手法による方が精度良くパラメータ同定されていること

(注4)：相対誤差は、推定値と真値との差の絶対値を真値で除算したものと定義している。

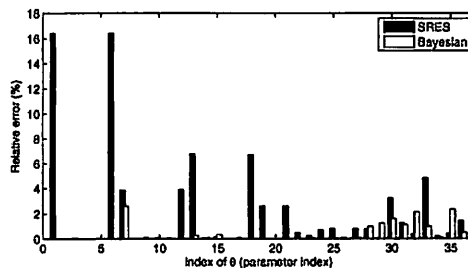


図6 代謝パスウェイモデルを用いた性能評価結果

が分かる。

5. むすび

本報告では、ベイズ推定法に基づく生化学反応系のパラメータ同定法を提案した。提案手法の特徴は、実験データを最も良く再現するパラメータの同定のみだけでなく、従来手法ではあまり対象とされていなかったパラメータ推定値の信頼区間の表現や類似したシミュレーション挙動を示すパラメータ空間の可視化という問題まで含めて統一的に扱うことができる点である。現在のところパラメータ推定問題のみを扱っているが、将来的には、ベイズのモデル選択基準を利用して化学反応構造を同定できる手法へと拡張する予定である。また、現在、提案手法をより広い問題に適用できるようにSBML[15]互換のライブラリ開発を進めているところであり、近々その報告をする予定である。

文 献

- [1] Bhalla, U. S. and Iyengar, R. *Science* **387**, 283-381 (1999).
- [2] Kuroda, S. et al. *The Journal of Neuroscience* **21**(15), 5693-5702 (2001).
- [3] Doi, T. et al. *The Journal of Neuroscience* **25**(4), 950-961 (2005).
- [4] Moles, C. G. et al. *Genome Research* **13**, 2467-2474 (2003).
- [5] Krennling, A. et al. *Genome Research* **14**, 1773-1785 (2004).
- [6] Sugimoto, M. et al. *BioSystems* **80**, 155-164 (2005).
- [7] Fall, C. P. et al., editors. *Computational Cell Biology*. Springer, (2000).
- [8] Gamerman, D. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall, (1997).
- [9] Metropolis, N. et al. *Journal of Chemical Physics* **21**, 1087-1091 (1953).
- [10] Hastings, W. K. *Biometrika* **57**, 97-109 (1970).
- [11] Andrieu, C. et al. *Machine Learning* **50**, 5-43 (2003).
- [12] MacKay, D. C. *Information theory, Inference, and Learning algorithms*. Cambridge University Press, (2003).
- [13] Gilks, W. R. et al. *The Statistician* **43**(1), 179-189 (1994).
- [14] Silverman, B. W. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, (1986).
- [15] Hucka, M. et al. *Systems Biology Mark-up Language Level-1*. ERATO Kitano Symbiotic Systems Project, (2001).