

## 解 説

## 対数線形モデルによるマルチニューロンスパイクデータ解析

—一定常スパイク相関解析から動的スパイク相関解析への橋渡し—

島 崎 秀 昭

理化学研究所 脳科学総合研究センター\*

Analysis of Multiple Neural Spike Train Data using the Log-linear Model  
—From Stationary to Time-Varying Spike Correlation—

Hideaki Shimazaki

RIKEN Brain Science Institute\*

## 概要

神経細胞集団は協調して情報処理を行っています。特に比較的短い時間内に複数の細胞が発火する同期発火現象はセル・アセンブリ仮説やシンファイア・チェーンといった細胞集団による情報処理の仮説との関係から注目を集めてきました。近年実データの同期的発火構造を調べるために対数線形モデルを用いた解析が広く行われています。しかしこれまでの対数線形モデルは定常性を仮定しており行動下の動物から得られるスパイクデータへ適用することはできませんでした。本稿は対数線形モデルの基本事項、パラメータの基本的な推定法（最尤推定・MAP 推定）を経て、最終的に動的な相関構造の推定を可能にする非線形フィルタの構築に至るまでを紹介します。

## 1. はじめに

神経細胞は集団で協調して情報を処理していると考えられています。複数の細胞が依存しあって活動を行っているとなると、観測された同時記録スパイク時系列には同期的な集団発火が観測されるかもしれません。Schneidman らの論文<sup>1)</sup>を皮切りに、同期発火活動の統計モデルとして統計物理学のイジングモデル（スピングラスモデル）を援用するデータ解析が広く行われています。さらに最近では2次のスパイク相関では説明のできない、高次のスパイク相関構造にも注目が集まっています<sup>2)</sup>。そこで、本稿では低次から高次までのスパイク相関を統一的に扱うことのできる対数線形モデルを用いたマルチニューロンのスパイクデータ解析を説明します。

多くの対数線形モデルによる解析は定常性、すなわち個々の細胞の発火頻度や相関構造が時間に依存しないことを仮定しています。そのため、これまでの実データの解析はスライス網膜神経節細胞（Retinal ganglion cell）や皮質培養細胞、麻酔下のサル（*Macaca mulatta*）の皮質神経細胞な

ど、この仮定が成り立つと期待できる条件下で記録されたスパイクデータに対して適用されてきました。一方で行動下の動物の神経細胞の活動はよりダイナミックです。刺激や行動に応じて個々の細胞の発火頻度は時間的に変動します。同一の試行を繰り返し行い peristimulus time histogram (PSTH) やカーネル密度推定法などで個々の神経細胞の発火頻度の刺激依存のダイナミクスを調べることは電気生理実験の基本的な手法です。さらに発火頻度だけでなくスパイク相関構造もダイナミックに変動する可能性があるとするればどうでしょうか。実際 Cross correlogram, Joint PSTH, Unitary Event 法を用いた解析で皮質や海馬体の2つの神経細胞の相関が行動に依存して動的に変動することが示されています<sup>3–6)</sup>。こうしたデータに対して通常の対数線形モデルを用いた解析を行うことはできません。そこで本稿では定常モデルから出発して、こうしたダイナミックなデータを扱うための状態空間モデルによるアプローチ<sup>7)</sup>を駆け足で紹介します。

## 2. 高次スパイク相関とはなにか

議論の始めとして、まず本稿で扱う同時刻スパイク

\* 〒 351-0198 埼玉県和光市広沢 2 番 1 号

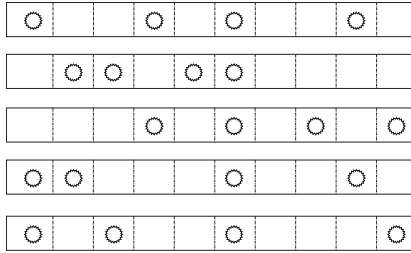


図1 秘密の暗号で仲間探し.

相関とはなにかを明らかにしなければいけません. ここではまだ上述のようなダイナミクスは考えないことにしましょう. 2次の相互(交互)作用項までを考えた対数線形モデルは多くの文献がありますので, 以下では高次スパイク相関とはなにかを紹介しようと思います.

それでは始めに次のような例から考えて見ましょう. ある学校のクラスの5人の生徒たちに先生が紙切れを配りこう言いました.「この紙には秘密の暗号で君達の仲間の情報が書かれています. 暗号を解読して自分の仲間を見つけてください.」紙切れにはパンチカードのように所々穴が開けられています(図1). どうやら同じ箇所穴が開けられていると仲間である確信が持てそうなのですが…まず生徒は隣同士で紙を見せ合ってみます.すると1~2箇所は穴が一致しています. 隣の子は仲間なのでしょう. 紙切れをよく見るとそれぞれの紙には10個の目目のうち4つに穴が開けられています. 2つの紙にランダムに穴を開けても同じ箇所に穴が開いている確率は $0.4 \times 0.4 = 0.16$ ですから10箇所のうち1~2箇所は穴が一致していても不思議ではなさそうです. それでは次に3人が集まって紙を見せ合ってみます. 1箇所ですべての紙に穴が開いています. しかしやはりランダムに穴が開けられていると仮定した場合,  $(0.4)^3 = 0.064$ の確率で3箇所すべてが一致することがあります. 穴が一致した頻度は偶然から考えられるより多いですが, 3人が仲間であると確信は持てません. そこで5人が集まってみたところ5人の全員の紙に同じ箇所で穴が開いていました. ランダムに穴を開けた場合その確率は $(0.4)^5 = 0.0102$ ですから, これは故意に同じ場所に穴を開けたと考えるても良さそうです. どうやら5人全員が仲間だという結論に達しました.

部分だけを見てもわからない依存関係, それが高次の依存関係です. 神経スパイクの話に戻りましょう. 複数の神経細胞が低頻度でばらばらと独立発火してい

るなかで, 非常にまれに全ての神経細胞が同期発火する状況を想像してください. このようなスパイク時系列では低次のスパイク相関が存在せず高次のスパイク相関だけが存在する可能性があります. その理由は, 全ニューロンの同期発火という異常事態が部分集団で見た場合には偶然に起こりうる同期発火と見分けがつかないからです. 私を含めて発火頻度や2次のスパイク相関だけでなく高次のスパイク相関も神経細胞が協調して情報処理を行う様子を明らかにするために重要であると考えられる研究者がいます. なぜ, 高次スパイク相関に興味があるのでしょうか. このことを説明するために, もう一度例を挙げて説明してみましょう.

上述のパンチカードの穴をTwitterのようなネット上での発信(発言や会話)に置き換えて考えてみましょう. ある適当な時間区間(例えば10分程度)の中で一回でも発信をしていれば1, 発信がなければ0とする二値時系列を考えます. 二人の人が時間区間に定期的に会話をしていればその二値時系列は高い2次相関を持つことが期待できます. また, 各人が勝手に発言していれば独立な二値時系列です. さて, ここで独立に発言していた人々が突然一斉に発言を始めたとしましょう. この現象をどう解釈するのが適当でしょうか(発言内容は見ることはできずとします). 1つの解釈として, 何か特定の事件が起きてそれに各人が独立に反応したとみなすことができそうです. すなわち観測できない共通の外部入力による影響が高次相関に現れる可能性があります.

どんなに観測技術が発達しても私達が観測できるのは全体のごく一部です. その一部の情報を使って全体の様子を最大限推察したいという希望が高次相関解析の動機の一つに挙げられます. こうした高次のスパイク依存関係を簡単に作り出してくれるのが, 次に紹介する対数線形モデルです.

### 3. 対数線形モデル

本節では対数線形モデルの定義と良く知られている性質を簡単にまとめます. 本節は多くの優れた文献を参考に本稿の目的である動的スパイク相関の推定手法の構築に必要な諸性質のみをまとめたものです. より詳しい内容は原著論文<sup>8-10)</sup>を, より包括的な文脈での解説は参考文献<sup>11-13)</sup>等をご覧ください.

$N$ 個のニューロンによる同時刻スパイクパターンを考えます. ひとつひとつのニューロンの状態は二値の確率変数(binary random variable)  $X_i$  ( $i = 1, \dots, N$ )で表されるとします. ここでは, '1'はスパイク, '0'はスパイクが生成されなかったことを表すとします.

$N$  個のニューロンの集団スパイクパターンはベクトル  $\mathbf{X} = (X_1, X_2, \dots, X_N)$  で表されます。すべての可能なパターンの数は  $2^N$  個です。

$N$  個のニューロンの同時確率関数 (joint probability mass function) を考えます。ニューロンがとりうる値として二値変数  $x_i = 1$  or  $0$  ( $i = 1, \dots, N$ ) を使って同時確率関数を

$$P(X_1 = x_1, \dots, X_n = x_n) = p(\mathbf{x}). \quad (1)$$

と書くことにしましょう。右辺では簡単のためにベクトル  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  を使いました。確率関数としての制約  $\sum p(\mathbf{x}) = 1$  がありますから、すべてのスパイクパターンの確率は  $2^N - 1$  個のパラメータを決めれば定まります。当然  $2^N - 1$  個のスパイクパターンのそれぞれについて、直接そのパターンが生じる確率  $p(\mathbf{x})$  を決めれば同時確率関数は定まります。しかし、それ以外にも同時確率関数を定める便利なパラメータの組があります。情報幾何の分野では、 $2^N - 1$  個のパラメータをすべての分布  $\{p(\mathbf{x})\}$  をカバーする  $2^N - 1$  次元の‘座標’と見なします。座標系の一点・一点が特定の確率分布を表しています。座標系はある確率分布から別の確率分布へ移動するときの道先案内人の役を担います<sup>11)</sup>。ここでよく登場するのが  $\theta$ -座標系と  $\eta$ -座標系です。

### 3.1 $\theta$ -座標系と $\eta$ -座標系

同時確率関数の対数を次のような展開式で表します。

$$\log p(\mathbf{x}) = \sum_i \theta_i x_i + \sum_{i < j} \theta_{ij} x_i x_j + \dots \\ \dots + \theta_{1\dots N} x_1 \dots x_N - \psi(\theta). \quad (2)$$

ここで  $\theta = (\theta_1, \theta_2, \dots, \theta_{12}, \theta_{13}, \dots, \theta_{1\dots N})'$ 。本稿ではプライム記号で転置を表します。 $\psi(\theta) = -\log p(\{0, \dots, 0\})$  は  $\sum_{\mathbf{x}} p(\mathbf{x}) = 1$  を満たすように決められる (対数) 正規化項です。そのため  $\psi(\theta)$  は  $\theta$  に関する複雑な式になっています。式 (2) が対数線形モデル (log-linear model) です。パラメータ  $\theta_i, \theta_{ij}, \dots, \theta_{1\dots N}$  は指数分布族の自然パラメータあるいは正準パラメータと呼ばれ、‘ $\theta$ -座標系’を構成します。

もうひとつ重要な座標系が  $\eta$ -座標系です。 $\eta$ -座標系は次のように構成されます：

$$\eta_i = E[X_i], \quad i = 1, \dots, N \\ \eta_{ij} = E[X_i X_j], \quad i < j \\ \vdots \\ \eta_{1\dots N} = E[X_1 \dots X_N]. \quad (3)$$

これらのパラメータは同時スパイク生成の期待値を表

します。これを同時スパイクレートと呼ぶことにします。どのニューロン同士の同期発火の期待値なのかは添え字で表現されています。以下ではこれらをベクトル  $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_{12}, \eta_{13}, \dots, \eta_{1\dots N})'$  を用いて表します。

式 (2) 及び式 (3) は次のような‘マルチインデックス’表記でコンパクトに書くことができます。 $\Omega_k$  を 1 から  $N$  までの自然数から  $k$  個を取り出してきた時のすべての組としましょう： $\Omega_1 = \{1, 2, \dots, N\}$ ,  $\Omega_2 = \{12, 13, \dots\}$ ,  $\Omega_3 = \{123, 124, \dots\}$ , etc. この要素  $I \in \{\Omega_1, \dots, \Omega_N\}$  を使って、自然パラメータ・期待値パラメータを  $\theta_I, \eta_I$  と表すことができます。同様に式 (2) の相互 (交互) 作用項を

$$F_i(\mathbf{x}) = x_i, \quad i = 1, \dots, N \\ F_{ij}(\mathbf{x}) = x_i x_j, \quad i < j \\ \vdots \\ F_{1\dots N}(\mathbf{x}) = x_1 \dots x_N. \quad (4)$$

とすれば、 $F_I(\mathbf{x})$  ( $I \in \{\Omega_1, \dots, \Omega_N\}$ ) を使って式 (2)・(3) は簡潔に

$$\log p(\mathbf{x}) = \sum_I \theta_I F_I(\mathbf{x}) - \psi(\theta), \quad (5)$$

$$\eta_I = E[F_I(\mathbf{X})]. \quad (6)$$

と書くことができます。上記の式からわかるように式 (5) の対数線形モデルは連結関数 (ここでは多変量ロジット関数) をとった後にパラメータに対して線形和であれば良く、特徴関数と呼ばれる  $F_I(\mathbf{x})$  の部分はデータを非線形に加工しても良いわけです。そうすると式 (4) による展開以外にもっと良いデータの加工の仕方があっても良いように思えます。とても興味深いところですが、今回は式 (4) による展開を用いた相関解析に焦点を当てて先を急ぎましょう。

### 3.2 対数線形モデルのキュミュラント

先ほど分布の正規化に用いた関数  $\psi(\theta)$  は  $\theta$  に関する複雑な関数だと書きました。実際に  $\psi(\theta)$  を書くとき次のようになります。

$$\psi(\theta) = \log \left[ \sum_{\mathbf{x}} e^{\sum_I \theta_I F_I(\mathbf{x})} \right]. \quad (7)$$

一般に指数分布族のキュミュラント母関数  $K(\mathbf{u})$  は  $K(\mathbf{u}) = \psi(\theta + \mathbf{u}) - \psi(\theta)$  で与えられ、確率変数  $F_I(\mathbf{X})$  の高次のキュミュラントは  $\psi(\theta)$  の高階導関数で表されます。 $\theta_I$  による変微分を  $\partial_I$  と書くことにします。対数正規化項  $\psi(\theta)$  を  $\theta_I$  に関して一階微分すると

$$\begin{aligned} \partial_I \psi(\boldsymbol{\theta}) &= E[F_I(\mathbf{X})] \\ &= \eta_I. \end{aligned} \tag{8}$$

次に  $\psi(\boldsymbol{\theta})$  の二階導関数は二次のキュミュラントを与えます。

$$\begin{aligned} \partial_I \partial_J \psi(\boldsymbol{\theta}) &= E[(F_I(\mathbf{X}) - \eta_I)(F_J(\mathbf{X}) - \eta_J)] \\ &= \eta_{I \cup J} - \eta_I \eta_J. \end{aligned} \tag{9}$$

ここで記号  $\cup$  は本稿だけで使う特別な表記です。マルチインデックスの要素を集合とし、和集合を再びマルチインデックス表記にしたものとします。たとえば  $I = 12$  及び  $J = 234$  のとき  $I \cup J = 1234$  とします。 $\partial_I \partial_J \psi(\boldsymbol{\theta})$  はフィッシャー計量で  $\partial_I \partial_J \psi(\boldsymbol{\theta}) = g_{IJ}$  と書くことにします。これから様々なシチュエーションで対数線形モデルの2次近似式を使います。そこで上記の式 (8)・(9) はとても重要な役割を担います。

### 3.3 部分モデルと高次スパイク相関

対数線形モデルの高次相互作用項の  $\theta$ -パラメータは高次のスパイク相関を表します。ただし  $\theta$ -座標系によるスパイク相関は対応する  $\eta$ -座標系との関係において解釈されます。具体的には非ゼロの  $r$  次以上の相互作用項  $\{\theta_I\}$  ( $I \in \{\Omega_r, \dots, \Omega_N\}$ ,  $r \leq N$ ) は  $r$  次以上の同期スパイクレート  $\{\eta_I\}$  ( $I \in \{\Omega_r, \dots, \Omega_N\}$ ) が  $r$  次未満の同期スパイクレート  $\{\eta_i\}$  ( $i \in \{\Omega_1, \dots, \Omega_{r-1}\}$ ) から偶発的に起こるとして期待される値 (チャンスレベル) より多いかあるいは少なくなることを表しています。甘利ら<sup>8-10)</sup> は  $\theta$ -座標系と  $\eta$ -座標系の双対性を巧みに利用した情報幾何の枠組みによりこれを簡潔に説明しました。ここでは情報幾何の詳細には踏み込まず、甘利<sup>9)</sup> に従ってそのエッセンスだけを紹介します。

$N$  個のニューロンのスパイクパターンの分布  $q(\mathbf{x})$  を考え、その分布の  $r$  次以下の同期スパイクレートを  $\{\eta_i\}$  ( $i \in \{\Omega_1, \dots, \Omega_r\}$ ) とします。次に  $r$  次以下の同期スパイクレートが  $q(\mathbf{x})$  と同じモデル  $\tilde{p}(\mathbf{x})$  を考えます。この制約の下で  $\tilde{p}(\mathbf{x})$  は  $r$  次より大きい同期スパイクレート  $\{\tilde{\eta}_I\}$  ( $i \in \{\Omega_{r+1}, \dots, \Omega_N\}$ ) を自由にとることができ、これによって  $\tilde{p}(\mathbf{x})$  のスパイクの高次相関構造が変わります。図2上部ではこのモデル群を  $M_r$  と表しています。この中で  $\tilde{p}(\mathbf{x})$  のエントロピーを最大にするモデルがあるはずで、これを  $p^*(\mathbf{x})$  としておきます。一方、対数線形モデルにおいて使用する相互作用項を1次から  $r$  次までのみに制限した部分モデルを考え  $E_r$  と書くことにします (図2下)。  $E_r$  に含まれるモデルをここでは  $p(\mathbf{x})$  とします。再び上図に戻り  $M_r$  上で  $p(\mathbf{x})$  に最も近い分布を探します (ここで‘近い’・‘遠い’ という概念にはカルバック距離を

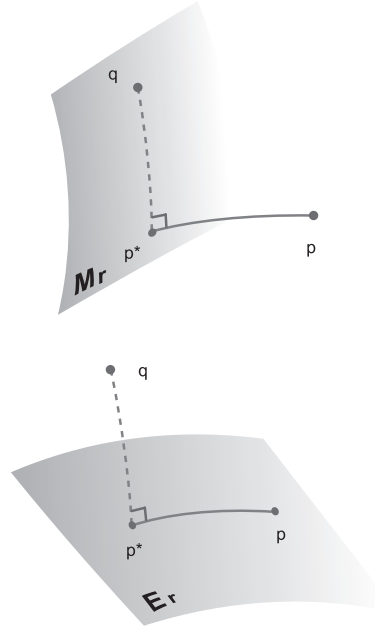


図2 高次スパイク相関の幾何的な説明。

使います)。するとこの作業は  $\tilde{p}(\mathbf{x})$  のエントロピーを最大にすることに等しく、 $p(\mathbf{x})$  に最も近い  $M_r$  上のモデルは最大エントロピーモデル  $p^*(\mathbf{x})$  であることが示せます。次に下図の  $E_r$  の中で  $q(\mathbf{x})$  に最も近いモデル  $p^{**}(\mathbf{x})$  を探します。すると  $p^{**}(\mathbf{x})$  は最大エントロピーモデル  $p^*(\mathbf{x})$  に一致することが ( $\theta$ -座標系と  $\eta$ -座標系の双対性に基づくピタゴラスの定理によって) 示されます。

$q(\mathbf{x})$  をデータ分布 (経験分布) とし、これに最も近い  $E_r$  上の分布を探す作業を最尤法といいます。従って  $r$  次以下の同期スパイクレートをデータ分布に一致させるという制約の下で最大エントロピーモデルを求めることは、データに対して  $r$  次までの対数線形モデルを最尤法によってフィッティングすることと同じです。多くの文献では最尤推定を行っており、そのため対数線形モデルは最大エントロピーモデルと呼ばれることも多いです。  $r$  次の最大エントロピーモデルでは  $r+1$  次以上の同期スパイクレートに対して制約が設けられていないため、これらは ( $r$  次以下の同時スパイクレートの制約の下で) 偶発的に生じます。逆に非ゼロの  $r+1$  次以上の相互作用項  $\{\theta_I\}$  はこの偶然に起こる頻度からのずれを表します。これが対数線形モデルの相互作用項によって定義される高次スパイク相関の意味になります。

#### 4. 定常モデルのデータへの適用

##### 4.1 最尤推定

次に  $r$  次の定常対数線形モデルのデータへのフィッティングに話を移しましょう. まず考えられるのが上で挙げた最尤法によるフィッティングです.

同時記録された複数のニューロンのスパイク時系列を幅  $\Delta$ [s] のビンに区切ります. ビンの数は  $n$  個あったとします (観測時間が  $n\Delta$ ).  $t$  番目のビンの観測を二値データで表し  $\mathbf{X}^t = (X_1^t, X_2^t, \dots, X_N^t)$  とします. ここで '1' の値を含むビンはその中にひとつあるいはそれ以上のスパイクが存在したことを示します. また '0' の値はそのビンの中にスパイクがひとつもなかったことを示します. 対数線形モデルの式にデータを投入してパラメータの関数と見る尤度関数を考えます.

$$\begin{aligned} p(\mathbf{X}^1, \dots, \mathbf{X}^n | \boldsymbol{\theta}) &= \prod_{t=1}^n p(\mathbf{X}^t | \boldsymbol{\theta}) \\ &= \prod_{t=1}^n \exp\left[\sum_{I \in \Omega_{1:r}} \theta_I F_I(\mathbf{X}^t) - \psi(\boldsymbol{\theta})\right] \\ &= \exp\left[n\left(\sum_{I \in \Omega_{1:r}} \theta_I \frac{1}{n} \sum_{t=1}^n F_I(\mathbf{X}^t) - \psi(\boldsymbol{\theta})\right)\right] \\ &= \exp\left[n\left(\sum_{I \in \Omega_{1:r}} k_I \theta_I - \psi(\boldsymbol{\theta})\right)\right]. \end{aligned} \quad (10)$$

始めの等式ではデータの定常性を仮定して, かつビン毎に  $n$  回独立なサンプリングを行ったという仮定を用いています. 実際にビン毎にサンプルが独立かというところかなり怪しい面があることは否めません. 神経細胞には応答期が存在し, ある時刻におけるスパイクの発生率は過去のスパイク, なかでも直前のスパイクからの経過時間に大きく依存すると考えられています. また複数神経細胞を考えた場合, 直接的なシナプス結合及びネットワークを介した間接的な影響で過去の他の神経細胞のスパイクの活動履歴がある神経細胞のスパイク生成率に影響を与えることが考えられます. 実際, スパイク履歴の影響を導入しその代わりに同時刻の条件付独立を仮定するモデルが発展しています<sup>14)</sup>. 一方で本稿のように複数ニューロンの同時刻での同時分布を考えたい場合, 加えて観測履歴もうまく導入するためには少し腰をすえて考える必要があります. 二兎追うものは一兎も得ず... とは言いませんが, 今回はビン毎の独立性を仮定して話を進めましょう.

最後の等式で出てくる  $k_I$  ( $I \in \{\Omega_1, \dots, \Omega_r\}$ ) は次のように定義しました.

$$k_I = \frac{1}{n} \sum_{t=1}^n F_I(\mathbf{X}^t). \quad (11)$$

つまり  $k_I$  は同時スパイクレート (単位は  $\Delta$ ) の観測値であり,  $\eta_I$  の推定量です. 式 (10) を見ると  $r$  次までの相互作用項を入れたモデルの推定には  $r$  次の同時スパイクレート  $\mathbf{k}$  ( $\mathbf{k} = [k_1, \dots, k_{12}, \dots, k_{1\dots r}, \dots]'$ ) があれば十分であることがわかります. 例えば 2 次までの相互作用を考えるモデルの最尤推定には, 発火レートと二つのニューロン間の同時スパイクレート (ただしすべての組について) があれば十分で, それ以上の高次の同時スパイクレートはパラメータの推定値に影響を与えません. このような性質を持つデータの縮約量を十分統計量といいます.

最尤推定量を  $\boldsymbol{\theta}_{\text{MLE}}$  とすると形式的に

$$\boldsymbol{\theta}_{\text{MLE}} = \arg \max_{\boldsymbol{\theta}} \log p(\mathbf{k} | \boldsymbol{\theta}). \quad (12)$$

与えられますが, 実際に最適解を求めるには次のようになります. 目的関数を

$$\begin{aligned} f(\boldsymbol{\theta}) &= -\log p(\mathbf{k} | \boldsymbol{\theta}) \\ &= -n \{ \mathbf{k}'\boldsymbol{\theta} - \psi(\boldsymbol{\theta}) \}. \end{aligned} \quad (13)$$

として  $f(\boldsymbol{\theta})$  を最小にする  $\boldsymbol{\theta}$  を求めることを考えます.  $f(\boldsymbol{\theta})$  は  $\boldsymbol{\theta}$  に関して凸関数になることが保証されています. そこで  $f(\boldsymbol{\theta})$  を簡単な二次近似の式に直してから近似式の最小値を求めることを繰り返していきます.

$$\begin{aligned} \boldsymbol{\theta}^{\text{new}} &= \min_{\boldsymbol{\theta}} \{ f(\boldsymbol{\theta}^{\text{old}}) + \langle \nabla f(\boldsymbol{\theta}^{\text{old}}), \boldsymbol{\theta} - \boldsymbol{\theta}^{\text{old}} \rangle \\ &\quad + \frac{1}{2\epsilon} \|\boldsymbol{\theta} - \boldsymbol{\theta}^{\text{old}}\|^2 \}. \end{aligned} \quad (14)$$

近似式の最小値はその一回微分が  $\mathbf{0}$  になるところです.

$$\nabla f(\boldsymbol{\theta}^{\text{old}}) + \frac{1}{\epsilon} (\boldsymbol{\theta} - \boldsymbol{\theta}^{\text{old}}) = \mathbf{0}. \quad (15)$$

従って更新式

$$\boldsymbol{\theta}^{\text{new}} = \boldsymbol{\theta}^{\text{old}} - \epsilon \nabla f(\boldsymbol{\theta}^{\text{old}}) \quad (16)$$

が得られます. この式是最急降下法 (Gradient descent) の公式です. もう少し 2 次の項の近似を良くするように二階微分 (ヘシアン) を入れます.

$$\begin{aligned} \boldsymbol{\theta}^{\text{new}} &= \min_{\boldsymbol{\theta}} \{ f(\boldsymbol{\theta}^{\text{old}}) + \langle \nabla f(\boldsymbol{\theta}^{\text{old}}), \boldsymbol{\theta} - \boldsymbol{\theta}^{\text{old}} \rangle \\ &\quad + \frac{1}{2\epsilon} (\boldsymbol{\theta} - \boldsymbol{\theta}^{\text{old}})' [\nabla \nabla f(\boldsymbol{\theta}^{\text{old}})]^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}^{\text{old}}) \}. \end{aligned} \quad (17)$$

この式から自然勾配法<sup>15)</sup>と呼ばれる式が得られます.

$$\boldsymbol{\theta}^{\text{new}} = \boldsymbol{\theta}^{\text{old}} - \epsilon [\nabla \nabla f(\boldsymbol{\theta}^{\text{old}})]^{-1} \nabla f(\boldsymbol{\theta}^{\text{old}}). \quad (18)$$

$\epsilon = 1$  のときは Newton-Raphson 法です。

最尤法の場合の具体的な更新式を求めるために  $f(\boldsymbol{\theta})$  の勾配とヘシアンを求めます。式 (8)・(9) は部分モデルでも成り立つことに注意して、 $f(\boldsymbol{\theta})$  の勾配は

$$\begin{aligned} \nabla f(\boldsymbol{\theta}) &= -n \{\mathbf{k} - \nabla \psi(\boldsymbol{\theta})\} \\ &= -n(\mathbf{k} - \boldsymbol{\eta}). \end{aligned} \quad (19)$$

ここで式 (8) ( $\nabla \psi(\boldsymbol{\theta}) = \boldsymbol{\eta}$ ) を用いました。次に  $f(\boldsymbol{\theta})$  のヘシアンは

$$\begin{aligned} \nabla \nabla f(\boldsymbol{\theta}) &= n \nabla \nabla \psi(\boldsymbol{\theta}) \\ &\equiv n\mathbf{G}. \end{aligned} \quad (20)$$

ここで式 (9) の要素を持つフィッシャー情報行列  $\mathbf{G}$  を定義しました。

従って最尤法の場合の更新式は最急降下法・自然勾配法それぞれ

$$\boldsymbol{\theta}^{\text{new}} = \boldsymbol{\theta}^{\text{old}} + \epsilon n (\mathbf{k} - \boldsymbol{\eta}^{\text{old}}). \quad (21)$$

$$\boldsymbol{\theta}^{\text{new}} = \boldsymbol{\theta}^{\text{old}} + \epsilon \mathbf{G}^{-1} (\mathbf{k} - \boldsymbol{\eta}^{\text{old}}). \quad (22)$$

となります。自然勾配法ではフィッシャー情報行列も  $\boldsymbol{\eta}^{\text{old}}$  を用いて計算します。これらの最適化公式は誤差項  $\mathbf{k} - \boldsymbol{\eta}$  を最小化するように  $\boldsymbol{\theta}$  を更新します。極値である  $\nabla f(\boldsymbol{\theta}) = 0$  の解が  $\boldsymbol{\eta} = \mathbf{k}$  であり、最尤法ではモデルの同時スパイクレート  $\eta_I$  が実測値  $k_I$  と一致するように選ばれます。

さて、どちらの式も  $\boldsymbol{\theta}^{\text{old}}$  に対応する  $\boldsymbol{\eta}^{\text{old}}$  を計算する必要があります。周辺量である  $\boldsymbol{\eta}$  を直接得るためには対数正規化項  $\psi(\boldsymbol{\theta})$  の計算が必要になります。10 程度の少数ニューロンならばすべての確率関数を計算してしまっても十分できます。しかしニューロン数が増えた場合、実はこの変換は容易ではありません。対数線形モデルの難しさはほとんどすべてこの座標変換が容易でないことに起因するといっても過言ではありません。通常はモンテカルロ法を使って正規化項を計算することなく  $\boldsymbol{\eta}$  を求めることが行われています。ここでは  $\boldsymbol{\eta}$  を求めることができるとして先に進みましょう。

最尤法では最急降下法を用いるのが一般的です（その理由は後で説明します）。それではこの方法を使ってデータに対数線形モデルをフィッティングしてみましょう。するとどうでしょう、いくつかの  $\theta_I$  についてその最尤推定量が負の非常に大きな値になってしまいます！ この現象はデータサイズ  $n$  が小さいとき、加えて高次スパイク相関をモデルに組み込んだときに顕

著に現れます。なにが起きたのでしょうか？ 理由は「データにスパイク同期発火がなかったから」です。例えば 2 つのニューロンで考えた場合、この 2 ニューロンが観測期間中に一度も同時に発火しなかったとき  $\theta_{12} \rightarrow -\infty$  となります。つまり最尤法では、2 つのニューロンは非常に強い負の相関関係にあって両者が同時に発火するのを拒んでいるという解釈になります。これは与えられたデータに対しては邪念のない解釈です。しかし私達が扱うスパイクデータの発火率は低く（データが 0 に偏っている）、たとえニューロンが独立に発火していても同期発火確率は非常に小さくなります。従って短いデータ区間では同期発火がたまたま見られなかった可能性もあり、最尤推定量によるモデルは新しいデータを説明するには適さないかもしれません。この問題は良く知られたモデルの過剰適合（オーバーフィッティング）の問題です。この問題を回避するためには、適切な次数の対数線形モデルを選ぶための交差検定や赤池情報量基準（AIC）によるモデル選択、あるいはパラメータの正則化を行う必要があります。ここでは先を急ぎ事前分布を用いたベイズの定理に基づく正則化を考えます。

#### 4.2 MAP 推定

ベイズの定理を用いる推定法ではパラメータ  $\boldsymbol{\theta}$  に分布（事前分布）を想定します。ここでは事前分布として平均  $\mathbf{0}$ 、共分散行列  $\lambda^{-1}\mathbf{I}$  の正規分布を考えましょう。

$$p(\boldsymbol{\theta}|\lambda) = \left(\frac{\lambda}{2\pi}\right)^{d/2} \exp\left(-\frac{1}{2}\lambda\|\boldsymbol{\theta}\|\right). \quad (23)$$

データが与えられた下でのパラメータ  $\boldsymbol{\theta}$  の密度分布（事後確率密度関数）は

$$p(\boldsymbol{\theta}|\mathbf{k}, \lambda) = \frac{p(\mathbf{k}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\lambda)}{p(\mathbf{k}|\lambda)} \quad (24)$$

で与えられます。この事後分布の最頻値（Maximum a posteriori, MAP 推定量）

$$\boldsymbol{\theta}_{\text{MAP}} = \arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{k}, \lambda) \quad (25)$$

を求めたいのですが、これを探すのに事後分布の分母の計算は必要ありません。そこで目的関数を

$$\begin{aligned} f(\boldsymbol{\theta}) &= -\log p(\mathbf{k}|\boldsymbol{\theta}) - \log p(\boldsymbol{\theta}|\lambda) \\ &= -n \{\mathbf{k}'\boldsymbol{\theta} - \psi(\boldsymbol{\theta})\} - \frac{d}{2} \log \frac{\lambda}{2\pi} + \frac{\lambda}{2} \|\boldsymbol{\theta}\|^2 \end{aligned} \quad (26)$$

として  $f(\boldsymbol{\theta})$  を最小化する  $\boldsymbol{\theta}$  を求めます。最小化の手順は最尤法のとおりで、式 (16) もしくは式 (18) に対して上記の  $f(\boldsymbol{\theta})$  を用います。そこで  $f(\boldsymbol{\theta})$  の勾

配及びヘシアンを求めておきます。

$$\nabla f(\boldsymbol{\theta}) = -n(\mathbf{k} - \boldsymbol{\eta}) + \lambda \boldsymbol{\theta}, \quad (27)$$

$$\nabla \nabla f(\boldsymbol{\theta}) = n\mathbf{G} + \lambda \mathbf{I}. \quad (28)$$

最尤法における式(19)・(20)と比べると式(27)・(28)には正則化項が第二項として加えられています。式(27)を見るとこの項によって過剰適合が避けられることがわかります。一方MAP推定量によるモデルの同時スパイクレートは実測値からずれることがわかります。また前節の最尤法では、モデルに含まれていながら観測されない同期スパイクがあるときに自然勾配法(式(18))による更新はうまくいきません。フィッシャー情報量行列の対応する行(列)の値がすべて0に近づき逆行列を求められなくなるからです。事前分布を用いた正則化を行うと式(28)第二項の存在によりこの問題点も回避することができます。

事後分布を用いた推定では分布の広がりを見ることで信用区間(信頼区間のベイズ版)を得ることができます。特に事後分布を正規分布で近似するとその平均はMAP推定量 $\boldsymbol{\theta}_{\text{MAP}}$ 、共分散行列 $\mathbf{W}$ は

$$\begin{aligned} \mathbf{W} &= -\left[\nabla \nabla \log p(\boldsymbol{\theta}|\mathbf{k}, \lambda)|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\text{MAP}}}\right]^{-1} \\ &= \left[\nabla \nabla f(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\text{MAP}}}\right]^{-1} \\ &= [n\mathbf{G} + \lambda \mathbf{I}]^{-1} \end{aligned} \quad (29)$$

で与えられます。この近似はラプラス近似といい後述の動的スパイク相関の推定においても使用します。

尤度と事前分布のどちらを重視するかは超パラメータ $\lambda$ によって決まります。通常 $\lambda$ は周辺尤度のパラメータとして

$$p(\mathbf{k}|\lambda) = \int p(\mathbf{k}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\lambda) d\boldsymbol{\theta} \quad (30)$$

を最大にするものが選ばれます(経験ベイズ法・TypeIIの最尤推定)。周辺尤度はラプラス近似を用いて近似式を得ることができます。次数の異なるモデルの比較も周辺尤度比(ベイズ因子)を用いて行うことができます。上記の事前分布ではひとつの超パラメータを考えましたが、パラメータ毎に異なる複数の超パラメータを考えることもできます。このような方法は関連ベクトルマシンと呼ばれスパース推定の文脈で研究されています。この場合の異なるモデル間の比較では超パラメータ数がパラメータ次元に依存するため、周辺尤度を用いた赤池ベイズ情報量基準(ABIC, 単にAICと呼ぶことも多いようです)や、さらに上位の階層ベイズによる周辺化を考える必要があります。

## 5. 動的スパイク相関の推定

ここまで同時確率関数 $p(\mathbf{x})$ は時間によらず一定であることを仮定してきました。しかし行動中の動物から記録されたスパイク時系列データの発火頻度や同期発火頻度は変動します。本節ではこうした実データに対応できるよう状態空間モデルの枠組みを用いて対数線形モデルを時間依存の形に拡張します<sup>7)</sup>。準備が長くなりましたがここまでくれば動的スパイク相関の推定に必要な手段は揃っています。

### 5.1 動的スパイク相関の状態空間モデル

状態空間モデルでは観測されない状態変数を考え、隠れ状態の時間的遷移を仮定します。ここでは対数線形モデルの相互作用パラメータ $\boldsymbol{\theta}$ を隠れ状態とし、 $\boldsymbol{\theta}$ が時刻(ビン) $t (= 1, \dots, T)$ に依存するとします。時刻 $t$ における相互作用パラメータを $\boldsymbol{\theta}_t = [\theta_1^t, \dots, \theta_{12}^t, \dots, \theta_{1\dots r}^t, \dots]^T$ とし、変動の様子を記述する状態方程式を仮定します：

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\xi}_t, \quad (31)$$

( $t = 2, \dots, T$ )。ここで $\boldsymbol{\xi}_t$ は平均 $\mathbf{0}$ ・共分散 $\mathbf{Q}$ の多変量正規分布に従う乱数とします。初期値は $\boldsymbol{\theta}_1 \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ とします。これは観測期間にわたる隠れ状態 $\boldsymbol{\theta}_{1:T} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_T\}$ の同時分布として、事前分布

$$p(\boldsymbol{\theta}_{1:T}) = p(\boldsymbol{\theta}_1|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \prod_{t=2}^T p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) \quad (32)$$

を想定していることとなります。ここで $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})$ は状態遷移を表す確率密度で式(31)から正規分布 $\mathcal{N}(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{Q})$ となります。共分散行列 $\mathbf{Q}$ は相互作用パラメータが時間的にどの程度速く変動するかを決める平滑化パラメータの役割を果たします。事前分布に含まれる超パラメータをまとめて $\mathbf{w} = [\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{Q}]$ とします。

状態空間モデルではさらに時刻(ビン)毎の条件付独立を仮定します。すなわち時刻 $t$ の $\boldsymbol{\theta}_t$ が決まったらこの同時確率関数に従ってデータが生成されるとします。従って時刻 $t$ の条件付対数線形モデルは式(10)と同じように

$$p(\mathbf{x}|\boldsymbol{\theta}_t) = \exp\left[\sum_{I \in \Omega_{1,r}} \theta_I^t F_I(\mathbf{x}) - \psi(\boldsymbol{\theta}_t)\right] \quad (33)$$

となります。このときモデルの同時発火レートは

$$\eta_I^t = E[F_I(\mathbf{x})|\boldsymbol{\theta}_t], \quad (34)$$

( $I \in \{\Omega_1, \dots, \Omega_r\}$ )。これをベクトル $\boldsymbol{\eta}_t$ で表します、 $\boldsymbol{\eta}_t = [\eta_1^t, \dots, \eta_{12}^t, \dots, \eta_{1\dots r}^t, \dots]^T$ 。

電気生理実験では同じ試行（トライアル）を繰り返して行い、その加算平均を取ることがしばしば行われます。そこでマルチニューロンの同時記録スパイク時系列が  $n$  回の同一試行により得られたとします。このスパイク時系列を適当な幅  $\Delta$  [s] のビンに区切りビン中に 1 つ以上のスパイクがあれば '1' とし、なければ '0' とします。ビンの数を  $T$  個とします。この二値データを  $\mathbf{X}^{t,l} = (X_1^{t,l}, X_2^{t,l}, \dots, X_N^{t,l})$  であらわします。要素  $X_i^{t,l}$  は  $i (= 1, \dots, N)$  番目のニューロンの  $l (= 1, \dots, n)$  試行目の  $t (= 1, \dots, T)$  番目のビンのスパイクの有無を表します。ここで定常モデルのときとサンプル数  $n$  の定義が変わっていることに注意してください。定常性を仮定できるのは同一試行を繰り返したときの同一時刻のサンプルであるとするからです。そこで時刻  $t$  の観測同期スパイクレートを

$$k_I^t = \frac{1}{n} \sum_{l=1}^n F_I(\mathbf{X}^{t,l}) \quad (35)$$

( $I \in \{\Omega_1, \dots, \Omega_r\}$ ) とします。これを用いて全観測データ  $\mathbf{k}_{1:T} = \{\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_T\}$  によるパラメータ  $\boldsymbol{\theta}_{1:T} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_T\}$  の尤度関数は

$$\begin{aligned} p(\mathbf{k}_{1:T}|\boldsymbol{\theta}_{1:T}) &= \prod_{t=1}^T \prod_{I \in \Omega_{1:r}} \exp[\sum_{I \in \Omega_{1:r}} \theta_I^t F_I(\mathbf{X}^{t,l}) - \psi(\boldsymbol{\theta}_t)] \\ &= \prod_{t=1}^T \exp[n(\sum_{I \in \Omega_{1:r}} \theta_I^t k_I^t - \psi(\boldsymbol{\theta}_t))] \\ &= \prod_{t=1}^T \exp[n(\mathbf{k}'_t \boldsymbol{\theta}_t - \psi(\boldsymbol{\theta}_t))] \end{aligned} \quad (36)$$

と書くことができます。

本稿での目標はデータが与えられた下で相互作用パラメータの事後分布を求めることです。尤度関数（式 (36)）と事前分布（式 (31)・(32)）を与えたので原理的にはベイズの定理から事後分布は

$$p(\boldsymbol{\theta}_{1:T}|\mathbf{k}_{1:T}, \mathbf{w}) = \frac{p(\mathbf{k}_{1:T}|\boldsymbol{\theta}_{1:T})p(\boldsymbol{\theta}_{1:T}|\mathbf{w})}{p(\mathbf{k}_{1:T}|\mathbf{w})} \quad (37)$$

で与えられます。事後分布がわかればその MAP 推定量を得ることでデータに照らして最も適切なスパイク相関のダイナミクスの経路がわかります。また分布を求めることで推定の確からしさを表す信用区間を得ることもできます。ただし事後分布は超パラメータ  $\mathbf{w}$  の選択に依存します。超パラメータは定常モデルのときと同じように周辺尤度

$$p(\mathbf{k}_{1:T}|\mathbf{w}) = \int p(\mathbf{k}_{1:T}, \boldsymbol{\theta}_{1:T}|\mathbf{w}) d\boldsymbol{\theta}_{1:T} \quad (38)$$

を最大化するものを選びます。長くなるので割愛しますがこの式もまたラプラス近似により近似式を求めることができます。また異なる次数の相関を組み込んだモデルの適合度の比較にも周辺尤度を用いることができます。ただし超パラメータがパラメータ数に依存する場合、超パラメータの次元のペナルティーを入れた赤池ベイズ情報量基準等でモデル選択をすることになります。

実際の運用では式 (37) のような大掛かりな事後分布の構築と超パラメータ最適化を同時に行うことはせず、EM アルゴリズムを用いて交互に行うことが普通です<sup>16-18)</sup>。この方法では超パラメータ  $\mathbf{w}$  を決めて事後分布を構築し (E-step)、この事後分布を用いて  $\mathbf{w}$  を最適化する (M-step) ことを交互に行います。E-step における事後分布の構築は時間前方へ再帰的に事後分布を構築するフィルタリングを行い、その後時間後方に向けて事後分布を構築する平滑化 (スムージング) を行うことで完了します<sup>19)</sup>。本稿では最も重要なフィルタリングとスムージングについて説明します。

## 5.2 フィルタ分布・平滑化分布

まず初めに一期先予測分布  $p(\boldsymbol{\theta}_t|\mathbf{k}_{1:t-1})$  を計算します。これは時刻  $t-1$  までの観測が与えられたもとで事前分布に基づいて時刻  $t$  の状態を予測する分布です。一期先予測分布は Chapman-Kolmogorov の式として次のように書くことができます。

$$p(\boldsymbol{\theta}_t|\mathbf{k}_{1:t-1}) = \int p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})p(\boldsymbol{\theta}_{t-1}|\mathbf{k}_{1:t-1})d\boldsymbol{\theta}_{t-1}. \quad (39)$$

ここで右辺の遷移確率  $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})$  は事前分布である状態モデル（式 (31)）で与えられる平均  $\boldsymbol{\theta}_{t-1}$ ・共分散  $\mathbf{Q}$  の正規分布です。また右辺 2 つめの密度関数は時刻  $t-1$  におけるフィルタ密度関数  $p(\boldsymbol{\theta}_{t-1}|\mathbf{k}_{1:t-1})$  です。ここで時刻  $t-1$  におけるフィルタ分布を仮に平均  $\boldsymbol{\theta}_{t-1|t-1}$ ・共分散  $\mathbf{W}_{t-1|t-1}$  の正規分布とすると一期先予測分布も正規分布となり、その平均  $\boldsymbol{\theta}_{t|t-1}$  及び共分散行列  $\mathbf{W}_{t|t-1}$  は次式で与えられます。

$$\boldsymbol{\theta}_{t|t-1} = \boldsymbol{\theta}_{t-1|t-1}, \quad (40)$$

$$\mathbf{W}_{t|t-1} = \mathbf{W}_{t-1|t-1} + \mathbf{Q}. \quad (41)$$

次にフィルタ分布  $p(\boldsymbol{\theta}_t|\mathbf{k}_{1:t})$  を求めます。フィルタ分布は時刻  $t$  までのスパイク時系列を観測した下での、時刻  $t$  における  $\boldsymbol{\theta}_t$  の事後分布です。フィルタ分布は



ベイズの定理を使って尤度と一期先予測分布を用いて次のように書くことができます。

$$\begin{aligned}
 p(\boldsymbol{\theta}_t | \mathbf{k}_{1:t}) &= \frac{p(\mathbf{k}_t | \boldsymbol{\theta}_t) p(\boldsymbol{\theta}_t | \mathbf{k}_{1:t-1})}{p(\mathbf{k}_t | \mathbf{k}_{1:t-1})} \\
 &\propto \exp \left[ n(\mathbf{k}'_t \boldsymbol{\theta}_t - \psi(\boldsymbol{\theta}_t)) \right. \\
 &\quad \left. - \frac{1}{2} (\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t|t-1})' \mathbf{W}_{t|t-1}^{-1} (\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t|t-1}) \right].
 \end{aligned} \tag{42}$$

ここで MAP 推定量を求めるため定常モデルのときと同じように目的関数を

$$f(\boldsymbol{\theta}) = -\log p(\mathbf{k}_t | \boldsymbol{\theta}) - \log p(\boldsymbol{\theta} | \mathbf{k}_{1:t-1}) \tag{43}$$

とおきます。目的関数の勾配およびヘシアンは

$$\nabla f(\boldsymbol{\theta}) = -n(\mathbf{k}_t - \boldsymbol{\eta}) + \mathbf{W}_{t|t-1}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_{t|t-1}) \tag{44}$$

$$\nabla \nabla f(\boldsymbol{\theta}) = n\mathbf{G} + \mathbf{W}_{t|t-1}^{-1} \tag{45}$$

で得られます。式 (16) もしくは (18) に対して上述の導関数を適用し MAP 推定量を得ることができます。

さらにここでは事後分布を正規分布によって近似します (ラプラス近似)。これには MAP 推定量をフィルタ分布の期待値  $\boldsymbol{\theta}_{t|t}$  とし、共分散行列を  $\mathbf{W}_{t|t} = [\nabla \nabla f(\boldsymbol{\theta}_{t|t})]^{-1}$  とします。フィルタ分布を正規分布で近似することで一期先予測分布の導出で行った正規分布の仮定と矛盾が生じません。このようにして一期先予測の式 (40)・(41) とあわせて  $t = 1, \dots, T$  について再帰的にフィルタ分布の近似分布を求めて行きます。

上述のアルゴリズムは次のような非線形再帰フィルタ公式を解いていることになります<sup>7)</sup>。

$$\boldsymbol{\theta}_{t|t} = \boldsymbol{\theta}_{t|t-1} + n\mathbf{W}_{t|t-1}(\mathbf{k}_t - \boldsymbol{\eta}_{t|t}), \tag{46}$$

$$\mathbf{W}_{t|t}^{-1} = \mathbf{W}_{t|t-1}^{-1} + n\mathbf{G}_{t|t}. \tag{47}$$

ここで式 (46) は  $\nabla f(\boldsymbol{\theta}) = \mathbf{0}$  から得ました。式 (46) で  $\boldsymbol{\theta}_{t|t}$  及び  $\boldsymbol{\eta}_{t|t}$  は同じ分布の異なる座標系による表記ですからこの式は非線形方程式であり、上述のように解かなければなりません。しかしながら式 (46)・(47) による表記はこの非線形近似フィルタがなにを行っているかを端的に表しています (図 3 参照)。まず式 (46) の第一項は時刻  $t$  のデータが与えられる前の事前分布に基づいた期待値の一期先の予測です。これに対し第二項は時刻  $t$  の観測が行われたことによる修正を表します。同様に式 (47) の第一項は事前分布によって得られる一期先予測の推定の正確さを表します。これに対

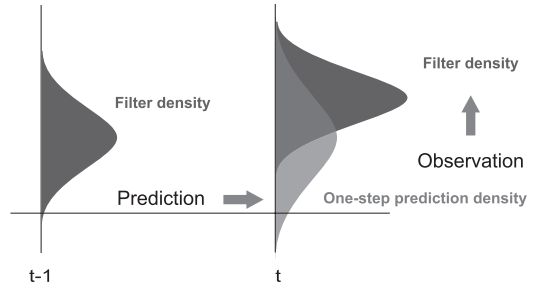


図3 フィルタ分布構築の模式図。

して第二項のフィッシャー情報行列は時刻  $t$  の観測によって得られる推定精度の向上を示します。以上を模式的に示すと図 3 のようになります。一般に時刻  $t$  の  $\boldsymbol{\theta}$  の一期先予測分布の裾野は時刻  $t-1$  におけるフィルタ分布に比べて広がります (式 (41) 参照)。時刻  $t$  においてデータを観測することにより一期先予測による分布の期待値は修正され (式 (46)), 新たな事後分布は分散も小さくなります (式 (47))。

最後に平滑化事後分布<sup>16, 19)</sup>

$$\begin{aligned}
 p(\boldsymbol{\theta}_t | \mathbf{k}_{1:T}) &= p(\boldsymbol{\theta}_t | \mathbf{k}_{1:t}) \\
 &\int \frac{p(\boldsymbol{\theta}_{t+1} | \mathbf{k}_{1:T}) p(\boldsymbol{\theta}_{t+1} | \boldsymbol{\theta}_t)}{p(\boldsymbol{\theta}_{t+1} | \mathbf{k}_{1:t})} d\boldsymbol{\theta}_{t+1}
 \end{aligned} \tag{48}$$

の構築を後方への再帰的アルゴリズムを用いて行います。これにはフィルタリングにより得られた  $\boldsymbol{\theta}_{T|T}$  及び  $\mathbf{W}_{T|T}$  から始めて平滑化分布の平均と分散を

$$\boldsymbol{\theta}_{t|T} = \boldsymbol{\theta}_{t|t} + \mathbf{A}_t (\boldsymbol{\theta}_{t+1|T} - \boldsymbol{\theta}_{t+1|t}), \tag{49}$$

$$\mathbf{W}_{t|T} = \mathbf{W}_{t|t} + \mathbf{A}_t (\mathbf{W}_{t+1|T} - \mathbf{W}_{t+1|t}) \mathbf{A}'_t \tag{50}$$

(ただし  $\mathbf{A}_t = \mathbf{W}_{t|t} \mathbf{F}' \mathbf{W}_{t+1|t}^{-1}$ ) に従って  $t = T-1, T-2, \dots, 1$  の順に求めて行きます。これにより全観測データを用いて時刻  $t$  のスパイク相関状態を推定することができます。

## 6. おわりに

本稿は対数線形モデルを用いたマルチニューロンスパイクデータ解析を紹介しました。このモデルによって明らかにできること、特に高次スパイク相関について紹介しました。次に定常モデルのパラメータの具体的な推定法 (最尤推定・MAP 推定) を経て、動的なスパイク相関構造の推定を可能にする状態空間モデルによる定式化までを順を追って紹介しました。それで

は高次のスパイク相関は本当に動物の行動とともに動的に生成されるのでしょうか。それだけでなく動物の高次の認知機能に積極的に関与しているのでしょうか。その答えは... 近々日の目をみるであろう私達の論文にぜひ目を通してみて下さい。

**謝辞** 対数線形モデルを教えて頂いた甘利俊一先生・中原裕之先生に感謝いたします。また豊泉太郎氏との議論により理解が深まりました。第2章で用いたパンチカードの例のヒントを与えてくれたElad Schneidman氏にも感謝致します。

### 参 考 文 献

- 1) Schneidman, E., Berry, M.J., Segev, R., and Bialek, W. (2006): Weak pairwise correlations imply strongly correlated network states in a neural population, *Nature*, Vol.440, No.7087, pp.1007–1012
- 2) Santos, G.S., Gireesh, E.D., Plenz, D., and Nakahara, H. (2010): Hierarchical interaction structure of neural activities in cortical slice cultures, *J. Neurosci.*, Vol.30, No.26, p.8720
- 3) Riehle, A., Grün, S., Diesmann, M., and Aertsen, A. (1997): Spike synchronization and rate modulation differentially involved in motor cortical function, *Science*, Vol.278, No.5345, pp.1950–1953
- 4) Sakurai, Y. and Takahashi, S. (2006): Dynamic synchrony of firing in the monkey prefrontal cortex during working-memory tasks, *J. Neurosci.*, Vol.26, No.40, pp.10141–10153
- 5) Fujisawa, S., Amarasingham, A., Harrison, M.T., and Buzsáki, G. (2008): Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex, *Nat. Neurosci.*, Vol.11, No.7, pp.823–833
- 6) Ito, H., Maldonado, P.E., and Gray, C.M. (2010): Dynamics of stimulus-evoked spike timing correlations in the cat lateral geniculate nucleus, *J. Neurophysiol.*, Vol.104, No.6, pp.3276–3292
- 7) Shimazaki, H., Amari, S.-I., Brown, E.N., and Grün, S. (2009): State-space analysis on time-varying correlations in parallel spike sequences, In Proc. IEEE ICASSP2009, pp.3501–3504
- 8) Amari, S.-I. and Nagaoka, H. (2000): *Methods of Information Geometry* (Translations of Mathematical Monographs). American Mathematical Society
- 9) Amari, S.-I. (2001) Information geometry on hierarchy of probability distributions, *Information Theory, IEEE Transactions on*, Vol.47, No.5, pp.1701–1711
- 10) Nakahara, H. and Amari, S.-I. (2002): Information-geometric measure for neural spikes, *Neural Comput.*, Vol.14, No.10, pp.2269–2316
- 11) 赤穂昭太郎 (2005)：情報幾何と機械学習，計測と制御，Vol.44, No.5, pp.299–306
- 12) 岡田真人 (2009)：脳の計算論（シリーズ脳科学1），chapter スパイクの確率論，pp.223–251. 東京大学出版会
- 13) 中原裕之 (2009)：脳の計算論（シリーズ脳科学1），chapter 意思決定とその学習理論，pp.159–221. 東京大学出版会
- 14) Truccolo, W., Eden, U.T., Fellows, M.R., Donoghue, J.P., and Brown, E.N. (2005): A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects, *J. Neurophysiol.*, Vol.93, No.2, pp.1074–1089
- 15) Amari, S.-I. (1998): Natural gradient works efficiently in learning, *Neural Comput.*, Vol.10, pp.251–276
- 16) Kitagawa, G. and Gersch, W. (1996): *Smoothness Priors Analysis of Time Series* (Lecture Notes in Statistics). Springer
- 17) Brown, E.N., Frank, L.M., Tang, D., Quirk, M.C., and Wilson, M.A. (1998): A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells, *J. Neurosci.*, Vol.18, No.18, pp.7411–7425
- 18) Smith, A.C. and Brown, E.N. (2003): Estimating a state-space model from point process observations, *Neural Comput.*, Vol.15, No.5, pp.965–991
- 19) Koyama, S. and Paninski, L. (2010): Efficient computation of the maximum a posteriori path and parameter estimation in integrate-and-fire and more general state-space models, *J. Comput. Neurosci.*, Vol.89, No.105, pp.1–2