

る。このような連続性の拘束条件を、独立性を仮定した式 (6.29) に導入できれば、図 6.2 のような過渡状態でも発火率を推定できる可能性がある。本章の後半では、過渡状態での発火率を推定する枠組みを議論する。

6.2 時間変化する発火率のベイズ推定

この節では各時間ビンごとに独立にスパイクが生成されると仮定して、時間的に変化する発火率を推定する方法を考えてみよう。

6.2.1 ヒストグラム法を用いた発火率の推定

まず定常性を仮定して、すべての時間ビンが同一の発火率を持つとする。この場合は、すべての時間ビンで生成されたスパイクを数え上げて、時間ビン数 T でわった発火率を求めればよい。発火率が一定の場合は、あえて時間ビンを導入する必要はなく、スパイクを測定しているすべての時間で一定と仮定した発火率を求めるわけである。

この考えを発火率が時間変化する場合に拡張する。これまでは時間ビンにただか1つのスパイクしか入らないようにするために、時間ビンを 1 ms とした。ここでは図 6.2 の過渡状態のように、発火率は滑らかに連続的に変化すると仮定する。 1 ms に対して比較的広めの時間ビン、たとえば数十から 100 ms 程度のビン幅を導入し、その時間ビンでは定常性を仮定し発火率を一定と見なすことにしよう。このようにして求めたビン内の発火率の平均を、そのビンの発火率とするのである。この手法はヒストグラム法とよばれ、実際の電気生理学のスパイクのデータ解析に広く使われている。通常は1回の測定ではなく、複数回の測定結果を足し合わせて発火率を推定する。どれぐらいの大きさのビン幅を用いるかに関しては、理論的な考察も行われているが (Shimazaki & Sinomori, 2007)、現状では多くの場合は天下りの決められている。

この手法はビン内の定常性を仮定しているが、隣り合ったビン間がよく似た発火率を持つなどの条件を考慮していない。ビン内で定常性が仮定できるほど、発火率が連続的に変化しているのであれば、隣り合ったビン間の発火率が似ているという条件を取り入れる方が望ましい。そこで次にベイズ推定に基づき、時間ビン間に滑らかさを導入して時間変動する発火率を推定する手法を考える。

6.2.2 画像修復の正則化理論のヒストグラム法への応用

まず前の 6.2.1 項のヒストグラム法の要領で、与えられた時間幅でビンを区切り、各ビン内での平均発火率 λ_t を計算する。 $t = 1, \dots, T$ を時間ビンを指定する番号とする。ここで T は時間ビン数である。6.1 節ではスパイクが時間ビン 1 つにつき、たかだか 1 個しか存在しないように決めたが、ここでは多くのスパイクが 1 つのビンに存在するように時間幅をさだめる。そのようにして求めた λ_t をそのまま発火率の推定値として用いるのがヒストグラム法であった。ここでは時間ビン t での発火率 η_t を以下の方法で求める。すべての時間ビンでの発火率 $\{\eta_t\}$ のエネルギー関数 $E(\{\eta_t\}; \epsilon)$ を以下のように定義する、

$$E(\{\eta_t\}; \epsilon) = \frac{1}{2} \sum_{t=1}^T (\eta_t - \lambda_t)^2 + \frac{\epsilon}{2} \sum_{t=1}^{T-1} (\eta_{t+1} - \eta_t)^2, \quad \epsilon > 0. \quad (6.34)$$

ϵ と $\{\lambda_t\}$ が与えられたときに、 $E(\{\eta_t\}; \epsilon)$ を最小にする $\{\eta_t\}$ を発火率の推定値とする方法を考える。エネルギー関数 $E(\{\eta_t\}; \epsilon)$ は $\{\eta_t\}$ の 2 次関数である。式 (6.34) の第 1 項はデータフィット項である。もし $\epsilon = 0$ であれば、 $\eta_t = \lambda_t$ がエネルギー関数 $E(\{\eta_t\}; 0)$ の最小値を与える。このとき、 $\eta_t = \lambda_t$ が推定値となり、ヒストグラム法と同じ結果になる。第 1 項は推定値 η_t を観測データ λ_t にフィットさせる効果を持つ。第 2 項は滑らかさの拘束条件を表す。隣り合った η_{t+1} と η_t が近い値を持つ方が式 (6.34) の第 2 項は小さくなる。 $\epsilon > 0$ の場合、 η_t は観測で得られるヒストグラム法の結果である λ_t にほどほどに近く、しかも隣り合ったビンの η_{t+1} と η_{t-1} にほどほどに近い値をとる。これは先ほど述べた、時間ビン間に滑らかさが導入されたことに対応している。このような手法は正則化と呼ばれており、ノイズが存在するデータにある種の拘束条件をかけて、データ修復をする代表的な手法の 1 つである。式 (6.34) の 2 番目の項は、発火率がビン間で滑らかに変化するという拘束条件を表している。実はここで議論している発火率推定はデータ修復の典型例である画像修復 (田中, 2006) の 1 次元版とみなすことができる。発火率推定の時間ビンが画像のピクセルに対応している。通常の画像はピクセルが 2 次元的に並んでいるが、ここでの発火率推定ではピクセルが 1 次元的に並んでいる。 λ_k がノイズ下で観測した画素値の観測データに対応している。画像修復においては、式 (6.34) の 2 番目の項

は、隣り合ったピクセルの画素値は近いという拘束条件を表す。

式 (6.34) の $E(\{\eta_t\}; \epsilon)$ を最小化する手法は数多く存在する。1 つの方法は、逆行列を用いる方法である。 $E(\{\eta_t\}; \epsilon)$ は 2 次関数なので、1 次微分が 0 になる条件は連立方程式になる。この連立方程式を解くことは、対応する行列の逆行列を求めることに対応し、その計算量は T^2 のオーダーである。後の 6.2.4 項で紹介するように、このエネルギー関数 $E(\{\eta_t\}; \epsilon)$ の時刻に関して周期的境界条件、 $\eta_{T+t} = \eta_t$, $\lambda_{T+t} = \lambda_t$ を導入すると、フーリエ変換を用いて行列を対角化でき、高速に最小値を与える $\{\eta_t\}$ を求めることができる。これらはいずれも非局所的な計算である。一方、式 (6.34) の $E(\{\eta_t\}; \epsilon)$ を最小にする $\{\eta_t\}$ を求める代表的な局所的な手法の 1 つは最急降下法である。最急降下法では $E(\{\eta_t\}; \epsilon)$ の η_t に関する偏微分を求め、 $E(\{\eta_t\}; \epsilon)$ が小さくなるように η_t を反復的に更新する。 ϵ は更新幅を制御するパラメータである。最急降下法では、まず $\{\eta_t\}$ に対する初期値 $\{\eta_t^{(0)}\}$ を適当に決める。次にエネルギーの偏微分を用いて、エネルギーが小さくなるように以下の式を使って $\{\eta_t^{(n)}\}$ の値が収束するまで更新する、

$$\eta_t^{(n+1)} = \eta_t^{(n)} - \kappa \frac{\partial E(\{\eta_t\}; \epsilon)}{\partial \eta_t} \Big|_{\eta_t = \eta_t^{(n)}} \quad (6.35)$$

$$\frac{\partial E(\{\eta_t\}; \epsilon)}{\partial \eta_t} \Big|_{\eta_t = \eta_t^{(n)}} = (\eta_t^{(n)} - \lambda_t) + \epsilon(-\eta_{t+1}^{(n)} + 2\eta_t^{(n)} - \eta_{t-1}^{(n)}). \quad (6.36)$$

ここで κ は十分小さな正の定数である。式 (6.36) の右辺の第 1 項は $\eta_t^{(n)} = \lambda_t$ のときに 0 となり、第 2 項は $\eta_t^{(n)}$ が隣接する $\eta_{t+1}^{(n)}$ と $\eta_{t-1}^{(n)}$ の平均値に等しいときに 0 になる。このように式 (6.35) と (6.36) から、 $E(\{\eta_t\}; \epsilon)$ を最小化することにより、得られる推定値 η_t は観測で得られる λ_t にほどほどに近く、しかも隣り合ったピクセルの η_{t+1} と η_{t-1} にほどほどに近い値をとることがわかる。最急降下法は η_t の更新の際に両隣の η_{t+1} と η_{t-1} の値しか用いない局所的な計算である。最急降下法では 1 次微分のみを用いるので、計算結果が局所的な極小値に収束する可能性がある。しかし、ここで議論しているエネルギーは $\{\eta_t\}$ の 2 次関数なので、局所的な極小値は存在せず、任意の初期値から計算をして最小値に収束する。

実は正則化のもっとも大きな問題点は、これら解を求める方法を議論することではなく、 ϵ の値をどのように決めるかである。これまでの議論からわかる

ように、どのような推定値が得られるかは ϵ に強く依存する。 ϵ はパラメータである推定値の $\{\eta_t\}$ を決めるパラメータなのでハイパーパラメータと呼ばれる。ハイパーパラメータを決める方法の1つとして、次にベイズ推定に基づく手法を説明しよう。

6.2.3 ベイズ推定の枠組みの適用

ここでは式(6.34)のエネルギ関数 $E(\{\eta_t\}; \epsilon)$ をベイズ推定の枠組みで導出する。真の発火率 $\{\eta_t\}$ が存在すると考え、以下のような方法で真の発火率 η_t を推定する枠組みがベイズ推定である。以下の式(6.37)の最初の式に示すように、発火率の観測値 λ_t は真の発火率 η_t に平均0分散 γ^2 のガウスノイズ n_t が重畳されて観測されると仮定する。以下の式(6.37)の2番目の式のように、このデータ生成過程は真の値 η_t を条件とする観測値 λ_t の条件付き確率 $P(\lambda_t|\eta_t)$ で書ける、

$$\lambda_t = \eta_t + n_t, \quad P(\lambda_t|\eta_t) = \frac{1}{\sqrt{2\pi\gamma^2}} \exp\left(-\frac{1}{2\gamma^2}(\lambda_t - \eta_t)^2\right). \quad (6.37)$$

ここでノイズ n_t はピンごとに独立であると仮定すると、発火率の観測値全体 $\{\lambda_t\}$ は真の発火率を $\{\eta_t\}$ を条件とした条件付き確率は、式(6.37)の2番目の式を t に関してかけたものになる、

$$P(\{\lambda_t\}|\{\eta_t\}) = \prod_{t=1}^T P(\lambda_t|\eta_t) = \frac{1}{(\sqrt{2\pi\gamma^2})^T} \exp\left(-\frac{1}{2\gamma^2} \sum_{t=1}^T (\lambda_t - \eta_t)^2\right). \quad (6.38)$$

これを生成モデルと呼ぶことにする。この式が式(6.34)のエネルギ関数 $E(\{\eta_t\}; \epsilon)$ の第1項に似ていることに注意しよう。

次に真の発火率 $\{\eta_t\}$ は以下の式(6.39)のガウス分布に従う確率 $P(\{\eta_t\})$ で生成されるとする、

$$P(\{\eta_t\}) = \frac{1}{Z_0(\beta)} \exp\left(-\beta \sum_{t=1}^{T-1} (\eta_{t+1} - \eta_t)^2\right),$$

$$Z_0(\beta) = \int \left(\prod_{\tau=1}^T d\eta_\tau\right) \exp\left(-\beta \sum_{t=1}^{T-1} (\eta_{t+1} - \eta_t)^2\right) \quad (6.39)$$

ここで $Z_0(\beta)$ は規格化定数であり、式 (6.21) で導入された分配関数に相当する。この $P(\{\eta_k\})$ を事前確率と呼ぶ。式 (6.39) は、隣り合った真の発火率 $\{\eta_{t+1}\}$ と $\{\eta_t\}$ は同じ値をとる確率が一番大きいことを意味する。この式は、滑らかさを表すエネルギー関数 $E(\{\eta_t\}; \epsilon)$ の第 2 項に対応している。

事前確率を導入したことで問題の定式化が、さきほどの正則化とは質的に変化する。6.2.2 項の正則化では真の発火率 $\{\eta_t\}$ は何らかの方法で与えられており、確率的に記述されるのは発火率の観測値 $\{\lambda_t\}$ のみであった。ここでは真の発火率 $\{\eta_t\}$ を事前確率 $P(\{\eta_t\})$ に従う確率変数と考えている。このように考えると真の値と観測値の同時確率 $P(\{\eta_t\}, \{\lambda_t\})$ を考えることができる。この同時確率は式 (6.10) で述べたベイズの定理から、式 (6.38) の生成モデルと式 (6.39) の事前確率の積で表現することができる。一方、形式的に同時確率 $P(\{\eta_t\}, \{\lambda_t\})$ は、観測値の事前確率 $P(\{\lambda_t\})$ と観測値 $\{\lambda_t\}$ を与えたときの真の値の条件付き確率 $P(\{\eta_t\}|\{\lambda_t\})$ の積で表現できる。

$$P(\{\eta_t\}, \{\lambda_t\}) = P(\{\lambda_t\}|\{\eta_t\})P(\{\eta_t\}) \quad (6.40)$$

$$= P(\{\eta_t\}|\{\lambda_t\})P(\{\lambda_t\}). \quad (6.41)$$

式 (6.41) から観測値 $\{\lambda_t\}$ を与えたときの真の値の条件付き確率 $P(\{\eta_t\}|\{\lambda_t\})$ を求めると、

$$P(\{\eta_t\}|\{\lambda_t\}) = \frac{P(\{\lambda_t\}|\{\eta_t\})P(\{\eta_t\})}{P(\{\lambda_t\})}, \quad (6.42)$$

となる。これは数学的には単なる式変形に過ぎないが、因果律の観点では質的な違いを表現する式になっている。式 (6.40) は因果律を表している。まず観測値の $\{\lambda_t\}$ の原因である $\{\eta_t\}$ が事前確率 $P(\{\eta_t\})$ で生成される。その後我々が実際に $\{\eta_t\}$ を観測しようとするときに、観測過程の不完全さによりノイズが生じて $\{\lambda_t\}$ を観測する。式 (6.40) は実際に物事が起こる順番である因果律をそのまま表している。科学では、我々は観測値から原因を推定する。観測値 $\{\lambda_t\}$ が与えられたときに、因果律をさかのぼり、原因である $\{\eta_t\}$ を推測するわけである。そのような観点で式 (6.42) の条件付き確率 $P(\{\eta_t\}|\{\lambda_t\})$ は、事が起こった後の確率という意味で事後確率 (posterior probability) と呼ばれている。ベイズ推定とは一言でいうと、この事後確率を使って未知の変数を推定する数学的枠組みである。観測データ $\{\lambda_t\}$ の生成過程 $P(\{\lambda_t\}|\{\eta_t\})$ にはノイズが存

在するので、ノイズを取り去るには何らかの別の条件が必要である。その別の条件を表すのが、真の値 $\{\eta_t\}$ の事前確率 $P(\{\eta_t\})$ であり、ここでは具体的に隣り合ったビンの発火率は似通っているという事前知識を表している。

ここでは事後確率を用いた代表的な2つの推定法を紹介する。1つは、最大事後確率 (Maximum a posteriori: MAP) 推定である。MAP 推定では観測値 $\{\lambda_t\}$ が与えられた条件下で、式 (6.42) の事後確率 $P(\{\eta_t\}|\{\lambda_t\})$ を最大にする $\{\eta_k\}$ を推定値とする。もう1つは事後確率に関する、真の値 η_t の平均値、

$$\langle \eta_t \rangle \equiv \int \left(\prod_{\tau=1}^T d\eta_\tau \right) \eta_t P(\{\eta_t\}|\{\lambda_t\}), \quad (6.43)$$

を求め、その平均値 $\langle \eta_t \rangle$ を推定値とする事後平均 (Posterior Mean: PM) 推定である。MAP 推定と PM 推定は一般には異なった結果を与えるが、次の6.2.4項で示すように、この節で議論しているようなガウス分布に従うモデルでは同じ結果を与える。

以下に示すように、前の6.2.2項で述べた正則化に関する式 (6.34) のエネルギー関数最小化と MAP 推定は等価である。事後確率の対数 $\log P(\{\eta_t\}|\{\lambda_t\})$ は以下のように式 (6.34) のエネルギー関数を用いて書くことができる、

$$\log P(\{\eta_t\}|\{\lambda_t\}) = -E(\{\eta_k\}; \frac{\beta}{\gamma^2}) - \frac{N}{2} \log(2\pi\gamma^2) - \log Z_0(\beta) - \log P(\{\lambda_t\}). \quad (6.44)$$

式 (6.44) の右辺で推定値 $\{\eta_t\}$ に依存するのは、第1項の $E(\{\eta_k\}; \beta/\gamma^2)$ だけである。対数は単調増加関数なので、事後確率最大化とエネルギー関数の最小化は同じ推定値 $\{\eta_t\}$ を与える。式 (6.34) のエネルギー関数と式 (6.44) を比較すると、

$$\epsilon = \beta\gamma^2, \quad (6.45)$$

の関係があることがわかる。正則化の枠組みで天下りの与えていた滑らかさの拘束条件の強さを表す ϵ は、ベイズ推定の枠組みでは生成モデルと事前確率の分散の比として解釈できる。式 (6.39) の β が大きいほど、隣り合った2つの発火率の差は小さくなる。生成モデルのノイズの分散 γ が大きい場合、 ϵ は大きくなり、第1項のデータフィット項よりも滑らかさを優先する。これはデータ生成過程のノイズが大きいので、観測値そのものを信用しないことに対応する。

逆に生成過程のノイズが小さいと、 γ^2 が小さくなり、 ϵ が小さくなるので、観測値を信用するようになる。これら生成過程と事前確率のトレードオフによって、滑らか拘束の ϵ が決まる。天下りの的に与えられた ϵ に因果律に基づく根拠が与えられたわけである。

ベイズ推定はさらに、このような解釈レベルでのご利益以上のものをもたらす。観測データだけからハイパーパラメータである β や γ を決定する枠組みが存在するのである。これをハイパーパラメータ推定という。式 (6.42) の右辺の分母の確率 $P(\{\lambda_t\})$ は、式 (6.42) の右辺の分子である尤度 $P(\{\{\lambda_t\}|\{\eta_t\}\})P(\{\eta_t\})$ を真の値 $\{\eta_t\}$ に関して積分して周辺化したものである、

$$P(\{\lambda_t\}) = \int \left(\prod_{\tau=1}^T d\eta_\tau \right) P(\{\lambda_t\}|\{\eta_t\})P(\{\eta_t\}) \quad (6.46)$$

そのため $P(\{\lambda_t\})$ は周辺尤度とよばれる。周辺尤度 $P(\{\lambda_t\})$ はハイパーパラメータである β や γ に依存する。この依存性を陽に表すために、ここでは周辺尤度を $P(\{\lambda_t\}; \beta, \gamma)$ と書こう。

この周辺尤度を用いてハイパーパラメータ β や γ を決定することを考える。いま我々は観測値は知っているので、手元にある $\{\lambda_t\}$ に関して周辺尤度をもっとも大きくする β と γ の値を求めることができる。この方法は周辺尤度最大化とよばれ、観測値を与えた条件で、周辺化尤度を最大化する値をハイパーパラメータの推定値とする方法である。ベイズ推定の枠組みを導入することで、正則化では天下りの的に与えられたハイパーパラメータを、観測データから推定する手法を自然に導入できたわけである。

以下に示すような条件下で、周辺尤度最大化で求めたハイパーパラメータの値は正しい値に一致する。周辺尤度を最大化するハイパーパラメータは、対数周辺化尤度を最大化するそれに等しいので、対数周辺尤度を考える。事前確率と生成モデルの真のハイパーパラメータが β_0 と γ_0 であったとしよう。この場合、対数周辺尤度の期待値は

$$\langle \log P(\{\lambda_t\}; \beta, \gamma) \rangle \quad (6.47)$$

$$= \int \left(\prod_{\tau=1}^T d\lambda_\tau \right) P(\{\lambda_t\}; \beta_0, \gamma_0) \log P(\{\lambda_t\}; \beta, \gamma) \quad (6.48)$$

$$\begin{aligned}
&= \langle \log P(\{\lambda_t\}; \beta_0, \gamma_0) \rangle \\
&\quad - \int \left(\prod_{\tau=1}^T d\lambda_\tau \right) P(\{\lambda_t\}; \beta_0, \gamma_0) \log \left(\frac{P(\{\lambda_t\}; \beta_0, \gamma_0)}{P(\{\lambda_t\}; \beta, \gamma)} \right) \quad (6.49)
\end{aligned}$$

となる．ここで式 (6.49) の第2項は $P(\{\lambda_t\}; \beta_0, \gamma_0)$ と $P(\{\lambda_t\}; \beta, \gamma)$ の間の Kullback-Leibler (KL) ダイバージェンスとよばれる非負の量である．これら2つの確率が等しいときのみ、KL ダイバージェンスは0となる．したがって対数周辺化尤度 $\langle \log P(\{\lambda_t\}; \beta, \gamma) \rangle$ は

$$\langle \log P(\{\lambda_t\}; \beta, \gamma) \rangle \leq \langle \log P(\{\lambda_t\}; \beta_0, \gamma_0) \rangle, \quad (6.50)$$

のように上限 $\langle \log P(\{\lambda_t\}; \beta_0, \gamma_0) \rangle$ をもち、 $\beta_0 = \beta$ 、 $\gamma_0 = \gamma$ のときに等号が成立する．このことから、周辺尤度最大化で求めたハイパーパラメータは、真の値をとることがわかる．さらに次の6.2.4項で示すように、周辺尤度最大化により推定されたハイパーパラメータは、平均2乗誤差最小の基準を満たす．

6.2.4 離散フーリエ変換を用いた理論解析

この節で議論した確率モデルは並進対称性を持つので、ガウス分布の共分散行列について、離散フーリエ変換で対角化を行うことができる．この結果を用いると、MAP 推定と PM 推定の関係や周辺尤度最大化の推定精度などに関して、より具体的に解析的な議論を行うことができる．

真の発火率 $\{\eta_t\}$ と測定値 $\{\lambda_t\}$ の時刻 t に関して周期的境界条件 $\eta_{T+t} = \eta_t$ 、 $\lambda_{T+t} = \lambda_t$ を仮定し、 η_t と λ_t の離散フーリエ変換 $\tilde{\eta}_\omega$ と $\tilde{\lambda}_\omega$ を以下のように定義する、

$$\tilde{\eta}_\omega = \frac{1}{\sqrt{T}} \sum_{t=1}^T \eta_t \exp(-2\pi i \frac{\omega t}{T}), \quad \tilde{\lambda}_\omega = \frac{1}{\sqrt{T}} \sum_{t=1}^T \lambda_t \exp(-2\pi i \frac{\omega t}{T}) \quad (6.51)$$

この場合の逆フーリエ変換は、

$$\eta_t = \frac{1}{\sqrt{T}} \sum_{\omega=1}^T \tilde{\eta}_\omega \exp(2\pi i \frac{\omega t}{T}), \quad \lambda_t = \frac{1}{\sqrt{T}} \sum_{\omega=1}^T \tilde{\lambda}_\omega \exp(2\pi i \frac{\omega t}{T}), \quad (6.52)$$

となる．ここで i は虚数単位である． η_t と λ_t は実数であるので、それらのフーリエ変換には以下のような関係式が成り立つ、

$$\tilde{\eta}_\omega^\dagger = \tilde{\eta}_{-\omega}, \quad \tilde{\lambda}_\omega^\dagger = \tilde{\lambda}_{-\omega}, \quad (6.53)$$

ここで†は複素共役を表す。

これらのフーリエ表示を用いると、式 (6.38) の生成モデルと式 (6.39) の事前確率は、

$$P(\{\tilde{\lambda}_\omega\}|\{\tilde{\eta}_\omega\}) = \prod_{\omega=1}^T P(\tilde{\lambda}_\omega|\tilde{\eta}_\omega) = \prod_{\omega=1}^T \frac{1}{\sqrt{2\pi\gamma^2}} \exp\left(-\frac{1}{2\gamma^2}|\tilde{\lambda}_\omega - \tilde{\eta}_\omega|^2\right). \quad (6.54)$$

$$\begin{aligned} P(\{\tilde{\eta}_\omega\}) &= \prod_{\omega=1}^T \frac{1}{Z_\omega^0(\beta)} \exp(-\beta G_\omega |\tilde{\eta}_\omega|^2), \\ Z_\omega^0(\beta) &= \prod_{\omega=1}^T \int d\tilde{\eta}_\omega \exp(-\beta G_\omega |\tilde{\eta}_\omega|^2) \\ G_\omega &= 2 \left(1 - \cos \frac{2\pi\omega}{T}\right) \end{aligned} \quad (6.55)$$

と書ける。ここで $|Z|$ は複素数 Z の絶対値を表し、 $|Z|^2 = ZZ^\dagger$ で与えられる。フーリエ表示を用いることにより、生成モデルだけでなく事前確率も ω に関して因数分解した形で表現されていることがわかる。これはフーリエ変換で求められた基底ベクトルが、並進対称性を持つ行列の固有ベクトルになることに由来している。

結果的に事後確率も ω に関して因数分解した形で求めることができる、

$$P(\{\tilde{\eta}_\omega\}|\{\tilde{\lambda}_\omega\}) = \prod_{\omega=1}^T P(\tilde{\eta}_\omega|\tilde{\lambda}_\omega) \quad (6.56)$$

$$P(\tilde{\eta}_\omega|\tilde{\lambda}_\omega) = \frac{1}{Z_\omega(\beta, \gamma)} \exp\left(-F_\omega |\tilde{\eta}_\omega - \frac{1}{2\gamma^2 F_\omega} \tilde{\lambda}_\omega|^2\right), \quad (6.57)$$

$$Z_\omega(\beta, \gamma) = \int d\tilde{\eta}_\omega \exp\left(-F_\omega |\tilde{\eta}_\omega - \frac{1}{2\gamma^2 F_\omega} \tilde{\lambda}_\omega|^2\right), \quad (6.58)$$

$$F_\omega = \beta G_\omega + \frac{1}{2\gamma^2}. \quad (6.59)$$

事後確率 $P(\{\tilde{\eta}_\omega\}|\{\tilde{\lambda}_\omega\})$ に関するこれらの式から、 $\tilde{\eta}_\omega$ に関する PM 推定の推定値 $\langle \tilde{\eta}_\omega \rangle$ は、

$$\langle \tilde{\eta}_\omega \rangle = \frac{1}{2\gamma^2 F_\omega} \tilde{\lambda}_\omega, \quad (6.60)$$

となることがわかる．さらにこの PM 推定値は事後確率 $P(\{\tilde{\eta}_\omega\}|\{\tilde{\lambda}_\omega\})$ を最大化する．この例から，確率がガウス分布のような対称性をもたない場合は，MAP 推定と PM 推定の結果が一致しないことがわかる．

このモデルでは2つの結果が一致するので，この6.2.4項の今後の議論では2つの推定結果を区別せずに，単に推定値とよぶことにする．式(6.52)の逆フーリエ変換を用いると， η_t に関する推定値 $\langle \eta_t \rangle$ を求めることができる，

$$\langle \eta_t \rangle = \frac{1}{\sqrt{T}} \sum_{\omega=1}^T \langle \tilde{\eta}_\omega \rangle \exp(2\pi i \frac{\omega t}{T}) = \frac{1}{T} \sum_{\tau=1}^T \sum_{\omega=1}^T \frac{\cos(2\pi\omega(t-\tau)/T)}{2\gamma^2 F_\omega} \lambda_\tau. \quad (6.61)$$

この式は画像処理に用いられるウィナーフィルターに対応している(田中, 2006)．このモデルでは PM 推定と MAP 推定の結果は同じである．このモデルでは MAP 推定をするために最急降下法などの繰り返し計算をする必要はなく，式(6.61)のフィルターで繰り返し計算なしで，言わば一撃計算で結果が求まる．

次に周辺尤度最大化によるハイパーパラメータ推定をフーリエ変換を用いて議論する．発火率生成過程と観測過程の真のハイパーパラメータが β_0 と γ_0 であったとしよう．フーリエ表示された周辺尤度の対数を取り，式(6.49)と同様に対数周辺尤度の期待値を計算すると，

$$\langle \log P(\{\tilde{\lambda}_\omega\}; \beta, \gamma) \rangle \quad (6.62)$$

$$= \int \left(\prod_{\omega'=1}^T d\lambda_{\omega'} \right) P(\{\tilde{\lambda}_\omega\}; \beta_0, \gamma_0) \log P(\{\tilde{\lambda}_\omega\}; \beta, \gamma) \quad (6.63)$$

$$= \frac{1}{2} \sum_{\omega=1}^T \left[\log \frac{\beta_0 G_\omega}{2\gamma_0^2 F_\omega^0} - \log \pi - 1 - \left\{ \left(\frac{\beta}{\gamma^2 F_\omega} \right) \left(\frac{\beta_0}{\gamma_0^2 F_\omega^0} \right)^{-1} - 1 - \log \left(\frac{\beta}{\gamma^2 F_\omega} \right) \left(\frac{\beta_0}{\gamma_0^2 F_\omega^0} \right)^{-1} \right\} \right] \quad (6.64)$$

$$F_\omega^0 = \beta_0 G_\omega + \frac{1}{2\gamma_0^2}, \quad F_\omega = \beta G_\omega + \frac{1}{2\gamma^2} \quad (6.65)$$

となる．式(6.49)の結果と同様に， $\beta_0 = \beta$ ， $\gamma_0 = \gamma$ のときに，対数周辺尤度の期待値は最大になる．

周辺尤度最大化と平均 2 乗誤差 L の関係を述べる．平均 2 乗誤差 L は実表示とフーリエ表示を用いて，

$$L = \frac{1}{T} \sum_{t=1}^T (\langle \eta_t \rangle - \eta_t)^2 = \frac{1}{T} \sum_{\omega=1}^T (\langle \tilde{\eta}_\omega \rangle - \tilde{\eta}_\omega)^2, \quad (6.66)$$

と書ける．ハイパーパラメータの真の値が β_0, γ_0 である仮定して，平均 2 乗誤差 L の期待値を計算すると，

$$\begin{aligned} \langle L \rangle &= \int \left(\prod_{\omega'=1}^T d\tilde{\eta}_{\omega'} d\tilde{\lambda}_{\omega'} \right) P(\{\tilde{\eta}_\omega\}, \{\tilde{\lambda}_\omega\}; \beta_0, \gamma_0) L \\ &= \sum_{\omega=1}^T \frac{1}{2TF_\omega^0} \left\{ 1 + \frac{\gamma_0^2 (F_\omega^0)^2}{2\beta_0 G_\omega} \left(\frac{1}{\gamma_0^2 F_\omega^0} - \frac{1}{\gamma^2 F_\omega} \right)^2 \right\}, \end{aligned} \quad (6.67)$$

となる．したがって平均 2 乗誤差 $\langle L \rangle$ は以下の条件のときに最小化される，

$$\beta\gamma^2 = \beta_0\gamma_0^2. \quad (6.68)$$

この条件は $\beta_0 = \beta, \gamma_0 = \gamma$ を含んでいるので，対数周辺尤度の期待値を最大化するハイパーパラメータは，平均 2 乗誤差最小の基準を満たすことがわかった．このことは，仮定した事前確率と生成モデルの関数型が正しい場合，スパイクの平均発火率の観測データ $\{\lambda_t\}$ だけから，2 乗誤差の観点からもっとも良い推定を行えることを意味する．

6.2.5 変動発火率のワンショット推定

前の 6.2.3 項の議論をもとに，1 個のニューロンの発火率をスパイク列 1 回の測定からワンショットでベイズ推定する枠組みを考えてみよう (Smith & Brown, 2003; Koyama & Shinomoto, 2005)．最初に戻り 6.1 節のように，1 つの時間ビンにたかだかスパイクが 1 個含まれる場合を議論する．ニューロンが 1 個なので，そのニューロンの時刻 t でのスパイクのあるなしを $S_t = 0, 1$ で表す．時刻 t での発火率 η_t が与えられるとき，スパイク S_t は式 (6.1) で生成される．各時間ビンで発火率 η_t が与えられたときに，スパイクは独立に生成されると仮定する．スパイク全体 $\{S_t\}$ は発火率を $\{\eta_t\}$ を条件とした条件付き確率で生成される，

$$P(\{S_t\}|\{\eta_t\}) = \prod_{t=1}^T \eta_t^{S_t} (1 - \eta_t)^{(1-S_t)}. \quad (6.69)$$

6.2.3 項では複数回の試行で発火率の測定値 $\{\lambda_t\}$ を測定し、それらの測定値はガウス分布に従うと仮定して議論した。式 (6.69) からわかるように、スパイク $\{S_t\}$ はガウス分布に従わないので、6.2.3 項とは異なった取り扱いが必要である。その観点で、本節での議論は生成モデルが非ガウスである場合のベイズ推定の良い例である。

式 (6.69) の生成過程の条件である発火率 $\{\eta_t\}$ の事前確率は 6.2.3 項と同じ

$$P(\{\eta_t\}) = \frac{1}{Z_0(\beta)} \exp \left(-\beta \sum_{t=1}^{T-1} (\eta_{t+1} - \eta_t)^2 \right),$$

$$Z_0(\beta) = \int_0^1 \left(\prod_{\tau=1}^T d\eta_\tau \right) \exp \left(-\beta \sum_{t=1}^{T-1} (\eta_{t+1} - \eta_t)^2 \right), \quad (6.70)$$

を用いる。ここで β は発火率 η_t の滑らかさを表すハイパーパラメータであり、 $Z_0(\beta)$ は規格化定数である。式 (6.69) と (6.70) からベイズの定理を用いて発火率の事後確率 $p(\{\eta_t\}|\{S_t\})$ を、

$$P(\{\eta_t\}|\{S_t\}) = \frac{P(\{S_t\}|\{\eta_t\})P(\{\eta_t\})}{\int_0^1 \left(\prod_{\tau} d\eta_\tau \right) P(\{S_t\}|\{\eta_t\})P(\{\eta_t\})}, \quad (6.71)$$

として計算する。前の 6.2.3 項と同様に式 (6.42) を用いて、発火率を最大事後確率推定する。式 (6.71) の分子の対数を取りエネルギーを定義する、

$$E(\{\eta_t\}|\{S_t\}) = - \sum_{t=1}^T (S_t \log \eta_t + (1 - S_t) \log(1 - \eta_t))$$

$$+ \beta \sum_{t=1}^{T-1} (\eta_{t+1} - \eta_t)^2. \quad (6.72)$$

式 (6.72) の最小値を求めることが MAP 推定に対応する。 $E(\{\eta_t\}|\{S_t\})$ の最小値はどのような η_t によって与えられるか考えてみよう。式 (6.72) の最初の和に関しては、その和を大きくするとエネルギーは小さくなる、これは、0 または 1 の 2 値をとる S_t に対して、それぞれ $\log(1 - \eta_t)$ と $\log \eta_t$ を最大にする η_t

を選ぶことに対応する．発火率の範囲が $0 \leq \eta_t \leq 1$ に制限されるので，エネルギーの最初の部分は， $S_t = 1$ に対応する時刻 t にスパイクがある場合は $\eta_t = 1$ と推定し， $S_t = 0$ に対応するスパイクがない場合は $\eta_t = 0$ と推定する寄与を持つ．つまりスパイクのあるなしに対応して，ビンごとに発火率が離散的に 0 または 1 をとってしまう．これは 6.2.2 項の画像修復において，第 1 項のデータフィット項だけは推定値が観測ノイズを含んだ観測値と同じになってしまうことに対応している．

それを避けるのが画像修復における滑らかさの拘束条件に対応する式 (6.72) の 2 つ目の和である．これは，隣り合った時間ビンの発火率の差 $\eta_{t+1} - \eta_t$ が小さいほど小さな値をとる．スパイクのあるなしで発火率を 0 と 1 にする項と時間的に隣り合った発火率を滑らかにしようとする項の競合の結果，観測データであるスパイクのあるなしを考慮に入れた比較的滑らかな発火率 $\{\eta_t\}$ が推定されるようになっている．

式 (6.72) の $E(\{\eta_t\}|\{S_t\})$ を最小にする $\{\eta_t\}$ を 6.2.2 項と同様に最急降下法で求める，

$$\eta_t^{(n+1)} = \eta_t^{(n)} - \kappa \frac{\partial E(\{\eta_t\}|\{S_t\})}{\partial \eta_t} \Big|_{\eta_t = \eta_t^{(n)}} \quad (6.73)$$

$$\frac{\partial E(\{\eta_t\}|\{S_t\})}{\partial \eta_t} = -\frac{S_t}{\eta_t} + \frac{1 - S_t}{1 - \eta_t} + \kappa(-\eta_{t+1} + 2\eta_t - \eta_{t-1}). \quad (6.74)$$

ここで κ は十分小さな正の定数である．式 (6.74) の右辺の第 1 項と第 2 項は式 (6.72) の最初の和に対応する．これら 2 つの項は，ビン内にスパイクがあれば発火率 η_t を増加させ，スパイクがなければ減少させるはたらきを持つ．第 3 項は $\eta_t^{(n)}$ が隣接する $\eta_{t+1}^{(n)}$ と $\eta_{t-1}^{(n)}$ の平均値に等しいときに 0 になる．このように式 (6.73) と (6.74) から， $E(\{\eta_t\}|\{S_t\})$ を最小化することにより，スパイクのあるなしを考慮に入れた比較的滑らかな発火率 $\{\eta_t\}$ が推定されるようになっている．6.2.2 項と同様に，この最急降下法も η_t の更新の際に両隣の η_{t+1} と η_{t-1} の値しか用いない局所的な計算である．対数の単調増加性から，式 (6.72) の $E(\{\eta_t\}|\{S_t\})$ の 1 つ目の和は凹関数である．6.2.2 項と同様に，2 つ目の和は 2 次関数なので凹関数である．凹関数と凹関数の和は凹関数なので，式 (6.72) の $E(\{\eta_t\}|\{S_t\})$ は凹関数であり，局所的な最小値は存在せず，任意の初期値から計算をしても最小値に収束する．

次に PM 推定を説明する．後に例で示すように，このモデルでは MAP 推定と PM 推定の結果にはずれが生じる．PM 推定では，真の値 η_t の平均値，

$$\langle \eta_t \rangle = \int_0^1 \left(\prod_{\tau=1}^T d\eta_\tau \right) \eta_t P(\{\eta_t\} | \{S_t\}) \quad (6.75)$$

$$= \int_0^1 d\eta_t \eta_t P(\eta_t | \{S_t\}) \quad (6.76)$$

$$P(\eta_t | \{S_t\}) \equiv \int_0^1 \left(\prod_{\tau=1}^{t-1} d\eta_\tau \right) \left(\prod_{\tau=t+1}^T d\eta_\tau \right) P(\{\eta_t\} | \{S_t\}) \quad (6.77)$$

を推定値とする．式 (6.75) からわかるように PM 推定には T 次の積分が必要になる．式 (6.76) の $P(\eta_t | \{S_t\})$ は，事後確率 $P(\{\eta_t\} | \{S_t\})$ を，推定したい η_t 以外の確率変数で周辺化しているので，周辺化事後確率とよばれる． $P(\{\eta_t\} | \{S_t\})$ はガウス分布に従わないので，前項のようなフーリエ変換を用いて対角化を行い計算量を削減することはできない．

一方，ここで議論している系は 1 次元系なので，以下に示すように転送行列法を用いて， T 次の積分を $O(T)$ の計算量で数値的に厳密に計算することが可能である (Watanabe et al., 2009)．式 (6.69)–(6.72) より，式 (6.76) の周辺化事後確率を，

$$P(\eta_t | \{S_t\}) \propto \int_0^1 \prod_{\tau=1}^{t-1} d\eta_\tau W(\eta_{\tau+1}, \eta_\tau) \int_0^1 \prod_{\tau=t+1}^{T-1} d\eta_\tau W(\eta_{\tau+1}, \eta_\tau), \quad (6.78)$$

$$W(\eta_{t+1}, \eta_t) = \exp \left(S_t \log \eta_t + (1 - S_t) \log(1 - \eta_t) - \beta(\eta_{t+1} - \eta_t)^2 \right), \quad (6.79)$$

$$W(\eta_{T+1}, \eta_T) = W(\eta_T) = \exp (S_T \log \eta_T + (1 - S_T) \log(1 - \eta_T)), \quad (6.80)$$

と表すことができる．式 (6.78) の右辺は， $\tau = 1$ から $\tau = t - 1$ までの左側の積分と $\tau = t + 1$ から $\tau = T - 1$ までの右側の積分に分離している．左側の積分は $\tau = 1$ から $t - 1$ まで τ が大きくなる方向へ順番に積分していくことができる．同様に右側の積分は $\tau = T$ から $t + 1$ へと τ が小さくなる方向へ順番に積分していくことができる．したがって，これら 2 つの積分は以下のような漸化式で計算できる．

$$P(\eta_t|\{S_t\}) \propto L_{t-1}^t(\eta_t)R_{t+1}^t(\eta_t). \quad (6.81)$$

$$\begin{cases} L_{\tau-1}^\tau(\eta_\tau) = \int_0^1 d\eta_{\tau-1} L_{\tau-2}^{\tau-1}(\eta_{\tau-1})W(\eta_\tau, \eta_{\tau-1}) \\ L_0^1(\eta_1) = 1 \end{cases} \quad (6.82)$$

$$\begin{cases} R_{\tau+1}^\tau(\eta_\tau) = \int_0^1 d\eta_{\tau+1} R_{\tau+2}^{\tau+1}(\eta_{\tau+1})W(\eta_{\tau+1}, \eta_\tau) \\ R_{T+1}^T(\eta_T) = W(\eta_T). \end{cases} \quad (6.83)$$

漸化式 (6.82) と (6.83) の 1 つの数値積分の計算量は数値積分の刻みの 2 乗であり、時間ビンの数 T には依存しない。したがって式 (6.81) を求めるための計算量は T に関する 1 次のオーダーである。このようにして数値的に求めた $L_{t-1}^t(\eta_t)$ と $R_{t+1}^t(\eta_t)$ を使って、確率の規格化条件を考慮して周辺事後確率 $P(\eta_t|\{S_t\})$ を求めると、

$$P(\eta_t|\{S_t\}) = \frac{L_{t-1}^t(\eta_t)R_{t+1}^t(\eta_t)}{\int_0^1 d\eta_t L_{t-1}^t(\eta_t)R_{t+1}^t(\eta_t)}, \quad (6.84)$$

となる。この式を式 (6.76) に代入し、数値的に厳密に PM 推定値を求めることができる。転送行列法は確率的推論で用いられている確率伝播法と同じものである。今回のような 1 次元系だけでなく、同時確率が隣接した確率変数の同時確率の積で表されるような場合に適用可能である。さらに、そのような場合以外にも近似的に適用されて成功を収めている。

図 6.3 は MAP 推定と PM 推定の結果である。図の点線は式 (6.70) の発火率の事前確率 $P(\{\eta_t\})$ から生成した発火率の系列 $\{\eta_t\}$ を表す。事前確率のハイパーパラメータは $\beta = 50$ を使い、ビン数は $T = 400$ である。図 6.3 の破線は MAP 推定の結果であり、実線は PM 推定の結果である。図 6.3 から、2 つの結果が少し異なることがわかる。

ハイパーパラメータ推定に必要な、式 (6.71) の分母に対応する周辺尤度 $P(\{S_t\})$,

$$P(\{S_t\}) = \frac{1}{Z_0(\beta)} \int_0^1 \prod_{\tau=1}^T d\eta_\tau W(\eta_{\tau+1}, \eta_\tau) \quad (6.85)$$

$$Z_0(\beta) = \int_0^1 \prod_{\tau=1}^N d\eta_\tau W^0(\eta_{\tau+1}, \eta_\tau) \quad (6.86)$$

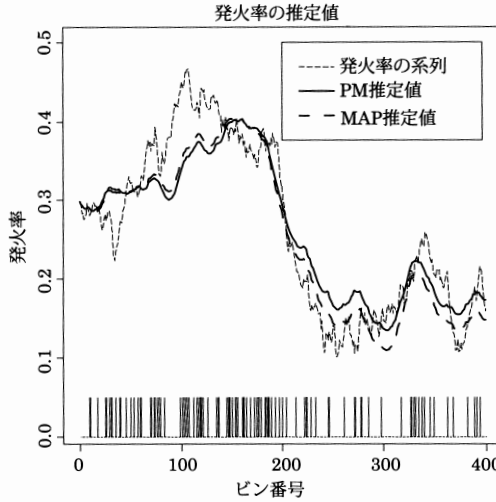


図 6.3 式 (6.70) の事前分布から生成した発火率 $\{\eta_t\}$ を推定した例

$$W_0(\eta_{\tau+1}, \eta_\tau) = \exp(-\beta(\eta_{\tau+1} - \eta_\tau)^2) \quad (6.87)$$

$$W_0(\eta_{\tau+1}, \eta_\tau) = 1, \quad (6.88)$$

も以下のように転送行列法で計算することができる,

$$P(\{S_t\}) = \frac{\int_0^1 d\eta_T L_{T-1}^T(\eta_T) W(\eta_T)}{\int_0^1 d\eta_T L_{T-1}'^T(\eta_T)}, \quad (6.89)$$

$$L_{\tau-1}'^\tau(\eta_\tau) = \int_0^1 d\eta_{\tau-1} L_{\tau-2}'^{\tau-1}(\eta_{\tau-1}) W_0(\eta_\tau, \eta_{\tau-1}). \quad (6.90)$$

このように数値的に求められた周辺尤度を用いて、数値的にハイパーパラメータ β を決めることができる (Watanabe et al., 2009).

最後にここで紹介した手法の最近の発展を簡単に紹介する．式 (6.70) は隣り合った発火率は似ているという拘束条件を表している．しかしながら、神経科学では突然刺激が切り替わる状況も考えられ、その場合は式 (6.70) の条件は適切でないことが考えられる．最近我々は、画像処理のラインプロセスの考え方 (田中, 2006) を発火率推定に適用し、このような状況に対処できる手法を提案した (Takiyama et al., 2009).

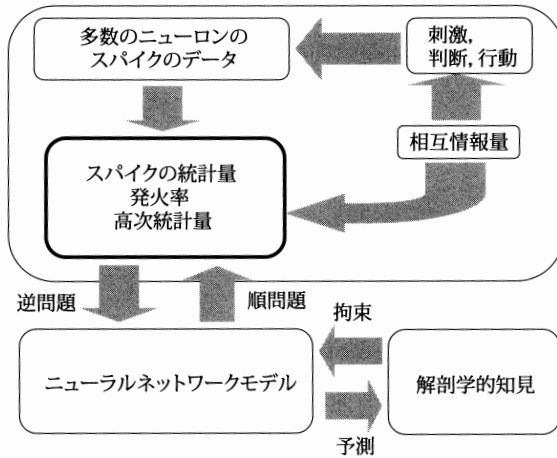


図 6.4 脳の計算論のフローチャート

6.3 まとめと関連事項

本章を終わるにあたって、ここで議論されたスパイクの確率的記述が神経科学でどのように位置づけられるかを述べる。

図 6.4 は、脳の理論的アプローチのフローチャートである。ニューロンの応答を測定するときは、動物に画像などの刺激を与えたり、判断や行動を伴うタスクを行わせる。これと同時にニューロンの応答であるスパイクを計測する。そのスパイクデータからスパイクの 1 次統計量である発火率を計算したり、2 次以上の高次統計量を計算する。本章での解説はこの過程に対応し、図 6.4 の太線で囲われた部分に対応する。

次に得られたスパイクの統計量が意味あるものかどうかを吟味する必要がある。たとえば、用意したすべての視覚刺激に対して、ニューロンの応答がまったく同じであれば、そのニューロンの応答をはかっても意味がないことは理解できるであろう。用意した刺激に対して、いろいろな応答が得られれば、その応答からどんな刺激が動物に与えられたかを予測できるようになる。このような予測特性を表す指標の 1 つが相互情報量である。与えられた刺激とスパイクの間の相互情報量を計算して、相互情報量をもつスパイク統計量だけに注目す

る必要がある (Oizumi et al, 2009).

次に意味のあるスパイク統計量が、どのようなニューラルネットワークから生成されたかを考える。ここで重要な点は、得られる統計量の数はニューラルネットワークの自由度に比べて著しく少ないことである。因果の観点では、ニューラルネットワークを定めると、その結果として統計量が得られる。つまりニューラルネットワークから統計量を求めるのが順問題である。この順問題を解くための基礎理論が、本書で詳細に説明されている非線形動力学や確率過程の理論である。統計量からニューラルネットワークを決めることは逆問題となり、一般的には解けない。そこで順問題を解いておいて、それをもとに逆問題の解を推定する戦略をとる。具体的には、まず解剖学的知見などからニューラルネットワークの構造を予測する。予測したニューラルネットワークから、典型的にどのようなスパイク統計量が出てくるかを、理論もしくは計算機シミュレーションで計算しておく。これが順問題を解くことに相当する。いくつかの典型的なニューラルネットワークに関して同様なことを行い、スパイク統計量のデータベースのようなものをつくっておく。このデータベースと実験から得られたスパイク統計量を比較することで、いま問題にしている領野のニューラルネットワークに関する情報が得られる。さらには、そこから解剖学的な構造を予測することにもつながる。この順問題と逆問題のループを回すことが、スパイクの統計量とニューラルネットワークの構造の関係を解明する鍵になる。

刺激、判断、行動などとスパイクの統計量の間の関係を明確にすることと、スパイクの統計量とニューラルネットワークの構造の関係を解明することで、脳で行われる情報処理を深く理解することができる。

謝辞

本章の執筆にあたっては、東京大学大学院新領域創成科学研究科博士課程飯田宗徳氏、同修士課程瀧山健氏に多大なるご尽力を賜った。また、紹介した研究内容の一部は複数の共同研究者との成果に基づいている。これらの研究は共同研究者の方々のご尽力がなければ完成にはいたらなかった。ここに深く感謝する。