

状態遷移モデルにおける
ベイズ予測アルゴリズムに関する研究
A Study on Bayes Prediction Algorithms
for State Transition Models

2009年3月

須子 統太

目次

第1章	序論	1
1.1	研究背景	1
1.2	研究の目的と構成	3
第2章	準備	6
2.1	確率モデルを用いた予測	6
2.1.1	予測問題の定式化	6
2.1.2	予測に対する評価基準	7
2.1.3	ベイズ予測	9
2.1.4	ベイズ予測の計算	9
2.1.5	ベイズ予測の課題	10
2.2	状態遷移モデル	11
2.2.1	確率モデルの定義	11
2.2.2	モデルの分類	12
2.3	状態遷移モデルに対するベイズ予測	14
2.3.1	離散状態で状態遷移既知の状態遷移モデルに対するベイズ予測	14
2.3.2	離散状態で状態遷移未知の状態遷移モデルに対するベイズ予測	16
2.4	本研究の位置づけ	20
第3章	離散状態・状態遷移未知の場合の近似ベイズ予測アルゴリズム	21
3.1	計算量を削減した近似ベイズ予測アルゴリズムの提案	21

3.2	外れ値を含む回帰モデルへの応用	24
3.2.1	外れ値解析	24
3.2.2	外れ値を含む回帰モデル	25
3.2.3	提案近似アルゴリズムの適用	29
3.2.4	シミュレーションによる評価	30
3.2.5	シミュレーションに対する考察	35
3.3	本章のまとめ	36
第 4 章	連続状態・状態遷移未知の場合の効率的ベイズ予測アルゴリズム	39
4.1	ベイズ予測と情報源符号化	39
4.1.1	情報源符号化	39
4.1.2	ユニバーサル情報源符号化	40
4.1.3	ベイズ符号	42
4.2	i.p.i.d. 情報源モデル	43
4.3	i.p.i.d. 情報源に対するユニバーサル符号化アルゴリズム	44
4.4	i.p.i.d. 情報源モデルに対するベイズ符号	44
4.5	効率的逐次ベイズ符号化アルゴリズムの提案	46
4.5.1	m の事前分布	46
4.5.2	効率的アルゴリズム	46
4.6	アルゴリズムの最適性	49
4.7	本章のまとめ	50
第 5 章	連続状態・状態遷移既知の場合のベイズ予測の評価	51
5.1	i.p.i.d. 情報源に対するユニバーサル符号の漸近評価	51
5.2	変化パターンが既知の場合のベイズ符号	52
5.3	変化パターンのクラスの定義	53

5.3.1	クラスの定義	53
5.3.2	ベルヌーイ分布に従いパラメータ変化が起きる場合	53
5.4	リスクの評価	55
5.5	本章のまとめ	58
第 6 章	結論	60
6.1	まとめ	60
6.2	今後の展望	61
6.2.1	離散状態で状態遷移未知の状態遷移モデルに対して	61
6.2.2	i.p.i.d. 情報源モデルに対して	62
	謝辞	63
	参考文献	65
	研究業績	69

第1章 序論

1.1 研究背景

与えられたデータから次に発生するデータを予測する問題は、様々な工学分野における本質的な問題の一つである。例えば、株価予測や需要予測などの時系列データに対する予測の他にも、手書き文字認識や音声認識などのパターン認識や、信号処理の分野における信号推定、情報理論におけるデータ圧縮など、予測問題と捉えられる問題は多岐にわたる。そのため、従来から様々な分野で予測問題は取り扱われてきた。これらを統一的に捉え、より広い視点から体系化することは非常に重要である。

予測問題は、 n 個の時系列データが与えられたもとの $n+1$ 番目のデータを予測する問題や、 n 組の説明変数と目的変数の組と $n+1$ 番目の説明変数が与えられたもとの、 $n+1$ 番目の目的変数を予測する問題などが含まれる。また、情報理論や制御理論などでは、 n 個の時系列データが与えられたもとの $n+1$ 番目のデータの出現する確率分布や、 n 個のデータが出現した確率分布を推定する問題が扱われている。本研究ではこれらも同様に予測問題として扱う。

予測問題の評価基準としては、予測値と真の値との誤差を表す損失関数があげられる。損失関数はデータの関数として得られるため、与えられたデータにより値が変わる。そのため、損失関数をデータの出現確率で期待値を取ったりリスク関数が評価基準として用いられることも多い。従来から様々な予測問題に対し、提案された予測法についてこれら損失関数やリスク関数により予測性能の評価が行われている。

また、確率モデルを用いた予測法として、ベイズ決定理論に基づく予測法（以下ベイズ予測）がある。ベイズ予測は、ベイズ基準のもとで損失を最小にする予測法である。近年、学習理論や情報理論において注目を集めており、その有効性が示されている。ベイズ基準のもとで最適とは、リスク関数を確率モデルにおける未知パラメータの事前分布で期待値をとったベイズリスク関数を考え、そのベイズリスク関数を最小にするという意味である。ベイズ予測の値は、定義された損失関数により異なる。しかし、多くの損失関数に対しベイズ予測は、予測分布と呼ばれる確率分布の関数として与えられる。そのため、ベイズ予測を行うには、予測分布の計算が中心的な役割を果たす。一般の確率モデルに対し、この予測分布の計算は積分が含まれ解析的に解けない、和計算の計算量が膨大になる、などの問題点があることが知られている。そのため、ベイズ予測を実用化するには何らかの計算量削減アルゴリズムが必要となる。また、ベイズ予測はベイズ基準のもとで損失を最小にすることを保証するが、一般の確率モデルに対し、ベイズ基準以外の評価基準における予測性能は保証されとは限らない。そのためベイズ予測を実用化するにあたって、ベイズ基準以外の視点から予測性能を評価する必要がある。従来、学習理論や情報理論の分野では、損失やリスクの漸近評価により予測性能を評価する研究が多く行われている。本研究では、これらベイズ予測を実用化するにあたっての課題に着目する。

他方、予測問題に用いる確率モデルとして、マルコフモデルや隠れマルコフモデルなど、離散時間で状態遷移が起きる確率モデルが学習理論・情報理論を中心に用いられており、実用的にも有効性が示されている。本研究ではこれら離散時間で状態遷移の起きる確率モデルを状態遷移モデルと呼ぶ。状態遷移モデルは状態が連続であるか離散であるか、状態遷移が既知であるか未知であるかで様々なモデルに分類することができる。例えば、状態が離散で状態遷移系列が既知のモデルとしては、マルコフモデルやパターン認識で用いられる決定木モデル、情報理論で用いられる Context Tree

モデルなどがある．また，状態が離散で状態遷移系列が未知のモデルとしては，隠れマルコフモデルや混合分布モデルなどがある．さらに，状態が連続のモデルとしては，情報理論で用いられている i.p.i.d.(independent piecewise-identically-distributed) 情報源モデルなどがある．

1.2 研究の目的と構成

本研究では，状態遷移モデルに含まれるいくつかのモデルに対しベイズ予測を適用することを目的とする．そのもとで，ベイズ予測の実用化に伴う前述の課題に対し，計算量削減アルゴリズムの提案とリスクの漸近評価を行う．

以下第 2 章以降の構成を示す．

第 2 章では，準備として，予測問題の分類及びベイズ予測の定式化と，状態遷移モデルについて定義する．また，ベイズ予測に関する従来研究について述べる．

第 3 章では，離散状態の状態遷移モデルで状態遷移が未知の場合に対し，計算量を削減した近似ベイズ予測アルゴリズムを提案する．この場合の状態遷移モデルにおけるベイズ予測は，取り得る全ての状態遷移のパターンについて事後確率の重み付け和を計算することで求められる．状態遷移のパターンは状態数のデータ数乗のパターンが存在する．そのため，事後確率の重み付け和の計算には，データ数に対し指数オーダーの計算量がかかり計算量的に困難である．そこで提案アルゴリズムは，全ての状態遷移のパターンを重み付けるのではなく，事後確率の高い状態遷移パターンを近似的にいくつか求め，それらについての重み付け計算を行うことで計算量を削減する．また，従来外れ値解析で用いられていた外れ値を含む回帰モデルに提案アルゴリズムを適用し，シミュレーションにより評価を行う．ここで言う外れ値とは，解析対象とする母集団以外から発生したデータのことを表す．外れ値は少量でもデータに含まれると解析結果に大きな影響を与えることが知られており，従来より外れ値の取り扱いに関する研究が数多くなされている．外れ値の発生にはモデル

を仮定する場合と，仮定しない場合の研究があるが，外れ値を含む回帰モデルは前者の外れ値の発生にモデルを仮定した場合のモデルのひとつである．また，外れ値解析の目的は，外れ値の検出や，パラメータ推定，予測などが考えられるが，外れ値を含む回帰モデルに対し予測を目的とした研究は行われていない．そこで，本研究では外れ値を含む回帰モデルに対し，予測を目的とした解析法として，計算量を削減した近似ベイズ予測アルゴリズムを適用する．シミュレーションによる評価の結果として，最適なベイズ予測の計算量がデータ数 n に対し $O(2^n)$ の計算量が必要であったのに対し，提案アルゴリズムは計算量を $O(n^2)$ に削減しつつ，予測性能の劣化が少ないアルゴリズムであることを示す．また，一般的なベイズ予測の近似アルゴリズムである MCMC (Markov Chain Monte Carlo) 法や変分ベイズ法との予測性能を比較し，提案アルゴリズムは従来近似アルゴリズムよりも予測性能が高くなることを示す．

第 4 章と第 5 章では，連続状態の状態遷移モデルである，i.p.i.d. 情報源モデルを扱う．i.p.i.d. 情報源モデルとは，情報理論におけるデータ圧縮に用いられる非定常な確率モデルである．従来データ圧縮の分野において，i.p.i.d. 情報源に対する予測アルゴリズム（圧縮アルゴリズム）が提案されているが，ベイズ予測を適用した研究はあまりなされていない．そこで本研究では i.p.i.d. 情報源モデルにベイズ予測を適用した場合の効率的なアルゴリズムの提案とリスクの漸近評価を行う．

第 4 章では，i.p.i.d. 情報源モデルに対し状態遷移が未知である場合の効率的なベイズ予測アルゴリズムを示す．状態遷移が未知の i.p.i.d. 情報源に対するベイズ予測は，状態遷移の全てのパターンについての重み付け和の計算を行うことから， $O(2^n)$ の計算量が必要であった．そこで，状態遷移がベルヌーイ分布に従って起きるという仮定を置くことで，ベイズ最適性を保持したまま計算量を $O(n^2)$ に削減したアルゴリズムを提案する．その結果，状態遷移が未知の i.p.i.d. 情報源に対し，計算量とメモリ量の観点からベイズ予測が実用可能であることを示す．

第 5 章では, i.p.i.d. 情報源モデルに対し, ベイズ予測のリスクの漸近評価を行う. 従来 i.p.i.d. 情報源モデルにおける予測アルゴリズムを漸近評価する場合には, クラスに一部制約をおいたモデルに対し評価が行われている. そこで, 本研究では, 状態遷移が既知であり, データ数 n が無限になるとき状態遷移も無限回起きるという i.p.i.d. 情報源モデルのクラスを考える. このとき, 任意の予測法に対し, リスクとベイズリスクが漸近的に一致することから, ベイズ予測が漸近的にリスクを最小にすることを示す.

最後に第 6 章では, 以上の結果の結論として, まとめと今後の展望について述べる.

第2章 準備

本章では準備として、ベイズ予測の定式化とその課題について述べ、状態遷移モデルを定義する。また、状態遷移モデルに対するベイズ予測の適用について述べ、従来研究について整理する。そのもとで、本研究の位置づけを示す。

2.1 確率モデルを用いた予測

2.1.1 予測問題の定式化

予測問題は統計学のみならず、学習理論、データマイニング、パターン認識、情報理論など、様々な工学分野で扱われている基本的な問題のひとつである。また、予測対象に確率モデルを仮定することで、多くの予測手法が開発されている。本研究では様々な予測問題を統一的に捉え、予測問題の定式化を行う。

いま、 i 番目の説明変数を $x_i \in \mathcal{X}$ 、目的変数を $y_i \in \mathcal{Y}$ とする。1 番目から n 番目までの n 個の説明変数を $x^n = (x_1, x_2, \dots, x_n) \in \mathcal{X}^n$ とし、同様に 1 番目から n 番目までの n 個の目的変数を $y^n = (y_1, y_2, \dots, y_n) \in \mathcal{Y}^n$ と表す。本研究では予測問題を、説明変数と目的変数の n 個の組 (x^n, y^n) と $n + 1$ 番目の説明変数 x_{n+1} が与えられたもとで、 x_{n+1} に対応する目的変数 y_{n+1} を推定する問題として扱う。また、 y_{n+1} の値を予測する場合に以外にも、情報理論（情報源符号化）や制御理論では、 y_{n+1} の出現する確率分布を推定する問題や、 (x^n, y^n) が出現する確率分布を推定する問題などが考えられ、本研究ではこれらも同様に予測問題として扱う。さらに、時系列デー

タの予測は上記定式化における特殊な場合と考えられ、説明変数 x_i が y^{i-1} である場合に相当する。

\mathcal{X}, \mathcal{Y} は離散集合、連続集合いずれの場合も考えられる。回帰分析では \mathcal{X}, \mathcal{Y} ともに連続集合として扱っている、判別分析やパターン認識では \mathcal{X} を連続集合、 \mathcal{Y} を離散集合として扱っている。また、情報理論における情報源符号化（情報源符号化）では \mathcal{X}, \mathcal{Y} ともに離散集合として扱っている。本研究では \mathcal{X}, \mathcal{Y} については離散集合、連続集合どちらの場合においても同様の議論が可能であるが、以下本章では、特に断りが無い限り \mathcal{X}, \mathcal{Y} についてはいずれも連続集合として説明を行う。

また、 y_i は確率分布 $p(y_i|x_i, \theta)$ に従い発生すると仮定する。ここで、 $\theta \in \Theta$ は確率分布の未知パラメータとし、分布系は既知であると仮定する。 Θ についても連続集合の場合と離散集合の場合が考えられるが、以下本章では特に断りが無い限り連続集合として説明する。また、以降真のパラメータを θ^* と表記する。

2.1.2 予測に対する評価基準

予測に対する評価基準として、予測の誤差を損失関数 $Loss$ として定義する。損失関数は対象とする予測問題によって様々な場合が考えられる。 y_{n+1} の予測値を (x^n, y^n) と x_{n+1} の関数として $\hat{y}_{n+1}(x^n, y^n, x_{n+1})$ と表記する。例えば、回帰モデルにおける予測や学習理論などでは以下の二乗誤差損失が考えられる。

$$Loss(\hat{y}_{n+1}(x^n, y^n, x_{n+1}), y_{n+1}) = (y_{n+1} - \hat{y}_{n+1}(x^n, y^n, x_{n+1}))^2. \quad (2.1)$$

また、 y_{n+1} の出現確率で期待値を取った平均二乗誤差損失なども考えられる。 y_{n+1} と (x^n, y^n) が独立に生起している場合、平均二乗誤差損失は以下で与えられる。

$$Loss(\hat{y}_{n+1}(x^n, y^n, x_{n+1}), \theta^*) = \int_{\mathcal{Y}} (y_{n+1} - \hat{y}_{n+1}(x^n, y^n, x_{n+1}))^2 p(y_{n+1}|x_{n+1}, \theta^*) dy_{n+1}, \quad (2.2)$$

さらに，判別分析やパターン認識のように \mathcal{Y} が離散集合の場合には以下の 0-1 損失を考える場合が多い．

$$Loss(\hat{y}_{n+1}(x^n, y^n, x_{n+1}), y_{n+1}) = \begin{cases} 0 & y_{n+1} = \hat{y}_{n+1} \\ 1 & y_{n+1} \neq \hat{y}_{n+1} \end{cases} . \quad (2.3)$$

その他にも，データ圧縮などの確率分布を推定する場合には，対数損失などが考える場合もある． y_{n+1} の確率分布の推定値を \hat{p} とすると，対数損失は以下で与えられる．

$$Loss(\hat{p}(y_{n+1}|x_{n+1}, x^n, y^n), \theta) = \log p(y_{n+1}|x_{n+1}, \theta^*) - \log \hat{p}(y_{n+1}|x_{n+1}, x^n, y^n). \quad (2.4)$$

また， y^n の分布の推定値を \hat{P} とすると，同様に対数損失は以下で定義される．

$$Loss(\hat{P}(y^n|x^n), \theta) = \log P(y^n|x^n, \theta^*) - \log \hat{P}(y^n|x^n). \quad (2.5)$$

これら損失関数はいずれも得られた n 個のデータ (x^n, y^n) の関数である．そのため，実際に得られたデータの実現値によって損失の値が変わってしまい評価しにくい．そこで，損失関数をデータの出現確率で期待値を取ったリスク関数（危険関数）を評価基準とする事も考えられる．例えば，(2.2) 式の平均二乗誤差を損失関数とした場合，リスク関数 $Risk$ は以下で与えられる．

$$Risk(\hat{y}_{n+1}, \theta^*) = \int_{\mathcal{Y}^n} \int_{\mathcal{X}^n} Loss(\hat{y}_{n+1}(x^n, y^n), \theta^*) P(y^n|x^n, \theta^*) P(x^n) dy^n dx^n. \quad (2.6)$$

ここで $v(x^n)$ は x^n の出現する事前分布である．

一般に予測問題では，これらの損失関数やリスク関数を予測に対する評価基準としている．しかし，損失関数やリスク関数は θ^* の関数として与えられており，有限時点で任意の θ^* に対し損失関数やリスク関数を最小化する予測法が存在するとは限らない．

2.1.3 ベイズ予測

そこで、ベイズ決定理論ではリスク関数に対し、未知パラメータ θ の事前分布 $w(\theta)$ ¹ で期待値を取った、ベイズリスク関数を考える。例えば、(2.2) 式の平均二乗誤差を損失関数とした場合、ベイズリスク関数 BR は以下で与えられる。

$$BR(\hat{y}_{n+1}) = \int_{\Theta} Risk(\hat{y}_{n+1}, \theta) w(\theta) d\theta. \quad (2.7)$$

このベイズリスク関数を最小にする予測を、ベイズ基準の基で最適な予測（ベイズ予測）とする。

ベイズ予測は近年、特に学習理論や情報理論（データ圧縮）の分野でその有効性が示されており注目を集めている。

2.1.4 ベイズ予測の計算

ベイズリスクの最小化は、損失関数により異なるが、多くの場合以下の予測分布の計算が重要である。 Θ が連続集合の場合の予測分布は以下で与えられる。

$$\hat{p}^*(y_{n+1}|x_{n+1}, x^n, y^n) = \int_{\Theta} p(y_{n+1}|x_{n+1}, \theta) w(\theta|x^n, y^n) d\theta. \quad (2.8)$$

また、 Θ が離散集合の場合の予測分布は以下で与えられる。

$$\hat{p}^*(y_{n+1}|x_{n+1}, x^n, y^n) = \sum_{\Theta} p(y_{n+1}|x_{n+1}, \theta) w(\theta|x^n, y^n). \quad (2.9)$$

例えば、(2.2) 式の平均二乗誤差を損失関数とした場合、ベイズ予測は予測分布の期待値となり、次式で与えられる。

$$\hat{y}_{n+1}^* = \int_{y^n} y_{n+1} \int_{\Theta} p(y_{n+1}|x_{n+1}, \theta) w(\theta|x^n, y^n) d\theta dy_{n+1}. \quad (2.10)$$

¹ θ が連続のとき、正確には $w(\theta)$ は事前密度関数を表すため、 θ が連続であるか離散であるかで事前分布の表記を変えるべきである。しかし本稿では特に混乱の無い場合、 θ が連続、離散に関わらず事前分布を $w(\theta)$ と表記することとする。

このように，ベイズ予測を実現するには予測分布の計算ができれば，中心的な課題は解決する．しかし，一般にこの予測分布の計算は容易ではない． Θ が連続集合の場合， $p(y_i|x_i, \theta)$ が指数型の分布であれば，自然共役な事前分布を仮定することで，(2.8) 式の積分を解析的に解く事ができるが，少し複雑な分布系になると解析的に解けず，MCMC 法や変分ベイズ法などの近似計算が必要となる．また， Θ が離散集合の場合 (2.9) 式の計算は $|\Theta|$ 回の和計算が必要となる． $|\Theta|$ が大きい場合には予測分布の計算が，計算量的に困難になるため，計算量を削減した効率化アルゴリズムや近似アルゴリズムが必要となる．

2.1.5 ベイズ予測の課題

ベイズ予測の課題のひとつとしては，前述の予測分布の計算が挙げられる．ベイズ予測を実用化するためには，現実的な計算量でベイズ最適を達成する効率的な予測アルゴリズムや，計算量を削減した精度の高い近似アルゴリズムなどの開発が必要である．本研究ではこの点に着目し，ベイズ予測の計算量削減方法について議論する．

また，ベイズ予測はベイズ基準の基で損失を最小化する予測法であるが，一般に他の基準や評価方法の基ではその性質は明らかではない．そのため，ベイズ基準以外の評価基準を考えた場合，ベイズ予測がどのような性質を持つか調べることも重要な課題のひとつである．従来より学習理論や情報理論の分野では，予測アルゴリズムに対し損失やリスクの漸近評価を行う事が多い．これは，学習理論や情報理論では得られるデータ数が非常に多い場合が想定され，漸近評価の結果が実用的な状況と合致するためである．これらの分野ではいくつかのモデルに対し，ベイズ予測についても漸近的な損失やリスクの評価が行われており，その有効性が示されている．そこで本研究ではこの点に着目し，従来評価が行われていないモデルに対しベイズ予測の漸近的なリスクの評価を行う．

その他の重要なベイズ予測の課題としては、事前分布の設定方法の明確化が挙げられる。前述のようにベイズ予測においては、未知パラメータの事前分布 $w(\theta)$ を仮定したもとで議論が行われている。ベイズ予測を実用化する際、事前分布をどのように設定するかは、予測精度にも関わってくることから非常に重要な問題であると言える。しかし、一般にこの問題は非常に難しく、従来より数多くの議論がなされているが根本的な解決はいまだなされていない。しかしながら、情報理論や学習理論では事前分布の設定についてあまり議論されていない。その理由として、情報理論では対象となるデータの構造がある程度事前に分かっている場合が多い事や、アルゴリズムの構成のしやすさから自然共役事前分布を仮定する場合が多い事などが挙げられる。また、学習理論ではデータの事前情報が全く無い状況を想定される事から、暗に一様分布を仮定している事が多い。本研究では、情報理論や学習理論と同様の立場をとり、事前分布は既に設定されているものとし詳細な議論については特に行わないこととする。

2.2 状態遷移モデル

2.2.1 確率モデルの定義

次に本研究で対象とする確率モデルを定義する。マルコフモデルや隠れマルコフモデルなど、状態遷移を用いた確率モデルは、従来より様々な分野で用いられておりその有効性が示されている。本研究では離散時間で状態遷移の起きる確率モデルを状態遷移モデルと呼び統一的に扱う。

いま、 i 時点における状態を $s_i \in S$ とし、1 時点から n 時点までの状態遷移系列を $s^n = (s_1, s_2, \dots, s_n)$ とする。また、各状態 $s \in S$ のもとで従う確率分布のパラメータを $\theta_s \in \Theta_s$ とし、1 時点から n 時点までのパラメータベクトルを、 $\theta_{s^n} = (\theta_{s_1}, \theta_{s_2}, \dots, \theta_{s_n}) \in \Theta_s^n$ とする。以降、 $\theta_s \in \Theta_s$ は連続パラメータとする。このとき、

以下の確率分布に従いデータが発生するモデルを状態遷移モデルと定義する .

$$P(y^n|x^n, \theta_{s^n}, s^n) = \prod_{i=1}^n p(y_i|x_i, \theta_{s_i}). \quad (2.11)$$

また , i 時点 ($i = 1, 2, \dots$) における状態 s_i には , $i - 1$ 時点の状態 s_{i-1} によって定まる状態遷移確率 $v(s_i|s_{i-1})$ に従い遷移すると仮定する . 以降 , この状態遷移確率については既知であるとする .

2.2.2 モデルの分類

状態遷移が確率的に起きると仮定した場合 , 状態遷移モデルは状態が離散状態か連続状態かによってモデル进行分类することができる .

離散状態のモデル

状態集合 S が離散集合の時 , 特に状態集合が有限の場合は , 状態は有限の状態集合のなかで遷移を繰り返す . そのため , ある状態から他の状態へ遷移した後 , 再び同一の状態に戻ってくる可能性がある . パラメータベクトル θ_{s^n} は同一のパラメータ θ_s が複数回含まれるので , 確率モデルは以下のように表記することができる .

$$P(y^n|x^n, \theta_{s^n}, s^n) = P(y^n|x^n, \theta_S, s^n). \quad (2.12)$$

ここで , $\theta_S = \{\theta_s | s \in S\} \in \Theta_S$ である . 状態遷移系列が既知の場合 , 上記モデルに含まれるモデルとしては , i.i.d. モデルやマルコフモデル , 情報理論で用いられる Context tree モデル , パターン認識で用いられる決定木モデルなどが挙げられる . また , 状態遷移系列が未知の場合では , 隠れマルコフモデルや混合分布モデルなどが上記モデルに含まれる .

連続状態のモデル

状態集合 S が連続集合の場合で特に，ある状態から他の状態へ遷移した後，再び同一の状態に遷移しないモデルを考える．ある s^n のもとで j 回目の状態遷移が起きた時点 $t_j^{s^n}$ とし遷移時点の集合を $T_{s^n} = \{t_0^{s^n}, t_1^{s^n}, \dots, t_{C_{s^n}}^{s^n}\}$ とする．ここで， C_{s^n} は s^n における状態遷移の回数とする．また， $\theta_{t_j^{s^n}}$ を j 回目の遷移から $j+1$ 回目の遷移が起きるまでに従う確率分布のパラメータとし， $\theta_{T_{s^n}} = \{\theta_t | t \in T_{s^n}\}$ とする．このとき，状態遷移モデルを次のように書き換える事ができる．

$$P(y^n | x^n, \theta_{s^n}, s^n) = P(y^n | x^n, \theta_{T_{s^n}}, T_{s^n}). \quad (2.13)$$

上記で表現される状態が連続の場合のモデルとしては，情報理論におけるデータ圧縮で用いられる i.p.i.d. (independent piecewise-identically-distributed) 情報源モデルが挙げられる．また，制御理論等で用いられるカルマンフィルタは状態遷移系列が未知の場合に相当する．

以上より，状態遷移モデルは表 2.1 のように分類される．

表 2.1: 状態遷移モデルの分類

	離散状態	連続状態
状態遷移既知	i.i.d. モデル マルコフモデル Context tree モデル 決定木モデル	i.p.i.d. 情報源モデル
状態遷移未知	隠れマルコフモデル 混合分布モデル	i.p.i.d. 情報源モデル カルマンフィルタ

2.3 状態遷移モデルに対するベイズ予測

2.3.1 離散状態で状態遷移既知の状態遷移モデルに対するベイズ予測

効率的アルゴリズムに関する研究

離散状態で状態遷移既知の場合のベイズ予測については、情報理論における情報源符号化において多くの研究が行われている。計算量、メモリ量の効率化に関しては、特に [4, 10, 11] など、自然共役事前分布やモデルの構造を利用することで、ベイズ最適性を保持したまま計算量やメモリ量を削減する研究が行われている。

情報源符号化では、時系列データ系列 $x^n = x_1, x_2, \dots, x_n$ が得られたもとで、次のデータの出現確率 $p(x_{n+1})$ を推定する問題（逐次符号化）、もしくは $P(x^n)$ を推定する問題（一括符号化）が扱われている。情報源符号化で用いられる最も単純なモデルのひとつとして、定常無記憶情報源（i.i.d. モデル）がある。

いま、 $\mathcal{X} = \{0, 1, \dots, k\}$ を離散有限集合、 θ を k 次元パラメータベクトルとしたとき、定常無記憶情報源は以下であらわされる。

$$P(x^N|\theta) = \prod_{n=1}^N p(x_n|\theta). \quad (2.14)$$

すなわち、各時点 n において、 x_n は独立同分布に従い発生するモデルである。このとき予測分布は、

$$\hat{p}^*(x_{n+1}|x^n) = \int_{\Theta} p(x_{n+1}|\theta)w(\theta|x^n)d\theta, \quad (2.15)$$

で与えられるが、パラメータの事前分布 $w(\theta)$ に自然共役事前分布であるディリクレ分布を仮定することで、上記の積分計算を解析的に次式で計算することができる。

$$\hat{p}^*(x_{n+1}|x^n) = \frac{\nu(x_{n+1}|x_1^n) + \beta(x_{n+1})}{\sum_{a=0}^k (\nu(a|x^n) + \beta(a))}, \quad (2.16)$$

但し、 $\nu(a|x^n)$ は時点 1 から時点 n において、記号 a が出現した回数とする。また、 $\beta(a)$ はディリクレ分布のパラメータとする。

このように，自然共役事前分布を仮定することで，積分計算を解析的に解くことができる．しかし，自然共役事前分布は $p(x_n|\theta)$ が指数型分布の場合のみ存在し，一般の確率モデルではこのように簡単に計算することはできない．

また，情報源符号化においてベイズ予測の効率的アルゴリズムに関して，もっとも多く研究されているモデルのひとつに Context Tree モデル (CT モデル) がある．CT モデルは階層モデルとして表現される状態遷移モデルのひとつである．CT モデルは，モデル $m \in M$ とモデル m の元でのパラメータ $\theta_m \in \Theta_m$ の条件付確率 $P(x^n|\theta_m, m)$ で定義される．モデル m とパラメータ θ_m が未知のとき，一括符号化で用いられる予測分布は以下で表される．

$$\hat{P}^*(x^N) = \sum_{m \in M} \int_{\theta_m \in \Theta_m} P(x^N|\theta_m, m) w(\theta_m|m) d\theta_m v(m) \quad (2.17)$$

ここで， $w(\theta_m|m), v(m)$ はそれぞれ θ_m, m の事前分布である．また，逐次符号化の予測分布は，

$$\hat{P}^*(x_{n+1}|x^n) = \sum_{m \in M} \int_{\theta_m \in \Theta_m} P(x_{n+1}|x^n, \theta_m, m) w(\theta_m|m, x^n) d\theta_m v(m|x^n) \quad (2.18)$$

で与えられる．

CT モデルに対するベイズ予測は，上記予測分布を計算することで構成されるが，CT モデルの木の深さが未知の場合， m の数は系列長 N に対し，指数的に増大する．そのため，(2.17),(2.18) 式の m に関する和計算の計算量・メモリ量が膨大になってしまい，そのままでは計算することができない．そこで，ベイズ予測の実用化には何らかの計算量・メモリ量削減アルゴリズムが必要となる．

Willems らによって提案された CTW 法は， $w(\theta_m|m), v(m)$ にある特定の分布を仮定したもとの，一括符号化の効率的アルゴリズムとしてみなすことができる．[4, 5] これにより，計算量，メモリ量共に $O(N^2)$ 程度まで削減された．

それに対し Matsushima らは，CTW 法と同時期に逐次効率化における効率的アルゴリズムを提案した．[11] このアルゴリズムは，CTW 法で仮定されている事前分布

$w(\theta_m|m), v(m)$ を含むより一般的な事前分布のクラスを仮定している．計算量，メモリ量ともに CTW 法と同等のアルゴリズムである．

CTW 法や Matsushima のアルゴリズムにより，計算量，メモリ量共に大幅な削減がなされたが，実用的に $O(N^2)$ のメモリ量では直ぐにメモリオーバーフローを起こす場合がある．そこで野村らは，重み付ける m の数を制限することで，メモリ量を削減した近似ベイズ予測アルゴリズムを提案した．[26] このアルゴリズムは，メモリ量を定数オーダーに削減した近似アルゴリズムであるが，漸近的には正確なベイズ予測と損失が一致するという性質をもつ．また中野らにより，CT モデルの木表現を変更することで，近似なしでメモリ量を $O(N)$ まで削減するアルゴリズムが提案されている．[27]

漸近評価に関する研究

損失関数やリスク関数の漸近評価に関しても同様に情報源符号化の分野で多く研究されており，ベイズ予測の様々な有効性が示されている．

i.i.d. モデルに対して，Clark らによりベイズ予測のリスク関数の漸近評価が行われている．Clark らは，最尤推定量の漸近正規性を用いることで，充分大きな N に対し次式が成り立つことを示した．[9]

$$Risk(\hat{P}^*(x^N), \theta^*) = \frac{k}{2} \log \frac{N}{2\pi e} + \log \frac{\sqrt{\det I(\theta^*)}}{w(\theta^*)} + o(1). \quad (2.19)$$

CT モデルに対しても，Gotoh らにより同様の評価が行われている．[12]

2.3.2 離散状態で状態遷移未知の状態遷移モデルに対するベイズ予測

近似アルゴリズムに関する研究

前述のように，ベイズ予測は損失関数によらず予測分布の計算が中心的な役割を果たす．そこで，離散状態で状態遷移が未知の場合の状態遷移モデルについて予測

分布を示す .

まず , 状態遷移が未知である場合の予測問題について考える . S を離散集合とし , モデルの表記は (2.12) 式の表記を用いる . 状態遷移系列 s^n が未知である場合 , 予測対象となる y_{n+1} の状態 s_{n+1} も未知であるとする . 各 i 時点 ($1 \leq i \leq n+1$) での状態 s_i は , 1 時点前の状態 s_{i-1} に依存する状態遷移確率 $v(s_i|s_{i-1})$ によって確率的に定まると仮定する . 以降この状態遷移確率は既知であるとする . ここで , 長さ n の状態遷移系列の出現確率を $v(s^n)$ とし , 状態遷移遷移確率を用いて以下で計算されるとする .

$$v(s^n) = v(s_1) \prod_{i=2}^n v(s_i|s_{i-1}). \quad (2.20)$$

ここで $v(s_1)$ は初期状態 s_1 の事前確率とし既知であるとする . また , 状態 s_{n+1} の出現確率は以下で計算される .

$$v(s_{n+1}) = \sum_{s^n \in S^n} v(s_{n+1}, s^n) = \sum_{s^n \in S^n} v(s^{n+1}). \quad (2.21)$$

y_{n+1} の発生する確率分布は , 以下のように表される .

$$p(y_{n+1}|x_{n+1}, \theta_S) = \sum_{s_{n+1} \in S} v(s_{n+1})p(y_{n+1}|x_{n+1}, \theta_{s_{n+1}}). \quad (2.22)$$

また , $w(\theta_S)$ を θ_S の事前分布として仮定すると , 予測分布は次式で表される .

$$\hat{p}^*(y_{n+1}|x_{n+1}, x^n, y^n) = \int_{\Theta_S} p(y_{n+1}|x_{n+1}, \theta_S)w(\theta_S|x^n, y^n)d\theta_S. \quad (2.23)$$

ここで , $w(\theta_S|x^n, y^n)$ は θ_S の事後分布である . s^n が与えられたもとでの θ_S の事後分布を $w_S(\theta_S|x^n, y^n, s^n)$ とすると , $w(\theta_S|x^n, y^n)$ は次式で計算される .

$$w(\theta_S|x^n, y^n) = \sum_{s^n \in S^n} v(s^n|x^n, y^n)w_S(\theta_S|x^n, y^n, s^n). \quad (2.24)$$

いま, $\Gamma_s(s^n) = \{i | s_i = s, i = 1, 2, \dots, n\}$ とする. $x^{\Gamma_s(s^n)}, y^{\Gamma_s(s^n)}$ をそれぞれ, $\Gamma_s(s^n)$ に含まれる時点 i におけるデータ x_i, y_i を全て並べたベクトルとする. すると,

$$\begin{aligned} P(y^n | x^n, \theta_S, s^n) &= \prod_{i=1}^n p(y_i | x_i, \theta_{s_i}) \\ &= \prod_{s \in S} \prod_{i \in \Gamma_s(s^n)} p(y_i | x_i, \theta_s) \\ &= \prod_{s \in S} P(y^{\Gamma_s(s^n)} | x^{\Gamma_s(s^n)}, \theta_s), \end{aligned} \quad (2.25)$$

となる. また, $w(\theta_S) = \prod_{s \in S} w(\theta_s)$ を仮定し, s^n と θ_S は独立であると仮定し, $w_S(\theta_S | s^n) = w(\theta_S)$ とすると,

$$\begin{aligned} w_S(\theta_S | x^n, y^n, s^n) &= \frac{w_S(\theta_S | s^n) P(y^n | x^n, \theta_S, s^n)}{\int_{\Theta_S} w_S(\theta_S | s^n) P(y^n | x^n, \theta_S, s^n) d\theta_S} \\ &= \prod_{s \in S} \frac{w(\theta_s) P(y^{\Gamma_s(s^n)} | x^{\Gamma_s(s^n)}, \theta_s)}{\int_{\Theta_s} w(\theta_s) P(y^{\Gamma_s(s^n)} | x^{\Gamma_s(s^n)}, \theta_s) d\theta_s} \\ &= \prod_{s \in S} w(\theta_s | x^{\Gamma_s(s^n)}, y^{\Gamma_s(s^n)}), \end{aligned} \quad (2.26)$$

となる. よって, (2.24), (2.26) 式を用いて, (2.23) 式の予測分布は以下で展開される.

$$\begin{aligned} &\hat{p}^*(y_{n+1} | x_{n+1}, x^n, y^n) \\ &= \int_{\Theta_S} p(y_{n+1} | x_{n+1}, \theta_S) \sum_{s^n \in S^n} v(s^n | x^n, y^n) \prod_{s \in S} w(\theta_s | x^{\Gamma_s(s^n)}, y^{\Gamma_s(s^n)}) d\theta_S. \\ &= \sum_{s^n \in S^n} v(s^n | x^n, y^n) \\ &\quad \times \left\{ \sum_{s_{n+1} \in S} v(s_{n+1}) \int_{\Theta_s} p(y_{n+1} | x_{n+1}, \theta_{s_{n+1}}) w(\theta_{s_{n+1}} | x^{\Gamma_{s_{n+1}}(s^n)}, y^{\Gamma_{s_{n+1}}(s^n)}) d\theta_{s_{n+1}} \right\}. \end{aligned} \quad (2.27)$$

いま, $p(y_{n+1} | x_{n+1}, \theta_s)$ が指数型の分布族である場合, 自然共役な事前分布 $w(\theta_s)$ が

存在し，上式の $\int_{\Theta_s} p(y_{n+1}|x_{n+1}, \theta_s) w(\theta_s | x^{\Gamma_s(s^n)}, y^{\Gamma_s(s^n)}) d\theta_s$ の積分部分は解析的に求められることが知られている．

以上より，離散状態で状態遷移未知の状態遷移モデルの予測分布は (2.27) 式で解析的に求められることが分かった．しかし，一般に (2.27) 式の計算には全ての $s^n \in S^n$ に対する重み付け和の計算をする必要がある．² そのため計算量は $O(|S|^n)$ となり，データ数に対し指数的に増大してしまいベイズ予測の実用は困難である．

離散状態で状態遷移未知の状態遷移モデルに対して，予測分布を効率的に計算することは難しい．そこで従来，この場合の状態遷移モデルにベイズ予測を適用する際，(2.27) 式の近似計算を用いる手法が用いられている．予測分布を近似する一般的な手法としては，MCMC (Markov Chain Monte Carlo) 法や変分ベイズ法が知られている．[23]

MCMC 法では確率的なサンプリングを行うことで，予測分布を近似する．比較的簡単にアルゴリズムを構成することができる反面，精度の高い近似をするにはサンプリング数を多くする必要があり計算量がかかる．

変分ベイズ法は，予測分布を直接的に計算するのではなく，真の分布とのダイバージェンスを最小にするような，変分分布と呼ばれる比較的計算が容易な分布を求めることで近似する．比較的計算量は少なくすむが，アルゴリズムの導出が複雑である．

漸近評価に関する研究

離散状態で状態遷移未知の状態遷移モデルに対するベイズ予測の漸近評価は，前述の状態遷移既知の場合に比べて複雑になる．状態遷移既知の場合については最尤推定量の漸近正規性を用いることで，損失やリスクの評価を行っていたが，一般に，離散状態で状態遷移未知の状態遷移モデルに対しては，最尤推定量の漸近正規性が

²状態遷移確率の確率構造によっては，重み付ける s^n の個数を削減するなど，効率的に計算することができる場合もある．

成り立たない。これは、離散状態で状態遷移未知の状態遷移モデルが特定不能性を持つからである。特定不能性とは、ある確率モデルをそれに対して過大なパラメータ数を持つモデルで表現したとき、パラメータと分布の一対一対応が成り立たないという性質を指す。このような特定不能性を有する確率モデルに対するベイズ予測については、宅味らによりある特定の条件のもとでの損失およびリスクの漸近評価が行われている。[28]

2.4 本研究の位置づけ

本研究では、第 3 章で離散状態・状態遷移未知のモデルを対象とし、近似ベイズ予測アルゴリズムを提案する。また、第 4 章、第 5 章では連続状態を持つモデルの中で i.p.i.d. 情報源モデルに注目し、第 4 章では状態遷移未知の場合の計算量を削減した効率的なベイズ予測アルゴリズムを提案、第 5 章では状態遷移既知の場合のベイズ予測についてリスクの漸近評価を行う。

第3章 離散状態・状態遷移未知の場合 の近似ベイズ予測アルゴリズム

本章では, 離散状態で状態遷移が未知の状態遷移モデルに対し, ベイズ予測を適用する際の近似アルゴリズムを提案する. そのもとで, 外れ値解析で用いられるモデルに対し, 提案アルゴリズムを適用しシミュレーションにより評価を行う.

3.1 計算量を削減した近似ベイズ予測アルゴリズムの提案

離散状態で状態遷移が未知の場合の状態遷移モデルにベイズ予測を適用する場合, 2章で示したように (2.27) 式を計算することとなる. この計算には $O(|S|^n)$ の計算量が必要であった.

そこで, 計算量を削減し近似的に (2.27) 式を求めるアルゴリズムを提案する. 和計算を減らす方法として, 重み付ける s^n の集合を, S^n からそれより小さい何らかの集合に制限するという方法が考えられる. 最も単純な方法としては, 事後確率 $v(s^n|x^n, y^n)$ が最大となる s^n のみを用いて近似する方法や, $v(s^n|x^n, y^n)$ の高い順にいくつかの s^n だけを重み付けて近似する方法が考えられる. しかし, $v(s^n|x^n, y^n)$ を求めるにも指数オーダの計算量が必要となる.

他方, (2.12) 式で表される確率モデルのパラメータ推定法として EM アルゴリズムが知られている. [21] EM アルゴリズムは尤度関数の極大値を求めるアルゴリズム

で、特に状態数の少ない場合、パラメータの推定精度が高い事が知られている。そこで、EM アルゴリズムより求められたパラメータの推定値 $\hat{\theta}_S$ を利用した近似アルゴリズムを提案する。

EM アルゴリズムより求められたパラメータの推定値 $\hat{\theta}_S$ を用いて、 $v(s_i|x_i, y_i, \hat{\theta}_S)$ という分布を計算する。これは、パラメータの推定値が与えられたもとで、 i 番目のデータが状態 $s_i \in S$ から出現した確率を表した分布になっており、良いパラメータの推定値が与えられれば、 $v(s^n|x^n, y^n)$ の高い s^n を見つける指標になると考えられる。また、 $v(s_i|x_i, y_i, \hat{\theta}_S)$ は、EM アルゴリズムの反復計算の中で計算され、簡単に求める事ができる。提案近似アルゴリズムを以下に示す。

提案近似ベイズ予測アルゴリズム

step1: EM アルゴリズムを用いて $v(s_i|x_i, y_i, \hat{\theta}_S)$ を求める。

step1-1: パラメータの初期値 $\theta_S^{(0)} = (\theta_s^{(0)} | s \in S)$ を設定する。

step1-2: 以下の E-step と M-step を $\theta_S^{(l)}$ が収束するまで繰り返す。

E-step: l 回目の反復において、次式を計算する。

$$Q(\theta_S | \theta_S^{(l-1)}) = \sum_{i=1}^n \sum_{s_i \in S} v(s_i|x_i, y_i, \theta_S^{(l-1)}) \log p(y_i|x_i, \theta_S, s_i). \quad (3.1)$$

ここで、

$$v(s_i|x_i, y_i, \theta_S^{(l-1)}) = \frac{v(s_i)p(y_i|x_i, \theta_S^{(l-1)}, s_i)}{\sum_{s_i \in S} v(s_i)p(y_i|x_i, \theta_S^{(l-1)}, s_i)}, \quad (3.2)$$

である。

M-step: $v(s_i|x_i, y_i, \theta_S^{(l-1)})$ を重みとした、重み付き最小二乗法を用いて $Q(\theta_S | \theta_S^{(l-1)})$ の最大化を行い、

$$\theta_S^{(l)} = \arg \max_{\theta_S \in \Theta_S} Q(\theta_S | \theta_S^{(l-1)}), \quad (3.3)$$

とする .

step1-3: 収束した $\theta_S^{(l)}$ を $\hat{\theta}_S$ とし , (3.2) 式から $v(s_i|x_i, y_i, \hat{\theta}_S)$ を全ての $i = 1, 2, \dots, n$ について求める .

step2: \hat{s}_i を $i = 1, 2, \dots, n$ について次式で求め ,

$$\hat{s}_i = \arg \max_{s_i \in S} v(s_i|x_i, y_i, \hat{\theta}_S) \quad (3.4)$$

確信度 $r_i = v(\hat{s}_i|x_i, y_i, \hat{\theta}_S)$ を計算する .

step3: $i = 1, 2, \dots, n$ について , r_i が何番目に小さいかを表す関数を $\eta(r_i)$ とし , $\Omega^A = \{i|\eta(r_i) \leq A, i = 1, 2, \dots, n\}$ とする . この時 , 以下の集合を求める .

$$\tilde{S}^n(A) = \{(\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_n) | \tilde{s}_i \in \tilde{S}_i(A), i = 1, 2, \dots, n\}. \quad (3.5)$$

但し ,

$$\tilde{S}_i(A) = \begin{cases} S & i \in \Omega^A, \\ \{\hat{s}_i\} & \text{otherwise,} \end{cases} \quad (3.6)$$

とする .

step4: 次式で予測分布を計算 .

$$\begin{aligned} & \tilde{p}(y_{n+1}|x_{n+1}, x^n, y^n) \\ &= \sum_{s^n \in \tilde{S}^n(A)} \tilde{v}(s^n|x^n, y^n) \\ & \times \left\{ \sum_{s_{n+1} \in S} v(s_{n+1}) \int_{\Theta_s} p(y_{n+1}|x_{n+1}, \theta_{s_{n+1}}) w(\theta_{s_{n+1}}|x^{\Gamma_{s_{n+1}}(s^n)}, y^{\Gamma_{s_{n+1}}(s^n)}) d\theta_{s_{n+1}} \right\}. \end{aligned} \quad (3.7)$$

但し,

$$\tilde{v}(\tilde{s}^n | x^n, y^n) = \frac{v(\tilde{s}^n) \int_{\Theta} P(y^n | x^n, \theta_S, \tilde{s}^n) w_S(\theta_S | \tilde{s}^n) d\theta_S}{\sum_{\tilde{s}^n \in \tilde{S}^n(A)} v(\tilde{s}^n) \int_{\Theta} P(y^n | x^n, \theta_S, \tilde{s}^n) w_S(\theta_S | \tilde{s}^n) d\theta_S}, \quad (3.8)$$

である.

上記近似アルゴリズムは $\tilde{S}^n(A)$ を求める所のみ EM アルゴリズムを用いている. 具体的には, step3 において, 確信度の低い A 個の s_i に対しては \hat{s}_i の値が誤っている可能性が高いと考え, 全ての $s_i \in S$ のパターンを重み付ける集合に入れ, それ以外の s_i に対しては \hat{s}_i の値に固定している. また, 予測値自体は (3.7) 式のように, $\hat{\theta}_S$ を使わず計算している. そのため, ベイズ最適な予測値との違いは, 重み付ける s^n の集合を S^n から $\tilde{S}^n(A)$ に制限している部分のみである.

また, $|\tilde{S}^n(A)| = |S|^A$ となるので, 近似アルゴリズムでは $|S|^A$ 個の s^n を重み付けている事になる. そのため, 仮に $A = n$ とすると, 近似アルゴリズムによる予測値とベイズ最適な予測値は一致するという特徴がある.

3.2 外れ値を含む回帰モデルへの応用

3.2.1 外れ値解析

提案近似ベイズ予測アルゴリズムの性能評価を行うため, 外れ値解析で用いられているモデルに提案アルゴリズムを適用する.

統計解析を行う際, 得られたデータの中に外れ値が含まれることが多々ある. 外れ値は転記ミスのようにデータに例外的な事象が含まれてしまう場合や, 何らかの理由で別の母集団から発生したデータが混入してしまう場合など様々な要因によって発生する. 外れ値は少量であっても解析結果に大きく影響を与えることがあるた

め、従来から外れ値を含むデータに対する解析手法が数多く研究されている。解析手法は主に、外れ値の発生に確率モデルを仮定する場合と仮定しない場合とに分けることができる。[14] 本研究では前者の外れ値の発生に確率モデルを仮定する場合を扱う。

従来、線形回帰モデルに対し混合分布を用いて外れ値の発生をモデル化する研究が行われている。Box らは、正常値の発生する分布と外れ値の発生する分布の混合分布を用いることで外れ値の発生をモデル化した。[15] 同様のモデルに対し様々な研究が行われているが、いずれの研究も外れ値の検出や混合分布のパラメータ推定を目的としている。[16, 17, 18, 19] そのため、予測を目的とした研究はあまりない。

本研究では、外れ値データの発生を含む回帰モデルに対して、外れ値の検出ではなく予測を解析の目的とし、提案した近似予測アルゴリズムを適用、アルゴリズムの評価を行う。

3.2.2 外れ値を含む回帰モデル

i 番目のデータの説明変数を $\boldsymbol{x}_i = (1, x_{i1}, \dots, x_{ip-1})^t$ 、目的変数を y_i とする。但し、 t は行列の転置を表す事とし、 $x_{ij}, y_i \in \mathcal{R}$ とする。また、 β_0, σ_0^2 をそれぞれ正常値の従う p 次元回帰係数ベクトルおよび分散パラメータ、 β_1, σ_1^2 をそれぞれ外れ値の従う p 次元回帰係数ベクトルおよび分散パラメータとする。さらに、外れ値の発生する確率を α とした時、 y_i は確率 $1 - \alpha$ で正規分布 $N(\boldsymbol{x}_i^t \beta_0, \sigma_0^2)$ に従い発生し、確率 α で正規分布 $N(\boldsymbol{x}_i^t \beta_1, \sigma_1^2)$ に従い発生すると仮定する。以降、正常値の分布のパラメータを $\boldsymbol{\theta}_0 = (\beta_0, \sigma_0^2) \in \Theta_0$ 、外れ値の分布のパラメータを $\boldsymbol{\theta}_1 = (\beta_1, \sigma_1^2) \in \Theta_1$ で表記し、全体の分布のパラメータを $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma_0^2, \sigma_1^2) \in \Theta$ と表記する。この時、 y_i

の確率分布を以下で定義する．

$$\begin{aligned}
 p(y_i|\mathbf{x}_i, \boldsymbol{\theta}) &= (1 - \alpha)p_0(y_i|\mathbf{x}_i, \boldsymbol{\theta}_0) + \alpha p_1(y_i|\mathbf{x}_i, \boldsymbol{\theta}_1) \\
 &= (1 - \alpha) \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2\sigma_0^2}(y_i - \mathbf{x}_i^t \boldsymbol{\beta}_0)^2\right\} \\
 &\quad + \alpha \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{1}{2\sigma_1^2}(y_i - \mathbf{x}_i^t \boldsymbol{\beta}_1)^2\right\}. \tag{3.9}
 \end{aligned}$$

また， $\mathbf{x}^n = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, $y^n = (y_1, y_2, \dots, y_n)$ とする．各データは独立に生起するものと仮定すると， y^n の確率分布は以下で表される．

$$p(y^n|\mathbf{x}^n, \boldsymbol{\theta}) = \prod_{i=1}^n p(y_i|\mathbf{x}_i, \boldsymbol{\theta}). \tag{3.10}$$

上記モデルは，パラメータ $\boldsymbol{\theta}$ に制約を置くことで，外れ値データの発生する構造を表現する事ができる．外れ値の発生する構造は，実際にデータをサンプリングする状況に応じて様々な場合が考えられる．Box らは， $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0$, $\sigma_1^2 = k^2\sigma_0^2$ (k は正の定数で既知)，という制約を置き，外れ値は正常値よりも分散が大きいというモデルを表現している．[15] それに対し，Abraham らは， $\sigma_1^2 = \sigma_0^2$ という制約を置くことで，平均値のみがずれるモデルを表現している．[16] 本研究では外れ値データが，正常値データとは全く別の母集団から発生した場合を考え， $\boldsymbol{\theta}_0$ と $\boldsymbol{\theta}_1$ が独立である事を仮定する．これは例えば，本来解析対象としたいデータサンプルに，何かしらの要因で別の母集団から発生したデータサンプルが混ざってしまった場合などが考えられる．また，従来研究同様，外れ値の発生する確率 α は既知，パラメータ $\boldsymbol{\theta}$ は未知として扱う．

モデルとして外れ値データの発生を含む回帰モデルを仮定した場合，予測問題として大きく分けて二つの問題設定を考える事ができる．一つは，予測対象となる y_{n+1} が $p(y_{n+1}|\mathbf{x}_{n+1}, \boldsymbol{\theta})$ に従って発生すると仮定する問題と，もう一つは， $p_0(y_{n+1}|\mathbf{x}_{n+1}, \boldsymbol{\theta}_0)$ から発生すると仮定する問題である．前者の場合，予測対象となる y_{n+1} が外れ値の分布から発生する可能性があり，外れ値も予測する必要が出てくる．そのため，多く

の場合では後者の y_{n+1} は正常値の分布からのみ出現すると仮定する事が考えられる。本研究では、どちらの問題設定においても同様の議論を行う事ができるが、より一般的な式展開を示すため、特に断りの無い限り前者の問題設定で議論を進める。予測に対する評価基準として、次式の平均二乗誤差損失を仮定する。

$$Loss(\hat{y}_{n+1}(\mathbf{x}^n, y^n), \boldsymbol{\theta}) = \int_{\mathcal{R}} (y_{n+1} - \hat{y}_{n+1}(\mathbf{x}^n, y^n))^2 p(y_{n+1} | \mathbf{x}_{n+1}, \boldsymbol{\theta}) dy_{n+1}. \quad (3.11)$$

このとき、ベイズ予測は次式で計算される。

$$\hat{y}_{n+1}^* = \int_{\mathcal{R}^n} y_{n+1} \int_{\Theta} p(y_{n+1} | \mathbf{x}_{n+1}, \boldsymbol{\theta}) w(\boldsymbol{\theta} | \mathbf{x}^n, y^n) d\boldsymbol{\theta} dy_{n+1}. \quad (3.12)$$

以上のモデルを、状態遷移モデルとして表現するために、次式で表される隠れ変数 s_i を導入する。

$$s_i = \begin{cases} 0 & y_i \text{ が正常値の分布から発生,} \\ 1 & y_i \text{ が外れ値の分布から発生.} \end{cases} \quad (3.13)$$

これは状態遷移モデルにおける状態と捉えることができる。 s_i は通常未知であるが、もし既知であった場合、 y の分布は次式で表される。

$$p(y_i | \mathbf{x}_i, \boldsymbol{\theta}, s_i) = \begin{cases} p(y_i | \mathbf{x}_i, \boldsymbol{\theta}_0) & s_i = 0, \\ p(y_i | \mathbf{x}_i, \boldsymbol{\theta}_1) & s_i = 1. \end{cases} \quad (3.14)$$

また、 $s^n = (s_1, s_2, \dots, s_n) \in S^n$ とすると、

$$\begin{aligned} P(y^n | \mathbf{x}^n, \boldsymbol{\theta}, s^n) &= \prod_{i \in \Gamma_0(s^n)} p_0(y_i | \mathbf{x}_i, \boldsymbol{\theta}_0) \prod_{j \in \Gamma_1(s^n)} p_1(y_j | \mathbf{x}_j, \boldsymbol{\theta}_1) \\ &= p_0(y^{\Gamma_0(s^n)} | \mathbf{x}^{\Gamma_0(s^n)}, \boldsymbol{\theta}_0) p_1(y^{\Gamma_1(s^n)} | \mathbf{x}^{\Gamma_1(s^n)}, \boldsymbol{\theta}_1), \end{aligned} \quad (3.15)$$

となる。但し、 $\Gamma_0(s^n) = \{i | s_i = 0, i = 1, 2, \dots, n\}$ 、 $y^{\Gamma_0(s^n)} = (y_i | i \in \Gamma_0(s^n))$ 、同様に $\Gamma_1(s^n) = \{i | s_i = 1, i = 1, 2, \dots, n\}$ 、 $y^{\Gamma_1(s^n)} = (y_i | i \in \Gamma_1(s^n))$ 、とする。このように、外れ値を含む回帰モデルは状態遷移モデルとして表記する事ができる。

s_i, s^n の事前分布をそれぞれ以下で定義する .

$$v(s_i) = \begin{cases} 1 - \alpha & s_i = 0, \\ \alpha & s_i = 1, \end{cases} \quad (3.16)$$

$$v(s^n) = (1 - \alpha)^{|\Gamma_0(s^n)|} \alpha^{|\Gamma_1(s^n)|}. \quad (3.17)$$

これはある時点の状態が , それより前の時点の状態に依存せず独立に決まる状態遷移確率と捉えることができる . また , $w_0(\theta_0), w_1(\theta_1)$ をそれぞれ , θ_0, θ_1 の事前分布とし ,

$$w(\theta) = w_0(\theta_0)w_1(\theta_1), \quad (3.18)$$

を仮定する . このとき , (3.12) 式のベイズ予測は (2.27) 式と同様の展開をすることで , 次式のように書き換える事ができる .

$$\begin{aligned} \hat{y}_{n+1}^* &= \sum_{s^n \in S^n} v(s^n | \mathbf{x}^n, y^n) \\ &\times \left\{ (1 - \alpha) \int_{\mathcal{R}^n} y_{n+1} \int_{\Theta_0} p_0(y_{n+1} | \mathbf{x}_{n+1}, \theta_0) w_0(\theta_0 | \mathbf{x}^{\Gamma_0(s^n)}, y^{\Gamma_0(s^n)}) d\theta_0 dy_{n+1} \right. \\ &\left. + \alpha \int_{\mathcal{R}^n} y_{n+1} \int_{\Theta_1} p_1(y_{n+1} | \mathbf{x}_{n+1}, \theta_1) w_1(\theta_1 | \mathbf{x}^{\Gamma_1(s^n)}, y^{\Gamma_1(s^n)}) d\theta_1 dy_{n+1} \right\}. \quad (3.19) \end{aligned}$$

今 , (3.19) 式の波括弧内第一項 , $\int_{\Theta_0} p_0(y_i | \mathbf{x}_i, \theta_0) w_0(\theta | \mathbf{x}^{\Gamma_0(s^n)}, y^{\Gamma_0(s^n)}) d\theta_0$, は通常の線形回帰モデルにおける予測分布である . いま , $w_0(\theta_0)$ に正規分布に対する自然共役事前分布 ,

$$w_0(\theta_0) \propto (\sigma_0^2)^{-\frac{\nu_0'}{2}} \exp\left\{-\frac{1}{2}[\lambda_0' + (\beta_0 - \beta_0')^t C_0'(\beta_0 - \beta_0')]\right\}, \quad (3.20)$$

を仮定すると , この予測分布は t 分布で表される . [1] 同様に第二項も t 分布となる . 以上より , s^n を用いる事でベイズ最適な予測法は , t 分布の期待値の混合を s^n の事後確率で重み付けることで計算される事がわかる . ¹

¹Box らのパラメータ制約においても , 同様にベイズ最適な予測値を解析的に求める事ができるが , 本稿では議論の煩雑を避けるため , パラメータの独立性を制約した場合のみ示す .

また同様に， y_{n+1} が正常値の分布からのみ発生すると仮定した場合でも，

$$p(y_{n+1}|\mathbf{x}_{n+1}, \boldsymbol{\theta}) = p_0(y_{n+1}|\mathbf{x}_{n+1}, \boldsymbol{\theta}_0), \quad (3.21)$$

と書き換えることで，ベイズ最適な予測値が次式のように解析的に求まる．

$$\begin{aligned} \hat{y}_{n+1}^* &= \sum_{s^n \in S^n} v(s^n|\mathbf{x}^n, y^n) \\ &\times \int_{\mathcal{R}^n} y_{n+1} \int_{\Theta_0} p_0(y_{n+1}|\mathbf{x}_{n+1}, \boldsymbol{\theta}_0) w_0(\boldsymbol{\theta}|\mathbf{x}^{\Gamma_0(s^n)}, y^{\Gamma_0(s^n)}) d\boldsymbol{\theta}_0 dy_{n+1}, \end{aligned} \quad (3.22)$$

次に，計算量を削減した近似予測アルゴリズムを適用する．

3.2.3 提案近似アルゴリズムの適用

外れ値を含む回帰モデルにおける予測にたいし，step4 を以下のように置き換える事で提案近似アルゴリズムを適用できる．

step4': 次式で予測値を計算．

$$\begin{aligned} \tilde{y}_{n+1} &= \sum_{\tilde{s}^n \in \tilde{S}^n(A)} \tilde{v}(\tilde{s}^n|\mathbf{x}^n, y^n) \\ &\times \left\{ (1 - \alpha) \int_{\mathcal{R}^n} y_{n+1} \int_{\Theta_0} p_0(y_{n+1}|\mathbf{x}_{n+1}, \boldsymbol{\theta}_0) w_0(\boldsymbol{\theta}|\mathbf{x}^{\Gamma_0(\tilde{s}^n)}, y^{\Gamma_0(\tilde{s}^n)}) d\boldsymbol{\theta}_0 dy_{n+1} \right. \\ &\left. + \alpha \int_{\mathcal{R}^n} y_{n+1} \int_{\Theta_1} p_1(y_{n+1}|\mathbf{x}_{n+1}, \boldsymbol{\theta}_1) w_1(\boldsymbol{\theta}|\mathbf{x}^{\Gamma_1(\tilde{s}^n)}, y^{\Gamma_1(\tilde{s}^n)}) d\boldsymbol{\theta}_1 dy_{n+1} \right\}. \end{aligned} \quad (3.23)$$

また, y_{n+1} が正常値の分布から発生すると仮定した場合には, (3.23) 式を次式に置き代える事で, 同様に近似アルゴリズムを適用することができる.

$$\begin{aligned} \tilde{y}_{n+1} &= \sum_{\tilde{s}^n \in \tilde{S}^n(A)} \tilde{v}(\tilde{s}^n | \mathbf{x}^n, y^n) \\ &\times \int_{\mathcal{R}^n} y_{n+1} \int_{\Theta_0} p_0(y_{n+1} | \mathbf{x}_{n+1}, \boldsymbol{\theta}_0) w_0(\boldsymbol{\theta}_0 | \mathbf{x}^{\Gamma_0(\tilde{s}^n)}, y^{\Gamma_0(\tilde{s}^n)}) d\boldsymbol{\theta}_0 dy_{n+1}. \quad (3.24) \end{aligned}$$

3.2.4 シミュレーションによる評価

近似アルゴリズムの性能をシミュレーションにより評価する.

提案アルゴリズムは, 事後確率の高いいくつかの s^n のみを重み付けするという近似と, 事後確率の高いいくつかの s^n を近似的に求めるという二段階の近似を行っている. そこで, 実験 1 では, 事後確率の高いいくつかの s^n のみを重み付けた場合の一段階の近似をした場合の近似精度を評価し, さらに事後確率の高い s^n についても近似的に求めた二段階の近似を行った場合 (提案アルゴリズム) と比較する. 実験 2 では, 実験 1 で得られた提案アルゴリズムの挙動が, データ数の多い場合でも同様に見られるかについて調べる. また, 実験 3 では提案アルゴリズムと他のベイズ予測の近似手法との予測性能を比較する.

以下, y_{n+1} が正常値の分布から発生すると仮定した場合についてシミュレーションを行う.

実験 1: s^n の重み付け数の変化による評価 (データ数が少ない場合)

データ数が少ない場合には, 事後確率 $v(s^n | \mathbf{x}^n, y^n)$ を正確に計算する事ができる. そこで, 正確な $v(s^n | \mathbf{x}^n, y^n)$ が高い順に s^n をいくつか重み付ける場合と比較する. $v(s^n | \mathbf{x}^n, y^n)$ が高い順に重み付ける s^n の数を $1, 2, \dots, 2^n$ と増やしていった場合と,

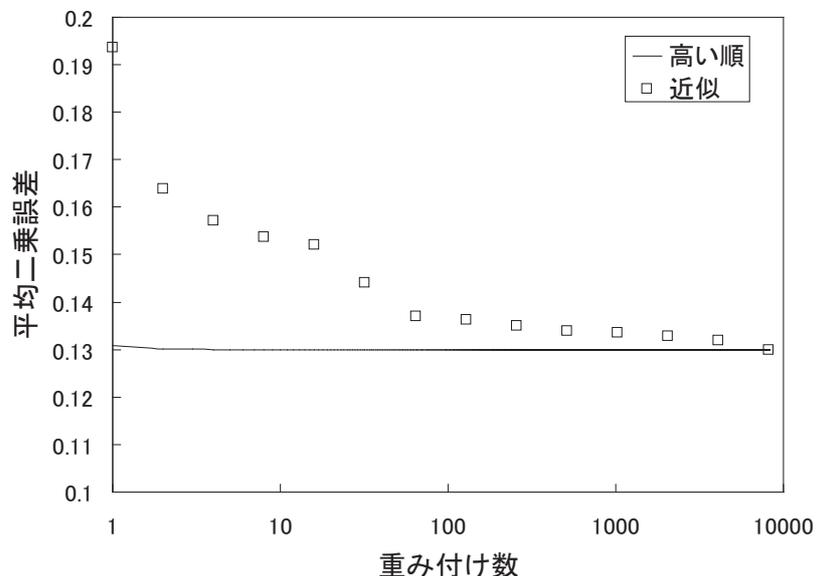


図 3.1: 実験 1: 重み付け数の変化による平均二乗誤差 ($n = 13, \alpha = 0.1$)

近似アルゴリズムにより $A = 0, 1, \dots, n$ と $\tilde{S}^n(A)$ の集合を大きくしていった場合の予測誤差について調べた。

データは人工データを用いた。データ数 $n = 13$ とし、1 回の実験で 1 組の θ を自然共役事前分布に従いランダムに発生させ、そのもとで 1 組の学習データ (x^n, y^n) と (x_{n+1}, y_{n+1}) を発生させた。各実験で予測値の二乗誤差を測り、30000 回の実験について平均を求めた。実験結果を図 3.1, 3.2 に示す。縦軸に平均二乗誤差、横軸に重み付ける s^n の個数を取りプロットした。図 3.1, 3.2 の近似アルゴリズムについてのプロットは、左から $A = 0, 1, \dots, 13$ と A の値を増やした時の予測誤差を点で示している。近似アルゴリズムと、事後確率の高い順に重み付けていった場合については、重み付け数 2^{13} の点は、ベイズ最適な予測値と等しい。

図 3.1, 3.2 の結果から正確な事後確率の高い順番で重み付けを行う場合、重み付け数が非常に少ない段階で平均二乗誤差がベイズ最適な予測とほぼ同じ値になっている。

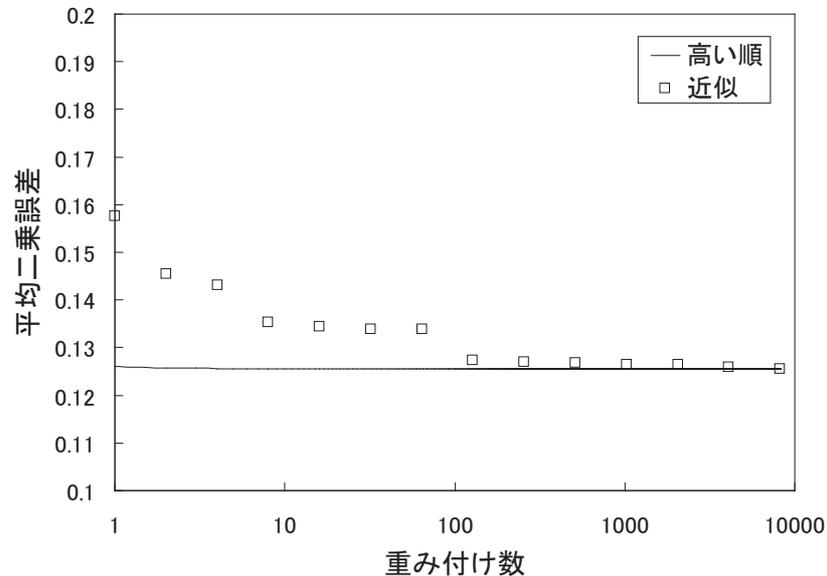


図 3.2: 実験 1: 重み付け数の変化による平均二乗誤差 ($n = 13, \alpha = 0.05$)

る．そのため，事後確率の大きい s^n を見つけることができれば，全ての s^n の重み付けをしなくてもベイズ最適に近い予測が可能であると考えられる．また，近似アルゴリズムは A がある一定の値になると平均二乗誤差が収束し，ベイズ最適な予測に近づいている．このことから，収束した時点の A を用いることができれば，計算量を削減したもとで精度の高い予測が可能であると言える．

実験 2: s^n の重み付け数の変化による評価 (データ数が多い場合)

近似アルゴリズムについて，データ数が多い場合に実験 1 と同様の傾向が見られるのかを調べた．

実験 1 と同様の方法で発生させたデータに対し，データ数 $n = 30, 45, 60$ とした時の近似アルゴリズムの予測誤差を，実験 1 と同様にプロットした．実験結果を図 3.3, 3.4 に示す．

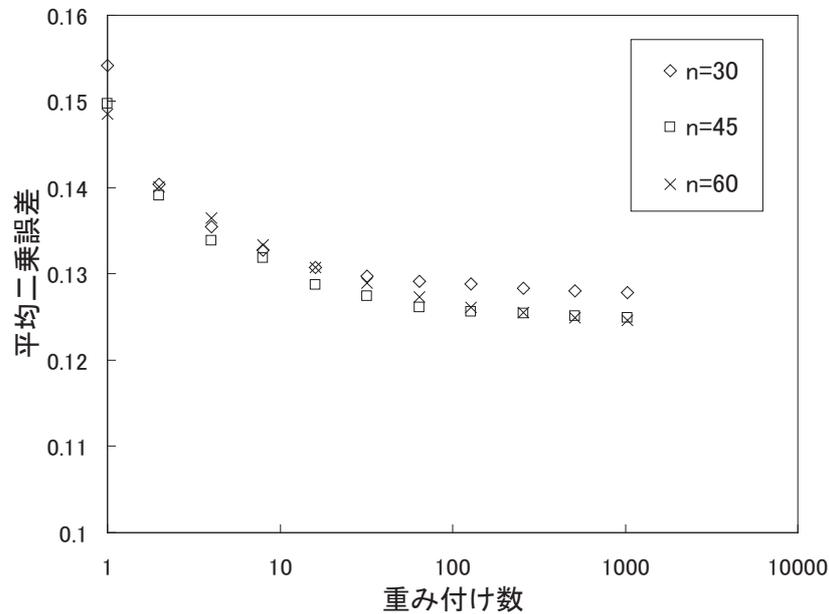


図 3.3: 実験 2: 重み付け数の変化による平均二乗誤差 ($n = 30, 45, 60, \alpha = 0.1$)

図 3.3, 3.4 の結果から、データ数の違いによらず、ほぼ同じ A で平均二乗誤差が収束している事がわかる。この事から、 A はデータ数 n によらず決める事ができ、 n が増えた場合でも、計算量を増やさずに精度の高い予測ができると考えられる。また、 α の値が小さい時の方が、 A が小さい段階で近似アルゴリズムとベイズ最適な予測の差がなくなり、近似アルゴリズムの効果が高い事が分かる。外れ値がデータに含まれる割合は、データに対して非常に小さい割合であると考えられるため、今回のような問題設定では、提案した近似アルゴリズムが有効であると言える。

実験 3：データ数の変化による評価

近似アルゴリズムについて、データ数が変化した場合の予測誤差の変化を調べた。

実験 1 と同様の方法で発生させたデータに対し、 $A = 0, 2, 6, 8, 10$ と固定したもので、縦軸に平均二乗誤差、横軸にデータ数を取りプロットした。また比較のため、外

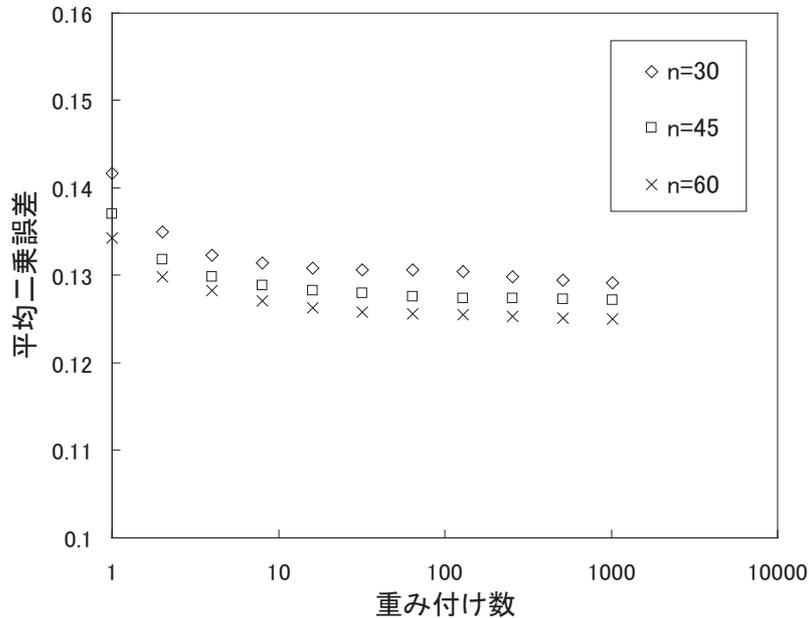


図 3.4: 実験 2: 重み付け数の変化による平均二乗誤差 ($n = 30, 45, 60, \alpha = 0.05$)

れ値が全て正確に分かっていたもとのベイズ予測，EM アルゴリズムによって求めた推定値 $\hat{\theta}$ を用いて， $\hat{y} = \mathbf{x}_{n+1}^t \hat{\beta}_0$ と予測した場合，MCMC 法の一つであるギブスサンプラー法による予測²，変分ベイズ法による予測についても実験を行った。³ 実験結果を図 3.5, 3.6 に示す。また，図 3.7 に近似アルゴリズム中に用いられる EM アルゴリズムの平均反復回数をデータ数ごとに示した。

図 3.5, 3.6 より，近似アルゴリズムは $A = 2$ 以上において，他の手法よりも予測精度が大幅に向上していることが分かる。また，ギブスサンプラー法はサンプル数が少ないため不安定な挙動を示している。

² 今回の実験では提案近似アルゴリズムにおける $A = 10$ の場合と計算量を同等にするため，ギブスサンプラー法におけるサンプリング数は 2^{10} 個で固定した。

³ ギブスサンプラー法と変分ベイズ法のシミュレーションについては，文献 [23] を参考にした。

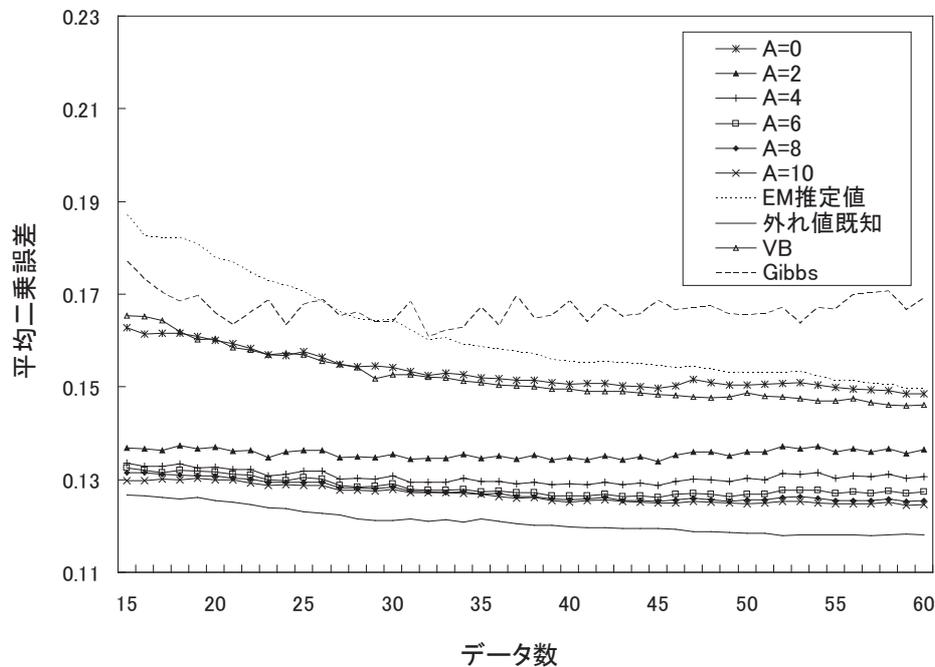


図 3.5: 実験 3 : データ数の変化による平均二乗誤差 ($\alpha = 0.1$)

3.2.5 シミュレーションに対する考察

実験の結果からも、提案近似アルゴリズムが良い近似性能を持っている事が分かったが、それ以外長所として、 $A = n$ と設定する事でベイズ最適に一致する点があげられる。変分ベイズ法の場合、いくら計算量をかけてもベイズ最適には一致しない。また、ギブスサンプラー法がベイズ最適に一致するには無限の計算量が必要となる。それに対し提案近似アルゴリズムは、計算量を多くかける事ができる場合にはベイズ最適な予測が可能であり、 A を変える事で計算量と予測精度のトレードオフを簡単に調整する事ができる。これは実用上有効な性質であると考えられる。

また、近似アルゴリズムの計算量について考察する。近似アルゴリズムは A を固定した場合、重み付ける s^n の個数は n によらず一定となる。そのため step1 以外に

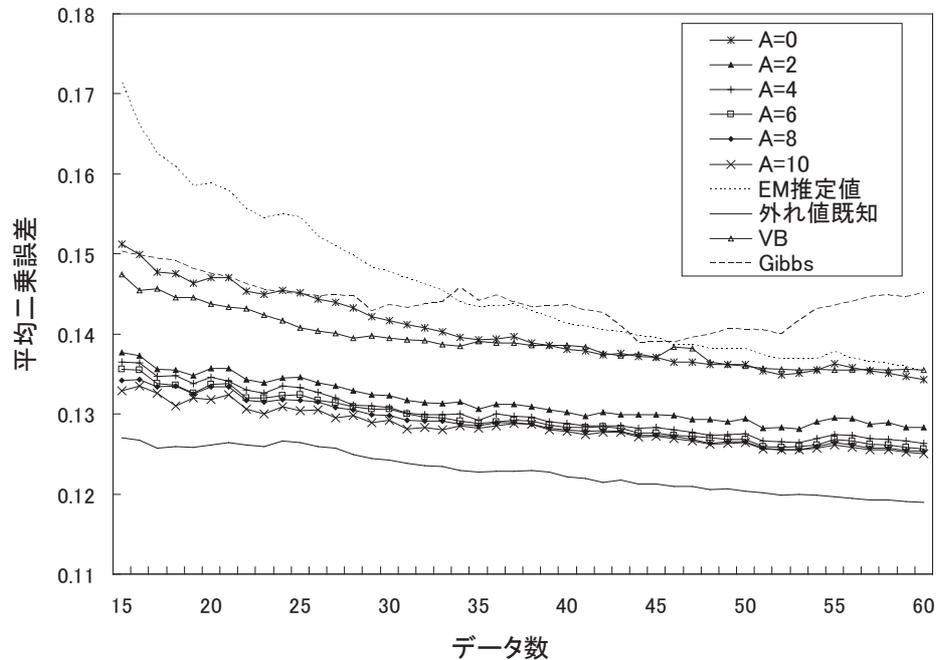


図 3.6: 実験 3 : データ数の変化による平均二乗誤差 ($\alpha = 0.05$)

かかる計算量はデータ数の線形オーダーとなる。また，step1 における EM アルゴリズムの反復回数は，図 3.7 から， α によって違いが出てくるものの，データ数に対しては，線形オーダーより少ない増え方しかしていない事がわかる。EM アルゴリズムの一回の反復にかかる計算量はデータ数の線形オーダーとなるので，近似アルゴリズム全体としてもデータ数の線形オーダー程度の計算量になると考えられる。

3.3 本章のまとめ

本章では，離散状態・状態遷移未知の状態遷移モデルに対するベイズ予測アルゴリズムを扱った。離散状態・状態遷移未知の場合，ベイズ予測はデータ数の指数オー

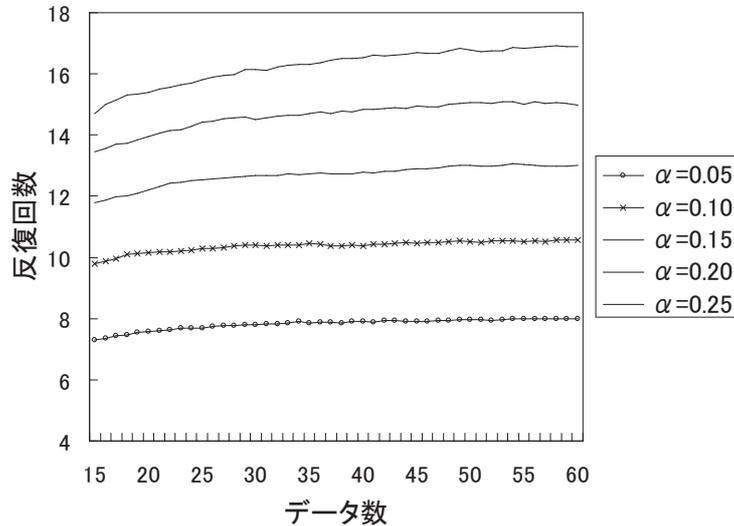


図 3.7: 実験 3: データ数の変化による EM 反復回数

ダの計算量がかかってしまうことから，計算量を削減した近似ベイズ予測アルゴリズムを提案した．提案した近似ベイズ予測アルゴリズムを，外れ値を含む回帰モデルに対し適用し，シミュレーションにより性能を評価を行った．その結果，提案アルゴリズムは計算量を大幅に削減しながらも，良い近似精度を示すことがわかった．

提案アルゴリズムが良い近似精度を示した理由としては，事後確率の高い s^n のみを重み付ける，事後確率の高い s^n を近似的に求める，という二段階の近似のそれぞれの近似精度が高かったことで挙げられる．の近似のみを行った場合の近似精度は， s^n の事後確率が偏っている場合に高くなると考えられる．つまり，事後確率の非常に大きい数個の s^n と，事後確率の非常に小さいほとんどの s^n に分かれる場合である．また，の近似精度が高くなるには，EM アルゴリズムによる推定値の精度が高い必要があると考えられる．EM アルゴリズムは状態数 $|S|$ が少ない場合のモデルに対し，推定値の精度が高くなることが知られている．つまりの近似精度は，状態数 $|S|$ が少ない場合に高くなると考えられる．以上より， s^n の事後確率

が偏りやすく，状態数 $|S|$ が少ない場合について提案アルゴリズムは有効であると
考えられる．

第4章 連続状態・状態遷移未知の場合 の効率的ベイズ予測アルゴリズム

本章では，連続状態の状態遷移モデルのひとつである，i.p.i.d. 情報源モデルに対するベイズ予測の適用について述べる．i.p.i.d. 情報源は情報源符号化で用いられる確率モデルのひとつである．そこで情報源符号化とベイズ予測の関係について述べた後，i.p.i.d. 情報源にベイズ予測を適用する際の，計算量を削減した効率化アルゴリズムを提案する．

4.1 ベイズ予測と情報源符号化

4.1.1 情報源符号化

ベイズ予測に対する理論的な研究が多くなされている代表的な分野に，情報理論における情報源符号化分野が挙げられる．情報源符号化とは情報自身もっている冗長性を除去することによってデータ系列をより短い系列に変換することを目的とした符号化のことである．情報源符号化は，圧縮したデータを復元した際，完全に元の情報と一致する符号化「無歪み情報源符号化」と，復元した際，完全には元の情報と一致しない符号化「有歪み情報源符号化」に分類される．無歪み情報源符号化は，プログラムや文書などのテキストデータの圧縮に主に用いられている．ま

た，有歪み情報源符号化は音声や画像などのデータに対する圧縮に用いられる．以下では，本研究の対象となる無歪み情報源符号化について説明する．

情報源符号化では，圧縮の対象となるデータを情報源系列と呼ぶ．この情報源系列は時系列データとして与えられる．また，情報源系列の発生に何らかの確率モデルを仮定し，これを情報源モデルと呼ぶ．情報源符号化は，ある情報源モデルから発生した情報源系列を，出来るだけ短い系列へと変換する規則（符号化と呼ぶ）を見つける問題と言える．評価基準としては，符号長（符号化後の系列の長さ）や符号長を情報源系列の出現確率で平均化した平均符号長などが考えられる．情報源の確率分布が既知である場合，ハフマン符号は平均符号長を最小にする符号として知られている．また算術符号はハフマン符号よりも符号長は若干長くなるものの，メモリ量や計算量を削減した符号として知られている．

4.1.2 ユニバーサル情報源符号化

情報源の確率分布が既知である場合にはハフマン符号を用いることで，平均符号長を最小にする符号化が可能である．しかし，実用上において，情報源の確率構造が既知であることは希である．情報源の確率分布が完全には分からない場合の問題設定をユニバーサル情報源符号化とよび，その時の符号化をユニバーサル符号と呼ぶ．従来，ユニバーサル符号については数多くの研究がなされており，様々な情報源モデルに対し符号化が提案されている．

情報源の確率分布の分布型が既知で，そのパラメータが未知である場合のユニバーサル符号を考える．今，離散集合 \mathcal{X} を情報源アルファベットと呼び， i 時点でのデータ $x_i \in \mathcal{X}$ をシンボルと呼ぶ．このとき長さ N の情報源系列を $x^N = (x_1, x_2, \dots, x_N) \in \mathcal{X}^N$ とし，情報源モデルを $P(x^N|\theta) = \prod_{n=1}^N p(x_n|x^{n-1}, \theta)$ とする． $\theta \in \Theta$ は情報源モデルの未知パラメータで，真のパラメータを θ^* と表記する．

もし，何かしらの $P(x^N|\theta^*)$ の予測値を決める事ができれば，前述のハフマン符号

等を用いることで符号化が可能となる．そのため，ユニバーサル情報源符号化の問題は，情報源系列 x^N が与えられたもとで，この $P(x^N|\theta^*)$ を推定する問題に帰着される． $P(x^N|\theta^*)$ の推定値を符号化確率と呼び， $\hat{P}(x^N)$ で表す．情報源符号化の評価基準は，以下の個別冗長度が用いられる．

$$Loss(\hat{P}(x^N), \theta^*) = \log P(x^N|\theta^*) - \log \hat{P}(x^N). \quad (4.1)$$

$-\log P(x^N|\theta^*)$ は真の確率分布を知っていたもとの符号長を表し， $-\log \hat{P}(x^N)$ は符号化確率を用いた時の符号長を表している．つまり，個別冗長度は真の確率分布を知らなかったために伸びた符号長の増分を表している．また，個別冗長度を N で割った，情報源シンボル 1 記号あたりの個別冗長度を評価基準とする場合もある．

$$Loss_N(\hat{P}(x^N), \theta^*) = \frac{1}{N} \left\{ \log P(x^N|\theta^*) - \log \hat{P}(x^N) \right\}. \quad (4.2)$$

さらに，個別冗長度を情報源系列の真の出現確率で平均化したものを平均冗長度と呼び，これを評価基準とする場合もある．

$$Risk(\hat{P}(x^N), \theta^*) = \sum_{x^N \in \mathcal{X}^n} P(x^N|\theta^*) \left\{ \log P(x^N|\theta^*) - \log \hat{P}(x^N) \right\}. \quad (4.3)$$

冗長度と同様に，平均冗長度を N で割った，情報源シンボル 1 記号あたりの平均冗長度を考える場合もある．

$$Risk_N(\hat{P}(x^N), \theta^*) = \frac{1}{N} \sum_{x^N \in \mathcal{X}^n} P(x^N|\theta^*) \left\{ \log P(x^N|\theta^*) - \log \hat{P}(x^N) \right\}. \quad (4.4)$$

従来，情報源符号化では，この 1 記号あたりの平均冗長度が漸近的に（系列長を無限にする）0 になる符号を弱ユニバーサル符号と呼び，弱ユニバーサル符号を設計する事が中心的な課題となっている．そのため，設計した符号に対し平均冗長度の漸近的な評価が行われている．また個別冗長度と平均冗長度は，それぞれ前述の対数損失と対数損失を仮定した場合のリスクに対応していることが分かる．

4.1.3 ベイズ符号

ベイズ予測を用いたユニバーサル符号のことをベイズ符号と呼び、ベイズ符号の符号化確率のことをベイズ符号化確率と呼ぶ。すなわち、ベイズ符号は (4.9) 式の個別冗長度をベイズ基準のもとで最小にする符号化確率を用いる符号である。ベイズ符号化確率は以下で与えられる。¹

$$\hat{P}^*(x^N) = \int_{\Theta} P(x^N|\theta)w(\theta)d\theta. \quad (4.5)$$

ここで、 $w(\theta)$ は θ の事前分布である。

また、上記ベイズ符号化確率は長さ N の情報源系列を一括で符号化する場合を表している。それに対し、情報源シンボル 1 記号ずつ符号化する逐次型のベイズ符号も考えられる。同様の手順から、逐次型のベイズ符号化確率は以下で導出される。

$$\hat{p}^*(x_n|x^{n-1}) = \int_{\Theta} p(x_n|x^{n-1}, \theta)w(\theta|x^{n-1})d\theta. \quad (4.6)$$

一括型の符号化確率と、逐次型の符号化確率には以下の関係が成り立つ事が知られている。

$$\hat{P}^*(x^N) = \prod_{n=1}^N \hat{p}^*(x_n|x^{n-1}). \quad (4.7)$$

すなわち、長さ N の系列を一括型のベイズ符号で符号化した場合と、逐次型のベイズ符号で符号化した場合とでは符号長が一致すると言える。

従来、情報源符号化では定常無記憶情報源 (i.i.d. 情報源) や Context tree モデルと呼ばれる情報源モデルに対し、ベイズ符号の効率的アルゴリズムや個別冗長度、平均冗長度の漸近評価が行われている。

¹1 記号あたりの個別冗長度を評価基準とした場合でも同様の符号化確率が導出される。

4.2 i.p.i.d. 情報源モデル

連続状態を持つ状態遷移モデルのひとつとして情報源符号化で用いられる, i.p.i.d. (independent piecewise-identically-distributed) 情報源モデルが挙げられる. 第 4 章, 第 5 章ではこの i.p.i.d. 情報源モデルに対するベイズ符号について扱う. i.p.i.d. 情報源は, 非定常な情報源を表現するモデルで, i.i.d. 情報源のパラメータが不連続に変化する. ある一定の区間だけを見れば, i.i.d. 情報源と見なすことができるが, 複数回のパラメータ変化が起きることで全体としては非定常な情報源とみなせる. 以下, i.p.i.d. 情報源を定義する.

情報源アルファベットを $x_i \in \mathcal{X} = \{0, 1, \dots, k\}$, 長さ N の情報源系列を $x_1^N : x_1, x_2, \dots, x_N$, また i 番目から j 番目までの部分系列を $x_i^j : x_i, x_{i+1}, \dots, x_{j-1}, x_j$ とする. 長さ N の状態遷移系列 s^N を m と表し, パラメータ変化パターンと呼ぶこととする. ここで, 本来 m は N の関数として表記されるべきだが, 表記の煩雑をさけるため単に m と記述する. i.p.i.d. 情報源はパラメータ変化パターン m と, その変化パターンのもとでのパラメータ集合 θ_m によって定まる確率分布 $P(x_1^N | \theta_m, m)$ で表される.

いま, ある変化パターン m において j 回目にパラメータの変化が起きた時点をも t_j^m とし, m における変化時点集合を, $T_m = \{t_0^m, t_1^m, \dots, t_{|T_m|-1}^m\}$, と表す. ただし便宜上, 任意の m に対し, $t_0 = 1, t_{|T_m|} = N + 1$, と定義する. 次に, 変化パターン m のもとでのパラメータベクトルを, $\theta_m = (\theta_{t_0^m}, \theta_{t_1^m}, \dots, \theta_{t_{|T_m|-1}^m}) \in \Theta_m$ とし, $p(\cdot | \theta)$ をパラメータ θ に従う i.i.d. 情報源の確率分布とする. このとき, i.p.i.d. 情報源における系列 x_1^N の出現確率を以下で定義する.

$$P(x_1^N | \theta_m, m) = \prod_{j=0}^{|T_m|-1} p(x_{t_j^m}^{t_{j+1}^m-1} | \theta_{t_j^m}). \quad (4.8)$$

以下, 表記の簡単のため特に断りがない限り, t_j^m は単に t_j と記述する. この時, $|T_m|$ は m における i.i.d. 区間の個数を表している.

パラメータ変化パターン m を，状態遷移モデルにおける遷移時点集合と考えると，i.p.i.d. 情報源は連続状態の状態遷移モデルの一種と捉えることができる．以下本章では，パラメータ変化パターン m ，およびパラメータ集合 θ_m は未知として扱う．

4.3 i.p.i.d. 情報源に対するユニバーサル符号化アルゴリズム

従来，パラメータ変化パターン m とパラメータ集合 θ_m が未知の i.p.i.d. 情報源に対して，いくつかのユニバーサル符号化アルゴリズムが提案されている．[3, 6, 7, 8]

そのなかで，ベイズ符号と関連する研究としては，[3] と [6] の研究がある．Merhav は， $|T_m|$ が定数で既知である場合について任意の符号化による符号長の下界を示し，下界を達成する符号がベイズ符号であることを示した．[3] $|T_m|$ が未知である場合については，Willems により一括型の符号化アルゴリズムが提案されている．[6] この符号化法はある特殊な m の事前分布を仮定した場合のベイズ符号とみなすことができる．

符号化確率を計算した後，算術符号等を用いて符号化することを考えると，ユニバーサル符号化では，一括型の符号化アルゴリズムより逐次型の符号化アルゴリズムの方が実用上有効である．そこで，本研究では m ， θ_m ，および $|T_m|$ が未知である i.p.i.d. 情報源に対する，逐次ベイズ符号化アルゴリズムについて考える．

4.4 i.p.i.d. 情報源モデルに対するベイズ符号

パラメータ変化パターン m とパラメータ集合 θ_m が未知の i.p.i.d. 情報源に対するベイズ符号化確率を導出する．以下，真のパラメータ変化パターンを m^* ，真のパラ

メータ集合を $\theta_{m^*}^*$ と表記する．損失関数として以下の個別冗長度を考える．

$$Loss(\hat{P}(x^N), \theta_{m^*}^*, m^*) = \log P(x_1^N | \theta_{m^*}^*, m^*) - \log \hat{P}(x^N). \quad (4.9)$$

リスク関数である平均冗長度は次式で与えられる．

$$Risk(\hat{P}(x^N), \theta_{m^*}^*, m^*) = \sum_{x^N \in \mathcal{X}^n} P(x_1^N | \theta_{m^*}^*, m^*) \left\{ \log P(x_1^N | \theta_{m^*}^*, m^*) - \log \hat{P}(x^N) \right\}. \quad (4.10)$$

パラメータ変化パターンを m の事前分布を $v(m)$, m が与えられたもとのパラメータ集合 θ_m の事前分布を $W(\theta_m | m)$ とする．このとき , ベイズリスク関数は次式で与えられる．

$$BR(\hat{P}(x^N)) = \sum_{m \in M} \int_{\Theta_m} Risk(\hat{P}(x^N), \theta_m, m) W(\theta_m | m) d\theta_m v(m). \quad (4.11)$$

このベイズリスクを最小にする符号化確率 , すなわち一括型のベイズ符号化確率は次式で与えられる．

$$\hat{P}^*(x^N) = \sum_{m \in M} \int_{\Theta_m} P(x_1^N | \theta_m, m) W(\theta_m | m) d\theta_m v(m). \quad (4.12)$$

同様の手順で , 逐次型のベイズ符号化確率は以下のように導出される．

$$\hat{p}^*(x_n | x_1^{n-1}) = \sum_{m \in M} \int_{\Theta_m} p(x_n | \theta_m, m) W(\theta_m | m, x_1^{n-1}) d\theta_m v(m | x_1^{n-1}). \quad (4.13)$$

一括型 , 逐次型いずれの場合においても , 自然共役事前分布を仮定することで , 式中の積分計算は解析的に解くことができる．しかし , 全てのパラメータ変化パターン m に対する重み付け和を計算する必要がある．ブロック長 N の情報源系列を符号化する際 , とりうる全ての m は 2^N 個存在するため , このまま (4.12) , (4.13) 式を計算するには , $O(2^N)$ の計算量およびメモリ量を必要とする．そこで , 最適性を保持したまま , 計算量 , メモリ量を削減した逐次型ベイズ符号の効率的アルゴリズムを提案する．

4.5 効率的逐次ベイズ符号化アルゴリズムの提案

4.5.1 m の事前分布

アルゴリズムの効率化に際し， m の事前分布のクラスに仮定を置く．ここで n 時点目での変化パターンの集合を M_n ，その要素を $m_n \in M_n$ と改めて表記する． n 時点目の符号化確率を計算した後， m_n の事後確率は，

$$v(m_n|x_1^n) = \frac{v(m_n|x_1^{n-1})\hat{p}_m^*(x_n|m_n, x_1^{n-1})}{\hat{p}^*(x_n|x_1^{n-1})}, \quad (4.14)$$

によって計算される．但し，

$$\hat{p}_m^*(x_n|m_n, x_1^{n-1}) = \int_{\Theta_{m_n}} p(x_n|x_1^{n-1}, \theta_{m_n}, m_n)W(\theta_{m_n}|m_n, x_1^{n-1})d\theta_{m_n}, \quad (4.15)$$

である．

$n + 1$ 時点目においてこの変化パターン m_n は， n 時点目までは m_n と同じ変化で，その後 $n + 1$ 時点目で変化が起きる m_{n+1}^a と変化が起きない m_{n+1}^b の 2 つの変化パターンに分けられる．いま，各時点で確率 α のベルヌーイ分布に従いパラメータの変化が起きる m のクラスを仮定する．この時， $v(m_{n+1}^a|x_1^n)$ および $v(m_{n+1}^b|x_1^n)$ は m_n の事後確率を用いて以下で計算することができる．

$$v(m_{n+1}^a|x_1^n) = \alpha v(m_n|x_1^n). \quad (4.16)$$

$$v(m_{n+1}^b|x_1^n) = (1 - \alpha)v(m_n|x_1^n). \quad (4.17)$$

4.5.2 効率的アルゴリズム

変化パターンが与えられたもとでのある区間におけるパラメータ $\theta_{t_c}^m$ を考える．今， $\theta_{t_c}^m$ の事前分布 $w(\theta_{t_c}^m)$ が，変化パターンおよび区間によらず全て等しいと仮定する．

次に，二種類の変化パターン m' と m'' を考える．但し， $m' \neq m''$ とする． m' ， m'' がそれぞれ，

$$m' = \{\dots, i, j, \dots\}, \quad (4.18)$$

$$m'' = \{\dots, i, j, \dots\}, \quad (4.19)$$

である．つまり，変化が起きていない同一の区間が存在するとする．さらに，変化パターン m が与えられたもとでの符号化確率を，

$$\hat{p}_m^*(x_n|x_1^{n-1}, m) = \int_{\Theta_m} P(x_n|x_1^n, \theta_m, m)W(\theta_m|m)d\theta_m, \quad (4.20)$$

とする．この時， $i \leq n \leq j - 1$ の区間において，

$$\hat{p}_m^*(x_n|x_1^{n-1}, m') = \hat{p}_m^*(x_n|x_1^{n-1}, m''), \quad (4.21)$$

が成立する．

また，時点 1 から n までの間で，最後にパラメータが変化した時点を τ_n とする．但し， $\tau_n = 1, 2, \dots, n$ である．この時， τ_n の事前分布 $V(\tau_n)$ を以下で定義する．

$$V(\tau_n) = \sum_{\{m:\tau_n \in T_m\}} v(m). \quad (4.22)$$

この時，効率的な適応的ベイズ符号化アルゴリズムを以下で示す．

【効率的アルゴリズム】

step-1.

x_n を読み込む．

step-2.

符号化確率を次式で計算する．

$$\hat{p}_{algo}(x_n|x_1^{n-1}) = \sum_{\tau_n=1}^n p_\tau(x_n|\tau_n, x_1^{n-1})V(\tau_n|x_1^{n-1}). \quad (4.23)$$

但し,

$$p_{\tau}(x_n|\tau_n, x_1^{n-1}) = \int_{\Theta} p(x_n|x_1^{n-1}, \theta_{\tau_n})w(\theta_{\tau_n}|x_1^{n-1})d\theta_{\tau_n}. \quad (4.24)$$

step-3.

$\tau_{n+1} = 1, 2, \dots, n+1$ について $V(\tau_{n+1}|x_1^n)$ を次式で計算する.

$\tau_{n+1} = 1, 2, \dots, n$ の時,

$$\begin{aligned} V(\tau_{n+1}|x_1^n) &= (1 - \alpha)V(\tau_n|x_1^n) \\ &= (1 - \alpha)\frac{p_{\tau}(x_n|\tau_n, x_1^{n-1})V(\tau_n|x_1^{n-1})}{\hat{p}_{algo}(x_n|x_1^{n-1})}. \end{aligned} \quad (4.25)$$

$\tau_{n+1} = n+1$ の時,

$$V(\tau_{n+1}|x_1^n) = \alpha. \quad (4.26)$$

step-4.

step-1 へ戻る.

□

(4.24) 式については, パラメータの事前分布にディリクレ分布を仮定し, 次式で計算する.

$$p_{\tau}(x_n|\tau_n, x_1^{n-1}) = \frac{\nu(x_n|x_{\tau_n}^{n-1}) + \beta(x_n|\tau_n)}{\sum_{a=0}^{l-1} (\nu(a|x_{\tau_n}^{n-1}) + \beta(a|\tau_n))}, \quad (4.27)$$

によって計算される. 但し, l は記号の数とし, $\nu(a|x_{\tau_n}^{n-1})$ は時点 τ_n から $n-1$ の区間において記号 a が出現した回数とする. また, $\beta(a|\tau_n)$ はベータ分布のパラメータとする.

提案アルゴリズムは, n 時点目の符号化において n 個の重み付け和計算を行うことから, n 時点での計算量およびメモリ量は $O(n)$ となる. よって, ブロック長 N 全体の符号化には, $O(N^2)$ の計算量およびメモリ量がかかる.

4.6 アルゴリズムの最適性

提案アルゴリズムが, (4.13) 式の逐次ベイズ符号化確率を正確に計算している事を示す. まず, (4.14)(4.22) 式より,

$$V(\tau_n|x_1^{n-1}) = \sum_{\{m:\tau_n \in m\}} v(m|x_1^{n-1}), \quad (4.28)$$

が成立する. また, (4.8) 式の定義より,

$$\hat{p}_m^*(x_n|x_1^{n-1}, m) = p_\tau(x_n|x_1^{n-1}, \tau_n). \quad (4.29)$$

よって, (4.28)(4.29) 式より,

$$\begin{aligned} \hat{p}^*(x_n|x_1^{n-1}) &= \sum_m \hat{p}_m^*(x_n|x_1^{n-1}, m)v(m|x_1^{n-1}) \\ &= \sum_m p_\tau(x_n|x_1^{n-1}, \tau_n)v(m|x_1^{n-1}) \\ &= \sum_{\tau_n=1}^n \left\{ p_\tau(x_n|x_1^{n-1}, \tau_n) \sum_{\{m:\tau_n \in m\}} v(m|x_1^{n-1}) \right\} \\ &= \sum_{\tau_n=1}^n p_\tau(x_n|x_1^{n-1}, \tau_n)V(\tau_n|x_1^{n-1}) \\ &= \hat{p}_{\text{algo}}(x_n|x_1^{n-1}). \end{aligned} \quad (4.30)$$

また, (4.16)(4.22) 式より, (4.25)(4.26) 式の計算により $P(m_{n+1}|x_1^n)$ が計算されていることは明らか.

以上より, 効率的アルゴリズムによる符号化確率がベイズ符号化確率と等しくなっていることが示された.

4.7 本章のまとめ

本章では、状態が連続の状態遷移モデルのひとつである i.p.i.d. 情報源モデルに対する、効率的なベイズ符号化アルゴリズムを提案した。実用上有効である逐次型のベイズ符号に対し、近似をすることなく、計算量およびメモリ量を $O(2^N)$ から $O(N^2)$ に削減したアルゴリズムを示した。その結果、計算量、メモリ量の観点から、i.p.i.d. 情報源に対しベイズ予測が実用可能であることを示した。

第5章 連続状態・状態遷移既知の場合 のベイズ予測の評価

4章に引き続き,本章ではi.p.i.d.情報源に対するベイズ予測について扱う.4章では状態遷移が未知のi.p.i.d.情報源を扱ったが,本章では状態遷移が既知である場合のi.p.i.d.情報源を考え,ベイズ予測を適用した場合の,リスク関数の漸近評価を行う.

5.1 i.p.i.d.情報源に対するユニバーサル符号の漸近評価

本章では,連続状態の状態遷移モデルにおけるベイズ予測に対するリスクの評価として,ベイズ符号の平均冗長度の漸近評価を行う.

従来,i.p.i.d.情報源に対し提案されたユニバーサル符号の評価をする際,変化パターン m のクラスに制約を置いたもとで,個別冗長度や平均冗長度の漸近評価を行っている.

パラメータの総変化回数を $|T_m|$, j 回目の変化が起きてから次の変化が起きるまでのi.i.d.区間の長さを $l_j = t_{j+1} - t_j$ とする.Merhav[3] や Willems[6] の研究では, $|T_m|$ が N に依らない定数となるクラスを仮定している.このクラスでは, $N \rightarrow \infty$ のとき, $l_j \rightarrow \infty$ となる変化パターンを考えていることになる.また,Shamir らの研究 [7, 8] では, l_j が $O(\log N)$ となる変化パターンのクラスを仮定している.このクラスでも同様に, $N \rightarrow \infty$ のとき, $l_j \rightarrow \infty$ となる変化パターンを考えていると言

える .

本研究においても変化パターンのクラスに制約を置いたもとでベイズ符号の漸近評価を行う . 具体的には $N \rightarrow \infty$ のとき , $|T_m| \rightarrow \infty$ となるクラスを考える . ただし , 従来研究では m が未知である場合の符号化法について評価を行っているが , 本章では m が既知である場合のベイズ符号に対し評価を行う .

表 5.1: 本章で制約する m のクラス

	$ T_m $ 定数	$ T_m \rightarrow \infty$	
		$l_j = O(\log N)$	
m, θ_m 未知	[3] , [6]	[7] , [8]	
m 既知, θ_m 未知		本章	

5.2 変化パターンが既知の場合のベイズ符号

系列 x_1^N に対する任意の符号化確率を $Q(x_1^N)$ とし , 符号化確率 $Q(x_1^N)$ を用いる符号を符号 Q と呼ぶ事とする . 本章では符号 Q に対する評価基準として , 以下の情報系列 1 記号あたりの個別冗長度を用いる .

$$Loss_N(\theta_m^*, m, Q, x_1^N) = \frac{1}{N} \log \frac{P(x_1^N | \theta_m^*, m)}{Q(x_1^N)}. \quad (5.1)$$

また , 1 記号あたりの平均冗長度は以下で表される .

$$Risk_N(\theta_m^*, m, Q) = \frac{1}{N} \sum_{x^N} P(x_1^N | \theta_m^*, m) \log \frac{P(x_1^N | \theta_m^*, m)}{Q(x_1^N)}. \quad (5.2)$$

以下 , 本研究ではこの平均冗長度についての評価を行っていく . また , ベイズ冗長度を以下で定義する .

$$BR_N(W(\theta_m), Q) = \int_{\Theta_m} R_N(\theta_m, m, Q) W(\theta_m) d\theta_m. \quad (5.3)$$

このベイズ冗長度を最小にする符号化確率をベイズ符号化確率呼び \hat{P}^* で表す．ベイズ符号化確率は次式で与えられる．

$$\hat{P}^*(x_1^N | m) = \int_{\Theta_m} P(x_1^N | \theta_m, m) W(\theta_m) d\theta_m. \quad (5.4)$$

5.3 変化パターンのクラスの定義

5.3.1 クラスの定義

本研究では，以下で定義する変化パターンのクラス M' について議論する．

定義 5.1 *i.p.i.d.* 情報源において， $N \rightarrow \infty$ のとき， $|T_m| \rightarrow \infty$ となり， l_j/\bar{l} が期待値 $\beta < \infty$ と分散 $\sigma^2 < \infty$ を持つ m のクラスを M' と定義する．ただし， \bar{l} は *i.i.d.* 区間の算術平均とし，以下で定義する．

$$\bar{l} = \frac{1}{|T_m|} \sum_{j=0}^{|T_m|-1} l_j = \frac{N}{|T_m|}. \quad (5.5)$$

5.3.2 ベルヌーイ分布に従いパラメータ変化が起きる場合

M' に含まれる最も単純な変化パターンの例として，確率 $0 < \alpha \leq 1$ のベルヌーイ分布に従いパラメータの変化が起きるモデルが考えられる．このとき， l_j は幾何分布に従うため， N によらない有限の値を取る．ベルヌーイ分布に従いパラメータの変化が起きる変化パターンのクラスを M'_α とすると以下の補題が成り立つ．

補題 5.1 任意の $m \in M'_\alpha$ と任意の符号 Q に対し，次式が成り立つ．

$$\lim_{N \rightarrow \infty} Risk_N(\theta_m^*, m, Q) > 0. \quad (5.6)$$

(証明) 表記の簡単のため, $x^{(j)} = x_{t_j}^{t_{j+1}-1}$ と定義する. 情報源の定義, および平均冗長度の定義より,

$$\begin{aligned} Risk_N(\theta_m^*, m, Q) &= \frac{1}{N} \sum_{\mathcal{X}^N} \prod_{j=0}^{|T_m|-1} p(x^{(j)}|\theta_{t_j}^*) \log \frac{\prod_{j=0}^{|T_m|-1} p(x^{(j)}|\theta_{t_j}^*)}{\prod_{j=0}^{|T_m|-1} Q(x^{(j)})} \\ &= \frac{1}{N} \sum_{j=0}^{|T_m|-1} \left\{ \sum_{\mathcal{X}^{l_j}} p(x^{(j)}|\theta_{t_j}^*) \log \frac{p(x^{(j)}|\theta_{t_j}^*)}{Q(x^{(j)})} \right\}, \end{aligned} \quad (5.7)$$

となる. いま, 上式括弧内を以下のように定義する.

$$r(\theta_{t_j}^*, Q) = \sum_{\mathcal{X}^{l_j}} p(x^{(j)}|\theta_{t_j}^*) \log \frac{p(x^{(j)}|\theta_{t_j}^*)}{Q(x^{(j)})}. \quad (5.8)$$

これは, i.i.d. 情報源に対するユニバーサル符号の冗長度とみなす事ができる. i.i.d. 情報源に対しては, 任意の Q , l_j と $\epsilon > 0$ に対し次式が成り立つことが知られている ([13], Theorem 1).

$$r(\theta_{t_j}^*, Q) \geq \left(\frac{1}{2} - \epsilon \right) k \log l_j. \quad (5.9)$$

また, $r(\theta_{t_j}^*, Q)$ の平均を \bar{r} とする.

$$\bar{r} = \frac{1}{|T_m|} \sum_{j=0}^{|T_m|-1} r(\theta_{t_j}^*, Q). \quad (5.10)$$

このとき, (5.9) 式より $\bar{r} > 0$ である. また,

$$\begin{aligned} Risk_N(\theta_m^*, m, Q) &= \frac{|T_m|}{N|T_m|} \sum_{j=0}^{|T_m|-1} r(\theta_{t_j}^*, Q) \\ &= \frac{|T_m|}{N} \bar{r}. \end{aligned} \quad (5.11)$$

M'_α の定義および大数の法則より,

$$\lim_{N \rightarrow \infty} \frac{|T_m|}{N} = \alpha, \quad (5.12)$$

が成り立つことから，次式が成り立つ．

$$\lim_{N \rightarrow \infty} Risk_N(\theta_m^*, m, Q) = \alpha \bar{r} > 0 \quad (5.13)$$

□

従来よりユニバーサル符号化の分野では，平均冗長度が漸近的に 0 に収束する符号（弱ユニバーサル性を有する符号）を設計する事が最大の目的の 1 つであるが，上記のように， M' には弱ユニバーサル性を有する符号が存在しない情報源のクラスも含む事が分かる．

5.4 リスクの評価

ベイズ符号の平均冗長度を評価するために，事前分布 $W(\theta_m)$ のクラスに対し，以下の仮定を置く．

仮定 5.1 θ_m における各 $\theta_t, t \in T_m$ は独立に同一の分布 $w(\theta_t)$ に従うと仮定する．

$$W(\theta_m) = \prod_{t \in T_m} w(\theta_t). \quad (5.14)$$

また，符号のクラス Q_m を以下で定義する．

定義 5.2 符号化確率 $Q(x_1^N)$ が，任意の符号化確率 $q_j(\cdot)$ を用いて，

$$Q(x_1^N) = \prod_{j=0}^{|T_m|-1} q_j(x_{t_j}^{t_{j+1}-1}), \quad (5.15)$$

と表す事ができ，

$$f_Q(\theta_{t_j}^*, l_j) = -\frac{1}{l} \sum_{x^{l_j}} p(x_{t_j}^{t_{j+1}-1} | \theta_{t_j}^*) \log q_j(x_{t_j}^{t_{j+1}-1}), \quad (5.16)$$

が期待値 $\mu_Q < \infty$ と分散 $\Sigma_Q^2 < \infty$ を持つ，符号 Q のクラスを Q_m と定義する．

このとき，次の定理が成り立つ．

定理 5.1 仮定 5.1 が成り立つもとで， $N \rightarrow \infty$ のとき，任意の $m \in M'$ および $Q \in \mathcal{Q}_m$ に対し次式が成り立つ．

$$Risk_N(\boldsymbol{\theta}_m^*, m, P_c) \leq Risk_N(\boldsymbol{\theta}_m^*, m, Q) \quad (5.17)$$

(証明) \mathcal{Q}_m に含まれる，任意の符号化確率 $Q(x_1^N)$ の平均冗長度を考える．

$$\begin{aligned} & Risk_N(\boldsymbol{\theta}_m^*, m, Q) \\ &= -\frac{1}{N} \sum_{\mathcal{X}^N} P(x_1^N | \boldsymbol{\theta}_m^*, m) \log Q(x_1^N) - \left\{ -\frac{1}{N} \sum_{\mathcal{X}^N} P(x_1^N | \boldsymbol{\theta}_m^*, m) \log P(x_1^N | \boldsymbol{\theta}_m^*, m) \right\}, \end{aligned} \quad (5.18)$$

となる．まず (5.18) 式の右辺第 2 項について考える． $l_j = t_{j+1} - t_j$ とすると，情報源の定義より，

$$\begin{aligned} & -\frac{1}{N} \sum_{\mathcal{X}^N} P(x_1^N | \boldsymbol{\theta}_m^*, m) \log P(x_1^N | \boldsymbol{\theta}_m^*, m) \\ &= -\frac{1}{N} \sum_{\mathcal{X}^N} \left\{ \prod_{j=0}^{|T_m|-1} p(x_{t_j}^{t_{j+1}-1} | \theta_{t_j}^*) \log \prod_{j=0}^{|T_m|-1} p(x_{t_j}^{t_{j+1}-1} | \theta_{t_j}^*) \right\} \\ &= \frac{1}{N} \sum_{j=0}^{|T_m|-1} \left\{ -\sum_{\mathcal{X}^{l_j}} p(x_{t_j}^{t_{j+1}-1} | \theta_{t_j}^*) \log p(x_{t_j}^{t_{j+1}-1} | \theta_{t_j}^*) \right\}. \end{aligned} \quad (5.19)$$

ここで，上式括弧内はパラメータ $\theta_{t_j}^*$ に従う i.i.d. 情報源のエントロピーである．パラメータ $\theta_{t_j}^*$ に従う i.i.d. 情報源のエントロピーを $H(X^{l_j}|\theta_{t_j}^*)$ と表記すると，

$$\begin{aligned}
 -\frac{1}{N} \sum_{\mathcal{X}^N} P(x_1^N | \theta_m^*, m) \log P(x_1^N | \theta_m^*, m) &= \frac{1}{N} \sum_{j=0}^{|T_m|-1} H(X^{l_j} | \theta_{t_j}^*) \\
 &= \frac{1}{N} \sum_{j=0}^{|T_m|-1} l_j H(X | \theta_{t_j}^*) \\
 &= \frac{|T_m|}{N|T_m|} \sum_{j=0}^{|T_m|-1} l_j H(X | \theta_{t_j}^*). \\
 &= \frac{1}{|T_m|} \sum_{j=0}^{|T_m|-1} \frac{l_j}{\bar{l}} H(X | \theta_{t_j}^*). \quad (5.20)
 \end{aligned}$$

定義 5.1，仮定 5.1 より， $N \rightarrow \infty$ のとき $|T_m| \rightarrow \infty$ ．また， l_j/\bar{l} および $H(X|\theta_{t_j}^*)$ は有限の期待値と分散を持つ確率変数の実現値であることから，大数の法則により次式が成り立つ．

$$\lim_{N \rightarrow \infty} \frac{1}{|T_m|} \sum_{j=0}^{|T_m|-1} \frac{l_j}{\bar{l}} H(X | \theta_{t_j}^*) = \int_{\Theta} \beta H(X|\theta) w(\theta) d\theta. \quad (5.21)$$

次に，(5.18) 式の右辺第 1 項について考える．情報源の定義および， Q_m の定義より，

$$\begin{aligned}
 &-\frac{1}{N} \sum_{\mathcal{X}^N} P(x_1^N | \theta_m^*, m) \log Q(x_1^N) \\
 &= -\frac{1}{N} \sum_{\mathcal{X}^N} \left\{ \prod_{j=0}^{|T_m|-1} p(x_{t_j}^{t_{j+1}-1} | \theta_{t_j}^*) \log \prod_{j=0}^{|T_m|-1} q(x_{t_j}^{t_{j+1}-1}) \right\} \\
 &= \frac{1}{N} \sum_{j=0}^{|T_m|-1} \left\{ - \sum_{\mathcal{X}^{l_j}} p(x_{t_j}^{t_{j+1}-1} | \theta_{t_j}^*) \log q(x_{t_j}^{t_{j+1}-1}) \right\} \\
 &= \frac{1}{|T_m|} \sum_{j=0}^{|T_m|-1} f_Q(\theta_{t_j}^*, l_j). \quad (5.22)
 \end{aligned}$$

ここで, $f_Q(\theta_{t_j}^*, l_j)$ は \mathcal{Q}_m の定義より有限の期待値と分散を持つ. よって大数の法則より次式が成り立つ.

$$\lim_{N \rightarrow \infty} \frac{1}{|T_m|} \sum_{j=0}^{|T_m|-1} f_Q(\theta_{t_j}^*, l_j) = \mu_Q. \quad (5.23)$$

(5.21), (5.23) 式共に, $\theta_m^* \in \Theta_m$ に依存しない値であることから, 漸近的に平均冗長度は $\theta_m^* \in \Theta_m$ に依存しなくなる事がわかる. また, ベイズ冗長度は平均冗長度を θ_m に関する期待値をとった値である. よって, 任意の符号化確率 $Q \in \mathcal{Q}_m$ に対し漸近的にベイズ冗長度と平均冗長度は一致すると言える. また, ベイズ符号は全ての N に対しベイズ冗長度を最小にする事から, 漸近的に平均冗長度も最小にする. 以上より定理は示された. \square

5.5 本章のまとめ

本章では, 連続状態の状態遷移モデルに対するベイズ予測の評価として, パラメータ変化パターン既知の場合の i.p.i.d. 情報源に対するベイズ符号の平均冗長度の評価を行った. パラメータの変化パターンのクラスに制約を置いたもとで, ベイズ符号は漸近的に平均冗長度を最小にすることを示した. その結果, このクラスの情報源に対するベイズ符号は, ベイズ基準に基づく最適性だけでなく, リスク関数の最小化という観点からも漸近的に最適性を有することを示した.

従来, 情報源符号化の分野では, リスク関数が漸近的に 0 に収束するかしないかについての議論が多くなされてきた. これは, リスク関数が漸近的に 0 に収束すれば, 必然的に漸的にリスク関数最小を示すことができるからである. しかし, 本章で扱ったモデルは, 漸的にリスク関数が 0 に収束するとは限らないモデルを扱っている. このようなリスク関数が 0 に収束しないモデルに対し, リスク関数を直接的に評価することは非常に難しい. しかし, 本章で扱ったモデルは, ある条件のも

とで、漸近的にリスク関数とベイズリスク関数が一致するという性質を持つことから、リスク関数最小を示すことができた。そのため、本章で扱ったモデルはベイズ予測が非常に効果的なモデルのひとつであると言える。

第6章 結論

6.1 まとめ

本研究では，状態遷移モデルに含まれるいくつかのモデルに対しベイズ予測を適用するために，計算量削減アルゴリズムの提案とリスクの漸近評価を行った．

第3章では，離散状態の状態遷移モデルで状態遷移が未知の場合に対し，計算量を削減した近似ベイズ予測アルゴリズムを提案した．この場合の状態遷移モデルにおけるベイズ予測は，取り得る全ての状態遷移のパターンについて事後確率の重み付け和を計算することで求められる．状態遷移のパターン数はデータ数に対し指数的に増えるため，事後確率の重み付け和の計算には，データ数に対し指数オーダーの計算量がかかり計算量的に困難であった．そこで提案アルゴリズムでは，全ての状態遷移のパターンを重み付けるのではなく，事後確率の高い状態遷移パターンを近似的にいくつか求め，それらについての重み付け和計算を行うことで計算量を削減した．また，従来外れ値解析で用いられていた外れ値を含む回帰モデルに提案アルゴリズムを適用し，シミュレーションにより評価を行った．その結果，最適なベイズ予測の計算量がデータ数 N に対し $O(2^N)$ の計算量が必要であったのに対し，提案アルゴリズムは計算量を $O(N^2)$ に削減しつつ，予測性能の劣化が少ないアルゴリズムであることを示した．また，一般的なベイズ予測の近似アルゴリズムである MCMC (Markov Chain Monte Carlo) 法や変分ベイズ法との予測性能を比較し，提案アルゴリズムは従来の近似アルゴリズムよりも予測性能が高いことを示した．

第4章と第5章では，連続状態の状態遷移モデルである，i.p.i.d. 情報源モデルを

扱った．従来データ圧縮の分野において，i.p.i.d. 情報源に対する予測アルゴリズム（圧縮アルゴリズム）が提案されていたが，ベイズ予測を適用した研究はあまりなされていなかった．そこで本論文では i.p.i.d. 情報源モデルにベイズ予測を適用した場合の効率的なアルゴリズムの提案とリスクの漸近評価を行った．

第 4 章では，i.p.i.d. 情報源モデルに対し状態遷移が未知である場合の効率的なベイズ予測アルゴリズムを示した．状態遷移が未知の i.p.i.d. 情報源に対するベイズ予測は，状態遷移の全てのパターンについての重み付け和の計算を行うことから， $O(2^N)$ の計算量が必要であった．そこで，状態遷移がベルヌーイ分布に従って起きるという仮定を置くことで，ベイズ最適性を保持したまま計算量を $O(N^2)$ に削減したアルゴリズムを提案した．その結果，状態遷移が未知の i.p.i.d. 情報源に対し，計算量とメモリ量の観点からベイズ予測が実用可能であることを示した．

第 5 章では，i.p.i.d. 情報源モデルに対し，ベイズ予測のリスクの漸近評価を行った．従来 i.p.i.d. 情報源モデルにおける予測アルゴリズムを漸近評価する場合には，クラスに一部制約をおいたモデルに対し評価が行われている．そこで本論文では，状態遷移が既知でありデータ数 N が無限になるとき，状態遷移も無限回起きるという i.p.i.d. 情報源モデルのクラスを定義した．このとき，任意の予測法に対しリスクとベイズリスクが漸近的に一致することから，ベイズ予測が漸近的にリスクを最小にすることを示した．

6.2 今後の展望

6.2.1 離散状態で状態遷移未知の状態遷移モデルに対して

また，3 章で提案したアルゴリズムは離散状態で状態遷移未知の状態遷移モデルであれば，他のモデルにも適用することができる．本論文では混合分布モデルに提案アルゴリズムを適用することで，外れ値解析に対し応用を行った．しかし，隠れ

マルコフモデルなど他のモデルに対し提案アルゴリズムを適用することで，音声認識など他の分野の問題へ応用が期待できる．これについては今後の課題のひとつとしたい．

6.2.2 i.p.i.d. 情報源モデルに対して

4章，5章で扱った i.p.i.d. 情報源モデルは，情報源符号化の立場から見ると，非定常情報源の中のある特別な情報源モデルに過ぎない．実データに対し圧縮率の高い符号化を目指すのであれば，より広いクラスの情報源モデルに対しても議論をする必要がある．情報源符号化において扱われている他の非定常情報源としては，安田らによる非定常情報源モデルがある．[29] これは毎時点パラメータの変化が起きるモデルであるが，パラメータの値は一時点前のパラメータの値を条件とした条件付確率によって決まる確率モデルで表される．i.p.i.d. 情報源はパラメータの値は変化前のパラメータと独立に決まるモデルである．i.p.i.d. 情報源と安田らのモデルを合わせることで，両者のモデルを拡張した新しい非定常モデルを考えることができる．このようなより広いクラスに対する，ベイズ符号のアルゴリズムや漸近評価，また実データでの検証等については今後の課題のひとつとしたい．

謝辞

本論文をまとめるにあたり主査として御指導頂いた，早稲田大学基幹理工学部 応用数理学科 松嶋敏泰教授に心より感謝いたします。松嶋研究室配属以来10年余り，松嶋教授には研究はもとより，教育や仕事，社会人としての姿勢など数え切れないほど多くの有益な御指導，御助言を賜りました。

また，副査として大変貴重な御時間を賜り御指導頂いた，早稲田大学創造理工学部 経営システム工学科 平澤茂一教授，早稲田大学基幹理工学部 応用数理学科 大石進一教授に深く感謝いたします。特に平澤教授には，大学院進学以来の長きに亘り，研究の細部に及ぶ適確な御指導，御助言を頂きました。

松嶋研究室の諸先輩方には，多くの御指導，御助言を頂き感謝いたします。特に，横浜商科大学 浮田善文准教授には学部時代より常に親身なアドバイスを頂きました。青山学院大学 野村亮助教には，研究に関する適切な御指導を頂きました。中央大学 吉田隆弘氏には，研究室内外の様々な場面でお世話になりました。

平澤研究室諸氏には，研究室の枠を超え多くの御助言を頂き感謝いたします。特に，湘南工科大学情報工学科 小林学准教授にはこの論文をまとめるにあたり貴重な御助言を頂きました。早稲田大学メディアネットワークセンター 石田崇助教には日頃から多くの御助言を頂きました。

また，青山学院大学 斎藤友彦助手，早稲田大学理工学研究所 小泉大城氏，早稲田大学メディアネットワークセンター 小林直人助手，早稲田大学理工学研究科博士後期課程 堀井俊佑氏，株式会社日立製作所 安井謙介氏，松嶋研究室 OB 仲川文隆氏には，研究に関する多くの議論や刺激を頂きました。本当にありがとうございます。

した。

本論文は以上をはじめとする，多くの方々の御指導，御支援の賜物です。お世話になった方々に心より御礼申し上げます。最後に，著者の研究生活に理解を示し支えてくれた，父，母，姉，義兄ら家族に心より感謝いたします。

2009年3月

参考文献

- [1] J. M. Bernardo and A. F. M. Smith, :Bayesian theory, *John Wiley&Sons*, 1994.
- [2] J.O. Berger, Statistical Decision Theory and Bayesian Analysis, *Springer*, 1993.
- [3] N. Merhav, “On the minimum description length principle for sources with piecewise constant parameters.” *IEEE Trans. Inf. Theory*, vol.39, No.6, page 1962, 1993.
- [4] F. M. J. Willems, Y. M. Shtarkov and T. J. Tjalkens, “The Context-Tree Weighting Method:Basic Properties,” *IEEE Trans. Inf. Theory*, vol.41, No.3, pp. 653–664, 1995.
- [5] Frans M. J. Willems, Y. M. Shtarkov and T. J. Tjalkens, “Context Weighting for General Finite-Context Sources,” *IEEE Trans. Inf. Theory*, vol.42, No.5, page 1514, 1996.
- [6] Frans M.J. Willems, “Coding for Binary Independent Piecewise-Identically-Distributed Source.” *IEEE Trans. Inf. Theory*, vol.42, No.6, page 2210, 1996.
- [7] G.I. Shamir and N. Merhav, “Low-Complexity Sequential Lossless Coding for Piecewise-Stationary Memoryless Sources.” *IEEE Trans. Inf. Theory*, vol.45, No.5, page 1498, 1999.
- [8] G.I. Shamir and D. J. Costello, Jr., “Asymptotically Optimal Low-Complexity Sequential Lossless Coding for Piecewise-Stationary Memoryless Sources-Part I:The Regular Case.” *IEEE Trans. Inf. Theory*, vol.46, No.7, page 2444, 2000.

-
- [9] B.S. Clark and A.R. Barron, "Information-theoretic asymptotics of Bayes methods," *IEEE Trans. Inf. Theory*, vol.IT-36, no.3, pp.453–471, May 1990.
- [10] T. Matsushima, H. Inazumi and S. Hirasawa, "A Class of Distortionless Codes Designed by Bayes Decision Theory," *IEEE Trans. Inf. Theory*, vol.37, No.5, pp. 1288–1293, 1991.
- [11] T. Matsushima and S. Hirasawa, "A Bayes coding algorithm for FSMX sources," *In Proc. Int. Symp. of Inf. Theory*, pp.388, 1995.
- [12] M. Gotoh, T. Matsushima, S.Hirasawa, "A generalization of B.S.Clark and A. R. Barron's asymptotics of Bayes codes for FSMX sources," *IEICE Trans. fundamentals*, vol.E81-A, no.10, pp.2123–2132, 1998.
- [13] J. Rissanen, "Universal Coding, Information, Prediction, and Estimation," *IEEE Trans. Inf. Theory*, vol.IT-30, no.4, pp.629–636, July 1984.
- [14] V. Barnett, *Outliers in statistical data*, John Wiley & Sons, 1994.
- [15] G. E. P. Box and G. C. Tiao, "A Bayesian approach to some outlier problems," *Biometrika*, vol.55, no.1, pp.119–129, 1968.
- [16] B. Abraham and G. E. P. BOX, "Linear Models and Scurious Observations," *Applied Statistics*, vol.27, no.2, pp.131–138, 1978.
- [17] 北川源四郎, "異常値解析ベイズモデル," *数理科学*, no.213, pp.62–66, 1981.
- [18] D. Pena and I. Guttman, "Comparing probabilistic methods for outlier detection in linear models," *Biometrika*, vol.80, no.3, pp.603–610, 1993.
- [19] J. Hoeting and A. E. Raftery and D. Madigan, "A Method for simultaneous Variable Selection and Outlier Identification in Linear Regression," *Computational Statistics and Data Analysis*, vol.22, no.3, pp.251–270, 1989.
- [20] 松嶋 敏泰, "帰納・演繹推論と予測 -決定理論による学習モデル-, " 第1回情報論的学習理論ワークショップ予稿集, pp.1–8, 1998.

- [21] F. Komaki, "Maximum likelihood from incomplete data via the EM algorithm," *J. of Royal Statistical Society Series B*, vol.39, pp.1–38, 1977.
- [22] J. Hoeting and A. E. Raftery and D. Madigan, "On asymptotic properties of predictive distributions," *Biometrika*, vol.83, no.2, pp.299–313, 1996.
- [23] 上田 修功, "ベイズ学習のアルゴリズム-高次元積分の近似手法-, " 人工知能学会誌, vol.19, no.6, pp.656–663, 2004.
- [24] S. R. Waterhouse and D. Mackay and A. J. Robinson, "Bayesian Methods for Mixtures of Experts," *Advances in Neural Information Processing Systems*, vol.8, pp.351–357, 1996.
- [25] W.K. Hastings, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, vol.57, no.1, pp.97–109, 1970.
- [26] 野村亮, 松嶋敏泰, 平澤茂一, "メモリ量を低減した近似ベイズ符号化アルゴリズム," 電子情報通信学会論文誌 (A), Vol.J86-A, No.1, pp.46-59, 2003.
- [27] 中野晶, 小林直人, 松嶋敏泰, "FSMX 情報源に対するベイズ符号のメモリ量削減アルゴリズム," 電子情報通信学会技術報告 IT2005-47, pp.47-52, 2005.
- [28] 宅味丈夫, 須子統太, 松嶋敏泰, "階層モデルにおけるベイズ予測の漸近評価に関する一考察," 第 27 回情報理論とその応用シンポジウム, pp.639-642, 2004 .
- [29] 安田豪毅, 野村亮, 松嶋敏泰, "パラメータが時間変化する情報源とその符号化に関する一考察," 電子情報通信学会技術研究報告, Vol.101, No177, IT2001-15(2001-7), pp25-30, 2001.
- [30] 有本卓, 確率・情報・エントロピー, 森北出版, 1980.
- [31] 韓太瞬, 小林欣吾, 情報と符号化の数理, 培風館, 1999.
- [32] 韓太瞬, 情報理論における情報スペクトル的方法, 培風館, 1998.
- [33] 平澤茂一, 情報理論, 培風館, 1996.

-
- [34] 今井秀樹, 情報理論, 昭晃堂, 1984.
- [35] T. M. Cover and J. A. Thomas, Elements of information theory, Wiley, 1991.
- [36] 須子統太, 松嶋敏泰, 平澤茂一, “外れ値データの発生を含む回帰モデルに対するベイズ予測アルゴリズム,” 情報処理学会論文誌数理モデル化と応用, Vol.1, No.1, pp.17-26, 2008.
- [37] T. Suko, T. Matsushima and Shigeichi Hirasawa “Asymptotic Property of Universal Lossless Coding for Independent Piecewise Identically Distributed Sources,” Pre-ICM International Convention on Mathematical Sciences, 2008.
- [38] 須子統太, 松嶋敏泰, 平澤茂一, “区間で一定なパラメータを持つ情報源におけるベイズ符号化法について,” 第 26 回情報理論とその応用シンポジウム予稿集, pp.165-168, 2003.
- [39] 須子統太, 松嶋敏泰, 平澤茂一, “BW 変換を用いたユニバーサル符号化アルゴリズムに関する研究,” 第 28 回情報理論とその応用シンポジウム予稿集, pp.343-346, 2005.
- [40] 須子統太, 松嶋敏泰, 平澤茂一, “区間で一定なパラメータを持つ非定常情報源の漸近的な性質について,” 第 31 回情報理論とその応用シンポジウム予稿集, pp.815-818, 2008.

研究業績

種類別	題名, 発表掲載誌名, 発表発行年月, 著者
1. 論文	外れ値データの発生を含む回帰モデルに対するベイズ予測アルゴリズム 情報処理学会論文誌数理モデル化と応用, Vol.1, No.1, pp.17-26 (2008-10) 須子統太, 松嶋敏泰, 平澤茂一
2. 論文	Asymptotic Property of Universal Lossless Coding for Independent Piece- wise Identically Distributed Sources Pre-ICM International Convention on Mathematical Sciences, (2008-12) Tota Suko, Toshiyasu Matsushima and Shigeichi Hirasawa
3. 論文	拡張された有本 Blahut アルゴリズムの大域的収束性について 電子情報通信学会論文誌 (A), J91-A No.9, pp.846-860 (2008-9) 安井謙介, 須子統太, 松嶋敏泰
4. 論文	An Algorithm for Computing the Secrecy Capacity of Broadcast Chan- nels with Confidential Messages Proceedings of IEEE International Symposium on Information Theory, pp.936-940, (2007-6) Kensuke Yasui, Tota Suko, and Toshiyasu Matsushima
5. 論文	Multiuser Detection Algorithm for CDMA Based on the Belief Propaga- tion Algorithm Proceedings of IEEE 10th Int. Symp. on Spread Spectrum Technical and Applications(ISSSTA'08), pp.194-199 Shunsuke Horii, Tota Suko, Toshiyasu Matsushima and Shigeichi Hira- sawa

種類別	題名, 発表掲載誌名, 発表発行年月, 著者
6. 講演	区間で一定なパラメータを持つ非定常情報源の漸近的な性質について 第31回情報理論とその応用シンポジウム予稿集, pp.815-818, (2008-10) 須子統太, 松嶋敏泰, 平澤茂一
7. 講演	外れ値データの発生を含む回帰モデルに対するベイズ予測アルゴリズム 情報処理学会研究報告 2007-MPS-67, pp.13-16, (2007-12) 須子統太, 松嶋敏泰, 平澤茂一
8. 講演	BW変換を用いたユニバーサル符号化アルゴリズムに関する研究 第28回情報理論とその応用シンポジウム予稿集, pp.343-346, (2005-11) 須子統太, 松嶋敏泰, 平澤茂一
9. 講演	外れ値データの発生を考慮にいた回帰モデルにおけるベイズ予測法について 2004年情報論的学習理論ワークショップ (IBIS2004) 予稿集, pp.34-39, (2004-11) 須子統太, 仲川文隆, 松嶋敏泰
10. 講演	区間で一定なパラメータを持つ非定常情報源におけるベイズ符号の冗長度について 第27回情報理論とその応用シンポジウム予稿集, pp.523-526, (2004-12) 須子統太, 松嶋敏泰, 平澤茂一
11. 講演	区間で定常なパラメータを持つ非定常情報源におけるベイズ符号の冗長度について 電子情報通信学会技術報告 IT2004-22, pp.23-28, (2004-7) 須子統太, 松嶋敏泰, 平澤茂一
12. 講演	区間で一定なパラメータを持つ情報源におけるベイズ符号化法について 第26回情報理論とその応用シンポジウム予稿集, pp.165-168, (2003-12) 須子統太, 松嶋敏泰, 平澤茂一
13. 講演	決定木モデルにおける予測アルゴリズムについて 電子情報通信学会技術報告 COMP2003-36, pp.93-98, (2003-8) 須子統太, 野村亮, 松嶋敏泰, 平澤茂一

種類別	題名, 発表掲載誌名, 発表発行年月, 著者
14. 講演	拡張された階層モデルにおける予測アルゴリズムについて 第25回情報理論とその応用シンポジウム予稿集, pp.755-758, (2002-12) 須子統太, 野村亮, 松嶋敏泰
15. 講演	マーキング仮定に基づくフィンガープリンティング符号のキャパシティについて 暗号と情報セキュリティシンポジウム, 2008 (SCIS 2008) (2008-01) 柴田大介, 須子統太, 松嶋敏泰
16. 講演	A Note on Multiuser Detection Algorithms for CDMA based on the Belief Propagation Algorithm 電子情報通信学会技術報告 IT2007-26, pp.7-12, (2008-1) S. Horii, T. Suko, T. Matsushima, and S. Hirasawa
17. 講演	Multiuser Detection Algorithms for CDMA based on the Message Passing Algorithms 電子情報通信学会技術研究報告 IT2006-16, pp.17-22, (2006-5) Shunsuke Horii, Tota Suko, and Toshiyasu Matsushima
18. 講演	秘密情報を持つ Broadcast Channel の Secrecy Capacity 計算アルゴリズム 第29回情報理論とその応用シンポジウム予稿集, (2006-11) 安井謙介, 須子統太, 松嶋敏泰
19. 講演	電子透かしにおける秘匿容量計算アルゴリズム 電子情報通信学会技術報告 IT2005-94, pp.177-182, (2006-3) 安井謙介, 須子統太, 松嶋敏泰
20. 講演	使用ユーザが変化する DS/CDMA システムにおけるベイズ最適なマルチユーザ検出について 第28回情報理論とその応用シンポジウム予稿集, pp.781-784, (2005-11) 堀井俊佑, 須子統太, 松嶋敏泰
21. 講演	電子透かしにおける秘匿容量の計算手法に関する研究 電子情報通信学会技術報告 IT2005-47, pp.29-34, (2005-7) 安井謙介, 須子統太, 松嶋敏泰

種類別	題名, 発表掲載誌名, 発表発行年月, 著者
22. 講演	階層モデルにおけるベイズ予測の漸近評価に関する一考察 第27回情報理論とその応用シンポジウム, pp.639-642, (2004-12) 宅味丈夫, 須子統太, 松嶋敏泰
23. 講演	ベイズ決定理論に基づく予測における近似手法について 第26回情報理論とその応用シンポジウム予稿集, pp.703-706, (2003-12) 江口公盛, 須子統太, 松嶋敏泰
24. 講演	相関のある時系列の状態空間によるモデル化と予測 電子情報通信学会技術報告 IT2003-38, pp.87-92, (2003-7) 鈴木悠哉, 須子統太, 松嶋敏泰