

博士学位論文

階層型の確率モデル族に基づく
ユニバーサル情報源符号化に関する研究

A Study on Universal Source Coding
Based on Hierarchical Probabilistic Model Class

2016年7月

宮 希望

Nozomi MIYA

博士学位論文

階層型の確率モデル族に基づく
ユニバーサル情報源符号化に関する研究

A Study on Universal Source Coding
Based on Hierarchical Probabilistic Model Class

2016年7月

早稲田大学大学院 基幹理工学研究科
数学応用数理専攻 情報理論研究

宮 希望

Nozomi MIYA

目次

第 1 章	序論	1
1.1	はじめに	1
1.2	本論文の目的	7
1.3	本論文の構成	8
第 2 章	問題設定および従来研究	9
2.1	はじめに	9
2.2	無歪み可変長情報源符号化	9
2.2.1	可変長情報源符号化の数理モデル	9
2.2.2	無歪み符号化とクラフトの不等式	10
2.2.3	平均符号語長とシャノンの情報源符号化定理	11
2.2.4	平均符号語長の理論的限界を達成する符号の構成	12
2.3	ユニバーサル情報源符号化	13
2.3.1	実用性と理論的明瞭性のトレードオフ関係	13
2.3.2	実用的なユニバーサル符号の代表例：ブロックソート法	14
2.3.3	確率モデルの定義	15
2.3.4	モデル未知の確率モデルの例 1：FSMX 情報源	17
2.3.5	モデル未知の確率モデルの例 2：区分定常情報源	18
2.3.6	評価関数の導入と非逐次型および逐次型の符号化	19
2.4	ベイズ符号に関する従来研究	21
2.4.1	逐次型の符号化と統計的予測問題との関連	21
2.4.2	ベイズ符号	22
2.4.3	ベイズ基準を導入するモチベーション	25
2.4.4	モデル既知の確率モデルに基づくベイズ符号の性能解析	26

2.4.5	モデル未知の確率モデルに基づくベイズ符号の性能解析	27
2.4.6	FSMX 情報源に対する効率的なベイズ符号化アルゴリズム	28
2.4.7	区分定常情報源に対する効率的なベイズ符号化アルゴリズム	30
第 3 章	表現不可能な確率モデルに基づくベイズ符号の性能解析	35
3.1	はじめに	35
3.2	確率モデルの表現可能性	35
3.3	本章の位置づけ	40
3.4	モデル既知の表現不可能な確率モデルに対する解析	40
3.5	階層型のモデル未知の表現不可能な確率モデルに対する解析	42
3.5.1	対数損失の漸近評価	42
3.5.2	カルバックライブラー情報量の漸近評価	43
3.5.3	考察	45
3.5.4	ベイズ符号の有効性を示す具体例	46
3.6	補題および定理の証明	48
3.6.1	補題 3.4.2 の証明	48
3.6.2	補題 3.4.3 の証明	49
3.6.3	補題 3.5.1 の証明	49
3.6.4	定理 3.5.1 の証明	52
3.6.5	系 3.5.1 の証明	52
第 4 章	BW 変換に基づくベイズ符号化アルゴリズム	54
4.1	はじめに	54
4.2	マルコフ情報源に対する BW 変換の性質	54
4.3	BW 変換に基づくベイズ符号化アルゴリズム	55
第 5 章	結論	59
5.1	本論文の成果	59
	謝辞	61
	参考文献	63

表 目 次

3.1 確率モデルの種類と表現可能性による分類	40
-----------------------------------	----

目 次

2.1 可変長情報源符号化の数理モデル	10
2.2 2元FSMX情報源モデル	18

第1章 序論

1.1 はじめに

現代の情報化社会において、デジタルデータの圧縮や機械学習に関する研究はますますその重要性を増しており、その基礎理論については情報理論の一分野である情報源符号化に関する問題として近年盛んに研究が行われている。情報源符号化の主な目的はなるべく効率よく情報を伝送する事であるが、Shannon が 1948 年に発表した論文に端を発するこの情報源符号化の問題は、次のように数理モデル化される [20]。まず、デジタルデータ等の情報は、ある確率構造を仮定した情報源から情報源系列として出力される。次に、出力された情報源系列は符号器により符号語と呼ばれるまた別の異なる系列に変換、すなわち符号化された後に伝送あるいは蓄積される。そして、伝送あるいは蓄積された符号語は復号器により、元の情報源系列に復号される。ただし、復号の際に誤りが生じ、元の情報源系列に正しく復号できない可能性がある。情報源符号化の代表的な問題設定として、符号語を元の情報源系列に誤りなく復号するといった事が考えられる。誤りがないというのは、符号語を復号した系列が元の情報源系列と異なる確率が、情報源系列の長さを十分大きくした場合に 0 に収束していくかあるいは任意の情報源系列の長さに対して 0 であるかのいずれかを指す場合が多い。特に後者は無歪み情報源符号化と呼ばれる。

無歪み情報源符号化において可変長、すなわち符号語長が情報源系列毎に異なる事を許容した場合、情報の伝送および蓄積の効率性を測る代表的な評価基準として平均符号語長が挙げられる。平均符号語長とは各情報源系列に対する符号語長の情報源の確率分布に関する期待値である。一方、任意の無歪み符号は、各情報源系列に対して定められる符号語長に関し、クラフトの不等式¹を満たさなければならない

¹情報源系列 x^n が離散有限集合 \mathcal{X}^n 上に値をとり、符号化関数 ϕ_n に対し、 x^n の符号語長を $|\phi_n(x^n)|$ と表した場合、クラフトの不等式は $\sum_{x^n \in \mathcal{X}^n} |\mathcal{X}|^{-|\phi_n(x^n)|} \leq 1$ である。ただし、 $|\mathcal{X}|$ は \mathcal{X} の要素数を

事が知られている（例えば，[8] 等）．従来，情報源に定常エルゴードな確率分布を仮定した場合，シャノンの情報源符号化定理 [20] として，任意の無歪み符号に対し，情報源 1 シンボルあたりの平均符号語長の下限が漸近的にエントロピーレートで与えられる事が知られ，さらに，情報源 1 シンボルあたりの平均符号語長が漸近的にエントロピーレートに等しくなる，ある無歪み符号が存在する事が知られている．ここで，エントロピーレートとは情報源の確率分布から一意に定まる量である．平均符号語長を評価基準とする無歪み情報源符号化においては，シャノンの情報源符号化定理により，平均符号語長の理論的限界が与えられているため，クラフトの不等式を満たし，かつ，平均符号語長がその理論的限界に到達する符号を構成する事が主な目的となる．実際，情報源の確率構造を利用する事により，情報源 1 シンボルあたりの平均符号語長がエントロピーレートに漸近する実用的な符号化アルゴリズムが提案されている．その代表的な例としてハフマン符号や算術符号等が挙げられる．ハフマン符号は独立定常な確率分布を持つ情報源に対し，情報源 1 シンボルを符号化した場合に平均符号語長が理論的限界値であるエントロピー²に達する事が知られている [13]．さらに，ハフマン符号を情報源系列の長さに応じて拡大して適用する事により，情報源 1 シンボルあたりの平均符号語長は漸近的に，すなわち情報源系列の長さを十分大きくした場合にエントロピーレートに到達するが，符号化における計算量は情報源系列の長さに応じて指数的に大きくなってしまふ．一方，算術符号は，情報源 1 シンボル毎ではなく情報源系列全体を一度に符号化する，より計算量の削減された実用的な符号化アルゴリズムである [18, 19]．いずれの符号においてもその符号語長は情報源の確率分布により定められ，情報源系列 x^n が生起する確率を $p(x^n)$ とした場合，ハフマン符号であれば情報源 1 シンボルあたりの符号語長は $\lceil -(1/n) \log p(x^n) + 1/n \rceil$ 以下で与えられ，算術符号であれば情報源系列 x^n に対する符号語長は $-\log p(x^n) + 2$ により定められる．ここで， $-\log p(x^n)$ は理想符号語長と呼ばれる．

一方，ユニバーサル情報源符号化と呼ばれる情報源の確率構造の一部が未知である情報源符号化の問題は実用的に重要な研究となっている．ユニバーサル情報源符号化の一分野においては，定常エルゴードな確率分布を持つ情報源に対し，無歪み

表す．

²独立定常な確率分布に対するエントロピーレートはエントロピーに帰着される．

の下，情報源 1 シンボルあたりの平均符号語長が情報源のエントロピーレートに漸近するような符号を構成する事が重要なテーマとなっている．例えば，テキストデータの圧縮等に応用されている LZ 符号 [32, 33] やネットワーク上でやり取りされるファイルの圧縮等にも利用される bzip2 に応用されているブロックソート法 [4] は，情報源に定常エルゴードな確率分布を仮定した場合，いずれも実験的には情報源系列の長さを十分大きくすれば平均符号語長がエントロピーレートに収束する事が知られている³．

ユニバーサル情報源符号化における代表的な問題設定として，情報源の確率構造に対し，例えば，情報源系列の出現確率が 1 次マルコフ過程と呼ばれる直前の値のみに依存する確率モデルに従う事は既知であるが遷移確率の確率パラメータは未知であるという様な仮定を置く事が挙げられる [6, 15, 24, 25, 28, 30]．さらに，出現確率が直前の長さ k の系列に依存する k 次マルコフ過程である事，すなわち確率モデルはマルコフ過程であると仮定されているが次数 k が未知，つまり何次のマルコフ過程かは未知であるという問題設定等も考えられている [3, 10, 15, 16, 21, 22]．本論文では，前者をモデル既知の確率モデル，後者をモデル未知の確率モデルとそれぞれ呼ぶ事にする．先に述べたハフマン符号や算術符号は，その符号語長が情報源の確率分布によって定まるものであった．つまり，符号語長を与える確率分布である符号化確率を何らかの形で決定できれば算術符号等を用いた符号化が可能であるという事である．したがって，確率モデルの仮定に基づくユニバーサル情報源符号化に関する一部の研究においては，符号化の良し悪しを測る基準である評価関数を導入した下でその評価関数を最適化する符号化確率を決定する事が主な目的とされている [3, 6, 10, 15, 16, 21, 22, 24, 25]．符号化確率の決定に関し，時点 1 から時点 n までに発生した情報源系列 x^n を得た下で，一括して x^n に対する符号化確率 $A_p(x^n)$ を決定する非逐次型の符号化 [28, 30] と各時点 t においてそれまでに発生した情報源系列 x^{t-1} を得た下で，次に発生する情報源シンボル x_t に対する符号化確率 $A_p(x_t | x^{t-1})$ を逐次的に決定し，それを時点 n まで繰り返す逐次型の符号

³LZ 符号に関しては情報源に定常エルゴードな確率分布を仮定した場合，平均符号語長がエントロピーレートに収束する事が理論的にも証明されているが [31]，その収束オーダーまでは明らかにされていない．ブロックソート法に関しては情報源の確率分布が有限次数の定常マルコフ過程により表される場合，平均符号語長がエントロピーレートにより上界される事が理論的にも証明されているが [2]，その収束オーダーまでは明らかにされていない．

化 [16, 21, 22] の二つが挙げられる．実際に計算機上で符号化確率を計算する場合，非逐次型の符号化では時点 n までの確率の積を計算する事になるため，情報源系列の長さが大きくなった場合，符号化確率の値が 0 に非常に近い値となってしまう，計算機上では 0 と扱われてしまう．一方，逐次型の符号化の場合，各時点 t において計算した符号化確率を用いて符号語長を定めるため，そのような問題は発生しない．また，平均符号語長の解析という点においては，両者は等価である．ただし，先に述べたように，任意の無歪みな符号はクラフトの不等式を満たさなければならないが，算術符号等のように符号化確率により符号語長を定める場合，符号語長は非逐次型の符号化であれば $-\log Ap(x^n)$ の形で与えられ，逐次型の符号化であれば各時点 t において $-\log Ap(x_t | x^{t-1})$ の形で与えられるため，クラフトの不等式は，何らかの手法により決定される符号化確率が劣確率分布である事を要求するものであると言い換えられる⁴．逐次型の符号化においては各時点 t における情報源の確率分布 $p(x_t | x^{t-1})$ と符号化確率 $Ap(x_t | x^{t-1})$ を用いて表される $\log(p(x_t | x^{t-1})/Ap(x_t | x^{t-1}))$ を時点 1 から時点 n まで足し合わせた $\log(p(x^n)/Ap(x^n))$ を評価関数として導入するのが一般的である⁵ [6, 10, 15, 16, 21, 22, 28, 30]．ここで， $\log(p(x^n)/Ap(x^n))$ は，符号化確率 $Ap(x^n)$ を用いた場合の符号語長と理想符号語長との差分を表す．この評価関数 $\log(p(x^n)/Ap(x^n))$ の情報源の確率分布に関する期待値は冗長度と呼ばれ，これは平均符号語長と情報源のエントロピーレートとの差分，すなわち相対エントロピーを表す．

一方，統計的予測問題における一部の研究 [11, 12] では，時系列的に観測されるデータに基づき，次の時点に発生するデータが何であるかを予測する，あるいはデータが従う確率分布そのものを推定する問題が考えられている．特に後者については，各時点 t におけるデータが従う確率分布 $p(x_t | x^{t-1})$ の推定に対し，推定の良し悪しを測る何らかの損失関数を導入し，その損失関数を最小にする決定関数 $Ap(x_t | x^{t-1})$ を求めるといった事が主に行われている．損失関数の一例として $\log(p(x_t | x^{t-1})/Ap(x_t | x^{t-1}))$ を時点 1 から時点 n まで足し合わせた対数損失 $\log(p(x^n)/Ap(x^n))$ が挙げられる．これは，ユニバーサル情報源符号化における符号化確率 $Ap(x^n)$ を用いた場合の符号語

⁴符号語長を $-\log Ap(x^n)$ で与えた場合，クラフトの不等式は $\sum_{x^n \in \mathcal{X}^n} Ap(x^n) \leq 1$ となる．

⁵一方，非逐次型の符号化においても同様に時点 n までの情報源の確率分布および符号化確率を用いた $\log(p(x^n)/Ap(x^n))$ が評価関数として導入される事が多い．

長と理想符号語長 $-\log p(x^n)$ との差分を表す. さらに, 対数損失 $\log(p(x^n)/Ap(x^n))$ のデータが従う確率分布に関する期待値はカルバックライブラー情報量と呼ばれ, これは先に述べた逐次型の符号化における冗長度と等価であるとみなせる. つまり, 冗長度すなわち平均符号語長を評価基準としたユニバーサル情報源符号化の問題において平均符号語長の最小化を図る事は, 統計的予測問題の中でも対数損失を導入した確率分布の推定問題においてカルバックライブラー情報量の最小化を図る事に等しいと言える [11, 12].

統計的予測問題の中でも確率分布を推定する問題においては, データの確率分布に対し, 先に述べたような確率モデルを導入し, ミニマックス基準やベイズ基準に基づいた推定を行う事が考えられているが [11], ベイズ基準に基づく推定をユニバーサル情報源符号化に応用したベイズ符号は, 任意の長さの情報源系列に対し, 平均符号語長をベイズ基準の下で最小にする符号であり [15], 理論的にも重要な符号である. ベイズ符号では確率モデルのパラメータに事前分布が導入されるが, 情報源符号化においては, 例えば圧縮対象のテキストの種類に応じて, ある単語の出現頻度がどの程度であり, さらにその頻度がそれまでの文脈にどの程度依存して変わってくるか等が事前に分かっている場合が多いため, パラメータに事前分布を仮定する事は比較的自然而である. また, ミニマックス基準に基づく推定を応用した符号も一部では研究されているが [3, 24, 25], ベイズ符号においてジェフリーズの事前分布と呼ばれるあるクラスの事前分布を仮定した場合, ベイズ符号は漸近的にそのような符号と等価になる事が知られている [7]. 特にベイズ符号の場合, その性能を理論的に容易かつ精密に解析する事が可能であり, 実際, 平均符号語長がエントロピーレートに漸近する事が明らかにされているが, その漸近式は $o(1)$ まで明らかにされており, 収束オーダーも明確である [6, 10]. 特にマルコフ過程のように高次のモデルが低次のモデルを含む入れ子構造を持つ階層型のモデル未知の確率モデルで表される情報源に対し, ベイズ符号の平均符号語長が精密に解析されており, ベイズ符号の有効性が理論的に証明されている [10]. 一方, 先に述べた, 確率モデルはマルコフ過程であると仮定されているが次数 k が未知, つまり何次のマルコフ過程かは未知であるものの一種である FSMX 情報源に対し, 効率的なベイズ符号化アルゴリズムが提案されている [16]. また, モデル未知の確率モデルで表される情報源の一種である区分定常情報源に対しても効率的なベイズ符号化アルゴリズムが提案され

ている [21]. 区分定常情報源とは, ある定常分布に従って情報源系列を発生させている情報源の確率パラメータがある時点で突然に変化し, その後またある時点まではその変化したパラメータを持つ定常分布に従って系列を発生させ, またパラメータが変化するという情報源である [29]⁶.

ユニバーサル符号化の問題においては先に述べた通り, 未知の情報源の確率分布に対し, 平均符号語長がエントロピーレートに収束する符号を構成する事が主な目的とされているが, これは, モデル既知あるいはモデル未知の確率モデルを仮定した上で, 情報源の確率分布がどのようなモデルあるいはどのようなパラメータにより表される確率分布であったとしても, 平均符号語長に関するそのような性質を満たす事が望まれている. つまり, 仮定した確率モデルは情報源の確率分布を表現できる事が前提とされている. しかし, このような前提の下で平均符号語長がそのような性質を満たすユニバーサル符号に対し, 確率モデルが表現できるとしたその前提が実際には成り立たなかった場合の符号化性能を評価する事も重要であると考えられる. 実用的な場面においては, 情報源系列の観測を始める前に確率モデルを仮定し, そのモデルに基づき, 以降は情報源系列を観測しながら符号化を行っていくため, 確率モデルが情報源の確率分布を表現できない場合は往々にして起こり得るはずである. しかし, これまでのところ, そのような問題設定についてはほとんど議論されていない. ただし, 確率モデルはマルコフ過程であると仮定されているが, 次数 k が未知, つまり何次のマルコフ過程かは未知であるといったモデル未知の確率モデルに対し, 任意の次数および任意の遷移確率パラメータに対して, それらによって規定される確率分布がデータが従う確率分布と等しくならない場合, データが従う確率分布を仮定した確率モデルでは表現できないと呼ぶ事にする. 従来, 統計学におけるモデル選択問題に関する研究では, データが従う確率分布を表現できない確率モデルについての研究が一部で行われており, 例えば, モデル選択基準の一つである AIC [1] をデータが従う確率分布を表現できない確率モデルまで拡張した TIC が提案されている [23]. AIC や TIC の解析においては, パラメータの最尤推

⁶区分定常情報源をモデル未知の確率モデルとみなした場合, これは統計学における変化点検出の問題と同様であるが, 変化点を検出するという目的と平均符号語長を最小にする符号化を行うという目的とでは, 評価関数や損失関数が異なる. したがってその目的に応じて異なった手法が必要であり, 目的を達成するための最適なアルゴリズムも異なってくる.

定量が用いられているが，仮定したモデル既知の確率モデルがデータが従う確率分布を表現できる場合，最尤推定量は，ある正則条件の下，その確率分布のパラメータに概収束し，漸近正規性を持つ事が知られている [14]．一方，データが従う確率分布を仮定したモデル既知の確率モデルでは表現できない場合，最尤推定量は，先の正則条件に対応する適切な条件の下，同様の漸近的性質を持つ事が知られている．すなわち，最尤推定量は，仮定した確率モデルの中の確率分布の内，データが従う確率分布とのカルバックライブラー情報量を最小にする確率分布のパラメータに概収束し，漸近正規性を持つ事が知られている [17, 27]．

1.2 本論文の目的

本論文ではまず，モデル既知の確率モデルに基づくベイズ符号において情報源の確率分布がその確率モデルでは表現されない分布だった場合の冗長度を精密に漸近評価する．さらに，その結果を応用し，マルコフ過程のように入れ子構造を持つ階層型のモデル未知の確率モデルに基づくベイズ符号において，情報源の確率分布がその確率モデルでは表現されない分布だった場合の冗長度を精密に漸近評価し，そのように仮定が崩れた場合でもベイズ符号が有効である事を論じる．すなわち，階層型のモデル未知の確率モデルに基づくベイズ符号は，そのように仮定が崩れた場合であっても最高次の確率モデルのみに基づくベイズ符号と比較し，冗長度の意味で同等あるいはそれ以上の性能を発揮する事を示す．

本論文ではさらに，区分定常情報源に関するベイズ符号の応用について考察を行う．先に述べたブロックソート法は，これまでのところ，その符号化性能が理論的に明確に解析されていない．本論文では，モデル未知の確率モデルの一種である，次数 k が未知のマルコフ過程で表される情報源に対し，ブロックソート法の特徴的前処理であるブロックソート後の系列が区分定常情報源からの出力系列とみなせる事を利用し，区分定常情報源のベイズ符号を応用する事により，実用性をほとんど失う事なく理論的にも明確な符号化アルゴリズムを提案し，その有効性を論じる．

1.3 本論文の構成

第 2 章では，問題設定および従来研究について述べる．まず，無歪み可変長情報源符号化の数理モデルと問題設定を述べた後，シャノンの情報源符号化定理とハフマン符号等の実用的な符号化アルゴリズムについて述べる．そして，実用的なユニバーサル符号の代表例であるブロックソート法について述べ，確率モデルに基づくユニバーサル符号について述べる．最後に，ベイズ符号を定義し，本論文に関連するベイズ符号の従来研究について述べる．第 3 章においては，確率モデルが情報源の確率分布を表現できない場合の問題を考える重要性について述べた後，本論文の位置づけについて述べる．そして，本論文の一方の成果である，情報源の確率分布が仮定した確率モデルでは表現されない分布だった場合のベイズ符号の冗長度を解析し，ベイズ符号の有用性について考察する．第 4 章では，本論文のもう一方の成果である，ブロックソート法に区分定常情報源のベイズ符号を応用した符号化アルゴリズムを提案し，その有効性について考察する．最後に，第 5 章において以上の成果をまとめる．

第2章 問題設定および従来研究

2.1 はじめに

本章では、本論文の問題設定および従来研究について述べる。まず、無歪み可変長情報源符号化の数理モデルと問題設定を述べた後、シャノンの情報源符号化定理とハフマン符号等の実用的な符号化アルゴリズムについて述べる。そして、実用的なユニバーサル符号の代表例であるブロックソート法について述べた後、確率モデルに基づくユニバーサル符号について述べる。最後に、確率モデルに基づくユニバーサル符号と統計的予測問題との関連について言及した後、ベイズ符号を定義し、本論文に関連するベイズ符号の従来研究について述べる。

2.2 無歪み可変長情報源符号化

本節では、無歪み可変長情報源符号化の数理モデルと問題設定を述べた後、シャノンの情報源符号化定理とハフマン符号等の実用的な符号化アルゴリズムについて述べる。

2.2.1 可変長情報源符号化の数理モデル

可変長情報源符号化の数理モデルを図 2.1 に示す。離散有限集合 \mathcal{X} を情報源アルファベットと呼ぶ。情報源アルファベット \mathcal{X} 上に値をとる確率変数を X とし、その実現値を x とする。さらに、 \mathcal{X}^n 上の確率変数列を $X^n = X_1 \cdots X_n$ とし、その実現値 $x^n = x_1 \cdots x_n$ を情報源系列と呼ぶ。情報源系列 x^n は、確率質量関数 $p(x^n)$ を持つ情報源の確率分布に従い発生する。離散有限集合 \mathcal{U} を符号語アルファベットと呼

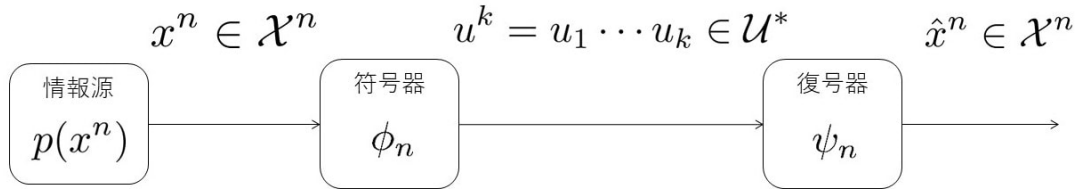


図 2.1: 可変長情報源符号化の数理モデル

び, \mathcal{U}^* を \mathcal{U} の要素からなる任意の長さの系列の集合とする. 符号化関数 ϕ_n および復号関数 ψ_n を次のように定義する.

$$\phi_n : \mathcal{X}^n \longrightarrow \mathcal{U}^*, \quad (2.1)$$

$$\psi_n : \mathcal{U}^* \longrightarrow \mathcal{X}^n. \quad (2.2)$$

ここで, $\phi_n(x^n)$ を符号語と呼び, $\phi_n(x^n) = u^k = u_1 \cdots u_k \in \mathcal{U}^*$ に対し, $|\phi_n(x^n)| = k / \log |\mathcal{U}|$ を符号語長と呼ぶ. 符号語長 $|\phi_n(x^n)|$ は一般には元の情報源系列長 n とは異なり, これは可変長符号化と呼ばれる.

2.2.2 無歪み符号化とクラフトの不等式

任意の長さの情報源系列に対し, 対応する符号語を復号した系列が元の情報源系列と一致しない確率が 0 である場合を無歪みと呼ぶ. 無歪みとはすなわち任意の $n = 1, 2, \dots$ に対し, 次の式が成り立つ事である.

$$\Pr \{ \psi_n(\phi_n(X^n)) \neq X^n \} = 0. \quad (2.3)$$

任意の無歪みな符号化において, 次の式が満たされなければならない事が知られている (例えば, [8] 等).

補題 2.2.1

$$\sum_{x^n \in \mathcal{X}^n} |\mathcal{X}|^{-|\phi_n(x^n)|} \leq 1. \quad (2.4)$$

ただし, $|\mathcal{X}|$ は \mathcal{X} の要素数を表す.

2.2.3 平均符号語長とシャノンの情報源符号化定理

無歪み可変長符号化における情報の伝送および蓄積の効率性を測る代表的な評価基準として平均符号語長が挙げられる。平均符号語長とは各情報源系列に対する符号語長の情報源の確率分布に関する期待値であり，次のように定義される。

$$E_{p^n} [|\phi_n(X^n)|] = \sum_{x^n \in \mathcal{X}^n} p(x^n) |\phi_n(x^n)| \quad (2.5)$$

ただし， E_{p^n} は $p(x^n)$ による期待値を表す。

従来，情報源に定常エルゴードな確率分布を仮定した場合，シャノンの情報源符号化定理として，任意の無歪みな符号化に対し，情報源 1 シンボルあたりの平均符号語長の下限が漸近的にエントロピーレートで与えられる事が知られ，さらに，情報源 1 シンボルあたりの平均符号語長が漸近的にエントロピーレートに等しくなる，ある無歪みな符号化が存在する事が知られている。すなわち，以下の定理が成り立つ事が知られている [20]。

定理 2.2.1 [20] 以下，確率質量関数 $p(x^n)$ は定常エルゴード性を満たすとする。このとき，任意の無歪みな符号化は次の式を満たさなければならない。

$$\lim_{n \rightarrow \infty} \frac{1}{n} E_{p^n} [|\phi_n(X^n)|] \geq H(\mathcal{X}). \quad (2.6)$$

一方，次の式を満たすある無歪みな符号化が存在する。

$$\lim_{n \rightarrow \infty} \frac{1}{n} E_{p^n} [|\phi_n(X^n)|] = H(\mathcal{X}). \quad (2.7)$$

ただし， $H(\mathcal{X})$ は情報源のエントロピーレートと呼ばれ，情報源の確率分布から一意に定まる量である。これは次のように定義される¹。

$$H(\mathcal{X}) = \lim_{n \rightarrow \infty} -\frac{1}{n} \sum_{x^n \in \mathcal{X}^n} p(x^n) \log p(x^n). \quad (2.8)$$

ここで， \log の底は $|\mathcal{X}|$ である事に注意する。

¹確率質量関数 $p(x^n)$ が定常エルゴード性を満たせば，極限值は必ず存在する。

2.2.4 平均符号語長の理論的境界を達成する符号の構成

平均符号語長を評価基準とする無歪み情報源符号化においては、シャノンの情報源符号化定理により、平均符号語長の理論的境界が与えられているため、クラフトの不等式を満たし、かつ、平均符号語長がその理論的境界に到達する符号を構成する事が主な目的となる。実際、ハフマン符号 [13] と呼ばれる、情報源の確率構造を利用した情報源 1 シンボルあたりの平均符号語長がエントロピーレートに漸近する符号化アルゴリズムが提案されている。ハフマン符号は定常無記憶情報源、すなわち情報源シンボルが独立にある定常な確率分布に従って発生する情報源に対して、情報源 1 シンボルを符号化した場合に平均符号語長が理論的境界値であるエントロピー²に達する事が知られている [13]。以下、符号語アルファベット \mathcal{U} を $\{0, 1\}$ とした場合の定常無記憶情報源に対するハフマン符号化アルゴリズムを示す。

- step-1: 情報源アルファベット \mathcal{X} の各要素を生起する確率 $p(x)$ が大きい順に並べる。
- step-2: 生起する確率 $p(x)$ が最も小さい二つの要素にそれぞれ 0 と 1 を割り当てる。そして、その二つの生起する確率の和を計算し、その二つの要素を、その和を生起する確率とした一つの要素とみなし、各要素を生起する確率が大きい順に並べる。
- step-3: 生起する確率の和が 1 になるまで step-2 を繰り返す。
- step-4: 各要素に割り当てられた系列をその符号語とする。

さらに、ハフマン符号を情報源系列の長さに応じて拡大して適用する事により、情報源 1 シンボルあたりの平均符号語長は $\lceil -(1/n) \log p(x^n) + 1/n \rceil$ 以下となるが、すなわち平均符号語長はエントロピーレートに漸近するが、符号化における計算量が情報源系列の長さに応じて指数的に大きくなってしまいう事が知られている。一方、算術符号 [18, 19] と呼ばれるより計算量の削減された実用的な符号化アルゴリズムも提案されており、情報源系列 x^n に対する符号語長は $-\log p(x^n) + 2$ により定められ

²定常無記憶情報源に対するエントロピーレートはエントロピー $H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$ に帰着される。

る事が知られている（例えば，[8] 等）．つまり，ハフマン符号も算術符号も符号語長は情報源の確率分布により定められる事が分かる．ここで， $-\log p(x^n)$ は理想符号語長と呼ばれる．

2.3 ユニバーサル情報源符号化

本節では，まず，実用的なユニバーサル符号の代表例であるブロックソート法について述べる．次に，確率モデルを定義し，その例として FSMX 情報源と区分定常情報源の二つを挙げる．最後に，確率モデルに基づくユニバーサル符号の中でも評価関数を導入した符号化について述べる．

2.3.1 実用性と理論的明瞭性のトレードオフ関係

情報源の確率構造の一部が未知である情報源符号化の問題はユニバーサル情報源符号化の分野として研究が行われており，実用的にも重要な研究である．ユニバーサル情報源符号化の多くの研究においては，定常エルゴードな確率分布を持つ情報源に対し，無歪みの下，情報源 1 シンボルあたりの平均符号語長が情報源のエントロピーレートに漸近するような符号を構成する事が重要なテーマとなっている．例えば，テキストデータの圧縮等に应用されている LZ 符号 [32, 33]，ネットワーク上でやり取りされるファイルの圧縮等に利用される bzip2 に应用されているブロックソート法 [4] は，情報源に定常エルゴードな確率分布を仮定した場合，いずれも実験的には情報源系列の長さを十分大きくすれば平均符号語長がエントロピーレートに収束する事が知られている．しかし，その収束オーダーがどの程度であるかといった事は理論的には明らかにされていない．LZ 符号に関しては情報源に定常エルゴードな確率分布を仮定した場合，平均符号語長がエントロピーレートに収束する事が理論的にも証明されているが [31]，その収束オーダーまでは明らかにされていない．ブロックソート法に関しては情報源の確率分布が有限次数の定常マルコフ過程により表される場合，平均符号語長がエントロピーレートにより上界される事が理論的にも証明されているが [2]，その収束オーダーまでは明らかにされていない．一方，確率構造の一部が未知である情報源に対して確率モデルを仮定し，それに基づいて符号化を行うといった手

法についても一部では研究が行われている [3, 16, 21, 22, 24, 25, 28, 30]. さらに, 確率モデルに加えて評価関数を導入し, 仮定した確率モデルに基づき評価関数の最適化を行う事により, 符号化を行うといった手法も提案されている [3, 16, 21, 22, 24, 25]. このような手法については, 時間計算量やメモリ計算量が膨大となり, 実用化においてはそれらを削減した効率的な符号化アルゴリズムを考案する必要があるが, 平均符号語長の評価をその収束オーダーまで含めて精密に行う事が可能である等, 理論的明瞭性は大きい.

2.3.2 実用的なユニバーサル符号の代表例: ブロックソート法

ブロックソート法ではまず, 情報源系列を 1 シンボルずつ巡回シフトした系列を全て辞書順にソートする事により, 元の情報源系列と同じ長さの別の系列に変換する. これは BW 変換と呼ばれる. そして, 情報源系列に BW 変換を施した系列を Move-To-Front 法 (以下, MTF 法) により符号語に変換する.

BW 変換のアルゴリズム [4] を以下に示す.

- step-1: 情報源系列 x^n を逆順に並べた系列を \bar{x}^n とし, \bar{x}^n の最後に終端記号 \$ を加えた系列 $\bar{x}^n\$$ を生成する.
- step-2: step-1 で生成した系列 $\bar{x}^n\$$ を左巡回シフトさせて生成した n 個の系列を縦に並べ, $(n+1)$ 行 $(n+1)$ 列の行列 M を生成する.
- step-3: step-2 で生成した行列 M の各行を 1 列目のシンボルから辞書順にソートし, ソート後の行列を \tilde{M} とする.
- step-4: \tilde{M} の第 $n+1$ 列を \$ を除いて上から $y_1y_2\cdots$ として y^n を出力し, さらに, y^n を出力する際に除いた \$ の行番号 r を出力する.

step-4 において y^n を出力する際に除いた \$ の行番号は, BW 変換後の系列 y^n を元の情報源系列 x^n に逆変換する際に必要となるが, 逆変換のアルゴリズムについては [4] を参照されたい.

例えば，情報源系列 x^9 が 110011010 である場合の BW 変換は次のように行われる．まず，step-1 において生成される系列 \bar{x}^9 は 010110011\$ となる．そして，step-2 および 3 で生成される行列 M および \tilde{M} は次のように与えられる．

$$M = \begin{bmatrix} 010110011\$ \\ 10110011\$0 \\ 0110011\$01 \\ 110011\$010 \\ 10011\$0101 \\ 0011\$01011 \\ 011\$010110 \\ 11\$0101100 \\ 1\$01011001 \\ \$010110011 \end{bmatrix}, \quad \begin{bmatrix} 0011\$01011 \\ 010110011\$ \\ 0110011\$01 \\ 011\$010110 \\ 10011\$0101 \\ 10110011\$0 \\ 110011\$010 \\ 11\$0101100 \\ 1\$01011001 \\ \$010110011 \end{bmatrix} = \tilde{M}. \quad (2.9)$$

最後に，step-4 で $(y^9, r) = (110100011, 2)$ を出力する．

ブロックソート法 [4] では，BW 変換後の系列 y^n に MTF 法を適用する事により，元の情報源系列 x^n に対する符号語が定められる．

2.3.3 確率モデルの定義

ユニバーサル情報源符号化の分野においては一方で，情報源の確率構造に対して確率モデルを仮定し，確率モデルに基づきユニバーサル符号化を行うといった手法についても研究が行われている．本論文では確率モデルに基づくユニバーサル符号を主に扱うため，まず，確率モデルを次のように定義する．

ある実数値パラメータ θ により規定される確率質量関数を $p(x^n | \theta)$ と表す．パラメータ θ は未知であり，あるパラメータ空間 $\Theta \subseteq \mathbf{R}^k$ 上に値を取るとすると，確率質量関数 $p(x^n | \theta)$ の族である $\{p(x^n | \theta) | \theta \in \Theta\}$ が定義される．この確率質量関数の族は確率モデルと呼ばれる．さらに，パラメータ空間 Θ がある確率モデルを表すインデックス m により区別される状況を考える．この時，確率質量関数は，確率モデルを表すインデックス m とその下でのパラメータ θ_m によって規定され， $p(x^n | \theta_m, m)$

と表される. ただし, $\theta_m = (\theta_{m,1}, \dots, \theta_{m,k_m})^T$ は, $\Theta_m \subseteq \mathbf{R}^{k_m}$ 上の k_m 次元実数値パラメータである. 以後, a^T および A^T はそれぞれベクトル a および行列 A の転置を表す. 確率モデルを表すインデックス m はある離散有限集合 \mathcal{M} 上に値を取り, パラメータ θ_m はインデックス m 毎に異なるパラメータ空間 Θ_m 上に値を取るため, 確率質量関数 $p(x^n | \theta_m, m)$ の族 \mathcal{H} は次のように定義される.

$$\mathcal{H} = \{p(x^n | \theta_m, m) \mid \theta_m \in \Theta_m, m \in \mathcal{M}\}. \quad (2.10)$$

以後, 本論文では \mathcal{H} をモデル未知の確率モデルと呼ぶ事にする. これは, 確率モデルを表すインデックス m が未知であり, さらにその下でのパラメータ θ_m も未知であるという事を表す.

一方, 確率モデルを表すインデックス m がある一つの既知である値に固定され, その下でのパラメータ θ_m が未知であるような場合も考えられる. この時, m は一つに固定されているため, 確率質量関数 $p(x^n | \theta_m, m)$ の族 \mathcal{H}_m は次のように定義される.

$$\mathcal{H}_m = \{p(x^n | \theta_m, m) \mid \theta_m \in \Theta_m\}. \quad (2.11)$$

以後, 本論文では \mathcal{H}_m をモデル既知の確率モデルと呼ぶ事にする.

今後は表記の簡略化のため, 確率モデルを表すインデックス m の事をモデル m と呼び, モデル m とは m がある一つの既知である値に固定されたモデル既知の確率モデル, すなわち, ある固定された m の下でのパラメータ θ_m が未知である場合の確率質量関数 $p(x^n | \theta_m, m)$ の族 \mathcal{H}_m を表す事にする.

また, モデル未知の確率モデル \mathcal{H} は入れ子構造を持つ事, すなわち, $k_{m_1} < k_{m_2} < \dots$ なる $m_1, m_2, \dots \in \mathcal{M}$ に対し,

$$\mathcal{H}_{m_1} \subset \mathcal{H}_{m_2} \subset \dots, \quad (2.12)$$

である事を許容する. 本論文では, 式 (2.12) が成り立つ \mathcal{H} を入れ子構造を持つ階層型のモデル未知の確率モデルと呼ぶ事にする.

マルコフ過程を例に挙げる. ここで, $\mathcal{M} = \{1, 2, 3\}$ に対し, モデル $m = 1$ は遷移確率パラメータが未知である 1 次マルコフ過程の確率モデル \mathcal{H}_1 を表す事にする. ただし, θ_1 は 1 次マルコフ過程の遷移確率パラメータに対応する. $m = 2, 3$ につい

でも同様である。この時、 m が \mathcal{M} 上のいずれの値を取るか未知である場合、モデル未知の確率モデルを仮定している事になる。一方、 $m = 1$ である事が既知である場合、モデル既知の確率モデル \mathcal{H}_1 を仮定している事になる。 $m = 2, 3$ についても同様である。ここで、 $k_1 < k_2 < k_3$ であり、 $\mathcal{H}_1 \subset \mathcal{H}_2 \subset \mathcal{H}_3$ であるから、 m が \mathcal{M} 上のいずれの値を取るか未知である場合、入れ子構造を持つ階層型のモデル未知の確率モデルを仮定している事になる。

2.3.4 モデル未知の確率モデルの例 1 : FSMX 情報源

FSMX 情報源とはある時点の情報源シンボルの生起する確率が過去の有限長の系列から定まる情報源でマルコフ過程の一種である。FSMX 情報源の確率分布はモデル m とモデル m の下でのパラメータ θ_m により定まるため、モデル未知の確率モデルとみなせる。

離散有限集合 S_m をモデル m における状態の集合とする。情報源系列 x^{t-1} から状態 $s \in S_m$ への写像を次のように表す。

$$s(x^{t-1}) : \mathcal{X}^{t-1} \longrightarrow S_m. \quad (2.13)$$

各状態 s における各情報源シンボルの生起確率は $|\mathcal{X} - 1|$ 次元パラメータ θ^s により定まる。情報源系列 x^{t-1} を得た下で x_t が生起する確率は次のように与えられる。

$$p(x_t | x^{t-1}, \theta_m, m) = p(x_t | \theta^{s(x^{t-1})}, s(x^{t-1})). \quad (2.14)$$

したがって、情報源系列 x^n が生起する確率 $p(x^n)$ は次のように与えられる。

$$p(x^n) = p(x^n | \theta_m, m) \quad (2.15)$$

$$= \prod_{t=1}^n p(x_t | x^{t-1}, \theta_m, m) \quad (2.16)$$

$$= \prod_{t=1}^n p(x_t | \theta^{s(x^{t-1})}, s(x^{t-1})). \quad (2.17)$$

ただし、空系列を λ とした場合、 $s(\lambda)$ は初期状態であり、 $s(x^0) = s(\lambda)$ である。

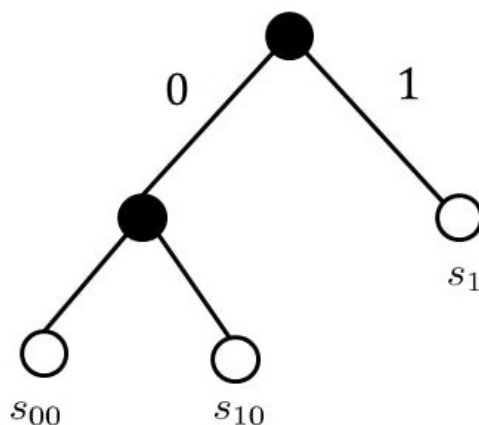


図 2.2: 2 元 FSMX 情報源モデル

FSMX 情報源のモデル m は完全木で表現する事が可能である. 完全木におけるそれぞれの枝はシンボル $x \in \mathcal{X}$ に対応する. また, 完全木における葉ノードから根ノードへの一つのパスを文脈あるいはポストフィクスと呼ぶ. 完全木における葉ノードは状態 s と一対一に対応しているため, 葉ノードは s と書く事ができる. したがって, FSMX 情報源のモデル m における状態の集合 S_m は完全木における葉ノードの集合とみなせる. 特に, モデル m を完全木で表現した場合に全ての葉ノードが深さ m のみに存在する時, その情報源は m 次マルコフ情報源と呼ばれる. これは入れ子構造を持つ階層型のモデル未知の確率モデルとみなせる.

FSMX 情報源のモデルの例を図 2.2 に示す. 文脈 $x^{t-1} = \dots 10$ により定まる状態を s_{10} とすると, 情報源系列 $x^5 = 10010$ において時点 $t = 2$ の状態は $s(1) = s_1$, 時点 $t = 3$ の状態は $s(10) = s_{10}$, 時点 $t = 4$ の状態は $s(100) = s_{00}$ となる. また, $S(m) = \{s_1, s_{10}, s_{00}\}$ である.

2.3.5 モデル未知の確率モデルの例 2 : 区分定常情報源

区分定常情報源は, 時点 1 からある時点 $t_1 - 1$ まではあるパラメータで支配される定常分布に従って情報源系列を発生させる. そして, 時点 t_1 においてそのパラメー

タの値が変化し、以後、時点 t_1 からある時点 $t_2 - 1$ まではその変化後のパラメータで支配される定常分布に従って情報源系列を発生させ、ある時点 t_2 において再びパラメータの変化が起こる。これを時点 n まで繰り返す。区分定常情報源の確率分布はパラメータの変化パターン ω とその変化パターン ω の下でのパラメータ ϕ_ω により定まるため、モデル未知の確率モデルとみなせる。

情報源系列 y^n の i 番目から j 番目までの部分系列を $y_i^j = y_i y_{i+1}, \dots, y_j$ と表す。パラメータの変化パターン ω はパラメータの変化時点の集合として表現される。すなわち、ある変化パターン $\omega \in \Omega$ を $\omega = (t_{\omega,1}, \dots, t_{\omega,C(\omega)})$ と表す。ここで、 $t_{\omega,0} = 1$ および $t_{\omega,C(\omega)+1} = n+1$ を仮定する。部分系列 $y_{t_{\omega,c}}^{t_{\omega,c+1}-1}$ を発生させる定常分布のパラメータを $\phi_{t_{\omega,c}}$ と表す。ただし、 $c = 0, \dots, C(\omega)$ である。すると、ある変化パターン ω のパラメータベクトルは $\phi_\omega = (\phi_{t_{\omega,0}}, \dots, \phi_{t_{\omega,C(\omega)}}) \in \Phi_\omega$ と書ける。部分系列 $y_{t_{\omega,c}}^{t_{\omega,c+1}-1}$ に含まれる各シンボルの条件付確率質量関数を $p(y_t | y^{t-1}, \phi_{t_{\omega,c}}, \omega)$ とすると、ある変化パターン ω が与えられた下での情報源系列 y^n の確率質量関数は以下のように書き表される。

$$p(y^n | \omega) = \prod_{c=0}^{C_\omega} \prod_{t=t_{\omega,c}}^{t_{\omega,c+1}-1} p(y_t | y^{t-1}, \phi_{t_{\omega,c}}, \omega). \quad (2.18)$$

2.3.6 評価関数の導入と非逐次型および逐次型の符号化

先に述べたハフマン符号や算術符号は、その符号語長が情報源の確率分布によって定まるものであった。つまり、符号化確率、すなわち、符号語長を与える確率分布を何らかの形で決定できれば算術符号等を用いた符号化が可能であるという事である。したがって、確率モデルの仮定に基づくユニバーサル情報源符号化に関する一部の研究においては、符号化の良し悪しを測る基準としてある評価関数を導入した下でその評価関数を最適化する符号化確率を決定する事が主な目的とされている [3, 6, 10, 15, 16, 21, 22, 24, 25].

符号化確率の決定に関し、時点 1 から時点 n までに発生した情報源系列 x^n を得た下で、一括して x^n に対する符号化確率 $Ap(x^n)$ を決定する非逐次型の符号化 [28, 30] と各時点 t においてそれまでに発生した情報源系列 x^{t-1} を得た下で、次に発生する情報源シンボル x_t に対する符号化確率 $Ap(x_t | x^{t-1})$ を逐次的に決定し、それを時

点 n まで繰り返す逐次型の符号化 [16, 21, 22] の二つが挙げられる. 実際に計算機上で符号化確率を計算する場合, 非逐次型の符号化では時点 n までの確率の積を計算する事になるため, 情報源系列の長さが大きくなった場合, 符号化確率の値が 0 に非常に近い値となってしまい, 計算機上では 0 と扱われてしまう. 一方, 逐次型の符号化の場合, 各時点 t において計算した符号化確率を用いて符号語長を定めるため, そのような問題は発生しない. また, 平均符号語長の解析という点においては, 両者は等価である. ただし, 先に述べたように, 任意の無歪みな符号はクラフトの不等式を満たさなければならないが, 算術符号等のように符号化確率により符号語長を定める場合, 符号語長は非逐次型の符号化であれば $-\log Ap(x^n)$ の形で与えられ, 逐次型の符号化であれば各時点 t において $-\log Ap(x_t | x^{t-1})$ の形で与えられるため, クラフトの不等式は, 何らかの手法により決定される符号化確率が劣確率分布である事を要求するものであると言い換えられる³.

逐次型の符号化においては次で定義される関数を評価関数として導入するのが一般的である [6, 10, 15, 16, 21, 22, 28, 30]⁴.

$$\sum_{t=1}^n \log \frac{p(x_t | x^{t-1})}{Ap(x_t | x^{t-1})} = \log \frac{p(x^n)}{Ap(x^n)}. \quad (2.19)$$

ここで, $\log(p(x^n)/Ap(x^n))$ は, 符号化確率 $Ap(x^n)$ を用いた場合の符号語長と理想符号語長 $-\log p(x^n)$ との差分を表す. さらに, この評価関数の情報源の確率分布に関する期待値は冗長度と呼ばれ, 次のように定義される.

$$E_{p^n} \left[\log \frac{p(X^n)}{Ap(X^n)} \right] = \sum_{x^n \in \mathcal{X}^n} p(x^n) \log \frac{p(x^n)}{Ap(x^n)}. \quad (2.20)$$

これは平均符号語長と情報源のエントロピーレートとの差分, すなわち相対エントロピーを表す.

³符号語長を $-\log Ap(x^n)$ で与えた場合, クラフトの不等式は $\sum_{x^n \in \mathcal{X}^n} Ap(x^n) \leq 1$ となる.

⁴一方, 非逐次型の符号化においても同様に時点 n までの情報源の確率分布および符号化確率を用いた $\log(p(x^n)/Ap(x^n))$ が評価関数として導入される事が多い.

2.4 ベイズ符号に関する従来研究

本節では、逐次型の符号化と統計的予測問題との関連について言及した後、ベイズ符号を定義し、本論文においてベイズ符号を対象とするモチベーションと本論文に関連するベイズ符号の従来研究について述べる。

2.4.1 逐次型の符号化と統計的予測問題との関連

統計的予測問題における一部の研究 [11, 12] では、時系列的に観測されるデータに基づき、次の時点に発生するデータが何であるかを予測する、あるいはデータが従う確率分布そのものを推定する問題が考えられている。特に後者については、各時点 t におけるデータが従う確率分布 $p(x_t | x^{t-1})$ の推定に対し、推定の良し悪しを測る何らかの損失関数を導入し、その損失関数を最小にする決定関数 $Ap(x_t | x^{t-1})$ を求めるといった事が主に行われている。損失関数の一例として $\log(p(x_t | x^{t-1})/Ap(x_t | x^{t-1}))$ を時点 1 から時点 n まで足し合わせた対数損失 $\log(p(x^n)/Ap(x^n))$ が挙げられる。これは、ユニバーサル情報源符号化における符号化確率 $Ap(x^n)$ を用いた場合の符号語長と理想符号語長 $-\log p(x^n)$ との差分を表す。さらに、対数損失 $\log(p(x^n)/Ap(x^n))$ のデータが従う確率分布に関する期待値はカルバックライブラー情報量と呼ばれ、これは 2.3.6 節で述べたユニバーサル情報源符号化における冗長度と等価であるとみなせる。つまり、冗長度すなわち平均符号語長を評価基準としたユニバーサル情報源符号化の問題において平均符号語長の最小化を図る事は、統計的予測問題の中でも対数損失を導入した確率分布の推定問題においてカルバックライブラー情報量の最小化を図る事に等しいと言える [11, 12]。

2.3.6 節までに述べた問題設定では \mathcal{X} が離散有限集合であった。一方、統計的予測問題においては \mathcal{X} は離散有限集合とは限らない一般の集合として議論される事が多い。そこで、 \mathcal{X} が一般の集合である場合の問題設定についても述べておく。後に紹介する従来研究の一部と本論文で述べる成果の一部は、 \mathcal{X} が一般の集合である場合の問題設定となっている。

まず、 $t = 1, \dots, n$ に対し、 \mathcal{X}_t を可測空間とし、 \mathcal{X}_t 上に値を取る確率変数 X_t を x_t と

する. さらに, \mathcal{X}_t 上のある測度 ν に対し, 確率密度 $p(x_t | x^{t-1})$ ⁵ が存在し, x_t は $p(x_t | x^{t-1})$ に従って発生すると仮定する. また, \mathcal{X}^n 上の確率変数列を $X^n = X_1 \cdots X_n$ とし, その実現値を $x^n = x_1 \cdots x_n$ とする. ここで, $p(x^n) = \prod_{t=1}^n p(x_t | x^{t-1})$ はデータが従う確率分布の密度関数である⁶. ただし, $p(x^n)$ は定常性かつエルゴード性を満たすものとする. 決定関数 $Ap(x_t | x^{t-1})$ を $\int_{\mathcal{X}_t} Ap(x_t | x^{t-1}) d\nu(x_t) = 1$ なる \mathcal{X}_t 上の確率密度とする⁷. 対数損失を導入した統計的予測問題においては, 以下で定義される損失関数を何らかの基準により最小化する $Ap(x_t | x^{t-1})$ を決定する事が主な目的となる.

$$\sum_{t=1}^n \log \frac{p(x_t | x^{t-1})}{Ap(x_t | x^{t-1})} = \log \frac{p(x^n)}{Ap(x^n)}. \quad (2.21)$$

ただし, \log は 2.3.6 節までとは異なり自然対数である. これは対数損失と呼ばれる. さらに, 対数損失の $p(x^n)$ に関する期待値はカルバックライブラー情報量として次のように定義される.

$$E_{p^n} \left[\log \frac{p(X^n)}{Ap(X^n)} \right] = \int_{\mathcal{X}^n} p(x^n) \log \frac{p(x^n)}{Ap(x^n)} d\nu(x^n), \quad (2.22)$$

以後, E_{p^n} は $p(x^n)$ に関する期待値を表す. これは, \mathcal{X}_t が離散有限集合であれば, 2.3.6 節で定義した冗長度に相当する. カルバックライブラー情報量の解析を行う事はユニバーサル符号化における冗長度, すなわち平均符号語長の解析を行う事に等しい.

また, 2.3.3 節で定義した確率モデルにおける確率質量関数は, 適宜, 確率密度関数と読み換える事にする.

2.4.2 ベイズ符号

統計的予測問題における一部の研究においてはデータが従う確率分布に確率モデルを仮定し, ミニマックス基準やベイズ基準に基づく決定が行われている [11]. 本

⁵もし, X_t が離散であれば, 各 $x_t \in \mathcal{X}_t$ に対し, $\nu(\{x_t\}) = 1$ であり, $p(x_t | x^{t-1})$ は \mathcal{X}_t 上の確率質量関数である.

⁶ただし, X^n が離散の場合は確率質量関数である.

⁷もし, X_t が離散であれば, $Ap(x_t | x^{t-1})$ は $\sum_{x_t} Ap(x_t | x^{t-1}) = 1$ なる \mathcal{X}_t 上の確率質量関数である. いずれもクラフトの不等式を等号で満たす場合を考えている事になる.

節ではベイズ基準に基づく決定について述べる. 2.4.1 節で定義した対数損失およびカルバックライブラー情報量は損失関数およびリスク関数として次のように書き換えられる. まず, 損失関数 $L(\theta_m, m, Ap(x^n), x^n)$ は次のように定義される.

$$L(\theta_m, m, Ap(x^n), x^n) = \log \frac{p(x^n | \theta_m, m)}{Ap(x^n)}. \quad (2.23)$$

そして, 損失関数を $p(x^n)$ に関して期待値を取ったリスク関数 $R(\theta_m, m, Ap(x^n))$ は次のように定義される.

$$R(\theta_m, m, Ap(x^n)) = \int_{\mathcal{X}^n} p(x^n | \theta_m, m) L(\theta_m, m, Ap(x^n), x^n) d\nu(x^n). \quad (2.24)$$

まず, モデル既知の確率モデルを仮定した場合について述べる. ベイズ基準に基づく決定を行う場合, パラメータの事前分布, すなわち Θ_m 上の事前確率密度を $w(\theta_m | m)$ と定義し, リスク関数をパラメータの事前分布に関し, 期待値を取った以下のベイズリスク関数 $BR_{\mathcal{H}_m}(Ap(x^n))$ を導入する.

$$BR_{\mathcal{H}_m}(Ap(x^n)) = \int_{\Theta_m} w(\theta_m | m) R(\theta_m, m, Ap(x^n)) d\theta_m. \quad (2.25)$$

ベイズ基準の下で最適 (以下, ベイズ最適) な決定関数, すなわちベイズリスク関数を最小にする決定関数 $Ap_m^*(x^n)$ は次のように与えられる [15].

$$Ap_m^*(x^n) = p(x^n | m) \quad (2.26)$$

$$= \int_{\Theta_m} p(x^n | \theta_m, m) w(\theta_m | m) d\theta_m. \quad (2.27)$$

一方, 各時点 t における逐次型のベイズ最適な決定関数 $Ap_m^*(x_t | x^{t-1})$ は次のように与えられる [15].

$$Ap_m^*(x_t | x^{t-1}) = p(x_t | x^{t-1}, m) \quad (2.28)$$

$$= \int_{\Theta_m} p(x_t | x^{t-1}, \theta_m, m) w(\theta_m | x^{t-1}, m) d\theta_m. \quad (2.29)$$

ただし, $w(\theta_m | x^{t-1}, m)$ は, x^{t-1} を得た下でのパラメータ θ_m の事後確率密度を表し, これは次のように定義される.

$$w(\theta_m | x^{t-1}, m) = \frac{p(x^{t-1} | \theta_m, m) w(\theta_m | m)}{\int_{\Theta_m} p(x^{t-1} | \theta_m, m) w(\theta_m | m) d\theta_m}. \quad (2.30)$$

統計的予測問題においては、パラメータ θ_m の推定量（例えば、最尤推定量）を $\hat{\theta}_m$ として $p(x^n | \hat{\theta}_m, m)$ が予測に用いられる事もあるが、ベイズ最適な決定では、パラメータの推定量を一つに定める事はせず、 $p(x^n | \theta_m, m)$ に対し、パラメータの取得る値全てをその事前分布あるいは事後分布で重みづけたものを予測に用いている事に相当する。ただし、これは損失関数として対数損失を導入している事によるものである。

次に、モデル未知の確率モデルを仮定した場合について述べる。この場合はさらに、モデルの事前分布、すなわち \mathcal{M} 上の事前確率を $P(m)$ と定義し、リスク関数をパラメータの事前分布とモデルの事前分布のそれぞれに関し、期待値を取った以下のベイズリスク関数 $BR_{\mathcal{H}}(Ap(x^n))$ を導入する。

$$BR_{\mathcal{H}}(Ap(x^n)) = \sum_{m \in \mathcal{M}} P(m) \int_{\Theta_m} w(\theta_m | m) R(\theta_m, m, Ap(x^n)) d\theta_m. \quad (2.31)$$

ベイズ最適な決定関数 $Ap^*(x^n)$ は次のように与えられる [15].

$$Ap^*(x^n) = \sum_{m \in \mathcal{M}} P(m) p(x^n | m) \quad (2.32)$$

$$= \sum_{m \in \mathcal{M}} P(m) \int_{\Theta_m} p(x^n | \theta_m, m) w(\theta_m | m) d\theta_m. \quad (2.33)$$

一方、各時点 t における逐次型のベイズ最適な決定関数 $Ap^*(x_t | x^{t-1})$ は次のように与えられる [15].

$$Ap^*(x_t | x^{t-1}) = \sum_{m \in \mathcal{M}} P(m | x^{t-1}) p(x_t | x^{t-1}, m) \quad (2.34)$$

$$= \sum_{m \in \mathcal{M}} P(m | x^{t-1}) \int_{\Theta_m} p(x_t | x^{t-1}, \theta_m, m) w(\theta_m | x^{t-1}, m) d\theta_m. \quad (2.35)$$

ただし、 $P(m | x^{t-1})$ は、 x^{t-1} を得た下でのモデル m の事後確率を表し、これは次のように定義される。

$$P(m | x^{t-1}) = \frac{P(m) p(x_t | x^{t-1}, m)}{\sum_{m \in \mathcal{M}} P(m) p(x_t | x^{t-1}, m)}. \quad (2.36)$$

統計的予測問題においては、何らかのモデル選択基準を用いる事により、モデルを一つに選択した下で予測が行われる事が多いが、ベイズ最適な決定では、モデルを一つに選択する事はせず、あらゆるモデル全てをその事前確率あるいは事後確率で重みづけたものを予測に用いている事に相当する。ただし、これは損失関数として対数損失を導入している事によるものである。

ベイズ符号では、ベイズ最適な決定関数を符号化確率として用いて算術符号化を行う事を試みる。また、時点 n におけるベイズ最適な非逐次型の符号化確率と時点 1 から時点 n までのベイズ最適な逐次型の符号化確率の積とは等価になるため [15]、理論解析においては両者を区別せずに議論を進めていく。特に、冗長さすなわち平均符号語長の解析においては非逐次型の符号化確率で議論し、符号化アルゴリズムにおいては逐次型の符号化確率で議論する事が多い。

以後、本論文では、 $Ap^*(x^n)$ をモデル未知の確率モデルに基づくベイズ最適な決定（あるいは符号化確率）と呼び、 $p(x^n | m)$ をモデル既知の確率モデルに基づくベイズ最適な決定（あるいは符号化確率）と呼ぶ事にする。

2.4.3 ベイズ基準を導入するモチベーション

ベイズ符号は、任意の長さの情報源系列に対し、平均符号語長をベイズ基準の下で最小にする符号であり [15]、理論的にも重要な符号であるが、本論文においてベイズ符号を対象とする理由は次の通りである。

情報源符号化においては、例えば圧縮対象のテキストの種類に応じて、ある単語の出現頻度がどの程度であり、さらにその頻度がそれまでの文脈にどの程度依存して変わってくるか等が事前に分かっている場合が多いため、パラメータに事前分布を仮定する事は比較的自然である。

また、ミニマックス基準に基づく決定を応用した符号も一部では研究されているが [3, 24, 25]、パラメータの事前確率密度が $w(\theta_m | m)$ がジェフリーズの事前確率密度である場合、モデル既知の確率モデルに対するベイズ最適な決定関数 $p(x^n | m)$ はミニマックスリスク基準の下でも最適な決定関数になる事が知られている [7]。すなわち、ベイズ符号はある事前分布の下、漸近的にミニマックス基準に基づく決定を応用した符号と等価になる事が言える。

特にベイズ符号の場合，その性能を理論的に容易かつ精密に解析する事が可能であり，実際，平均符号語長がエントロピーレートに漸近する事が明らかにされているが，その漸近式は $o(1)$ まで明らかにされており，収束オーダーも明確である [6, 10]. これについては次節で詳細に述べる．特にマルコフ過程のように高次のモデルが低次のモデルを含む入れ子構造を持つ階層型のモデル未知の確率モデルで表される情報源に対し，ベイズ符号の平均符号語長が精密に解析されており，ベイズ符号の有効性が理論的に証明されている [10]. これについては 2.4.5 節で詳細に述べる．

一方，先に述べた，確率モデルはマルコフ過程であると仮定されているが次数 k が未知，つまり何次のマルコフ過程かは未知であるものの一種である FSMX 情報源に対し，効率的なベイズ符号化アルゴリズムが提案されている [16]. また，モデル未知の確率モデルで表される情報源の一種である区分定常情報源に対しても効率的なベイズ符号化アルゴリズムが提案されている [21]. これらについては 2.4.6 節以降で詳細に述べる．

2.4.4 モデル既知の確率モデルに基づくベイズ符号の性能解析

従来，2.4.1 節の問題設定において，モデル既知の確率モデルに基づくベイズ最適な決定におけるカルバックライブラー情報量の漸近評価が行われている．これは，モデル既知の確率モデルに基づくベイズ符号の冗長度，すなわち平均符号語長の漸近評価と等価である．本節ではその主結果について述べる．

本節で扱う確率モデルはモデル既知のものであるため，2.3.3 節や 2.4.1 節で定義されているものに関し， m はある一つの値に固定しておく．さらに，本節において，確率分布は定常無記憶であるため， $p(x^n) = p(x^n | \theta_m^*, m) = \prod_{i=1}^n p(x_i | \theta_m^*, m)$ なるある $\theta_m^* \in \Theta_m$ が存在すると仮定する．ここで， θ_m^* を真のパラメータと呼ぶ事にする．

モデル既知の確率モデルに基づくベイズ最適な決定におけるカルバックライブラー情報量は次の定理により与えられる．

定理 2.4.1 [6]

適当な正則条件 [6] の下で次の式が成り立つ.

$$E_{p^n} \left[\log \frac{p(X^n)}{p(X^n | m)} \right] = \frac{k_m}{2} \log \frac{n}{2\pi e} + \log \frac{\sqrt{\det I_{\theta_m^*}}}{w(\theta_m^* | m)} + o(1).$$

ただし, $p(x^n | m)$ は 2.4.2 節において定義されたモデル既知の確率モデルに基づくベイズ最適な決定である. また, $I_{\theta_m^*}$ は次のように定義される.

$$I_{\theta_m^*} = E \left[\frac{\partial \log p(X | \theta_m^*, m)}{\partial \theta_m} \frac{\partial \log p(X | \theta_m^*, m)}{(\partial \theta_m)^T} \right]. \quad (2.37)$$

ここで, E は $p(x | \theta_m^*, m)$ による期待値を表す.

定理 2.4.1 により, モデル既知の確率モデルに基づくベイズ符号の冗長度は, $o(1)$ までの精密な評価が行われており, かつ $O(k_m \log n)$ で 0 に収束する, すなわち, 1 シンボルあたりの平均符号語長は $O(k_m \log n/n)$ でエントロピーレートに収束する事が分かる. また, 誤差項として真のパラメータ θ_m^* に依存する定数項が現れているのが見て取れる.

2.4.5 モデル未知の確率モデルに基づくベイズ符号の性能解析

従来, 2.3.6 節までの問題設定において, モデル未知の確率モデルに基づくベイズ符号における冗長度の漸近評価が行われている. 本節ではその主結果について述べる. そして, 主結果より導かれる性質として, モデル未知の確率モデルに基づくベイズ符号はモデル既知の確率モデルに基づくベイズ符号と比較し, 平均符号語長の意味でより良い性能を発揮する事について述べる.

式 (2.12) で定義される入れ子構造を持つ階層型のモデル未知の確率モデルにおいて, $p(x^n) = p(x^n | \theta_m^*, m)$ なるある $\theta_m^* \in \Theta_m$ およびある $m \in \mathcal{M}$ が存在すると仮定する. そして, 全ての θ_m^* の内, 最小次数のものを $\theta_{m^*}^* \in \Theta_{m^*}$ と定義する. すなわち, $m^* \in \mathcal{M}$ を次のように定義する.

$$m^* = \arg \min_{m \in \mathcal{M}} \{k_m | \exists \theta_m, \forall x^n, p(x^n | \theta_m, m) = p(x^n)\}. \quad (2.38)$$

ここで, $m^*, \theta_{m^*}^*$ をそれぞれ真のモデル, 真のパラメータと呼ぶ事にする.

このとき, モデル未知の確率モデルに基づくベイズ符号の冗長度は次の定理により与えられる.

定理 2.4.2 [10]

適当な正則条件 [10] の下で次の式が成り立つ.

$$E_{p^n} \left[\log \frac{p(X^n)}{Ap^*(X^n)} \right] = \frac{k_{m^*}}{2} \log \frac{n}{2\pi e} + \log \frac{\sqrt{\det I(\theta_{m^*}^* | m^*)}}{w(\theta_{m^*}^* | m^*)} - \log P(m^*) + o(1). \quad (2.39)$$

ただし, $Ap^*(x^n)$ は 2.4.2 節において定義されたモデル未知の確率モデルに基づくベイズ最適な符号化確率である. また, $I(\theta_m | m)$ は次の式で定義され, その極限值は存在すると仮定する.

$$I(\theta_m | m) = \lim_{n \rightarrow \infty} \frac{1}{n} E_{p^n} \left[\frac{\partial \log p(X^n | \theta_m, m)}{\partial \theta_m} \frac{\partial \log p(X^n | \theta_m, m)}{(\partial \theta_m)^T} \right]. \quad (2.40)$$

定理 2.4.2 より, モデル未知の確率モデルに基づくベイズ符号の冗長度は, $o(1)$ までの精密な評価が行われており, かつ $O(k_{m^*} \log n)$ で 0 に収束する, すなわち, 1 シンボルあたりの平均符号語長は $O(k_{m^*} \log n/n)$ でエントロピーレートに収束する事が分かる. その収束オーダーは真のモデルのパラメータの次数に依存し, さらに, 真のモデルや真のパラメータに依存する誤差項が現れているのが見て取れる.

一方, \mathcal{H}_{m^*} を $\mathcal{H}_{m'}$ の部分集合とすると, 定理 2.4.2 より, $Ap^*(x^n)$ を符号化確率として用いたベイズ符号における冗長度は $O(k_{m^*} \log n)$ で 0 に収束するが, モデル m' を既知として一つに固定した下で $p(x^n | m')$ を符号化確率としてモデル既知の確率モデルに基づくベイズ符号を適用した場合, 冗長度は $O(k_{m'} \log n)$ で 0 に収束する. したがって, m^* の定義より $k_{m^*} < k_{m'}$ である事から, モデル未知の確率モデルに基づくベイズ符号は, より次数の高いモデルを既知として一つに固定した下でのモデル既知の確率モデルに基づくベイズ符号と比較し, 平均符号語長のエントロピーレートへの収束オーダーが小さくなるといった意味でより良い性能を発揮する.

2.4.6 FSMX 情報源に対する効率的なベイズ符号化アルゴリズム

本節では, 2.3.6 節までの問題設定を考える.

従来, 2.3.4 節で定義された FSMX 情報源に対する効率的なベイズ符号化アルゴリズムが提案されている. ベイズ符号においては, 2.4.2 節で述べた通り, ベイズ基

準の下で最適な符号化確率を決定した上で算術符号化を行う事を考えるため、この符号化確率の決定が主要な問題となる。

FSMX 情報源において、モデルおよびその下でのパラメータは未知であるが、それらに関する事前確率関数 $P(m)$ および事前確率密度 $w(\theta_m | m)$ はそれぞれ既知であるとする。すると、FSMX 情報源から出力される情報源系列 x^{t-1} を得た下で、次の時点に発生するシンボル x_t のベイズ最適な逐次型の符号化確率は次のように与えられる [15]。

$$Ap^*(x_t | x^{t-1}) = \sum_{m \in \mathcal{M}} P(m | x^{t-1}) \int_{\theta_m \in \Theta_m} p(x_t | x^{t-1}, \theta_m, m) w(\theta_m | x^{t-1}, m) d\theta_m. \quad (2.41)$$

ただし、 $P(m | x^{t-1})$ および $w(\theta_m | x^{t-1}, m)$ はそれぞれ、 x^{t-1} を得た下でのモデル m の事後確率関数および x^{t-1}, m を得た下でのパラメータ θ_m の事後確率密度を表す。この符号化確率を効率的に計算するアルゴリズムについて述べていく。

まず、パラメータの事後確率密度に関する積分の部分は自然共役な事前確率密度を仮定する事により、代数的な計算に帰着する事ができ、その結果、時間計算量は $O(t)$ まで削減する事ができる。一方、モデルの事後確率関数に関する和計算の部分については、各時点 t において $O(2^{t-1})$ の時間計算量を要し、これは系列長と共に指数的に増大してしまう。そこで、モデルに関する事前確率関数にある特別な仮定を置く事により、この時間計算量を $O(t)$ まで削減する方法 [16] について述べる。

従来の方法では、各時点において文脈木を生成し、その文脈木を用いて符号化確率の計算が行われている。

文脈木の生成については次のように行う。部分系列を $x_i^j = x_i \cdots x_j$ とする。時点 t までに発生した情報源系列 x^t に対し、その文脈に対応するノード $s(\lambda), s(x_{t-1}^t), \cdots, s(x_1^t)$ の内、すでに文脈木に含まれるノードを除いたノードを文脈木に加える。時点 t までの文脈に対応するノードの集合を S_t と書く事にする。すなわち、 $S_t = \{s(\lambda), s(x_{t-1}^t), \cdots, s(x_1^t)\}$ である。このようにして生成される文脈木は、時点 t までに発生した情報源系列 x^t から考え得る全ての状態の集合である。

符号化確率の計算に関し、従来、文脈木における一つのノードの事前確率関数にそのノードを状態として含む FSMX 情報源のモデルの事前確率関数を対応させた次

の式が提案されている [16].

$$P(s) = \sum_{\{m|s \in S(m)\}} P(m). \quad (2.42)$$

この式を用いれば、符号化確率の計算において一つの文脈に属するノードのみを考えれば良い事になる⁸. 式 (2.42) およびパラメータの事前確率密度としてディリクレ分布の確率密度を仮定した場合、符号化確率は自然共役性から次のように計算される.

$$Ap(x_t | x^{t-1}) = \sum_{s \in S_t} p(x_t | x^{t-1}, s) P(s | x^{t-1}), \quad (2.43)$$

$$p(x_t | x^{t-1}, s) = \frac{n(x_t | x^{t-1}) + \beta(x_t | s)}{\sum_{a=0}^{|\mathcal{X}|-1} (n(a | x^{t-1}) + \beta(a | s))}. \quad (2.44)$$

ただし、 $n(a | x^{t-1})$ は状態 s の下での情報源シンボル $a \in \mathcal{X}$ が発生した回数であり、 $\beta(a | s)$ はディリクレ分布のパラメータである.

生成した文脈木と式 (2.43) を用いれば時点 t における文脈のみを考えて符号化確率を計算する事ができる. 一方、ノードの事後確率関数は以下のように更新される⁹.

$$P(s | x^t) = \frac{q(s | x^t)}{\sum_{s \in S_t} q(s | x^t)}, \quad (2.45)$$

$$q(s | x^t) = \begin{cases} \frac{p(x_t | x^{t-1}, s) P(s | x^{t-1})}{Ap(x_t | x^{t-1})}, & (s \in S_t) \\ P(s | x^{t-1}). & (\text{otherwise}) \end{cases} \quad (2.46)$$

ただし、 $P(s | x^0) = P(s)$ である.

モデルの事後確率関数の和計算に関する時間計算量は各時点 t において、モデル m の総数が 2^{t-1} である事から当初は $O(2^{t-1})$ であったところが $O(t)$ まで削減されている事が分かる.

2.4.7 区分定常情報源に対する効率的なベイズ符号化アルゴリズム

本節では、2.3.6 節までの問題設定を考える.

⁸式 (2.42) において、 $\sum_{s \in S_t} P(s) = 1$ が満たされる事に注意する.

⁹事後確率関数の更新式に関し、 $P(s | x^t) = \sum_{\{m|s \in S(m)\}} P(m | x^t)$ が成り立つ事に注意する.

従来, 2.3.5 節で定義された区分定常情報源に対する効率的なベイズ符号化アルゴリズムが提案されている¹⁰. ベイズ符号においては, 2.4.2 節で述べた通り, ベイズ基準の下で最適な符号化確率を決定した上で算術符号化を行う事を考えるため, この符号化確率の決定が主要な問題となる.

区分定常情報源において, 変化パターンおよびその下でのパラメータは未知であるが, それらに関する事前確率関数 $P(\omega)$ および事前確率密度 $w(\phi_\omega | \omega)$ はそれぞれ既知であるとする. すると, 区分定常情報源から出力される情報源系列 y^{t-1} を得た下で, 次の時点に発生するシンボル y_t のベイズ最適な逐次型の符号化確率は次のように与えられる [15].

$$Ap^*(y_t | y^{t-1}) = \sum_{\omega \in \Omega} P(\omega | y^{t-1}) \int_{\phi_\omega \in \Phi_\omega} p(y_t | y^{t-1}, \phi_\omega, \omega) w(\phi_\omega | y^{t-1}, \omega) d\phi_\omega. \quad (2.47)$$

ここで, $P(\omega | y^{t-1})$ および $w(\phi_\omega | y^{t-1}, \omega)$ はそれぞれ, y^{t-1} を得た下での変化パターン ω の事後確率関数および y^{t-1}, ω を得た下でのパラメータ ϕ_ω の事後確率密度を表す. 前節と同様, この符号化確率を効率的に計算するアルゴリズムについて述べていく.

まず, パラメータの事後確率密度に関する積分の部分は自然共役な事前確率密度を仮定する事により, 代数的な計算に帰着する事ができ, その結果, 時間計算量は $O(t)$ まで削減する事ができる. 一方, 変化パターンの事後確率関数に関する和計算の部分については, 各時点 t において $O(2^{t-1})$ の時間計算量を要し, これは系列長と共に指数的に増大してしまう. そこで, 変化パターンに関する事前確率関数にある特別な仮定を置く事により, この時間計算量を $O(t)$ まで削減する方法 [21] について述べる.

時点 1 から t までの間に最後にパラメータが変化した時点を τ_t とし, $\omega_t = (t_{\omega,1}, \dots, t_{\omega,C(\omega_t)})$ を $t_{\omega,C(\omega_t)} = \tau_t$ によって定義する. これは, ある変化パターンの部分パターン, すなわち, ω の中で t 時点までの変化時点を並べたものとなっている. この

¹⁰区分定常情報源をモデル未知の確率モデルとみなした場合, これは統計学における変化点検出の問題と同様であるが, 変化点を検出するという目的と平均符号語長を最小にする符号化を行うという目的とでは, 評価関数や損失関数が異なる. したがってその目的に応じて異なった手法が必要であり, 目的を達成するための最適なアルゴリズムも異なってくる.

部分パターンに対するパラメータ $\phi_{\omega_t} \in \Phi_{\omega_t}$ も同様に定義する. すると, ベイズの定理により, ある部分パターン ω_t の事後確率関数は以下のように計算される.

$$P(\omega_t | y^t) = \frac{P(\omega_t | y^{t-1})p(y_t | y^{t-1}, \omega_t)}{Ap^*(y_t | y^{t-1})}, \quad (2.48)$$

$$p(y_t | y^{t-1}, \omega_t) = \int_{\phi_{\omega_t} \in \Phi_{\omega_t}} p(y_t | y^{t-1}, \phi_{\omega_t}, \omega_t)w(\phi_{\omega_t} | y^{t-1}, \omega_t) d\phi_{\omega_t}. \quad (2.49)$$

ここで, ある部分パターン ω_{t+1} は次の 2 つの場合に分ける事ができる. 一方は部分パターン ω_t において時点 $t+1$ でパラメータの変化が起きる場合の ω_{t+1}^a であり, もう一方は変化が起きない場合の ω_{t+1}^b である. 時点 $t+1$ においてパラメータの変化が起きる確率を π とすれば, ある部分パターン ω_{t+1} の事後確率 $P(\omega_{t+1} | y^t)$ は次のように計算される.

$$P(\omega_{t+1}^a | y^t) = \pi P(\omega_t | y^t), \quad (2.50)$$

$$P(\omega_{t+1}^b | y^t) = (1 - \pi)P(\omega_t | y^t). \quad (2.51)$$

従来, 時点 t までの間に最後にパラメータが変化した時点が i である確率に関し, 次のような仮定が用いられている [21]. ただし, $i = 1, \dots, t$ である.

$$P(\tau_t = i) = \sum_{\{\omega | t_{C(\omega_t)} = i\}} P(\omega). \quad (2.52)$$

この仮定を用いた効率的なベイズ符号化アルゴリズムは次のように記述される.

符号化確率の計算は次のように行われる.

$$AP(y_t | y^{t-1}) = \sum_{i=1}^t p(y_t | y^{t-1}, \tau_t = i)P(\tau_t = i | y^{t-1}), \quad (2.53)$$

$$p(y_t | y^{t-1}, \tau_t = i) = \int_{\phi_{\tau_t}} p(y_t | y^{t-1}, \phi_{\tau_t}, \tau_t = i)w(\phi_{\tau_t} | y^{t-1}, \tau_t = i) d\phi_{\tau_t}. \quad (2.54)$$

式 (2.54) は, y_t が独立定常分布に従って発生し, さらに, パラメータの事前確率密度としてディリクレ分布の確率密度を仮定した場合は自然共役性から次のように計

Algorithm 1 従来アルゴリズム [21]

```

1: Input  $n$ 
2: for  $t = 1$  to  $n$  do
3:   calculate  $AP(y_t | y^{t-1})$  by Eq. (2.53)
4:   for  $i = 1$  to  $t$  do
5:     calculate  $P(\tau_{t+1} = i | y^t)$  by Eq. (2.57)
6:   end for
7:   calculate  $P(\tau_{t+1} = t + 1 | y^t)$  by Eq. (2.58)
8: end for
9: Output  $AP(y_n | y^{n-1})$ 

```

算される.

$$\begin{aligned}
& p(y_t | y^{t-1}, \tau_t = i) \\
&= \frac{\nu(y_t | y_i^{t-1}) + \beta(y_t | \tau_t = i)}{\sum_{a=0}^{|\mathcal{X}|-1} (\nu(a | y_i^{t-1}) + \beta(a | \tau_t = i))}. \tag{2.55}
\end{aligned}$$

ただし, $\nu(a | y_i^{t-1})$ は時点 i から $t-1$ までの間に情報源シンボル $a \in \mathcal{X}$ が発生した回数であり, $\beta(a | \tau_t = i)$ はディリクレ分布のパラメータである.

一方, 事後確率は次のように計算される.

$$P(\tau_{t+1} = i | y^t) = (1 - \pi)P(\tau_t = i | y^t) \tag{2.56}$$

$$= (1 - \pi) \frac{p(y_t | y^{t-1}, \tau_t = i)P(\tau_t = i | y^{t-1})}{AP(y_t | y^{t-1})}, \tag{2.57}$$

$$P(\tau_{t+1} = t + 1 | y^t) = \pi. \tag{2.58}$$

従来アルゴリズムにおいて計算される符号化確率はベイズ最適な符号化確率である事が次の定理により保証される.

定理 2.4.3 [21] 従来アルゴリズムにおいて次の式が成り立つ.

$$AP(y_t | y^{t-1}) = AP^*(y_t | y^{t-1}). \tag{2.59}$$

すなわち、従来アルゴリズムにおいて計算される符号化確率はベイズ最適な逐次型の符号化確率と等しい。ただし、 $AP(y_t | y^{t-1})$ は従来アルゴリズムで時点 t において計算される符号化確率である。

(証明) まず、式 (2.18), (2.52) および (2.57) より、

$$P(\tau_t = i | y^{t-1}) = \sum_{\{\omega | t_{C(\omega_t)}=i\}} P(\omega | y^{t-1}), \quad (2.60)$$

$$p(y_t | y^{t-1}, \omega) = p(y_t | y^{t-1}, \tau_t = i), \quad (2.61)$$

が成り立つ。したがって、

$$AP^*(y_t | y^{t-1}) = \sum_{\omega \in \Omega} p(y_t | y^{t-1}, \omega) P(\omega | x^{t-1}) \quad (2.62)$$

$$= \sum_{i=1}^t \sum_{\{\omega | t_{C(\omega_t)}=i\}} p(y_t | y^{t-1}, \omega) P(\omega | y^{t-1}) \quad (2.63)$$

$$= \sum_{i=1}^t \sum_{\{\omega | t_{C(\omega_t)}=i\}} p(y_t | y^{t-1}, \tau_t = i) P(\omega | y^{t-1}) \quad (2.64)$$

$$= \sum_{i=1}^t p(y_t | y^{t-1}, \tau_t = i) \sum_{\{\omega | t_{C(\omega_t)}=i\}} P(\omega | y^{t-1}) \quad (2.65)$$

$$= \sum_{i=1}^t p(y_t | y^{t-1}, \tau_t = i) P(\tau_t = i | y^{t-1}) \quad (2.66)$$

$$= AP(y_t | y^{t-1}). \quad (2.67)$$

を得る。ただし、 $AP(y_t | y^{t-1})$ は従来アルゴリズムで時点 t において計算される符号化確率である。□

従来アルゴリズムにおいて、変化パターンの事後確率関数の和計算に関する時間計算量は各時点 t において、部分パターン ω_t の総数が 2^{t-1} である事から当初は $O(2^{t-1})$ であったところが $O(t)$ まで削減されている事が分かる。

第3章 表現不可能な確率モデルに基づく ベイズ符号の性能解析

3.1 はじめに

本章では，本論文の一方の成果である情報源の確率分布を表現できない確率モデルに基づくベイズ符号の性能解析について述べる．まず，確率モデルの表現可能性の定義およびユニバーサル符号化問題における表現不可能な確率モデルについて議論する重要性を述べ，統計的予測問題においてデータが従う確率分布を仮定した確率モデルでは表現できない場合を考慮した問題を扱った従来研究について述べた後，本論文の位置づけについて述べる．次に，本論文の一つの成果である，モデル既知の確率モデルを仮定したベイズ符号において情報源の確率分布がその確率モデルでは表現されない分布だった場合の冗長度すなわち平均符号語長の精密な漸近評価について述べる．さらに，その結果を応用し，階層型の入れ子構造を持つモデル未知の確率モデルを仮定したベイズ符号において，情報源の確率分布がその確率モデルでは表現されない分布だった場合の冗長度の漸近評価を厳密に行い，そのように仮定が崩れた場合にもベイズ符号が有効である事について述べる．具体的には，モデル未知の確率モデルに基づくベイズ符号は，そのように仮定が崩れた場合であっても最高次の確率モデルのみに基づくベイズ符号と比較し，平均符号語長の意味で同等あるいはそれ以上の性能を発揮する事を示す．

3.2 確率モデルの表現可能性

モデル既知の確率モデルの場合は任意のパラメータ θ_m に対し，モデル未知の確率モデルの場合は任意のモデル m およびパラメータ θ_m に対し，それらによって規

定される確率分布がデータが従う確率分布と等しくならない場合、仮定した確率モデルではデータが従う確率分布を表現できないと呼ぶ事にする。一方、モデル既知の確率モデルの場合はある θ_m により、モデル未知の確率モデルの場合はあるモデル m およびあるパラメータ θ_m により規定される確率分布がデータが従う確率分布と等しくなる場合、仮定した確率モデルはデータが従う確率分布を表現できると呼ぶ事にする。例えば、確率モデルはマルコフ過程であると仮定されているが、次数 k が未知、つまり何次のマルコフ過程かは未知であるといったモデル未知の確率モデルに対し、任意の次数および任意の遷移確率パラメータに対して、それらによって規定される確率分布がデータが従う確率分布と等しくならない場合、データが従う確率分布を仮定した確率モデルでは表現できない事になる。

ユニバーサル符号化の問題においては先に述べた通り、未知の情報源の確率分布に対し、平均符号語長がエントロピーレートに収束する符号を構成する事が主な目的とされているが、これは、モデル既知あるいはモデル未知の確率モデルを仮定した上で、情報源の確率分布がどのようなモデルあるいはどのようなパラメータにより表される確率分布であったとしても、平均符号語長に関するそのような性質を満たす事が望まれている。つまり、仮定した確率モデルは情報源の確率分布を表現できる事が前提とされている。しかし、このような前提の下で平均符号語長がそのような性質を満たすユニバーサル符号に対し、確率モデルが表現できるとしたその前提が実際には成り立たなかった場合の符号化性能を評価する事も重要であると考えられる。実用的な場面においては、情報源系列の観測を始める前に確率モデルを仮定し、そのモデルに基づき、以降は情報源系列を観測しながら符号化を行っていくため、確率モデルが情報源の確率分布を表現できない場合は往々にして起こり得るはずである。しかし、これまでのところ、そのような問題設定についてはほとんど議論されていない。

従来、統計学におけるモデル選択問題に関する研究では、データが従う確率分布を表現できない確率モデルについての研究が一部で行われており、例えば、モデル選択基準の一つである AIC [1] をデータが従う確率分布を表現できない確率モデルまで拡張した TIC が提案されている [23]。AIC や TIC の解析においては、パラメータの最尤推定量が用いられているが、仮定したモデル既知の確率モデルがデータが従う確率分布を表現できる場合、最尤推定量は、ある正則条件の下、その確率分布の

パラメータに概収束し、漸近正規性を持つ事が知られている [14]. 一方、データが従う確率分布を仮定したモデル既知の確率モデルでは表現できない場合、最尤推定量は、先の正則条件に対応する適切な条件の下、同様の漸近的性質を持つ事が知られている. すなわち、最尤推定量は、仮定した確率モデルの中の確率分布の内、データが従う確率分布とのカルバックライブラー情報量を最小にする確率分布のパラメータに概収束し、漸近正規性を持つ事が知られている [17, 27]. 本論文の解析においても同様に最尤推定量の漸近的性質を利用しているため、まずはそれについて述べていく.

本節では、データが取り得る値の集合である \mathcal{X} については 2.4.1 節で述べた設定で議論を進める. まず、いくつかの諸定義について述べる. はじめに、最尤推定量とフィッシャー情報行列を定義する. 次に、データが従う確率分布とモデル未知の確率モデルに含まれるある確率分布とのカルバックライブラー情報量を定義する. 最後に、モデル既知の確率モデル \mathcal{H}_m あるいはモデル未知の確率モデル \mathcal{H} に含まれる確率分布の内、データが従う確率分布にカルバックライブラー情報量の意味で最も近い確率分布について定義する.

定義 3.2.1 情報源系列 x^n が与えられた下での対数尤度関数を $\log p(x^n | \theta_m, m)$ と定義し、各 $m \in \mathcal{M}$ に対し、最尤推定量 $\hat{\theta}_m$ を次のように定義する.

$$\hat{\theta}_m = \arg \max_{\theta_m \in \Theta_m} \log p(x^n | \theta_m, m). \quad (3.1)$$

各 $m \in \mathcal{M}$ に対し、フィッシャー情報行列 $\hat{J}(\theta_m | m)$, $J(\theta_m | m)$ および $I(\theta_m | m)$ を次のように定義する.

$$\hat{J}(\theta_m | m) = -\frac{1}{n} \frac{\partial^2 \log p(x^n | \theta_m, m)}{\partial \theta_m (\partial \theta_m)^T}, \quad (3.2)$$

$$J(\theta_m | m) = \lim_{n \rightarrow \infty} \frac{1}{n} E_{p^n} \left[-\frac{\partial^2 \log p(X^n | \theta_m, m)}{\partial \theta_m (\partial \theta_m)^T} \right], \quad (3.3)$$

$$I(\theta_m | m) = \lim_{n \rightarrow \infty} \frac{1}{n} E_{p^n} \left[\frac{\partial \log p(X^n | \theta_m, m)}{\partial \theta_m} \frac{\partial \log p(X^n | \theta_m, m)}{(\partial \theta_m)^T} \right], \quad (3.4)$$

各 $m \in \mathcal{M}$ に対し、 $J(\theta_m | m)$ や $I(\theta_m | m)$ に表れている極限は存在するものとする. また、 $I(\theta_m | m)$ の定義は 2.4.5 節で述べた定義に等しい.

定義 3.2.2 各 $m \in \mathcal{M}$ に対し, $p = p(x^n)$ と $p_{\theta_m, m} = p(x^n | \theta_m, m)$ との間のカルバックライブラー情報量 $D(p \| p_{\theta_m, m})$ を次のように定義する.

$$D(p \| p_{\theta_m, m}) = E_{p^n} \left[\log \frac{p(X^n)}{p(X^n | \theta_m, m)} \right]. \quad (3.5)$$

定義 3.2.3 各 $m \in \mathcal{M}$ に対し, カルバックライブラー情報量 $D(p \| p_{\theta_m, m})$ を最小にするパラメータ θ_m^0 を次のように定義する.

$$\theta_m^0 = \arg \min_{\theta_m \in \Theta_m} D(p \| p_{\theta_m, m}), \quad (3.6)$$

また, m^0 を次のように定義する.

$$m^0 = \arg \min_{m \in \mathcal{M}} \left\{ k_m \mid \forall m' \neq m \in \mathcal{M}, D(p \| p_{\theta_m^0, m}) \leq D(p \| p_{\theta_{m'}, m'}) \right\}. \quad (3.7)$$

ここで, $p(x^n | \theta_m^0, m)$ は各 $m \in \mathcal{M}$ に対し, \mathcal{H}_m に含まれる確率分布の内, $p(x^n)$ に最も近い確率分布の確率密度関数あるいは確率質量関数である. さらに, $p(x^n | \theta_{m^0}^0, m^0)$ は \mathcal{H}_{m^0} に含まれる確率分布の内, $p(x^n)$ に最も近い確率分布の中でも最小次数の確率分布の確率密度関数あるいは確率質量関数である.

次に, 後に示す最尤推定量の漸近的性質に関する補題が成り立つための諸条件について述べる. 次に述べる条件は, 後の主定理を示すためにも必要な条件となる.

条件 3.2.1 全ての $m \in \mathcal{M}$ に対し, Θ_m はコンパクトであり, θ_m^0 は Θ_m の内部に唯一存在する¹.

条件 3.2.2 データが従う確率分布 $p(x^n)$ に関し, $\log p(x^n)$ の $p(x^n)$ に関する期待値が存在する. また, $\log p(x^n | \theta_m, m)$ は θ_m について C^3 級であり, 各偏導関数, $\log p(x^n)$, および,

$$\frac{\partial \log p(x^n | \theta_m, m)}{\partial \theta_m} \cdot \frac{\partial \log p(x^n | \theta_m, m)}{(\partial \theta_m)^T}, \quad (3.8)$$

は全ての $x^n \in \mathcal{X}^n$ と全ての $m \in \mathcal{M}$ に対し, 有界である. 一方, 最尤推定量 $\hat{\theta}_m$ に関し, 任意の n について $\hat{J}(\hat{\theta}_m | m)$ の行列式および

$$|n(\hat{\theta}_m - \theta_m^0)^T \hat{J}(\hat{\theta}_m | m)(\hat{\theta}_m - \theta_m^0)|, \quad (3.9)$$

¹一般に, θ_m^0 は各 $m \in \mathcal{M}$ に対し, 常に唯一存在するとは限らない. 例えば, \mathcal{H}_m が混合正規分布の確率モデルである場合等が挙げられる.

は有界である. さらに, $w(\theta_m | m)$ は全ての $m \in \mathcal{M}$ に対し, θ_m について C^2 級であり, 任意の n について $|w(\hat{\theta}_m | m)|$ は有界である.

条件 3.2.3 情報行列 $I(\theta_m | m)$ および $J(\theta_m | m)$ が存在し, かつそれらは正則行列である. また, $J(\theta_m | m)$ は全ての $m \in \mathcal{M}$ に対し, θ_m^0 のある開近傍で行列の階数が一定である.

条件 3.2.4 全ての $\theta_m \in \Theta_m$ および全ての $m \in \mathcal{M}$ に対し, $w(\theta_m | m) > 0$ および $P(m^0) > 0$ が成り立つ.

注意 3.2.1 例えば, \mathcal{X} が離散有限集合で \mathcal{H} が有限次数の定常エルゴードマルコフ過程の確率モデルであり, 全ての $x \in \mathcal{X}$ に対し, その確率質量関数が 0 または 1 にならず, $w(\theta_m | m)$ がディリクレ分布の確率密度である場合, 条件 3.2.1–3.2.4 は満たされる [10].

以上の条件の下で, 重複対数の法則 [5, 17, 27] より, 最尤推定量の漸近的性質に関する次の補題が成り立つ.

補題 3.2.1 条件 3.2.1–3.2.3 が成り立つならば, 全ての $m \in \mathcal{M}$ に対し, 次式が成り立つ².

$$\hat{\theta}_m = \theta_m^0 + O\left(\sqrt{\frac{\log \log n}{n}}\right), \text{ a.s.}, \quad (3.10)$$

$$\hat{J}(\hat{\theta}_m | m) = J(\theta_m^0 | m) + O\left(\sqrt{\frac{\log \log n}{n}}\right), \text{ a.s.} \quad (3.11)$$

また, 最尤推定量の漸近正規性に関する次の補題が成り立つ.

補題 3.2.2 [5, 17, 27] 確率変数 S_n を $\sqrt{n}(\hat{\theta}_{m^0} - \theta_{m^0}^0)$ と定義すると, S_n は, 条件 3.2.1–3.2.3 の下, 平均ベクトル $\mathbf{0}$ および分散共分散行列 $J(\theta_{m^0}^0 | m^0)^{-1}I(\theta_{m^0}^0 | m^0)J(\theta_{m^0}^0 | m^0)^{-1}$ を持つ多変量正規分布に分布収束する.

² $\Pr\{\lim_{n \rightarrow \infty} X_n = X\} = 1$ が成り立つならば確率変数列 X_n は X に概収束するといいい, $X_n \rightarrow X, \text{ a.s.}$ と表す.

表 3.1: 確率モデルの種類と表現可能性による分類

確率モデル	表現できる	表現できない
モデル既知	[6] (2.4.4 節)	本論文 (3.4 節)
モデル未知	[10] (2.4.5 節)	本論文 (3.5 節)

3.3 本章の位置づけ

前節で述べた通り、統計的予測問題においては、データが従う確率分布を仮定した確率モデルでは表現できない場合についてしばしば議論されるが、その一方でユニバーサル情報源符号化においては、確率モデルに基づく符号化においてさえも情報源の確率分布を表現できない場合についての議論はあまりなされていない。

本章で述べる研究成果の位置づけをベイズ符号の性能解析という視点で、確率モデルの種類および表現可能性の二つの軸で整理したのが表 3.1 である。

3.4 モデル既知の表現不可能な確率モデルに対する解析

本節では、モデル既知の確率モデルを仮定したベイズ符号において情報源の確率分布がその確率モデルでは表現されない分布だった場合のベイズ最適な決定におけるカルバックライブラー情報量の漸近評価を行う。本節以降もデータが取り得る値の集合である \mathcal{X} については 2.4.1 節で述べた設定で議論を進める。本節で扱う確率モデルはモデル既知のものであるため、2.3.3 節や 2.4.1 節で定義されているものに関し、 m はある一つの値に固定しておく。

まず、モデル既知の確率モデルに基づくベイズ最適な決定 $p(x^n | m)$ の漸近式を示す次の補題を用意する。

補題 3.4.1 [6],[26] 条件 3.2.1–3.2.4 が成り立つならば、次の式が成り立つ。

$$-\log p(x^n | m) = -\log p(x^n | \hat{\theta}_m, m) + \frac{k_m}{2} \log \frac{n}{2\pi} + \log \frac{\sqrt{\det \hat{J}(\hat{\theta}_m | m)}}{w(\hat{\theta}_m | m)} + o(1). \quad (3.12)$$

(証明) 文献 [6],[26] を参照. \square

モデル既知の確率モデルに基づくベイズ最適な決定におけるカルバックライブラー情報量の漸近式を示すために補題 3.4.1 から導かれる次の二つの補題を用意する.

補題 3.4.2 条件 3.2.1–3.2.4 が成り立つならば, 次の式が成り立つ.

$$E_{p^n} \left[\log \frac{p(X^n | \hat{\theta}_m, m)}{p(X^n | m)} \right] = \frac{k_m}{2} \log \frac{n}{2\pi} + \log \frac{\sqrt{\det J(\theta_m^0 | m)}}{w(\theta_m^0 | m)} + o(1), \quad (3.13)$$

(証明) 3.6.1 節を参照. \square

補題 3.4.3 条件 3.2.1–3.2.3 が成り立つならば, 次の式が成り立つ.

$$E_{p^n} \left[\log \frac{p(X^n | \hat{\theta}_m, m)}{p(X^n | \theta_m^0, m)} \right] = \frac{1}{2} \text{tr} (I(\theta_m^0 | m) J(\theta_m^0 | m)^{-1}) + o(1), \quad (3.14)$$

(証明) 3.6.2 節を参照. \square

モデル既知の確率モデルに基づくベイズ最適な決定におけるカルバックライブラー情報量の漸近式は次の定理によって与えられる.

定理 3.4.1 条件 3.2.1–3.2.4 が成り立つならば, 次の式が成り立つ.

$$\begin{aligned} E_{p^n} \left[\log \frac{p(X^n)}{p(X^n | m)} \right] &= \frac{k_m}{2} \log \frac{n}{2\pi} + \log \frac{\sqrt{\det J(\theta_m^0 | m)}}{w(\theta_m^0 | m)} \\ &\quad - \frac{1}{2} \text{tr} (I(\theta_m^0 | m) J(\theta_m^0 | m)^{-1}) + D(p || p_{\theta_m^0, m}) + o(1), \end{aligned} \quad (3.15)$$

が成り立つ.

(証明)

$$\begin{aligned} E_{p^n} \left[\log \frac{p(X^n)}{p(X^n | m)} \right] &= \\ E_{p^n} \left[\log \frac{p(X^n | \hat{\theta}_m, m)}{p(X^n | m)} \right] &+ E_{p^n} \left[\log \frac{p(X^n | \theta_m^0, m)}{p(X^n | \hat{\theta}_m, m)} \right] + E_{p^n} \left[\log \frac{p(X^n)}{p(X^n | \theta_m^0, m)} \right], \end{aligned} \quad (3.16)$$

とし、補題 3.4.2 と補題 3.4.3 を用いる。□

定理 3.4.1 より、モデル既知の確率モデルに基づくベイズ符号において情報源の確率分布がその確率モデルでは表現されない分布だった場合、冗長度は $o(1)$ までの精密な評価がなされており、かつ $O(k_m \log n)$ で $D(p \| p_{\theta_m^0, m})$ に収束する、すなわち、1 シンボルあたりの平均符号語長は $O(k_m \log n/n)$ でエントロピーレートには収束せず、カルバックライブラー情報量の意味で情報源の確率分布に最も近い確率分布と情報源の確率分布とのカルバックライブラー情報量の分だけ冗長となる事が分かる。したがって、この場合のベイズ符号は、カルバックライブラー情報量の意味で情報源の確率分布に最も近い確率分布を符号化確率として用いた符号化と漸近的には同等の性能を発揮する事が分かる。

3.5 階層型のモデル未知の表現不可能な確率モデルに対する解析

本節では、2.4.1 節で述べた問題設定の下、前節で得られた結果を応用し、階層型の入れ子構造を持つモデル未知の確率モデルを仮定したベイズ最適な決定において、情報源の確率分布がその確率モデルでは表現されない分布だった場合の対数損失やカルバックライブラー情報量の漸近評価を厳密に行い、そのように仮定が崩れた場合にもベイズ符号が有効である事について述べる。具体的には、モデル未知の確率モデルに基づくベイズ符号は、そのように仮定が崩れた場合であっても最高次の確率モデルのみに基づくベイズ符号と比較し、冗長度すなわち平均符号語長の意味で同等あるいはそれ以上の性能を発揮する事を示す。

3.5.1 対数損失の漸近評価

本節では、入れ子構造を持つ階層型のモデル未知の確率モデルに基づくベイズ最適な決定における対数損失の漸近式を示す。

補題 3.2.1 および補題 3.4.1 から導かれる次の補題は、対数損失の漸近評価において非常に大きな役割を果たす。

補題 3.5.1 条件 3.2.1–3.2.4 が成り立つならば、次式が成り立つ.

$$\frac{\max_{m \neq m^0} p(x^n | m)}{p(x^n | m^0)} = o^+(1), \text{ a.s.} \quad (3.17)$$

ただし, $f_n = o^+(1)$ は関数列 $\{f_n\}$ に対し, $\lim_{n \rightarrow \infty} f_n = +0$ を表す.

(証明) 3.6.3 節を参照. □

対数損失の漸近式は次の定理によって与えられる.

定理 3.5.1 条件 3.2.1–3.2.4 が成り立つならば、次の式が成り立つ.

$$\begin{aligned} \log \frac{p(x^n)}{Ap^*(x^n)} &= \log \frac{p(x^n)}{p(x^n | \hat{\theta}_{m^0}, m^0)} \\ &+ \frac{k_{m^0}}{2} \log \frac{n}{2\pi} + \log \frac{\sqrt{\det \hat{J}(\hat{\theta}_{m^0} | m^0)}}{w(\hat{\theta}_{m^0} | m^0)} - \log P(m^0) + o(1), \text{ a.s.} \end{aligned} \quad (3.18)$$

(証明) 3.6.4 節を参照. □

定理 3.5.1 は入れ子構造を持つ階層型のモデル未知の確率モデルに基づくベイズ最適な決定における対数損失の漸近式を与えているが、主要項として $\log(p(x^n)/p(x^n | \hat{\theta}_{m^0}, m^0))$ および $(k_{m^0}/2) \log n$ が表れており、他に表れている項は定数項となっている.

3.5.2 カルバックライブラー情報量の漸近評価

本節では、入れ子構造を持つ階層型のモデル未知の確率モデルに基づくベイズ最適な決定におけるカルバックライブラー情報量の漸近式を示す. これはベイズ符号における漸近冗長度と等価であるため、本節の主旨は、モデル未知の確率モデルに基づくベイズ符号における平均符号語長の漸近評価について述べる事に相当する.

まず、補題 3.2.1 を用いる事により、定理 3.5.1 の系として次の系を得る.

系 3.5.1 条件 3.2.1–3.2.4 が成り立つならば、次の式が成り立つ.

$$\begin{aligned} E_{p^n} \left[\log \frac{p(X^n | \hat{\theta}_{m^0}, m^0)}{Ap^*(X^n)} \right] &= \frac{k_{m^0}}{2} \log \frac{n}{2\pi} + \log \frac{\sqrt{\det J(\theta_{m^0}^0 | m^0)}}{w(\theta_{m^0}^0 | m^0)} \\ &- \log P(m^0) + o(1). \end{aligned} \quad (3.19)$$

(証明) 3.6.5 節を参照. □

入れ子構造を持つ階層型のモデル未知の確率モデルに基づくベイズ最適な決定におけるカルバックライブラー情報量の漸近式は次式で与えられる.

定理 3.5.2 条件 3.2.1–3.2.4 が成り立つならば, 次式が成り立つ.

$$\begin{aligned}
 E_{p^n} \left[\log \frac{p(X^n)}{Ap^*(X^n)} \right] &= \frac{k_{m^0}}{2} \log \frac{n}{2\pi} + \log \frac{\sqrt{\det J(\theta_{m^0}^0 | m^0)}}{w(\theta_{m^0}^0 | m^0)} \\
 &\quad - \frac{1}{2} \text{tr} \left(I(\theta_{m^0}^0 | m^0) J(\theta_{m^0}^0 | m^0)^{-1} \right) - \log P(m^0) + D \left(p \| p_{\theta_{m^0}^0, m^0} \right) + o(1). \quad (3.20)
 \end{aligned}$$

ただし,

$$D \left(p \| p_{\theta_{m^0}^0, m^0} \right) = E_{p^n} \left[\log \frac{p(X^n)}{p(X^n | \theta_{m^0}^0, m^0)} \right], \quad (3.21)$$

である.

(証明)

$$\begin{aligned}
 E_{p^n} \left[\log \frac{p(X^n)}{Ap^*(X^n)} \right] &= E_{p^n} \left[\log \frac{p(X^n | \hat{\theta}_{m^0}, m^0)}{Ap^*(X^n)} \right] \\
 &\quad + E_{p^n} \left[\log \frac{p(X^n | \theta_{m^0}^0, m^0)}{p(X^n | \hat{\theta}_{m^0}, m^0)} \right] + E_{p^n} \left[\log \frac{p(X^n)}{p(X^n | \theta_{m^0}^0, m^0)} \right], \quad (3.22)
 \end{aligned}$$

とし, 系 3.5.1 および $m = m^0$ に対しても補題 3.4.3 が成り立つ事を用いる. □

定理 3.5.2 より, 入れ子構造を持つ階層型のモデル未知の確率モデルに基づくベイズ符号において情報源の確率分布がその確率モデルでは表現されない分布だった場合, 冗長度は, $o(1)$ までの精密な評価がなされており, かつ $O(k_{m^0} \log n)$ で $D(p \| p_{\theta_{m^0}^0, m^0})$ に収束する, すなわち, 1 シンボルあたりの平均符号語長は $O(k_{m^0} \log n/n)$ でエントロピーレートには収束せず, カルバックライブラー情報量の意味で情報源の確率分布に最も近い確率分布と情報源の確率分布とのカルバックライブラー情報量の分だけ冗長となる事が分かる. したがって, この場合のベイズ符号は, カルバックライブラー情報量の意味で情報源の確率分布に最も近い確率分布の内, 最小次数のものを符号化確率として用いた符号化と漸近的には同等の性能を発揮する事が分かる.

3.5.3 考察

本節では、定理 3.5.1 および定理 3.5.2 より導かれる結果について考察を行う。まず、 $Ap^*(x^n)$ を決定関数として用いた場合とある m が与えられた下で $p(x^n | m)$ を決定関数として用いた場合とでカルバックライブラー情報量の漸近的な性質について比較を行う。これは、入れ子構造を持つ階層型のモデル未知の確率モデルを仮定した場合に、その確率モデルに基づくベイズ符号を用いる場合とモデル未知の確率モデルの内のある 1 つの確率モデルを既知としたモデル既知の確率モデルに基づくベイズ符号を用いる場合との性能を比較するという事である。次に、前節で述べた結果が 2.4.5 節で述べた従来研究の主結果の一般化になっている事について述べる。

まず、情報源の確率分布を表現できないモデル未知の確率モデルを考える。さらに、この確率モデルは入れ子構造を持ち、 \tilde{m} は最大次数の確率モデル $\mathcal{H}_{\tilde{m}} \in \mathcal{H}$ を表すものとする。このような仮定の下では、 $p(x^n | \tilde{m})$ を決定関数として用いた場合、つまり最大次数の確率モデルを既知としたモデル既知の確率モデルに基づくベイズ符号を用いた場合の平均符号語長の収束オーダーは、 $Ap^*(x^n)$ を決定関数として用いた場合、つまりモデル未知の確率モデルに基づくベイズ符号を用いた場合より大きくなってしまふ。これは次のような理由による。もし、 $p(x^n | \tilde{m})$ を用いた場合、ベイズ最適な決定におけるカルバックライブラー情報量の漸近式の主要項の 1 つとして表れている $(k_{m^0}/2) \log n$ は $(k_{\tilde{m}}/2) \log n$ に等しくなる。よって、 \tilde{m} および m^0 の定義より、 $k_{\tilde{m}}$ は k_{m^0} と等しいかより大きくなるため、 $p(x^n | \tilde{m})$ を用いた場合の平均符号語長の収束オーダーは $(k_{\tilde{m}}/2) \log n$ となるが、 $Ap^*(x^n)$ を用いた場合の平均符号語長の収束オーダーは $(k_{m^0}/2) \log n$ となるため、前者の方が後者より大きくなるかあるいは両者とも等しくなる。ただし、 $k_{\tilde{m}} > k_{m^0}$ であれば $p(x^n | \tilde{m})$ を用いた場合の平均符号語長の収束オーダー $(k_{\tilde{m}}/2) \log n$ は $Ap^*(x^n)$ を用いた場合の $(k_{m^0}/2) \log n$ より真に大きくなる。さらに、 $k_{\tilde{m}} > k_{m^0}$ なる \tilde{m} に対し、 $p(x^n | \theta_{\tilde{m}}^0, \tilde{m})$ と $p(x^n | \theta_{m^0}^0, m^0)$ が共に $D(p||p_{\theta_m, m})$ を最小にする場合、 $p(x^n | \tilde{m})$ を用いた場合のカルバックライブラー情報量の漸近式に表れる $D(p||p_{\theta_{\tilde{m}}^0, \tilde{m}}^0)$ は、 $Ap^*(x^n)$ を用いた場合の漸近式に表れる $D(p||p_{\theta_{m^0}^0, m^0}^0)$ に等しくなる。このような場合が生じる具体例については 3.5.4 節で述べる。さらに、 $k_{m^0} < k_m < k_{\tilde{m}}$ であり、かつ $p(x^n | \theta_{m^0}^0, m^0)$ や $p(x^n | \theta_{\tilde{m}}^0, \tilde{m})$ だけでなく $p(x^n | \theta_m, m)$ も $D(p||p_{\theta_m, m})$ を最小にするような m についても同様の議

論が成り立つ.

一方, 逆に, $Ap^*(x^n)$ を用いた場合の平均符号語長の収束オーダーが $p(x^n | \tilde{m})$ を用いた場合より大きくなる事があるかどうかについて議論する. 先に述べたものと同様, $k_{\tilde{m}}$ を, $\mathcal{H}_{\tilde{m}}$ に含まれる, 情報源の確率分布にカルバックライブラー情報量の意味で最も近い確率分布の確率密度関数 (あるいは確率質量関数) $p(x^n | \theta_{\tilde{m}}^0, \tilde{m})$ が持つパラメータの次数とする. ただし, $p(x^n | \theta_{\tilde{m}}^0, \tilde{m})$ は, \tilde{m} の定義より \mathcal{H} にも含まれる. しかし, $\mathcal{H}_{\tilde{m}}$ は \mathcal{H} に含まれる最大次数の確率モデルであるため, $k_{m''} > k_{\tilde{m}}$ なる $p(x^n | \theta_{m''}^0, m'')$ は存在しない. ここで, $p(x^n | \theta_{m''}^0, m'')$ は情報源の確率分布に最も近い確率分布の確率密度関数 (あるいは確率質量関数) である. よって, $Ap^*(x^n)$ を用いた場合の平均符号語長の収束オーダーは $k_{m''}$ に依存する事はない. したがって, $Ap^*(x^n)$ を用いた場合の平均符号語長の収束オーダーが $p(x^n | \tilde{m})$ を用いた場合より大きくなる事はない.

最後に, モデル未知の確率モデルが情報源の確率分布を表現できる場合, 前節で述べた結果が 2.4.5 節で述べた従来研究の主結果に帰着される事について述べる. このような場合, $p(x^n)$ と $p(x^n | \theta_{m^*}^0, m^*)$ が等しくなるようなある m^* および $\theta_{m^*}^0$ が存在する. よって, m^0 は m^* に等しくなり, 式 (3.20) における $D(p || p_{\theta_{m^0}^0, m^0}^0)$ の項は 0 に等しくなる. また, $I(\theta_{m^0}^0 | m^0)$ は $J(\theta_{m^0}^0 | m^0)$ に等しくなるため, $I(\theta_{m^0}^0 | m^0)J(\theta_{m^0}^0 | m^0)^{-1}$ は単位行列に等しくなる. よって, $\text{tr}(I(\theta_{m^0}^0 | m^0)J(\theta_{m^0}^0 | m^0)^{-1})$ の値は k_{m^*} に等しくなる. したがって, 前節で述べた結果は, モデル未知の確率モデルが情報源の確率分布を表現できるかどうかという点について, 2.4.5 節で述べた従来研究の主結果の一般化になっている.

3.5.4 ベイズ符号の有効性を示す具体例

本節では, 3.5.3 節で述べた, $k_{\tilde{m}} > k_{m^0}$ なる \tilde{m} に対し, $p(x^n | \theta_{\tilde{m}}^0, \tilde{m})$ と $p(x^n | \theta_{m^0}^0, m^0)$ が共に $D(p || p_{\theta_{m,m}})$ を最小にする場合の具体例について述べる. まず, $\mathcal{X} = \{0, 1\}$ とし, \mathcal{H} を 1 次から 3 次までのマルコフ情報源の確率モデルとする. マルコフ情報源の定義については 2.3.4 節を参照されたい. ただし, $i = 1, 2, 3$ に対し, \mathcal{H}_{m_i} は i 次マルコフ情報源の確率モデルである. また, $Ap^*(x^n)$ および $p(x^n | m_i)$ をそれぞれ, モデル未知の確率モデル \mathcal{H} およびモデル既知の確率モデル \mathcal{H}_{m_i} に基づくベイズ

符号の決定関数とする. さらに, $p(x^n)$ を後述の仮定をみたす 4 次マルコフ情報源のある確率分布の確率質量関数とする. このような場合, モデル未知の確率モデルは情報源の確率分布を表現できない. 一方, x^t および $l = 1, 2, \dots$ に対し, $x_{t-l}^{t-1} = x_{t-l} \cdots x_{t-1}$ が与えられた下で次の時点に発生する情報源シンボル x_t が 0 に等しくなる確率を $\Pr(0 | x_{t-l}^{t-1})$ と表すものとする. 例えば, x^{t-1} が $\dots 0010$ に等しい場合, x_{t-4}^{t-1} が与えられた下で x_t が 0 に等しくなる確率は $\Pr(0 | 0010)$ である. ここで, $p(x^n)$ は次の式を満たすものとする.

$$\begin{aligned} \Pr(0 | 0000) &= \Pr(0 | 0100), \Pr(0 | 1000) = \Pr(0 | 1100), \\ \Pr(0 | 0010) &= \Pr(0 | 0110), \Pr(0 | 1010) = \Pr(0 | 1110), \\ \Pr(0 | 0001) &= \Pr(0 | 0101), \Pr(0 | 1001) = \Pr(0 | 1101), \\ \Pr(0 | 0011) &= \Pr(0 | 0111), \Pr(0 | 1011) = \Pr(0 | 1111). \end{aligned} \quad (3.23)$$

これは, 各時点 t において, $x_{t-1}, x_{t-2}, x_{t-4}$ の値が等しければ, x_t が発生する確率は x_{t-3} の値に依存しない事を意味する. このような場合, \mathcal{H}_{m_3} に含まれる, $p(x^n)$ にカルバックライブラー情報量の意味で最も近い確率分布の確率質量関数 $p(x^n | \theta_{m_3}^0, m_3)$ は次の式を満たさなければならない.

$$\begin{aligned} \Pr(0 | 000) &= \Pr(0 | 100), \Pr(0 | 010) = \Pr(0 | 110), \\ \Pr(0 | 001) &= \Pr(0 | 101), \Pr(0 | 011) = \Pr(0 | 111). \end{aligned} \quad (3.24)$$

これは, 各時点 t において, $p(x^n | \theta_{m_3}^0, m_3)$ は, x_t が発生する確率が x_{t-1}, x_{t-2} の値のみに依存するような確率質量関数である事を意味する. また, $p(x^n | \theta_{m_2}^0, m_2)$ は, \mathcal{H}_{m_2} に含まれる, $p(x^n)$ に最も近い確率分布の確率質量関数であり, 次の式をみたす.

$$\begin{aligned} \Pr(0 | 00) &= \Pr(0 | 000) = \Pr(0 | 100), \\ \Pr(0 | 10) &= \Pr(0 | 010) = \Pr(0 | 110), \\ \Pr(0 | 01) &= \Pr(0 | 001) = \Pr(0 | 101), \\ \Pr(0 | 11) &= \Pr(0 | 011) = \Pr(0 | 111). \end{aligned} \quad (3.25)$$

式 (3.24) および式 (3.25) より, $p(x^n | \theta_{m_3}^0, m_3)$ は $p(x^n | \theta_{m_2}^0, m_2)$ に等しくなる. よって, $p(x^n | \theta_{m_3}^0, m_3)$ が \mathcal{H} に含まれる, $p(x^n)$ に最も近い確率分布の確率質量関数である事から $p(x^n | \theta_{m_2}^0, m_2)$ も \mathcal{H} に含まれる, $p(x^n)$ に最も近い確率分布の確率質量関数である事が言える. つまり, \mathcal{H} に含まれる, $p(x^n)$ に最も近い確率分布の確率質量関数は \mathcal{H}_{m_2} および \mathcal{H}_{m_3} に存在するという事である. ただし, k_{m_2} は k_{m_3} より真に小さいため, 定義 3.2.3 より m^0 は m_2 に等しくなる. よって, $D(p||p_{\theta_{m^0}^0, m^0}^0) = D(p||p_{\theta_{m_2}^0, m_2}^0) = D(p||p_{\theta_{m_3}^0, m_3}^0)$ が成り立つ. また, k_{m^0} は k_{m_2} に等しくなる, すなわち, k_{m^0} は 2 次マルコフ情報源のパラメータの次数に等しくなる. したがって, $Ap^*(x^n)$ を用いた場合の平均符号語長の収束オーダーは, $p(x^n | m_3)$ を用いた場合に比べ, $((k_{m_3} - k_{m_2})/2) \log n$ だけ小さくなる事が分かる. ここで, m_3 は 3.5.3 節における \tilde{m} であり, k_{m_3} は 3 次マルコフ情報源のパラメータの次数に等しいため, $k_{m_3} > k_{m_2}$ すなわち $k_{\tilde{m}} > k_{m^0}$ である事が分かる. つまり, 本節で述べた例は, $k_{\tilde{m}} > k_{m^0}$ なる \tilde{m} に対し, $p(x^n | \theta_{\tilde{m}}^0, \tilde{m})$ と $p(x^n | \theta_{m^0}^0, m^0)$ が共に $D(p||p_{\theta_m, m})$ を最小にする場合の例であり, この時, $p(x^n | \tilde{m})$ を用いた場合のカルバックライブラー情報量の漸近式に表れる $D(p||p_{\theta_{\tilde{m}}^0, \tilde{m}}^0)$ は, $Ap^*(x^n)$ を用いた場合の漸近式に表れる $D(p||p_{\theta_{m^0}^0, m^0}^0)$ に等しくなる.

3.6 補題および定理の証明

3.6.1 補題 3.4.2 の証明

補題 3.4.1 より,

$$\log \frac{p(x^n | \hat{\theta}_m, m)}{p(x^n | m)} = \frac{k_m}{2} \log \frac{n}{2\pi} + \log \frac{\sqrt{\det \hat{J}(\hat{\theta}_m | m)}}{w(\hat{\theta}_m | m)} + o(1), \quad (3.26)$$

となる. ここで, 条件 3.2.4 より, $w(\theta_m | m)$ は連続であるから, 補題 3.2.1 より,

$$\log \frac{\sqrt{\det \hat{J}(\hat{\theta}_m | m)}}{w(\hat{\theta}_m | m)} = \log \frac{\sqrt{\det J(\theta_m^0 | m)}}{w(\theta_m^0 | m)} + o(1), \quad a.s., \quad (3.27)$$

が成り立つ. よって, 有界収束定理³を用いて題意を得る. \square

³確率変数列 X_n が X に概収束し, 任意の n に対し, $|X_n|$ が有界ならば, $\lim_{n \rightarrow \infty} E[X_n] = E[X]$ が成り立つ.

3.6.2 補題 3.4.3 の証明

$\nabla \log p(x^n | \hat{\theta}_m, m) = 0$ より, $\log p(x^n | \theta_m^0, m)$ にテーラーの定理を用いて,

$$\begin{aligned} \log p(x^n | \theta_m^0, m) &= \log p(x^n | \hat{\theta}_m, m) \\ &\quad - \frac{n}{2} (\hat{\theta}_m - \theta_m^0)^T \hat{J}(\hat{\theta}_m | m) (\hat{\theta}_m - \theta_m^0) + o(1), \end{aligned} \quad (3.28)$$

となるので,

$$\log \frac{p(x^n | \hat{\theta}_m, m)}{p(x^n | \theta_m^0, m)} = \frac{n}{2} (\hat{\theta}_m - \theta_m^0)^T \hat{J}(\hat{\theta}_m | m) (\hat{\theta}_m - \theta_m^0) + o(1), \quad (3.29)$$

となる. よって, 補題 3.2.1 より,

$$\begin{aligned} \frac{n}{2} (\hat{\theta}_m - \theta_m^0)^T \hat{J}(\hat{\theta}_m | m) (\hat{\theta}_m - \theta_m^0) &= \frac{n}{2} (\hat{\theta}_m - \theta_m^0)^T J(\theta_m^0 | m) (\hat{\theta}_m - \theta_m^0) \\ &\quad + o(1), \quad a.s., \end{aligned} \quad (3.30)$$

が成り立つので, 有界収束定理を用いて,

$$\begin{aligned} E_{p^n} \left[\frac{n}{2} (\hat{\theta}_m - \theta_m^0)^T \hat{J}(\hat{\theta}_m | m) (\hat{\theta}_m - \theta_m^0) \right] &= \frac{1}{2} \text{tr} \left(I(\theta_m^0 | m) J(\theta_m^0 | m)^{-1} \right) \\ &\quad + o(1), \end{aligned} \quad (3.31)$$

となることから, 題意を得る. \square

3.6.3 補題 3.5.1 の証明

まず, $-\log p(x^n | \theta_m^0, m)$ の項を評価する. ここで, $\nabla = \left(\frac{\partial}{\partial \theta_{m,1}}, \dots, \frac{\partial}{\partial \theta_{m,k_m}} \right)^T$ とすると, 全ての $m \in \mathcal{M}$ に対し, $\nabla \log p(x^n | \hat{\theta}_m, m)$ は $\mathbf{0}$ に等しく, $-\log p(x^n | \hat{\theta}_m, m)$ は $O(n)$ に等しいので, テーラーの定理より次の式を得る.

$$\begin{aligned} -\log p(x^n | \theta_m^0, m) &= -\log p(x^n | \hat{\theta}_m, m) + \frac{n}{2} (\theta_m^0 - \hat{\theta}_m)^T \hat{J}(\hat{\theta}_m | m) (\theta_m^0 - \hat{\theta}_m) \\ &\quad + O \left(n |\theta_m^0 - \hat{\theta}_m|^3 \right). \end{aligned} \quad (3.32)$$

さらに, 補題 3.2.1 より,

$$O\left(n|\hat{\theta}_m - \theta_m^0|^3\right) = O\left(\frac{(\log \log n)^{3/2}}{n^{1/2}}\right), \text{ a.s.}, \quad (3.33)$$

および

$$\frac{n}{2}(\theta_m^0 - \hat{\theta}_m)^T \hat{J}(\hat{\theta}_m | m)(\theta_m^0 - \hat{\theta}_m) = O(\log \log n), \text{ a.s.}, \quad (3.34)$$

が成り立つ. よって, 全ての $m \in \mathcal{M}$ に対し,

$$-\log p(x^n | \theta_m^0, m) = -\log p(x^n | \hat{\theta}_m, m) + O(\log \log n), \text{ a.s.}, \quad (3.35)$$

が成り立つ.

次に \mathcal{M}^- および \mathcal{M}^+ を次のように定義する.

$$\mathcal{M}^+ = \left\{ m \mid m \in \mathcal{M}, \forall m \neq m^0, D(p \| p_{\theta_m^0}, m) = D(p \| p_{\theta_{m^0}^0}, m^0) \right\}, \quad (3.36)$$

$$\mathcal{M}^- = \left\{ m \mid m \in \mathcal{M}, \forall m \neq m^0, D(p \| p_{\theta_m^0}, m) > D(p \| p_{\theta_{m^0}^0}, m^0) \right\}. \quad (3.37)$$

全ての $m \in \mathcal{M}$ に対し, $D(p \| p_{\theta_{m^0}^0}, m^0)$ は $D(p \| p_{\theta_m^0}, m)$ に等しいかあるいはより小さいので, $\mathcal{M}^- \cup \mathcal{M}^+ \cup \{m^0\}$ は \mathcal{M} に等しく, $\mathcal{M}^- \cap \mathcal{M}^+$ は ϕ に等しい. よって, 補題 3.4.1 より, 全ての $m \in \mathcal{M}$ に対し,

$$\begin{aligned} \log \frac{p(x^n | m)}{p(x^n | m^0)} &= \log \frac{p(x^n | \hat{\theta}_m, m)}{p(x^n | \hat{\theta}_{m^0}, m^0)} - \frac{k_m - k_{m^0}}{2} \log \frac{n}{2\pi} \\ &\quad + \log \frac{w(\hat{\theta}_m | m) \sqrt{\det \hat{J}(\hat{\theta}_{m^0} | m^0)}}{w(\hat{\theta}_{m^0} | m^0) \sqrt{\det \hat{J}(\hat{\theta}_m | m)}} + o(1), \end{aligned} \quad (3.38)$$

を得る. まず $m \in \mathcal{M}^-$ の場合を考える. 大数の法則より, 全ての $m \in \mathcal{M}$ に対し,

$$-\log p(x^n | \theta_m^0, m) = E_{p^n} [-\log p(X^n | \theta_m^0, m)] + o(n), \text{ a.s.}, \quad (3.39)$$

が成り立ち、さらに、条件 3.2.1 が成り立つならば、全ての $m \in \mathcal{M}$ に対し、

$$D(p \| p_{\theta_{m^0}, m^0}^0) < D(p \| p_{\theta_m^0, m}^0) \\ \iff E_{p^n} [-\log p(X^n | \theta_{m^0}^0, m^0)] < E_{q^n} [-\log p(X^n | \theta_m^0, m)], \quad (3.40)$$

が成り立つので、 n が大きくなったとしても $-\log p(x^n | \theta_{m^0}^0, m^0)$ は $-\log p(x^n | \theta_m^0, m)$ を超えない。よって、全ての $m \in \mathcal{M}$ に対し、 $-\log p(x^n | \theta_m^0, m)$ は $O(n)$ に等しいので、

$$\log \frac{p(x^n | \theta_m^0, m)}{p(x^n | \theta_{m^0}^0, m^0)} = -C_m n + o(n), \quad a.s., \quad (3.41)$$

が成り立つ。ただし、 C_m は m に依存する定数である。よって、式 (3.35) より、

$$\log \frac{p(x^n | \hat{\theta}_m, m)}{p(x^n | \hat{\theta}_{m^0}, m^0)} = \log \frac{p(x^n | \theta_m^0, m)}{p(x^n | \theta_{m^0}^0, m^0)} + O(\log \log n), \quad a.s. \quad (3.42)$$

$$= -C_m n + o(n), \quad a.s., \quad (3.43)$$

を得る。ここで、 $w(\hat{\theta}_m | m)$ が $O(1)$ に等しく、 $\hat{J}(\hat{\theta}_m | m)$ が $O(1)$ に等しいため、式 (3.38) より、全ての $m \in \mathcal{M}^-$ に対し、

$$\log \frac{p(x^n | m)}{p(x^n | m^0)} = -C_m n - \frac{k_m - k_{m^0}}{2} \log n \\ + o(n), \quad a.s., \quad (3.44)$$

$$\lim_{n \rightarrow \infty} \log \frac{p(x^n | m)}{p(x^n | m^0)} = -\infty, \quad a.s., \quad (3.45)$$

が成り立つ。次に $m \in \mathcal{M}^+$ の場合は、 $p(x^n | \theta_m^0, m)$ は $p(x^n | \theta_{m^0}^0, m^0)$ に等しいので、式 (3.35) より、

$$\log \frac{p(x^n | \hat{\theta}_m, m)}{p(x^n | \hat{\theta}_{m^0}, m^0)} = O(\log \log n), \quad a.s., \quad (3.46)$$

が成り立つ。よって、式 (3.38) および m^0 の定義より $k_m > k_{m^0}$ である事から、全ての $m \in \mathcal{M}^+$ に対し、

$$\log \frac{p(x^n | m)}{p(x^n | m^0)} = -\frac{k_m - k_{m^0}}{2} \log n \\ + O(\log \log n), \quad a.s., \quad (3.47)$$

$$\lim_{n \rightarrow \infty} \log \frac{p(x^n | m)}{p(x^n | m^0)} = -\infty, \quad a.s., \quad (3.48)$$

が成り立つ。したがって、全ての $m \in \mathcal{M}$ に対し、

$$\lim_{n \rightarrow \infty} \log \frac{\max_{m \neq m^0} p(x^n | m)}{p(x^n | m^0)} = -\infty, \text{ a.s.}, \quad (3.49)$$

$$\lim_{n \rightarrow \infty} \frac{\max_{m \neq m^0} p(x^n | m)}{p(x^n | m^0)} = +0, \text{ a.s.}, \quad (3.50)$$

を得る. □

3.6.4 定理 3.5.1 の証明

まず、

$$-\log \frac{Ap^*(x^n)}{P(m^0)p(x^n | m^0)} = -\log \left(1 + \frac{\sum_{m \neq m^0} P(m)p(x^n | m)}{P(m^0)p(x^n | m^0)} \right) \quad (3.51)$$

$$\leq 0, \quad (3.52)$$

が成り立ち、さらに、補題 3.5.1 より

$$-\log \frac{Ap^*(x^n)}{P(m^0)p(x^n | m^0)} \geq -\log \left(1 + \frac{\max_{m \neq m^0} p(x^n | m) \sum_{m \neq m^0} P(m)}{P(m^0)p(x^n | m^0)} \right) \quad (3.53)$$

$$= -\log(1 + o^+(1)), \text{ a.s.}, \quad (3.54)$$

が成り立つ。したがって、

$$-\log \frac{Ap^*(x^n)}{P(m^0)p(x^n | m^0)} = o(1), \text{ a.s.}, \quad (3.55)$$

が成り立つため、補題 3.4.1 を用いる事により、題意を得る. □

3.6.5 系 3.5.1 の証明

定理 3.5.1 より、

$$\log \frac{p(x^n | \hat{\theta}_{m^0}, m^0)}{Ap^*(x^n)} = \frac{k_{m^0}}{2} \log \frac{n}{2\pi} + \log \frac{\sqrt{\det \hat{J}(\hat{\theta}_{m^0} | m^0)}}{w(\hat{\theta}_{m^0} | m^0)} - \log P(m^0) + o(1), \text{ a.s.}, \quad (3.56)$$

を得る. 条件 3.2.2 が成り立つならば, $w(\theta_{m^0} | m^0)$ は連続であるため, 補題 3.2.1 より,

$$\log \frac{\sqrt{\det \hat{J}(\hat{\theta}_{m^0} | m^0)}}{w(\hat{\theta}_{m^0} | m^0)} = \log \frac{\sqrt{\det J(\theta_{m^0}^0 | m^0)}}{w(\theta_{m^0}^0 | m^0)} + o(1), \text{ a.s.}, \quad (3.57)$$

を得る. したがって, 有界収束定理を用いる事により, 題意を得る. \square

第4章 BW変換に基づくベイズ符号化アルゴリズム

4.1 はじめに

本章では、本論文のもう一方の成果であるBW変換に基づくベイズ符号化アルゴリズムについて述べる。まず、2.3.2節で述べたブロックソート法の一部に用いられているBW変換を階層型のモデル未知の確率モデルの一種であるマルコフ過程で表される情報源の出力系列に適用した場合、変換後の系列は2.3.5節で述べた区分定常情報源の出力系列とみなせる事について述べる。そして、2.4.7節で述べた区分定常情報源に対する効率的なベイズ符号化アルゴリズムをBW変換後の系列の符号化に用いる事により、マルコフ過程で表される情報源に対しては実用性をほとんど失う事無く、符号化の性能が理論的にも明確な符号化アルゴリズムを提案する。

4.2 マルコフ情報源に対するBW変換の性質

2.3.4節で述べたように、FSMX情報源においてモデル m を完全木で表現した場合に全ての葉ノードが深さ m のみに存在する時、その情報源は m 次マルコフ情報源と呼ばれる。この時、モデル m における状態は全て長さ m の文脈に対応する。本章では、FSMX情報源のモデル m に関し、深さ m のみに葉ノードを持つ完全木で表現されるもののみを考え、すなわち、 m 次マルコフ情報源のみを考え、 m は $M = \{1, \dots, m_{\max}\}$ 上に値をとるものとする。

2.3.2節におけるBW変換の例において、情報源系列 x^9 が1次マルコフ情報源から出力されたと仮定すると、 y_1^3 は文脈0の下で発生し、 y_4^8 は文脈1の下で発生している事になる。一方、情報源系列 x^9 が2次マルコフ情報源の出力されたと仮定する

と、 y_1 は文脈 00 の下で発生し、 y_2^3 は文脈 10 の下で発生し、 y_4^5 は文脈 01 の下で発生し、 y_6^7 は文脈 11 の下で発生している事になる。このように、マルコフ情報源の出力系列に BW 変換を施した系列は、ある区間の部分系列が同じ文脈から発生しているため、2.3.5 節で述べた区分定常情報源からの出力系列とみなせる事が分かる。

4.3 BW 変換に基づくバイズ符号化アルゴリズム

前節で述べたようにマルコフ情報源の出力系列に BW 変換を施した系列を区分定常情報源からの出力系列とみなした場合、その変化パターン ω はモデル m ，すなわち、マルコフ情報源の次数に依存する。そこで、 $t_{\omega,j}$ を $t_{m,j}$ と書き換える事にする。ここで、 m 次マルコフ情報源において発生し得る文脈の総数が 2^m である事から、 $C(\omega) \leq 2^m - 1$ が成り立つ。もし、 y^n ，において全ての文脈が発生したとすると、 $C(\omega)$ は $2^m - 1$ に等しくなる。また、2.4.7 節で定義した、ある部分パターン ω_t を得た下での部分パターン ω_{t+1} に対し、時点 $t+1$ においてパラメータが変化する確率 π は時点 t および j に依存する。ただし、 j はある変化パターン ω における $t_{\omega,j}$ の添え字である。この添え字 j はパラメータが何回変化したかという事を表している。そこで本節では、式 (2.52) の仮定に対応するものとして次のような仮定を用いる。ただし、 $i = 1, \dots, t$ である。

$$P(\tau_t = i, \tau_t = t_{\omega,j}) = \sum_{\{\omega | t_{C(\omega_t)} = i, t_{C(\omega_t)} = t_{\omega,j}\}} P(\omega). \quad (4.1)$$

すると、2.4.7 節で述べた従来アルゴリズムは次のように書き換える事ができる。

符号化確率の計算は次のように行われる。

$$AP(y_1 | y^0) = \sum_{l=0}^{m_{\max}} p(y_1 | y^0, \tau_1 = 1, \tau_1 = t_{m_{\max}-l,0}) P(\tau_1 = 1, \tau_1 = t_{m_{\max}-l,0} | y^0), \quad (4.2)$$

Algorithm 2 提案アルゴリズム

```

1: Input  $n, m_{\max}$ 
2: calculate  $AP(y_1 | y^0)$  by Eq. (4.2)
3: for  $t = 2$  to  $n$  do
4:   calculate  $AP(y_t | y^{t-1})$  by Eq. (4.3)
5:    $m \leftarrow m_{\max}$ 
6:   calculate  $P(\tau_{t+1} = 1, \tau_{t+1} = t_{m,0} | y^t)$  by Eq. (4.7)
7:   for  $i = 2$  to  $t$  do
8:     for  $j = 1$  to  $\min\{2^{m_{\max}}, i - 1\}$  do
9:       calculate  $P(\tau_{t+1} = i, \tau_{t+1} = t_{m,j} | y^t)$  by Eq. (4.7)
10:       $k \leftarrow j$ 
11:      while  $k$  is even do
12:         $m \leftarrow m - 1, k \leftarrow k/2$ 
13:        calculate  $P(\tau_{t+1} = i, \tau_{t+1} = t_{m,k} | y^t)$  by Eq. (4.7)
14:      end while
15:    end for
16:  end for
17:   $m \leftarrow m_{\max}$ 
18:  for  $j = 1$  to  $\min\{2^{m_{\max}}, t\}$  do
19:    calculate  $P(\tau_{t+1} = t + 1, \tau_{t+1} = t_{m,j} | y^t)$  by Eq. (4.8)
20:     $k \leftarrow j$ 
21:    while  $k$  is even do
22:       $m \leftarrow m - 1, k \leftarrow k/2$ 
23:      calculate  $P(\tau_{t+1} = t + 1, \tau_{t+1} = t_{m,k} | y^t)$  by Eq. (4.8)
24:    end while
25:  end for
26: end for
27: Output  $AP(y_n | y^{n-1})$ 

```

$$\begin{aligned}
AP(y_t | y^{t-1}) &= AP(y_1 | y^0) \\
&+ \sum_{l=0}^{\min\{\lfloor \log_2(t-1) \rfloor, m_{\max}\}} \sum_{i=2^l+1}^t \sum_{j=1}^{\min\{\lfloor (i-1)/2^l \rfloor, m_{\max}\}} p(y_t | y^{t-1}, \tau_t = i, \tau_t = t_{m_{\max}-l,j}) \\
&\quad \times P(\tau_t = i, \tau_t = t_{m_{\max}-l,j} | y^{t-1}). \quad (4.3)
\end{aligned}$$

$$\begin{aligned}
p(y_t | y^{t-1}, \tau_t = i, \tau_t = t_{m_{\max}-l,j}) \\
&= \int_{\theta_{t_{m_{\max}-l,j}}} p(y_t | y^{t-1}, \theta_{t_{m_{\max}-l,j}}, \tau_t = t_{m_{\max}-l,j}) \\
&\quad \times w(\theta_{t_{m_{\max}-l,j}} | y^{t-1}, \tau_t = t_{m_{\max}-l,j}) d\theta_{t_{m_{\max}-l,j}}. \quad (4.4)
\end{aligned}$$

式 (4.4) は、パラメータの事前確率密度にディリクレ分布を仮定した場合、自然共役性より次のように計算される。

$$\frac{\nu_{m,j}(y_t | y_i^{t-1}) + \beta_{m,j}(y_t | \tau_t = i)}{\sum_{a=0}^{|\mathcal{X}|-1} (\nu_{m,j}(a | y_i^{t-1}) + \beta_{m,j}(a | \tau_t = i))}. \quad (4.5)$$

ただし、 $\nu_{m,j}(a | y_i^{t-1})$ は時点 i から $t-1$ までの間にシンボル $a \in \mathcal{X}$ が発生した回数であり、 $\beta_{m,j}(a | \tau_t)$ はディリクレ分布のパラメータである。

一方、事後確率関数は次のように計算される。

$$\begin{aligned}
&P(\tau_{t+1} = i, \tau_{t+1} = t_{m,j} | y^t) \\
&= \left(1 - \frac{2^m - 1 - j}{n - t}\right) P(\tau_t = i, \tau_t = t_{m,j} | y^t) \\
&= \left(1 - \frac{2^m - 1 - j}{n - t}\right) \frac{p(y_t | y^{t-1}, \tau_t = i, \tau_t = t_{m,j}) P(\tau_t = i, \tau_t = t_{m,j} | y^{t-1})}{AP(y_t | y^{t-1})}, \quad (4.6) \\
&\quad (4.7)
\end{aligned}$$

$$P(\tau_{t+1} = t + 1, \tau_{t+1} = t_{m,j} | y^t) = \frac{2^m - j}{n - t}. \quad (4.8)$$

先に述べた通り、提案アルゴリズムは従来アルゴリズムを書き換えたものである。また、定理 2.4.3 により、従来アルゴリズムにおいて計算される符号化確率はベイズ最適な符号化確率に等しい。したがって、定理 2.4.3 と同様に次の定理が成り立つ。

定理 4.3.1 提案アルゴリズムにおいて計算される符号化確率はベイズ基準の下で最適な逐次型の符号化確率と等しい, すなわち,

$$AP(y_t | y^{t-1}) = AP^*(y_t | y^{t-1}), \quad (4.9)$$

が成り立つ. ただし, $AP(y_t | y^{t-1})$ は提案アルゴリズムで時点 t において計算される符号化確率である.

(証明) 定理 2.4.3 と同様. □

提案アルゴリズムにおいて, 変化パターンの和計算に関する時間計算量について考える. 提案アルゴリズムの式 (4.3) において, 十分大きな t に対し, 最も外側の和計算が $m_{\max} + 1$ 回であり, 最も内側の和計算が m_{\max} 回である. 一方, 中央の和計算は $l = 0, 1, \dots, m_{\max}$ に対し, t 回 $\rightarrow t - 2$ 回 $\rightarrow \dots \rightarrow t - 2^{m_{\max}}$ 回となっていく. したがって, 最大でも $O(m^2 t)$ であり, 当初の $O(2^{t-1})$ より削減されている事が分かる.

第5章 結論

5.1 本論文の成果

本論文では階層型の確率モデル族に基づくユニバーサル情報源符号化についての研究成果を述べた。

第3章においては、本論文の一方の成果である情報源の確率分布を表現できない確率モデルに基づくベイズ符号の性能解析について述べた。モデル既知の確率モデルを仮定したベイズ符号において情報源の確率分布がその確率モデルでは表現されない分布だった場合の平均符号語長の精密な漸近評価を行った。さらに、その結果を応用し、入れ子構造を持つ階層型のモデル未知の確率モデルを仮定したベイズ符号において、情報源の確率分布がその確率モデルでは表現されない分布だった場合の平均符号語長の漸近評価を厳密に行い、そのように仮定が崩れた場合にもベイズ符号が有効である事について述べた。具体的には、モデルが規定するパラメータの次数が最も大きいモデルのみに基づいたベイズ符号を用いる場合（ケースA）と、そのモデルだけでなくよりパラメータの次数が低いモデルまで全て含めてベイズ符号を用いる場合（ケースB）ではケースBの方がケースAと同等、状況によってはケースAより良い性能を発揮するという事を主張した。良い性能とはすなわち、平均符号語長のエントロピーレートへの収束オーダーがより小さいという事である。ケースBの方がケースAより良い性能を発揮する状況というのは、真の確率分布とのカルバックライブラー情報量を最小にする（真の分布に最も近い）確率分布を、パラメータの次数が最も大きくなるモデルだけではなくより小さい次数のパラメータを規定するモデルでも表せる場合である。主定理より、収束オーダーは、真の分布に最も近い分布のうち、パラメータの次数が最も小さい分布のパラメータの次数に依存する事が分かる。したがって、上記の状況においては、ケースAの場合、収束オーダーは各モデルが規定するパラメータの次数のうち最大のものに依存する。一

方，ケース B の場合，最大次数より小さい次数のパラメータでも真の分布に最も近い分布を表せるため，収束オーダーはその次数に依存する事になり，ケース B の方がより良い性能を発揮する事が分かった。

第 4 章では，本論文のもう一方の成果である BW 変換を利用したベイズ符号化アルゴリズムについて述べた。ブロックソート法の一部に用いられている BW 変換を階層型のモデル未知の確率モデルの一種であるマルコフ過程で表される情報源の出力系列に適用した場合，区分定常情報源の出力系列とみなせる。従来提案されている区分定常情報源に対する効率的なベイズ符号化アルゴリズムを BW 変換後の系列の符号化に用いる事により，マルコフ過程で表される情報源に対しては実用性をほとんど失う事無く，符号化の性能が理論的にも明確な符号化アルゴリズムを提案した。提案アルゴリズムは，従来のベイズ最適性が保証されているアルゴリズムを流用しているため，その性能は理論的にも明確であった。また，計算量についても，マルコフ過程で表される情報源の一種である FSMX 情報源に対する効率的なベイズ符号化アルゴリズムとほとんど同等であり，実用性もほとんど失っていない事が分かった。

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

に心より感謝致します。松嶋教授には学部生の頃より、自らがご多忙な状況にもかかわらず、研究だけでなく、社会人としての心構え等、数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、大変お忙しい中貴重なお時間を頂きご指導を賜りました、

早稲田大学基幹理工学部応用数理学科
谷口正信教授

早稲田大学基幹理工学部応用数理学科
大石進一教授

早稲田大学基幹理工学部応用数理学科
柏木雅英教授

早稲田大学理工学術院総合研究所
平澤茂一名誉教授

に心より感謝致します。

また、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。横浜商科大学 浮田善文教授には研究室内外の様々な場面でお世話になりました。北見工業大学 前田康成准教授、横浜商科大学 吉田隆弘准教授、専修大学 野村亮准教授、日本電気株式会社 峯松一彦氏、株式会社NTT データ 末永高志氏、小樽商科大学 小泉大城准教授、早稲田大学 須子統太准教授、東京都市大学 齋藤友彦講師、株式会社東芝 安田豪毅氏、株式会社NTT ドコモ 桑田修平氏、早稲田大学 堀井俊佑

助教には研究室のゼミを通じて多くの議論やご助言を戴きました。特に須子統太准教授と堀井俊佑助教には学部生の頃より研究の枠を超えて常に親身なアドバイスを戴きました。深く感謝致します。

松嶋研究室に所属する学生一同には、常日頃より大変お世話になりました。この場を借りて心よりお礼申し上げます。

本論文は以上をはじめとする多くの方々のご指導，ご支援の賜物です。お世話になりました皆様に深く感謝致します。最後に，著者の研究生生活をあらゆる面から支え，温かく見守ってくれた家族に心から感謝致します。

2016年7月

参考文献

- [1] H. Akaike, “A new look at the statistical model identification,” *IEEE Trans. Autom. Control*, vol.19, no.6, pp.716–723, Dec. 1974.
- [2] M. Arimura and H. Yamamoto, “Asymptotic optimality of the block sorting data compression algorithm,” *IEICE Trans. Fundamentals*, vol.E81-A, no.10, pp.2117–2122, Oct. 1998.
- [3] A.R. Barron, J. Rissanen, and B. Yu., “The minimum description length principle in coding and modeling,” *IEEE Trans. Inf. Theory*, vol.44, no.6, pp.2743–2760, Oct. 1998.
- [4] M. Burrows and D.J. Wheeler, “A block-sorting lossless data compression algorithm,” *Digital Syst. Res. Ctr.*, Palo Alto, CA, Tech. Rep. SRC 124, May 1994.
- [5] K.L. Chung, *Markov Chains with Stationary Transition Probabilities*, Springer-Verlag, 1967.
- [6] B.S. Clarke and A.R. Barron, “Information-theoretic asymptotics of Bayes methods,” *IEEE Trans. Inf. Theory*, vol.36, no.3, pp.453–471, May 1990.
- [7] B.S. Clarke and A.R. Barron, “Jeffreys’ prior is asymptotically least favorable under entropy risk,” *J. Statist. Plann. Inference*, vol.41, no.1, pp.37–60, Aug. 1994.
- [8] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, 2006.

-
- [9] M. Effros, K. Visweswariah, S.R. Kulkarni, and S. Verdú, “Universal lossless source coding with the Burrows Wheeler transform,” *IEEE Trans. Inf. Theory*, vol.48, no.5, pp.1061–1081, May 2002.
- [10] M. Gotoh, T. Matsushima, and S. Hirasawa, “A generalization of B.S. Clarke and A.R. Barron’s asymptotics of Bayes codes for FSMX sources,” *IEICE Trans. Fundamentals*, vol.E81-A, no.10, pp.2123–2132, Oct. 1998.
- [11] D. Haussler and M. Opper, “Mutual information, metric entropy and cumulative relative entropy risk,” *Ann. Statist.*, vol.25, no.6, pp.2451–2492, Jan. 1997.
- [12] D. Haussler, J. Kivinen, and M.K. Warmuth, “Sequential prediction of individual sequences under general loss functions,” *IEEE Trans. Inf. Theory*, vol.44, no.5, pp.1906–1925, Sept. 1998.
- [13] D.A. Huffman, “A Method for the Construction of Minimum-Redundancy Codes,” *Proceedings of the I.R.E.*, vol.20, no.3, pp.1098–1101, Sept. 1952.
- [14] E.L. Lehmann and G. Casella, *Theory of Point Estimation*, Springer-Verlag, 1998.
- [15] 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, Sept. 1991.
- [16] T. Matsushima and S. Hirasawa, “Reducing the space complexity of a Bayes coding algorithm using an expanded context tree,” in *Proc. IEEE Int. Symp. Inf. Theory*, pp.719–723, Seoul, June–July, 2009.
- [17] R. Nishii, “Maximum likelihood principle and model selection when the true model is unspecified,” *J. Multivariate Anal.*, vol.27, no.2, pp. 392-403, Nov. 1988.
- [18] J. Rissanen, “Generalized Kraft Inequality and Arithmetic Coding,” *IBM Journal Res. Develop*, vol.20, no.3, pp.198–203, May 1976.

-
- [19] J. Rissanen and G.G. Langdon, Jr, "Arithmetic Coding," IBM Journal Res. Develop, vol.23, no.2, pp.149–162, Mar. 1979.
- [20] C.E. Shannon, "A Mathematical Theory of Communication," Bell. Sys. Tech. Journal, vol.27, no.3, pp.379–423, July 1948.
- [21] 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 (SITA2003) 予稿集, pp.165–168, 2003年12月.
- [22] 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 (SITA2005) 予稿集, pp.315–318, 2005年11月.
- [23] 竹内啓, "情報統計量の分布とモデルの適切さの規準," 数理科学, no.153, pp.12–18, 1976.
- [24] J. Takeuchi and A.R. Barron, "Asymptotically minimax regret by Bayes mixtures," in Proc. IEEE Int. Symp. Inf. Theory, Boston, MA, Aug. 1998, p.318.
- [25] J. Takeuchi and A.R. Barron, "Robustly minimax codes for universal data compression," 第21回情報理論とその応用シンポジウム (SITA1998) 予稿集, pp.213–216, 1998年12月.
- [26] L. Tierney and J. Kadane, "Accurate approximations for posterior moments and marginal densities," J. Amer. Statist. Assoc., vol.81, no.393, pp.82–86, Mar. 1986.
- [27] H. White, "Maximum likelihood estimation of misspecified models," Econometrica, vol.50, no.1, pp.1–25, Jan. 1982.
- [28] 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, May 1995.

-
- [29] F.M.J. Willems, “Coding for binary independent piecewise-identically-distributed source,” *IEEE Trans. Inf. Theory*, vol.42, no.6, pp.2210–2217, Nov. 1996.
- [30] F.M.J. Willems, “The context-tree weighting method: extensions,” *IEEE Trans. Inf. Theory*, vol.44, no.2, pp.792–798, Mar. 1998.
- [31] A.D. Wyner and J. Ziv, “The sliding-window Lempel-Ziv algorithm is asymptotically optimal,” *Proceedings of the IEEE*, vol.82, no.6, pp.872–877, June 1994.
- [32] J. Ziv and A. Lempel, “A universal algorithm for sequential data compression,” *IEEE Trans. Inf. Theory*, vol.23, no.3, pp.337–343, May 1977.
- [33] J. Ziv and A. Lempel, “Compression of individual sequences via variable-rate coding,” *IEEE Trans. Inf. Theory*, vol.24, no.5, pp.530–536, Sept. 1978.

研究業績

種類別	題名, 発表・発行掲載誌名, 発表・発行年月, 連名者 (申請者含む)
1. 論文	<p>Evaluation of the Bayes Code from Viewpoints of the Distribution of Its Codeword Lengths IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol.E98-A, no.12, pp.2407-2414, Dec. 2015 Shota SAITO, Nozomi MIYA, Toshiyasu MATSUSHIMA</p>
2. 論文○	<p>Asymptotics of Bayesian Inference for a Class of Probabilistic Models under Misspecification IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol.E97-A, no.12, pp.2352-2360, Dec. 2014 Nozomi MIYA, Tota SUKO, Goki YASUDA, Toshiyasu MATSUSHIMA</p>
3. 国際会議 (査読有)	<p>Fundamental limit and pointwise asymptotics of the Bayes code for Markov sources 2015 IEEE International Symposium on Information Theory (ISIT), Hong Kong, June 2015 Shota Saito, Nozomi Miya, Toshiyasu Matsushima</p>

種類別	題名, 発表・発行掲載誌名, 発表・発行年月, 連名者 (申請者含む)
4. 国際会議 (査読有)	Evaluation of the minimum overflow threshold of Bayes codes for a Markov source 2014 International Symposium on Information Theory and Its Applications (ISITA), Melbourne, Oct. 2014 Shota Saito, Nozomi Miya, Toshiyasu Matsushima
5. 国際会議 (査読有)	Asymptotics of MLE-based Prediction for Semi-supervised Learning 2014 International Symposium on Information Theory and Its Applications (ISITA), Melbourne, Oct. 2014 Goki Yasuda, Nozomi Miya, Tota Suko, Toshiyasu Matsushima
6. 国際会議 (査読有) ○	Asymptotics of Bayesian estimation for nested models under misspecification 2012 International Symposium on Information Theory and Its Applications (ISITA), Honolulu, Oct. 2012 Nozomi Miya, Tota Suko, Goki Yasuda, Toshiyasu Matsushima
7. 講演	プライバシー保護機能を持つ分散型正則化ロジスティック回帰に関する一考察 電子情報通信学会パターン認識・メディア理解研究会 (PRMU), 大阪府, 2016年1月 増井秀之, 宮希望, 松嶋敏泰
8. 講演	一般情報源に対する Slepian-Wolf 符号化問題の2次の達成可能レート領域の別表現 電子情報通信学会情報理論研究会 (IT), 福岡県, 2015年3月 齋藤翔太, 宮希望, 松嶋敏泰

種類別	題名，発表・発行掲載誌名，発表・発行年月，連名者（申請者含む）
9. 講演	<p>非定常情報源に対する Bayes 符号のオーバーフロー確率における最小しきい値の評価</p> <p>第 37 回情報理論とその応用シンポジウム (SITA2014)，富山県，2014 年 12 月</p> <p>守屋貴司，齋藤翔太，宮希望，松嶋敏泰</p>
10. 講演	<p>定常エルゴードマルコフ情報源に対する Bayes 符号の符号語長の漸近正規性と重複対数の法則</p> <p>第 37 回情報理論とその応用シンポジウム (SITA2014)，富山県，2014 年 12 月</p> <p>齋藤翔太，宮希望，松嶋敏泰</p>
11. 講演	<p>パターン認識における AdaBoost の予測誤り率改善に関する一考察</p> <p>電子情報通信学会情報論的学習理論と機械学習研究会 (IBISML)，茨城県，2014 年 9 月</p> <p>増井秀之，都築遼馬，宮希望，松嶋敏泰</p>
12. 講演	<p>推薦対象ユーザのクラスが未知の推薦問題におけるマルコフ決定過程を用いた推薦システムに関する一考察</p> <p>電子情報通信学会情報理論研究会 (IT)，兵庫県，2014 年 7 月</p> <p>岩井秀輔，宮希望，前田康成，松嶋敏泰</p>
13. 講演	<p>真の分布を含むとは限らない階層モデル族に対するベイズ推定の漸近評価</p> <p>第 36 回情報理論とその応用シンポジウム (SITA2013)，静岡県，2013 年 11 月</p> <p>宮希望，須子統太，安田豪毅，松嶋敏泰</p>

種類別	題名，発表・発行掲載誌名，発表・発行年月，連名者（申請者含む）
14. 講演	半教師付き学習における一致推定量に基づく予測の漸近評価 第 36 回情報理論とその応用シンポジウム（SITA2013），静岡県， 2013 年 11 月 安田豪毅，宮希望，須子統太，松嶋敏泰
15. 講演	ベイズ符号のオーバーフロー確率における最小しきい値の評価 第 36 回情報理論とその応用シンポジウム（SITA2013），静岡県， 2013 年 11 月 齋藤翔太，宮希望，野村亮，松嶋敏泰
16. 講演	真のモデルを含まないパラメトリックモデル族に対するベイズ予測 の漸近評価 電子情報通信学会情報理論研究会（IT），岡山県，2011 年 7 月 宮希望，須子統太，安田豪毅，松嶋敏泰