

令和 6 年 5 月 13 日現在

機関番号：12608

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K12025

研究課題名（和文）情報量規準WBICの高精度化

研究課題名（英文）Information Criterion WBIC and its improvement

研究代表者

渡邊 澄夫（Watanabe, Sumio）

東京工業大学・情報理工学院・教授

研究者番号：80273118

交付決定額（研究期間全体）：（直接経費） 1,400,000円

研究成果の概要（和文）：統計的推測あるいは統計的学習において、確率モデル・事前分布・サンプルが与えられたときの自由エネルギー（マイナス対数周辺尤度）を求める課題において、以下の成果が得られた。(1) サンプルが交換可能である場合において従来理論の一般化を行い WBICが適用できることを示した。(2) ReLU関数を活性化関数として持つ畳み込み神経回路網においてスキップ接続を持つ場合と持たない場合における自由エネルギーの相違を明らかにし、スキップ接続が自由エネルギーを小さくする機能を持つことを示した。(3) データ生成分布が確率モデルによって実現できないとき階層構造や潜在変数を持つモデルの自由エネルギーの挙動を解明した。

研究成果の学術的意義や社会的意義

確率モデル・事前分布・サンプルが与えられたときに自由エネルギーの値を算出することは、データ生成分布に対する確率モデルと事前分布の適切さを判断する際に重要な役割を果たすことが知られているが、事後分布が正規分布で近似できない場合には、その値を求めるには大きな演算量が必要であった。本研究では、自由エネルギーの値を求めるために提案されていた情報量規準WBICについて研究を行い、より一般的な条件下で利用可能であるように拡張し、階層モデルを持つモデルの挙動を解明した。これらの結果は、実問題のデータ分析におけるモデリングの基盤のひとつであり、また人工知能アライメントのための数学的基礎を構成するものである。

研究成果の概要（英文）：In statistical inference and statistical learning, the free energy which is equal to the minus log marginal likelihood is defined for a given probabilistic model, a prior distribution, and a sample. In this research project, the following results were obtained. (1) The conventional theory was generalized for exchangeable random variables so that WBIC can be employed to estimate the free energy. (2) The free energies for the convolutional neural networks using a ReLU function with skip connection and without skip connection were clarified. And (3) the free energy of a learning model which has hierarchical structure or hidden variables were clarified when the data-generating distribution is not realizable.

研究分野：特異学習理論

キーワード：特異学習理論 周辺尤度 情報量規準 代数幾何 自由エネルギー

1. 研究開始当初の背景 統計的推論あるいは統計的学習において、サンプル・確率モデル・事前分布が与えられたときに定義される量として周辺尤度があり、マイナス対数周辺尤度のことを自由エネルギー、確率的複雑さ、あるいはベイズ符号長という。自由エネルギーのサンプルに関する平均値の大小は、未知のデータ生成分布と「確率モデル・事前分布から定義される同時分布」とのカルバック・ライブラ情報量の大小と等価であるため、自由エネルギーの値を求めることは確率モデルと事前分布のデータ生成分布に対する適切さを調べる際に重要な意義を有している。しかしながら、その値の定義におけるパラメータ集合上での積分を数値的に行う際に多大な演算量が必要になるという課題が知られていた。事後分布が多変量正規分布で近似できるときには、自由エネルギーの値を近似する量として BIC があり、また事後分布が正規分布で近似できない場合でも利用可能な量として WBIC が知られていた。研究当初の背景として次の課題が存在していた。(1) サンプルが独立でない場合でも WBIC が利用可能であるか不明であった。(2) 自由エネルギーを解明することで確率モデルの構造が推測精度に与える影響をどこまで取り出せるかは知られていなかった。特に階層構造を持つモデルの設計にどのような具体的な基盤を与えるかは不明であった。(3) 20 世紀の古いベイズ統計学では、データ生成分布が確率モデルで実現可能であるとユーザが信じていることが前提条件であるとされていたため、データ生成分布が確率モデルで実現可能でない場合を考察することができないという問題があった。
2. 研究の目的 本研究は統計的推論における自由エネルギー(マイナス対数周辺尤度)に関して次の目的の研究を行った。(1) 自由エネルギーの性質を従来よりも一般的な状況下において解明することにより、応用可能な用途を拡張する。(2) 階層構造を持つモデルの自由エネルギーを解析することで、確率モデルの構造が推測精度に与える影響を明らかにする。(3) データ生成分布が確率モデルで実現できるという前提条件を必要としない場合にベイズ統計学の基盤を拡張し、自由エネルギーの挙動を導出し、自由エネルギーの最小化と汎化誤差の最小化の相違を明らかにする。
3. 研究の方法 サンプルサイズを  $n$  とする。自由エネルギー(マイナス対数周辺尤度)  $F$  は、マイナス対数尤度関数をハミルトニアン関数として定義されており、その値はエネルギーの総和  $nE$  とエントロピー  $S$  の和に等しい。すなわち  $F=nE+S$  が成り立つ。次にエネルギーの総和  $nE$  はデータ生成分布と確率モデルとのカルバック・ライブラ距離の  $n$  倍であり、またエントロピー  $S$  はサンプルサイズの対数  $\log n$  と実対数閾値 の積で表される。すなわち  $S= \log n$ 。さらに実対数閾値はパラメータ空間の双有理不変量であることからハミルトン関数の特異点を双有理写像を用いて正規交差に変形したときの個々のパラメータの指数を用いて計算することができる。以上の数学的手段によって自由エネルギーの値を数学的に求めることができる。以上より  $F=nE+ \log n$  が成り立つ。これは特異学習理論によって基礎づけられている数学的方法である。この方法を次の各場合において適用する。(1)  $X_1, X_2, \dots, X_n$  が交換可能であるとは、それらの同時分布から定義される任意の部分集合の周辺分布が変数の置換に依存しない場合のことをいう。サンプルが独立でなくても交換可能な場合には、ある汎関数確率分布  $Q$  に従う「確率分布に値を取る確率変数  $q$ 」が存在して、データ生成分布は独立な確率分布  $q(x_1)q(x_2)\dots q(x_n)$  の  $Q$  による混合として表すことができるので、このことを利用して、各  $q$  ごとに従来の理論を適用することで、自由エネルギーの挙動を導出することができる。特にサンプルサイズ  $n$  を無限大にする極限で種々の確率変数の値は、 $q$  により条件づけられた平均に収束し  $Q$  による平均そのものには収束しない。これにより任意の  $q$  毎に WBIC により自由エネルギーの値を近似することができることが解明される。(2) 確率モデルが ReLU 関数を用いた畳み込み神経回路網であるときには、スキップ接続を持つ場合と持たない場合において、自由エネルギーの主要値を与える局所的なパラメータ集合が異なることから自由エネルギーの挙動も異なることを示すことができる。特に確率モデルが十分に複雑で階層数が大きくデータ生成分布を実現できると想定できる場合には、エントロピーの値を定める の値がスキップ接続がある場合にはない場合よりも小さくなることが導かれる。このことからスキップ接続の存在が自由エネルギーを小さくするためにどのように機能しているかを解明することができる。(3) ユーザが確率モデルと事前分布を想定するとサンプルが交換可能であることが導かれることから、未知の汎関数確率分布  $Q$  に従うデータ生成分布  $q$  の存在も同時に導出される。なぜならば、もしもユーザが  $Q$  も  $q$  も存在しないと想定すると確率モデルと事前分布を設定できることとの矛盾が導かれるからである。これより  $Q$  と  $q$  が存在することから、データ生成分布がモデルにより実現可能でない場合を調べることが可能になる。パラメータ集合  $W$  を局所的な部分集合  $W_k$  の和集合と考え、個々の部分集合に対して局所的な自由エネルギー  $F_k$  を定義する。このとき  $F=\min F_k + \text{定数}$  が成り立つ。事後確率を最も大きくする部分集合は、局所的な自由エネルギーを最も小さくする部分集合と等しい。これより自由エネルギーの主要

値は、事後確率を最大にする部分集合により定まることがわかる。事後確率を最大にする部分集合は局所的なエネルギー $nE_k$ とエントロピー $-k \log n$ の和を最小にすることにより定まる。以上のことから、サンプルサイズが増大するにつれて、事後確率を最大にする部分集合は、 $nE_k + -k \log n$ を最小にするように「ある特異点の近傍から別の特異点の近傍へとジャンプしながら」移動していくことを示すことができる。この方法により、データ生成分布が確率モデルにより実現できない場合を数学的に解析することができる。

4. 研究成果 本研究では、以下の成果が得られた。(1) サンプルが独立性の仮定を満たしていない場合においても、交換可能であれば、WBICにより自由エネルギーの近似値を求めることができることを明らかにした。(2) 畳み込み神経回路網においてスキップ接続を持つ場合と持たない場合においてWBICの主要値を定めている実対数閾値がどのように異なるかを明らかにし、スキップ接続が自由エネルギーを小さくするという意味において有用であることを示した。(3) 一般的な状況においては、複数の確率モデルを準備してモデル選択が適切に行われたとしても、未知のデータ生成分布はモデルにより実現可能にはならない。このため、あらかじめ用意したモデル族によりデータ生成分布が実現可能であることを前提とする20世紀の古いベイズ統計では階層型の神経回路網のような現代の学習モデルの性質を取り扱うことができなかったが、汎関数確率分布 $Q$ に従う確率分布 $q$ の設定を基盤として、確率モデルのパラメータ近傍のエネルギーとエントロピーのバランスによって事後分布が相転移を持つことを示すことができることを示し、階層構造や隠れた変数を持つモデルの統計的推論を数理的に扱うことができることを明らかにした。

## 5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 6件/うち国際共著 0件/うちオープンアクセス 3件）

1. 著者名 S.Nagayasu, S.Watanabe	4. 巻 222
2. 論文標題 Free Energy of Bayesian Convolutional Neural Network with Skip Connection	5. 発行年 2024年
3. 雑誌名 Proceedings of the 15th Asian Conference on Machine Learning	6. 最初と最後の頁 927-942
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Sumio Watanabe	4. 巻 47
2. 論文標題 Mathematical Theory of Bayesian Statistics where all models are wrong	5. 発行年 2022年
3. 雑誌名 Handbook of Statistics, Elsevier	6. 最初と最後の頁 209-238
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/bs.host.2022.06.001	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Sumio Watanabe	4. 巻 -
2. 論文標題 Mathematical Theory of Bayesian Statistics for Unknown Information Source	5. 発行年 2023年
3. 雑誌名 Philosophical Transactions of the Royal Society A	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1098/rsta.2022.0151	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Sumio Watanabe	4. 巻 -
2. 論文標題 Recent Advances in Algebraic Geometry and Bayesian Statistics	5. 発行年 2022年
3. 雑誌名 Information Geometry	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s41884-022-00083-9	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Sumio Watanabe	4. 巻 4
2. 論文標題 Information criteria and cross validation for Bayesian inference in regular and singular cases	5. 発行年 2021年
3. 雑誌名 Japanese Journal of Statistics and Data Science	6. 最初と最後の頁 1-19
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s41237-021-00133-z	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 渡辺澄夫	4. 巻 31
2. 論文標題 代数幾何と統計理論	5. 発行年 2021年
3. 雑誌名 応用数理	6. 最初と最後の頁 7-14
掲載論文のDOI (デジタルオブジェクト識別子) 10.11540/bjsiam.31.3_7	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

[学会発表] 計3件 (うち招待講演 1件 / うち国際学会 1件)

1. 発表者名 Sumio Watanabe
2. 発表標題 Singularity Theory in Statistical Science
3. 学会等名 MSJ-SI: Deepening and Evolution of Applied Singularity Theory (招待講演) (国際学会)
4. 発表年 2022年

1. 発表者名 広瀬青、渡辺澄夫
2. 発表標題 入力分布が低次元超平面上にあるときの縮小ランク回帰の実対数閾値
3. 学会等名 信学技報, vol. 121, no. 419, IBISML2021-32,
4. 発表年 2022年

1. 発表者名 Takumi Watanabe, Sumio Watanabe
2. 発表標題 Asymptotic Behavior of Bayesian Generalization Error in Multinomial Mixtures
3. 学会等名 arxiv.org
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関