

研究終了報告書

「相互作用系を活用したサンプリング手法の開発」

研究期間：2019年10月～2022年3月

研究者：二見 太

1. 研究のねらい

ベイズ推論は機械学習の中核をなす推論方法であるが、ベイズ事後分布の厳密な評価は計算量の観点から難しいため、多くの場合何らかの近似に頼る必要がある。近年では取り扱うデータは巨大になり、またモデルも高次元化しているため、モデルの次元、データ双方に対してスケラブルなアルゴリズムの開発が強く求められている。大規模なデータに対しては並列型の計算システムを使うということが不可欠になっている。ランジュバンモンテカルロを使った方法はこの2つの要請を満たすものであるが、応用上、複数のモンテカルロ鎖を並列に回すことで大量のサンプルを得ることが良く行われる。こうした既存手法の問題点として、複数のサンプルが同時に得られるにも関わらず、理論的には誤差が並列でない場合と比較して減少しないことがあげられる。これは並列の互いのサンプルの位置に関しての情報を陽に使わないため、並列性を有効に活用できず必ずしも位相空間の探索が効率的に行われてこなかったためであると考えられる。我々はこの問題を克服するために、並列なランジュバンダイナミクスの間に適切な相互作用を陽に入れ、位相空間をより効率的に探索できるようなアルゴリズムの開発を目標とした。今回の研究により、相互作用を導入することで、探索がより効率的になり、定常分布への収束が早まることで最終的な誤差の減少につなげるという利点を得ることが期待される。これは従来の計算量を減らすことを目的とした並列計算とは異なる利点を持つ全く新しい手法であり、学術的に重要な貢献を生み出すことを目指すものである。

一方でベイズ推論などの場合には、定常分布への収束に加えて、汎化能力が実応用上では重要であることを踏まえ、サンプリングや最適化によって得られた複数のモデルから統合して予測を行う際に、モデルのばらつきが汎化能力にあたえる影響を解明することを目指した。さらにこの解明により、汎化能力を高めるための適切な相互作用の構築を行い、新たなアンサンブル学習の手法を開発する。さらにこうしたベイズ推論における、並列間のばらつきや相互作用が加速および汎化性能にもたらす影響を多角的にとらえることで、確率的推論におけるアンサンブルに対する革新的な理論を構築することを目指すものである。

2. 研究成果

(1) 概要

研究期間の前半及び後半でそれぞれ以下のテーマを扱った。

前半で行ったものが当初の研究計画である、1) 並列モンテカルロにおける相互作用の導入により、定常分布への収束の加速法の解析、及びアルゴリズムの開発である。

後半で行ったものが、前半の計画をふまえて当初の計画を変更しさらに並列のモデルに関して機械学習の立場から理論を深化させた、2) 機械学習モデルのアンサンブル等における、相互作用の導入による汎化誤差への影響の解析、および、及びアルゴリズムの開発である。以下はこれらについてそれぞれ概略を述べる。



1)については、当初の研究計画に従い、ベイズ推論のサンプリングという問題に関して特に、ランジュバンダイナミクス(LD)と相互作用を使ったサンプリングにおける、相互作用の構成の仕方、およびハミルトンモンテカルロ系への拡張、またこれらの技術の凸最適化問題への応用といった研究を行った。より具体的には、どのような相互作用が定常分布への収束を早めるために理論的に望ましいのかということ、目的のベイズモデルのポテンシャル関数に注目することで、収束の加速が発生するための十分条件を導き、その条件を満たす相互作用を構成するアルゴリズムの開発を行った。またより実用面で優れた、ハミルトンモンテカルロの手法と組み合わせた並列アルゴリズムの開発を行った。

2)に関しては、ベイズ推論などの場合には、定常分布への収束に加えて、汎化能力が実応用上では重要であることを踏まえ、サンプリングや最適化によって得られた複数のモデルから統合して予測を行う際に、モデルのばらつきが汎化能力に与える影響を解明し、新しいアルゴリズムの開発を行った。特に我々はモデルのばらつきに関する情報を自然に取り込んだ、新しい汎化誤差上界の不等式を導出した。これにより我々は、汎化誤差を効果的に縮めるためにどのようなモデル間のばらつきが望ましいのかを明らかにした。さらに新たに得られた汎化誤差上界と既存のアンサンブル学習の手法が直接関係していることを明らかにした。一方でこの新しい汎化誤差を直接最適化することで、汎化誤差を効果的に縮めるための新しいアンサンブル学習を開発した。

(2) 詳細

研究テーマ1「並列モンテカルロにおける相互作用の導入による定常分布への収束の加速に関する研究」

分散されたシステムにおける並列な LD の間に適切な相互作用を陽に入れ、位相空間をより効率的に探索できるようなアルゴリズムの開発を目指すことが当初の本研究における最終ゴールであった。それを達成するために、研究テーマ1では分散システム間におけるコミュニケーションコストという問題が発生しない、単純な1台の計算機のみという設定において、相互作用を含む並列な LD の理論及び数値的挙動に関する研究を行った。そして相互作用を含む並列な LD の基本的な理論・数値的挙動を明らかにした。

図 1

■ 目的分布 $\pi(x) \propto e^{-\beta U(x)}$ LDサンプリングは

$$X_{k+1} = X_k - \nabla U(X_k)h + \sqrt{2h\beta^{-1}}\epsilon_k, \epsilon_k \sim N(0, I)$$

既存方法(N並列)

$$X_t^{\otimes N} := (X_t^{(1)}, \dots, X_t^{(N)})^\top \quad \nabla U^{\otimes N}(X_t^{\otimes N}) := (\nabla U(X_t^{(1)}), \dots, \nabla U(X_t^{(N)}))^\top$$

$$X_{k+1}^{\otimes N} = X_k^{\otimes N} - \nabla U^{\otimes N}(X_k^{\otimes N})h + \sqrt{2\beta^{-1}h}\epsilon_k$$

提案方法

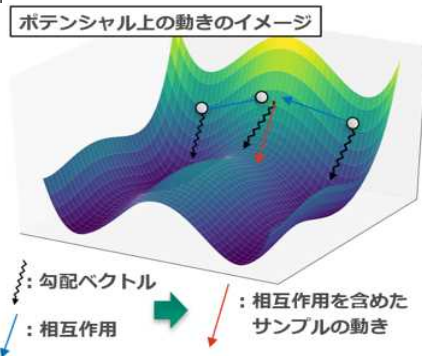
$$X_{k+1}^{\otimes N} = X_k^{\otimes N} - (I + \alpha J) \nabla U^{\otimes N}(X_k^{\otimes N})h + \sqrt{2\beta^{-1}h}\epsilon_k$$



我々は LD において図1および図2のとおり勾配ベクトルを通して相互作用を導入する枠組みに注目を行った【代表的な論文2】。その際には相互作用をどのように設計するかが重要な課題となる。まずそのために、相互作用を入れても定常分布の形が変化しない条件について理論的に導出を行った。この条件

件は直感的には、相互作用 により、あるサンプルを動かしたら、その分だけ他のサンプルを

図 2



動かして全体の動きを相殺することを要求する。アルゴリズムとしてこの条件をできるだけ計算量およびメモリーを少なく満たすために、我々は反対称行列 $\times$ 勾配で設計を行うことを考えた。図1における J が反対称行列である。さらにこうした相互作用を導入すると定常分布を変化させないだけではなく、ポテンシャル関数の実効的なヘシアン行列の固有値の最小値が変化し、加速が発生することを理論的に明らかにした。

直感的には図3のとおり反対称行列が加わることで、ポテンシャル関数の鞍点 (saddle point)

図 3



および Global minima, local minima 周辺のダイナミクスに対する幾何的な構造 (ヘシアン行列の最小の固有値) が変化することで、加速が発生することを理論的に明らかにした。

また反対称行列 J をどのように構成すれば加速が発生するのかその条件を明らかにした。その条件を満たすために、我々はある種のランダム行列を活用す

れば高い確率で加速を発生させることを明らかにした。また反対称行列の保存のメモリーをできるだけ小さくするため、クロネッカー積を使った構造を導入することで加速の保証をしつつ必要なメモリー量を並列数の 2 乗に抑えるアルゴリズムを開発した。また提案法により関数の期待値を計算する際に、Non-asymptotic な設定でどの程度の期待値の誤差が発生するのか理論解析を行った。こうした理論を踏まえて実際にアルゴリズムを構成し、多くのベイズモデルにおいて数値的に有用性を検証した。またより実用的なサンプリングとしてLDの発展形である、ハミルトンモンテカルロにおいても、上述の理論を発展させ、相互作用付きのハミルトンモンテカルロの理論、およびアルゴリズムを開発し、同様に数値実験によってその有用性を検証した【「その他の成果物(著作物1)」】。

また「その他の成果物(著作物2)」のとおりサンプリングだけではなく、相互作用を凸最適化に応用する手法を ACT-X 領域外の研究者と開発し、論文として発表を行った。

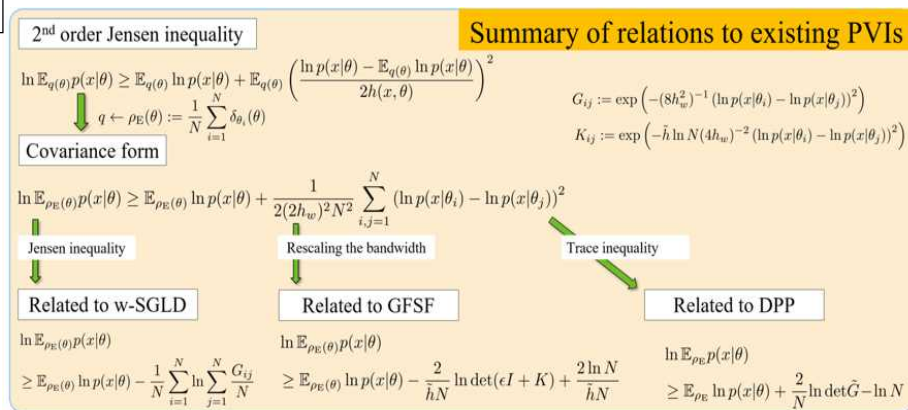
研究テーマ2「ベイズ推論におけるモデルのばらつきと汎化に関する研究」

研究テーマ1では、相互作用する並列化により、より効率的にパラメータ、モデル空間を探索することで、サンプリングの収束速度を高めることが最終目的であったが、一方で手法を機械学習におけるベイズ推論で実装を行う際には、収束速度よりも未知のデータに対する汎化性能を高めることもまた非常に重要な課題である。こうした観点に立つと並列化、特にアンサンブル学習や、ベジアンモデルアンサンブルのように並列な複数のモデルを統合して予測を行う場合に、そのモデル同士の間のばらつき、多様性が汎化や学習にとってどういう役割を果たすのか、多くの点が未解明であることが分かった。そこで我々は当初の計画を変更し、研

究テーマ1の手法を分散システムへの拡張を行う前に、汎化能力とモデルのばらつきに関して理論解析を行い並列のモデルを使うことの利点を説明することを目的とした研究を行った。

上述の目的を達成するため、我々は汎化誤差をまずは Kullback Leibler 距離で、かつ PAC-ベイズの枠組みでとらえることにした。そして並列モデル間の多様性をとらえるために、我々は Jensen の不等式を拡張し並列モデル間のばらつきの情報を含む 2 次の Jensen の不等式を導出した。図4はその概要である。図4において、 x はデータ、 θ はモデル $p(x|\theta)$ のパラメータであり、 $q(\theta)$ が何らかの手法で得られた事後分布もしくはその近似である。2次の Jensen の不等式はばらつきの項の分だけ通常の Jensen の不等式よりも強い不等式になっていることがわかる。次に 2 次の Jensen の不等式をさらに図4のとおり様々な不等式を使うことでさら

図 4



に変形をすることができ、この変形を行って出てくる不等式たちが既存の様々な手法の目的関数となっていることを確認した。また汎化誤差について、データ数が D の時には、任意の事前分布 π に対して、データを i.i.d. で引いたときに確率 $1 - \xi$ で以下の PAC-Bayesian の汎化誤差が成立すること説明した【代表的な論文1】。

$$CE \leq -\mathbb{E}_{\nu, q} [\ln p(x|\theta) + R(x, h)] \leq -\mathbb{E}_q \frac{1}{D} \sum_{d=1}^D [\ln p(x_d|\theta) + R(x_d, h_m)] + \frac{KL(q, \pi) + \frac{\ln \xi^{-1} + \Psi''_{\pi, \nu}(c, D)}{3}}{cD}$$

ここで R は図4の 2 次の Jensen の不等式のばらつきの項である。これにより、複数モデルを学習する際には、負の対数尤度を最小化することによるモデルフィッティングとばらつきの項を大きくすることによるモデル間の多様性を増やすことのトレードオフを制御することが必要であることを説明した。最後にこの PAC-Bayesian の汎化誤差上界を直接最適化することでこのトレードオフを制御するアルゴリズムを開発した。実データを使った多くの場合において、この手法は既存の並列モデル最適化手法と同等、もしくはそれ以上の性能を発揮することが分かった。

3. 今後の展開

本研究を得て着想を得たベイズ推論でない、確率的モデリングの研究を今後推し進める予定である。研究テーマ2で得られたばらつきを加味した確率的推論はギブス事後分布の一種であり、厳密にはベイズ推論ではない。しかしギブス事後分布は近年ベイズ事後分布より理論的、数値的に優れた性能を示すことが知られている。さらにこうしたギブス分布の持つ性質についてより研究を進める予定である。また Model misspecification とモデルのばらつきの関係をより深化させる予定である。

4. 自己評価

まず研究の達成状況について、研究テーマ1により相互作用ありのモンテカルロ法について、当初の計画におけるコミュニケーションコストのない場合の設定での理論的かつ数値的な検証は達成できたと考えている。一方で当初の計画では含めていなかった研究テーマ2の内容により相互作用の汎化能力に関する解析、及びアルゴリズムの開発という機械学習分野における重要な貢献を達成できたと考えている。

次に研究の進め方について、本研究プロジェクトは個人型であることを踏まえ、研究者本人を主担当者として、適宜それぞれの研究テーマに精通している企業もしくは大学との共同研究として研究を進めた。研究者本人の独自性を発揮しつつ、研究の客観的な批評が生かされた適切な体制であったと考えられている。

研究成果の科学技術及び社会・経済への波及効果について、本研究は基礎研究であるため、直接社会・経済への波及効果をもたらすものではないものの、本研究を通し、複数のモデルからの情報を統合し予測を行うというプロセスをえて機械学習を実行する際に、その学習の速さや、汎化性能がどのように異なるのか、これまでの機械学習で検討がされてこなかった知見を多く得ることができた。これにより3. 今後の展開で記述したような研究テーマや、現在応用で積極的に研究されている、エッジコンピューティングにおける機械学習システムの導入への応用研究などへの展開が強く期待される。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 5件

1. Futami, F., Iwata, T., Ueda, N., Sato, I., & Sugiyama, M. Loss function based second-order Jensen inequality and its application to particle variational inference. Advances in Neural Information Processing Systems 35, 2021. (NeurIPS2021)

汎化能力とモデルのばらつきに関して理論解析を行った。汎化誤差を KL 距離で測る場合に注目し、アンサンブル間の多様性をとらえるために、モデル間のばらつきの情報を含む 2 次の Jensen の不等式を導出した。これを用い、ばらつきの情報を含んだ 2 次の PAC ベイズの汎化誤差上界を導出した。そして多くの既存のアンサンブル法がこの上界から導かれること

を示し、またこの上界を直接最適化すると数値的に優れた性質を示すことを確認した。

2. Futami, F., Sato, I., & Sugiyama, M. Accelerating the diffusion-based ensemble sampling by non-reversible dynamics. In H. Daumé III and A. Singh (Eds.), Proceedings of 37th International Conference on Machine Learning (ICML2020), Proceedings of Machine Learning Research, vol. 119, pp. 3337–3347, online, Jul. 13–18, 2020.

並列 LD において勾配ベクトルを通して相互作用を導入する枠組みを提案した。そのために、相互作用を入れても定常分布の形が変化しない条件について理論的に導出を行い反対称行列を使えばその条件が満たされることを示した。また提案法により関数の期待値を計算する際に、Non-asymptotic な設定でどの程度の期待値の誤差が発生するのか理論解析を行った。こうした理論を踏まえて実際にアルゴリズムを構成し、多くのベイズモデルにおいて数値的に有用性を検証した。

3. Futami, F., Iwata, T., Ueda, N., Sato, I., & Sugiyama, M. Predictive variational Bayesian inference as risk-seeking optimization. In Artificial Intelligence and Statistics, 2022. (AISTATS 2022).

代表的な論文1の手法をさらに詳細に理論解析を行った、特に Model misspecification の場合を含めてパラメータ性能の性質である Consistency をもつことを明らかにした。また汎化誤差として KL 距離を考えた際に、そのほかの指標尺度、たとえば最小 2 乗距離の意味でも適切な性能を示すことの理論的な解明を行った。さらに Over-parametrized なモデルで代表的な論文1が良い性能を示すことを Risk sensitive learning との関係から明らかにし、深層生成モデルへの拡張を行った。

(2) 特許出願

なし

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

(著作物)

1. Futami, F., Iwata, T., Ueda, N., & Sato, I. Accelerated Diffusion-Based Sampling by the Non-Reversible Dynamics with Skew-Symmetric Matrices. Entropy 2021, 23(8), 993.
2. Futami, F., Iwata, T., Ueda, N., & Yamane, I. Skew-symmetrically perturbed gradient flow for convex optimization. In Balasubramanian, V. N. and Tsang, I. (Eds.), Proceedings of The 13th Asian Conference on Machine Learning, volume 157 of Proceedings of Machine Learning Research, pp.721–736. PMLR, 17–19 Nov2021.

(学会発表)

1. The 35th Conference on Neural Information Processing Systems, オンライン, Nov. 8 2021
2. The 13th Asian Conference on Machine Learning, オンライン, Nov. 18 2021
3. 第 24 回情報論的学習理論ワークショップ, オンライン, Nov. 10, 2021
4. 第 23 回情報論的学習理論ワークショップ, オンライン, Oct. 25, 2020
5. The 37th International Conference on Machine Learning, オンライン, Jul. 14, 2020