

## 論文の内容の要旨

論文題目 Robust and Computationally-Efficient Approximate Bayesian Inference  
(ロバストで計算効率の良い近似ベイズ推論に関する研究)

氏 名 二見 太

**序論 (1章):** 観測データを元にしてそのデータの生成メカニズムに関し仮説を立てることは自然科学から社会科学まで広く有効な手段として行われている。しかしデータ生成の真の仕組みを完全に記述することは、完全に制御されたシミュレーションや実験環境等以外では難しく、そうした不確からしさを表現するため確率を利用してモデルを作るということがしばしば行われる。特に統計的推論では、必ずしもデータ生成の真のメカニズムを求めることに重きを置くのではなく、手元にある限られた有限個のデータを用いて、まだ観測していない未知のデータに関する予測をうまく行うことが可能な汎化能力の高いモデルを得るということを目的としている。推論の代表的な方法として、例えば、モデルを調整することで、観測データを尤もらしく生成する仕組みを得るという手法などがあげられる。

推論を行う際には観測ノイズやサンプルの数が有限であることによる**不確実性**が発生する。こうした不確実性を評価しつつ予測を行う方法として**ベイズ推論**は有効な方法である。特にベイズ推論では確率を**確信の度合い**と考える。ベイズ推定では、パラメータは事前知識や専門的知見を反映する事前分布に従う確率変数として取り扱われ、観測データの尤もらしさを表す尤度関数及び**ベイズの定理**を用いて、事前分布が事後分布に置き換わり確信の度合いが更新される。近年ベイズ推論は社会科学や医療の分野をはじめとして実問題への活用が飛躍的に増加している。この理由としてベイズ推論は、結果（観測データ）が判明している時に、それをもたらした原因の確率を求める逆問題に対して有用でことがあげられる。これは原因をまず事前確率に従うと仮定し、ある原因の下での結果が尤度関数を用いて表現されるとすれば、結果が与えられた時の原因の確率は事後分布として自然に求められるからである。この他ベイズ推論ではモデルや予測に関する不確実性についても、確率を通じて考えることができ、予測をどの程度信頼できるのかといった情報を得ることが可能になる。

ベイズ推論はこの他にも様々な利点を持つが、実用の際には事後分布や予測分布を解析的に求めることや厳密に数値的に評価することが難しく、何らかの**近似**を使う必要がある。ベイズ推論自体の歴史に比べて、その実問題への活用が近年始まった理由として、こうした近似を適切に行うための、計算機の進展や優れたサンプリングの手法の進展、サンプリング以外の計算効率の良い近似手法の開発が必要であったことがあげられる。つまり優れた近似手法の開発はベイズ推論にとって、本質的に重要である。現在は様々な近似手法が開発されているが、近似手法の選択の際にはモデルやデータの特性を考慮して選ぶ必要がある。特に尤度関数と事前分布、さらに計算量や事後分布として取り出したい情報を踏まえて近似手法を選択する必要がある、これらの適切な組み合わせを行うことがベイズ推論

を実際に活用する際には重要になる。

この博士論文では、ベイズの近似手法について、特に観測データに**外れ値**が存在する場合に着目して議論を行った。外れ値のあるデータとは、観測データを生成するメカニズムに異常があり、私たちが本来解析したい現象に関するデータに加えて、それとはあまり関係のない汚染が観測データに加わっているという状況である。そうした外れ値は、私たちが解析したい現象を必ずしも反映しておらず、大多数のデータとは非常に異なった振る舞いを示す。こうした外れ値は観測データの分布の端に多く発生するため、たとえ少量でも推定結果に大きな影響をもたらすことが知られている。特に近年取り扱われるようなデータ、例えばセンサーからのデータには外れ値が存在することが多く、またクラウドサービスを活用したデータにはラベル付けミスが含まれていることから、こうした外れ値に対して**ロバストなアルゴリズムの開発**は、実問題において非常に重要である。

そしてそうした外れ値の影響を軽減するためのロバスト推論と近似手法に関して、本博士論文では、2つの新しい研究成果（3章、4章）、及び既存の近似手法の利点を組み合わせた性質を持つ新しい近似手法（5章）を提案した。

**準備と関連研究（2章）：** 2章では、論文全体を通して必要となるベイズ推論とロバスト推論に関する基礎概念について例を取り上げつつ導入を行った。具体的にはまずベイズ推論に関する導入をし、更にベイズ推論において解析的に事後分布の計算ができる場合について説明を行った。その後解析的に計算が不可能な場合に使われる近似手法について、代表的な手法を紹介し、特にパラメトリックな近似として、**変分推論**、**仮定密度フィルタリング**、**期待値伝搬法**について説明を行った。またサンプルを使った近似として**確率的ランジュバン法**について導入を行った。これらの手法は論文全体を通して数値計算等において使われるのと同時に、先行研究として我々の提案する手法の土台として扱われた。次に外れ値、およびロバスト性について具体的なモデルを用いつつ解説を行った。さらにロバスト推論で広く使われる**影響関数**の概念を導入し、どのような場合にロバスト性を得ることが可能なのかについて議論を行った。

**t 指数分布族とそれに対する期待値伝搬法（3章）：** ベイズ推論の枠組みの中で外れ値に対するロバスト性を得る手段としてモデル自体を修正する手法が多く採用されてきた。例えばモデル中のガウス分布を、より裾が長いために外れ値に強い**スチューデント t 分布**に入れ替えてロバスト性を得る方法などである。しかしスチューデント t 分布はガウス分布と違い、**指数分布族**と呼ばれる分布族に属しておらず、その様々な有用な性質をもっていないため、指数分布族に属するガウス分布のように計算効率の高い近似アルゴリズムを開発することが困難になる。そこで本博士論文では特殊な代数（**q-代数**）を使うことで、こうしたスチューデント t 分布においても指数分布族の持つ有用な性質を扱えるようにし、解析的な計算が可能であることを示した。そして、指数分布族に対する期待値伝搬法を、スチューデント t 分布を含む、一般化された指数分布族（t 指数分布族）に対して適用可能になるように一般化を行い、スチューデント t 分布に対しても計算量の観点から効率的なアルゴリズムを構成した。さらにこの効率的なアルゴリズムにより新しいスチューデント

t 分布を使ったモデリングである、スチューデント t 過程分類器を提案した。

**ロバストな距離尺度を用いた変分推論（4章）：** 上述のモデルを変更することでロバスト性を得る方法は確かに有効ではあるものの、例えば現代の機械学習において使われる複雑なモデル（例えばニューラルネットワークモデルなど）ではそもそもどのようにモデルを変更することがロバスト性を得るために適切なのか自明ではなく、またその探索のために、モデルの変更を色々と試行錯誤することは、非常に時間がかかるため望ましくない。そこでモデルを変更するのではなく、ベイズにおける学習の方法を変更することでロバスト性を系統的に得る手法を開発した。ベイズ推論ではベイズの定理が中心的役割を果たすが、まずベイズの定理自体がある種の最適化の問題の解と捉えなおすことができる。その解釈をもとにすると、ベイズの定理では全ての訓練データ点を同じ重みで取り扱っているために、外れ値にも本当のデータ点と同じ重みで取り扱うことになり、外れ値の影響を受けやすくなっているということが分かった。そこで外れ値に対しては小さな重みを与えるような枠組みを**ロバストな距離尺度**によって導入し、外れ値をその重みによって自動的に無視する手法を開発した。このような手法で得られる事後分布はベイズの定理で得られる事後分布と違い、外れ値の影響が自動的に無視されていることになるため、モデルの変更を行わなくても、外れ値に強い学習が行われたことになる。更にこの手法について変分推論を用いることで効率よく計算できるアルゴリズムの構成を行い、そのアルゴリズムのロバスト性について影響関数を用いて議論を行った。

**フランク・ウルフのアルゴリズムを用いた逐次近似ベイズ推論（5章）：** 論文3章、4章の近似手法は真の事後分布を、何らかのパラメトリックな分布で近似する手法である。こうしたパラメトリックな近似は一般に計算量の観点から優れており、高次元にも適用ができる方法であるが、強い仮定をおくため、真の事後分布からの偏りが発生し、この偏りの大きさを理論的に保証することは難しい。この他の近似としてサンプリングによる近似があげられるが、この手法では十分たくさんのサンプルさえ用意できれば事後分布を十分な精度で近似でき、その精度は理論保証できる。しかし高次元や多峰性を持つ分布では、計算量が莫大になってしまうという欠点を抱えている。そこで我々はこうしたそれぞれの近似の性質の利点を組み合わせた、偏りについて理論保証がありかつ計算効率の良い手法を開発した。そのために我々は近似分布をサンプリングの手法のように有限個の点による経験分布で構成し、ある最適化問題を解くことでそれら有限個の点を推定するアルゴリズムを考案した。具体的には我々は最大平均距離と呼ばれる距離尺度を再生核ヒルベルト空間上で最小化する、制約付き凸最適化問題を構成し、それを**フランク・ウルフのアルゴリズム**を使い逐次的に解くことで理論保証付きのアルゴリズムの構成を行った。

**まとめと今後の展望（6章）：** 最後の6章ではこれまでのまとめ及び今後の展開、特に4章で行ったベイズ推論の再定義を基にした一般化ベイズに関する更なる進展や5章で提案した近似手法に基づく新しいサンプリング手法などに関して議論を行った。