

特集にあたって

黒田 正博 (岡山理科大学)

本特集「統計計算アルゴリズム」は、データ解析のパラメータ推定に用いられる計算アルゴリズムの紹介を目的としている。

パラメータ推定アルゴリズムとして、まず思いつくのが最小二乗法とニュートン-ラプソン法ではないかと思う。たとえば、一般線形モデルのパラメータ推定において、使われるアルゴリズムが最小二乗法である。それに対して、一般化線形モデルのパラメータ推定においては、ニュートン-ラプソン法が使われる。これは、一般線形モデルのパラメータ推定問題が線形方程式の解を求めるのに対して、一般化線形モデルでは非線形方程式を解くことため、ニュートン-ラプソン法による反復解法が必要になるためである。

ニュートン-ラプソン法は二次収束性を持ち、非常に少ない反復回数で解を求めることができる。しかしながら、収束は初期値に依存することが知られており、初期値の選択が適切ではないとき、アルゴリズムが収束しないことや、異なる解に収束することがある。また、反復関数の導出において、対数尤度関数の一階微分と二階微分の計算が必要であり、計算コストが高くなる。したがって、対数尤度関数の形が複雑である、あるいは、高次元データであるとき、この解法の適用が困難となることが予想される。このため、これらの弱点を補うアルゴリズムが開発されている。

計算コストの減少については、対数尤度関数の二階微分の代わりに二階微分の期待値を用いる Fisher Scoring 法がある。また、対数尤度関数が複雑となる場合として、欠測が含まれる観測データ (不完全データ) からのパラメータ推定がある。この場合、欠測部分に何らか値を補完 (imputation) して疑似的に完全データ (欠測がなくすべてが観測されたデータ) を生成し、従来の枠組みで推定を行う。最尤推定に基づくものに Expectation-Maximization (EM) アルゴリズムがある。一方、バイズ推定による方法に多重補完法 (Multiple Imputation) やマルコフ連鎖モンテカルロ法がある。高次元データに対しては、パラメータ推定に加えて観測個体のクラスタリングも解析の目的となることがあり、そのためのアルゴリズムが開発されている。

今回の特集では、次の統計計算アルゴリズムを紹介する。まずは、伝統的な統計計算アルゴリズムである最小二乗法とその拡張アルゴリズムである交互最小二乗法である。次に、ニュートン-ラプソン法を改良した Fisher Scoring 法を取り上げる。不完全データからのパラメータ推定法として、EM アルゴリズムと多重補完法を紹介する。最後は、高次元データのクラスタリングのためのクラスタリングアルゴリズムである。

森裕一氏 (岡山理科大学) は、最小二乗法と交互最小二乗法の紹介を行っている。前半では、誤差論の観点から、最小二乗法の原理と解法の説明をしている。後半では、目的関数が複数の制約条件をもつ場合などに適用される交互最小二乗法を紹介している。

高井啓二氏 (関西大学) は、Fisher Scoring 法と不完全データへの拡張を紹介している。完全データに対する Fisher Scoring 法について解説した後、不完全データへの拡張について導出から収束スピードまでを解説している。さらに、Fisher Scoring 法と EM アルゴリズムの関係にも言及している。

黒田は、EM アルゴリズムの導出と収束性などの基本的な内容を概説している。また、統計モデルの確率分布が指数型分布族に属する場合のパラメータの十分統計量の方程式と EM アルゴリズムの反復関数の関係について示している。

阿部貴行氏 (京都女子大学) は、多重補完法の紹介を行っている。多重補完法は理論的妥当性をバイズ理論におき、幅広い種類の欠測データの解析に使用できる汎用的な手法の一つである。多重補完法のアルゴリズムに加え、その特徴や理論的根拠を解説している。

宇野光平氏 (名古屋大学) は、クラスタリングアルゴリズムの紹介を行っている。アルゴリズムには、非階層的クラスタリングの k-means クラスタリング、階層的クラスタリングでは近年注目されて convex クラスタリングを取り上げて説明している。

最後に、本特集にご協力いただいた各著者の方々、ならびに学会機関誌編集委員の先生方に心より感謝申し上げます。本特集が読者の皆様にとって、今後のご参考となれば甚だ幸いである。