

外れ値が多い場合でもバイアスの小さいロバスト推定

藤澤洋徳

統計数理研究所 数理・推論研究系

概要

ロバスト統計では外れ値に強い推定かどうかを、潜在バイアスで測ることが一般的である。しかし、潜在バイアスを直接に評価するのは難しい。そのため、ロバスト統計では、潜在バイアスの代替指標として、外れ値の割合が小さいという想定の下に潜在バイアスを近似的に表現できる影響関数や、潜在バイアスが破綻しない最大の外れ値の割合としての破局点、という二つの指標がしばしば用いられる。本研究では、潜在バイアスを直接的に評価して、たとえ外れ値の割合が大きくても、外れ値の外れ値らしさが確かであれば、潜在バイアスを小さくできるロバスト推定を紹介する。それを可能にする手法は、ガンマ・ダイバージェンスに基づく。また、適当な仮定の下では、外れ値の割合が大きくても潜在バイアスを十分に小さくできる手法は、ガンマ・ダイバージェンスに基づく手法だけであることも証明できる。手法の良さは、ピタゴリアン関係によって自然に説明できる。パラメータ推定アルゴリズムは、ピタゴリアン関係を利用して構築でき、ガンマ・ダイバージェンスに対応するロス関数を単調に減少させる。近年に関連研究が加速している。ダイバージェンスとモデルを拡張することで、外れ値の割合をも推定することも可能である。パラメータ推定アルゴリズムを、ピタゴリアン関係からではなく、Majorization-Minimization アルゴリズムから構築することで、様々なモデルに拡張可能になり、線形回帰モデル・ロジスティック回帰モデル・ポアソン回帰モデル・ガウシアングラフィカルモデルなどにも利用可能である。