

# 統計的モデル選択の問題

山口 健太郎 (Kentaro Yamaguchi)  
京都大学大学院文学研究科

統計的推論は、コイン投げのような確率的な振る舞いをする現象を記述するために用いられる。この統計的推論の手続きは、(1)現象を記述するための統計的モデルの族(確率分布)をどのように特定するのかと、(2)モデルの族が与えられたときにどのようにしてモデルのもつパラメータを推論するのか、との2つに分けることができる。このような手続きを踏んだ結果として、現象を記述するモデルを推論することが可能となる。この際問題となるのは、(3)その推論のよさは何に基づいているのか、ということである。20世紀前半に始まる推計統計学の整備の過程で生じた、「どの立場による統計的推論が優れているのか」という論争は、おもに(2)や(3)の問題を中心に行われていたのだった。他方で、(1)の問題も重要であったにもかかわらず、それが論争の中心になることは少なかった。

赤池が提示した赤池情報量基準(AIC)は、いくつかの準備された統計的モデルの族から、最適なモデルを選択するための基準である。つまりAICによって、モデルの族の特定(すなわち(1))だけでなく、パラメータの推定(2)に相当)を一挙に行うことが可能となったのである。AIC以降、さまざまな統計的モデル選択基準が登場したが、(3)の問題はずっと複雑なものとなった。というのも、AICに基づいている統計的推論は、NeymanやPearsonによる頻度主義的な統計的推論と、Fisherによる尤度に基づく統計的推論の融合したものだが、それらを統一的に特徴づける概念的基礎がないからである。そのため、統計学の哲学において従来から論じられてきた、NeymanやPearsonによる頻度主義、SavageやLindleyらによるベイズ主義、Fisherによる尤度主義のうち、どの統計学の立場による推論が妥当なのかという観点から、統計的モデル選択基準の問題を捉えることはできにくくなってしまった。

そこで本発表では、このような統計的モデル選択基準がもつ統計的推論の問題の複雑さを明らかにし、統計的推論のよさは何によって与えられるのかについて再考してみたいと思う。具体的には、このような問題に対し、従来の統計学の哲学の論点、つまりどの立場による統計的推論が優れているのかという点から分析することが可能なのか、もし可能でないとすればどのような観点をとれば解決の糸口を見出すことができるのかについて、SoberやForsterらの論文を引用しながら検討してみたいと思う。