

Recent Developments in Extreme Quantile Regression

鹿児島大学 吉田拓真

経済損失, 自然災害の規模, 医療コストなどリスクに関する分野では, 平均値などのデータの中心的指標よりも, 最大値や最小値, またはそれに近い分布の裾に興味があることがある。その場合, 0.01%や99.9%分位点など境界に近いところの分位点 (extreme quantile, EQ) の予測が重要となる。ここでは, 便宜上, 高位な分位点 (最大値付近) のみに焦点を充てるが, 低位 (最小値付近) の場合も議論は同じである。

さて, EQは最大値に近い分位点を推定するわけであるが, 重要なのは, 現在得られているデータの最大値よりも大きいデータが得られる可能性を考慮しなければならない点である。通常の方分位点回帰の方法を用いても, その部分がうまく考慮されず, 極端に大きい分位点においては推定精度が良くないことが知られている。これを克服するためには, “分布の裾である” という情報を適切に盛り込んだ推定方法を考案しなければならない。そこで, 登場するのが極値統計学である。例えば, 独立同一分布から生成された n 個の確率変数の最大値を考えてみてほしい。この最大値が従う分布はもはや元のデータが従う分布と必ずしも等しくはならないだろう。では, どのような分布に従うのか? 極値統計学はこのような問題を解決してくれる。このような観点を今回の話に当てはめると, データを生成する真の分布があるとして, 高位分位点周りのデータのみを持ってくるとそれらはもはや元の分布ではなく, そこから導かれる分布 (極値分布と呼ぶ) に従うと見なして解析を行うべきである。ということである。極値分布を考慮すると, 上で述べた “現在得られているデータの最大値よりも大きいデータが得られる可能性” を視野にいれた推定量が構築できる。というわけである。

近年の研究では, ターゲットとなる変数の分布をより詳細に理解するために, 説明変数が導入されることが多い。その場合, 高位/低位な分位点回帰 (extreme quantile regression, EQR) が目的となる。本講演では, 分位点回帰と極値統計学を融合させた高位分位点予測のための方法, 理論, 実例を解説する。また, 最新の研究を紹介する。