

Heavy tail 分布の裾近似と POT 法における一意的な閾値決定法

鈴木佑輔 清水泰隆
早稲田大学 基幹理工学研究科

大地震などのまれにしか起こらないが損害額が大きい事故は保険会社に大きな打撃を与えるので平均値よりも最大値に興味の対象がおかれることが多い。また、損害額の分布は裾が重いことが経験的に知られている。

裾が重い分布の閾値 u に対する超過分布関数は、 u を十分大きくとると、一般化パレート分布で近似出来ることが知られている。この中の未知パラメータ推定法として” POT 法 (Peaks-Over-Threshold method)” があり、閾値 u の選択が推定に大きな影響を与えることが分かっている。しかし、従来の閾値 u の決め方は最終的にはグラフから視覚的に判断することが求められ、結果が主観的になるという欠点がある。

この問題を解消する新しい閾値決定法として、Gonzalez *et al.*(2013) により提案された、Kullback-Leibler 情報量を用いた手法がある。この論文では、Kullback-Leibler 情報量

$$I(g; f_{\hat{\theta}}) := E_G \left[\log \frac{g(Z)}{f(Z|\hat{\theta})} \right] = E_G[\log g(Z)] - E_G[\log f(Z|\hat{\theta})]$$

の右辺第 2 項の推定量として

$$\hat{f}_n(u) = \frac{1}{n} \sum_{i=1}^n \log \hat{T}_u(X_i)$$

を用いている。ただし、 \hat{T}_u は閾値 u 以前をカーネル密度推定、以降を一般化パレート分布で最尤推定したセミパラメトリックな密度関数である。しかし、この推定量はカーネル密度推定する際のバンド幅に大きく影響されるという問題がある。

本発表では閾値以前の分布にパラメトリックモデルを仮定することによりバンド幅の影響を解消し、AIC 型の情報量規準を用いたモデル選択によって閾値を定める手法を提案する。また、別の手法として hyperexponential 分布と全データで最尤推定した一般化パレート分布の交点を閾値とする手法も提案する。

最後には数値シミュレーションの結果を報告し、それぞれの方法の比較をする。

参考文献

- [1] Gonzalez, J.; Rodriguez, D. and Sued, M. (2013). Threshold selection for extremes under a semiparametric model, *Statistical Methods & Applications*, **22**, 481-500.
- [2] Feldmann, A. and Whitt, W. (1996). Fitting mixtures of exponentials to long-tail distributions to analyze network performance models, *Performance Evaluation*, **31**, 245-279.