—研究概要—

等相関正規ランダム行列と高次元統計学

東北大学大学院 理学研究科 数学専攻 アティナ・フスナキラティ

昨今の大規模高次元データの数理統計学ではデータの次元 p と標本の大きさ n が $p/n \to c > 0$ を満たすように無限大にいく設定 (*Kolmogorov* **極限**) が有用で、Kolmogorov 極限における、標本分散行列と標本相関行列 $\mathbf{R}$ の固有値の極限分布が、ランダム行列理論を用いて研究されています (Bai-Silverstein 2010; Jiang 2004)。数理経済学・ファイナンスでは、頻繁に、「大規模高次元データは変数の間に密に関係がある因子モデルが生成される」と仮定し、サンプルの倍化・移動に関して不変である標本相関行列 $\mathbf{R}$ を用いて解析しますが、因子モデルの中でも、変数の間に一定の相関係数 $\rho > 0$ があるモデル (**等相関正規モデル**) は単純かつ基本的です。そこで、等相関正規モデルで生成されたサンプルの $\mathbf{R}$ の固有値のバルク分布の Kolmogorov 極限をプリンストン大の Fan とミネソタ大の Jiang は問いました (Fan-Jiang 2019)。

本研究では、[1] において Fan と Jiang の問に解答しました。$\mathbf{R}$ の固有値のバルク分布のそのような極限分布は、標準的な Marčenko-Pastur 分布 (Marčenko-Pastur 1967) を $1-\rho$ 倍に収縮した分布であることを証明しました。それにより、高次元大規模データの統計学におけるある推論の質が、変数の間に互いに一定の正の相関係数 ρ が存在する場合に著しく低下するという経験則を数学的に解明しました。具体的には、統計ソフトウェアは、p 次元データの圧縮先の次元の基準として、$\mathbf{R}$ の p 個の固有値の平均より大きな固有値の個数 d を取るという、Guttman-Kaiser 基準 (Kaiser 1992) を採用していますが、データの次元 p の標本の大きさ n が比例的に無限大にいきその上で比例係数 p/n が 0 に近づく二重極限では d/p が、$\rho = 0$ の時は 1/2 に収束し、一方で、$\rho > 0$ の時は 0 に収束する、という極限定理の相転移を証明しました。$\rho = 0$ の時の d/p の 1/2 への収束は、(Yeomans-Golder 1982) のシミュレーションで示唆されていたものです。

1947 年に、 プリンストン大の Walsh は、1 次元標本変数 X_i $(1 \le i \le n)$ が互いに正の定数 ρ により相関している場合に仮説検定が劣化することを非漸近的に与えました。 その際に、Walsh は $(X_1, \ldots, X_n)$ を 1-因子モデル $X_i = \sqrt{1-\rho}Y_i + \sqrt{\rho}Z$ であることを用いました。ここで Y_i $(1 \le i \le n)$ と Z は独立な標準正規変数です。正規変数のこのような分解により、我々は、$\mathbf{R}$ の極限固有値分布を与えましたが、さまざまな代表的なランダム行列にも適用できることが確認できます [2]。

参考文献

[1] Y. Akama and A. Husnaqilati. A dichotomous behavior of Guttman-Kaiser criterion from equi-correlated normal population. *J. Indones. Math. Soc.*, 28(3):272–303, 2022.

[2] A. Husnaqilati. Limiting spectral distribution of random matrices from equi-correlated normal population. Preprint, 2022.