統計数理研究所 75 周年記念事業（2019）

オイラー標数法によるウィシャート行列の最大固有値分布の近似

The Euler characteristic method for approximating the distributions of the largest eigenvalues of Wishart matrices

数理・推論研究系 栗木 哲 (Satoshi Kuriki)

要　　旨

実対称ランダム行列の最大固有値は，行列の 2 次形式として定義される確率場の最大値であるため，オイラー標数法（チューブ法）によってその近似分布を与えることができる．本稿では，ウィシャート行列の行列サイズと自由度パラメータが無限大に発散する状況でも，オイラー標数法近似が真の極限分布の裾確率を精確に近似することを確認する．

キーワード：チューブ法, オイラー標数法, ランダム行列, Tracy-Widom 分布

1. 最大固有値とエクスカーション集合のオイラー標数

$n\times n$ 実対称行列 A が自由度 N のウィシャート分布 $W_n(N,I_n)$ に従うとする．その最大固有値は，M をランク 1 の $n\times n$ 直交射影行列の全体とするとき

$$\lambda_1(A)=\max_{\|h\|=1} h^\top Ah=\max_{U\in M}\mathrm{tr}(UA),$$

すなわち M を添字集合とする確率場 $\{\mathrm{tr}(UA)\}_{U\in M}$ の最大値である．これより，エクスカージョン集合 $M_x=\{U\in M\mid \mathrm{tr}(UA)\geq x\}$ が定義される．このランダム集合のオイラー標数 $\chi(M_x)$ の期待値が，最大固有値のオイラー標数法近似である（栗木, 2019）：

$$\Pr(\lambda_1(A)\geq x)\approx E\bigl[\chi(M_x)\bigr]\quad (x \text{ が大きいとき}).$$

$\widetilde{A}$ が自由度 N の $n\times n$ 複素ウィシャート分布に従う場合も，添字集合 $\widetilde{M}$ を適当にとることにより同じ議論ができる．ラゲール多項式を $L_n^{(\alpha)}(x)=x^{-\alpha}e^x(d/dx)^n(x^{n+\alpha}e^{-x})/n!$ とおく．

定理 1. (i) $A\sim W_n(N,I_n)$, $\alpha=N-n$.

$$E[\chi(M_x)]=\frac{\sqrt{\pi}(-1)^{n-1}(n-1)!}{2^{\frac{N+n-1}{2}}\Gamma(\frac{N}{2})\Gamma(\frac{n}{2})}\int_x^\infty \lambda^{\frac{N-n-1}{2}}e^{-\frac{\lambda}{2}}L_{n-1}^{(\alpha)}(\lambda)\,d\lambda.$$

(ii) $\widetilde{A}\sim CW_n(N,I_n)$, $\alpha=N-n$.

$$E\bigl[\chi(\widetilde{M}_x)\bigr]=\frac{n!}{\Gamma(N)}\int_x^\infty \lambda^{N-n}e^{-\lambda}\bigl\{L_{n-1}^{(\alpha)}(\lambda)L_{n-1}^{(\alpha+1)}(\lambda)-L_n^{(\alpha)}(\lambda)L_{n-2}^{(\alpha+1)}(\lambda)\bigr\}d\lambda.$$

(i) と同等な表現は，Kuriki and Takemura（2001, 2008）で与えられている．

2. エッジ極限

Marchenko-Pastur 則より，ウィシャート行列の固有値の極限分布は適当なスケーリングのもとで有限サポートをもつ．サポートの上界 μ_+ を拡大するスケーリング（エッジ極限）を行う．

$$x \mapsto s = \frac{x-\mu_+}{\sigma} \quad ただし \quad \frac{N}{n} \to \gamma, \ \ \mu_+ = (1+\sqrt{\gamma})^2 n, \ \ \sigma = (1+\sqrt{\gamma})\left(1+\frac{1}{\sqrt{\gamma}}\right)^{\frac{1}{3}} n^{\frac{1}{3}}.$$

ラゲール多項式の極限定理（Johnstone, 2001）によって次を得る．

定理 2. (i) $A \sim W_n(N, I_n)$, $N, n \to \infty$ s.t. $N/n \to \gamma$ のとき

$$E[\chi(M_x)]\Big|_{x=\mu_+ + \sigma s} \to \frac{1}{2}\int_s^\infty \mathrm{Ai}(x)\,dx.$$

ここで Ai は第一種エアリー関数．

(ii) $\widetilde{A} \sim CW_n(N, I_n)$, $N, n \to \infty$ s.t. $N/n \to \gamma$ のとき

$$E\big[\chi(\widetilde{M}_x)\big]\Big|_{x=\mu_+ + \sigma s} \to \int_s^\infty \big\{\mathrm{Ai}'(x)^2 - \mathrm{Ai}(x)^2\big\} dx.$$

図 1 より，オイラー標数法近似は，真の極限分布である Tracy-Widom 分布の上側裾確率を精確に近似することがわかる．実際それらの相対誤差は，$s \to \infty$ のとき

$$\Delta(s) \sim -2^{-5}\pi^{-\frac{1}{2}} s^{-\frac{9}{4}} e^{-\frac{2}{3}s^{3/2}} \ \ (実の場合), \qquad 2^{-10}\pi^{-1} s^{-\frac{9}{2}} e^{-\frac{4}{3}s^{3/2}} \ \ (複素の場合).$$

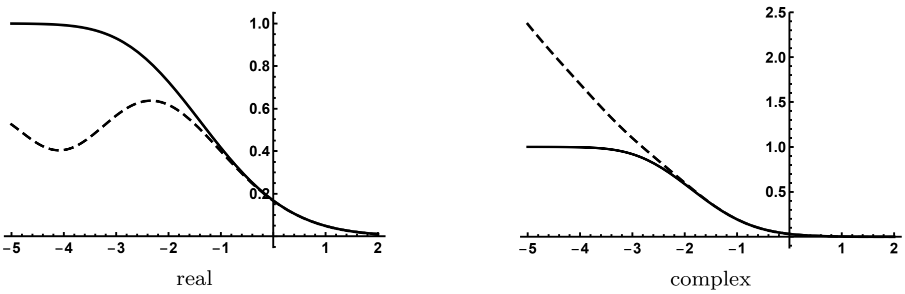

図 1．極限分布（Tracy-Widom 分布）の上側確率（実線）とそのオイラー標数法近似（破線）．
左：実ウィシャートの場合，右：複素ウィシャートの場合．

定理 2 で扱う極限では，オイラー標数法の確率場の添字集合の次元と体積が無限に発散するため，オイラー標数法が有効に働くかどうかは自明ではない．無限次元の添字集合を直接扱うことのできるオイラー標数法の開発は，これからの大きな研究課題である．

参 考 文 献

Johnstone, I. (2001). On the distribution of the largest eigenvalue in principal components analysis, *The Annals of Statistics*, **29** (2), 295–327.

栗木哲 (2019). チューブ法の理論・応用とその周辺, 「統計数理」, 投稿中.

Kuriki, S. and Takemura, A. (2001). Tail probabilities of the maxima of multilinear forms and their applications, *The Annals of Statististics*, **29** (2), 328–371.

Kuriki, S. and Takemura, A. (2008). Euler characteristic heuristic for approximating the distribution of the largest eigenvalue of an orthogonally invariant random matrix, *Journal of Statistical Planning and Inference*, **138** (11), 3357–3378.