ランダム行列におけるモーメントのゆらぎの推定

須田羽奈子　(指導教員：吉田裕亮)

1　はじめに

近年，データの蓄積量が膨大になっており，それに伴い少ない計算量でデータ解析の精度をあげることが求められている．従来のランダム行列に関する研究では，ランダム行列の固有値の経験分布が調べられており，最近では，時系列データ解析への応用も研究されている．この固有値の経験分布の性質を用いたデータ解析への応用は知られているが，固有値のゆらぎを用いた応用は希少である．ゆらぎを正確に解析することにより，統計的検定への応用が可能になる．

本研究では，固有値のモーメントのゆらぎを表すモーメントの同時分布の共分散行列を推定し，それが今得られている予想に正しいかどうかを確かめた．

2　ランダム行列

ランダム行列とは行列要素が確率変数として与えられている行列のことである．各要素が確率変数であるため，固有値も確率変数である．また，ランダム行列の固有値の漸近的性質に関する理論はランダム行列理論 (RMT) と呼ばれている．代表的なものに $Wigner$ 半円則がある．

Wigner の半円則

実対称ランダム行列の固有値経験分布は，行列のサイズ $n \to \infty$ で密度関数

$$\frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - x^2}$$

に収束する．(σ^2 は分散)

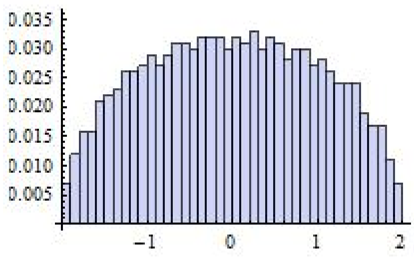

図 1: wigner の半円則

3　固有値のモーメントとそのゆらぎ

$n \times n$ のランダム行列 A の k 次モーメントは $\mathrm{tr}[A^k]$，すなわち A の固有値の k 乗の平均であり，k 次モーメントは $n \to \infty$ で，以下の理論値に収束することが知られている [1].

$$m_k = \begin{cases} 0 & \text{if } k = 2n+1, \\ \dfrac{(2n)!}{(n+1)!n!} & \text{if } k = 2n. \end{cases}$$

($\frac{(2n)!}{(n+1)!n!}$ はカタラン数である)

有限サイズでモーメントを求めると必ず理論値からのずれが発生し，このずれの挙動のことをモーメントのゆらぎと呼ぶ．各モーメントにどれくらい誤差があるかは求めることができるが，1 次から k 次モーメントまでの同時分布としての挙動はまだ知られていない部分もあり，重要である．これが本研究で予測と一致するかを調べるモーメントのゆらぎである．

以下に 1 次から k 次モーメントまでの同時分布としての挙動，つまり同時分布の共分散行列の予測値を示す．

生成するランダム行列 X

$$X_{i,i} \sim N(0,\sigma^2) \quad and \quad X_{i,j} \sim N(0,\eta^2)$$

共分散行列の予測値 $\alpha_{p,q}$

$$\alpha_{p,q} = \begin{cases} \dfrac{pq}{p+q}\dbinom{p}{\frac{p}{2}}\dbinom{q}{\frac{q}{2}}\eta^{p+q} \\ \qquad\qquad\qquad \text{if } p \text{ and } q \text{ are even,} \\ \dfrac{2pq}{p+q}\dbinom{p-1}{\frac{p-1}{2}}\dbinom{q-1}{\frac{q-1}{2}}\sigma^2\eta^{p+q-2} \\ \qquad\qquad\qquad \text{if } p \text{ and } q \text{ are odd,} \\ 0 \quad \text{otherwise.} \end{cases}$$

4　実験概要

本研究においては, 先の共分散行列の予測値が正しいかどうかの数値実験を

(対角成分σ^2，非対角成分η^2) $= (6,4),(2,1)$

の 2 つの場合に分けて行った．これを実験 1，実験 2 とする．(2,1) の場合のゆらぎは既知であるが，追認の意味で数値実験を行うこととした．

まず，行列のサイズ $n = 50r$ $(1 \leq r \leq 20)$ についてランダム行列 X を 500 回生成した．それぞれの X について，1 次から 8 次モーメントまでを求めて 8 変量データを得る．これにより，8 変量同時分布の 500 サンプルが得られる．行列のサイズとモーメントの次数，およびランダム行列の生成個は本研究で用いた値である．

本研究では，実験で得られた 8 変量の 500 サンプルの共分散行列を K とした．$K_{p,q}$ はどの程度ばらつきながらゆらぎが収束するかを表している．この K の各 (p,q) 成分は n が大きくなるにつれて指数的に 0 に収束する．

漸近的に K の各成分 (p,q) について

$$K_{p,q} \approx \frac{\alpha_{p,q}}{n^2}$$

より，対数をとり

$$\log K_{p,q} \approx -2\log n + \log\alpha_{p,q}$$

がいえる．$K_{p,q}, n$ は実験で得られるため，線形回帰により切片を各 (p,q) 成分について求めると，予測値 $\log \alpha_{p,q}$ に近い値が得られるはずである．

$(x,y) = (\log n, \log K_{p,q})$ とすると，各成分 (p,q) における回帰直線は $y = -2x + b$ となり，e^b は $\alpha_{p,q}$ の予測値である．得られた (x,y) から線形回帰を行うと傾きは-2 に近づくが，多少のずれが生じる．今回は，より精度を上げるために直線の傾きを-2 に固定して回帰を行った．図 1 は (1,1) 成分における (x,y) 点を示したものとその回帰直線である．多少のデータのずれで切片が大きく変わることが予想できる．

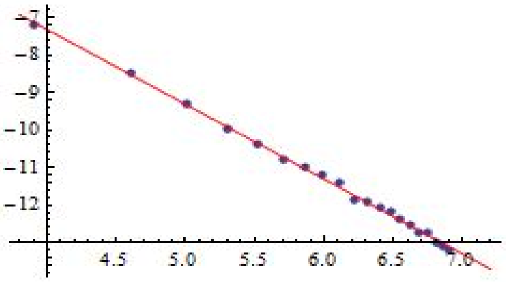

図 2: (1,1) 成分における回帰直線

5 Bootstrap 法

さらに本研究では，より収束の精度を上げるために，Bootstrap 法を用いた．

Bootstrap 法はリサンプリング法のひとつであり，B.Efron(1979) によって定式化された [2]． この方法により，得られた実験データから再び実験データを再標本化することで外れ値の影響を小さくできると考えた．本研究ではブートストラップ回帰の考え方を用いて得られたデータをリサンプリングする手法とサンプルから得られた予測値の残差をリサンプリングする手法を用いて Bootstrap 法を用いずに予測値を求めた場合と比較した．

5.1 データの再標本による Bootstrap 回帰

1) 得られたサンプル s 個の中から s 個のサンプルを重複を許してランダムに選び，k 変量 s サンプルを再標本する．
2) 再標本した k 変量 s サンプルの共分散行列を求める．
3) 1) と 2) を B 回繰り返す．
4) 3) の平均を新たな予測値とする．

5.2 残差による Bootstrap 回帰

1) r 個の (x,y) についてそれぞれ b を求める．
2) r 個の中から r 個重複を許してランダムに b を選ぶ．
3) r 個の平均を新たな予測値とする．

6 実験結果

実験 1 と実験 2 について Bootstrap 法 1,2 を用いたときと用いなかったときの予測値を (3,3) 成分と (4,2) 成分について比較してみる．

Bootstrap 法を用いることで精度が大きく上がり予測値が近づくことが期待されたが，成分 (p,q) によりばらつきがあり，一概に精度があがるとは言えない結果となった．また，Bootstrap 法 1,2 についても同様で，この 2 つについてはどの成分 (p,q) についてもほとんど差が見られない結果となった．

(3,3)	予測値	回帰のみ	5.1	5.2
実験 1	1152	1247.23	1243.02	1242.63
実験 2	24	24.5491	24.4597	24.4635

表 1: (3,3) 成分の数値結果

(4,2)	予測値	回帰のみ	5.1	5.2
実験 1	1024	1037.25	1033.02	1033.80
実験 2	16	16.5678	16.5016	16.5047

表 2: (4,2) 成分の数値結果

また，各 (p,q) 成分について予測値との分散を求めると p,q が大きくなるにつれて分散も大きくなることが確認された．以下に分散の分布を色分けした状態で示す．

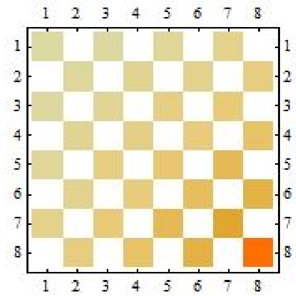

図 3: 実験 1 の分散

7 考察と今後の展望

実験により，モーメントのゆらぎの予測値は正しいと考えられる．この証明も推測を進めることができそうである．しかし，より予測値に近づけようと試みた Bootstrap 法は多少の改善しかみられず，また p,q が小さいところでは Bootstrap 法を用いない方が予想に近いものもあり，今回精度を上げる方法として有効ではなかったと思われる．これは，行列の要素自体はランダムに発生させているが，モーメントをとるとほとんど差がないからだと思われる．しかし，Bootstrap 法を用いなくても予測値に近い値が得られた．

今後は，このモーメントのゆらぎが予測値になることの証明を進めるとともに，実際にこのゆらぎの行列を用いてデータ解析の精度が上がることを確かめたい．

参考文献

[1] ランダム行列の数理と科学・渡辺澄夫, 永尾太郎, 樺島祥介, 田中利幸, 中島伸一

[2] 計算統計　確率計算の新しい手法・汪金芳, 田栗正章, 手塚集, 樺島祥介, 上田修功