Scaling in the eigenvalue fluctuations of correlation matrices

Udaysinh T. Bhosale,* S. Harshini Tekur,† and M. S. Santhanam‡
Indian Institute of Science Education and Research, Dr. Homi Bhabha Road, Pune 411 008, India

(Received 30 July 2018; published 26 November 2018)

The spectra of empirical correlation matrices, constructed from multivariate data, are widely used in many areas of sciences, engineering, and social sciences as a tool to understand the information contained in typically large data sets. In the past two decades, random-matrix-theory-based tools such as the nearest-neighbor eigenvalue spacing and eigenvector distributions have been employed to extract the significant modes of variability present in such empirical correlations. In this work we present an alternative analysis in terms of the recently introduced spacing ratios, which does not require the cumbersome unfolding process. It is shown that the higher-order spacing ratio distributions for the Wishart ensemble of random matrices, characterized by the Dyson index β, are related to the first-order spacing ratio distribution with a modified value of codimension β'. This scaling is demonstrated for the Wishart ensemble and also for the spectra of empirical correlation matrices drawn from the observed stock market and atmospheric pressure data. Using a combination of analytical and numerics, such scalings in spacing distributions are also discussed.

DOI: 10.1103/PhysRevE.98.052133

I. INTRODUCTION

Large multivariate data sets are commonly encountered in many areas of sciences [1,2], engineering [3], and social sciences [4]. Some common examples include the data generated from the financial markets [5], atmospheric and climate parameters [6], and communication networks [7]. Analysis of the spectra of empirical correlation matrices constructed from large data sets provides detailed and graded information about the systems being studied. In the past two decades, tools and results from random-matrix theory (RMT) have been widely applied to make sense of the information provided by detailed spectra, namely, the eigenvalues and the eigenvectors, of the empirical correlation matrices [8,9]. Originally, RMT was conceived as a model for the energy spectra of complex many-body quantum systems such as nuclei and atoms [10–12]. These applications of RMT expanded its scope well beyond its original domain of quantum spectra.

The eigenvalues E_i, $i = 1, 2, \ldots, N$, of the empirical correlation matrix of order N are positive definite, i.e., $E_i \geqslant 0$. Typically, the corresponding eigenmodes fall into two broad groups: (i) eigenmodes of the top and bottom small eigenvalues (in magnitude) that carry most of the information embedded in the original data set and (ii) the bulk of the rest which represents random correlations. It is the latter group that displays broad agreement with random-matrix-based results. For instance, the bulk of the correlation matrix spectra obtained from the time series of the largest stocks in the United States, including the ones that make up the S&P index, was shown to be in agreement with the random-matrix averages [13–16], and some studies have argued that they contain finer correlated structures [17,18]. The density of the bulk of the eigenvalues follows the Marčenko-Pastur distribution [19] and can be used to identify the top eigenvalues that carry significant information. The analysis of eigenvectors, in terms of its agreement with the Porter-Thomas distribution [20], indicates the stocks that are strongly correlated [14]. A similar approach for the analysis of atmospheric data can distinguish physically relevant modes of atmospheric variability from the those that are noisy [6]. By now, many applications [21–25] including in biology [26], image processing [27], and network traffic [28] abound.

An often demonstrated property of the bulk of eigenvalues is that the distribution of the spacings $s_i = E_{i+1} - E_i$, $i = 1, 2, \ldots$, between consecutive eigenvalues (after unfolding) follows the celebrated Wigner distribution $P(s) = (\pi/2)s \exp(-\pi s^2/4)$ [10]. This signifies level repulsion, the tendency of the eigenvalues to repel one another. This continues to be a popular test for RMT-like behavior, especially for the claim that spectral fluctuations of empirical correlation matrices display universal characteristics irrespective of the data set or system considered for analysis. A major impediment to computing the spacing distribution is the requirement to unfold the eigenvalues, a somewhat unreliable numerical procedure that approximately separates the system-dependent eigenvalue properties from the generic ones. This problem can be circumvented by considering the ratio of consecutive spacings $r_i = s_{i+1}/s_i$, $i = 1, 2, \ldots$, which is independent of local eigenvalue density and hence does not depend on the system [29–31]. In this work, scaling properties relating to the distribution of higher-order spacings and spacing ratios to the nearest-neighbor spacing properties are demonstrated.

The elements of the empirical correlation matrix represent the pairwise Pearson correlation among the N variables, each one being a time series of length T. From the point of view of random-matrix theory, correlation matrices fall within the class of the Laguerre-Wishart ensemble [32] of

*udaybhosale0786@gmail.com
†harshini.t@gmail.com
‡santh@iiserpune.ac.in

random-matrix theory represented by $W = D_R D_R^S$, where D_R represents the standardized data matrix of order N by T with real, complex, or quaternion elements depending on the Dyson index $\beta = 1, 2, 4$ of the ensemble and X^S represents self-dual operation on matrix X. For the Laguerre-Wishart ensemble indexed by β, the random-matrix average for the spacing ratios is not yet known, though in the limit of matrix size $N \to \infty$ it is well approximated by that for the Gaussian ensembles given by

$$P(r) = C_\beta \frac{(r + r^2)^\beta}{(1 + r + r^2)^{1+(3/2)\beta}}, \tag{1}$$

where C_β is a constant whose form is given in Ref. [31]. However, results for spacing ratios and spacing distributions beyond the nearest neighbors are not yet known. The higher-order spacing statistics provide a finer test for the universality of spectral fluctuations. In addition, the deviations from these will quantify the timescales up to which random-matrix-type universality can be expected to be valid in empirical cases. Indeed, long-range correlations such as Δ_3 statistics have indicated limitations of RMT assumptions at longer timescales [6,15,28].

The structure of the paper is as follows. In Sec. II a scaling relation is given for the higher-order spacing ratios for the Wishart ensemble of random matrices. In Sec. III this scaling relation is tested on various systems, including the spectra of the Wishart random-matrix ensembles and the spectra of empirical correlation matrices drawn from the stock market and atmospheric data set. This relation has also been shown to hold analytically for the first few orders for the spacing distribution. Section IV summarizes the work.

II. DISTRIBUTION OF HIGHER-ORDER SPACING RATIOS

Consider a sample correlation matrix C of order $N \gg 1$, constructed from time series of N variables and each of length T, whose ordered eigenvalues are $E_1 \leqslant E_2 \leqslant \cdots \leqslant E_N$. The density of eigenvalues are given, in the limit $T \geqslant N \gg 1$, by the Marčenko-Pastur distribution, which predicts an upper and a lower bound for the eigenvalues [19]. In nearly all of the earlier studies involving spectra of empirical correlation matrices, eigenvalue density and spacing distributions had been widely studied [33]. In contrast, in this work we study the kth-order spacing ratios defined by

$$r_i^{(k)} = \frac{s_{i+k}^{(k)}}{s_i^{(k)}} = \frac{E_{i+2k} - E_{i+k}}{E_{i+k} - E_i}, \tag{2}$$

where kth-order spacings can also be defined as $s^{(k)} = s_{i+k} - s_i$. If $k = 1$, this reduces to the standard nearest-neighbor spacing ratio. The distribution of kth-order spacing ratio is denoted by $P^k(r, \beta)$, where $\beta = 1, 2, 4$ is the codimension that represents the Wishart ensemble. Note that the higher-order spacing ratios are not uniquely defined [31]. In Eq. (2) we take them such that no common spacings are shared between the numerator and denominator. Using Eq. (2), the main result of this paper can be stated as follows: For the random matrix of order $N \gg 1$ and $T \geqslant N \gg 1$, from the Wishart ensemble with $\beta = 1, 2$ and 4, the kth-order spacing ratio distribution is related to the nearest-neighbor ($k = 1$) spacing ratio distribution statistics by

$$P^k(r, \beta) = P(r, \beta'), \quad \beta = 1, 2, 4, \tag{3}$$

$$\beta' = \frac{k(k+1)}{2}\beta + (k - 1), \quad k \geqslant 1. \tag{4}$$

Here $4 \leqslant \beta' < \infty$, though for these values of β' explicit matrix forms for the Wishart ensemble are not known. Though the eigenvalue density given by the Marčenko-Pastur distribution depends on the ratio T/N [19], the statistics of fluctuations represented by Eq. (4) can be expected to be independent of T/N. A similar scaling relation had been postulated for the spacing distribution of Gaussian ensembles [34] and its generalizations [35], and recently numerical evidence was provided for the ratios [36]. Both $P^k(r, \beta)$ and $P(r, \beta')$ have identical functional forms and the modified parameter β' depends on the order k of the spacing ratio and codimension β. Further, we present strong numerical evidence from Wishart matrices as well as from empirical correlation matrix spectra computed from the observed stock market and atmospheric data.

A rigorous proof of Eqs. (3) and (4) is mathematically challenging, but we give an intuitive argument why β' should be greater than β. The eigenvalues of the Wishart ensemble correspond to the charged particles of a two-dimensional Coulomb gas [37]. In this physical picture, the degree of repulsion between eigenvalue pairs beyond the nearest neighbors is greater than that for consecutive eigenvalue pairs. Hence, it appears physically reasonable to expect that β' for the case of $k > 1$ is greater than β for $k = 1$. For the special case of $k = 2$ in the context of the circular orthogonal ensemble ($\beta = 1$) of RMT, a limited analytical proof was derived in Ref. [38]. Thus, Eqs. (3) and (4) represent a generalization of this result for the spacing ratios of the Wishart ensemble. In the next section we apply the scaling relation in Eqs. (3) and (4) to the spectra of various systems.

III. RESULTS

A. Random-matrix spectra

Now we consider the spectra obtained from an ensemble of Wishart matrices with $\beta = 1$ and test the validity of Eqs. (3) and (4) by computing the higher-order spacing ratios. In Fig. 1, the kth-order spacing ratio distributions are shown as histograms for two cases, namely, $N = T$ and $N \neq T$. The validity of the scaling in Eq. (3) can be clearly inferred from the excellent agreement of the histogram with a solid curve representing $P(r, \beta')$, where β' given by Eq. (4). An additional layer of quantitative verification can be performed as follows. Let the cumulative distributions corresponding to the computed histograms $P^k(r_i, \beta)$ and $P(r, \beta')$ be represented, respectively, by $I^k(r_i, \beta)$ and $I(r_i, \beta')$. In Eq. (4), β' is treated as a tunable parameter and the difference between cumulative distributions

$$D(\beta') = \sum_i |I^k(r_i, \beta) - I(r_i, \beta')| \tag{5}$$

is computed. The minimum of $D(\beta')$ is the best value of β' that fits the histogram data. As seen in the insets in Fig. 1, the

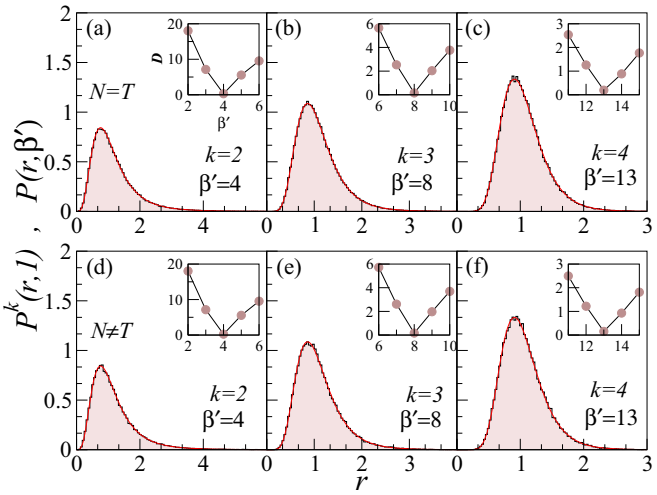

FIG. 1. Histograms of the kth-order spacing ratio distribution for the spectra of the random Wishart matrix for $\beta = 1$ with (a)–(c) $N = T = 40\,000$ and (d)–(f) $N = 20\,000$ and $T = 30\,000$. The computed histograms display good agreement with $P(r, \beta')$, shown as a solid line. Here β' is given by Eq. (4). The inset shows that the minimum of $D(\beta')$ corresponds to the value of β' predicted by Eq. (4).

minimum of $D(\beta')$ precisely coincides with the value of β' postulated in Eq. (4).

The results displayed in Fig. 2 show that the higher-order spacing ratio distributions computed from the spectra from Wishart matrices with $\beta = 2$ and 4 are consistent with the scaling relation postulated in Eqs. (3) and (4). The elements of Wishart matrices with $\beta = 2$ and 4 are, respectively, complex numbers and quaternions, and empirical correlations with such elements are rarely encountered in practice. The symbols in this figure represent the histograms and solid curves

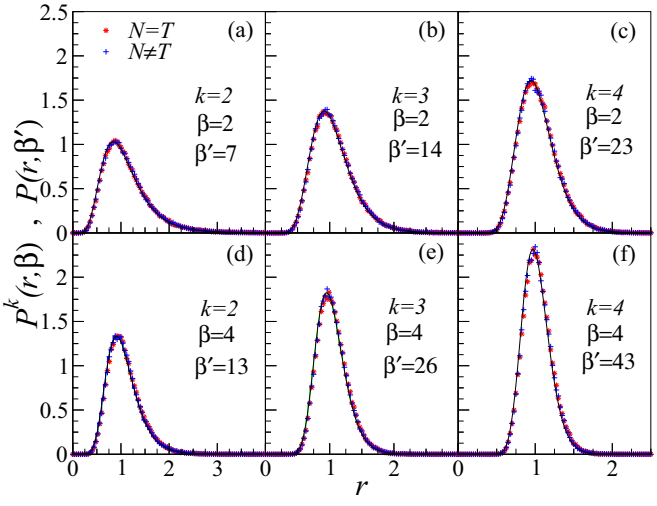

FIG. 2. Histograms of the kth-order spacing ratio distribution for the spectra of the random Wishart matrix with (a)–(c) $\beta = 2$ and (d)–(f) $\beta = 4$. For the $N = T$ case, $N = T = 20\,000$; for the $N \neq T$ case, $N = 10\,000$ and $T = 20\,000$. The computed histograms display good agreement with $P(r, \beta')$, shown as a solid line [β' is given by Eq. (4).

TABLE I. Mean values of spacing ratio for Wishart matrix $\langle r \rangle_w$, atmospheric correlations data $\langle r \rangle_{\text{atm}}$, and stock market correlations $\langle r \rangle_{\text{fin}}$. The expected theoretical value is $\langle r \rangle_{th}$.

β	k	β'	$\langle r \rangle_{th}$	$\langle r \rangle_w$ $(n = m)$	$\langle r \rangle_w$ $(n \neq m)$	$\langle r \rangle_{\text{atm}}$	$\langle r \rangle_{\text{fin}}$
	2	4	1.174	1.177	1.176	1.214	1.330
1	3	8	1.085	1.086	1.085	1.123	1.237
	4	13	1.052	1.053	1.052	1.089	1.211

represent $P(r, \beta')$. The results are shown for both $N = T$ and $N \neq T$ and, as anticipated, the agreement with Eqs. (3) and (4) is good irrespective of the relative values of N and T. Another form of evidence in Table I for the mean ratio $\langle r \rangle$ shows good agreement between the theoretically expected value based on Eqs. (3) and (4) and that obtained from computed Wishart spectra.

B. Stock market and atmospheric data

Next we demonstrate the validity of the scaling relation (3) and (4) for the spectra of empirical correlation matrices drawn from two different domains, namely, the stock market and an atmospheric data set. To begin with, the data of the time series of stocks that are part of the S&P500 index for the years 1996–2009 are considered [17]. This data set continues to be extensively used to understand the ramifications of how an RMT-based approach might work in the context of empirical correlation matrices. The data consist of daily (log) returns for $T = 3400$ days for $N = 396$ assets. The elements of the correlation matrix denote the Pearson correlation between pairs of stocks averaged over time. Note that $T \geqslant N$, implying that the correlations can be assumed to have converged. The statistical properties of its spectra have been reported in Refs. [13–16].

In Figs. 3(a)–3(c) we display the spacing ratio distribution for various orders. Figure 3(a) shows the nearest-neighbor spacing ratio distribution and it agrees with the analytical result in Eq. (1) obtained for the case of the Gaussian orthogonal ensemble [30]. The higher-order spacing ratio distributions are displayed in Figs. 3(b) and 3(c) and we notice good agreement with the postulated theoretical distribution $P(r, \beta')$, with β' as given by Eq. (4). Further, we consider the time series of monthly mean sea level pressure over the North Atlantic Ocean. The monthly data are drawn from National Centers for Environmental Prediction reanalysis archives [39] and is available over equally spaced latitude and longitude grids for the North Atlantic region bounded by (0–90°N, 120°W–30°E) for the years 1948–2017. Thus, in this case, $N = 434$ grid points and $T = 840$ months, satisfying the condition $T/N > 1$. An analysis of the climate phenomenon of the North Atlantic oscillation was performed by constructing an empirical correlation matrix from this data and using RMT statistics such as the spacing and eigenvector distributions [6]. In Fig. 3(d) the spacing ratio distributions for the nearest-neighbor spacings obtained from the spectra of this correlation matrix are shown. The computed histogram is seen to be well described by the theoretical distribution in Eq. (1) obtained for Gaussian ensembles. The higher-order spacing

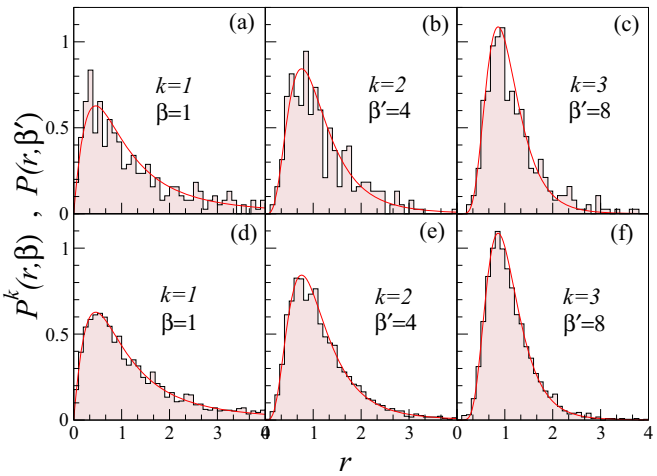

FIG. 3. Histograms of the kth-order spacing ratio distribution for the spectra of the correlation matrix from (a)–(c) S&P500 stock market data and (d)–(f) mean sea level pressure data. The computed histograms display good agreement with $P(r, \beta')$, shown as a solid line. Here β' is given by Eq. (4).

ratio distributions shown in Figs. 3(e) and 3(f) display good agreement with $P(r, \beta')$, as anticipated by Eq. (4).

Both these empirical correlation matrix spectra are computed from a relatively short sequence of time series compared to the length of time series used in computing Wishart spectra for Fig. 1. Hence, the noise level for the correlations is higher than for the Wishart case, and it is evident in the higher-order spectral statistics shown in Fig. 3. This also manifests as poor agreement with the $\langle r \rangle$ values shown in Table I. Finally, we point out that violating the conditions $N \gg 1$ and $T \geqslant N \gg 1$ leads to deviations from the scaling relation (3) and (4) due to finite-size effects as shown in Fig. 4. Figures 4(a)–4(c) show that the minima of $D(\beta')$ converge to the predicted value $\beta' = 76$ on considering a (relatively) small range of eigenvalues in the bulk, where the density of states may be

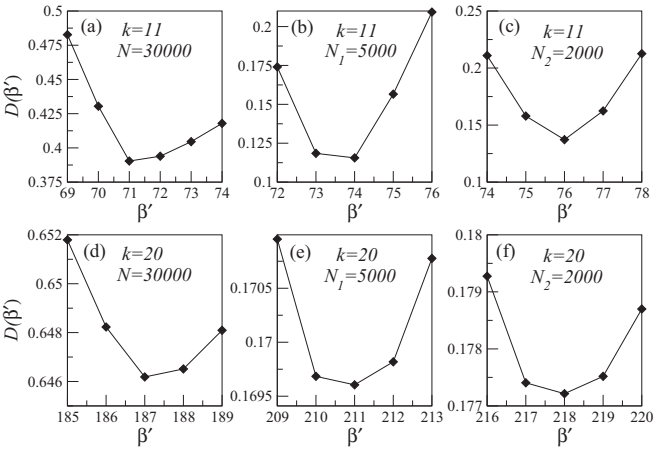

FIG. 4. Plot of $D(\beta')$ for (a)–(c) $k = 11$ and (d)–(f) $k = 20$ and (a) and (d) $N = 30\,000$, (b) and (e) $N_1 = 5000$, and (c) and (f) $N_2 = 2000$ eigenvalues, for matrices of dimensions $N = T = 30\,000$. Here N_1 and N_2 represent a subset of eigenvalues taken from the bulk of the spectrum.

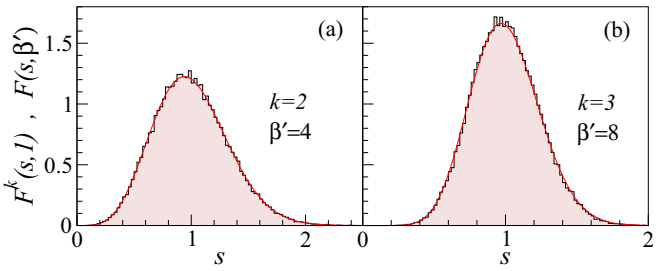

FIG. 5. Histograms of the kth-order spacing distribution for the spectra of the Wishart matrix. The histograms display good agreement with $F(s, \beta')$, shown as a solid line. Here β' is given by Eq. (4).

assumed to be constant. This finite-size or convergence effect has also been discussed for Gaussian ensembles in Ref. [36]. In contrast, Figs. 4(d)–4(f) show that the minima of $D(\beta')$ do not converge to the predicted value $\beta' = 229$ even on considering a (relatively) small range of eigenvalues in the bulk. To obtain a constant local density of states over a larger energy scale thus requires a random matrix of larger dimensions.

C. Spacing distributions

We have also studied the validity of the scaling relation (4) for the more popular eigenvalue spacing distribution [10,37]. In order to examine this, the kth-order nearest-neighbor spacing is defined as $s^{(k)} = (s_{i+k} - s_i)/\langle s \rangle$, $i = 1, 2, \ldots$. Based on the analytical result (see the Appendix), it is postulated that the second-order (third-order) spacing distribution is $F(s, \beta') = A_{\beta'} s^{\beta'} e^{-B_{\beta'} s^2}$, a form that is reminiscent of the Wigner surmise, where $\beta' = 3\beta + 1$ ($\beta' = 6\beta + 2$). This β' agrees with Eq. (4) for $k = 2$ ($k = 3$). The constants $A_{\beta'}$ and $B_{\beta'}$ depend on β' and are given in Ref. [11].

In Fig. 5 we verify this claim for the computed Wishart matrix spectra. The computed histograms display excellent agreement with the spacing distributions $F(s, \beta' = 4)$ [Fig. 5(a)] and $F(s, \beta' = 8)$ [Fig. 5(b)]. Figures 6(a) and 6(b) display next-nearest-neighbor ($k = 2$) spacing distribution for the data drawn from mean sea level pressure and S&P500 stocks. In both these cases, good agreement with the anticipated $F(s, \beta')$ is evident. For $k > 3$, it does not appear

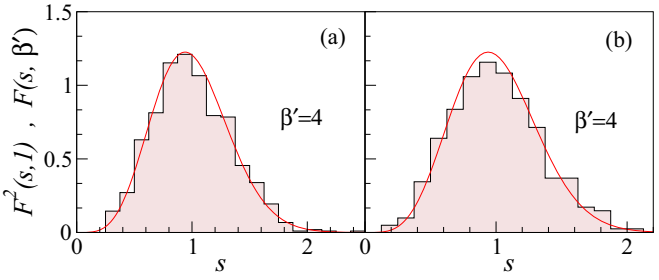

FIG. 6. Histograms of the kth-order spacing distribution for the spectra of the correlation matrix from (a) mean sea level pressure data and (b) S&P500 stock market data. The computed histograms display good agreement with $F(s, \beta')$, shown as a solid line. Here β' is given by Eq. (4).

straightforward to extend these results due to pronounced finite-size effects and the limitations of pushing the spacing distributions postulated based on $s \to 0$ results well beyond their regime of validity.

IV. CONCLUSION

The empirical correlation matrices are widely used in many areas of sciences and engineering as tools to extract information from large data sets. Typically, this is done by analyzing their spectra, the bulk of which are known to follow the random-matrix-theory predictions, especially for the eigenvalue density and the popular nearest-neighbor spacing distribution. Computation of the spacing distribution involves unfolding the spectra through an ambiguous fitting procedure. In recent years, the spacing ratio has become a popular alternative to spacing distributions since the former does not depend on the eigenvalue density and hence it does not require unfolding. In this work, for the Wishart matrices of order $N \gg 1$, we focused on the higher-order spacing statistics and showed that the kth-order spacing ratio distribution $P^k(r, \beta)$ can be obtained in terms of the corresponding nearest-neighbor ($k = 1$) distribution $P(r, \beta')$, where $\beta' > \beta$ and β' depends on k and β. We have used the correspondence of the Wishart eigenvalues with the charged particles of a two-dimensional Coulomb gas to explain $\beta' > \beta$. Further, using analytical and simulation results, a similar scaling with a limited scope was obtained for the spacing distributions of Wishart matrix spectra and empirical correlation matrices.

We demonstrated the validity of scaling in eigenvalue fluctuations using the spectra drawn from an ensemble of Wishart matrices. As an application with observed data sets, the scaling in fluctuations was also shown for the spectra of empirical correlation matrices obtained from S&P500 stock market data and mean sea level pressure data over the North Atlantic Ocean; both of these had earlier been analyzed from the RMT point of view. It would be interesting to obtain these results exactly for the Wishart ensemble. This presents the opportunity for tests for the claims of universality of eigenvalue fluctuations and further it can potentially determine the timescales over which RMT-like fluctuations hold for empirical correlation matrices.

ACKNOWLEDGMENTS

The authors thank Dr. Giacomo Livan for providing us S&P 500 correlation matrix data for S&P 500 stocks [17], whose eigenvalues are analyzed in this work. U.T.B. acknowledges the funding received from the Department of Science and Technology, India under the scheme Science and Engineering Research Board National Post Doctoral Fellowship Grant No. PDF/2015/00050.

APPENDIX: SPACING DISTRIBUTIONS FOR SECOND AND THIRD ORDER

In this Appendix the analytical result leading to the second- and third-order spacing distributions is derived. Consider the random Wishart matrix W of order N and specialized to the case of the next-nearest-neighbor ($k = 2$) spacing distribution that can be obtained from the Wishart matrix of order $T = 3$ with three eigenvalues $\{E_1, E_2, E_3\}$. Then the joint probability density function (JPDF) of the eigenvalues $E_l \geqslant 0$, $l = 1, 2, 3$, for the Wishart ensemble is given as

$$f(\{E_l\}) = \frac{1}{W_{a\beta T}} \prod_{i=1}^{T} E_i^{\beta a/2} e^{-\beta E_i/2} \prod_{1 \leqslant j < p \leqslant T} |E_p - E_j|^{\beta},$$

where $a = N - T + 1 - 2/\beta$ and $W_{a\beta T}$ is a constant [37]. Further, with $T = 3$, N and β are chosen such that $a = 0$. Then the JPDF can be obtained as

$$f(E_1, E_2, E_3) = \frac{3!}{W_{0\beta 3}} \prod_{i=1}^{3} e^{-\beta E_i/2} \prod_{1 \leqslant j < p \leqslant 3} |E_p - E_j|^{\beta}. \tag{A1}$$

Using the transformation $x = E_2 - E_1$ and $y = E_3 - E_2$, we obtain $E_2 = E_1 + x$, $E_3 = E_1 + x + y$, and

$$f(E_1, x, y) = \frac{3!}{W_{0\beta 3}} x^{\beta} y^{\beta} (x + y)^{\beta} e^{-c\beta(3E_1 + 2x + y)/2}. \tag{A2}$$

Letting $K_1 = 3!/W_{0\beta 3}$ and integrating over E_1, we obtain

$$f(x, y) = \frac{2K_1}{3\beta c} x^{\beta} y^{\beta} (x + y)^{\beta} e^{-c\beta(2x+y)/2}. \tag{A3}$$

It can be seen that $0 \leqslant x + y = E_3 - E_1 = s$ and $y = s - x$. After some algebra, the next-nearest-neighbor ($k = 2$) spacing distribution $\widetilde{F}^2(s)$ can be obtained as

$$\widetilde{F}^2(s) = \frac{s^{3\beta+1} e^{-c\beta s/2}}{2^{-1} K_1^{-1} 3\beta c} \sum_{q=0}^{\beta} \sum_{n=0}^{\infty} \binom{\beta}{q} \frac{s^n (-1)^{\beta - q} (-c\beta/2)^n}{n!(2\beta - q + n + 1)}. \tag{A4}$$

In the limit of $s \to 0$, the leading behavior is proportional to $s^{\beta'}$, where $\beta' = 3\beta + 1$. The result derived above can be extended easily for the case of $k = 3$ as well, resulting in $\beta' = 6\beta + 2$. Thus, based on these analytical results and in the spirit of the scaling relation (3) and (4), it is postulated that the second-order (third-order) spacing distribution is $F(s, \beta') = A_{\beta'} s^{\beta'} e^{-B_{\beta'} s^2}$, where $\beta' = 3\beta + 1$ ($\beta' = 6\beta + 2$). The constants $A_{\beta'}$ and $B_{\beta'}$ (given in Ref. [11]) depend on β' [Eq. (4)].

[1] J. Wishart, Biometrika **20A**, 32 (1928).
[2] J. Kwapien and S. Drozdz, Phys. Rep. **515**, 115 (2012).
[3] Z. Bai, Y. Chen, and Y.-C. Liang, *Random Matrix Theory and Its Applications: Multivariate Statistics and Wireless Communications* (World Scientific, Singapore, 2009).
[4] D. J. Bartholomew, F. Steele, J. Galbraith, and I. Moustaki *Analysis of Multivariate Social Science Data* (Chapman and Hall/CRC, London, 2008).
[5] C. Biely and S. Thurner, Quant. Finance **8**, 705 (2008).
[6] M. S. Santhanam and P. K. Patra, Phys. Rev. E **64**, 016102 (2001).

[7] S. Barlowe, T. Zhang, Y. Liu, J. Yang, and D. Jacobs, in *Proceedings of the 2008 IEEE Symposium on Visual Analytics Science and Technology, Columbus, OH, USA* (IEEE, Piscataway, 2008), pp. 147–154.
[8] J. Bun, J.-P. Bouchaud, and M. Potters, Phys. Rep. **666**, 1 (2017); M. A. Nowak and W. Tarnowsk, J. Stat. Mech. (2017) 063405.
[9] S. S. Wilks, *Mathematical Statistics* (Princeton University Press, Princeton, 1947).
[10] M. L. Mehta, *Random Matrices* (Academic, New York, 2004).
[11] T. Guhr, A. Müller-Groeling, and H. A. Weidenmüller, Phys. Rep. **299**, 189 (1998).
[12] G. Livan, M. Novaes, and P. Vivo, *Introduction to Random Matrices: Theory and Practice* (Springer, Berlin, 2018).
[13] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, Phys. Rev. Lett. **83**, 1471 (1999).
[14] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, Phys. Rev. Lett. **83**, 1467 (1999).
[15] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, Phys. Rev. E **65**, 066126 (2002).
[16] R. K. Pan and S. Sinha, Phys. Rev. E **76**, 046116 (2007).
[17] G. Livan, S. Alfarano, and E. Scalas, Phys. Rev. E **84**, 016113 (2011).
[18] J. Kwapień, S. Drożdż, and P. Oświęcimka, Physica A **359**, 589 (2006)
[19] V. A. Marčenko and L. A. Pastur, Mat. USSR Sb. **1**, 457 (1967).
[20] C. E. Porter and R. G. Thomas, Phys. Rev. **104**, 483 (1956).
[21] *The Oxford Handbook of Random Matrix Theory*, edited by G. Akemann, J. Baik, and P. Di Francesco (Oxford University Press, New York, 2011).
[22] J. Novembre and M. Stephens, Nat. Genet. **40**, 646 (2008).
[23] N. Patterson, A. L. Preis, and D. Reich, PLoS Genet. **2**, e190 (2006).
[24] G. Akemann, Nucl. Phys. B **507**, 475 (1997).
[25] K. Johansson, Commun. Math. Phys. **209**, 437 (2000).
[26] P. Śeba, Phys. Rev. Lett. **91**, 198104 (2003).
[27] K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Elsevier, New York, 1990).
[28] M. Barthélemy, B. Gondran, and E. Guichard, Phys. Rev. E **66**, 056110 (2002).
[29] V. Oganesyan and D. A. Huse, Phys. Rev. B **75**, 155111 (2007).
[30] Y. Y. Atas, E. Bogomolny, O. Giraud, and G. Roux, Phys. Rev. Lett. **110**, 084101 (2013).
[31] Y. Y. Atas, E. Bogomolny, O. Giraud, P. Vivo, and E. Vivo, J. Phys. A: Math. Theor. **46**, 355204 (2013).
[32] I. Dumitriu and A. Edelman, J. Math. Phys. **43**, 5830 (2002).
[33] A. Patil and M. S. Santhanam, Phys. Rev. E **92**, 032130 (2015).
[34] P. B. Kahn and C. E. Porter, Nucl. Phys. **48**, 385 (1963); A. Y. Abul-Magd and M. H. Simbel, Phys. Rev. E **60**, 5371 (1999).
[35] P. J. Forrester, Commun. Math. Phys. **285**, 653 (2009).
[36] S. H. Tekur, U. T. Bhosale, and M. S. Santhanam, Phys. Rev. B **98**, 104305 (2018).
[37] P. J. Forrester, *Log-Gases and Random Matrices* (Princeton University Press, Princeton, 2010).
[38] M. L. Mehta and F. J. Dyson, J. Math. Phys. **4**, 713 (1963).
[39] NCEP reanalysis data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at www.esrl.noaa.gov/psd/