A study on the empirical distribution of the scaled Hankel matrix eigenvalues

Hossein Hassani\textsuperscript{a,b,}\footnote{Corresponding author. Address: Tel.: +44 1202968708; fax: +44 1202968124. E-mail address: hhassani@bournemouth.ac.uk (H. Hassani). Peer review under responsibility of Cairo University.}, Nader Alharbi\textsuperscript{a}, Mansi Ghodsi\textsuperscript{a}

\textsuperscript{a} The Statistical Research Centre, Bournemouth University, Bournemouth BH8 8EB, UK
\textsuperscript{b} Institute for International Energy Studies (IIES), 65, Sayeh St., Vali-e-Asr Ave., Tehran 1967743 711, Iran

ABSTRACT

The empirical distribution of the eigenvalues of the matrix $XX^T$ divided by its trace is evaluated, where $X$ is a random Hankel matrix. The distribution of eigenvalues for symmetric and nonsymmetric distributions is assessed with various criteria. This yields several important properties with broad application, particularly for noise reduction and filtering in signal processing and time series analysis.

© 2014 Production and hosting by Elsevier B.V. on behalf of Cairo University.

Introduction

Consider a one-dimensional series $Y_N = (y_1, \ldots, y_N)$ of length $N$. Mapping this series into a sequence of lagged vectors with size $L$, $X_1, \ldots, X_K$, with $X_i = (y_{i+L}, \ldots, y_{i+L-1})^T \in \mathbb{R}^L$ provides the trajectory matrix $X = (x_{ij})_{i,j=1}^{L,K}$, where $L(2 \leq L \leq N/2)$ is the window length and $K = N-L+1$;

$$X = [X_1, \ldots, X_K] = (x_{ij})_{i,j=1}^{L,K} = \begin{bmatrix} y_1 & y_2 & \cdots & y_K \\ y_2 & y_3 & \cdots & y_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & \cdots & y_N \end{bmatrix}.$$
Furthermore, such Hankel matrix $X$ naturally appears in multivariate analysis and signal processing, particularly in Singular Spectrum Analysis, where each of it column represents the L-lagged vector of observations in $\mathbb{R}^d$ [11,12]. Accordingly, the aim was to determine the accurate dimension of the system, that is the smallest dimension with which the filtered series is reconstructed from a noisy signal. In this case, the main analysis is based on the study of the eigenvalues and corresponding eigenvectors. If the signal component dominates the noise component, then the eigenvalues of the random matrix $X$ have a few large eigenvalues and many small ones, suggesting that the variations in the data takes place mainly in the eigenspace corresponding to these few large eigenvalues. Note that the number of correct singular values, $r$, for filtering and noise reduction, is increased with the increased $L$ which makes the comparison among different choices $(L, r)$ more difficult. Furthermore, despite the fact that several approaches have been proposed to identify the values of $r$ [13], due to a lack of substantial theoretical results, none of them consider the distribution of singular values of $X$. Here, we study the empirical distribution of singular values of $X$ for different situations considering various criteria. Accordingly, the theoretical results on the eigenvalues of $XX^T$ divided by its trace with a new view is considered in Main results. The empirical results using simulated data are presented in The empirical distribution of $\xi_i$. Some conclusions and recommendations for future research are drawn in Conclusion.

Main results

The singular values of $X$ are the square root of the eigenvalues of the $L$ by $L$ matrix $XX^T$, where $X^T$ is the conjugate transpose. For a fixed value of $L$ and a series with length $N$, the trace of matrix $XX^T$, $tr(XX^T) = ||X||^2_F = \sum_{i=1}^L \lambda_i$, where $|||_F$ denotes the Frobenius norm, and $\lambda_i (i = 1, \ldots, L)$ are the eigenvalues of $XX^T$. Note that the increase of sample size $N$ leads to the increase of $\lambda_i$ which makes the situation more complex. To overcome this issue, we divide $XX^T$ by its trace $(XX^T/\sum_{i=1}^L \lambda_i)$, which provides the following properties.

**Proposition 1.** Let $\xi_1, \ldots, \xi_L$ denote eigenvalues of the matrix $(XX^T/\sum_{i=1}^L \lambda_i)$, where $X$ is a Hankel trajectory matrix with $L$ rows, and $\lambda_i (i = 1, \ldots, L)$ are the eigenvalues of $XX^T$. Thus, we have the following properties:

1. $0 \leq \xi_L \leq \ldots \leq \xi_1 \leq 1$,
2. $\sum_{i=1}^L \xi_i = 1$,
3. $\xi_1 \geq 1/L$,
4. $\xi_L \leq 1/L$.

**Proof.** The first two properties are simply obtained from matrix algebra and thus not provided here. The outermost inequalities are attained as equalities when, for example, $y_i = 1$ for all $i$. To prove the third property, the first two properties are used as follows. The second part confirms $\xi_1 + \xi_2 + \ldots + \xi_L = 1$. Thus, using the first property, $\xi_1 \geq \xi_i (i = 2, \ldots, L)$, we obtain $\xi_1 + \xi_2 + \ldots + \xi_L = L\xi_1 \geq 1 \Rightarrow \xi_1 \geq 1/L$. Similarly, for the fourth property, it is straightforward to show that $\xi_2 + \xi_L + \ldots + \xi_L = L\xi_L \leq 1 \Rightarrow \xi_L \leq 1/L$, since $\xi_L \leq \xi_i (i = 1, 2, \ldots, L - 1)$, and $\sum \xi_i = 1$. Note also that if $y_L = 1$ and $y_i = 0$ for $i \neq L$ then $\xi_1 = \ldots \xi_L = 1/L$. Rational number theory can also aid us to provide more informative inequalities (for more information see [14]).

Let us now evaluate the empirical distribution of $\xi_i$. In doing so, a series of length $N$ from different distributions, is generated $m$ times. For consistency and comparability of the results, a fixed value of $L$, here 10, is used for all examples and case studies throughout the paper. For point estimation and comparing the mean value of eigenvalues, the average of each eigenvalue in $m$ runs is used; $\xi_i$ as defined before, $i = 1, \ldots, L$, and $m$ is the number of the simulated series. Here we consider eight different cases that can be seen in real life examples:

(a) White Noise; WN.
(b) Uniform distribution with mean zero; $U(-\pi, \pi)$.
(c) Uniform distribution; $U(0, \pi)$.
(d) Exponential distribution; Exp($\pi$).
(e) $\beta + \exp(\pi)$.
(f) $\beta + t$.
(g) Sine wave series; $\sin(\pi)$.
(h) $\beta + \sin(\pi) + \sin(\theta)$.

where $\alpha = 1, \beta = 2, \varphi = 2\pi/12, \theta = 2\pi/5$, and $t$ is the time which is used to generate the linear trend series.

The effect of $N$

In this section, we consider the effect of the sample size, $N$ on $\xi_i$. Fig. 1 demonstrates $\xi_i$ for different values of $N$ for cases (a)–(c) considered in this study. In Fig. 1, $\xi_i$ has a decreasing pattern for different values of $N$. It can be seen that, for a large $N$, $\xi_i \rightarrow 1/10$ for cases (a) and (b). Thus, increasing $N$ clearly affects the values of $\xi_i$ for the white noise (a) and uniform distribution (b). However, there is no obvious effect on $\xi_i$ for other cases. For example, for case (c), $\xi_1$ is approximately equal to 0.8 for different values of $N$, and $\xi_{i\neq1}$ is less than 1/10 (see Fig. 1 (right)).

Although the pattern of $\xi_i$ for the uniform distribution (c) is similar to exponential case (d), but for case (c), $\xi_i$ is greater than $\xi_i$ comparing to the case (d), whilst other $\xi_i$ are smaller. It has been observed that $\xi_i$ has similar patterns for cases (e), (f), and (g). The values of $\xi_i$ for cases (a) and (b), where $Y_N$ generated from a symmetric distribution, are approximately the same. The results clearly indicate that increasing $N$ does not have a significant influence on the mean of $\xi_i$ for all cases except (a) and (b). As a result, if $Y_N$ is generated from WN or $U(-1, 1)$, then increasing $N$ will affect the value of $\xi_i$ significantly.

The patterns of $\xi_i$

Let us now consider the patterns of $\xi_i$ for $N = 10^5$. For the white noise distribution (a) and trend series (f), $\xi_i$ has different pattern. It is obvious that, for the white noise series, $\xi_i$ converges asymptotically to 1/10, whilst for the trend series $\xi_i$ is approximately equal to 1, and $\xi_{i\neq1}$ tends to zero. Similar results were obtained for the uniform distributions, cases (b) and (c), respectively.
Both samples generated from exponential distribution have similar patterns for $\xi_i$. However, it is noticed that adding an intercept $\beta$ to the exponential distribution, increases the value of $\xi_1$ and decreases other $\xi_i$. The results indicate that $\xi_1 \approx 0.6$ and $\xi_2 \approx 0.4$, whilst, other $\xi_i$ are not zero, whereas other $\xi_i$ tend to zero. It was noticed that the value of $\xi_1$ for sine wave (h) is greater than its value for sine case (g), whilst the value of $\xi_2$ is less.

The empirical distribution of $\xi_i$

The distribution of $\xi_i$ was assessed for different values of $L$. It was observed that the histograms of $\xi_i$ are similar for different values of $L$ (the results are not presented here). Therefore, for graphical aspect, and visualization purpose, $L = 10$ is considered here. The results are provided only for $\xi_1$, $\xi_5$ and $\xi_{10}$, for the cases ((a), (b), (c)), as similar results are observed for other $\xi_i$. Fig. 2 shows histogram of $\xi_i$ for $i = 1, 5, 10$ for $L = 10$, and $m = 5000$ simulations. It appears that the histogram of $\xi_1$, is skewed to the right for samples taken from WN (a) and uniform distributions (b), whilst for the data generated from the uniform (c) and exponential (d) distributions, might be symmetric. For the middle $\xi_i$, the histogram might be symmetric for the four cases (the results only provided for $\xi_5$), whilst the distribution of $\xi_{10}$ is skewed to the left.

For cases, exponential distribution (e), trend series (f), and sine wave series (g) and complex series (h), we have standardized $\xi_i$ to have conveying information about their distributions. Fig. 3 shows the density of $\xi_i$ for $i = 1, 2, 3, 5, 6, 10$ for those cases. It is clear that $\xi_i$ has different histogram for these cases, and also different from what was achieved for the white noise and uniform distributions with zero mean. Remember that, if $Y_N$ generated from a symmetric distribution, like case (a) and (b), $\xi_1$ has a right skewed distribution. Moreover, it is interesting that $\xi_{10}$ has a negative skewed distribution for all cases except the trend series and sine cases (g) and (h).

Additionally, it should be noted that, for sine series (g), both $\xi_1$ and $\xi_2$ have similar distributions, whereas other $\xi_i$ have right skewed distributions. It is obvious that the distribution of $\xi_i$ for sine series (h) becomes skewed to the right for $\xi_i$ ($i = 6, \ldots, 10$). Remember that the sine wave (h) was generated from an intercept and two pure sine waves. This means that the components related to the first few eigenvalues create the sine series (h). The results confirm that adding even an intercept alone will change the pattern of $\xi_i$. Note that an intercept can be considered as a trend in time series analysis. Generally, if we add more non stochastic components to the noise series, for instance trend, harmonic and cyclical components, then the first few eigenvalues are related to those components and as soon as we reach the noise level the pattern of eigenvalues will be similar to those found for the noise series.

Usually every harmonic component with a different frequency produces two close eigenvalues (except for frequency 0.5 which provides one eigenvalues). It will be clearer if $N$, $L$, and $K$ are sufficiently large [15]. In practice, the eigenvalues of a harmonic series are often close to each other, and this fact simplifies the visual identification of the harmonic components [15]. Thus, the results obtained here are very important for signal processing and time series techniques where noise reduction and filtering matter.

Generally, it is not easy to judge visually if $\xi_i$ has a symmetric distribution, thus it is necessarily to consider other criteria like statistical test. We calculate the coefficient of skewness
which is a measure for the degree of symmetry in the distribution of a variable. Table 1 represents the coefficient of skewness for \( \zeta_i \) for all cases. Bulmer [16] suggests that, if skewness is less than \(-1\) or greater than \(+1\), the distribution is highly skewed; if skewness is between \(-1\) and \(-1/2\) or between \(+1/2\) and \(+1\), the distribution is moderately skewed, and finally if skewness is between \(1/2\) and \(+1\), the distributions are approximately symmetric. Therefore, we can say that, for instance, the distribution of \( \zeta_4 \) for cases ((c),…,(f)), and \( \zeta_5 \) for all cases might be symmetric.

D’Agostino–Pearson normality test [17] is applied here to evaluate this issue properly. It is also known as the omnibus test because it uses the test statistics for both the skewness and kurtosis to come up with a single \( p \)-value and quantify how far from Gaussian the distribution is in terms of asymmetry and shape. The \( p \)-value of D’Agostin test was significant, greater than 0.05 for \( \zeta_4 \), for cases ((c),…,(f)), whereas, it is less than 0.05 for other cases ((a), (b), (g), (h)). Therefore, we accept the null hypothesis that the data of \( \zeta_4 \) for cases ((c),…,(f)) are not skewed and as a result are symmetric. Moreover, \( \zeta_5 \) has a symmetric distribution for all cases, except the trend series and sine waves. The distribution of \( \zeta_5(k = 2, 4) \), for the exponential case (d) is symmetric, whereas skewed for the exponential case with intercept (e).

In terms of the distribution of \( \zeta_i \) for the trend series and sine wave (g), the distributions of \( \zeta_i \) for the trend series are totally different to the distributions of other \( \zeta_i \), which becomes skewed distribution. Note that the distribution of \( \zeta_i \) (i = 1, 2) for the trend series is symmetric, whilst skewed for sine wave (g). For sine series (h), the distribution of \( \zeta_i \) (i = 1, …, 5) is different from the distribution of \( \zeta_i \) (i = 6, …, 10). It is obvious from the figure that \( \zeta_i \) (i = 6, …, 10) has a right skewed distribution.

Table 1 The coefficient of skewness for \( \zeta_i \), \( i = 1, \ldots, 10 \) for all cases.

| WN       | \( U(-1, 1) \) | \( U(0, 1) \) | \( Exp(1) \) | \( 2 + Exp(1) \) | \( \sin(q) \) | \( 2 + \sin(q) + \sin(\delta) \) | \( 2 + t \) |
|----------|----------------|----------------|-------------|-----------------|--------------|---------------------------------|-------------|
| \( \zeta_1 \) | 0.991          | 0.450          | 0.005       | -0.003          | -0.126       | 0.186                           | -0.764      | 0.466                      |
| \( \zeta_2 \) | 0.692          | 0.733          | 0.428       | 0.330           | 0.230        | -0.186                          | 0.273       | -0.544                     |
| \( \zeta_3 \) | 0.461          | 0.502          | 0.224       | 0.280           | 0.154        | 0.691                           | 0.025       | 0.995                      |
| \( \zeta_4 \) | 0.401          | 0.234          | 0.075       | 0.092           | 0.154        | 0.623                           | -0.096      | 0.781                      |
| \( \zeta_5 \) | 0.099          | 0.021          | 0.055       | 0.077           | 0.153        | 0.624                           | -0.045      | 0.915                      |
| \( \zeta_6 \) | -0.140         | -0.130         | -0.001      | 0.071           | 0.154        | 0.649                           | 0.775       | 0.835                      |
| \( \zeta_7 \) | -0.37          | -0.230         | -0.041      | -0.102          | 0.145        | 0.690                           | 0.632       | 1.020                      |
| \( \zeta_8 \) | -0.503         | -0.460         | -0.033      | -0.139          | 0.110        | 0.855                           | 0.716       | 1.135                      |
| \( \zeta_9 \) | -0.577         | -0.520         | -0.162      | -0.226          | 0.021        | 1.970                           | 1.020       | 1.484                      |
| \( \zeta_{10} \) | -0.810         | -0.790         | -0.371      | -0.480          | -0.036       | 1.880                           | 1.459       | 2.030                      |

Fig. 3 The density of \( \zeta_i \), \( i = 1, \ldots, 6, 10 \) for cases ((e),…,(h)).

Conclusions

The pattern of the eigenvalues of the matrix \( \mathbf{X}\mathbf{X}'/\sum_{i=1}^{10} \lambda_i \), generated from different distributions was studied, and several properties were introduced. We have considered symmetric, nonsymmetric distributions, trend and sine wave series. The results indicate that for a large sample size \( N, \zeta_i N \rightarrow 1/L \) for the symmetric distributions (the white noise and the uniform distributions with zero mean), whilst this convergence has not been observed for other cases. The results also indicate that, for the symmetric cases, the pattern of the first eigenvalue is skewed, whilst it can be symmetric for the trend and nonsymmetrical distributions. Furthermore, for all cases under this study, the distribution of the middle \( \zeta_i \) for \( L = 10 \), can be symmetric except the pattern of \( \zeta_5 \) for the trend case and both sine series. It is found that the last eigenvalue has a positive skewed distribution, for all cases except the trend series and sine waves. For future research, the theoretical distribution of the matrix \( \mathbf{X}\mathbf{X}'/\sum_{i=1}^{10} \lambda_i \) is of our interest.

Furthermore, we aim to evaluate the applicability of the results found here for noise reduction of the chaotic series. Additionally, we are applying the properties obtained here as extra criteria for filtering series with complex structure. We may also consider a test to evaluate the \( k \) largest eigenvalues, to decide whether the distribution of the eigenvalues can resemble the particular distribution of the eigenvalues. In addition, the distribution of the smallest eigenvalue is as well of great interest, for example, because its behavior is used to prove its convergence to the circular law. Accordingly, the study of the local properties of the spectrum as well as the related distribution is of interest.
Conflict of Interest

The authors have declared no conflict of interest.

Compliance with Ethics Requirements

This article does not contain any studies with human or animal subjects.

References

[1] Hassani H, Soofi A, Zhigljavsky A. Predicting inflation dynamics with Singular Spectrum Analysis. J Roy Stat Soc – Ser A 2013;176(3):743–60.
[2] Hassani H, Heravi H, Zhigljavsky A. Forecasting European industrial production with Singular Spectrum Analysis. Int J Forecast 2009;25(1):103–18.
[3] Sanei S, Lee TKM, Abolghasemi V. A new adaptive line enhancer based on Singular Spectrum Analysis. IEEE Trans Biomed Eng 2012;59(2):428–34.
[4] Sanei S, Ghodsi M, Hassani H. An adaptive singular spectrum analysis approach to murmur detection from heart sounds. Med Eng Phys 2011;33(3):362–7.
[5] Peller V. Hankel operators and their applications. New York: Springer; 2003.
[6] Hassani H, Thomakos D. A review on Singular Spectrum Analysis for economic and financial time series. Stat Interface 2010;3(3):377–97.
[7] Chugunov VN. On the parametrization of classes of normal Hankel matrices. Comput Math Math Phys 2011;51(11):1823–36.
[8] Pastur LA. A simple approach to the global regime of Gaussian ensembles of random matrices. Ukrainian Math J 2005;57(6):936–66.
[9] Naronic P. On the universality of the distribution of the generalized eigenvalues of a pencil of Hankel random matrices. Random Matrices: Theory Appl 2013;2(1):1–14.
[10] Edelman A, Plamen K. Eigenvalue distributions of beta-Wishart matrices. Random Matrices: Theory Appl 2014;3(2):1–11.
[11] Hassani H, Mahmoudvand R. Multivariate singular spectrum analysis: a general view and new vector forecasting approach. Int J Energy Stat 2013;01:55–83.
[12] Sanei S, Ghodsi M, Hassani H. An adaptive singular spectrum analysis approach to murmur detection from heart sounds. Med Eng Phys 2011;33:362–7.
[13] Golyandina N, Nekrutkin V, Zhigljavsky A. Analysis of time series structure: SSA and related techniques. Chapman & Hall/CRC; 2001.
[14] Niven I. Irrational numbers, ch. VII, pp. 83–88; also p. 157. The Mathematical Association of America; 2005.
[15] Hassani H. Singular spectrum analysis: methodology and comparison. J Data Sci 2007;5:239–57.
[16] Bulmer M. Principles of statistics. New York: Dover; 1979.
[17] D’Agostino RB. In: D’Agostino RB, Stephens MA, editors. Tests for normal distribution in goodness-of-fit techniques. Marcel Dekker; 1986.