

11-1-2002

# Exploration Of Distributions Of Ratio Of Partial Sum Of Sample Eigenvalues When All Population Eigenvalues Are The Same

Moonseong Heo  
*Cornell University*

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

## Recommended Citation

Heo, Moonseong (2002) "Exploration Of Distributions Of Ratio Of Partial Sum Of Sample Eigenvalues When All Population Eigenvalues Are The Same," *Journal of Modern Applied Statistical Methods*: Vol. 1 : Iss. 2 , Article 51.

DOI: 10.22237/jmasm/1036109400

Available at: <http://digitalcommons.wayne.edu/jmasm/vol1/iss2/51>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

## Exploration Of Distributions Of Ratio Of Partial Sum Of Sample Eigenvalues When All Population Eigenvalues Are The Same

Moonseong Heo

Department of Psychiatry  
Weill Medical College  
Cornell University

---

This paper explores empirically the first two moments of ratio of the partial sum of the first two sample eigenvalues to the sum of all eigenvalues when the population eigenvalues of a covariance matrix are all the same. Estimation of the first two moments can be practically crucial in assessing non-randomness of observed patterns on planar graphical displays based on lower rank approximations of data matrices. For derivation of the moments, exact and large sample asymptotic distributions of the sample ratios are reviewed but neither can be applicable to derivation of the moments. Therefore, I rely on simulations, where data matrices  $\mathbf{X}$  with order  $n \times m$  element-wise independent normal distribution with mean 0 and variance  $\sigma^2$  are assumed, that is,  $\mathbf{X} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{nm})$ , and then derive formulas for estimates of means and standard deviations of the sample ratios within a range of order of the data matrix. The derivations are based on the biplot graphical diagnostic methods proposed by Bradu and Gabriel (1976).

Keywords: Bias, biplot, eigenvalues, multivariate Gaussian; Schönemann-Lingoes-Gower coefficient.

---

### Introduction

Lower rank approximations of data matrices  $\mathbf{X}$  ( $n$  rows for individuals,  $m$  columns for variables) are much used in data analysis. The closeness of their fit to  $\mathbf{X}$  is frequently measured by the ratio of the sum of the first  $s$  ( $< m$ ) eigenvalues of  $l_1^2, l_2^2, \dots, l_s^2$  of  $\mathbf{X}^T \mathbf{X}$  to the total of all the eigenvalues  $l_1^2, l_2^2, \dots, l_m^2$  of  $\mathbf{X}^T \mathbf{X}$ , where  $s$  is the rank of the approximation. In particular, the rank  $s$  is usually chosen to be 2 for planar graphical displays, by which data analysts often want to see if they reveal any patterns in population expectations  $E(\mathbf{X}) = \Xi$  and/or covariance structure.

---

The author is very grateful to Dr. K. Ruben Gabriel for his valuable insights and comments, and to John T. Hutchens for the manuscript preparation. This study was supported in part by NIH grants P30DK26687 and P30MH49762. E-mail: moh2002@med.cornell.edu.

Accordingly, confirmation of such visual assessments is usually based on the quantities of the closeness of the planar displays to the data matrix measured by  $r_{(2)}^2 = (l_1^2 + l_2^2) / \sum_{i=1}^m l_i^2$ . This closeness coefficient is equal to the Schönemann - Lingoes - Gower coefficient  $r_{(2)}^2 = \text{trace} \left\{ (\tilde{\mathbf{X}}^T \mathbf{X} \tilde{\mathbf{X}})^{1/2} \right\} / \|\tilde{\mathbf{X}}\| \|\mathbf{X}\|$  (Gower 1971; Lingoes & Schönemann, 1974) as noted by Heo (1996), where  $\tilde{\mathbf{X}}$  is the Euclidian minimum distance rank 2 approximation of  $\mathbf{X}$ .

It has not been clear, however, how large value of  $r_{(2)}^2$  can play the role of a threshold for signaling non-random patterns on the planar displays, which are not overwhelmed by sampling variations. Furthermore, the threshold will depend on the order of data matrices,  $m$  and  $n$ . First, with respect to dependence on  $m$ ,  $r_{(2)}^2$  has its algebraic minimum  $2/m$  because the sample eigenvalues  $l_1^2, l_2^2, \dots, l_m^2$  are ordered in a

descending manner. Secondly, the larger  $n$ , the less will be sampling variations of the patterns of graphical displays. Therefore, observed patterns on graphical displays with  $r_{(2)}^2 = 0.45$  when  $m = 5$  may be less meaningful than those with  $r_{(2)}^2 = 0.45$  when  $m = 30$  for the same  $n$  — the former is relatively much closer to its minimum. One example of the latter case can be found in the biplot of  $n = 100$  archetypal patients with  $m = 30$  psychiatric variables (Strauss et al., 1979; Heo & Gabriel, 2001), where five distinctive clusters of patients of the same diagnosis within each cluster are displayed well enough to convince a data analyst that the patterns on the biplot may indeed represent patterns of population expectation, despite of the moderate  $r_{(2)}^2 = 0.45$ .

The significance of non-random pattern, however, must be inferred based on a sampling distribution of  $r_{(2)}^2$ . Specifically, if an observed  $r_{(2)}^2$  is above the 95 or 97.5 percentile of the sampling distribution, it may indicate that the pattern on planar displays may not be random and may be revealing patterns of population characteristics. Therefore, to provide such thresholds or critical values, I attempt to draw the sampling distribution of  $r_{(2)}^2$  under an  $m$ -variate null Gaussian model:

$$\mathbf{X} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{nm}). \tag{1}$$

In this situation, planar displays of  $\tilde{\mathbf{X}}$  show patterns solely due to random noise  $\sigma^2$ , not due to  $E(\mathbf{X}) = \mathbf{\Xi}$ , and all the eigenvalues of  $E(\mathbf{X}^T \mathbf{X})$ ,  $\lambda_1^2, \lambda_2^2, \dots, \lambda_m^2$ , are the same as  $\sigma^2$ .

I review what is known about the exact and asymptotic distribution of the sample eigenvalues  $l_1^2, l_2^2, \dots, l_s^2$  of  $\mathbf{X}^T \mathbf{X}$  under the null Gaussian model (1) and try to derive sampling distributions of  $r_{(2)}^2$  thereof. However, based on this review and to my knowledge, currently existing normal theories do not seem to be either practical or applicable for derivations of the sampling distribution. Therefore, relying on computer simulations under the null model (1), I attempt to derive empirical models for estimates of  $E(r_{(2)}^2)$

and  $SD(r_{(2)}^2)$ , the first two moments, through assessments of a relative bias  $\beta^2 = E(r_{(2)}^2)/(2/m)$  in comparison to the algebraic minimum and its SD. These two moments can provide basis for normal approximations to the sampling distributions and eventually for the thresholds, or the critical values. I use biplot for a model diagnostic tool as demonstrated in Gabriel and Braud (1971). Issues concerning normal approximation, practical meaning of non-random patterns displayed on the planar spaces and a justification of the null model (1) are discussed.

### Methods

#### Exact distribution

When all the population eigenvalues  $\lambda_i^2$  are equal, i.e.  $\lambda_i^2 = \lambda^2$  for all  $i$ , the exact joint distribution of the sample eigenvalues  $l_i^2$  can be expressed as (e.g., James, 1964):

$$f(\underline{l}^2) = \left(\frac{n}{2\lambda^2}\right)^{nm/2} \frac{\pi^{m^2/2}}{\Gamma_m(n/2)\Gamma_m(m/2)} \exp\left(-\frac{n}{2\lambda^2} \sum_i l_i^2\right) \prod_{i=1}^m l_i^{(n-m-1)} \prod_{i<j} (l_i^2 - l_j^2)$$

where  $\underline{l}^2 = (l_1^2, \dots, l_m^2)^T$  and

$$\Gamma_m(\cdot) = \pi^{m(m-1)/4} \prod_{i=1}^m \Gamma(\cdot - (i-1)/2).$$

Based on this, the exact density of  $r_{(2)}^2 = (l_1^2 + l_2^2) / \sum_{i=1}^m l_i^2$  under the null model, can be obtained by using the change of variable technique. Also, Krishnaiah and Waikar (1971) studied the exact marginal distribution of each individual sample eigenvalue, when all the population eigenvalues are equal, by applying Laplace's expansion to the Vandermonde determinant  $\prod_{i<j} (l_i^2 - l_j^2)$ .

Nevertheless, whichever way is used for calculation of the moments of  $r_{(2)}^2$  under the Null Gaussian model (1), the calculation will be very complicated and tedious, even by numerical computations. Therefore, application of asymptotic or approximation theories might be

preferred for a derivation of the sample moments of  $r_{(2)}^2$  as follows.

Asymptotic distributions

Under the assumption of simplicity (or at least two different multiplicities) of the population eigenvalues, asymptotic (representations for) distributions of the sample eigenvalues were extensively discussed in the 1960s and 70s (e.g., Muirhead, 1978). The joint distributions of sample eigenvalues, under that assumption, involve hypergeometric functions expressed in integral representations. On these integrals are focused the approximations, which are basically determined by the maximum values of the integrands involving ‘linkage factors’ of  $(l_i^2 - l_j^2)^{-1}$ . Such approximations are, therefore, inapplicable to the joint (or marginal) asymptotic behaviors of sample eigenvalues when all the population eigenvalues are equal. Hence, the derivation of an asymptotic distribution of  $r_{(2)}^2$  under multiplicity from the asymptotic joint distribution of sample eigenvalues under the simplicity would be misleading. The following are such examples.

An asymptotic distribution of  $l_i^2$  is  $l_i^2 \sim N(\lambda_i^2, 2\lambda_i^4/(n-1))$  (Anderson, 1963) and  $Cov(l_i^2, l_j^2) \approx 0$  for  $i \neq j$ , provided all eigenvalues are distinct. Under the Null Gaussian model (1), it might become  $l_i^2 \sim N(\sigma^2, 2\sigma^2/(n-1))$  for all  $i$ , if the multiplicity of  $\lambda_i^2$  is ignored, i.e., the fact that  $\lambda_i^2 = \lambda^2$  for all  $i$  is ignored. Applying Taylor approximation to each  $l_i^2$  about each corresponding  $\lambda_i^2$ :

$$\frac{l_1^2 + l_2^2}{\sum l_i^2} = \frac{\lambda_1^2 + \lambda_2^2}{\sum \lambda_i^2} + \frac{1}{\sum \lambda_i^2} \sum (l_i^2 - \lambda_i^2) \left[ I\{i \leq 2\} - \frac{\lambda_1^2 + \lambda_2^2}{\sum \lambda_i^2} \right]$$

where  $I\{\cdot\}$  is an indicator function. Under the Null Gaussian model (1), the right hand side can be reduced to  $2/m + (l_1^2 + l_2^2)/m\sigma^2 - 2\sum l_i^2/m^2\sigma^2$ , which is asymptotically Gaussian with mean  $2/m$  and variance  $4(m-2)/(n-1)m^3$ . This shows very

roughly that the distribution of  $r_{(2)}^2$  does not depend asymptotically on  $\sigma^2$ , as it should not, because  $r_{(2)}^2$  is a studentized ratio. However, the asymptotic expectation  $2/m$  is wrong, since  $(l_1^2 + l_2^2)/\sum l_i^2$  is greater than  $2/m$  with probability one because the sample eigenvalues  $l_i^2$  are ordered in a descending manner.

Asymptotic distributions of functions of sample eigenvalues were investigated by several authors (e.g., Fang & Krishiniah, 1982). Fujikoshi (1980), for example, showed that the distribution functions of functions of sample eigenvalues can be expanded up to the order of  $n^{-1/2}$ , when certain assumptions (including the simplicity of the population eigenvalues) are met. Based on his approximation for the multivariate Gaussian  $\mathbf{X}$ ,  $E(r_{(2)}^2) \approx R_2^2 + a/n$  and  $Var(r_{(2)}^2) = \zeta^2/n$ , where  $R_2^2 = (\lambda_1^2 + \lambda_2^2)/\sum \lambda_i^2$ ,  $a = \sum_{i \neq j} T_i (\lambda_i^2 - \lambda_j^2)^{-1} \lambda_i^2 \lambda_j^2 + \sum T_{ii} \lambda_i^4$ ,  $T_i = I\{i \leq 2 - R_2^2\} / \sum \lambda_i^2$ ,  $T_{ij} = -(T_i + T_j) / \sum \lambda_i^2$ , and  $\zeta^2 = 2 \sum T_i^2 \lambda_i^4$ . Then, apply Fujikoshi's approximations to the set of population eigenvalues such that  $\lambda_i^2 = \lambda_{i+1}^2 + \varepsilon$ , for  $i = 1, \dots, m-1$ , and  $\lambda_m^2 = 1$ , where the difference  $\varepsilon$  of the consecutive population eigenvalues is very small. Numerical evaluations of the expectation of  $r_{(2)}^2/R_2^2$  and its standard deviation are tabulated in Table 1 for  $\varepsilon = 0.001$ . It is clear from this table that the approximation formulae do not work for these settings of population eigenvalues.

It follows that either exact or asymptotic normal theory does not seem to be applicable to the case of equal eigenvalues. This inapplicability leads us to simulation-based studies, which are described in the following, for empirical exploration of the behavior of the expectation and SD of  $r_{(2)}^2$  under the null Gaussian model (1).

Results

Bias, standard deviation, and simulation fit

The  $n$ -by- $m$  data matrices  $\mathbf{X}$  with  $3 \leq m \leq 30$  and  $30 \leq n \leq 1000$  ( $m \leq n$ ) under the null Gaussian model (1) are randomly generated for 1000 times for each combination of  $n$  and  $m$ , and then  $r_{(2)}^2$  is computed for each data matrix  $\mathbf{X}$ .

Table 1: Asymptotic expectation and (SD) of  $r_{(2)}^2/R_2^2$ :  $\varepsilon=0.001$ .

| $n$  | $M$                |                    |                    |                    |                         |                         |
|------|--------------------|--------------------|--------------------|--------------------|-------------------------|-------------------------|
|      | 3                  | 5                  | 10                 | 15                 | 20                      | 30                      |
| 30   | 26.0<br>(0.1<br>1) | 49.7<br>(0.1<br>4) | 77.1<br>(0.1<br>7) | 92.3<br>(0.1<br>7) | 102.<br>9<br>(0.1<br>7) | 118.<br>1<br>(0.1<br>8) |
| 60   | 13.5<br>(0.0<br>7) | 25.3<br>(0.1<br>0) | 39.1<br>(0.1<br>2) | 46.7<br>(0.1<br>2) | 52.0<br>(0.1<br>2)      | 59.6<br>(0.1<br>3)      |
| 90   | 9.3<br>(0.0<br>6)  | 17.2<br>(0.0<br>8) | 26.4<br>(0.0<br>9) | 31.4<br>(0.1<br>0) | 35.0<br>(0.1<br>0)      | 40.1<br>(0.1<br>0)      |
| 120  | 7.3<br>(0.0<br>5)  | 13.2<br>(0.0<br>7) | 20.0<br>(0.0<br>8) | 23.8<br>(0.0<br>9) | 26.5<br>(0.0<br>9)      | 30.3<br>(0.0<br>9)      |
| 150  | 6.0<br>(0.0<br>5)  | 10.7<br>(0.0<br>6) | 16.3<br>(0.0<br>7) | 19.3<br>(0.0<br>8) | 21.4<br>(0.0<br>8)      | 24.4<br>(0.0<br>8)      |
| 500  | 2.5<br>(0.0<br>3)  | 3.9<br>(0.0<br>4)  | 5.6<br>(0.0<br>4)  | 6.5<br>(0.0<br>4)  | 7.1<br>(0.0<br>4)       | 8.0<br>(0.0<br>4)       |
| 1000 | 1.8<br>(0.0<br>2)  | 2.5<br>(0.0<br>2)  | 3.3<br>(0.0<br>3)  | 3.7<br>(0.0<br>3)  | 4.1<br>(0.0<br>3)       | 4.5<br>(0.0<br>3)       |

The sample bias  $B^2$  of  $r_{(2)}^2$  is calculated for each data matrix  $\mathbf{X}$  of the same order by the ratio to its absolute lower bound  $2/m$ , that is,  $B^2 = mr_{(2)}^2/2$ . Table 2 contains averages of  $B^2$  and standard deviations  $SD(B^2)$  from 1,000 simulations for each combination of  $m$  and  $n$ .

Table 2: Averages and (SD) of  $B^2$  from 1000 simulations for each combination of  $m$  and  $n$ .

| $n$  | $M$                |                    |                    |                    |                    |                    |
|------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
|      | 3                  | 5                  | 10                 | 15                 | 20                 | 30                 |
| 30   | 1.19<br>(0.0<br>6) | 1.46<br>(0.0<br>9) | 1.96<br>(0.1<br>3) | 2.38<br>(0.1<br>5) | 2.75<br>(0.1<br>6) | 3.43<br>(0.1<br>8) |
| 60   | 1.13<br>(0.0<br>4) | 1.32<br>(0.0<br>7) | 1.65<br>(0.0<br>9) | 1.92<br>(0.1<br>0) | 2.16<br>(0.1<br>0) | 2.58<br>(0.1<br>2) |
| 90   | 1.11<br>(0.0<br>4) | 1.26<br>(0.0<br>5) | 1.52<br>(0.0<br>7) | 1.73<br>(0.0<br>8) | 1.92<br>(0.0<br>8) | 2.23<br>(0.0<br>8) |
| 120  | 1.09<br>(0.0<br>3) | 1.22<br>(0.0<br>5) | 1.45<br>(0.0<br>6) | 1.62<br>(0.0<br>7) | 1.78<br>(0.0<br>7) | 2.04<br>(0.0<br>7) |
| 150  | 1.08<br>(0.0<br>3) | 1.20<br>(0.0<br>4) | 1.40<br>(0.0<br>5) | 1.55<br>(0.0<br>6) | 1.68<br>(0.0<br>6) | 1.92<br>(0.0<br>6) |
| 500  | 1.05<br>(0.0<br>2) | 1.11<br>(0.0<br>2) | 1.21<br>(0.0<br>3) | 1.29<br>(0.0<br>3) | 1.36<br>(0.0<br>3) | 1.47<br>(0.0<br>3) |
| 1000 | 1.03<br>(0.0<br>1) | 1.08<br>(0.0<br>2) | 1.15<br>(0.0<br>2) | 1.20<br>(0.0<br>2) | 1.25<br>(0.0<br>2) | 1.32<br>(0.0<br>2) |

It shows that  $B^2$  seems to converge slowly to 1 as  $n$  increases and that the bias depends on the order of  $\mathbf{X}$ ; it goes down with  $n$  but up with  $m$ .

Fit of bias

I first fit averages of  $B^2$ , an estimate of the expected bias  $\beta^2 = E(r_{(2)}^2)/(2/m)$  by taking  $n$  and  $m$  as factor levels. The biplot is used for a model diagnostic tool (Bradu & Gabriel, 1978). The biplot of the data matrix of the averages of  $B^2$  in Table 2 minus the grand mean of the averages is displayed in Figure 1.

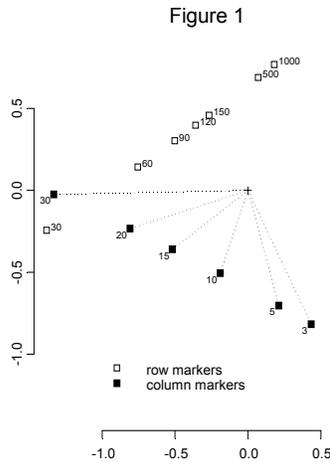


Figure 1: A biplot of  $\beta^2$  with rank 2 goodness of fit greater than 0.99.

This figure shows that the data matrix of  $B^2$  in Table 2 is virtually of rank 2 based on the goodness of fit greater than 0.99. Moreover, it is immediately seen that the sets of column and row markers are both collinear. This suggests that the data matrix must be closely fitted by means of Tukey's Degree of Freedom For Non-Additivity model (DOFNA; Tukey, 1949), i.e.,

$$\beta_{ij}^2 = \mu + \alpha a_i + \delta d_j + \tau a_i d_j + e_{ij} \quad (2)$$

subject to  $\sum a_i = \sum d_j = 0$  and  $\sum a_i^2 = \sum d_j^2 = 1$ . The subscripts  $i$  and  $j$  represent the levels of  $n$  and  $m$ , respectively. (Still, a rank 1 multiplicative model may be an alternative choice. However, a biplot of the data matrix without centering on the grand mean, though not presented herein, shows that the multiplicative model does not fit well.)

A summary graphic of the DOFNA model fit is shown in Figure 2. The residual sum of squares is 0.0037 with df 29, which means that the fit is almost perfect. In short, Figure 2 shows that: (a) There exists a clear interaction between row and column effects, which means that the coefficient  $\tau$  is significantly different from 0:  $\hat{\tau} = 1.84$ ,  $p < 0.001$ ; that is, the magnitude of the bias increases as  $m$  for a fixed  $n$  but the rate of increment is not constant over  $n$ ; (b)  $\beta^2$  seems to converge to 1 as  $n$  increases, as can be seen in Table 2; (c) Roughly, the effect of

the number of columns is close to linear but that of the number of rows is not; the intervals between consecutive row effects are not constant when the magnitude of the number of rows is taken into consideration.

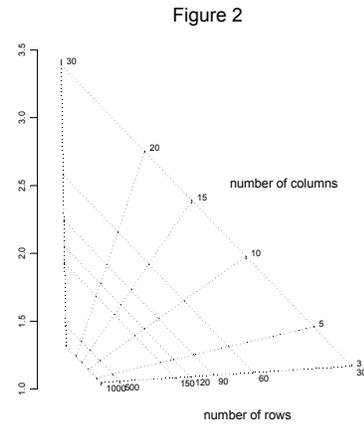


Figure 2: DOFNA fit to  $\beta^2$  with residual sum of squares 0.0037.

It should be recalled, however, that I am trying to formulate a function, which relates this model's parameters to the values (not the factor levels) of  $n$  and  $m$ . For this purpose, on the basis of plots of column effects versus  $m$  and row effects versus  $n$ , we modeled row and column effects as  $\alpha a_i = \gamma_3 / \sqrt{n}$  and  $\delta d_j = \gamma_1 m + \gamma_2 m^2$ , respectively. In light of the DOFNA model (2), this yields the following model:

$$\beta^2 = \eta + \gamma_1 m + \gamma_2 m^2 + (\gamma_3 + \gamma_4 m + \gamma_5 m^2) / \sqrt{n} + e$$

The least-square fit with significant ( $p$ -values  $< 0.001$ ) coefficients results in the following:

$$\hat{\beta}^2 = 1.0301 - 0.0068m + (-0.8319 + 0.6652m - 0.0060m^2) / \sqrt{n} \quad (3)$$

The residual sum of squares of this fit is 0.036 with df 37 and the multiple  $R^2$  is greater than 0.99. All of the fitted values of  $\beta^2$  are greater than

1 over the ranges of  $m$  and  $n$  considered:  $30 \leq n \leq 1000$  and  $3 \leq m \leq 30$ .

Fit of standard deviation

The biplot in Figure 3 with goodness of fit greater than 0.99 shows that the data matrix of  $SD(B^2)$  in Table 2 is also virtually of rank 2 and that the column markers are collinear. On the basis of Bradu and Gabriel (1976), the data matrix of  $SD(B^2)$  must be closely fitted by Mandel's row regression model (Mandel, 1961), that is,

$$SD(B^2) = \mu + \alpha a_i + \delta d_j + \theta c_i d_j + e_{ij}$$

subject to  $\sum a_i = \sum c_i = \sum d_j = 0$  and  $\sum a_i^2 = \sum c_i^2 = \sum d_j^2 = 1$ . The resulting residual sum of squares is  $0.89 \times 10^{-4}$  with df 24, which shows that this is an almost perfect fit.

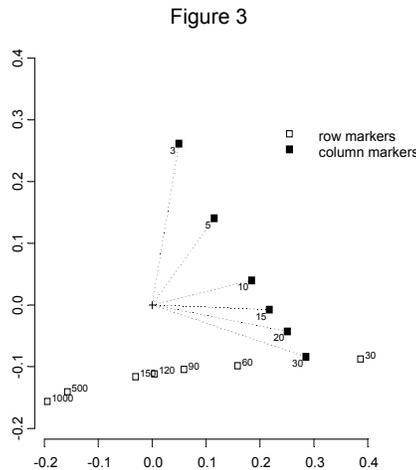


Figure 3: A biplot of  $SD(B^2)$  with rank 2 goodness of fit greater than 0.99.

The biplot in Figure 3, however, shows that the row markers are also virtually collinear. Furthermore, it was observed, though not presented herein, that the  $a_i$ 's and  $c_i$ 's are very similar up to a scale factor. These strongly suggest that Tukey's DOFNA model in a form of (2) can be an alternative fit to the data matrix of  $SD(B^2)$  in Table 2. The DOFNA fit results in a residual sum of squares of  $1.75 \times 10^{-4}$  with df 29. A summary graphic of this DOFNA fit is presented in Figure 4.

Figure 4

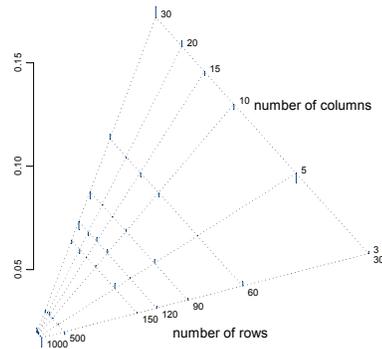


Figure 4: DOFNA fit to  $SD(B^2)$  with residual sum of squares 0.00018.

The structural relationship between  $SD(B^2)$  and the order of  $\mathbf{X}$  is clear;  $SD(B^2)$  seems to vanish slowly as  $n$  increases, which implies  $\beta^2$  converges in probability. Mandel's model is significantly better than the DOFNA model in fitting  $SD(B^2)$  data matrix with an approximated  $F$  ratio 4.65 and  $p$ -value 0.004. This DOFNA model, however, is simpler and easy to see graphically as shown in Figure 4, and its fit is also nearly perfect, which I chose for a functional model construction. Again, based on plots of column effects versus  $m$  and row effects versus  $m$ , I modeled column and row effects as follows:  $\delta d_j = \gamma_1 \log m$  and  $\alpha a_i = \gamma_2 / \sqrt{n}$ , respectively. It follows that

$$SD(B^2) = \eta + \gamma_1 \log m + (\gamma_2 + \gamma_3 \log m) / \sqrt{n} + e$$

(Nevertheless, Mandel's model yields the same form of this model.) The least-square fit with significant ( $p$ -values  $< 0.001$ ) coefficients results in the following:

$$\widehat{SD}(B^2) = 0.0128 - 0.0094 \log m + 0.3123 \log m / \sqrt{n} \tag{4}$$

The residual sum of squares of this fit is  $5.98 \times 10^{-4}$  with df 39 and the multiple  $R^2$  is greater than 0.99. All of the fitted values of

$SD(B^2)$  are positive over the ranges of  $m$  and  $n$  considered:  $30 \leq n \leq 1000$  and  $3 \leq m \leq 30$ .

### Discussion

Regarding features of the distribution of  $r_{(2)}^2$  under the null Gaussian model (1), I observe from the simulation that it is slightly skewed to the right for almost all combinations of  $m$  and  $n$ , but particulars of the asymptotic distributions are unknown. It follows that normal approximation of the distribution of  $r_{(2)}^2$  under the null Gaussian model with the expectation and standard deviation obtained from the formulae (3) and (4) is rather crude. Hypothesis testing based on this normal approximation would, therefore, be conservative. One might consider power transformations of  $r_{(2)}^2$  to have better approximations to normal distributions, or application of “delta” method to the first two moments.

Nevertheless, the crude normal approximation provides an idea of what the distribution of  $r_{(2)}^2$  might be under the null model. For example, to see how many multiples of  $SD(B^2)$  below the mean ensures  $B^2$  to be greater than 1, I calculate a multiple  $c$  from the fitted  $\beta^2$  and  $SD(B^2)$  in the following way:  $c = (\beta^2 - 1)/SD(B^2)$ . From formulae (3) and (4), the estimated minimum  $c$  over the considered ranges is 3.26 when  $m = 3$  and  $n = 30$ . This confirms that  $r_{(2)}^2$  is distributed well above the algebraic minimum of  $2/m$ . Moreover, the multiple  $c$  increases with  $m$ , implying that farther above  $2/m$   $r_{(2)}^2$  is distributed for bigger  $m$ . Indeed, as calculated based on the formulae (3) and (4), the percentiles of  $r_{(2)}^2 = 0.45$  are >99% and 1.4% when  $m$  are 30 and 5, respectively, for the same  $n=100$ . This confirms that observed patterns on graphical displays with  $r_{(2)}^2 = 0.45$  when  $m = 5$  may be less meaningful than those with  $r_{(2)}^2 = 0.45$  when  $m = 30$  for the same  $n$ , as stated in the introduction section.

It has been, however, suspected that  $r_{(2)}^2$  tends to locate between  $\rho_{(2)}^2$  and the absolute

minimum  $2/m$ , where  $\rho_{(2)}^2$  is the “actual” goodness of fit of  $\tilde{\mathbf{X}}$  to the expectation  $\Xi$ , that is  $\rho_{(2)}^2 = \text{trace}\left\{\left(\tilde{\mathbf{X}}^T \Xi \Xi^T \tilde{\mathbf{X}}\right)^{1/2}\right\} / \|\tilde{\mathbf{X}}\| \|\Xi\|$ , which should be a more appropriate measure for the “usefulness” of the lower rank approximation than the measure  $r_{(2)}^2$  of the closeness of  $\tilde{\mathbf{X}}$  to the data  $\mathbf{X}$  themselves, because patterns of the population expectations are to be inferred rather than patterns of data matrix. A simulation study of approximations using data generated under the  $m$ -variate Gaussian model  $\mathbf{X} \sim N(\Xi, \sigma^2 \mathbf{I}_{nm})$  with affine rank 2 expectation matrix  $\Xi$  has shown that  $r_{(2)}^2$  indeed underestimates  $\rho_{(2)}^2$  for many situations (Heo and Gabriel, 2001). Thus, non-significant  $r_{(2)}^2$  (less than 95- or 97.5%-tiles of the “null” sampling distribution) implies that the noise  $\sigma$  is much larger relative to the magnitude of  $\Xi$  — large enough so that  $\|\Xi\|/\sigma$  is approximately 0. This is the situation where the limiting distribution of  $\mathbf{X} \sim N(\Xi, \sigma^2 \mathbf{I}_{nm})$  can be approximated by  $\mathbf{X} \sim N(0, \sigma^2 \mathbf{I}_{nm})$  because  $\Xi$  reaches its zero limit relative to  $\sigma$ . That is, although it maintains all the time its rank, the expectation matrix  $\Xi$  tends to zero as the magnitude of  $\sigma$  increases, and at the limit it would not have any rank. Therefore, the null distribution  $\mathbf{X} \sim N(0, \sigma^2 \mathbf{I}_{nm})$  is valid for inferences of the critical values for significant  $r_{(2)}^2$ , which indicates that a planar display reveals patterns of population expectation of  $\Xi$  with a higher  $\rho_{(2)}^2$ .

In sum, the present study shows that there are clear structural patterns of expectation and variance of  $r_{(2)}^2$  under the null Gaussian model (1) as the order of data matrix  $\mathbf{X}$  varies. Construction of formulae for the expectations and standard deviations is elaborated through model diagnosis by use of the biplot. Similar application of the biplot diagnostic method can be extended to exploration of distributions of other ratios of partial sums of sample eigenvalues from data

matrices with bigger orders. The simulation-based approach employed in this paper seems appealing, since any large sample asymptotic theory does not seem to be applicable when all the population eigenvalues are the same. Therefore, the estimated first two moments of  $r_{(2)}^2$  may be useful in judging non-randomness of patterns of population expectations of data matrices displayed in a 2-dimensional space.

#### References

- Anderson, T. W. (1964). Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics*, 34, 122-148.
- Bradu, D., & Gabriel, K. R. (1978). The biplot as a diagnosis tool for models of two-way tables. *Technometrics*, 20, 46-68.
- Fang, C., & Krishniah, P. R. (1982). Asymptotic distributions of functions of the eigenvalues of some matrices for nonnormal populations. *Journal of Multivariate Analysis*, 12, 39-63.
- Fujikoshi, Y. (1980). Asymptotic expansions for the distributions of the sample roots under nonnormality. *Biometrika*, 67, 45-51.
- Gabriel, K. R. (1971). The biplot graphical display of matrices with application to principal component analysis. *Biometrika*, 58, 453-467.
- Gower, J. C. (1971). Statistical methods of comparing different multivariate analyses of the same data. In F.R. Hodson, D.G. Kendall, and P. Tautu (Eds), *Mathematics in the Archeological and Historical Sciences*, (pp. 138-149). Edinburgh: University Press.
- Heo, M. (1996). On the fit of sample graphical displays to patterns in population. *Unpublished Dissertation*, University of Rochester.
- Heo, M., & Gabriel, K. R. (2001). The fit of graphical displays to population patterns. *Computational Statistics and Data Analysis*, 36, 47-67.
- James, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Annals of Mathematical Statistics*, 35, 465-501.
- Krishnaiah, P. R. & Waikar, V. B. (1971). Exact joint distributions of any few ordered roots of a class of random matrices. *Journal of Multivariate Analysis*, 1, 308-315.
- Lingoes, J. C., & Schönemann, P. H. (1974). Alternative measures for fit for the Schönemann-Carroll matrix fitting algorithm. *Psychometrika*, 39, 423-427.
- Mandel, J. (1961). Non-additivity in two-way analysis of variance. *Journal of the American Statistical Association*, 56, 878-888.
- Muirhead, R. J. (1978). Latent roots and matrix variates: A review of some asymptotic results. *Annals of Statistics*, 6, 5-33.
- Strauss, J. S., Gabriel, K. R., Kokes, R. F., Ritzler, B. A., VanOrd, A., & Tarana, E. (1979). Do psychiatric patients fit their diagnosis? Patterns of symptomatology as described with the biplot. *Journal of Nervous and Mental Disease*, 167, 105-113.
- Tukey, J. W. (1949). One degree of freedom for non-additivity. *Biometrics*, 5, 232-242.