Journal of Multivariate Analysis 147 (2016) 265–272

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On the Tracy–Widom approximation of studentized extreme eigenvalues of Wishart matrices

Rohit S. Deo

44 West 4th Street, New York University, NY, 10012, USA

ARTICLE INFO

Article history:
Received 12 November 2014
Available online 13 February 2016

AMS 2000 subject classifications:
62H10
62H15
62E20

Keywords:
Tracy–Widom
Wishart matrix
Ratio of largest eigenvalue to trace
Sphericity

ABSTRACT

The few largest eigenvalues of Wishart matrices are useful in testing numerous hypotheses and are typically studentized as the noise variance is unknown. Specifically, the largest eigenvalue is studentized using the average trace of the matrix. However, this ratio has a distribution poorly approximated by its asymptotic one when either the sample size or dimension is not too large, making inference problematic. We present a simple variance adjustment that significantly improves the approximation and theoretically demonstrate the increase in power that this adjustment delivers compared to the power of the uncorrected studentized eigenvalue. We propose a bias corrected consistent estimator of the noise variance when studentizing the $(k+1)$st largest eigenvalue in the presence of exactly k spikes and a variance correction for the resulting studentized eigenvalue is proposed.

1. Introduction

The sample covariance matrix $\mathbf{S}_n = n^{-1}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i'$, based upon n *i.i.d.* zero-mean p dimensional real Gaussian vectors $\mathbf{X}_i$ follows the well-known Wishart distribution and, being an estimator of the population covariance matrix $\mathbf{\Sigma} \equiv \text{var}(\mathbf{X}_i)$, is a natural candidate upon which to base inference about the form of $\mathbf{\Sigma}$. Many such inference problems involve testing hypotheses about the eigenvalues $l_1 \geq \cdots \geq l_p > 0$ of $\mathbf{\Sigma}$ and are based upon the eigenvalues $\{\hat{l}_i\}$ of the sample covariance matrix, thus requiring an understanding of their distributions. Examples include tests for sphericity [13] and tests for detecting signals buried in noise under the spiked covariance model [5]. See also the comprehensive review paper by Paul and Aue [12].

The behavior of the sample eigenvalues $\{\hat{l}_i\}$ has long been known in the situation where the dimension p is fixed and not large relative to the sample size n. However, with the increasing availability of large data, attention has shifted to the high dimensional case where p is large and of the same order as n. Starting with the seminal paper by Johnstone [5], there has been a growing literature that has studied the asymptotic behavior of the first few largest eigenvalues of such high dimensional Wishart covariance matrices. A significant proportion of this literature (see, for example, [1,2,11], amongst others) has focused on the spiked covariance model where the population covariance matrix $\mathbf{\Sigma}$ is parametrized as

$$\mathbf{\Sigma} = \sum_{i=1}^{k}\lambda_i\boldsymbol{\theta}_i\boldsymbol{\theta}_i' + \sigma^2\mathbf{I}_p,$$

where the vectors $\boldsymbol{\theta}_i$ are orthonormal, $\lambda_1 \geq \cdots \geq \lambda_k > 0$ for some $k < p$, and $\sigma^2 > 0$. Such a parametrization would arise, for example, if the data vector $\mathbf{X}$ had a factor structure of the form

$$\mathbf{X} = \mathbf{A}\mathbf{u} + \mathbf{e},$$

E-mail address: rdeo@stern.nyu.edu.

http://dx.doi.org/10.1016/j.jmva.2016.01.010
0047-259X/

where $\mathbf{e} \sim \mathbf{N}_p\left(0, \sigma^2\mathbf{I}_p\right)$ and is independent of $\mathbf{u} \sim \mathbf{N}_k\left(0, \mathbf{I}_k\right)$, while $\mathbf{A}_{p\times k}$ is a deterministic matrix such that $\mathbf{A}\mathbf{A}'$ has spectral factorization given by $\sum_{i=1}^{k} \lambda_i\boldsymbol{\theta}_i\boldsymbol{\theta}_i'$. Under this spiked covariance model, the eigenvalues of $\boldsymbol{\Sigma}$ are $l_i = \lambda_i + \sigma^2$ for $i = 1, \ldots, k$ and $l_i \equiv \sigma^2$ for $i = k+1, \ldots, p$ and hypotheses about the rank of the factor loading matrix $\mathbf{A}$ can be posed equivalently in terms of null values for the "spikes" λ_i.

The fundamental null case where $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_p$ (i.e. $\lambda_i \equiv 0$ for all i) was considered by Johnstone [5], who showed in this case that the distribution of the normalized largest sample eigenvalue $\hat{l}_1/\sigma^2$, after appropriate centering and scaling, converges to the appropriate Tracy Widom distribution. However, the noise variance σ^2 is typically unknown and feasible versions of the normalized eigenvalue need to be "studentized", where σ^2 is replaced by an estimator of the noise variance σ^2, typically defined by the average trace of $\mathbf{S}_n$. Though the distribution of the infeasible normalized largest eigenvalue $\hat{l}_1/\sigma^2$ has been shown [7] to be well approximated, even in small samples, by the asymptotic distribution, Nadler [8] observed that this is no longer the case for the studentized version, which can be substantially undersized. Nadler [8] proposed a finite sample adjustment to the asymptotic cumulative distribution function that depends on the second derivative of the asymptotic distribution and showed through simulations that this adjustment provided a better approximation to the distribution of the studentized eigenvalue. However, this adjustment necessitates the calculation of adjusted critical values for every (p, n) combination. In Section 2 we motivate and propose instead a simple variance correction for the studentized statistic itself that is trivial to compute and also significantly improves the finite sample approximation. In addition, in Section 3 we study theoretically the impact this adjustment has on the power, increasing it substantially compared to that of the unadjusted studentized statistic which is undersized and at risk of losing power.

The issue of studentization also arises if the working assumption is that exactly k values of λ_i are positive and one is considering the distribution of the normalized $(k+1)$st sample eigenvalue $\hat{\lambda}_{k+1}/\sigma^2$. In Section 3 we propose a bias corrected consistent estimator for the noise variance by accounting for the k nuisance parameters λ_i. An analogous variance correction is proposed for the studentized $(k+1)$st eigenvalue and found to once again improve the finite sample approximation to the asymptotic distribution.

Throughout this paper, we will focus only on the case where $\mathbf{X}$ is real valued, though similar results can be obtained for complex valued data. We will also assume throughout that $\mathbf{X}$ has a Gaussian distribution, though the results of Soshnikov [14] imply that our results will continue to hold under less restrictive assumptions about the underlying distribution.

2. Non-spiked covariance matrix

In this section we consider the situation where $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_p$. Under this assumption, it is known [5] that as $\min(n, p) \to \infty$ such that $\lim \gamma \equiv p/n \in (0, \infty)$ we get,

$$(\hat{l}_1/\sigma^2 - \widetilde{\mu}_{n,p})/\widetilde{\sigma}_{n,p}, \xrightarrow{D} W_1 \tag{1}$$

where $\widetilde{\mu}_{n,p} = n^{-1}\left(\sqrt{n-1} + \sqrt{p}\right)^2, \widetilde{\sigma}_{n,p} = (\widetilde{\mu}_{n,p}/n)^{1/2}\left(1/\sqrt{n-1} + 1/\sqrt{p}\right)^{1/3}$ and W_1 obeys the Tracy Widom distribution of order 1.

Ma [7] proved that replacing the centering and scaling coefficients $\widetilde{\mu}_{n,p}$ and $\widetilde{\sigma}_{n,p}$ in (1) by $\mu_{n,p} = n^{-1}\left(\sqrt{n-1/2} + \sqrt{p-1/2}\right)^2$ and $\sigma_{n,p} = (\mu_{n,p}/n)^{1/2}\left(1/\sqrt{n-1/2} + 1/\sqrt{p-1/2}\right)^{1/3}$ resulted in an order of magnitude improvement in the finite sample approximation of the distribution of $\left(\hat{l}_1/\sigma^2 - \mu_{n,p}\right)/\sigma_{n,p}$ by that of W_1. Ma [7] also provided a detailed simulation study that demonstrated the quality of this improvement.

However, in practical situations the noise variance σ^2 is not known and one has to studentize $\hat{l}_1$ by using an estimate of σ^2. The usual estimator of σ^2 in the null case of $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_p$ is simply the average trace of $\mathbf{S}_n$,

$$\hat{\sigma}_0^2 \equiv p^{-1}trace\left(\mathbf{S}_n\right) = p^{-1}\sum_{i=1}^{p}\hat{l}_i.$$

Since every diagonal element of $\mathbf{S}_n$ is distributed as an independent $\sigma^2\chi_n^2/n$ random variable, it is easy to see that $\mathrm{E}\left(\hat{\sigma}_0^2\right) = \sigma^2$ while $\mathrm{var}\left(\sigma_0^2\right) = 2\sigma^4/(np)$. On the other hand, $\mu_{n,p} = O(1)$ while $\mathrm{var}\left(\hat{l}_1\right) = O\left(n^{-1}p^{-1/3}\right)$ and so it is easy to show that studentizing by replacing σ^2 by $\hat{\sigma}_0^2$ results in the same limiting distribution as in (1), viz.

$$\frac{\hat{l}_1/\hat{\sigma}_0^2 - \mu_{n,p}}{\sigma_{n,p}} \xrightarrow{D} W_1. \tag{2}$$

Unfortunately, in finite samples, the quality of this approximation for the statistic in (2) breaks down due to the studentization, as pointed out by Nadler [8], who noted that the test statistic in (2) is undersized under the asymptotic