On the distribution of the ratio of the largest eigenvalue to the trace of a Wishart Matrix

Boaz Nadler

Department of Computer Science and Applied Mathematics
Weizmann Institute of Science
Rehovot, Israel, 76100

Abstract

The ratio of the largest eigenvalue divided by the trace of a $p \times p$ random Wishart matrix with n degrees of freedom and identity covariance matrix plays an important role in various hypothesis testing problems, both in statistics and in signal processing. In this paper we derive an approximate explicit expression for the distribution of this ratio, by considering the joint limit as both $p, n \to \infty$ with $p/n \to c$. Our analysis reveals that even though asymptotically in this limit the ratio follows a Tracy-Widom (TW) distribution, one of the leading error terms depends on the *second derivative* of the TW distribution, and is non-negligible for practical values of p, in particular for determining tail probabilities. We thus propose to explicitly include this term in the approximate distribution for the ratio. We illustrate empirically using simulations that adding this term to the TW distribution yields a quite accurate expression to the empirical distribution of the ratio, even for small values of p, n.

Keywords: Ratio of largest eigenvalue to trace; Principal components analysis; Wishart matrices; Tracy-Widom distribution.

subject classification: 62H10, 62H15, 62E20, 62H25

1. Introduction

Let $x_1, \ldots, x_n$ be n i.i.d. p-dimensional observations from either a real valued or a complex valued Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, where the population covariance matrix is assumed to be of the form $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_{p \times p}$ with an unknown scaling factor σ^2. Denote the sample covariance matrix by

$$\mathbf{S}_n = \frac{1}{n} \sum_i x_i x_i^H \tag{1.1}$$

and let ℓ_j denote its eigenvalues, sorted in decreasing order, $\ell_1 \geq \ell_2 \geq \ldots \geq \ell_p$. In this setting, the matrix $\mathbf{S}_n$, upon division by the unknown factor σ^2, follows

a Wishart distribution with n degrees of freedom and with identity covariance matrix. For ease of notation we denote its average trace by

$$T = \frac{1}{p}\sum_{j=1}^{p} \ell_j = \frac{1}{p}\,\mathrm{Tr}(\mathbf{S}_n).$$

The focus of this paper is on the distribution of the ratio of the largest eigenvalue of $\mathbf{S}_n$ divided by its average trace, namely the distribution of the following random variable,

$$U = \frac{\ell_1}{\frac{1}{p}\sum_j \ell_j} = \frac{\ell_1}{T}. \tag{1.2}$$

Note that U is scale invariant and does not depend on the unknown noise level σ. Hence, for the rest of the paper, we assume w.l.o.g. that $\sigma = 1$.

The random variable U plays a key role in various scale independent hypothesis testing procedures, both in some classical problems in statistics as well as in some modern applications in signal processing. Classical examples include testing for the presence of interactions in multi-way data [8] and testing for equality of the population covariance to a scaled identity matrix [11]. Some modern signal processing applications include testing for the presence of signals in cognitive radio as well as non-parametric signal detection in array processing [2, 3]. Normalized Wishart matrices (e.g., with trace equal to one) are also a common model for random density matrices in quantum information channels, see for example [14, 20]. Hence, the largest eigenvalue of such matrices also follows the distribution of U.

Regretfully, despite its importance there is no simple-to-compute expression for the exact distribution of U. Various authors derived exact formulas for U in terms of high dimensional integrals or as inverses of certain Laplace transforms, which could then be evaluated numerically for small values of p, see [5, 18]. More recently, [12] developed asymptotic expansions for tail probabilities of U by considering the extrema of certain random fields. The resulting expressions, however, seem difficult to evaluate unless p is very small.

The difficulty in obtaining a simple closed form expression for the distribution of U is related, of course, to a similar difficulty regarding simply the largest sample eigenvalue ℓ_1. Whereas the exact distribution of ℓ_1 also has no simple to evaluate explicit expression, in the joint limit as both $p, n \to \infty$ with $p/n \to c$, various authors have shown that upon proper centering and scaling [7, 9],

$$\Pr\left[\frac{\ell_1 - \mu_{np}}{\sigma_{np}} < s\right] \to TW_\beta(s) \tag{1.3}$$

where TW_β denotes the Tracy-Widom distribution of order β, and $\beta = 1, 2$ for real valued or complex valued observations, respectively. Moreover, for carefully chosen centering and scaling functions μ_{np} and σ_{np}, the convergence rate in Eq. (1.3) is $O(p^{-2/3})$. For real valued observations, suitable expressions for μ_{np} and

σ_{np} are [13]

$$\mu_{np} = \frac{1}{n}\left(\sqrt{n-1/2}+\sqrt{p-1/2}\right)^2, \tag{1.4}$$

$$\sigma_{np} = \sqrt{\frac{\mu_{n,p}}{n}}\left(\frac{1}{\sqrt{n-1/2}}+\frac{1}{\sqrt{p-1/2}}\right)^{1/3}. \tag{1.5}$$

For the complex case, similar expressions appear in [6].

The results above imply that asymptotically U also follows a Tracy-Widom distribution. For a detailed proof we refer the reader to [3]. Here we provide an intuitive explanation for this: In the joint limit $p, n \to \infty, p/n \to c$, we have that $\mu_{np} \to (1+\sqrt{c})^2 = O(1)$, whereas $\sigma_{np} = O(\frac{1}{\sqrt{n}\,p^{1/6}}) = O(p^{-1/2-1/6})$. In contrast, the denominator $T = \mathrm{Tr}(\mathbf{S}_n)/p$ in the definition of U, Eq. (1.2) is distributed as a $\chi^2_{\beta np}/\beta np$ random variable, and thus has mean $\mathbb{E}[T] = 1$ and fluctuations of the order $O(1/\sqrt{pn}) = O(p^{-1})$. The key point is that asymptotically the fluctuations in T are negligible compared to those of ℓ_1. Combining these properties with Eq. (1.3), it then follows that as both $p, n \to \infty$,

$$\Pr\left[\frac{U-\mu_{np}}{\sigma_{np}} < s\right] \to TW_\beta(s). \tag{1.6}$$

Indeed, based on this analysis some recent works use Eq. (1.6) either to set the threshold corresponding to a given false alarm rate for various detection procedures, or to analyze them, see for example [3, 10]. Since these thresholds depend on tail probabilities of U, an interesting question is how accurate is the approximation in Eq. (1.6) for finite values of n and p, and in particular for the setting common in some modern signal processing applications, where $p = O(10)$ and $n \gg p$.

Before providing a theoretical analysis of this question it is instructive to first look at some simulation results. In figure 1, we compare the exact TW density with the empirical density of the largest eigenvalue ℓ_1 and of the ratio U, both centered and scaled by μ_{np} and σ_{np} as described above, for $\beta = 1$, $p = 20$ and $n = 500$. As shown in the figure, the empirical distribution of ℓ_1 is very well approximated by the limiting TW distribution. In contrast, approximating the distribution of U by the TW distribution (Eq. (1.6)) is quite inaccurate for small and even moderate values of p. In particular, tails probabilities may exhibit relative errors of 100%, even for quite large values of p. As an example, at a false alarm of $\alpha = 1\%$ which gives $s = 0.4776$ in the complex case, with $p = 20$ and $n = 500$ we have $1 - TW_2(s) = 0.01$, whereas $\Pr\left[\frac{U-\mu_{np}}{\sigma_{np}} > s\right] = 0.0054$.

In this paper we elucidate the reason for this observed behavior, and propose a simple correction term to Eq. (1.6) suitable for finite values of p, n and s. First, we show that even though the convergence rate in (1.6) is still $O(p^{-2/3})$, there are two main error terms in the difference

$$\Pr\left[\frac{U-\mu_{np}}{\sigma_{np}} < s\right] - TW_\beta(s).$$

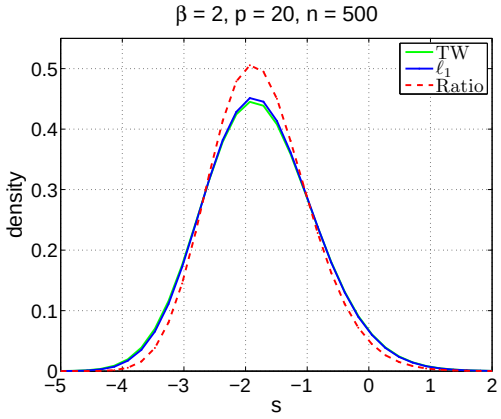

Figure 1: Comparison of Tracy-Widom density with empirical density of largest eigenvalue ℓ_1 and of ratio U, after centering and scaling by μ_{np} and σ_{np} respectively.

The first term is the approximation error already present in Eq. (1.3), namely $\Pr\left[\frac{\ell_1-\mu}{\sigma} < s\right] - TW_\beta(s)$. The second term depends on the second derivative of $TW_\beta(s)$. While asymptotically both terms are $O(p^{-2/3})$, empirically the first term is quite small even for small values of p. In contrast, for the second term we show theoretically that it is non-negligible, in particular for tail probabilities, unless $p \gg 10$.

This second term is the source of the relatively large difference between tail probabilities of U and of ℓ_1. The main result of this paper is the following explicit approximate formula for tail probabilities of U:

$$\Pr\left[\frac{U-\mu_{np}}{\sigma_{np}} > s\right] \approx 1 - TW_\beta(s) + \frac{1}{2}\left(\frac{2}{\beta np}\right)\left(\frac{\mu_{np}}{\sigma_{np}}\right)^2 TW''_\beta(s). \tag{1.7}$$

As shown in the simulation section, compared to the Tracy-Widom approximation, this formula provides a significantly better fit to the empirical density of U, with an error comparable to that of $\Pr\left[\frac{\ell_1-\mu_{np}}{\sigma_{np}} < s\right] - TW_\beta(s)$. We remark that our analysis is valid both for real and for complex valued data, $\beta = 1, 2$. This modified expression for the distribution of U should be useful both for practitioners (to set the threshold for a required false alarm rate) as well as for theoretical purposes. An example of the latter is the performance analysis of various detection tests in signal processing that depend on the distribution of this random variable [15, 16].

2. Distribution of the ratio of largest eigenvalue to the trace

We first introduce the following notation. Let $F_{np}(s)$ be the finite sample distribution function of the largest eigenvalue ℓ_1 properly centered and scaled by

μ_{np} and σ_{np} such that the convergence rate in Eq. (1.3) is $O(p^{-2/3})$. Similarly, let $H_{np}(s)$ denote the distribution function of U, also centered and scaled by the same parameters. That is,

$$F_{np}(s) = \Pr\left[\frac{\ell_1 - \mu_{np}}{\sigma_{np}} < s\right], \quad H_{np}(s) = \Pr\left[\frac{U - \mu_{np}}{\sigma_{np}} < s\right].$$

We also denote the respective densities by $f_{np}(s)$ and $h_{np}(s)$.

For our main result we first need the following lemma.

Lemma 1. *In the joint limit as both $p, n \to \infty$, $p/n = c$, not only does Eq. (1.3) hold, but also both*

$$|F'_{np}(s) - TW'_\beta(s)| \to 0 \qquad \text{and} \qquad |F''_{np}(s) - TW''_\beta(s)| \to 0 \tag{2.1}$$

In the appendix we outline the proof of this lemma for the complex valued case $\beta = 2$. The proof is similar to the proof of convergence of Eq. (1.3), e.g., using the Fredholm determinant representation. We conjecture that the lemma holds also in the real valued case $\beta = 1$. A proof in this case may be considerably more difficult as the distributions are now represented by regularized determinants of operators with matrix kernels, so that even the proof of Eq. (1.3) is much more involved, see [13].

Our main result regarding the distribution of the ratio of the largest eigenvalue to the average trace can be stated as follows:

Theorem 2.1. *Assume that in the joint limit as both $p, n \to \infty$ with $p/n \to c$, the following two conditions hold:*

(i) uniformly in p and s, $H_{np}(s)$ is a smooth function with bounded third derivative, $|H'''_{np}(s)| < C$.

(ii) Eq. (2.1) holds.

Then, in the joint limit as both $p, n \to \infty$,

$$H_{np}(s) - TW_\beta(s) = [F_{np}(s) - TW_\beta(s)] - \frac{1}{2}\left(\frac{2}{\beta np}\right)\left(\frac{\mu_{np}}{\sigma_{np}}\right)^2 TW''_\beta(s) + o(p^{-2/3}). \tag{2.2}$$

As discussed above, condition (ii) indeed holds for $\beta = 2$. We conjecture it holds also for $\beta = 1$. Condition (i), which seems a reasonable assumption, is required for our proof of the Theorem. It remains an open question whether the theorem can be proven without this assumption, or whether this assumption can be proven itself.

Before proving the theorem, let us discuss the two terms on the r.h.s. of Eq. (2.2). The first term is the error in approximating the distribution of the largest eigenvalue ℓ_1 by the TW distribution. While in principle this term is $O(p^{-2/3})$, empirically it has been shown to be very small even for small values of p, n, see [6, 9, 13]. Next, consider the second term, in particular in the context

of right tail probabilities, which are the most relevant for hypothesis testing applications. First, note that as $p, n \to \infty$, with $p/n \to c$

$$\frac{2}{\beta np}\left(\frac{\mu_{np}}{\sigma_{np}}\right)^2 = \frac{2}{\beta}(1+\sqrt{c})^{4/3}\frac{1}{p^{2/3}}(1+o(1)).$$

Hence the second term on the r.h.s. of Eq. (2.2) is also $O(p^{-2/3})$. Regarding the accuracy in approximating right tail probabilities by the TW distribution, the key quantity to analyze is the *relative* size of this second correction term with respect to the leading order term, $1 - TW_\beta(s)$. Since for $s \gg 1$, $1-TW(s) \sim C\exp(-as^{3/2})/s^{3/2}$ for some constants a, C, it follows that $|TW''(s)|/(1-TW(s))$ becomes increasingly large as $s \to \infty$. However, even for small values of s the second term in Eq. (2.2) is non-negligible, unless $p \gg 1$. To see this, consider for example $s = -0.2325$, where $1 - TW_2(s) \approx 5\%$. At this value of s, $|TW_2''(s)|/(1 - TW_2(s)) \approx 7$. Furthermore, for $n \gg p$, we have that $\frac{1}{np}(\mu_{np}/\sigma_{np})^2 \approx 1/p^{2/3}$. Hence, for the second term in Eq. (2.2) to have at most a 10% relative error, namely

$$\frac{1}{2}\frac{2}{\beta np}\left(\frac{\mu_{np}}{\sigma_{np}}\right)^2 \frac{|TW_2''(s)|}{1 - TW_2(s)} \leq 0.1$$

we need $p \gtrsim (35)^{3/2} \approx 200$. To correct for these potentially large relative errors, our proposed approximation for tail probabilities of the ratio is thus

$$\Pr\left[\tfrac{U-\mu_{np}}{\sigma_{np}} > s\right] \approx 1 - TW_\beta(s) + \frac{1}{2}\left(\frac{2}{\beta np}\right)\left(\frac{\mu_{np}}{\sigma_{np}}\right)^2 TW_\beta''(s). \tag{2.3}$$

Simulation results shown in fig. 2 and summarized in the tables below illustrate that Eq. (2.3) indeed provides a much more accurate fit to the empirical distribution of U.

Proof. The proof consists of two main steps. First, we show that the density $f_{np}(s)$ is approximately given by a convolution of $h_{np}(s)$ with a Gaussian. In the second step we approximately invert this deconvolution, showing that the overall error is $o(p^{-2/3})$.

The starting point for our analysis is the well known fact that the two random variables U and T are *independent*, see for example [8], Eq. (5.4), or [2]. Thus, suppressing the dependence on p, n, their joint density can be written as

$$f(u,t) = h(u)g(t).$$

Next, recall that the random variable T follows a $\chi^2_{\beta np}/(\beta np)$ distribution. Hence its density $g(t)$ is known explicitly,

$$g(t) = \beta np\frac{1}{2^{\beta np/2}\Gamma\left(\frac{\beta np}{2}\right)}(\beta npt)^{\beta np/2-1}e^{-\beta npt/2} = C_{\beta,n,p}\frac{e^{-\eta(t-1-\ln t)}}{t}$$

where $\eta = \beta np/2$ and $C_{\beta,n,p} = \eta^\eta e^{-\eta}/\Gamma(\eta)$.

Finally, to relate the distribution of U to that of ℓ_1 we consider the following equality,

$$\Pr[\ell_1 < x] = \int_0^{px} \Pr\left[U < \frac{x}{t}\right] g(t)dt. \tag{2.4}$$

Note that the upper limit of integration is px since by definition $\Pr[U < \frac{1}{p}] = 0$. At this point it is convenient to perform a change of variables $x = \mu_{np} + \sigma_{np}s$. The equation above can then equivalently be written as

$$F_{np}(s) = \Pr\left[\frac{\ell_1 - \mu_{np}}{\sigma_{np}} < s\right] = \int_0^{p(\mu_{np}+s\sigma_{np})} H_{np}\left(\frac{s}{t} + \frac{\mu_{np}}{\sigma_{np}}\frac{1-t}{t}\right) g(t)dt.$$

Furthermore, upon differentiation w.r.t. s, and using the fact that

$$H_{np}\left(\frac{s}{t} + \frac{\mu_{np}}{\sigma_{np}}\frac{1-t}{t}\right)\Bigg|_{t=p(\mu_{np}+s\sigma_{np})} = \Pr\left[U < \frac{1}{p}\right] = 0$$

we obtain a similar equation for the densities,

$$f_{np}(s) = \int_0^{p(\mu_{np}+s\sigma_{np})} h_{np}\left(\frac{s}{t} + \frac{\mu_{np}}{\sigma_{np}}\frac{1-t}{t}\right) g(t)dt.$$

According to assumption (ii), Eq. (2.1), the left hand side converges to $TW'_\beta(s)$. We thus study the behavior of the integral on the r.h.s. as both $p, n \to \infty$. In this limit, $g(t)$ becomes increasingly concentrated around a value of 1 with fluctuations of the order of $\epsilon = 1/\sqrt{\eta}$. Hence, as in Laplace's method for the asymptotic expansion of integrals, we make the change of variables $t = 1 + \epsilon z$, keeping in mind that $\epsilon \to 0$ as $n, p \to \infty$. Note that via a Taylor expansion,

$$\exp(-\eta(t - 1 - \ln t)) = \exp(-\tfrac{z^2}{2}) \exp(-\tfrac{\epsilon}{3}\tfrac{z^3}{(1+\epsilon\theta(z))^3})$$

where $\theta(z) \in [0, z]$. In addition, from the asymptotics of the Gamma function we have that

$$C_{\beta,n,p} = \frac{1}{\sqrt{2\pi\epsilon}}(1 + O(\epsilon^2)).$$

Thus,

$$\begin{aligned} f_{np}(s) \quad = \quad & \frac{1}{\sqrt{2\pi}} \int_{-1/\epsilon}^{(p\mu_{np}+ps\sigma_{np}-1)/\epsilon} h_{np}\left(\frac{s}{1+\epsilon z} - \epsilon\frac{\mu_{np}}{\sigma_{np}}\frac{z}{1+\epsilon z}\right) \exp\left(-\frac{z^2}{2}\right) \times \\ & \frac{1}{1+\epsilon z} \cdot \exp\left(-\frac{\epsilon}{3}\frac{z^3}{(1+\epsilon\theta(z))^3}\right) dz(1 + O(\epsilon^2)) \end{aligned}$$

Note that the lower and upper limits of integration converge to $\pm\infty$ as $\epsilon \to 0$. Furthermore, since the Gaussian function decays exponentially fast, up to exponentially small errors in p, n, we have that

$$f_{np}(s) = \int_{-\infty}^{\infty} h_{np}\left(s - \epsilon\frac{\mu_{np}}{\sigma_{np}}z - \epsilon sz + O\left(\frac{\mu_{np}}{\sigma_{np}}\epsilon^2\right)\right) \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz(1 + O(\epsilon)). \tag{2.5}$$

We remark that up to now, the above algebraic manipulations were nothing but the approximation, for large values of k, of a χ^2_k/k random variable by $1+\sqrt{\frac{2}{k}}Z$ where $Z \sim N(0,1)$.

Next, recall that in the joint limit $p, n \to \infty$, with $p/n \to c$, we have that $\epsilon = O(p^{-1})$ whereas $\epsilon\frac{\mu_{np}}{\sigma_{np}} = O(p^{-1/3})$, so that the latter is the leading order correction term inside the parentheses in Eq. (2.5) above. Denoting $\delta_p = \epsilon\frac{\mu_{np}}{\sigma_{np}}$, up to order $O(p^{-1})$ terms, we have that

$$f_{np}(s) = \int_{-\infty}^{\infty} h_{np}(s - \delta_p z)\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz. \tag{2.6}$$

This equation shows that the density of the largest eigenvalue, $f_{np}(s)$, is approximately the convolution of the required density of the ratio, $h_{np}(s)$, convolved with a Gaussian. The required solution is thus the inverse operation, e.g., a deconvolution.

To perform the deconvolution and obtain an expression for $h_{np}(s)$ and thus for $H_{np}(s)$ it is convenient to work in Fourier space. Since both f_{np} and h_{np} are probability density functions, their Fourier transforms are well defined. We use the following definition of the Fourier transform, $\hat{g}(\omega) = \int g(x)e^{-i\omega x}dx$, and denote by $\hat{tw}(\omega)$ the Fourier transform of the Tracy-Widom density. Conveniently, the convolution operator translates into multiplication in Fourier space, and the Fourier transform of a Gaussian is again a Gaussian. Therefore,

$$\hat{f}_{np}(\omega) = \hat{h}_{np}(\omega)e^{-\delta_p^2\omega^2/2} \tag{2.7}$$

or equivalently $\hat{h}_{np}(\omega) = e^{\delta_p^2\omega^2/2}\hat{f}_{np}(\omega)$. Simple algebraic manipulations give

$$\hat{h}_{np}(\omega) = \left(1 + \tfrac{1}{2}\delta_p^2\omega^2\right)\hat{tw}(\omega) + \left(1 + \tfrac{1}{2}\delta_p^2\omega^2\right)\left(\hat{f}_{np}(\omega) - \hat{tw}(\omega)\right) + \hat{q}(\omega) \tag{2.8}$$

where

$$\begin{aligned}\hat{q}(\omega) &= \left(e^{\delta_p^2\omega^2/2} - 1 - \tfrac{1}{2}\delta_p^2\omega^2\right)\hat{f}_{np}(\omega) \\ &= \left(1 - e^{-\delta_p^2\omega^2/2} - \tfrac{1}{2}\delta_p^2\omega^2 e^{-\delta_p^2\omega^2/2}\right)\hat{h}_{np}(\omega).\end{aligned} \tag{2.9}$$

Taking an inverse Fourier transform of Eq. (2.8) gives that

$$h_{np}(s) = tw(s) - \tfrac{1}{2}\delta_p^2\, tw''(s) \; + (1 - \tfrac{1}{2}\delta_p^2\tfrac{d^2}{ds^2})(f_{np}(s) - tw(s)) + q(s) \tag{2.10}$$

where $q(s)$ is the inverse Fourier Transform of $\hat{q}(\omega)$. Integrating from $-\infty$ to s gives

$$H_{np}(s) = TW(s) - \frac{1}{2}\delta_p^2 TW''(s) + \left[1 - \frac{1}{2}\delta_p^2\frac{d^2}{ds^2}\right](F_{np}(s) - TW(s)) + Q(s), \tag{2.11}$$

where

$$Q(s) = H_{np}(s) - \int_{-\infty}^{\infty} H_{np}(s+\delta_p z)\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz + \frac{1}{2}\delta_p^2 \int_{-\infty}^{\infty} H''_{np}(s+\delta_p z)\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz \tag{2.12}$$

Since $\delta_p^2 = O(p^{-2/3})$, Eq. (2.1), implies that the term $\delta_p^2 \frac{d^2}{ds^2}[F_{np}(s) - TW(s)] = o(p^{-2/3})$ and is thus negligible w.r.t. the term $\delta_p^2 TW''(s)$ in Eq. (2.11). To conclude the proof we thus need to show that the last term $Q(s)$ is also $o(\delta_p^2) = o(p^{-2/3})$. To this end, we expand $H_{np}(s+\delta_p z)$, as well as $H''_{np}(s+\delta_p z)$ in a Taylor series,

$$\begin{aligned} H_{np}(s+\delta_p z) &= H_{np}(s) + \delta_p z H'_{np}(s) + \frac{1}{2}\delta_p^2 z^2 H''_{np}(s) + \frac{1}{6}\delta_p^3 z^3 H'''_{np}(s+\delta_p\theta_1(z)) \\ H''_{np}(s+\delta_p z) &= H''_{np}(s) + \delta_p z H'''_{np}(s+\delta_p\theta_2(z)) \end{aligned}$$

Inserting these expansions into Eq. (2.12) gives that

$$Q(s) = \delta_p^3 \int_{-\infty}^{\infty} \left[\frac{1}{6}z^3 H'''_{np}(s+\delta_p\theta_1(z)) + \frac{1}{2}z H'''_{np}(s+\delta_p\theta_2(z))\right]\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz$$

Finally, using assumption (i) that $|H'''_{np}(s)|$ is bounded it follows that $|Q(s)| \leq C\delta_p^3 = o(p^{-2/3})$. □

Remarks: i) Eq. (2.6) shows that the density of the ratio U is approximately a deconvolution of the density of the largest eigenvalue ℓ_1 with a Gaussian. Since deconvolution is in general an ill-conditioned inverse operation, some a-priori regularity conditions, such as the condition that $|H''_{np}(s)| \leq C$ must be imposed on the solution to prove the Theorem.
ii) The deconvolution Eq. (2.6), or its equivalent in the Fourier space, Eq. (2.7), may also be used to numerically compute the distribution $H_{np}(s)$, given an accurate approximation of $F_{np}(s)$, simply by computing $\hat{f}_{np}(\omega)$ and performing an inverse Fourier transform of Eq. (2.7). Rather than using the TW approximation, the exact distribution F_{np} may be accurately evaluated numerically for finite values of p, n, via its Fredholm determinant representation by the methods developed in [4], for example.
iii) Note that Eq. (2.2) implies that to leading order in p,

$$\mathbb{E}[U] = \mathbb{E}[\ell_1]$$

where the latter may be approximated by $\mu_{n,p} + a_\beta \sigma_{n,p}$, with a_β the mean of a TW-distributed random variable. The reason is that by definition

$$\mathbb{E}[U] = \mu_{np} + \sigma_{np}\int_{-\infty}^{\infty} \Pr\left[\tfrac{U-\mu_{np}}{\sigma_{np}} > s\right] ds$$

and the integral of the correction term containing $TW''(s)$ vanishes, as $TW'(s)$ vanishes as $s \to \pm\infty$.

$\beta = 1$ $p = 10$ $n = 100$	$s(\alpha)$	$\Pr\left[\frac{\ell_1-\mu}{\sigma} > s(\alpha)\right]$	$\Pr\left[\frac{U-\mu}{\sigma} > s\right]$	$\tilde{s}(\alpha)$	$\Pr\left[\frac{U-\mu}{\sigma} > \tilde{s}(\alpha)\right]$
$\alpha = 10\%$	0.4501	0.0969	0.0600	0.1624	0.0937
$\alpha = 5\%$	0.9793	0.0479	0.0243	0.6015	0.0470
$\alpha = 1\%$	2.023	0.0097	0.0031	1.4303	0.0104

Table 1: Results of $5 \cdot 10^6$ simulations for real valued data, $\beta = 1$.

iv) A final remark on universality and non-Gaussianity is in place: As proven in [17, 19], under certain regularity conditions on the underlying distribution, the largest eigenvalue of a sample covariance matrix of non-Gaussian multivariate random variables asymptotically also follows a TW distribution. However, for a finite and relatively small number of dimensions p considered here, the deviations in the distribution of ℓ_1 from the TW distribution may be quite significant. Hence, Eq. (1.3) itself may potentially be not very accurate for tail probabilities, and may be much larger than the difference between the distribution of the two random variables ℓ_1 and U.

3. Simulation Results

In tables 1 and 2 and in figure 2 we present simulation results for the empirical tail probabilities $\Pr\left[\frac{\ell_1-\mu}{\sigma} > s(\alpha)\right]$ and $\Pr\left[\frac{U-\mu}{\sigma} > s(\alpha)\right]$, where $TW(s(\alpha)) = 1-\alpha$, for various values of α, p, n.

We compare the empirical probability of U to the theoretical formula in Eq. (2.3). In addition, we use Eq. (2.3) to find a modified threshold $\tilde{s}(\alpha)$, such that $1 - H_{np}(\tilde{s}(\alpha)) = \alpha$. As seen in the simulations and predicted by our analysis, the TW distribution is a relatively poor approximation for tail probabilities of the random variable U, whereas Eq. (2.3) is far more accurate.

All simulations were performed in Matlab. The TW distribution and density were computed numerically using the RMLab package by Dieng[1]. The derivative of the TW density, $tw'(s)$, was computed numerically via a standard central differencing scheme with $\Delta s = 10^{-3}$. This provided sufficient accuracy for our purposes. We remark that if needed, more accurate numerical methods for evaluating the TW distribution and its derivatives are available, see e.g. [4].

Acknowledgments

The author would like to thank Ofer Zeitouni for interesting discussions, and the two anonymous referees for valuable suggestions which greatly improved the exposition of the paper.

[1] Available at `math.arizona.edu/~momar/assets/docs/RMLab002.zip`

$\beta = 2$ $p = 10$ $n = 100$	$s(\alpha)$	$\Pr\left[\frac{\ell_1 - \mu_{np}}{\sigma_{np}} > s(\alpha)\right]$	$\Pr\left[\frac{U - \mu_{np}}{\sigma_{np}} > s\right]$	$\tilde{s}(\alpha)$	$\Pr\left[\frac{U - \mu_{np}}{\sigma_{np}} > \tilde{s}(\alpha)\right]$
$\alpha = 10\%$	-0.5969	0.0969	0.0602	-0.8042	0.0966
$\alpha = 5\%$	-0.2325	0.0474	0.0237	-0.5087	0.0486
$\alpha = 1\%$	0.4776	0.0092	0.0028	0.0287	0.0113

Table 2: Results of $5 \cdot 10^6$ simulations for complex valued data, $\beta = 2$.

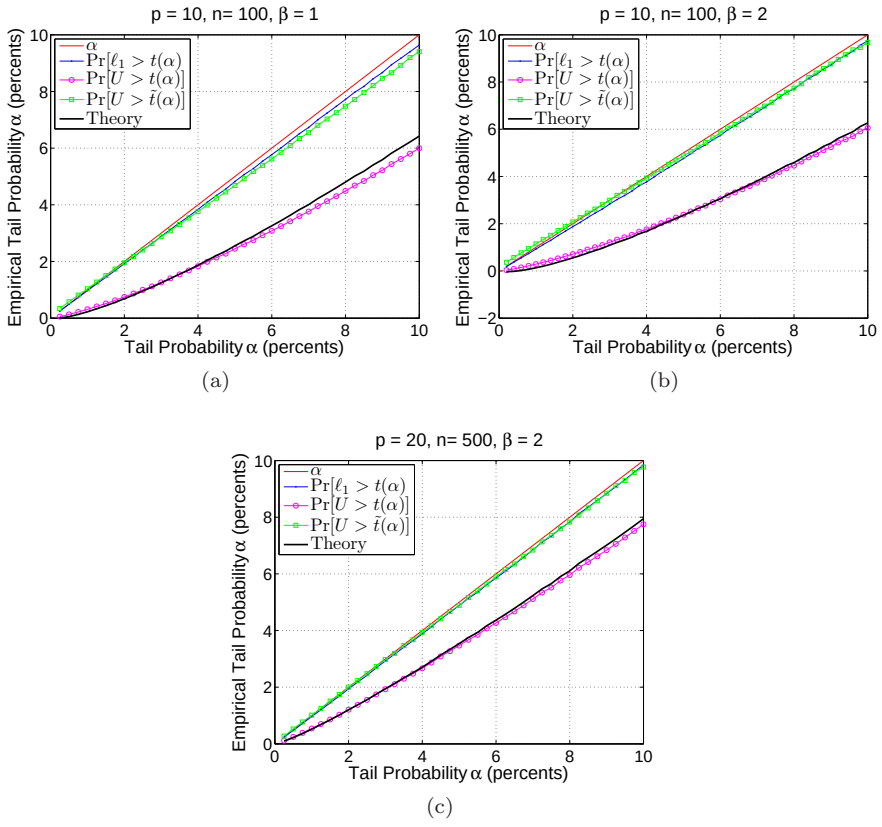

Figure 2: Comparison of empirical tail probabilities for the ratio U vs. theoretical approximation, Eq. (2.3), with $t(\alpha) = \mu_{np} + s(\alpha)\sigma_{np}$ and $\tilde{t}(\alpha) = \mu_{np} + \tilde{s}(\alpha)\sigma_{np}$.

Appendix A. Proof of Lemma

We provide an outline of the proof of the Lemma. For simplicity, we prove only convergence of the density of the largest eigenvalue ℓ_1 to the TW density. The proof for convergence of its derivative is similar. Furthermore, we present the proof only for the complex case $\beta = 2$.

Our analysis follows the notation and proof of convergence of Eq. (1.3) as described in [1]. Let X be an $n \times p$ matrix with i.i.d. complex Gaussian $N(0,1)$ entries, and let $W = 1/nX^*X$ be the sample covariance matrix (a scaled Wishart matrix). Let $I = [s, s']$ be a finite interval. The starting point of the analysis is the Fredholm determinant representation

$$F_{np}(I) = \Pr[\text{no eigenvalues of } W \text{ in } \mu_{np} + \sigma_{np} I] = \det(I - K_p \chi_I) \tag{A.1}$$

where χ_I is the indicator function for the interval I, and K_p is an operator with corresponding kernel

$$K_p(x,y) = \sum_{j=1}^{p} \phi_j(x)\phi_j(y)$$

where ϕ_j are scaled Laguerre polynomials. Eq. (A.1) can also be written as

$$\Pr[\text{no eigenvalues of } W \text{ in } \mu_{n,p} + \sigma_{n,p} I] = \Delta_I(K_p) = \sum_{k=1}^{p} \frac{(-1)^k}{k!} \Delta_{k,I}(K_p) \tag{A.2}$$

where

$$\Delta_{k,I}(K) = \int_I \cdots \int_I \det_{i,j=1}^{k} K(x_i, x_j) \prod_{i=1}^{k} dx_i$$

Using the asymptotics of the Laguerre polynomials, it is possible to show that [9]

$$\lim_{n,p\to\infty} \Pr[\text{no eigenvalues of } W \text{ in } [s,s']] = \det(I - \bar{K}\chi_{[s,s']}) = \Delta_I(\bar{K})$$

where $\bar{K}$ is the limiting operator corresponding to the Airy kernel, and further one may take the limit $s' \to \infty$, thereby proving Eq. (1.3).

We now consider the density of the largest eigenvalue ℓ_1. Taking the derivative with respect to s in Eq. (A.2) gives

$$\frac{\partial}{\partial s}\Delta_{[s,s']}(K_p) = \sum_{j=1}^{p} \frac{(-1)^k}{k!} k \int_s^{s'} \cdots \int_s^{s'} \det_{i,j=1}^{k} K_p(x_i, x_j) \Big|_{x_1 = s} \prod_{i=2}^{k} dx_i$$

We note that each of the terms in the sum above is known as a Fredholm adjugant, see [1].

We now claim that in the joint limit as both $p, n \to \infty$, the finite kernel K_p can be replaced by the limiting kernel $\bar{K}$. To this end, we use Lemma 3.4.2 from [1]: For any two kernels, $F(x,y)$ and $G(x,y)$,

$$\left| \det_{i,j=1}^{k} F(x_i, x_j) - \det_{i,j=1}^{k} G(x_i, x_j) \right| \le k^{1+k/2} \|F - G\|_\infty \max(\|F\|_\infty, \|G\|_\infty)^{k-1}$$

This lemma implies that

$$\left|\frac{\partial}{\partial s}\Delta_{k,[s,s']}(F) - \frac{\partial}{\partial s}\Delta_{k,[s,s']}(G)\right| \leq k^{1+k/2}\max(\|F\|_\infty, \|G\|_\infty)^{k-1}|s'-s|^{k-1}$$

It also implies that for the limiting kernel $\bar{K}$, the order of differentiation and summation in $\frac{\partial}{\partial s}\Delta_{[s,s']}(\bar{K})$ can be switched,

$$\frac{\partial}{\partial s}\Delta_{[s,s']}(\bar{K}) = \sum_k \frac{(-1)^k}{k!}\frac{\partial}{\partial s}\Delta_{k,[s,s']}(\bar{K})$$

as the sum of differentiated terms converges uniformly in s. Hence,

$$\begin{aligned}\left|\frac{\partial}{\partial s}\Delta_{[s,s']}(K_p) - \frac{\partial}{\partial s}\Delta_{[s,s']}(\bar{K})\right| \quad \leq \quad & \left(\sum_{k=1}^{\infty}\frac{k^{1+k/2}}{(k-1)!}\max(\|K_p\|, \|\bar{K}\|)^{k-1}|s'-s|^{k-1}\right)\times \\ & \|K_p - \bar{K}\|_\infty \end{aligned} \tag{A.3}$$

Note that the sum is convergent. Hence, if $\|K_p - \bar{K}\|_\infty \to 0$ then the difference above converges to zero. Indeed, as shown in [9], there is uniform convergence of the kernel K_p to the limiting Airy kernel $\bar{K}$ on compact sets. Furthermore, the bounds

$$|K_p(x,y)| \leq Ce^{-(x+y)}, \quad |\bar{K}(x,y)| \leq Ce^{-(x+y)}$$

for all $x, y > s$ imply that one can take the limit $s' \to \infty$. This yields,

$$\left|\frac{d}{ds}F_{np}(s) - TW_2'(s)\right| \to 0.$$

□.

References

[1] G.W. Anderson, A. Guionnet, O. Zeitouni, *An introduction to random matrices*, Cambridge University Press, 2009.

[2] O. Besson, L.L. Scharf, CFAR matched direction detector, *IEEE Trans. Sig. Proc.*, 54(7):2840–2844, 2006.

[3] P. Bianchi, M. Debbah, M. Maida and J. Najim, Performance of statistical tests for source detection using random matrix theory, *submitted*, 2009.

[4] F. Bornemann, On the numerical evaluation of distributions in random matrix theory, to appear, 2010.

[5] A.W. Davis, On the ratios of the individual latent roots to the trace of a Wishart matrix, *J. Mult. Anal.*, vol. 2, pp. 440–443, 1972.

[6] N. El Karoui, A rate of convergence result for the largest eigenvalue of complex white Wishart matrices. *Ann. Probab.*, 34:2077-2117, 2006.

[7] K. Johansson, Shape fluctuations and random matrices, *Comm. Math. Phys.*, vol. 209, no. 2, pp. 437–476, 2000.

[8] D.E. Johnson, F.A. Graybill, An analysis of a two-way model with interaction and no replication. *J. Amer. Stat. Assoc.,* 67:862-868, 1972.

[9] I.M. Johnstone, On the distribution of the largest eigenvalue in principal component analysis, *Ann. Stat.*, vol. 29, pp. 295–327, 2001.

[10] S. Kritchman, B. Nadler, Non-parametric detection of the number of signals, *IEEE Trans. Sig. Proc.*, vol. 57, pp. 3930–3941, 2009.

[11] W.J. Krzanowski, *Principles of multivariate analysis*, Oxford University Press, 1988.

[12] S. Kuriki, A. Takemura, Tail probabilities of the maxima of multilinear forms and their applications, *Ann. Stat.*, 29(2):328–371, 2001.

[13] Z. Ma, Accuracy of the Tracy-Widom limit for the largest eigenvalue in white Wishart matrices, *submitted*, 2008.

[14] C. Nadal, S.N. Majumdar and M. Vergassola, Statistical distribution of quantum entanglement for a random bipartite state, arXiv:1006.4091, 2010.

[15] B. Nadler, Detection of signals by information theoretic criteria: accurate performance analysis and an improved estimator, *IEEE Trans. Sig. Proc.*, in press, 2010.

[16] B. Nadler, F. Penna, R. Garello, Performance of eigenvalue-based signal detectors with known and unknown noise level, in preparation.

[17] S. Peche, Universality results for the largest eigenvalues of some sample covariance matrix ensembles, *Probab. Theory Relat. Fields* 143:481-516, 2009.

[18] F.J. Schuurmann, P.R. Krishnaiah, A.K. Chattopadhyay, On the distributions of the ratios of the extreme roots to the trace of the Wishart matrix, *J. Mult. Anal.*, 3(4):445–453, 1973.

[19] A. Soshnikov, A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Stat. Phys.* 108:1033-1056, 2002.

[20] K. Zyczkowski, H.J. Sommers, Induced measures in the space of mixed quantum states, *J. Phys. A.* 35:7111–7125, 2001.