Journal of Multivariate Analysis 165 (2018) 132–142

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

On the domain of attraction of a Tracy–Widom law with applications to testing multiple largest roots

Didier Chételat [a], Rajendran Narayanan [b], Martin T. Wells [c,*]

[a] Département de mathématiques et de génie industriel, École Polytechnique de Montréal, Canada
[b] School of Liberal Studies, Azim Premji University, India
[c] Department of Statistical Science, Cornell University, United States

ARTICLE INFO

Article history:
Received 9 November 2015
Available online 13 December 2017

AMS 2010 subject classifications:
60G70
62E20
62H15

Keywords:
Characteristic root
Equality of covariance matrices
Greatest root statistic
Gumbel distribution
MANOVA
Multiple testing
Multiway data
Tracy–Widom laws
Union-intersection test

ABSTRACT

The greatest root statistic arises as a test statistic in several multivariate analysis settings. Suppose there is a global null hypothesis $\mathcal{H}_0$ that consists of m different independent sub null hypotheses, i.e., $\mathcal{H}_0 = \mathcal{H}_{01} \cap \cdots \cap \mathcal{H}_{0m}$, and suppose the greatest root statistic is used as the test statistic for each sub null hypothesis. Such problems may arise when conducting a batch MANOVA or several batches of pairwise testing for equality of covariance matrices. Using the union-intersection testing approach, and by letting the problem dimension $p \to \infty$ faster than $m \to \infty$, we show that $\mathcal{H}_0$ can be tested using a Gumbel distribution to approximate the critical values. Although the theoretical results are asymptotic, simulation studies indicate that the approximation is accurate even for small to moderate dimensions. The results are general and can be applied in any setting where the greatest root statistic is used, not just for the two methods discussed for illustrative purposes.

1. Introduction

Assuming the data generating process is multivariate Gaussian, a statistic for hypothesis testing using the union–intersection approach arising in several multivariate analysis techniques is the largest eigenvalue of the multivariate Beta distribution. More formally, suppose X is an $n_1 \times p$ data matrix with each row being an independent copy of $\mathcal{N}_p(0, \Sigma)$; then, $A = X^\top X \sim \mathcal{W}_p(\Sigma, n_1)$ has a p-dimensional Wishart distribution with n_1 degrees of freedom. Let $B \sim \mathcal{W}_p(\Sigma, n_2)$ be another Wishart distribution with n_2 degrees of freedom independent of A with the same scale matrix Σ. If $n_1 > p$ then A^{-1} exists and the non-zero eigenvalues of the matrix $A^{-1}B$ generalize the univariate F statistic. The scale matrix has no effect on the distribution of these eigenvalues and so without loss of generality we can set $\Sigma = I_p$. The distribution of the random matrix $(A+B)^{-1}B$ is a generalization of the univariate Beta distribution and is called the multivariate Beta distribution or the Jacobi ensemble. The largest eigenvalue θ_p of $(A+B)^{-1}B$, also denoted $\theta(p, n_1, n_2)$, is a random variable called the *greatest root statistic* and since A is positive definite, $\theta_p \in (0, 1)$. We can also obtain θ_p as the largest root of the determinantal equation

$$\det\{B - \theta(A+B)\} = 0.$$

* Corresponding author.
E-mail address: mtw1@cornell.edu (M.T. Wells).

https://doi.org/10.1016/j.jmva.2017.11.008

The greatest root statistic arises as the null hypothesis distribution for the union–intersection test for several classical techniques such as MANOVA, tests for equality of covariance matrices, canonical correlations, and so on; see [16].

We consider the following problem. Suppose there is a global null hypothesis $\mathcal{H}_0$ that consists of m different independent sub null hypotheses, i.e., $\mathcal{H}_0 = \mathcal{H}_{01} \cap \cdots \cap \mathcal{H}_{0m}$. Such hypotheses arise when one is integrating datasets or assessing effects across various treatment levels. Consider a union–intersection type testing approach where the global null hypothesis is true if each and every component sub null hypothesis is true. In such a setting the global null hypothesis would be rejected if the maximum of the test statistics arising from each sub null hypothesis falls in the appropriate rejection region. In particular, suppose the test statistic from each sub null hypothesis is the greatest root statistic, i.e., $\theta_{p,1}, \ldots, \theta_{p,m}$, where for each $k \in \{1, \ldots, m\}$, $\theta_{p,k}$ is the greatest root statistic from the kth component sub null hypothesis. Then the decision rule is to reject the global null hypothesis $\mathcal{H}_0$ if $\max(\theta_{p,1}, \ldots, \theta_{p,m}) > c$ for some appropriately chosen constant c.

We show that the maximum of an iid sequence of the greatest root statistic falls in the Gumbel domain of attraction as $m \to \infty$ and hence the Gumbel distribution can be used to construct a test statistic to do inference for the global null hypothesis. Our approximation relies on two levels of asymptotics. The matrix dimension of each component multivariate Beta distribution goes to infinity, and also the number of sub null hypotheses under consideration goes to infinity; however, we let the matrix dimension go to infinity faster than the number of sub null hypotheses under consideration. In other words, $p \to \infty$ faster than $m \to \infty$ in the sense to be made precise in Section 3.

From a practical perspective, the setting we consider can be used to model array data, in which case the m dimension represents the various faces of the arrays. Multiway data analysis [14] aims to capture multilinear structures in higher-order datasets, where data have more than two modes. Standard two-way methods commonly applied on matrices often fail to find underlying structures in multiway arrays. With an ever increasing number of application areas, multiway data analysis has become popular as an analysis tool. Multiway-array-variate random variables are useful for multiply labeled random variables that can naturally be arranged in array form. Some examples include response from multi-factor experiments, two- or three-dimensional image and video data, multi-channel electroencephalogram data, relationships in social network data, spatial–temporal data, repeated measures data. In this context, the methods of this article can be used to test homogeneity across the faces of the array, and the results are applicable in many multivariate analysis settings where the greatest root statistic plays a role.

In particular, consider the following hypothesis testing framework to conduct pairwise testing of equality of covariance matrices arising from a multivariate normal sample. Let

$$\mathcal{H}_{01} : \Sigma_{11} = \Sigma_{12}, \ldots, \mathcal{H}_{0m} : \Sigma_{m1} = \Sigma_{m2}.$$

Define the global hypothesis $\mathcal{H}_0$ as $\mathcal{H}_0 = \mathcal{H}_{01} \cap \cdots \cap \mathcal{H}_{0m}$. For each $k \in \{1, \ldots, m\}$, let n_{k1}, n_{k2} denote the sample sizes for the kth hypothesis test and let S_{k1}, S_{k2} denote the covariance estimators for the kth hypothesis test. Assuming that the underlying data generating process for each of the m situations is a multivariate normal sample, then under $\mathcal{H}_{0k}$, $S_{k1} \sim \mathcal{W}_p(\Sigma_k, n_{k1})$ and $S_{k2} \sim \mathcal{W}_p(\Sigma_k, n_{k2})$ is independent of S_{k1}, where Σ_k is the common covariance matrix under $\mathcal{H}_{0k}$. Thus the test statistic for $\mathcal{H}_{0k}$ is $\theta_{p,k}$, which is the largest eigenvalue of $(n_{k1}S_{k1} + n_{k2}S_{k2})^{-1}n_{k2}S_{k2}$. Then $\max(\theta_{p,1}, \ldots, \theta_{p,m})$ can be used to test $\mathcal{H}_0$. We will discuss this covariance testing problem in more detail in Section 3.1.

Dumitriu and Koev [6] review the fact that the exact null distribution of the greatest root statistic $\theta(p, n_1, n_2)$ is difficult to compute. Deriving the exact distribution of the largest eigenvalue relies on performing a complicated $(p-1)$-dimensional integral with a Vandermonde term in the integrand. Constantine [4] showed that the marginal distribution of the largest eigenvalue can be expressed in terms of a hypergeometric function with a matrix argument. The cumulative distribution function of the greatest root statistic is

$$\Pr(\theta_{p,1} < x) = C_{1,p} x^{pn_1/2} {}_2F_1\Big(\frac{n_1}{2}, \frac{-n_2+p+1}{2}; \frac{n_1+p+1}{2}; xI\Big),$$

where

$$C_{1,p} = \frac{\Gamma_p^{(1)}\left(\frac{n_1+n_2}{2}\right)\Gamma_p^{(1)}\left(\frac{p+1}{2}\right)}{\Gamma_p^{(1)}\left(\frac{n_1+p+1}{2}\right)\Gamma_p^{(1)}\left(\frac{n_2}{2}\right)}$$

and ${}_2F_1(\cdot, \cdot; \cdot, xI)$ denotes the hypergeometric function with a matrix argument, which in this case is considered to be the identity matrix. Gupta and Richards [8] gave exact Pfaffian expressions for hypergeometric functions with a matrix argument when the arguments are multiples of the identity matrix and also showed that the cumulative distribution function (cdf) of the greatest root statistic can be expressed as a Pfaffian of a skew-symmetric matrix whose entries are double integrals. Koev and Edelman [13] exploited the recursion relations of Jack functions to develop efficient MATLAB implementations to evaluate the hypergeometric functions with a matrix argument. More recently, Butler and Paige [2] provided computational implementations of the theoretical framework advanced by Gupta and Richards [8]. Butler and Paige [2] expressed the double integrals of the Pfaffian in terms of series expansions that are computed using the Maple software. There is an extensive literature on the algorithmic and computational aspects of dealing with the hypergeometric functions with a matrix argument. An elegant treatment on the topic can be found in [6] and the references therein.

Recently, Chiani [3] developed an efficient algorithm for evaluating the cumulative null distribution function of the greatest root statistic. In theory, it could be used to provide an exact test for our problem, but the algorithm relies on the computation of the Pfaffian of a skew-symmetric matrix, a numerically difficult and computationally intensive problem [7,19,25]. Mathematica [26] code at http://sites.google.com/site/marcochianigroup/ demonstrates that the exact method from [3] can be applied to moderately large dimensions. Our scalable approach, based on asymptotic arguments developed in the next sections, circumvents the numerical challenges of the computation of Pfaffians and is both extremely accurate and more straightforward to apply than the exact results in [3] for the specific multiple testing scenario considered in this article.

Moving away from the issue of computational techniques to evaluate the hypergeometric function with a matrix argument, in the remarkable paper of Johnstone [10], it was shown that the greatest root statistic with suitable centering and scaling converges to the now ubiquitous Tracy–Widom distribution [22,23] In particular, Johnstone [10] showed that assuming p is even and that p, $n_1(p)$ and $n_2(p) \to \infty$ together in such a way that

$$\lim_{p\to\infty} \frac{\min(p, n_2)}{n_1+n_2} > 0, \quad \lim_{p\to\infty} \frac{p}{n_1} < 1,$$

then $T_p = \mathrm{logit}(\theta_p) = \ln(\theta_p/1-\theta_p)$ is approximately distributed according to the Tracy–Widom law, i.e., $(T_p - \mu_p)/\sigma_p \rightsquigarrow F_1$, where F_1 is the cdf of the Tracy–Widom distribution arising as a limiting distribution of the largest eigenvalue of Gaussian orthogonal ensembles; here, μ_p and σ_p are centering and scaling factors to make the asymptotics work. We focus on the asymptotics as opposed to exact evaluation of the greatest root statistic owing to the second order rate of convergence $O(p^{-2/3})$ of the greatest root statistic to the Tracy–Widom law. As Johnstone and Ma [12] show, this convergence rate can be guaranteed for appropriate centering and scaling factors. As illustrated by Johnstone [11], the Tracy–Widom approximation is fairly sharp even for small values of p and works quite well for many applied data analysis questions.

Our work aims to bridge two asymptotic regimes of extremes. From classical extreme-value theory we know that the maximum of an iid sequence of random variables converges to one of three distributions depending on whether the random variables are light-tailed, heavy-tailed or have a finite support. For light-tailed random variables it is well known that the maximal domain of attraction is the Gumbel distribution and the Tracy–Widom distribution appears as the limiting distribution of random matrices with light-tailed iid entries. This prompts us to study the asymptotic maximal behavior of iid extremal eigenvalues arising from a sequence of random matrices having light-tailed entries.

2. Tracy–Widom distribution

An important question of theoretical and practical interest is understanding the behavior of the largest eigenvalue of various classes of random matrices. If we consider a diagonal matrix with Gaussian entries, then the largest eigenvalue of such a matrix would converge to the Gumbel distribution as the matrix dimension goes to infinity. This is because the maximal domain of attraction of the Gaussian distribution is the Gumbel distribution. However, when we consider a symmetric matrix with each entry being a real-valued Gaussian random variable or a symmetric Hermitian random matrix with each entry being a complex-valued Gaussian random variable, then the largest eigenvalue converges to the Tracy–Widom distribution. It is indeed a remarkable fact that this distribution arises as the limiting distribution of a large class of random matrices and in fact the limit distribution of the largest eigenvalue has the Tracy–Widom law even if the assumption of iid Gaussian entries of the random matrix is relaxed; see, e.g., [20]. However, as shown in [21], when the matrix entries are heavy-tailed then the joint distribution of the edge eigenvalues converges weakly to the inhomogeneous Poisson random point process.

Let F_1 denote the cdf of the Tracy–Widom distribution arising from the Gaussian orthogonal ensemble (GOE) and let F_2 be the cdf of the Tracy–Widom distribution arising from the Gaussian unitary ensemble (GUE). Then from [22,23] we know that

$$F_2(x) = \exp\left\{-\int_x^\infty (y-x)q^2(y)dy\right\} \tag{1}$$

and

$$F_1(x) = \{F_2(x)\}^{1/2} \exp\left\{-\frac{1}{2}\int_x^\infty q(y)dy\right\}, \tag{2}$$

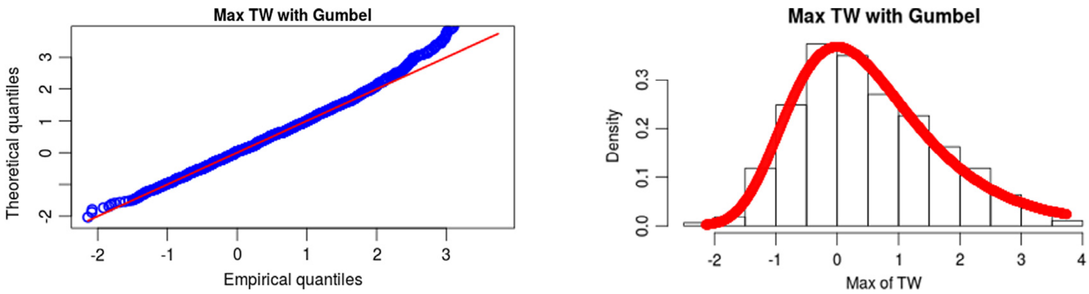
where $q(x)$ is the solution of the classical Painlevé non-linear second order differential equation

$$q''(x) = xq(x) + 2q^3(x), \quad q(x) \sim \mathrm{Ai}(x) \quad \text{as } x \to \infty$$

and $\mathrm{Ai}(x)$ denotes the Airy function. Johnstone [9,10] proved a universality property by showing that the largest eigenvalues of the Wishart matrix and the multivariate Beta matrix both converge to the Tracy–Widom distribution, subject to some growth conditions on the size of the design matrix. Narayanan and Wells [17] showed that the standardized maximum of an iid sequence of random variables having the Tracy–Widom distribution arising from the Gaussian unitary ensembles as in (1) belongs to the Gumbel domain of attraction.

If we take an iid sequence of random variables having the Tracy–Widom (TW) distribution arising from the Gaussian orthogonal ensemble, as in (2), then the maximum of such a sequence asymptotically converges to the Gumbel distribution.

Fig. 1(a) shows a QQ plot of simulated maxima of Tracy–Widom random variables and the standard Gumbel distribution. Fig. 1(b) depicts a histogram of simulated maxima of TW random variables overlaid with a standard Gumbel distribution. A Kolmogorov–Smirnov test to check equality of normalized maximum of iid Tracy–Widom random variables with the Gumbel



(a) QQ plot of the maximum of Tracy–Widoms against a Gumbel.

(b) Histogram of the maximum of Tracy–Widoms against a Gumbel.

Fig. 1. Simulated maximums of TW with Gumbel.

distribution fails to reject the null hypothesis at a p-value of 0.4658. The plots are based on 1000 simulation runs: in each simulation run, we generate an iid sequence, $TW_1, \ldots, TW_k$, of Tracy–Widom random variables of length $k = 10{,}000$ and we store the value of $M = \max(TW_1, \ldots, TW_k)$ from each run. The maxima are then normalized and compared with simulated Gumbel random numbers. The exercise was done on the `R` platform using the `RMTSTAT` package. The location term for normalization is $b_n = U(n)$, where $U(n)$ is the left continuous inverse of $1/(1 - F_1)$. The constant $U(n)$ can be treated as the $100(1 - 1/n)$th quantile of the distribution, which can be easily obtained using the quantile function of the software. The scaling term for normalization is $a_n = 1/nF_1'(b_n)$. Since the cumulative distribution function is continuous, $F_1'(b_n)$ is just the density evaluated at b_n, which can also be easily obtained using RMTSTAT.

3. Main result

For every $p \geq 1$, let $\theta_{1,1}(p, n_1, n_2), \ldots, \theta_{m,1}(p, n_1, n_2)$ denote an iid sequence of largest eigenvalues obtained from an iid sequence of multivariate Beta random matrices $A_{p,1}, \ldots, A_{p,m}$. The meanings of n_1 and n_2 are as in Section 1; we take $n_1 \geq p$. For each $k \in \{1, \ldots, m\}$, let $T_{p,k} = \text{logit}\{\theta_{k,1}(A_{p,k})\}$ be the logit transformed largest eigenvalues. Let also F_1 be the cumulative distribution function of the Tracy–Widom distribution for the real case as in (2), and define the following normalization constants:

$$b_m = F_1^{-1}\left(1 - \frac{1}{m}\right), \quad a_m = \frac{1}{mF_1'(b_m)}. \tag{3}$$

Then from [10] we know that $(T_{p,k} - \mu_{p,n_{1p},n_{2p}})/\sigma_{p,n_{1p},n_{2p}} \rightsquigarrow Z_1 \sim F_1$, where

$$\mu_{p,n_{1p},n_{2p}} = 2\ln\tan\left(\frac{\phi_{p,n_{1p},n_{2p}}}{2} + \frac{\gamma_{p,n_{1p},n_{2p}}}{2}\right), \quad \sigma^3_{p,n_{1p},n_{2p}} = \frac{16/(n_{1p} + n_{2p} - 1)^2}{\sin^2(\phi_{p,n_{1p},n_{2p}} + \gamma_{p,n_{1p},n_{2p}})\sin\phi_{p,n_{1p},n_{2p}}\sin\gamma_{p,n_{1p},n_{2p}}}$$

and

$$\phi_{p,n_{1p},n_{2p}} = 2\arcsin\left(\frac{1}{\sqrt{2}}\sqrt{\frac{2\max(p, n_{1p}) - 1}{n_{1p} + n_{2p} - 1}}\right), \quad \gamma_{p,n_{1p},n_{2p}} = 2\arcsin\left(\frac{1}{\sqrt{2}}\sqrt{\frac{2\min(p, n_{1p}) - 1}{n_{1p} + n_{2p} - 1}}\right).$$

Theorem 1. *Let $p, n_{1p}, n_{2p}, m_p \to \infty$, with $n_{1p} \geq p$, $\lim_{p\to\infty}\min(p, n_{2p})/(n_{1p} + n_{2p}) > 0$ and $\lim_{p\to\infty} m_p/p^{2/3} < \infty$. Let X_k^p denote the centered and scaled value obtained from the logit transform of the greatest root statistic, viz. $X_k^p = (T_{p,k} - \mu_{p,n_{1p},n_{2p}})/\sigma_{p,n_{1p},n_{2p}}$. Then, as $p \to \infty$,*

$$Y^p = \frac{q}{a_{m_p}}\left(\max_{k\in\{1,\ldots,m_p\}} X_k^p - b_{m_p}\right) \rightsquigarrow \mathcal{GU}(0, 1).$$

Before proving the main result, we present a few lemmas. Lemma 1 is an analogue of Theorem 1 of Narayanan and Wells [17].

Lemma 1. *Let $Z_1, Z_2, \ldots$ be a sequence of iid random variables having the Tracy–Widom distribution arising from a Gaussian orthogonal ensemble (GOE) with cumulative distribution function F_1 as given in (2). Let $x^* = \sup\{x \in \mathbb{R} : F_1(x) < 1\}$ denote the right-end point of F_1. Here $x^* = \infty$. Then for $m_p \to \infty$ as $p \to \infty$, we have, as $p \to \infty$,*

$$\frac{1}{a_{m_p}}\left(\max_{k\in\{1,\ldots,m_p\}} Z_k - b_{m_p}\right) \rightsquigarrow \mathcal{GU}(0, 1).$$

Proof. We call on von Mises' condition to show the validity of our claim; see [5,18] for further details. Namely, if

$$L(x) = \lim_{x \to x^*} \frac{\{1 - F_1(x)\}F_1''(x)}{F_1'(x)^2} = -1,$$

then F_1 is in the domain of attraction of the Gumbel distribution. To simplify calculations, we obtain the following from [24]:

$$E(x) = \exp\left\{-\frac{1}{2}\int_x^\infty q(s)ds\right\}, \quad F(x) = \exp\left\{-\frac{1}{2}\int_x^\infty (s-x)q^2(s)ds\right\}.$$

Observe that $F_1(x) = F(x)E(x)$ and $F_2(x) = F^2(x)$. It can also be easily seen that

$$E'(x) = \frac{E(x)}{2}q(x), \quad F'(x) = \frac{F(x)}{2}\int_x^\infty q^2(s)ds.$$

Therefore $F_1'(x) = F_1(x)R_1(x)$, where $R_1(x) = \left\{q(x) + \int_x^\infty q^2(s)ds\right\}/2$. Now,

$$L(x) = \frac{\{1 - F_1(x)\}F_1''(x)}{F_1'(x)^2} = \frac{1 - F_1(x)}{F_1(x)} + \frac{1 - F_1(x)}{F_1(x)}\frac{R_1'(x)}{R_1^2(x)}.$$

We are interested in finding $\lim_{x\to\infty} L(x)$. From Section 1.1.1 of [24] it follows that

$$F_1(x) = 1 - \left(\frac{e^{-\frac{2}{3}x^{3/2}}}{4\sqrt{\pi}x^{3/4}} + \frac{e^{-\frac{4}{3}x^{3/2}}}{32\pi x^{3/2}} - \frac{e^{-2x^{3/2}}}{128\pi^{3/2}x^{9/4}}\right)\{1 + O(x^{-3/2})\}.$$

Unfortunately, there is a typographical error in [24] in their expression for this expansion: there should be $x^{3/4}$ in the denominator. If we take this into account, we can write

$$1 - F_1(x) \sim \frac{e^{-\frac{2}{3}x^{3/2}}}{4\sqrt{\pi}x^{3/4}}. \tag{4}$$

From [1] and Lemma 3 of [17] we get,

$$\frac{1}{2}\left\{q(x) + \int_x^\infty q^2(s)ds\right\} = \frac{e^{-\frac{2}{3}x^{3/2}}}{4\sqrt{\pi}x^{1/4}}\{1 + O(x^{-3/2})\} + \frac{e^{-\frac{4}{3}x^{3/2}}}{16\pi x}\{1 + O(x^{-3/2})\}$$

which yields the following asymptotic expression for $R_1(x)$:

$$R_1(x) \sim \frac{e^{-\frac{2}{3}x^{3/2}}}{4\sqrt{\pi}x^{1/4}}.$$

Again using the asymptotic expansion from [1], we get

$$\frac{1}{2}\{q'(x) - q^2(x)\} = -\frac{x^{1/4}e^{-\frac{2}{3}x^{3/2}}}{4\sqrt{\pi}}\{1 + O(x^{-3/2})\} - \frac{e^{-\frac{4}{3}x^{3/2}}}{8\pi x^{1/2}}\{1 + O(x^{-3/2})\}.$$

This yields

$$R_1'(x) \sim -\frac{x^{1/4}e^{-\frac{2}{3}x^{3/2}}}{4\sqrt{\pi}}.$$

Since $F_1(x)$ is a cdf, $F_1(x) \sim 1$ as $x \to \infty$, and hence

$$\frac{1 - F_1(x)}{F_1(x)}\frac{R_1'(x)}{R_1^2(x)} \sim -1$$

as $x \to \infty$. Thus $L(x) \to -1$ as $x \to \infty$, which establishes that the maximum of an iid sequence from the Tracy–Widom distribution arising from a GOE is in the Gumbel domain of attraction, where $F_{\mathcal{GU}(0,1)}(x) = \exp(-e^{-x})$ for all $x \in \mathbb{R}$. Therefore, for the normalizing constants we defined, we have, as $p \to \infty$,

$$\frac{1}{a_{m_p}}\left(\max_{k\in\{1,\ldots,m_p\}} Z_k - b_{m_p}\right) \rightsquigarrow \mathcal{GU}(0,1)$$

as desired. □

Lemma 2. *We have, as $m \to \infty$,*

$$a_m \sim \left(\frac{4}{3}\right)^{1/2}\left(\frac{3}{4}\right)^{1/6}\ln^{-1/3}\left(\frac{m^2}{12\pi}\right), \quad b_m \sim \left(\frac{3}{4}\right)^{2/3}\ln^{2/3}\left(\frac{m^2}{12\pi}\right).$$

Then for $m \to \infty$ and fixed $y \in \mathbb{R}$, we have $\lim_{m\to\infty} a_m = 0$ and $\lim_{m\to\infty} b_m = \infty$.

Proof. As stated earlier, $b_m = U(m)$, where $U(m)$ is the left continuous inverse of $1/(1-F_1)$. Thus using (4) we can write

$$m = \frac{1}{1-F_1(b_m)} \sim h(b_m) = 4\sqrt{\pi}\, b_m^{3/4} \exp(2b_m^{3/2}/3c).$$

Let $g(x) = \ln h(e^x) = \ln(4\sqrt{\pi}) + \frac{3}{4}x + \frac{2}{3}\exp(3x/2)$. For any $y \in \mathbb{R}$,

$$\left|\frac{dg^{-1}(y)}{dy}\right| = \left|\frac{1}{g'\{g^{-1}(y)\}}\right| = \frac{1}{3/4 + \exp\{3g^{-1}(y)/2\}} \leq \frac{4}{3}.$$

Then, by the Mean Value Theorem,

$$|\ln h^{-1}(m) - \ln b_m| = |g^{-1}(\ln m) - g^{-1} \circ g(\ln b_m)| = |g^{-1}(\ln m) - g^{-1}\{\ln h(b_m)\}| \leq \frac{4}{3}|\ln(m) - \ln h(b_m)| \xrightarrow{x\to\infty} 0.$$

Therefore,

$$b_m \sim h^{-1}(m) = \left\{\frac{3}{4} W\left(\frac{m^2}{12\pi}\right)\right\}^{2/3} \sim \left\{\frac{3}{4}\ln\left(\frac{m^2}{12\pi}\right)\right\}^{2/3},$$

where W is the Lambert W function. Second, defining $d_m = \exp(b_m^{3/2})$, we also have

$$m = \frac{1}{1-F_1(\ln^{2/3} d_m)} \sim \tilde{h}(d_m) = 4\sqrt{\pi}\ln^{1/2}(d_m) d_m^{2/3}.$$

Now for $\tilde{g}(x) = \ln \tilde{h}(e^x) = \ln(4\sqrt{\pi}) + \frac{1}{2}\ln x + \frac{2}{3}x$, which is invertible as a function $[0,\infty) \to \mathbb{R}$,

$$\left|\frac{dg^{-1}(y)}{dy}\right| = \left|\frac{1}{g'\{g^{-1}(y)\}}\right| = \frac{1}{\frac{1}{2g^{-1}(y)} + \frac{2}{3}} \leq \frac{3}{2}$$

for any $y \in \mathbb{R}$. Then, by the Mean Value Theorem,

$$|\ln \tilde{h}^{-1}(m) - \ln d_m| = |g^{-1}(\ln m) - g^{-1} \circ g(\ln d_m)| = |g^{-1}(\ln m) - g^{-1}\{\ln \tilde{h}(d_m)\}| \leq 3\,|\ln(m) - \ln \tilde{h}(d_m)|/2 \xrightarrow{m\to\infty} 0.$$

So

$$\exp(b_m^{3/2}) = d_m \sim \tilde{h}^{-1}(m) = \exp\left\{\frac{3}{4}W\left(\frac{m^2}{12\pi}\right)\right\} \sim \left\{\frac{m^2/12\pi}{\ln(m^2/12\pi)}\right\}^{3/4}.$$

Therefore,

$$a_m = \frac{1}{mF_1'(b_m)} = \frac{1}{mF_1(b_m)R_1(b_m)} \sim \frac{4\sqrt{\pi}\, b_m^{1/4}\exp\left(\frac{2}{3}b_m^{3/2}\right)}{m} \sim \frac{4\sqrt{\pi}\left\{\frac{3}{4}\ln\left(\frac{m^2}{12\pi}\right)\right\}^{1/6}\left\{\frac{m^2/12\pi}{\ln(m^2/12\pi)}\right\}^{1/2}}{m}$$
$$\sim \frac{\left(\frac{4}{3}\right)^{1/2}\left(\frac{3}{4}\right)^{1/6}}{\ln^{1/3}(m^2/12\pi)},$$

as desired. □

We now prove Theorem 1.

Proof. We use Theorem 1 from [10]. The conditions required are that

$$\lim_{p\to\infty}\frac{\min(p, n_{2p})}{n_{1p}+n_{2p}} > 0, \quad \lim_{p\to\infty}\frac{p}{n_{2p}} < 1,$$

which are satisfied by the assumptions of the theorem. Then $\mu_{p,n_{1p},n_{2p}}$ and $\sigma_{n,p}$ are defined as in Eq. (5) on p. 2641, and so, under the null hypothesis, by Theorem 1 in [10] with $s_0 = 0$ there must be a constant $C > 0$ such that

$$|\Pr(X_k^p \leq x) - \Pr(Z_k \leq x)| \leq \frac{C}{p^{2/3}} e^{-x/2}$$

for all $x \geq 0$. Second, for any fixed $y \in \mathbb{R}$, $\lim_{p\to\infty} a_{m_p}y + b_{m_p} = \infty$ by Lemma 2, so there is some $P(y) > 0$ such that for all $p \geq P(y)$, $a_{m_p}y + b_{m_p} > 0$. Then, for $Z_1, Z_2, \ldots$ a sequence of independent real Tracy–Widom random variables, we have for all $p \geq P(y)$,

$$\begin{aligned}
|\Pr(Y^p \leq y) - \exp\left(-e^{-y}\right)| &\leq \left|\Pr\left(\max_{k\in\{1,\ldots,m_p\}} X_k^p \leq a_{m_p}y + b_{m_p}\right) - \Pr\left(\max_{k\in\{1,\ldots,m_p\}} Z_k \leq a_{m_p}y + b_{m_p}\right)\right| \\
&\qquad + \left|\Pr\left(\max_{k\in\{1,\ldots,m_p\}} Z_k \leq a_{m_p}y + b_{m_p}\right) - \exp\left(-e^{-y}\right)\right| \\
&\leq \sum_{k=1}^{m_p} \left|\prod_{\ell=1}^{k-1} \Pr(Z_\ell \leq a_{m_p}y + b_{m_p}) \prod_{\ell=k}^{m_p} \Pr(X_\ell^p \leq a_{m_p}y + b_{m_p}) \right. \\
&\qquad \left. - \prod_{\ell=1}^{k} \Pr(Z_\ell \leq a_{m_p}y + b_{m_p}) \prod_{\ell=k+1}^{m_p} \Pr(X_\ell^p \leq a_{m_p}y + b_{m_p})\right| \\
&\qquad + \left|\Pr\left(\max_{k\in\{1,\ldots,m_p\}} Z_k \leq a_{m_p}y + b_{m_p}\right) - \exp\left(-e^{-y}\right)\right| \\
&\leq m_p \left|\Pr(X_1^p \leq a_{m_p}y + b_{m_p}) - \Pr(Z_1 \leq a_{m_p}y + b_{m_p})\right| \\
&\qquad + \left|\Pr\left(\max_{k\in\{1,\ldots,m_p\}} Z_k \leq a_{m_p}y + b_{m_p}\right) - \exp\left(-e^{-y}\right)\right| \\
&\leq C\frac{m_p}{p^{2/3}} e^{-\frac{1}{2}(a_{m_p}y + b_{m_p})} + \left|\Pr\left(\max_{k\in\{1,\ldots,m_p\}} Z_k \leq a_{m_p}y + b_{m_p}\right) - \exp\left(-e^{-y}\right)\right|.
\end{aligned}$$

Thus since $\lim_{p\to\infty} m_p/p^{2/3} < \infty$ and $\lim_{p\to\infty} a_{m_p}y + b_{m_p} = \infty$, using Lemma 1 we get

$$\lim_{p\to\infty} |\Pr(Y^p \leq y) - \exp\left(-e^{-y}\right)| \leq \lim_{p\to\infty} \left|\Pr\left(\max_{k\in\{1,\ldots,m_p\}} Z_k \leq a_{m_p}y + b_{m_p}\right) - \exp\left(-e^{-y}\right)\right| \leq 0.$$

Since this is true for any $y \in \mathbb{R}$, the result follows. □

3.1. *Approximate α level test*

As a motivating example from multivariate analysis, consider the following hypothesis testing framework to conduct pairwise testing of equality of covariance matrices arising from a multivariate normal sample. Let

$$\mathcal{H}_{01} : \Sigma_{11} = \Sigma_{12}, \ldots, \mathcal{H}_{0m} : \Sigma_{m1} = \Sigma_{m2}.$$

Define the global hypothesis $\mathcal{H}_0$ as $\mathcal{H}_0 = \mathcal{H}_{01} \cap \cdots \cap \mathcal{H}_{0m}$. This implies that $\mathcal{H}_0$ is true if each of the component hypotheses $\mathcal{H}_{0k}$ is true. Thus, we *accept* $\mathcal{H}_0$ when every component hypothesis $\mathcal{H}_{0k}$ is accepted. We can equivalently say that we reject $\mathcal{H}_0$ if any component hypothesis $\mathcal{H}_{0k}$ is rejected.

Let R_k denote the rejection region corresponding to the kth hypothesis test, so that $R = R_1 \cup \cdots \cup R_m$ is the rejection region corresponding to $\mathcal{H}_0$. For each $k \in \{1, \ldots, m\}$, let n_{k1}, n_{k2} and S_{k1}, S_{k2} denote the sample sizes and covariance estimators, respectively, for the kth hypothesis test. By construction, S_{k1} and S_{k2} will be independent. If we further assume that each of the m samples follows a multivariate normal distribution, then under $\mathcal{H}_{0k}$ we would have $S_{k1} \sim \mathcal{W}_p(\Sigma_k, n_{k1})$ and $S_{k2} \sim \mathcal{W}_p(\Sigma_k, n_{k2})$, where Σ_k is the common covariance matrix under $\mathcal{H}_{0k}$. Thus the test statistic for $\mathcal{H}_{0k}$ is $\theta_{p,k}$, which is the largest eigenvalue of $(n_{k1}S_{k1} + n_{k2}S_{k2})^{-1}n_{k2}S_{k2}$. Then $\max(\theta_{p,1}, \ldots, \theta_{p,m}) \rightsquigarrow \mathcal{G}U_p$ as $m \to \infty$, where $\mathcal{G}U_p$ denotes the cdf of a univariate Gumbel distribution, where we explicitly write the dependence on the dimension p.

Using this, we can construct an approximate, high-dimensional α-level test for $\mathcal{H}_0 : \forall_{k\in\{1,\ldots,m\}} \; \Sigma_{1k} = \Sigma_{2k}$ using the union–intersection approach. Indeed, we could reject $\mathcal{H}_0$ when $\max(\theta_{p,1}, \ldots, \theta_{p,m}) > c_\alpha$, where

$$c_\alpha = [1 + \exp[\sigma_{p,n_{1p},n_{2p}} a_{m_p} \ln\{-\ln(1-\alpha)\} - \sigma_{p,n_{1p},n_{2p}} b_{m_p} - \mu_{p,n_{1p},n_{2p}}]]^{-1}.$$

This would an approximate α-level test because for $p_n/n \to (0, \infty)$ and $m_p/p^{2/3} \to (0, \infty)$, $\Pr(\text{Reject } \mathcal{H}_0 \mid \mathcal{H}_0) \to \alpha$ as $p \to \infty$. To see this, note that in the notation of Theorem 1,

$$\begin{aligned}
\Pr(\text{Reject } \mathcal{H}_0 \mid \mathcal{H}_0) &= \Pr\left(\max_{k\in\{1,\ldots,m_p\}} \theta_{p,1} > c_\alpha \,\middle|\, \mathcal{H}_0\right) \\
&= \Pr\left[\frac{\max_{k\in\{1,\ldots,m_p\}} \operatorname{logit}\theta_{p,1}\{(n_{k1}S_{k1} + n_{k2}S_{k2})^{-1}S_{k2}\} - \mu_{p,n_{1p},n_{2p}}}{\sigma_{p,n_{1p},n_{2p}}} > \frac{\operatorname{logit} c_\alpha - \mu_{p,n_{1p},n_{2p}}}{\sigma_{p,n_{1p},n_{2p}}} \,\middle|\, \mathcal{H}_0\right]
\end{aligned}$$

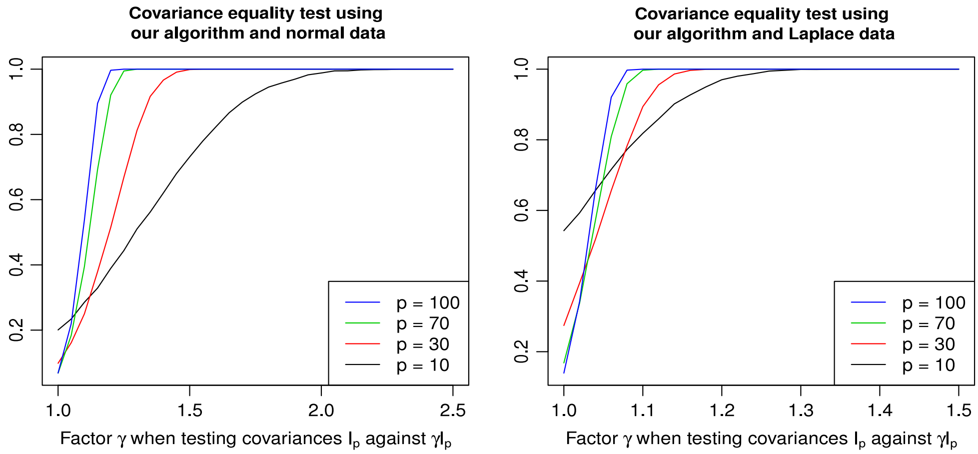


Fig. 2. Power curves for simultaneous covariance equality tests.

$$= \Pr\left(\max_{k\in\{1,\ldots,m_p\}} X_k^p > \frac{\operatorname{logit} c_\alpha - \mu_{p,n_{1p},n_{2p}}}{\sigma_{p,n_{1p},n_{2p}}} \,\middle|\, \mathcal{H}_0\right)$$
$$= \Pr\left(Y^p > \frac{\operatorname{logit} c_\alpha - \mu_{p,n_{1p},n_{2p}} - \sigma_{p,n_{1p},n_{2p}} b_{m_p}}{\sigma_{p,n_{1p},n_{2p}} a_{m_p}} \,\middle|\, \mathcal{H}_0\right),$$

so according to this same theorem, one gets the desired conclusion, i.e.,

$$\lim_{p\to\infty} \Pr(\text{Reject } \mathcal{H}_0 \mid \mathcal{H}_0) = 1 - \exp\left\{\exp\left(-\frac{\operatorname{logit} c_\alpha - \mu_{p,n_{1p},n_{2p}} - \sigma_{p,n_{1p},n_{2p}} b_{m_p}}{\sigma_{p,n_{1p},n_{2p}} a_{m_p}}\right)\right\} = \alpha.$$

4. Simulations

To explore the finite (m, n, p) behavior of our theoretical domain of attraction results, we carry out two numerical studies in this section. We consider two different large-scale inferential problems: pairwise testing for equality of covariance matrices and multivariate analysis of variance. In each simulation setting, we compute the power curves for different dimensions over one-dimensional spaces of alternatives.

4.1. Equality of covariance matrices

The theory behind this test was discussed in Section 3.1. We have m independent population pairs. For the kth population pair, with $k \in \{1, \ldots, m\}$, let k_1 be the index of the first population in the kth pair and k_2 be the index of the second population in the kth pair. Let n_{k1} and n_{k2} be the sample sizes of the first and the second population in the kth pair. Let Σ_{k1}, Σ_{k2} be the corresponding covariance matrices for the kth pair.

We simulated two independent p-dimensional multivariate datasets that form the two design matrices of dimensions $n_{k1} \times p$ and $n_{k2} \times p$, respectively. The test statistic to test the kth null hypothesis is the largest eigenvalue $\theta_{p,k}$ of $(n_{k1}S_{k1} + n_{k2}S_{k2})^{-1} n_{k2}S_{k2}$, where S_{k1} and S_{k2} are the sample covariance matrix analogues of Σ_{k1} and Σ_{k2}, respectively. With generated the data from a multivariate normal distribution, and to study deviations form normality, we performed the experiment again with data from a matrix with iid Laplace entries.

We considered two different regimes for generating covariance matrix pairs that need to be tested for equality. In the first regime, for each $k \in \{1, \ldots, m\}$ we set $\Sigma_{k1} = I_p$ and $\Sigma_{k2} = \gamma I_p$, where I_p denotes the p-dimensional identity matrix and $\gamma \in [1, 2.5]$ is a non-negative scalar giving rise to a one-parameter family of alternatives. We then performed 8000 simulations to test for simultaneous equality of $m = 500$ covariance matrix pairs for each value of γ in the grid. We repeated the exercise for matrix dimensions ranging from $p = 10$ to 100, while the sample sizes for each pair were chosen as $n_1 = n_2 = 2p$. We then computed the resulting approximations to the true power curves. The results for $p \in \{10, 30, 70, 100\}$ are plotted in Fig. 2.

Note that when $\gamma = 1$, both the null and the alternative hypothesis represent the identity matrix. As can be seen from Fig. 2, our approximation does very well in detecting departures from the null hypothesis. The power curve approaches 1 very quickly and gets much sharper even for a moderate values of p and mild increase of γ from 1. This supplements the theoretical asymptotic results rather well. Note that as our test relies on high-dimensional asymptotics, we only obtain a 0.05 significance level when p gets large enough. From the simulations, we see that $p = 30$ in the normal case and $p = 70$ in the Laplace case seem sufficient to achieve the desired significance level.

4.2. Manova

Our second set-up involved m independent batches. Within each batch, we had r different groups each of which contained n iid samples from a p-dimensional distribution. Between groups of the same batch, we had equal covariances but potentially unequal means. Let

$$\underbrace{\begin{array}{c} Y_{111},\ldots,Y_{11n}\sim\mathcal{N}_p(\mu_{11},\Sigma_1),\\ \vdots \\ Y_{1r1},\ldots,Y_{1rn}\sim\mathcal{N}_p(\mu_{11},\Sigma_1),\end{array}}_{\text{Batch } 1} \quad\cdots\quad \underbrace{\begin{array}{c} Y_{m11},\ldots,Y_{m1n}\sim\mathcal{N}_p(\mu_{1r},\Sigma_m),\\ \vdots \\ Y_{mr1},\ldots,Y_{mrn}\sim\mathcal{N}_p(\mu_{1r},\Sigma_m).\end{array}}_{\text{Batch } m}$$

We wanted to test the global null hypothesis of equality of group means across independent batches, viz.

$$\mathcal{H}_0:\begin{array}{c}\mu_{11}=\cdots=\mu_{1r},\\ \vdots\\ \mu_{m1}=\cdots=\mu_{mr}.\end{array}$$

It is to be emphasized that each row in the above null hypothesis expression is a p-dimensional vector. For each batch $k\in\{1,\ldots,m\}$, we computed the matrices

$$A_k=\sum_{\ell=1}^{r}\sum_{i=1}^{n}(Y_{k\ell i}-\bar{Y}_{k\ell})(Y_{k\ell i}-\bar{Y}_{k\ell})^\top,\quad B_k=n\sum_{\ell=1}^{r}(\bar{Y}_{k\ell}-\bar{Y}_k)(\bar{Y}_{k\ell}-\bar{Y}_k)^\top,$$

where $\bar{Y}_{k\ell}=(Y_{k\ell 1}+\cdots+Y_{k\ell n})/n$ and $\bar{Y}_k=(\bar{Y}_{k1}+\cdots+\bar{Y}_{kr})/r$. That is, for the kth batch, A_k was the *within group* covariance matrix and B_k was the *between group* covariance matrix. Under the null hypothesis, we had $A_k\sim\mathcal{W}_p[r(n-1),\Sigma_k]$ independent of $B_k\sim\mathcal{W}_p(r-1,\Sigma_k)$ so that

$$\begin{array}{c}\theta_1=\lambda_1\{(A_1+B_1)^{-1}B_1\}\sim\theta_{1,1}\big(p,r(n-1),r-1\big),\\ \vdots\\ \theta_m=\lambda_1\{(A_m+B_m)^{-1}B_m\}\sim\theta_{m,1}\big(p,r(n-1),r-1\big),\end{array}$$

where p refers to the dimension, $r(n-1)$ refers to the "error" degrees of freedom and $r-1$ is the "hypothesis" degrees of freedom for each batch. Furthermore, $\theta_1,\ldots,\theta_m$ were independent because the batches were independent. Consider the following argument: write $n_1=r(n-1)$ and $n_2=r-1$, and suppose that for fixed $n,p,r,m\to\infty$ with $\lim_{p\to\infty}m/p^{2/3}<\infty$ and $\lim_{p\to\infty}p/r>0$. Then n_1 and $n_2\to\infty$ and

$$\lim_{p\to\infty}\frac{\min(p,n_2)}{n_1+n_2}=\lim_{p\to\infty}\frac{\min(p/r,1-1/r)}{n-1/r}=\frac{1}{n}\min\Big(\lim_{p\to\infty}p/r,1\Big)>0.$$

Then, according to Theorem 1, we would find that

$$Z=\frac{\max\limits_{k\in\{1,\ldots,m\}}\operatorname{logit}\big(\theta_k\big)-\mu_{p,r-1,r(n-1)}-b_m\sigma_{p,r-1,r(n-1)}}{a_m\sigma_{p,r-1,r(n-1)}}\rightsquigarrow\mathcal{GU}(0,1),$$

where a_m, b_m are defined as in Eq. (3). Hence, an approximate α-test for testing $\mathcal{H}_0$ could be given by rejecting when $Z>F^{-1}_{\mathcal{GU}(0,1)}(1-\alpha)$. As an aside, in some situations it could be convenient to work with the following reparametrization outlined in [15]:

$$\theta_{k,1}[p,r(n-1),r-1]\stackrel{d}{=}\theta_{k,1}[r-1,r(n-1)+r-1-p,p]\stackrel{d}{=}\theta_{k,1}(r-1,rn-1-p,p).$$

It can be easily shown that the asymptotic regime and hence the simulation results are invariant under the above reparametrization. Now in order to generate the power curves for our hypothesis testing framework, we tested against the one-parameter family of alternatives

$$\mathcal{H}_1(\gamma):\begin{array}{c}(\mu_{11},\ldots,\mu_{1r})=\left(\begin{bmatrix}1^\gamma\\ \vdots\\ 1^\gamma\end{bmatrix},\begin{bmatrix}2^\gamma\\ \vdots\\ 2^\gamma\end{bmatrix},\ldots,\begin{bmatrix}r^\gamma\\ \vdots\\ r^\gamma\end{bmatrix}\right),\\ \cdots\\ (\mu_{m1},\ldots,\mu_{mr})=\left(\begin{bmatrix}1^\gamma\\ \vdots\\ 1^\gamma\end{bmatrix},\begin{bmatrix}2^\gamma\\ \vdots\\ 2^\gamma\end{bmatrix},\ldots,\begin{bmatrix}r^\gamma\\ \vdots\\ r^\gamma\end{bmatrix}\right),\end{array}$$

for $\gamma\in[0,1]$. We performed 8000 simulation runs for each p,r,γ combination, both for normal data and for data from a Laplace distribution. This was done for p ranging from 10 to 100, and for each such choice of p we set $r=2p$ and $m=\lfloor(p)^{2/3}\rfloor$.

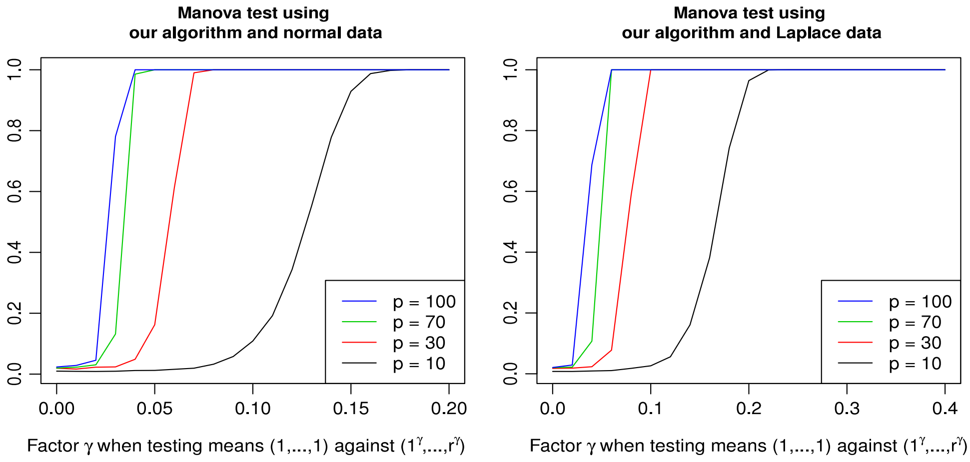


Fig. 3. Power curves for simultaneous MANOVA tests.

We then computed approximations to the true power curves based on these simulations. The results for $p \in \{10, 30, 70, 100\}$ are plotted in Fig. 3.

Note that when $\gamma = 0$, the alternate and the null hypothesis coincided. Just like in the first simulation setting, the results are good even for small to moderate values of p and for very mild departures from the null hypothesis, as evidenced by small positive values of γ yielding power close to 1. Further, as expected, the power curves even steeper as the problem dimension increases from $p = 10$ to $p = 100$. This is in agreement with our theoretical findings.

5. Discussion

The greatest root statistic arises as the test statistic in several multivariate statistical analysis settings. We explored the problem of several independent multivariate analysis testing problems when each hypothesis instance is the greatest root statistic. It is not difficult to fathom casting batch MANOVA or batch pairwise testing for equality of covariance matrices in our hypothesis testing framework. In this article, we prove that the maximal domain of attraction of an iid sequence of greatest root statistics arising out of such batch testing settings is the Gumbel distribution. We present the efficacy of the asymptotic results through two canonical multivariate analysis techniques.

In simulations, we showed that our method is somewhat robust to deviations from normality, by obtaining similar results under a Laplace distribution. We also attempted simulations with a heavy-tailed distribution, namely the Student t distribution, but the tests failed to produce the right significance level. Thus it is reasonable to believe that our method is robust to deviations from normality as long as the tails are not too heavy.

Acknowledgments

The first author's research was partially supported by NSF grant CCF-0808864. The third author's research was partially supported by grants NSFDMS-1208488, NSF-DMS 1611893, and NIH U19 AI111143.

References

[1] A.P. Bassom, P.A. Clarkson, C. Law, J.B. McLeod, Application of uniform asymptotics to the second Painlevé transcendent, Arch. Ration. Mech. Anal. 143 (1998) 241–271.
[2] R.W. Butler, R.L. Paige, Exact distributional computations for Roy's statistic and the largest eigenvalue of a Wishart distribution, Statist. Comput. 21 (2011) 147–157.
[3] M. Chiani, Distribution of the largest root of a matrix for Roy's test in multivariate analysis of variance, J. Multivariate Anal. 143 (2016) 467–471.
[4] A. Constantine, Some non-central distribution problems arising in multivariate analysis, Ann. Math. Statist. 34 (1963) 1270–1285.
[5] L. de Haan, A. Ferreira, Extreme Value Theory: An Introduction, Springer Verlag, Berlin, 2006.
[6] I. Dumitriu, P. Koev, Distributions of the extreme eigenvalues of Beta–Jacobi random matrices, SIAM J. Matrix Anal. 30 (2008) 1–6.
[7] C. González-Ballestero, L. Robledo, G.F. Bertsch, Numeric and symbolic evaluation of the Pfaffian of general skew-symmetric matrices, Comput. Phys. Comm. 182 (2011) 2213–2218.
[8] R. Gupta, D. Richards, Hypergeometric functions of scalar matrix argument are expressible in terms of classical hypergeometric functions, SIAM J. Math. Anal. 16 (1985) 852–858.
[9] I. Johnstone, On the distribution of the largest eigenvalue in principal component analysis, Ann. Statist. 29 (2001) 295–327.
[10] I. Johnstone, Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence, Ann. Statist. 36 (2008) 2638–2716.
[11] I. Johnstone, Approximate null distribution of the largest root in multivariate analysis, Ann. Appl. Stat. 3 (2009) 1616–1633.

[12] I. Johnstone, Z. Ma, Fast approach to the Tracy–Widom law at the edge of GOE and GUE, Ann. Appl. Probab. 22 (2012) 1962–1988.
[13] P. Koev, A. Edelman, The efficient evaluation of the hypergeometric function of a matrix argument, Math. Comp. 75 (2006) 833–846.
[14] P.M. Kroonenberg, Applied Multiway Data Analysis, Wiley, Hoboken, NJ, 2008.
[15] K. Mardia, J.T. Kent, J. Bibby, Multivariate Analysis, Academic Press, New York, 1979.
[16] R.J. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.
[17] R. Narayanan, M.T. Wells, On the maximal domain of attraction of Tracy–Widom distribution for Gaussian unitary ensembles, Statist. Probab. Lett. 83 (2013) 2364–2371.
[18] S.I. Resnick, Extreme Values, Extreme Values Regular Variation and Point Processes, Springer, New York, 2008.
[19] J. Rubow, U. Wolff, A factorization algorithm to compute Pfaffians, Comput. Phys. Comm. 182 (2011) 2530–2532.
[20] A. Soshnikov, A note on the universality of the fistribution of the largest eigenvalue in certain sample covariance matrices, J. Stat. Phys. 108 (2002) 1033–1056.
[21] A. Soshnikov, Poisson statistics for the largest eigenvalues in random matrix ensembles, in: J. Asch, A. Joye (Eds.), Mathematical Physics of Quantum Mechanics, Springer Berlin Heidelberg, 2006, pp. 351–364.
[22] C. Tracy, H. Widom, Level spacing distributions and the Airy kernel, Comm. Math. Phys. 159 (1994) 151–174.
[23] C. Tracy, H. Widom, On orthogonal and symplectic matrix ensembles, Comm. Math. Phys. 177 (1996) 727–754.
[24] C. Tracy, H. Widom, The distributions of random matrix theory and their applications, Stanford Institute for Theoretical Economics, 2008.
[25] M. Wimmer, Algorithm 923: Efficient numerical computation of the Pfaffian for dense and banded skew-symmetric matrices, ACM Trans. Math. Software (TOMS) 38 (2012) 30.
[26] Wolfram Research Inc., Mathematica, Version 11.2, Champaign, IL, 2017.