Exact and approximate computation of critical values of largest root test in high dimension

Gregory Ang Tai Xiang, Zhidong Bai, Kwok Pui Choi, Yasunori Fujikoshi and Jiang Hu

June 16, 2020

Abstract
The difficulty to efficiently compute the null distribution of the largest eigenvalue of a MANOVA matrix has hindered the wider applicability of Roy's the largest root test (RLRT) though it was proposed over six decades ago. Recent progress made by Johnstone (2009), Butler and Paige (2011) and Chiani (2016) has greatly simplified the approximate and exact computation of the critical values of RLRT. When datasets are high dimensional (HD), Chiani's numerical algorithm of exact computation may not give reliable results due to truncation error, and Johnstone's approximation method via Tracy-Widom distribution is likely to give good approximation. In this paper, we conduct comparative studies to study in which region the exact method gives reliable numerical values, and in which region Johnstone's method gives good quality approximation. We formulate recommendations to inform practitioners of RLRT. We also conduct simulation studies in high dimensional setting to examine the robustness of RLRT against normality assumption in populations. Our study provides support of RLRT robustness against non-normality in HD.

Keywords Roy's largest root test · MANOVA · critical values · high dimension · Tracy-Widom distribution · Robustness

Mathematics Subject Classification (2010) 62H10 · 62H15 · 62E20

1 Introduction

For $1 \leq i \leq q+1$, let p-vectors $\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,n_i}$ denote a random sample of size n_i drawn from the i-th population with multivariate normal distribution $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$. Let $n = n_1 + \cdots + n_{q+1}$ denote the total sample size. One common form of hypothesis testing involves testing the equality of group means:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = ... = \boldsymbol{\mu}_{q+1} \text{ against } H_1 : H_0 \text{ is false.}$$

In the univariate parametric setting, there is essentially only one test, the analysis of variance (ANOVA) test. However, in multivariate parametric setting, there is no unique test. The different statistical tests on the equality of mean vectors from populations is broadly referred to as Multivariate Analysis of Variance (MANOVA). There are four MANOVA tests that have been given the most attention: (i) the Wilk's Lambda (WL) (also known as likelihood ratio test), (ii) the Lawley-Hotelling trace criterion (LHTC), (iii) the Bartlett-Nanda-Pillai criterion (BNPC), and (iv) Roy's largest root test (RLRT) [1, p.334]. These four tests are functions of the non-zero eigenvalues of

$\mathbf{S}_b\mathbf{S}_e^{-1}$ or $\mathbf{S}_b\mathbf{S}_t^{-1}$, where $\mathbf{S}_b$, $\mathbf{S}_e$ and $\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_e$ are respectively matrices of sums of squares and products due to between-groups, within-groups and total variation. Moreover, these four tests are also invariant. For a test to be invariant, it is necessary for the test to be a function of the the non-zero eigenvalues of $\mathbf{S}_b\mathbf{S}_e^{-1}$ [1, pp.327-328] or [10, p.387]. See [10] for a recent review of "naive" tests of signicance for high-dimensional mean vectors and covariance matrices.

This paper concerns the exact and approximate computation of the right-tail probability of the largest eigenvalue of RLRT under the high dimensional (HD) framework; and their robustness against departure from normality. The right-tail probability gives the p-value of RLRT; and HD refers to the situation when the total sample size (n) and the number of variables (p) in each observations of a random sample both tend to infinity such that p/n tends to a constant $c \in (0, 1)$. We shall call p the data dimension. When exact computation becomes impractical or not feasible (see Section 3.1 for further details), we provide a guideline when the present approximation methods are reasonably accurate.

MANOVA, and hence RLRT, finds many applications in different areas. A researcher could be interested if a treatment is effective by comparing the means of several dependent variables between the (multiple) treatment and control groups, such as differences in beliefs between male and female high school students [21].

RLRT can also be applied in growth curve analysis; to determine (i) if a linear growth is appropriate, (ii) if the population is to be stratified into different groups and (iii) if confidence bands can be obtained from the expected growth curves [27]. RLRT can be used in noise signal detection [15, 22] as well as calculating the probability that a given information rate is not supported due to variable channel capacity [16]. RLRT is the preferred test statistic for a few researchers [8]. When the outcome variables are highly inter-correlated, WL, LHTC and RLRT are more powerful than BNPC [28, 32]. Simulation studies demonstrated that RLRT is the most powerful when the matrix of mean vectors $[\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_{q+1}]$ is of rank 1 [23, 30].

Kritchman and Nadler (2009) derived asymptotic optimality of using the largest sample covariance eigenvalue in testing the two hypotheses: no signals versus one signal of known strength with known noise variance. Based on the sample eigenvalues, we can have an indication whether the rank-one alternative in a given problem is satisfied. Moreover, Johnstone and Nadler (2017) derived approximate power and sample size calculations for RLRT for rank-one alternatives, and Hou et al. (2019) improved these approximate power functions to high-dimensional and finite-ranked cases.

When populations are normally distributed, the joint probability density function (pdf) of the positive eigenvalues is explicitly known. In theory, the cumulative distribution function (CDF) of the largest eigenvalue is a multiple integration problem. If the CDF of the largest eigenvalue can be determined, then the desired quantile can be computed. However, when the number of groups, $q+1$, is not small, the CDF of the largest eigenvalue becomes nontrivial to compute. This problem has been studied by a number of authors [3, 4, 13] and references therein. Chiani (2016) provided an algorithm for exact computation of the CDF using the incomplete Beta functions, Butler and Paige (2011) computed the CDF by summing an infinite series, and Johnstone (2009) established the Tracy-Widom (TW) approximation of the logit-transformed of the largest root.

In the HD framework, the exact methods proposed by Chiani (2016) and Butler and Paige (2011) suffer from numerical stability issues. Chiani's method involves computing the determinant of a

matrix with very small entries computed from the incomplete beta function. In addition, Butler and Paige (2011) method may suffer from long computation times. Some computations in Table 1 from Butler and Paige (2011) are reported to take two hours to compute. While Johnstone (2009) TW approximation is accurate in high dimension, in practice the dimension is finite and thus the approximation may not be accurate, especially in moderate dimensions.

The first goal is to provide recommendations on which method is appropriate under which conditions; and to inform the reader of telltale signs in which the method fails. The second goal is to examine the robustness of the critical values against departure from normality in the HD framework. A common assumption is that the distributions of the populations from which the samples are drawn are normal with a common covariance matrix. While robustness has been studied [23, 30], these are done under the large sample framework, where p is fixed and n is large. Without assuming normality in the population, Bai, Choi and Fujikoshi (2018) showed that for the WL, LHTC and BNPC, a finite fourth moment was sufficient to establish robustness under the HD framework. They proved that the limiting distribution of suitably scaled and centered statistic (WL, LHTC and BNPC) converges to the same limiting distribution as does the populations that are normally distributed. To the best of the authors' knowledge, there is no similar result for RLRT.

One contribution is to provide empirical evidence of the robustness of the critical values of RLRT against departure of normality under the HD framework. A significant implication of the robustness is that it broadens the applicability of RLRT under HD to more populations which do not satisfy the normality assumption. The critical values of the largest root test can be computed based on Chiani's (2016) algorithm or Johnstone's (2009) Tracy-Widom approximation, both methods assume normality in the populations.

We introduce notation used and two R functions for exact and approximate quantile computation in Section 2. The methods for exact or approximate computation of the critical values are briefly sketched in Section 3. Section 4 describes simulation studies for non-normal populations. Section 5 summarizes the results from our comparative studies. The paper ends with recommendation when to compute exactly or approximately the desired critical values in Section 6.

2 Notation

Recall the notation in the beginning of Section 1, we denote the ith group sample mean vector by $\bar{\mathbf{x}}_i = \frac{1}{n_i}\sum_{k=1}^{n_i}\mathbf{x}_{ik}$ and the overall sample mean vector by $\bar{\mathbf{x}} = \frac{1}{n}(\bar{\mathbf{x}}_1 + \ldots + \bar{\mathbf{x}}_n)$. Let $n\mathbf{S}_b := \sum_{i=1}^{q+1} n_i(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})'$ and $n\mathbf{S}_e := \sum_{i=1}^{q+1}\sum_{k=1}^{n_i}(\mathbf{x}_{ik} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ik} - \bar{\mathbf{x}}_i)'$ be the matrices of sums of squares and products due to between-groups and within-groups respectively. Let $\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_e$, then $n\mathbf{S}_t$ denote the matrix of sums of squares and products due to the total variation. Apart from RLRT, WL, LHTC and BNPC respectively are based on $-\log(|\mathbf{S}_e|/|\mathbf{S}_t|)$, $\mathrm{tr}(\mathbf{S}_b\mathbf{S}_e^{-1})$ and $\mathrm{tr}(\mathbf{S}_b\mathbf{S}_t^{-1})$. RLRT is based on the largest root Θ_1 of $\mathbf{S}_b\mathbf{S}_t^{-1}$ or the largest root ℓ_1 of $\mathbf{S}_b\mathbf{S}_e^{-1}$. Here, $\Theta_1 > \cdots > \Theta_q$ and $\ell_1 > \cdots > \ell_q$ are the non-zero eigenvalues of $\mathbf{S}_b\mathbf{S}_t^{-1}$ and $\mathbf{S}_b\mathbf{S}_e^{-1}$, respectively. For more details about these tests, see Anderson (2003) and Fujikoshi, Ulyanov and Shimizu (2010). Let us denote the largest root test criterion as follows:

$$T = \Theta_1(p, q, n, F), \tag{2.1}$$

where $\Theta_1(p,q,n,F)$ denotes the largest root of $\mathbf{S}_b\mathbf{S}_t^{-1}$ for data-dimension p, $q+1$ populations with total sample size n where individual sample observation is drawn from population distribution F. When F is multivariate normal, we will sometimes suppress the dependence of Θ_1 on F; and when the context is clear, we simply refer it as Θ_1. Under the normal populations assumption, computing the exact null distribution of T, i.e. Θ_1, for given p, q and n, has been studied by many authors; for example, Krishnaiah and Chang (1971), Butler and Paige (2011) and Chiani (2016) and references therein.

It is well-known that under the normal populations assumption, that $n\mathbf{S}_b \sim W_p(q,\mathbf{\Sigma})$, $n\mathbf{S}_e \sim W_p(n-q-1,\mathbf{\Sigma})$, $n\mathbf{S}_t \sim W_p(n-1,\mathbf{\Sigma})$, and $\mathbf{S}_b$ and $\mathbf{S}_e$ are independent [1, 7]. We may assume, without loss of generality, that $\mathbf{\Sigma}=\mathbf{I}$. Denote the non-zero roots of $\mathbf{S}_b\mathbf{S}_t^{-1}$ by $0<\Theta_{p\wedge q}<\cdots<\Theta_1<1$, where $p\wedge q=\min(p,q)$. The exact joint pdf of $(\Theta_1,\ldots,\Theta_{p\wedge q})$, with support $0<y_{p\wedge q}<\cdots<y_1<1$, is given by (Johnstone 2009):

$$f(y_1,\ldots,y_{p\wedge q}) = C_1(p\wedge q,a,b)\prod_{k=1}^{p\wedge q} y_k^{a-1}(1-y_k)^{b-1}\prod_{1\le i<j\le p\wedge q}(y_i-y_j), \tag{2.2}$$

where

$$p\wedge q \overset{\triangle}{=} q; \tag{2.3}$$

$$a = \frac{|q-p|+1}{2} \overset{\triangle}{=} \frac{p-q+1}{2}; \tag{2.4}$$

$$b = \frac{n-p-q}{2}; \tag{2.5}$$

$$C_1(p\wedge q,a,b) \overset{\triangle}{=} C_1(q,a,b) = \pi^{q/2}\prod_{i=1}^{q}\frac{\Gamma(\frac{i+2a+2b+q-2}{2})}{\Gamma(\frac{i}{2})\Gamma(\frac{i+2a-1}{2})\Gamma(\frac{i+2b-1}{2})}, \tag{2.6}$$

where $\overset{\triangle}{=}$ is used since we assume that the data dimension (p) is always greater than the number of groups less 1 (q), i.e., $p\ge q$.

Our assumption of $p\ge q$ is motivated by HD MANOVA consideration. Butler and Paige (2011), Chiani (2016) and (to a certain extent) Johnstone (2009) assume without loss of generality that $p\le q$. Either assumption is acceptable as

$$\Theta_1(p,q,n) \overset{D}{=} \Theta_1(q,p,n)$$

which can be derived from Equations (2.2) to (2.5). Here $\overset{D}{=}$ denotes equal in distribution. Another equivalence links $\Theta_{p\wedge q}$,the smallest root of $\mathbf{S}_b\mathbf{S}_t^{-1}$, to Θ_1, which states

$$\begin{aligned}\Theta_1(p,q,n) &\overset{D}{=} 1-\Theta_q(n-p-1,q,n), \qquad p\ge q,\\ &\overset{D}{=} 1-\Theta_p(p,n-q-1,n), \qquad q\ge p,\end{aligned}$$

which allows us to compute the CDF of $\Theta_{p\wedge q}$. The TW approximation gives a general good approximation of the left tail of $\Theta_{p\wedge q}$.

Different authors adopt different notations. To assist the reader and facilitate better understanding, we present the links between our notation and theirs in the Supplementary Appendix Table 2. For the remaining sections, unless stated otherwise, we always assume $p\ge q$.

3 Comparative Studies

In this section, we compare and contrast the quality of numerical results of methods M1 to M4, explained in Subsections 3.1-3.4 below, in computing the null distribution of Θ_1 under a high-dimensional asymptotic framework in which $p, n \to \infty$ satisfying $p/n \to c \in (0, 1)$. From Equation (2.2), the CDF of Θ_1, $F_{\Theta_1}(\theta_1)$, is expressed as

$$F_{\Theta_1}(\theta_1) = \int \dots \int_{0 \le y_q < \dots < y_1 \le \theta_1} f(y_1, y_2, ..., y_q)\, dy_1 \dots dy_q. \tag{3.7}$$

It is effectively an integration problem to compute $F_{\Theta_1}(\theta_1)$, doable in principle but computationally non-trivial when q is large.

We focus on the methods of Butler and Paige (2011), Chiani (2016) and Johnstone (2009). Butler and Paige (2011), and Chiani (2016) provided exact computations by expressing the CDF of Θ_1 as a function of the square root of the determinant of a skew-symmetric matrix via de Bruijn's identity (1955). To evaluate the terms in the matrix, Butler and Paige (2011) expressed the terms as a sum of an infinite series while Chiani (2016) used incomplete beta functions. Johnstone (2009) established the Tracy-Widom (TW) approximation of the logit-transformed of the largest root Θ_1.

The comparison study is necessary as Chiani's method is applicable for moderate a and b whereas TW approximation is good when a and b are large. We are interested in the upper quantiles of Θ_1, particularly, the 0.9, 0.95, 0.99 quantiles. We consider $c = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.6, 0.7, 0.8, 0.9$, $q = 1, 3, 7$, and equal group sample size with $n_i = 20, 40, 60, 80, 100$.

To compare the different methods, two different R functions are used: (i) the `doubleWishart`(θ_1, q, a, b) function from the rootWishart package written by Turgeon (2018) which implements Chiani's (2016) algorithm to provide exact computations; and (ii) the `ptw`(θ_1) function from the RMTstat package written by Johnstone et al. (2014) which implements his approximations. The `doubleWishart` function uses multiple precision library and hence is able to provide more decimal points than base R, increasing accuracy and reducing round off errors. The `ptw` function uses a lookup table. The values were pre-computed at 769 values uniformly spaced between -10 and 6 using MATLAB's `bvp4c` solver to a minimum accuracy of about 3.4×10^{-08}.

3.1 rootWishart

Butler and Paige (2011), and Chiani (2016) show that Equation (3.7) can be rewritten as

$$F_{\Theta_1}(\theta_1) = P(\Theta_1 \le \theta_1) = C_1(q, a, b)\sqrt{|\mathbf{A}(\theta_1)|} \tag{3.8}$$

where $C_1(q, a, b)$ is given by Equation (2.6), $|\cdot|$ denotes the determinant of a matrix, and $\mathbf{A}(\theta_1)$ is a skew-symmetric matrix of order size $q \times q$ if q is even, and $(q + 1) \times (q + 1)$ if q is odd.

Chiani (2016) gave an efficient algorithm to compute the CDF, from which we can compute the critical values. The entries of the skew-symmetric matrix $\mathbf{A}(\theta_1)$ involves incomplete beta functions, $\mathcal{B}(x; c, d) = \int_0^x t^{c-1}(1 - t)^{d-1}\, dt$. His algorithm assumes the input of incomplete beta function, and he derives recurrence relations of entries of $\mathbf{A}(\theta_1)$. For more details, see Chiani (2016).

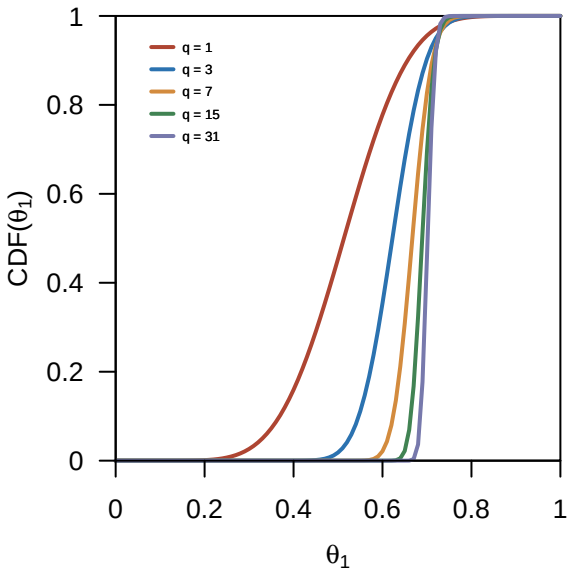

Figure 1: CDF of Θ_1 when $n_i = 20, c = 0.5$, $0 \leq \theta_1 \leq 1$

We point out two computational technicalities inherent in the high-dimensional setting. First, the normalization constant,

$$C_1(q, a, b) = \pi^{q/2} \prod_{k=1}^{q} \frac{\Gamma\left(\frac{2a+2b+q+k-2}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{2a+k-1}{2}\right)\Gamma\left(\frac{2b+k-1}{2}\right)},$$

poses problem for R software for large a and b. To overcome this technicality, we suggest users first compute $\log C_1(q, a, b)$ as the software handles $\log\Gamma(x)$ for very large x well. Second, for large a and b, as in our high-dimensional case, we need to guard the accuracy of incomplete beta functions due to truncation because $x^{a-1}(1-x)^{b-1}$ will be very small for large a and b. To illustrate our second point, we plotted the CDF of Θ_1 for a fixed $n_i = 20$ and $c = 0.5$, varying q.

From Figure 1 (also Figure 6 in the supplementary material), the distribution of Θ_1 becomes very "peaked" as q increases. Hence when q is very large, a small error in estimating the quantile leads to substantial error in estimating the probability, resulting in numerical instability. In addition, in order to ensure a high accuracy in estimating the quantile, any root finding algorithm would take a much longer computational time due to the level of precision required. In fact, straightforward implementation of his algorithm in R leads to false zero entries of $\mathbf{A}(\theta_1)$.

Chiani (2016) implemented his algorithm in Wolfram Mathematica. In his website, `https://sites.google.com/site/marcochianigroup/articles`, the codes can be found. As Wolfram Mathematica allows detailed control over precision, there were no issues in terms of precision to our knowledge. However, Wolfram Mathematica is not freely available. Implementing Chiani's (2016) algorithm in programming languages like Python and R would require multi-precision libraries. For example, when we tried to compute the 95th percentile for $q = 7, n_i = 40, p = 32$ using Base R, we were unable to get an estimate using Chiani's algorithm. Fortunately, Turgeon (2018) created an R package, rootWishart, implementing Chiani's (2016) algorithm using multi-precision linear alge-

bra. As the package only provides the CDF, the `uniroot` function in R was used to obtain the percentiles. Using rootWishart gives us an estimate of 0.216909 which agrees with our simulation results and the Tracy-Widom approximation, both methods are to be described below. We denote the rootWishart method as M1.

3.2 Tracy-Widom Approximation

Under the assumption that p is even and that $p, q = q(p), n = n(p) \to \infty$ satisfying $\lim_{p\to\infty} \frac{q}{n} > 0$ and $\lim_{p\to\infty} \frac{p}{n-q} < 1$, Johnstone (2009) showed that for $a \geq -1/2$, $b \geq 0$,

$$\frac{\log(\frac{\Theta_1(p,q,n)}{1-\Theta_1(p,q,n)}) - \mu(p,q,n)}{\sigma(p,q,n)} \xrightarrow{D} TW_1$$

where $\xrightarrow{D}$ denotes convergence in distribution and TW_1 is a random variable that follows the Tracy-Widom (TW) distribution of order 1 (Johnstone, 2009). The parameters $\mu(p,q,n)$ and $\sigma(p,q,n)$ are given by

$$\begin{aligned}
\mu &= 2\log\tan(\frac{\gamma+\phi}{2}), \\
\sigma &= \sqrt[3]{\frac{16}{(n-2)^2}\frac{1}{\sin^2(\gamma+\phi)\sin(\gamma)\sin(\phi)}}, \\
\gamma &= \arccos(\frac{n-2q-1}{n-2}), \\
\phi &= \arccos(\frac{n-2p-1}{n-2}).
\end{aligned}$$

Johnstone et al. (2014) created an R package RMTstat to compute the density, distribution and quantile functions as well as a random number generator for the TW distribution. We compute the centering and scaling terms according to the expressions above, and the upper percentage point of TW_1 with this package. We refer this as method M2.

3.3 Monte Carlo Method

As computing the CDF of Θ_1 is a multi-dimensional integration problem, one possible solution would be to apply Monte Carlo methods to approximate the integral. We can rewrite the multiple integral as the expectation of a function of independent Beta random variables as follows

$$\begin{aligned}
&F_{\Theta_1}(\theta_1) \\
&= \int\dots\int_{0\leq y_q<\dots<y_1\leq\theta_1} C_1(q,a,b)\prod_{k=1}^{q} y_k^{a-1}(1-y_k)^{b-1}\prod_{1\leq i<j\leq q}(y_i-y_j)\,dy_1\dots dy_q \\
&= E\Big[I\{0\leq Y_q<\dots<Y_1\leq\theta_1\}\Big(\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\Big)^q C_1(q,a,b)\prod_{1\leq i<j\leq q}(Y_i-Y_j)\Big]
\end{aligned}$$

where $Y_1, Y_2, \ldots, Y_q$ are i.i.d. Beta(a, b) random variables and $\Gamma(\cdot)$ is the Gamma function.

To approximate the CDF of Θ_1, first we generate q i.i.d. Beta(a, b) random variables $y_1, y_2, \ldots, y_q$. If $\max_{1\le i\le q} y_i > \theta_1$, return 0; else compute $\left(\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\right)^q C_1(q, a, b) \prod_{1\le i<j\le q}(y_i - y_j)$. We repeat these steps 10,000 times and take the mean of the estimates to obtain an approximation for the CDF of Θ_1. Similar to M1, the `uniroot` function in `R` was used to obtain the percentiles. We denote the Monte Carlo method as M3.

3.4 Simulation

We also conducted simulations where the underlying populations are simulated under the standard normal distribution, labelled as M4. Further details can be found in Section 4.

We will use M1 as the benchmark for the comparative studies. We consider M4 because it gives us a "rough" benchmark to gauge if rootWishart is giving a sensible value. Though in theory rootWishart gives the accurate value, it may still possibly give erroneous result when a and b are large due to truncation.

4 Simulation studies for non-normal populations

Motivated by relevance to the MANOVA testing problem, we are interested in how the upper-tail quantiles differ given different population distributions. We investigate the 90th, 95th and 99th percentiles of $\Theta_1(p, q, n, F)$ for various (p, q, n) in a high-dimensional framework $p/n \to c \in (0, 1)$. Seven distributions are considered: (i) N(0,1), (ii) t_3, (iii) t_4, (iv) t_5, (v) χ^2_3, (vi) Exponential with mean 1, and (vii) Poisson with mean 1. Their standardized versions (i.e., mean 0 and variance 1) are referred to as $F_1, \ldots F_7$ accordingly. Distributions (i) to (iv) are symmetric, whereas (v) to (vii) are skewed. Moreover, the finite fourth moment assumption in [2] does not hold for F_2 and F_3. We shall restrict ourselves to consider equal group sample size, i.e., $n_1 = n_2 = \ldots = n_{q+1}$; and hence the total sample size is $n = (q + 1)n_1$. The choice of our simulation parameters are as follows: $c = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35,\ 0.4, 0.45, 0.5, 0.6, 0.7, 0.8, 0.9,\ n_1 = 20, 40, 60, 80, 100$ and $q = 1, 3, 7$.

The simulation steps are as follows: for a fixed $j = 1, \ldots, 7, n_1, q$ and c, let $p = nc$. We first simulate a random matrix $\mathbf{Y}$ of size $p \times n$. The entries of the matrix are simulated independently from F_j (recall that F_j is standardized to mean 0 and variance 1). Let $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]$. Compute the overall mean vector $\bar{\mathbf{y}} := (\mathbf{y}_1 + \mathbf{y}_2 + \ldots + \mathbf{y}_n)/n$ and the $q+1$ population mean vectors $\bar{\mathbf{y}}_i := \frac{1}{n_i}\sum_{k\in \text{Group } i} \mathbf{y}_k$, $i = 1, \ldots, q + 1$. From $\mathbf{Y}$, we compute $\mathbf{S}_t = \frac{1}{n}\sum_{l=1}^{n}(\mathbf{y}_l - \bar{\mathbf{y}})(\mathbf{y}_l - \bar{\mathbf{y}})'$. We represent $\mathbf{S}_b = \mathbf{X}\mathbf{X}'$ where

$$\mathbf{X} = \left[\sqrt{\frac{n_1}{n}}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}), \sqrt{\frac{n_2}{n}}(\bar{\mathbf{y}}_2 - \bar{\mathbf{y}}), \ldots, \sqrt{\frac{n_{q+1}}{n}}(\bar{\mathbf{y}}_{q+1} - \bar{\mathbf{y}})\right]$$

It can be shown that $\mathbf{S}_b\mathbf{S}_t^{-1}$ and $\mathbf{X}'\mathbf{S}_t^{-1}\mathbf{X}$ have the same set of q positive eigenvalues. However, computationally it is much easier to deal with the second matrix as it is of order $(q + 1) \times (q + 1)$

whereas the first is of order $p \times p$ (note that p is much greater than q and is increasing with n). We then extract the largest eigenvalue θ_1. We repeat this procedure 10,000 times.

5 Results

In this section, we present the results of the comparison tests among the different methods M1 to M4. Recall M1 uses the rootWishart package implemented by Turgeon (2018) which is based on Chaini's (2016) algorithm. M2 uses the RMTstat package based on the Tracy-Widom approximation (Johnstone 2009). M3 uses Monte Carlo and M4 uses simulations. We end this section by presenting the results of our simulations over 7 different distributions.

5.1 Comparison of Method 1 with Butler and Paige

We compared the runtimes for Butler and Paige's (2011) method against M1. In Butler and Paige (2011) Table 1, when $p = q = 24, n = 55$ (our notation) or $m = k = 24, n = 30$ (Butler and Paige's notation), the estimated 95th percentile is 0.99161 with a runtime of almost two hours. Using M1, the average of ten runtimes is 12.55 seconds (min = 11.12 secs, max = 13.74 secs). The estimated 95th percentile from M1 is 0.9916111, which agrees with Butler and Paige up to the fifth decimal places.

When q is large, R package rootWishart, even with multi-precision libraries, fails. For example, warnings are raised when attempting to compute the 95th percentile for $q = 127, n_i = 100, c = 0.05$ using M1.

5.2 Comparison of Methods 1 to 4

We compared the estimation of the 90th, 95th and 99th percentiles using M2 to M4, with M1 as the benchmark. We considered $c = 0.05, 0.1, 0.15, 0.2,\ 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.6, 0.7, 0.8, 0.9,\ q = 1, 3, 7,\ n_i = 20, 40, 60, 80, 100$. The absolute percentage errors, $100|\hat{\theta}_x(p, q, n) - \hat{\theta}_{M1}(p, q, n)|/\hat{\theta}_{M1}(p, q, n)$, are computed. Here $\hat{\theta}_x(p, q, n)$ is the quantile generated using method M2, M3 or M4. See `Excel v2 Tab 1 chiani_algo` in Supplementary Material for the absolute percentage errors.

For M2, when $q = 1$, the absolute percentage errors were above 1% and increased with c, the percentile and decreases with n_i. The absolute percentage errors decreased when $q = 3$ compared to $q = 1$. When $q = 7$, almost all the absolute percentage errors were below 1%.

For M3, when $q = 1$, most of the absolute percentage errors were below 1%. When $q = 3$, the absolute percentage errors increased. Particularly, many of the absolute percentage errors for the 99th percentile were very high. When $q = 7$, M3 suffers even higher absolute percentage errors. These suggest that M3 should not be used.

Among the three methods, the quantiles obtained from M4 were the most accurate, where most of the absolute percentage errors were below 1% and all of the absolute percentage errors were below

3%. For details, refer to `Excel v2 Tab 2 Large_q_rootwishart_TW_error` in the Supplementary Material.

We further considered $q \in \{15, 31, 63\}$ for (i) M1 and (ii) M2 with the same c and n_i parameters. We reported the absolute percentage error, defined as $100|\hat{\theta}_{M1}(p,q,n) - \hat{\theta}_{M2}(p,q,n)|/\hat{\theta}_{M1}(p,q,n)$ where $\hat{\theta}_{M1}(p,q,n)$ and $\hat{\theta}_{M2}(p,q,n)$ are respectively the estimated quantile using M1 and M2. M2 returned an estimate for all the possible combinations of our parameters without warnings or errors but M1 returned errors for some combinations of the parameters.

We observed that when $q = 15$, the absolute percentage error was always less than 1% except for 1 out of 210 cases, indicating that M2 was good. In that one case, when $n_i = 20$, $c = 0.05$ and the 99th percentile, the absolute percentage error was 1.36%. When $q = 31$, the approximation was good (absolute percentage error was less than 1%) except when $c \geq 0.8$. When $c \geq 0.8$, M1 returned either a similar value to M2, a value very different or an error depending on the choice of n_i and c. When $q = 63$, $n_i \in \{20, 40\}$ and $c \in \{0.05, 0.1\}$, the estimates were very similar although the absolute percentage error for the 95th and 99th percentiles for $q = 63, n_i = 40, c = 0.1$ were 1.39% and 2.94% respectively. The absolute percentage error when $q = 63$, $n_i = 20$, $c = 0.15$ was also below 1%. For other choice of parameters when $q = 63$, M1 was numerically unstable in estimating the quantiles. Details can be found in `Excel v2 Tab 3 quantiles_TW_q2047` in Supplementary Material.

We tested M2 for values of q up to $2047 = 2^{11} - 1$ and no errors messages or warnings were reported. For large values of q, the Tracy-Widom distribution was a good approximation.

We end this subsection with a note. Chiani (2016, p.470) described an approximation to the Tracy-Widom distribution using a shifted-Gamma distribution. We observed that the Tracy-Widom approximation was very similar to the shifted-Gamma distribution. All the absolute percentage errors between those two approximations were very similar and decreased as q increased.

5.3 The Normality Assumption

From the simulations described in Section 4, we measure the variations of the quantiles across different distributions in our simulation studies in two ways: the coefficient of variation (CV) and the distance between the theoretical quantile and the empirical quantile over seven distributions for data dimension p, number of populations $q + 1$ and fixed batch size n_i. Let $\bar{\theta}_{M4}(p,q,n) = \sum_{j=1}^{7} \hat{\theta}_{M4}(p,q,n,F_j)/7$. We then computed the CV (or relative standard deviation):

$$\mathrm{CV}(p,q,n) = \frac{\left\{\sum_{j=1}^{7}\left[\hat{\theta}_{M4}(p,q,n,F_j) - \bar{\theta}_{M4}(p,q,n)\right]^2/6\right\}^{1/2}}{\bar{\theta}_{M4}(p,q,n)}.$$

The distance between the theoretical quantile and the empirical quantiles generated by F_j for $1 \leq j \leq 7$ for data dimension p, number of populations $q + 1$ and fixed batch size n_i is defined by

$$d_j(p,q,n) = |\hat{\theta}_{M4}(p,q,n,F_j) - \theta_{M1}(p,q,n,F_1)|,$$

where $\theta_{M1}(p,q,n,F_1)$ is computed using the rootWishart package.

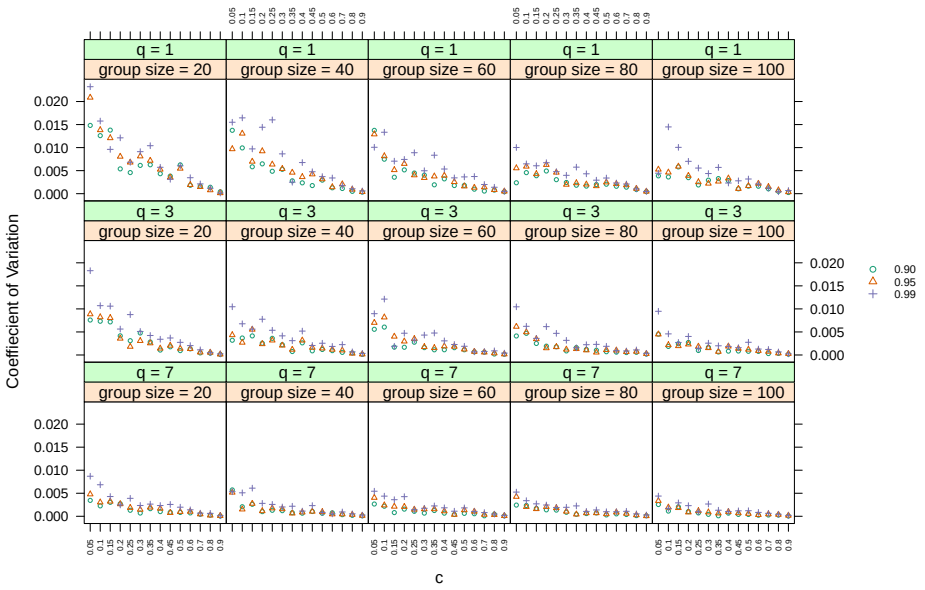

Figure 2: Coefficient of Variation computed based on the empirical percentiles generated from the 7 distributions for the 90th, 95th and 99th percentiles.

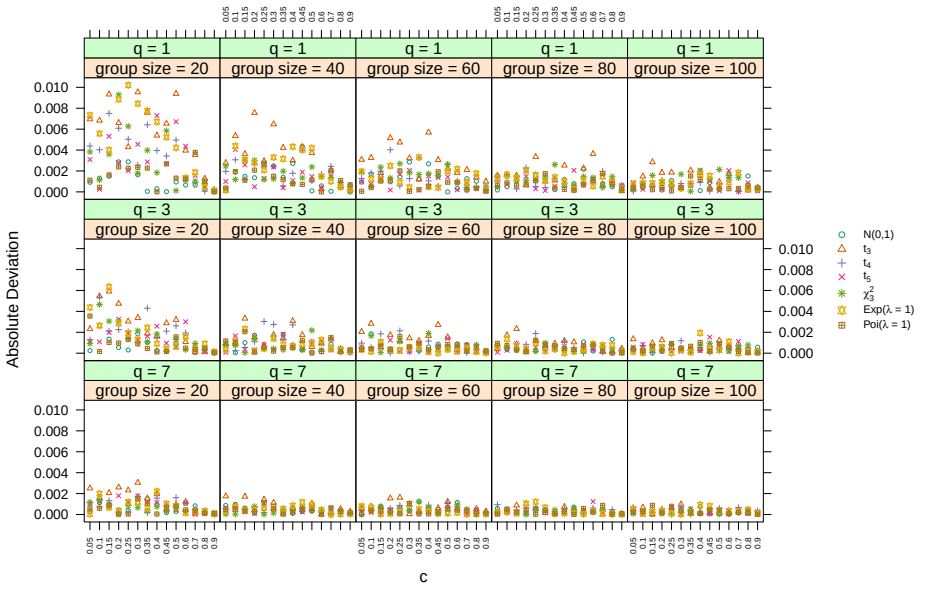

Figure 3: Absolute deviation computed based on the 95% empirical percentiles generated from the 7 distributions.

From Figure (2), we observed that the CV decreases with an increase in n_i, q, c and the quantiles. From Figure (3), we plotted the absolute deviations for the 95th percentile. Due to space consideration, the absolute deviations for the 90th and 99th percentiles were reported in the Supplementary Material. The results were similar across the 3 different percentiles. Similar to the CV, the absolute deviations decreased with an increase in n_i, q, c and the percentiles. We observed that the percentiles generated by the t_3 distribution, which does not have a finite fourth moment, had the highest absolute deviation in many of the cases. Skewness as in χ^2_3 and Exponential with mean 1 also seemed to affect the absolute error. Expectedly, the percentiles generated from the N(0,1) distribution had the least absolute deviation in many of the cases since $\theta_{M1}(p, q, n, F_1)$ was computed assuming the distribution was normal. Nonetheless, we observed that the errors were very small (except when both n_i and q were small) which suggests that Θ_1 was robust against the underlying distribution. A consequence of the robustness was that we could use Chiani's (2016) algorithm or the Tracy-Widom approximation to compute the critical values even the underlying population distribution may not be normal.

6 Recommendations and Concluding Remarks

We recommend using Chiani's (2016) algorithm with multi-precision libraries to compute the quantiles whenever it is possible. This can be done via the R package rootWishart implemented by Turgeon (2018). For large parameters (q, a and/or b) resulting from a large sample size and/or large data dimension, we can use the Tracy-Widom distribution to obtain the quantiles via the R package RMTstat. Specifically, our recommendations to compute the critical values of RLRT in MANOVA are

- when $q \leq 15$, use rootWishart;
- when $16 \leq q \leq 31$ and $p < 0.8n$, use rootWishart;
- when $16 \leq q \leq 31$ and $p \geq 0.8n$, use RMTstat;
- when $q \geq 32$, use RMTstat.

In addition, there are some telltale signs we can use to assess whether rootWishart fails to give sensible values. When these signs happen, we recommend to use the estimate given by RMTstat instead. The first sign would be to evaluate the CDF at $\theta_1 = 1$, which is 1. However, due to numerical instability, it is possible that 0 or some value greater than 1 is returned. Table 1 shows some of the problems we encountered when we implemented Chiani's (2016) algorithm on Base R without using multi-precision libraries.

Table 1: Issues with numerical stability

p	q	n	θ_1	$F_{\Theta_1}(\theta_1)$ in Base R	$F_{\Theta_1}(\theta_1)$ in RootWishart
432	7	481	1	18430.81	1
720	7	801	1	0	1

The second sign would be to compute the absolute percentage error
$100|\hat{\theta}_{M1}(p,q,n) - \hat{\theta}_{M2}(p,q,n)|/\hat{\theta}_{M1}(p,q,n)$ where $\hat{\theta}_{M1}(p,q,n)$ is the estimated quantile using M1 and $\hat{\theta}_{M2}(p,q,n)$ is the estimated quantile using M2. For $q \geq 15$, our calculations show that the percentage error should be below 1%. If the percentage error (when $q \geq 15$) is greater than 1%, we think that the Tracy-Widom approximation is more reliable. For large values of q, the Tracy-Widom distribution is a good approximation.

In conclusion, we are interested in computing the multiple integral of the joint probability density function of the q non-zero roots of $\mathbf{S}_b\mathbf{S}_t^{-1}$. This integration problem is motivated by Roy's Largest Root test, where the p-value of the test is the right tail probability of the largest root of $\mathbf{S}_b\mathbf{S}_t^{-1}$. The solution to this problem can be applied to compute the left tail probability of the smallest root of $\mathbf{S}_b\mathbf{S}_t^{-1}$ (Johnstone 2009) and leads us to compute the joint probability of the largest and smallest root of $\mathbf{S}_b\mathbf{S}_t^{-1}$. In addition, although we focused on the problem of testing the equality of the p dimensional mean vectors for $q+1$ groups, there are two other hypothesis testing problems in multivariate analysis where the multiple integral of Equation (2.2) with suitably chosen a and b is of interest (Pillai 1955), namely,

1. Testing the equality of covariance matrices of two p-variate normal populations: letting $a = (n_1 - p)/2$ and $b = (n_2 - p)/2$ where n_1 and n_2 denote sample sizes drawn from the first and second populations.
2. Testing independence between a p-set and a q-set of variates from a random sample of size n of a $(p+q)$-variate normal population with $p \leq q$ where $a = (q-p+1)/2$ and $b = (n-p-q)/2$.

Our simulation results provide support of robustness in the HD MANOVA setting. This leads us to surmise that the above two problems may also be robust against non-normality assumption, although further simulation studies is warranted.

Acknowledgements

ZD Bai was partially supported by NSFC (No.11571067). KP Choi acknowledges support from Singapore Ministry of Education Academic Research Funds Tier 1 grant R-155-000-188-114. J Hu was partially supported by NSFC (No.11771073, 11971097).

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Anderson, T.W.: An Introduction to Multivariate Statistical Analysis, 3 edn. John Wiley & Sons, Hoboken, NJ (2003)

[2] Bai, Z., Choi, K.P., Fujikoshi, Y.: Limiting behavior of eigenvalues in high-dimensional MANOVA via RMT. Ann. Statist. **46**(6A), 2985–3013 (2018)

[3] Butler, R.W., Paige, R.L.: Exact distributional computations for Roy's statistic and the largest eigenvalue of a Wishart distribution. Statistics and Computing **21**, 147–157 (2011)

[4] Chiani, M.: Distribution of the largest root of a matrix for Roy's test in multivariate analysis of variance. Journal of Multivariate Analysis **143**(C), 467–471 (2016)

[5] Davis, A.W.: On the effects of moderate multivariate nonnormality on roy's largest root test. Journal of the American Statistical Association **77**(380), 896–900 (1982)

[6] de Bruijn, N.G.: On some multiple integrals involving determinants. Journal of the Indian Mathematical Society. New Series **19**, 133–151 (1955)

[7] Fujikoshi, Y., Ulyanov, V.V., Shimizu, R.: Multivariate statistics: High-dimensional and large-sample approximations, vol. 760. John Wiley & Sons (2011)

[8] Harris, R.J.: A primer of multivariate statistics. Psychology Press (2001)

[9] Hou, Z., Liu, Y., Bai, Z., Hu, J.: Approximation of the power functions of Roy's largest root test under general spiked alternatives. To appear in Random Matrice: Theory and Applications.

[10] Hu, J., Bai, Z.: A review of 20 years of naive tests of signicance for high-dimensional mean vectors and covariance matrices. Science China Mathematics **59**, 2281–2300 (2009)

[11] Ito, P.K.: 7 robustness of anova and manova test procedures. In: Analysis of Variance, *Handbook of Statistics*, vol. 1, 199 – 236. Elsevier (1980)

[12] Johnstone, I.M.: Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy-Widom limits and rates of convergence. Ann. Stat. **36**(6), 2638–2716 (2008)

[13] Johnstone, I.M.: Approximate null distribution of the largest root in multivariate analysis. Ann. Appl. Stat. **3**(4), 1616–1633 (2009)

[14] Johnstone, I.M., Ma, Z., Perry, P.O., Shahram, M.: RMTstat: Distributions, Statistics and Tests derived from Random Matrix Theory (2014). R package version 0.3

[15] Johnstone, I.M., Nadler, B.: Roy's largest root test under rank-one alternatives. Biometrika **104**(1), 181–193 (2017)

[16] Kang, M., Alouini, M.S.: Water-filling capacity and beamforming performance of mimo systems with covariance feedback. 556 – 560 (2003)

[17] Krishnaiah, P.R.: 24 computations of some multivariate distributions. In: Analysis of Variance, *Handbook of Statistics*, vol. 1, 745 – 971. Elsevier (1980)

[18] Krishnaiah, P.R., Chang, T.C.: On the exact distributions of the extreme roots of the wishart and manova matrices. Journal of Multivariate Analysis **1**(1), 108 – 117 (1971)

[19] Kritchman, S., Nadler, B.: Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory. IEEETrans. Sig. Proces **57**, 3930-3941 (2009)

[20] Mardia, K.V., Bibby, J.M., Kent, J.T.: Multivariate analysis. Probability and Mathematical Statistics. Acad. Press (1982)

[21] Mason, L.: High school students' beliefs about maths, mathematical problem solving, and their achievement in maths: A cross-sectional study. Educational Psychology - EDUC PSYCHOL-UK **23**, 73–85 (2003)

[22] Nadler, B., Johnstone, I.M.: Detection performance of Roy's largest root test when the noise covariance matrix is arbitrary. 2011 IEEE Statistical Signal Processing Workshop (SSP) 681–684 (2011)

[23] Olson, C.L.: Comparative robustness of six tests in multivariate analysis of variance. Journal of the American Statistical Association **69**(348), 894–908 (1974)

[24] Olson, C.L.: On choosing a test statistic in multivariate analysis of variance. Psychological Bulletin **83**(4), 579–586 (1976)

[25] Olson, C.L.: Practical considerations in choosing a manova test statistic: A rejoinder to stevens. Psychological Bulletin **86**, 1350–1352 (1979)

[26] Pillai, K.C.S.: Some new test criteria in multivariate analysis. Ann. Math. Statist. **26**(1), 117–121 (1955)

[27] Potthoff, R.F., Roy, S.N.: A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika **51**(3/4), 313–326 (1964)

[28] Rencher, A.C.: Methods of Multivariate Analysis (2nd Edition). Wiley, New York, NY (2002)

[29] SAS: Big data: What it is and why it matters. `https://www.sas.com/en_sg/insights/big-data/what-is-big-data.html` (2020). Accessed: 2020-04-14

[30] Schatzoff, M.: Sensitivity comparisons among tests of the general linear hypothesis. Journal of the American Statistical Association **61**(314), 415–435 (1966)

[31] Schervish, M.J.: Theory of Statistics. Springer, New York, NY (1995)

[32] Tabachnick, B.G., Fidell, L.S.: Using Multivariate Statistics (5th Edition). Allyn and Bacon, Inc., USA (2006)

[33] Turgeon, M.: rootWishart: Distribution of Largest Root for Single and Double Wishart Settings (2018). R package version 0.4.1

Supplementary Material

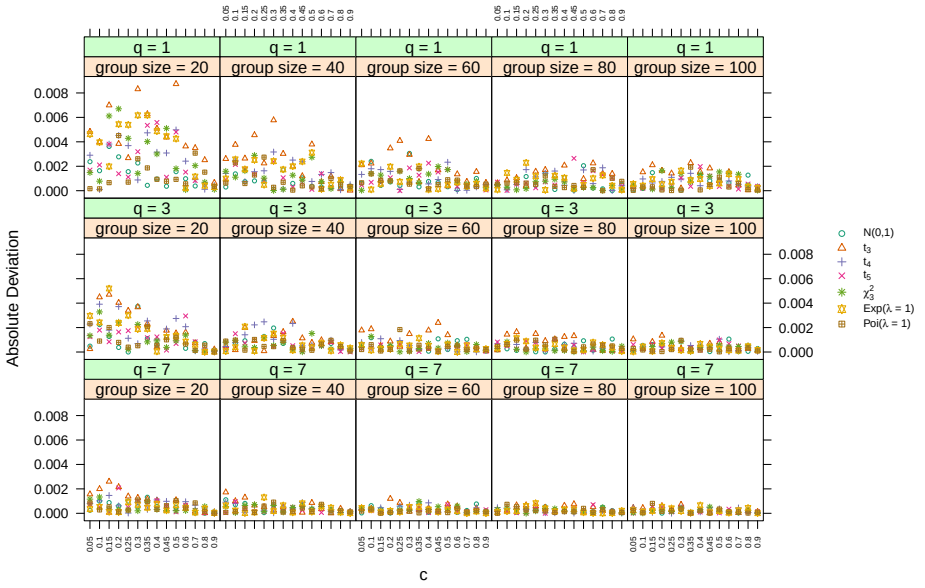

Figure 4: Absolute deviation computed based on the 90% empirical quantiles generated from the 7 distributions

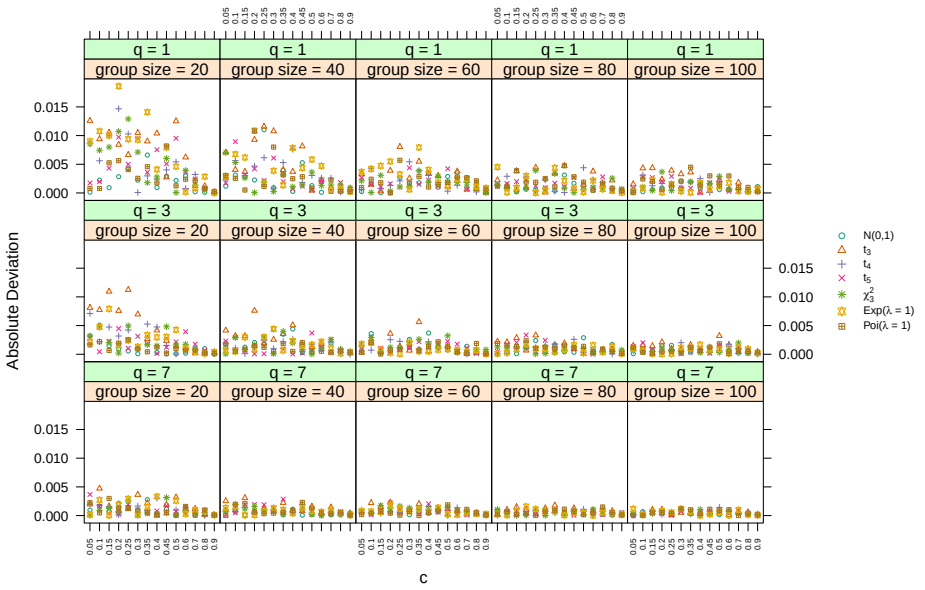

Figure 5: Absolute deviation computed based on the 99% empirical quantiles generated from the 7 distributions

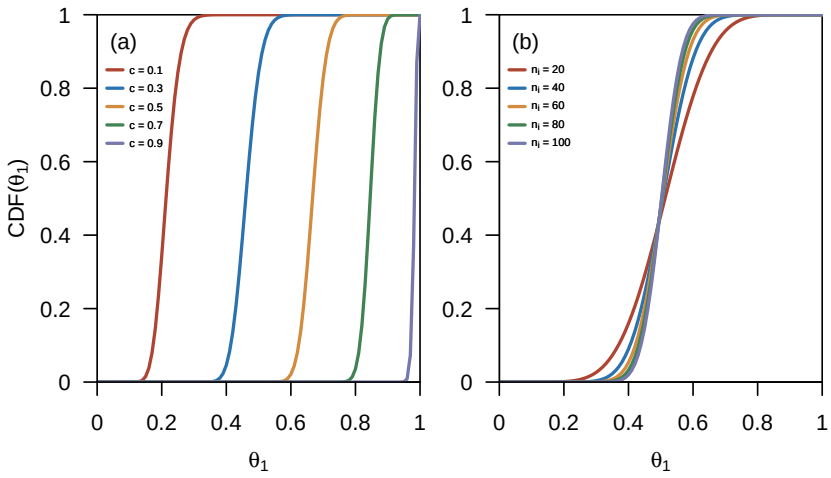

Figure 6: CDF of Θ_1 when (a) $n_i = 20$, $q = 7$, (b) $c = 0.5$, $q = 1$

Table 2: Comparison of Notations for MANOVA setting

	Ours	Butler & Paige	Chiani	Johnstone
Data dimension	p	k	s	p
No. of groups	$q+1$	$m+1$	$2m+s+2$	$n+1$
Group size	n_i	-	-	-
Sample size	$n = \sum_{i=1}^{q+1} n_i$	$n+m+1$	$2(m+n+s+1)+1$	$m+n+1$
Shape Parameter a	$(p-q+1)/2$	$(m-k+1)/2$	$m+1$	$(n-p+1)/2$
Shape Parameter b	$(n-p-q)/2$	$(n-k+1)/2$	$n+1$	$(m-p+1)/2$
Assumption	$p \geq q$	$k \leq m$	$-1/2 \leq m$	$p \leq n$
	$n-q-1 \geq p$	$k \leq n$	$-1/2 \leq n$	$p \leq m$