Journal of Multivariate Analysis 143 (2016) 467–471

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Distribution of the largest root of a matrix for Roy's test in multivariate analysis of variance

Marco Chiani

DEI, University of Bologna, V.le Risorgimento 2, 40136 Bologna, Italy

ARTICLE INFO

Article history:
Received 7 February 2014
Available online 26 October 2015

AMS subject classifications:
primary 62H10
secondary 62H15
62J10

Keywords:
Roy's test
Random matrices
Multivariate analysis of variance (MANOVA)
Characteristic roots
Largest eigenvalue
Tracy–Widom distribution
Wishart matrices

ABSTRACT

Let $\mathbf{X}$, $\mathbf{Y}$ denote two independent real Gaussian $\mathsf{p} \times \mathsf{m}$ and $\mathsf{p} \times \mathsf{n}$ matrices with $\mathsf{m}, \mathsf{n} \geq \mathsf{p}$, each constituted by zero mean independent, identically distributed columns with common covariance. The Roy's largest root criterion, used in multivariate analysis of variance (MANOVA), is based on the statistic of the largest eigenvalue, Θ_1, of $(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}$, where $\mathbf{A} = \mathbf{X}\mathbf{X}^T$ and $\mathbf{B} = \mathbf{Y}\mathbf{Y}^T$ are independent central Wishart matrices. We derive a new expression and efficient recursive formulas for the exact distribution of Θ_1. The expression can be easily calculated even for large parameters, eliminating the need of pre-calculated tables for the application of the Roy's test.

1. Introduction

The joint distribution of s non-null eigenvalues of a multivariate real beta matrix in the null case can be written in the form [14, page 112], [2, page 331],

$$f(\theta_1, \dots, \theta_s) = C(s, m, n) \prod_{i=1}^{s} \theta_i^m (1-\theta_i)^n \cdot \prod_{i<j}^{s} (\theta_i - \theta_j) \tag{1}$$

where $1 > \theta_1 \geq \theta_2 \cdots \geq \theta_s > 0$, and $C(s, m, n)$ is a normalizing constant given by

$$C = C(s, m, n) = \pi^{s/2} \prod_{i=1}^{s} \frac{\Gamma\left(\frac{i+2m+2n+s+2}{2}\right)}{\Gamma\left(\frac{i}{2}\right)\Gamma\left(\frac{i+2m+1}{2}\right)\Gamma\left(\frac{i+2n+1}{2}\right)}. \tag{2}$$

This distribution arises in multivariate analysis of variance (MANOVA) and, with the notation introduced above, is the distribution of the eigenvalues of $(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}$ with parameters

$$s = \mathsf{p}, \qquad m = (\mathsf{n}-\mathsf{p}-1)/2, \qquad n = (\mathsf{m}-\mathsf{p}-1)/2. \tag{3}$$

The marginal distribution of the largest eigenvalue, Θ_1, is of basic importance in testing hypotheses and constructing confidence regions in multivariate analysis of variance (MANOVA) according to the Roy's largest root criterion (see e.g.

E-mail address: marco.chiani@unibo.it.

http://dx.doi.org/10.1016/j.jmva.2015.10.007

[2, page 333] and references therein), and is generally considered difficult to compute. For this reason, extensive studies have produced tables of upper percentage points for few specific (small) values of s and some combinations of m, n (see e.g. [8,15], [2, Table B.4]). The most efficient numerical algorithm to compute the CDF of Θ_1 is provided in [3] based on [7], but for nonintegers m or n it requires infinite series expansions which can results in computational time of several hours. Approximations based on the Tracy–Widom distribution are discussed in [12].

In this paper we derive a simple expression for the exact CDF of Θ_1 for arbitrary s, n, m, and an iterative algorithm for its fast evaluation. The algorithm needs only the incomplete beta function, and does not rely on numerical integration or series expansion. For instance, all results in [2, Table B.4] or in [12, Table 1] can be easily computed instantaneously[1]; even for the most challenging cases analyzed in literature [3, Table 1], which with previous methods required hours of computational time, no more than a second is needed. Finally, we discuss some approximations based on the Tracy–Widom distribution and its approximation [12,4].

We remark that we study the case of real matrices: the complex analogous of our problem, i.e., the case where $\mathbf{X}, \mathbf{Y}$ are independent complex Gaussian, is much easier and has been solved in [13].

Throughout the paper we indicate with $\Gamma(\cdot)$ the gamma function, with $B(a, b)$ the beta function, with $\mathcal{B}(x; a, b) = \int_0^x t^{a-1}(1-t)^{b-1}dt$ the incomplete (lower) beta function [1, Ch. 6], and with $|\cdot|$ the determinant.

2. Exact distribution of the largest eigenvalue for multivariate beta matrices in the null case

The following is the main result of the paper.

Theorem 1. *The CDF of the largest eigenvalue Θ_1 for a multivariate beta matrix in the null case is:*

$$F_{\Theta_1}(\theta_1) = \Pr\{\Theta_1 \le \theta_1\} = C\sqrt{|\mathbf{A}(\theta_1)|}. \tag{4}$$

When s is even, the elements of the $s \times s$ skew-symmetric matrix $\mathbf{A}(\theta_1)$ are:

$$a_{i,j}(\theta_1) = \mathcal{E}(\theta_1; m+j, m+i) - \mathcal{E}(\theta_1; m+i, m+j) \quad i, j = 1, \ldots, s \tag{5}$$

where

$$\mathcal{E}(x; a, b) \triangleq \int_0^x t^{a-1}(1-t)^n \, \mathcal{B}(t; b, n+1)\, dt. \tag{6}$$

When s is odd, the elements of the $(s+1) \times (s+1)$ skew-symmetric matrix $\mathbf{A}(x_1)$ are as in (5), with the additional elements

$$a_{i,s+1}(\theta_1) = \mathcal{B}(\theta_1; m+i, n+1) \quad i = 1, \ldots, s \tag{7}$$

$$a_{s+1,j}(\theta_1) = -a_{j,s+1}(\theta_1) \quad j = 1, \ldots, s \tag{8}$$

$$a_{s+1,s+1}(\theta_1) = 0. \tag{9}$$

Note that $a_{i,j}(\theta_1) = -a_{j,i}(\theta_1)$ and $a_{i,i}(\theta_1) = 0$.

Moreover, the elements $a_{i,j}(\theta_1)$ can be computed iteratively, starting from the beta function, without numerical integration or series expansion.

Proof. The proof is based on the approach introduced in [4] for Wishart and GOE matrices.

Denoting $\xi(x) = x^m(1-x)^n$, $\mathbf{x} = [x_1, x_2, \ldots, x_s]$, and with $\mathbf{V}(\mathbf{x}) = \left\{x_j^{i-1}\right\}$ the Vandermonde matrix, we have for the eigenvalues in ascending order

$$f(x_s, \ldots, x_1) = C\prod_{i<j}(x_j - x_i)\prod_{i=1}^{s}\xi(x_i) = C\,|\mathbf{V}(\mathbf{x})|\prod_{i=1}^{s}\xi(x_i) \tag{10}$$

where now $0 < x_1 \le \cdots \le x_s < 1$.

The CDF of the largest eigenvalue is then

$$F_{\Theta_1}(\theta_1) = \int \cdots \int_{0 \le x_1 < \cdots < x_s \le \theta_1} f(x_s, \ldots, x_1) d\mathbf{x} \tag{11}$$

$$= C \int \cdots \int_{0 \le x_1 < \cdots < x_s \le \theta_1} |\mathbf{V}(\mathbf{x})| \prod_{i=1}^{s} \xi(x_i)\, d\mathbf{x}. \tag{12}$$

[1] On a current desktop computer in less than 0.1 s.