Numerical computation on distributions of statistics expressed by generalized hypergeometric function.

Yuichi Takeda
Kanagawa Institute of Technology,
1030 Shimo-ogino, Atsugi,
Kanagawa 243-0292, Japan
E-mail: y-takeda@ctr.kanagawa-it.ac.jp

Hiroki Hashiguchi
Saitama University,
255 Shimo-Okubo, Sakura-ku, Saitama City,
Saitama 338-8570, Japan
E-mail: hiro@ms.ics.saitama-u.ac.jp

Takakazu Sugiyama
Soka University,
1-236 Tangimachi, Hachioji
Tokyo 192-8577, Japan
E-mail: stakakaz@gmail.com

1.Introduction

James (1960) obtained the joint distributions of the characteristic roots of the covariance matrix on Wishart matrix by constituting the zonal polynomials. It was epoch-making in the history of multivariate distribution theory. The many density functions and moments in multivariate analysis have been expressed by the zonal polynomials and the generalized hypergeometric function of symmetric matrix argument. Examples include the noncentral distributions of the characteristic roots in multiple discriminant analysis (Constantine (1963)), the distributions of the largest characteristic root and the corresponding characteristic vector of a Wishart matrix (Sugiyama (1966,1967)) and the distributions of the largest and smallest root of a Wishart distribution of a multivariate beta distribution (Constantine (1963)). Sugiyama (1979) obtained the coefficients of zonal polynomials up to degree 200 in the case of order 2 and the programming to compute them, expressed by a linear combination of monomial symmetric function. At the practical point of view, it is very important that the cumulative density function, moments, and so on are possible to compute in the case of order 3, and also higher order.

The distributions expressed by the generalized hypergeometric function may be written by the zonal polynomial series. Sugiyama, Fukuda and Takeda (1999) derived the partial differential equation to obtain the recurrence relations of coefficients of the generalized hypergeometric function for their computations. It is possible to calculate distribution expressed by generalized hypergeometric function order 7. However, at the present computers it is very difficult to calculate the coefficients of order more than 5. Hashiguchi and Niki (1997) derived the coefficients of zonal polynomials to elementary symmetric polynomials of order 3. They showed some results of the exact values of the percentile points of order 3.

In this paper, we discuss how to calculate zonal polynomials and generalized hypergeometric functions by using a partially differential equation. The densities of the smallest characteristic roots are calculated numerically up to 5 dimensions. Some percentile points are also shown for these statistics.

2. Zonal polynomials

The zonal polynomials are an eigenfunction of the Laplace Beltrami operator,

$$\Delta = (\det G)^{-\frac{1}{2}} \sum_{k=1}^{n} \frac{\partial}{\partial x_k} (\det G)^{\frac{1}{2}} \sum_{i=1}^{n} g^{ik} \frac{\partial}{\partial x_i}, \tag{1}$$

where $x_1, \cdots, x_n$ are coordinates of a point in a space with metric differential form

$$(ds)^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} dx_i dx_j \;, \quad G = (g_{ij}) \quad \text{and} \quad (g^{ij}) = G^{-1}.$$

On substituting the part of the Laplace Beltrami operator concerned with the roots, we have the operator

$$\Delta = \sum_{i=1}^{m} \left[y_i^2 \frac{\partial^2}{\partial y_i^2} - \frac{1}{2}(m-3) y_i \frac{\partial}{\partial y_i} \right] + \sum_{\substack{i,\, j\, =\, 1 \\ i \neq j}}^{m} \frac{y_i^2}{y_i - y_j} \frac{\partial}{\partial y_i}. \tag{2}$$

In detail we may see in James (1968).

The zonal polynomials of a matrix are defined in terms of partitions of positive integers. Let k be a positive integer. A partition κ of k is written as $\kappa = (k_1, k_2, \ldots)$, where $\sum_i k_i = k$. We order the partitions of k lexicographically, that is, if $\kappa = (k_1, k_2, \ldots)$ and $\lambda = (l_1, l_2, \ldots)$ are two partitions of k, we write $\kappa \succ \lambda$ if $k_i > l_i$ for the first index i for which the parts are unequal. Let $y_1, \cdots, y_m$ be m variables. If $\kappa \succ \lambda$ we say that the monomial $y_1^{k_1} \cdots y_m^{k_m}$ is of higher weight than the monomial $y_1^{l_1} \cdots y_m^{l_m}$.

Let Y be an $m \times m$ symmetric matrix with the characteristic roots $y_1, \cdots, y_m$ and let $\kappa = (k_1, \cdots, k_m)$ be a partition of k into not more than m parts. The zonal polynomial of Y corresponding to κ, denoted by $C_\kappa(Y)$, is a symmetric homogeneous polynomial of degree k in the characteristic roots $y_1, \cdots, y_m$ such that;

(i) The term of the highest weight in $C_\kappa(Y)$ is $y_1^{k_1} \cdots y_m^{k_m}$, that is

$$C_\kappa(Y) = c_{\kappa,\kappa} y_1^{k_1} \cdots y_m^{k_m} + \text{terms of lower weight}, \tag{3}$$

where $c_{\kappa,\kappa}$ is a constant.

(ii) $C_\kappa(Y)$ is an eigenfunction of the differential operator Δ_Y given by

$$\Delta_Y = \sum_{i=1}^{m} y_i^2 \frac{\partial^2}{\partial y_i^2} + \sum_{\substack{i,\, j\, =\, 1 \\ i \neq j}}^{m} \frac{y_i^2}{y_i - y_j} \frac{\partial}{\partial y_i}. \tag{4}$$

(iii) As κ varies over all partitions of k the zonal polynomials have unit coefficients in the expansion of $(\text{tr } Y)^k$, that is

$$(\text{tr } Y)^k = (y_1 + y_2 + \cdots + y_m)^k = \sum_{\kappa} C_\kappa(Y). \tag{5}$$

So the zonal polynomial of Y corresponding to the partition κ satisfies the partial differential equation

$$\Delta_Y C_\kappa(Y) = [\rho_\kappa + k(m-1)] C_\kappa(Y), \tag{6}$$

where $\rho_\kappa = \sum_{i=1}^{m} k_i(k_i - i)$.

The monomial symmetric function $M_\kappa(Y)$ of the matrix argument for a partition κ is a homogeneous polynomial of degree k, defined as

$$M_\lambda(Y) = y_1^{k_1} y_2^{k_2} \cdots y_m^{k_m} + \text{symmetric terms}.$$

On substituting the formula (3) of the zonal polynomial we have the recurrence relation between the coefficients $c_{\kappa,\lambda}$,$c_{\kappa,\mu}$ of the monomial symmetric functions $M_\lambda(Y)$,$M_\mu(Y)$ as follows

$$c_{\kappa,\lambda} = \sum_{\lambda \prec \mu \preceq \kappa} \frac{(l_i + r) - (l_j - r)}{\rho_\kappa - \rho_\lambda} c_{\kappa,\mu} \tag{7}$$

where

$$\rho_\lambda = \sum_{i=1}^{m} l_i(l_i - i), \quad \lambda = (l_1, \cdots, l_m) \quad \text{and} \quad \mu = (l_1, \cdots, l_i + r, \cdots, l_j - r, \cdots, l_m).$$

The elementary symmetric function $\mathcal{E}_\kappa(Y)$ of the matrix argument for a partition κ is also a homogeneous polynomial of degree k, defined as

$$\mathcal{E}_\kappa(Y) = e_1^{k_1 - k_2}\, e_2^{k_2 - k_3} \cdots e_{m-1}^{k_{m-1} - k_m}\, e_m^{k_m},$$

where $e_1 = y_1 + y_2 + \cdots + y_m$, $e_2 = y_1 y_2 + \cdots + y_{m-1} y_m$, $\ldots$, $e_m = y_1 y_2 \cdots y_m$. Hashiguchi *et al.* (2000) have given a transition algorithm from $C_\kappa(Y)$ to $\mathcal{E}_\mu(Y)$, providing the recurrence relations shown in the following proposition between the coefficients $q[\kappa, \mu]$ in

$$C_\kappa(Y) = \sum_{\mu \preceq \kappa} q[\kappa, \mu]\, \mathcal{E}_\mu(Y), \tag{8}$$

where the symbol $\preceq$ signifies the reverse lexicographic ordering.

$$q[\kappa, \mu] = \begin{cases} q_0[\kappa], & \text{if } \kappa = \mu; \\ \dfrac{1}{d[\kappa] - d[\mu]} \displaystyle\sum_{\mu \prec \nu \preceq \kappa} b[\nu, \mu]\, q[\kappa, \nu], & \text{if } \kappa \succ \mu \text{ and } d[\kappa] \neq d[\mu], \\ 0, & \text{otherwise}; \end{cases} \tag{9}$$

where

$$q_0[\kappa] = 2^k\, k! \left(\prod_{i=1}^{m} (2k_i - i + m)!! \right)^{-1} \prod_{1 \leq i < j \leq m} \frac{(2\,k_i - 2\,k_j - i + j)!!}{(2\,k_i - 2\,k_j - i + j - 1)!!} \tag{10}$$

and each $b[\nu, \mu]$ is determined by

$$D_m\, \mathcal{E}_\nu(Y) = \sum_{\mu \preceq \nu} b[\nu, \mu]\, \mathcal{E}_\mu(Y). \tag{11}$$

3. Numerical results

This section is based on Hashiguchi and Niki (2006) that provided the exact computation on the distribution of the smallest characteristic root. Let W be distributed as $W_m(n, \Sigma)$ and let η_n denote the smallest characteristic root of W/n. This statistic might be considered as the most natural estimator of the smallest characteristic root of Σ.

Table 1 shows $\mathrm{E}(\eta_n)$, $\mathrm{SD}(\eta_n)$ and the upper 5% point $x_{.95}(\eta_n)$ for several combinations of m, n, and Σ, with some graphs of densities in Fig. 1.

Estimation of population's smallest characteristic root using η_n from those numerical results might be limited in those cases where the population's smallest characteristic root is sufficiently separated from any other (larger, say) characteristic root of Σ, m is very small and n is sufficiently large. It is noteworthy that, when $\Sigma = \boldsymbol{I}_m$, even upper 5% points are smaller than the population value 1, for $m = 3$, 4, and $n \leq 30$.

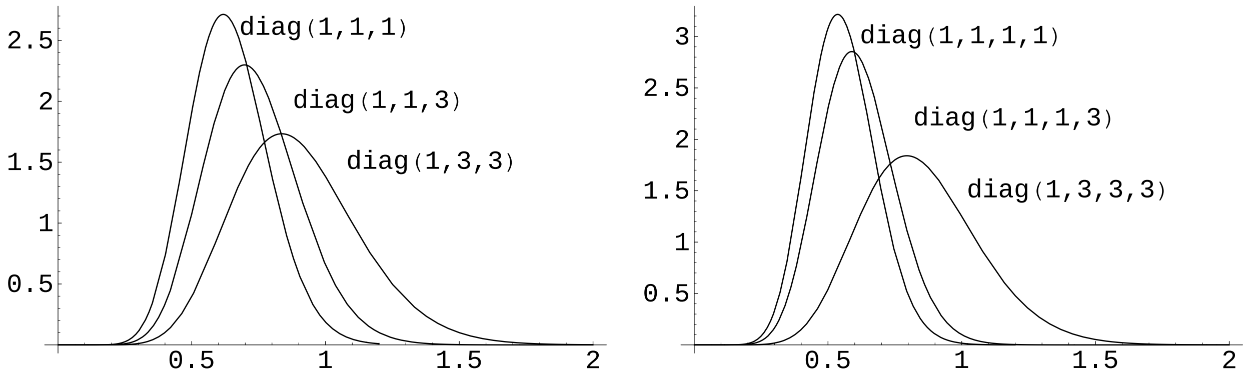
Table 1: Properties of the distribution of η_n.

(a) $\Sigma = \boldsymbol{I}_m$

m	n	$\mathrm{E}(\eta_n)$	$\mathrm{SD}(\eta_n)$	$x_{.95}(\eta_n)$
2	9	0.5937	0.2791	1.1117
	19	0.7162	0.2179	1.1042
	25	0.7518	0.1963	1.0972
	31	0.7767	0.1803	1.0911
3	10	0.4144	0.1904	0.7664
	20	0.5685	0.1665	0.8629
	24	0.6031	0.1582	0.8802
	30	0.6421	0.1477	0.8981
4	9	0.2570	0.1376	0.5154
	19	0.4484	0.1375	0.6917
	25	0.5099	0.1307	0.7378

(b) $\Sigma = \begin{pmatrix} \boldsymbol{I}_{m-1} & \boldsymbol{0} \\ 0 & 3 \end{pmatrix}$

m	n	$\mathrm{E}(\eta_n)$	$\mathrm{SD}(\eta_n)$	$x_{.95}(\eta_n)$
2	9	0.8160	0.3983	1.5604
	19	0.9163	0.3012	1.4605
	25	0.9374	0.2681	1.4168
	31	0.9500	0.2436	1.3827
3	10	0.5042	0.2354	0.9406
	20	0.6624	0.2005	1.0190
	24	0.6951	0.1893	1.0289
	30	0.7308	0.1753	1.0371
4	9	0.2999	0.1615	0.6036
	19	0.5051	0.1574	0.7848
	25	0.5676	0.1486	0.8275



(a) When $m = 3$ and $n = 30$. (b) When $m = 4$ and $n = 31$.

Figure 1: Probability density functions of η_n for several Σ's.

REFERENCES

Hashiguchi, H. and Niki, N. (1997). Algebraic algorithm for calculating coefficients of zonal polynomials of order three, *Journal of the Japanese Society of Computational Statistics*, **10**, 1, 41–46.

Hashiguchi, H., Nakagawa, S. and Niki, N. (2000). Simplification of the Laplace-Beltrami operator, *Mathematics and Computers in Simulation*, **51**, 5, 489–496.

Hashiguchi, H. and Niki, N. (2006) Numerical computation on distributions of the largest and the smallest latent roots of the Wishart Matrix, *Journal of the Japanese Society of Computational Statistics*, **19**, 1, 45–56.

James, A. T. (1960). The distribution of the latent roots of the covariance matrix. *The Annals of Mathematical Statistics*, **31**, 151–158.

James, A. T. (1964). Distribution of matrix variates and latent roots derived from normal samples. *The Annals of Mathematical Statistics*, **35**, 475-497.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* John Wiley, New York.

Sugiyama, T. (1967). On the distribution of the largest latent roots of the covariance matrix. *The Annals of Mathematical Statistics*, **38**, 1148–1151.

Sugiyama, T., Takeda, Y. and Fukuda, M. (1998). On the Numerical computation of confluent hypergeometric function with zonal polynomials of order 3. *Journal of the Japanese Society of Computational Statistics*, **11**, 1–8.

Sugiyama, T., Fukuda, M. and Takeda, Y. (1999). Recurrence relations of coefficients of the generalized hypergeometric function in multivariate analysis. *Communication in Statistics, Theory and Methods*, **28**, 3&4, 825-837.