CONSTANT REGRESSION POLYNOMIALS AND THE WISHART DISTRIBUTION[1,2]

Donald St. P. Richards
Department of Statistics
University of North Carolina
Chapel Hill, NC 27514

[1]Research supported by the National Science Foundation, Grant MCS-8403381.
[2]A part of this work was completed during a recent leave of absence spent at the University of Wyoming.

AMS 1980 Subject Classifications: Primary 62H05, 62H10; Secondary 33A65, 43A80.

Key Words and Phrases: Wishart distribution, constant regression, polynomial statistics, hyperbolic differential operator, unitary representation, generalized Bessel function.

ABSTRACT

Results are obtained for the problems of constructing and characterizing scalar-valued polynomial statistics which have constant regression on the mean of a random sample of Wishart matrices. The construction procedure introduced by Heller (J. Multivariate Anal., 14 (1984), 101-104) is generalized to show that certain polynomials in the principal minors of the sample matrices have zero regression on the mean. The zero regression polynomials are characterized through expectations involving certain matrix-valued Bessel functions of Gross and Kunze (J. Functional Anal, 22 (1976), 73-105).
It is shown that the zero regression property characterizes the Wishart distributions within a wide family of mixtures of Wishart distributions.

1. Introduction

Suppose that $X_1, X_2, \ldots, X_n$ are independent, identically distributed $m \times m$ random matrices, each having the Wishart distribution $W_m(\Sigma, N)$ with positive definite (symmetric) covariance matrix Σ and N degrees of freedom. A scalar-valued polynomial statistic $P(\underset{\sim}{X})$, $\underset{\sim}{X}=(X_1, \ldots, X_n)$, is said to have constant regression on $L=X_1 +\ldots+ X_n$ if the conditional expectation

$$E(P(\underset{\sim}{X})|L) = \beta \tag{1.1}$$

almost everywhere, for some constant β. Without loss of generality, we may suppose that $\beta = 0$ in which case the polynomial $P(\underset{\sim}{X})$ is called a zero regression polynomial. In this paper, we consider the problems of constructing and characterizing the zero regression polynomials.

In the classical case, Lukacs and Laha [8] obtained necessary and sufficient conditions for a quadratic polynomial statistic to have zero regression. Their results have been presented, within the more general context of characterizations of probability distributions, by Kagan, Linnik and Rao [7]. Although these problems were originally motivated by a purely academic desire to determine the theoretical distribution of the parent population from hypothetical properties of a particular statistic, recent applications [9] include the construction of testing procedures.

Recently, Heller [4] has constructed zero regression polynomials for the Wishart distribution by appropriately generalizing the methods of [8]. Her results are among a small number which extend the classical developments to the setting of matrix distributions.

The main tool used in [4] is the hyperbolic differential operator of Herz [5]. Using more complex operators, we generalize (in Section 2) the construction procedure of [4]. As a consequence, we find that certain polynomials in the principal minors of the sample matrices are zero regression polynomials.

Section 3 presents several characterizations of the zero regression polynomials using unconditional expectations. The main result characterizes the zero regression polynomials through expectations involving the matrix-valued Bessel functions of Gross and Kunze [2]. This result presents a new and unexpected link between harmonic analysis and multivariate statistical theory; further, it demonstrates the full complexity of the zero regression problem in higher dimensions.

Following on the results of [4], [8], a natural question is: Given that a polynomial $P(\underset{\sim}{X})$ has zero regression on L, is the underlying distribution necessarily Wishart? We obtain a partial answer, showing that if the parent distribution belongs to a wide class of Wishart mixtures, then the distribution is necessarily Wishart if $P(\underset{\sim}{X})$ is any one of a family of polynomials. In particular, this class of polynomials contains all the examples constructed in [4].

As a final comment, we remark that since the homogeneous zero regression polynomials with fixed degree form a finite dimensional vector space, it would perhaps be more useful to determine the dimension of the space along with a "natural" basis. Presently, no non-trivial results are available for these questions.

2. Construction of Zero Regression Polynomials.

A partition $\kappa=(k_1,k_2,\ldots,k_m)$ is a vector of non-negative integers, with $k_1 \geq k_2 \geq \ldots \geq k_m$. For any complex number t,

$$(t)_\kappa = \prod_{i=1}^{m} (t-\tfrac{1}{2}(i-1))_{k_i}$$

where

$$(t)_j = \Gamma(t+j)/\Gamma(t), \quad j=0,1,2,\ldots$$

Further

$$\Gamma_m(t) = \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma(t-\tfrac{1}{2}(i-1)), \; \mathrm{Re}(t) > \tfrac{1}{2}(m-1).$$

For any symmetric $m \times m$ matrix $T = (t_{ij})$, we define

$$|T|_\kappa = \prod_{i=1}^{m} |T_i|^{k_i-k_{i+1}} \quad , \quad k_{m+1} \equiv 0,$$

where T_i is the i-th principal minor of T, and $|\cdot|$ denotes determinant. With the matrix T, we associate a matrix of differential operators

$$\frac{\partial}{\partial T} := (\tfrac{1}{2}(1+\delta_{ij}) \frac{\partial}{\partial t_{ij}}) \quad ,$$

where δ_{ij} is Kronecker's delta. Whenever the variables of differentiation are clear from the context, we denote $\partial/\partial T$ by D; further, we shall use the notation

$$|D|_\kappa = \prod_{i=1}^{m} |D_i|^{k_i-k_{i+1}}$$

where D_i is the i-th principal minor of D. For example, if $\kappa = (4,2,0,\ldots,0)$, then

$$|D|_\kappa = \frac{\partial^2}{\partial t_{11}^2} (\frac{\partial^2}{\partial t_{11}\partial t_{22}} - \frac{\partial^2}{4\partial t_{12}^2})^2.$$

The operator $|D| \equiv |D|_{(1,1,\ldots,1)}$ is the hyperbolic operator of Herz [5].

If a random matrix X_1 has the $W_m(\Sigma, N)$ distribution then it is known that the characteristic function of X_1 is

$$\phi(T) := E(e^{i\,\mathrm{tr}(TX_1)}) = |I-2iT\Sigma|^{-N/2} \quad ,$$

where T is a symmetric matrix. Our first result determines the effect of the $|D|_\kappa$ operators on $\phi(T)$.

2.1. <u>Proposition</u>. $|D|_\kappa\, \phi(T) = i^k (N/2)_\kappa\, |(\tfrac{1}{2}\Sigma^{-1}-iT)^{-1}|_\kappa\, \phi(T)$, where $k = k_1 + k_2 + \ldots + k_m$.

In the course of proving Proposition 2.1., we will use the following result which is due to Bellman [1].

2.2 <u>Lemma</u>. Let Z be a complex, symmetric $m \times m$ matrix with $\mathrm{Re}(Z) > 0$ (positive definite symmetric). Then,

$$\int_{X>0} e^{-\mathrm{tr}(XZ)}\, |X|^{t-p}\, |X|_\kappa\, dX = (t)_\kappa\, \Gamma_m(t)\, |Z|^{-t}\, |Z^{-1}|_\kappa \quad ,$$

with absolute convergence for $\mathrm{Re}(t) > p-1$, $p=(m+1)/2$.

<u>Proof of Proposition 2.1</u>. Let

$$C = (|2\Sigma|^{N/2}\, \Gamma_m(N/2))^{-1}$$

be the normalizing constant for the Wishart density. Then,

$$|D|_\kappa \phi(T) = C\, |D|_\kappa \int_{X>0} e^{i\,\mathrm{tr}(TX)} e^{-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}X)} |X|^{(N-m-1)/2} dX.$$

Since

$$|D|_\kappa\, e^{i\,\mathrm{tr}(TX)} = i^k |X|_\kappa\, e^{i\,\mathrm{tr}(TX)} \quad ,$$

then

$$|D|_\kappa \phi(T) = C\, i^k \int_{X>0} e^{-\mathrm{tr}(\frac{1}{2}\Sigma^{-1}-iT)X} |X|^{(N-m-1)/2} |X|_\kappa dX.$$

Applying Lemma 2.2 and simplifying completes the proof.

If $\kappa = (k_1,\ldots,k_m)$ and $\lambda = (l_1,\ldots,l_m)$ are partitions, we define $\kappa+\lambda = (k_1+l_1,\ldots k_m+l_m)$.

2.3 Theorem. Let $\{a_{ij}\}_{i,j=1}^{n}$ be real numbers with $\sum\sum_{i\neq j} a_{ij}\neq 0$. For any two partitions κ,λ define

$$b_{\kappa,\lambda} = (N/2)_\kappa (N/2)_\lambda / (N/2)_{\kappa+\lambda}.$$

If $X_1,\ldots,X_n$ is a random sample from $W_m(\Sigma,N)$, then the polynomial

$$P(\underset{\sim}{X}) = \sum_{i\neq j}^{n}\sum^{n} a_{ij}[|X_i|_\kappa \; |X_j|_\lambda - b_{\kappa,\lambda}\; |X_i|_{\kappa+\lambda}] \tag{1}$$

is a zero regression polynomial.

Proof. A necessary and sufficient criterion [4] for a polynomial $P(\underset{\sim}{X})$ to have zero regression on L is that

$$E(e^{i\;tr(TL)}\; P(\underset{\sim}{X})) = 0 \tag{2}$$

for all symmetric T. In order to apply (2) to the polynomial (1), we shall use the results:

$$|D|_\kappa\; \phi(T) = i^k\; E(e^{i\;tr(TX)}\; |X|_\kappa),$$

$$|D|_\kappa |D|_\lambda\; \phi(T) = i^{k+l}\; E(e^{i\;tr(TX)}\; |X|_\kappa |X|_\lambda),$$

where $k = k_1+\ldots+k_m$, $l = l_2+\ldots+l_m$. Then, applying (2) to (1), we obtain

$$E(e^{i\;tr(TL)}\; P(\underset{\sim}{X})) =$$

$$i^{-k-l} \sum_{r\neq s}\sum a_{rs}(\phi(T))^{n-2}\; [(|D|_\kappa\; \phi(T))(|D|_\lambda\; \phi(T)) - b_{\kappa,\lambda}\; \phi(T)|D|_{\kappa+\lambda}\; \phi(T)]$$

However, from Proposition 2.1, we also have

$$(|D|_\kappa \phi)(|D|_\lambda \phi) - b_{\kappa,\lambda} \phi |D|_{\kappa+\lambda} \phi \equiv 0,$$

and hence $P(\underset{\sim}{X})$ has zero regression on L.

The construction given in Theorem 3 leads to a generalization of the results in [4]. To recover the zero regression polynomials of [4] from our result, we need only set $\kappa = \lambda = (1^m)$; this shows that

$$P(\underset{\sim}{X}) = \sum_{i \neq j}\sum a_{ij}[|X_i||X_j| - b_1 |X_i|^2], \qquad (3)$$

has zero regression on L, where $\sum_{i \neq j}\sum a_{ij} \neq 0$ and

$$b_1 = (N/2)_{(1^m)}(N/2)_{(1^m)}/(N/2)_{(2^m)}.$$

3. Characterizations of Zero Regression Polynomials.

This section characterizes the zero regression polynomials through several unconditional expectations, including generalizations of (2). First, we need to introduce notation and results pertinent to Fourier analysis on compact topological groups; an introductory account of this theory is given in [12].

Let G be a compact (Hausdorff) group. A representation ρ of G is a continuous homomorphism of G into the group of invertible linear transformations on a complex vector space V_ρ. Since G is compact, then V_ρ is necessarily of finite dimension, say d_ρ. Further, we lose no generality in assuming that the representations of G are given by unitary matrices.

A representation ρ is irreducible if the only proper invariant (under ρ) subspace of V_ρ is {0}, the trivial subspace.

Two representations ρ_1 and ρ_2 are unitarily equivalent if there exists a unitary matrix u such that $u\rho_1(g) = \rho_2(g)u$ for all g in G. The notion of unitary equivalence defines an equivalence relation among the representations of G. We denote by $\hat{G}$ the set of equivalence classes of irreducible representations of G. $\hat{G}$ is called the dual of G.

Let dg be the unique Haar measure on G, normalized so that G has total mass one. If $f: G \to \mathbb{C}$ is continuous, the Fourier transform of f is defined by

$$\rho(f) = \int_G f(g)\rho(g)dg \quad , \quad \rho \in \hat{G}.$$

The Fourier inversion formula then states that

$$f(g) = \sum_{\rho \in \hat{G}} d_\rho \, tr(\rho(g)^*\rho(f)) \quad , \quad g \in G, \tag{4}$$

where * denotes the transpose of complex conjugate. The sum in (4) is over a set of representatives for the equivalence classes in $\hat{G}$, one for each class. Further, (4) converges pointwise on G.

Next, we specialize to the case $G = SO(m)$, the special orthogonal group. The group G consists of all $m \times m$ orthogonal matrices g having determinant one.

3.1. Definition ([2], [3]). For symmetric $m \times m$ matrices S,T,

$$J_\rho(S,T) = \int_{SO(m)} e^{i\,tr(g'Sg\,T)}\,\rho(g)dg$$

is the generalized Bessel function of order ρ.

The generalized Bessel functions were introduced and studied in [2]. In the jargon of [2], [3] $J_\rho(S,T)$ is the generalized Bessel function of order ρ arising from the two-sided action of $SO(m)$ on the space of symmetric matrices.

Note that $J_\rho(S,T)$ is the Fourier transform of the function $g \to e^{i\,tr(g'Sg\,T)}$, $g \in SO(m)$. Further, $J_\rho(S,T)$ is a $d_\rho \times d_\rho$ matrix. The Fourier inversion formula (4) shows that for fixed S,T,

$$e^{i\,tr(g'Sg\,T)} = \sum_{\rho\in\hat{G}} d_\rho\,tr(\rho(g)^*\,J_\rho(S,T)). \tag{5}$$

Since the function $g \to e^{i\,tr(g'Sg\,T)}$ is infinitely differentiable, the series (5) converges uniformly on $SO(m)$.

Now, we can state the main result of the paper. As before $L = X_1 + \ldots + X_n$ where the X_i are i.i.d. Wishart matrices. Also, the polynomial $P(\underset{\sim}{X})$, $\underset{\sim}{X} = (X_1,\ldots,X_n)$, has zero regression on L.

3.2. Theorem. The following conditions are equivalent:

(a) $P(\underset{\sim}{X})$ has zero regression on L;

(b) $E(e^{i\,tr(TL)}p(L)P(\underset{\sim}{X})) = 0$ for all polynomials $p(L)$ and symmetric matrices T;

(c) $E(p(L)P(\underset{\sim}{X})) = 0$ for all polynomials $p(L)$;

(d) $E(P(\underset{\sim}{X})J_\rho(T,L)) = 0$ for all $\rho \in (SO(m))^\wedge$ and symmetric T.

Proof. In Section 2, we noted that the zero regression property is equivalent to

$$E(e^{i\ \mathrm{tr}(TL)}P(\underset{\sim}{X})) = 0 \tag{6}$$

for all T. Then, it is immediate that (a) ⇒ (b) since

$$E(e^{i\ \mathrm{tr}(TL)}p(L)P(\underset{\sim}{X})) = p(\tfrac{1}{i}\tfrac{\partial}{\partial T})\ E(e^{i\ \mathrm{tr}(TL)}P(\underset{\sim}{X})).$$

Further, (b) ⇒ (6) trivially, so that (a) and (b) are equivalent. To see that (c) ⇒ (6), note that $e^{i\ \mathrm{tr}(TL)}$ can be expressed as an absolutely convergent series,

$$e^{i\ \mathrm{tr}(TL)} = \sum_{\alpha \in A} q_\alpha(T)p_\alpha(L)$$

where A is a countable index set, and p_α and q_α are polynomials for all α. Taking expectations shows that

$$E(e^{i\ \mathrm{tr}(TL)}P(\underset{\sim}{X})) = \sum_{\alpha \in A} q_\alpha(T)\ E(p_\alpha(T))P(\underset{\sim}{X})) = 0,$$

where the interchange of summation and expectation is permitted by the Dominated Convergence Theorem. Since (b) ⇒ (c), then we have established the equivalence of (a), (b) and (c).

Finally, we prove that (6) ⇔ (d). First, Fubini's theorem and the definition of $J_\rho(S,T)$ show that

$$E(P(\underset{\sim}{X})J_\rho(S,T)) = \int_{SO(m)} E(e^{i\ \mathrm{tr}(g'Tg\ L)}P(\underset{\sim}{X}))\rho(g)dg.$$

Therefore, (6) ⇒ (d). Conversely if (d) holds, then by the Fourier inversion formula (5) and the Dominated Convergence Theorem,

$$E(e^{i\ \mathrm{tr}(g'Tg\ L)}P(\underset{\sim}{X})) = \sum_\rho d_\rho \mathrm{tr}(\rho(g)^*\ E(P(\underset{\sim}{X})J_\rho(T,L))) = 0,$$

for any g in SO(m) and symmetric T. Hence (d) ⇒ (6), and this completes the proof.

On reviewing the proof of Theorem 3.2, we notice that nowhere do we use the fact that the X_i are Wishart matrices. Therefore the result holds for _any_ symmetric matrix distribution, as long as the various interchanges of sums and integrals remain valid. In particular, it holds for the multivariate beta and F distributions [6].

To end this section, we make some remarks on the generalized Bessel functions. The _spherical Bessel function_ is

$$J_0(S,T) = \int_{SO(m)} e^{i\,tr(g'Sg\,T)}dg,$$

which corresponds to the trivial one-dimensional representation $\rho(g) \equiv 1$. When $m=2$, it can be shown using the techniques of [3] that

$$J_0(S,T) = e^{i\,tr(S)tr(T)/2}\,\tilde{J}_0(\Delta(S)\Delta(T)/2)$$

where $\tilde{J}_0(\cdot)$ denotes the classical Bessel function of the first kind of order zero, and $\Delta(S) = |s_1-s_2|$ where s_1 and s_2 are the two eigenvalues of S. For any odd $m \geq 3$, it has been shown [11] that $J_0(S,T) = {}_0F_0(iS,T)$, where ${}_0F_0(\cdot,\cdot)$ is James' hypergeometric function of two matrix arguments. For even $m \geq 4$, $J_0(\cdot,\cdot)$ can be again related to ${}_0F_0(\cdot,\cdot)$ but the result is more complicated. For general ρ the evaluation of the matrix elements of $J_\rho(S,T)$ requires powerful machinery from the theory of harmonic analysis [2], [3]; however, we also remark that for the case $m=2$, $J_\rho(S,T)$ can be computed explicitly using the techniques of [3].

4. A Characterization of the Wishart Distribution.

Both here and in [4], it has been shown that the polynomial (3) has zero regression on $L = X_1 + \ldots + X_n$. Given the converse, that (3) has zero regression of the sum of an i.i.d. sequence of random matrices, we wish to determine whether the underlying population is necessarily Wishart. From the proof of Theorem 2.3, we observe that the polynomial (3) has zero regression on L if and only if the characteristic function $\phi(T) = E(e^{i\,tr(TX_1)})$ satisfies the differential equation

$$(|D|\phi(T))^2 = b_1\,\phi(T)\,|D|^2\,\phi(T), \tag{7}$$

for all T. Therefore, it is enough to find all solutions of (7) which are simultaneously characteristic functions.

Consider the class of characteristic functions $\phi(T)$ of the form

$$\phi(T) = \int_H |I-2i\,T\,\Sigma|^{-N/2}\,d\nu(\Sigma) \tag{8}$$

where $\nu(\cdot)$ is a Borel probability measure on the hypersurface $H = \{\Sigma : \Sigma$ is positive definite and $|\Sigma| = 1\}$. A random matrix with characteristic function of the form (8) may be regarded as a "covariance mixture" of Wishart distributions; these mixtures have arisen [10] in the context of hyperspherical distributions. We now show that the zero regression polynomial (3) characterizes the Wishart distribution within the class of Wishart mixtures typified by (8).

4.1. Theorem. Let $X_1,\ldots,X_n$ be a random sample from a population with characteristic function (8). If the polynomial $P(\underset{\sim}{X})$ in (3) has zero regression on L, then X_1 has a Wishart distribution.

Proof. Since $P(\underset{\sim}{X})$ has zero regression on L, then (7) holds. Applying the $|D|$ and $|D|^2$ operators to $\phi(T)$ in (8) shows that

$$(\int_H |I-2iT\Sigma|^{-(N+2)/2} d\nu(\Sigma))^2 = \int_H |I-2iT\Sigma|^{-N/2} d\nu(\Sigma) \cdot \int_H |I-2iT\Sigma|^{-(N+4)/2} d\nu(\Sigma), \quad (9)$$

for all symmetric T. By considering discrete approximations to these latter integrals, it is evident that (9) implies $\nu(\cdot)$ to be the Dirac measure at some "point" Σ_o. Consequently, $X_1 \sim W_m(N,\Sigma_o)$.

Stephen Marron has kindly remarked that the above conclusion concerning $\nu(\cdot)$ can be also obtained by working with moment generating functions rather than characteristic functions. Indeed, the generating function analogue of (9) is

$$(\int_H |I-2T\Sigma|^{-(N+2)/2} d\nu(\Sigma))^2 = \int_H |I-2T\Sigma|^{-N/2} d\nu(\Sigma) \cdot \int_H |I-2T\Sigma|^{-(N+4)/2} d\nu(\Sigma), \quad (9')$$

and this holds for all T in a sufficiently small neighbourhood of the zero matrix. The Cauchy-Schwartz inequality guarantees that the left-hand-side of (9') is never larger than the right-hand-side. Then, the conclusion that $\nu(\cdot)$ is Dirac follows immediately from the criterion for equality in the Cauchy-Schwartz inequality.

Finally, we remark that results similar to Theorem 4.1 can be obtained for the general polynomials constructed in Theorem 2.3.

Acknowledgements. I am deeply grateful to Kenneth Gross for his patient explanations of harmonic analysis in general, and the generalized Bessel functions in particular.

REFERENCES

1. R.E. Bellman (1956). A generalization of some integral identities due to Ingham and Siegel. Duke Math. J., 23, 571-577.

2. K.I. Gross and R.A. Kunze (1976). Bessel functions and representation theory. J. Functional Anal., 22, 73-105.

3. K.I. Gross, W.J. Holman, III, and R.A. Kunze (1979). A new class of Bessel functions and applications in harmonic analysis. Proc. Symposia in Pure Math., XXXV, Amer. Math. Soc., 407-415.

4. B. Heller (1984). Use of the hyperbolic operator to find a scalar statistic which has constant regression on the mean of a sample of Wishart matrices. J. Multivariate Anal., 14, 101-104.

5. C.S. Herz (1955). Bessel functions of matrix argument. Ann. Math., 61, 474-523.

6. A.T. James (1964). Distributions of matrix variates and latent roots derived from normal samples. Ann. Math. Statist., 35, 475-501.

7. A.M. Kagan, Y.V. Linnik, and C.R. Rao (1973). Characterization Problems in Mathematical Statistics. Wiley, New York.

8. E. Lukacs and R.G. Laha (1964). Applications of Characteristic Functions. Griffin, London.

9. G.S. Mudholkar and C.C. Lin (1977). A simple test for normality. Bull. Inst. Math. Statist. 6, 273-274.

10. D. St. P. Richards (1984). Hyperspherical models, fractional derivatives and exponential distributions on matrix spaces. Sankhyā, Series A, 46, (to appear).

11. D. St. P. Richards (1984). Solution to problem 84-1: An integral on SO(3). SIAM Review, (to appear).

12. G. Weiss (1976). Harmonic analysis on compact groups, in Studies in Harmonic Analysis (J.M. Ash, ed.) MAA, Providence, 198-223.