MacMahon’s Master Theorem, Representation Theory, and Moments of Wishart Distributions

I-Li Lu

Medical College of Virginia

and

Donald St. P. Richards

School of Mathematics, Institute for Advanced Study, Princeton, New Jersey; and
Department of Statistics, University of Virginia, Charlottesville, Virginia

Received January 16, 2001; accepted March 28, 2001

DEDICATED TO DOMINIQUE FOATA ON THE OCCASION OF HIS 65TH BIRTHDAY

D. Foata and D. Zeilberger (1988, SIAM J. Discrete Math. 4, 425–433) and
a generalization of MacMahon’s master theorem. In this article we apply their
result to obtain an explicit formula for the moments of arbitrary polynomials in the
entries of $X$, a real random matrix having a Wishart distribution. In the case of
the complex Wishart distributions, the same method is applicable. Furthermore, we
apply the representation theory of $\text{GL}(d, \mathbb{C})$, the complex general linear group, to
derive explicit formulas for the expectation of Kronecker products of any complex
Wishart random matrix. © 2001 Elsevier Science

Key Words: complex Wishart distribution; Laplace-Beltrami operator; invariant
differential operator; irreducible representation; random matrix; Schur function;
Schur’s lemma; Weyl’s unitary trick; zonal polynomial.

1Research supported in part by a grant to the Institute for Advanced Study from the Bell
Fund and by the National Science Foundation under Grant DMS-9703705.
1. INTRODUCTION

The Wishart distributions are among the most prominent probability distributions in multivariate statistical analysis, arising naturally in applied research and as a basis for theoretical models. We refer to Anderson [1], among others, for extensive literature on the theory and applications of the Wishart distributions.

Because of widespread use of these distributions, a recurring problem in multivariate statistics is the calculation of moments of the Wishart distributions. This problem has generated a large body of literature, cf. [8, 13–15, 17, 20], for different approaches to this problem. These articles also contain many additional references to the literature.

Throughout the paper, \( d \) will denote a fixed positive integer. We let \( \Sigma \) be a fixed \( d \times d \) positive-definite symmetric matrix, denoted \( \Sigma > 0 \), and we denote by \( \mathbb{R}^{d \times d}_+ \) the space of positive-definite \( d \times d \) matrices. For any \( d \times d \) matrix \( g \), we also denote by \( \det(g) \) and \( \text{tr}(g) \) the determinant and trace, respectively, of \( g \).

Let \( X = (X_{i,j}) \) be a real, symmetric, positive-definite, \( d \times d \), random matrix. Furthermore, let \( \beta \in \mathbb{R} \) with \( \beta > (d - 1)/2 \) and let \( \Sigma \in \mathbb{R}^{d \times d}_+ \). The random matrix \( X \) will be said to have a Wishart distribution with parameters \( \beta \) and \( \Sigma \) if its probability density function, relative to Lebesgue measure on the space \( \mathbb{R}^{d \times d}_+ \), is

\[
f(x) = \frac{1}{\Gamma_d(\beta)} (\det \Sigma)^{-\beta} (\det x)^{\beta - (d + 1)/2} \exp(-\text{tr} \Sigma^{-1} x),
\]

where the multivariate gamma function is defined by

\[
\Gamma_d(\beta) := \pi^{d(d-1)/4} \prod_{j=1}^d \Gamma\left(\beta - \frac{1}{2}(j - 1)\right).
\]

Whenever (1.1) holds we shall write, simply, \( X \sim W_d(\beta, \Sigma) \).

Let \( P \) be a polynomial function on \( \mathbb{R}^{d \times d}_+ \). In many instances the evaluation of \( \mathbb{E}P(X) \), the expectation of \( P(X) \) with respect to the Wishart density function in (1.1), can be calculated using harmonic analysis on the space of positive-definite matrices. For example, Richards [18] applied harmonic analysis on the space of positive-definite symmetric matrices to calculate these expectations for cases in which the polynomial \( P \) is an eigenfunction of certain “invariant” differential operators. At the heart of these results is a fundamental result due to Selberg [22], who proved that if the function \( P(X) \) is an eigenfunction of the invariant differential operators in \( X \), then \( P(X) \) is also an eigenfunction of certain invariant integral operators. Thus, once it is observed (cf. [18]) that these integral operators include the
operator of expectation with respect to the Wishart distributions, we then obtain the value of those expectations.

A general example following from Selberg’s theory is the case in which $P(X)$ is a zonal polynomial [10, 16], usually denoted by $C_n(X)$. In this case, it is well known that $P(X)$ is an eigenfunction of the Laplace–Beltrami operator on $\mathbb{R}^{d\times d}_+$, and then it follows as a direct consequence that

$$EC_n(X) = c_n(\beta, d)C_n(\Sigma),$$

(1.3)

where the constant $c_n(\beta, d)$ depends on $\kappa$, $d$, and $\beta$, but not on $\Sigma$. The constant $c_n(\beta, d)$ may be written explicitly in terms of $\Gamma_d(\cdot)$, the multivariate gamma function in (1.2). We refer the interested reader to [16] for a comprehensive account of applications of zonal polynomials and to [7] for further details on the connections with harmonic analysis on matrix spaces.

The problem of calculating Wishart moments is also linked directly to the theory of group representations. In Section 3, we note the relevance of results of [5] for the calculation of $EP(X)$ for the case in which $P$ is a multiplicative function on $\mathbb{R}^{d\times d}_+$. In this paper, our main consideration is a new approach to evaluating the expectations $EP(X)$ when $P$ is an arbitrary polynomial. We shall apply to this problem an extension of MacMahon’s master theorem, derived independently by Foata and Zeilberger [3] and Vere-Jones [23]. Foata and Zeilberger [3] have made applications to combinatorial problems and Vere-Jones [24] gave applications to problems in statistics and probability.

In the literature, the evaluation of $EP(X)$ for a given polynomial $P$ of degree $k$ often has been done inductively, requiring knowledge of the expectations $EQ(X)$ for all polynomials $Q$ of degree less than $k$. By applying the extension of MacMahon’s master theorem, however, we find that inductive considerations can be circumvented. Indeed, we find that the evaluation of $EP(X)$ can be carried out directly using a symbolic manipulation computer software package, e.g., MAPLE. These results are given in Section 3.

In Section 4, we consider the problem of calculating the moments of the complex Wishart matrix, a random matrix taking values in $\mathbb{C}^{d\times d}_+$, the space of Hermitian, positive-definite, $d \times d$ matrices, and having the probability density function (4.1). Here again we can evaluate the expectations of arbitrary polynomials by application of the extension of the master theorem. However, by applying the representation theory of $GL(d, \mathbb{C})$, we obtain results which are even more explicit. Not only do we derive explicit formulas for the expectation of arbitrary tensor (or Kronecker) products of $X$, but we also obtain equally explicit formulas for the expected values of arbitrary tensor (or Kronecker) products of $X^{-1}$. These latter results are complete solutions to the problems of calculating expectations of polynomials of Wishart and inverse Wishart matrices.
2. THE $\beta$-EXTENSION OF THE MASTER THEOREM

We denote by $S_k$ the group of permutations on $k$ symbols, $k \in \mathbb{N}$. For any permutation $w \in S_k$, we also denote by $\text{cyc}(w)$ the number of disjoint cycles of the permutation $w$.

Let $A = (a_{i,j})$ be a real or complex $k \times k$ matrix. We shall denote the norm of $A$ by $\|A\| := \max \{ |\theta_1|, \ldots, |\theta_k| \}$ where $\theta_1, \ldots, \theta_k$ are the eigenvalues of $A$.

For $\beta \in \mathbb{C}$, the $\beta$-permanent of the $k \times k$ matrix $A$ is defined to be the function

$$\text{per}_\beta(A) := \sum_{w \in S_k} \beta^{\text{cyc}(w)} \prod_{j=1}^{k} a_{j,w(j)}.$$  \hspace{1cm} (2.1)

As observed by Vere-Jones [23], it follows directly from (2.1) that $\text{per}_{-1}(A) = (-1)^k \det(A)$, so that $\text{per}_{-1}(\cdot)$ is essentially the determinant function. Also $\text{per}_1(A) = \text{per}(A)$, the classical permanent function. It is also clear that, for any $c \in \mathbb{R}$, $\text{per}_\beta(cA) = c^k \text{per}_\beta(A)$; i.e., $\text{per}_\beta$ is homogeneous of degree $k$.

Suppose $T = (t_{i,j})$ is a $d \times d$ matrix, $k \in \mathbb{N}$, and $i_1, \ldots, i_k$ are positive integers with $1 \leq i_r \leq d$ for all $r = 1, \ldots, k$. Then we denote by $T(i_1, \ldots, i_k)$ the $k \times k$ matrix with $(r,s)$th entry $t_{i_r,i_s}$; i.e.,

$$T(i_1, \ldots, i_k) = \begin{pmatrix} t_{i_1,i_1} & \cdots & t_{i_1,i_k} \\ \vdots & \ddots & \vdots \\ t_{i_k,i_1} & \cdots & t_{i_k,i_k} \end{pmatrix}. \hspace{1cm} (2.2)$$

The $\beta$-extension of the master theorem [3, 23] states that if $Z = \text{diag}(z_1, \ldots, z_d)$ and $T = (t_{i,j})$ is a $d \times d$ matrix such that $\|ZT\| < 1$ then

$$\det(I_d - ZT)^{-\beta} = 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \times \sum_{i_1,\ldots,i_k=1}^{d} \text{per}_\beta(T(i_1, \ldots, i_k)) z_{i_1} z_{i_2} \cdots z_{i_k}. \hspace{1cm} (2.3)$$

In the inner sum in (2.3), the integers $i_1, \ldots, i_k$ are allowed to appear with repetition. For the case in which $-\beta \in \mathbb{N}$, the series (2.3) terminates because $\text{per}_\beta(T(i_1, \ldots, i_k)) \equiv 0$ for sufficiently large $k$. 


3. MOMENTS OF THE WISHART DISTRIBUTIONS

Let $X$ be a random matrix having the probability density function in (1.1). Consider the problem of calculating $\mathbb{E}P(X)$, the expectation of a monomial

$$P(X) = \prod_{1 \leq i \leq j \leq d} X_{i,j}^{k_{i,j}},$$

where the $k_{i,j}$ are given nonnegative integers. By applying the well-known transformation

$$X = \Sigma^{1/2} Y \Sigma^{1/2},$$

we find that $Y \sim W_d(\beta, \Sigma)$ and also that $P(X)$ is transformed to a linear combination of monomials in $Y$. By linearity of the expectation operator we may henceforth assume, with no loss of generality, that $\Sigma = I_d$.

**Definition 3.1.** Let $\{k_{i,j} : 1 \leq i < j \leq d\}$ be a given set of nonnegative integers, $k := \sum_{i \leq j} k_{i,j}$, and $A = (a_{i,j})$ be any symmetric $k \times k$ matrix. For $i_1, \ldots, i_k \in \{1, \ldots, d\}$, we shall say that the $k$-tuple $(i_1, \ldots, i_k)$ is admissible if there exists a permutation $\tau \in S_k$ such that, as polynomials in the matrix $A$,

$$\prod_{l \leq j} a_{i_l,j}^{k_{i_l,j}} = \prod_{j=1}^{k} a_{i_{\tau(j)},i_j},$$

Whenever (3.3) holds, we say that the permutation $\tau$ is $(i_1, \ldots, i_k)$-admissible.

For each $k$-tuple $(i_1, \ldots, i_k)$, we denote by $\mathcal{A}(i_1, \ldots, i_k)$ the set of all $\tau \in S_k$ which are $(i_1, \ldots, i_k)$-admissible. Furthermore, we denote by $\mathcal{A}$ the set of all admissible $k$-tuples $(i_1, \ldots, i_k)$.

Now we have the following combinatorial formula for the moments of monomials of Wishart-distributed matrices (cf. [8] for combinatorial expressions for moments of eigenvalues of random matrices).

**Theorem 3.2.** Suppose that $X = (X_{i,j}) \sim W_d(\beta, I_d)$. Then

$$\mathbb{E} \prod_{1 \leq i \leq j \leq d} X_{i,j}^{k_{i,j}} = \frac{2^{-p}}{k!} \left( \prod_{1 \leq i \leq j \leq d} k_{i,j}! \right) \sum_{(i_1, \ldots, i_k) \in \mathcal{A}} \sum_{w \in \mathcal{A}(i_1, \ldots, i_k)} \beta^{\text{cyc}(w)},$$

where $k := \sum_{i \leq j} k_{i,j}$, $p := \sum_{i < j} k_{i,j}$.

**Example 3.3.** Consider the case in which $d \geq 2, k_{1,1} = 1, k_{1,2} = 2$, and $k_{i,j} = 0$ for all other $i, j$. Then $k = 3$, and there are $d^3$ possible 3-tuples which arise in (3.4), viz., $(i_1, i_2, i_3)$ with each $i_j \in \{1, \ldots, d\}$. However, it is straightforward to show that only three of these are admissible; in fact,
\[ O \cap \{1, 1, 2\} \cap \{1, 2, 1\} \cap \{2, 1, 1\} \]. This implies, with no loss of generality, that we may assume \( d = 2 \). Then the sets \( O(i_1, i_2, i_3) \) are

\[ O(1, 1, 2) = \{(13), (23), (123), (132)\}, \]
\[ O(1, 2, 1) = \{(12), (23), (123), (132)\}, \]

and

\[ O(2, 1, 1) = \{(12), (13), (123), (132)\}, \]

and we obtain from Theorem 3.2 the result

\[ E(X_{1,1}X_{1,2}^2) = 4\beta(\beta + 1). \]

As we observe from this example, and more generally from Theorem 3.4, the moments of monomials in the entries of \( X \) are polynomials in \( \beta \). It is an open problem to characterize the roots of this polynomial.

It is also interesting to observe in the foregoing example that the admissible \( k \)-tuples are permutations of each other. As the following result shows, such a phenomenon is to be expected.

**Proposition 3.4.** Suppose that the \( k \)-tuple \( (i_1, \ldots, i_k) \) is admissible and \( \tau \in O(i_1, \ldots, i_k) \). Then,

(i) For every \( w \in S_k \), the \( k \)-tuple \( w \cdot (i_1, \ldots, i_k) := (i_{w(1)}, \ldots, i_{w(k)}) \) is admissible.

(ii) For each \( w \in S_k \), \( w^{-1} \tau w \in O(w \cdot (i_1, \ldots, i_k)) \).

(iii) \( \tau^{-1} \in O(i_1, \ldots, i_k) \).

**Proof.** Since \( \tau \in O(i_1, \ldots, i_k) \) then, as polynomials in the \( k \times k \) matrix \( A = (a_{i,j}) \),

\[ \prod_{i \leq j} a_{i,j}^{k_{i,j}} \equiv \prod_{j=1}^{k} a_{i,j,i_{\tau(j)}}, \]  

(3.5)

For \( w \in S_k \), to deduce that \( w \cdot (i_1, \ldots, i_k) \) is admissible, we need to exhibit a permutation \( \rho \in S_k \) such that

\[ \prod_{i \leq j} a_{i,j}^{k_{i,j}} \equiv \prod_{j=1}^{k} a_{i_{\rho(j)},i_{\rho(j)}}, \]  

\[ \prod_{j=1}^{k} a_{i_{\rho^{-1}(j)},i_{\rho^{-1}(j)}}, \]  

(3.6)

where the last equality follows by replacing the index \( j \) in the product by \( w^{-1}(j) \). By choosing \( \rho = w^{-1} \tau w \) we have the desired permutation. This proves (i).

Next, the proof of (ii) follows immediately from (3.6).
To prove (iii), we replace the index \( j \) by \( \tau^{-1}(j) \) in the product on the right-hand side of (3.5) and apply the symmetry of the matrix \( A = (a_{i,j}) \). Then we obtain

\[
\prod_{i \leq j} a_{k_{i,j}} = \prod_{j=1}^{k} a_{t_{\tau^{-1}(j)}},
\]

therefore \( \tau^{-1} \in \Sigma(i_1, \ldots, i_k) \).

We turn now to the proof of Theorem 3.2. Recall (cf. [16]) that the moment-generating function of \( X \) is defined to be

\[
E \exp \left( \text{tr} TX \right).
\]

Thus, with \( T = (t_{i,j}) \) and \( X = (X_{i,j}) \),

\[
E \exp(\text{tr} TX) = E \exp \left( \sum_{i=1}^{d} t_{i,i} X_{i,i} + 2 \sum_{1 \leq i < j \leq d} t_{i,j} X_{i,j} \right);
\]

so (3.7) is, in actuality, the joint moment-generating function of the set of random variables \( \{X_{i,j} : 1 \leq i \leq d\} \), where \( \delta_{i,j} \) denotes Kronecker's symbol.

This definition of the moment-generating function of a symmetric random matrix \( X \) is standard throughout multivariate statistical analysis. Define the matrix of differential operators

\[
\frac{\partial}{\partial T} = \left( \frac{1}{2} (1 + \delta_{i,j}) \frac{\partial}{\partial t_{i,j}} \right);
\]

the matrix \( \partial/\partial T \) is the unique differential operator for which

\[
P \left( \frac{\partial}{\partial T} \right) \exp(\text{tr} TX) = P(X) \exp(\text{tr} TX)
\]

for all symmetric \( d \times d \) matrices \( T \) and \( X \) and all polynomials \( P \).

By (3.8) it follows that, for any polynomial \( P(X) \),

\[
E P(X) = P \left( \frac{\partial}{\partial T} \right) E \exp(\text{tr} TX) \bigg|_{T=0}.
\]

In particular, for nonnegative integers \( k_{i,j} \) where \( 1 \leq i \leq j \leq d \),

\[
E \prod_{i \leq j} X_{i,j}^{k_{i,j}} = \prod_{i \leq j} \left( \frac{1}{2} (1 + \delta_{i,j}) \frac{\partial}{\partial t_{i,j}} \right)^{k_{i,j}} E \exp(\text{tr} TX) \bigg|_{T=0}.
\]

By a straightforward calculation from (1.1), it is well-known that the moment-generating function of \( X \) exists for \( \|T\| < 1 \); explicitly,

\[
E \exp(\text{tr} TX) = \det(I_d - T)^{-\beta}.
\]
By the extension (2.3) of the master theorem, we deduce that

\[ E \exp(\text{tr} \, TX) = 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{i_1, \ldots, i_k = 1}^{d} \text{per}_{\beta}(T(i_1, \ldots, i_k)). \] (3.11)

In this expansion it must be kept in mind that the matrix \( T \) is symmetric, whereas in the general expansion (2.3) \( T \) is allowed to be any square matrix.

Now the expectation of the monomial (3.1) is obtained by differentiating the moment-generating function. Noting (3.11), we deduce from (3.9) and (3.10) the result

\[ E \prod_{i \leq j} X_{i,j}^{k_{i,j}} = \prod_{i \leq j} \left( \frac{1}{2} (1 + \delta_{i,j}) \frac{\partial}{\partial t_{i,j}} \right)^{k_{i,j}} \cdot \det(I_d - T)^{-\beta} \bigg|_{T=0} \]

\[ = \frac{2^{-p}}{k!} \prod_{i \leq j} \left( \frac{\partial}{\partial t_{i,j}} \right)^{k_{i,j}} \sum_{i_1, \ldots, i_k = 1}^{d} \text{per}_{\beta}(T(i_1, \ldots, i_k)) \bigg|_{T=0}, \] (3.12)

where \( k \equiv \sum_{i \leq j} k_{i,j} \) and \( p \equiv \sum_{i<j} k_{i,j} \).

Finally, the proof is completed by inserting in (3.12) the formula (2.1) for the beta-permanent, reducing the set of monomials appearing in the expansion of the beta-permanent to those subsets determined by admissibility, and then performing the differentiations.

\[ \text{Remark 3.5.} \] We remark that (3.12) reduces the calculations of the moments of monomials in the \( X_{i,j} \) to a straightforward computation. In special cases in which the \( k_{i,j} \) are small, this calculation can be done by hand; for large \( k_{i,j} \), it is more efficient to utilize a symbolic algebra package such as MAPLE. Also, in (3.12), because of the homogeneity of the beta-permanents, there is no need to evaluate the outcome of differentiation at \( T = 0 \).

Despite the appealing nature of (3.12), the calculation of moments through (3.4) represents a significant further reduction in computational effort over (3.12). Such a reduction in effort stems from the fact that the condition of admissibility can reduce considerably the set of \( k \)-tuples arising in the calculations.

The admissible \( k \)-tuples \((i_1, \ldots, i_k)\) are precisely those for which the term \( P(T) = \prod_{i \leq j}^{k_{i,j}} \) appears in the monomial expansion of the beta-permanent \( \text{per}_{\beta}(T(i_1, \ldots, i_k)) \).

The problem of calculating the moments of polynomials in the entries of \( X^{-1} \), the inverse of a Wishart matrix, has also undergone scrutiny in multivariate statistics; cf. [20]. Suppose that \( X \sim W_d(\beta, I_d) \) and \( T \in \mathbb{R}_+^{d \times d} \).
then the moment-generating function of $X^{-1}$ is

$$
E \exp(\text{tr} TX^{-1}) = \frac{1}{\Gamma_d(\beta)} \int_{\mathbb{R}^{d \times d}_+} \left( \det \Sigma \right)^{-\beta} \left( \det x \right)^{\beta-(d+1)/2} \exp(-\text{tr} \Sigma^{-1} x) \, dx
$$

$$
= \frac{1}{\Gamma_d(\beta)} (\det T)^{\beta} B_{\beta}(T),
$$

(3.13)

where $B_{\beta}(\cdot)$ is a Bessel function of matrix argument of the second kind [9, Sect. 5]. Having evaluated the moment-generating function, we can also apply results of Herz [9, Sect. 9] to evaluate the moments of certain polynomials, including determinants or products of powers of principal minors, and zonal polynomials in $X^{-1}$. However, our methods do not yield explicit formulas for the moments of arbitrary monomials in the entries of $X^{-1}$; hence we refer to von Rosen [20] and others for additional results in this matter.

To complete this section we comment on a connection, established by Godement [5] and apparently not known to the statistical community, between the problem of calculating the moments of the Wishart distributions and the theory of group representations. Denote by $\text{GL}_+(d, \mathbb{R})$ the group of $d \times d$ real matrices with positive determinant. Suppose $\lambda$ is an irreducible representation of $\text{GL}_+(d, \mathbb{R})$ in a finite-dimensional vector space $\mathcal{F}_\lambda$. For any $\Sigma \in \mathbb{R}^{d \times d}_+$, define the integral

$$
H_{\lambda}(\Sigma) := \int_{\mathbb{R}^{d \times d}_+} \lambda(x) \exp(-\text{tr} \Sigma x)(\det x)^{-(d+1)} \, dx.
$$

Godement [5] gave a necessary and sufficient condition on the representation $\lambda$ for $H_{\lambda}(\Sigma)$ to be finite. Denote by $\alpha_{\lambda}$ the power to which the determinant occurs in the highest weight of $\lambda$; we assume that $\alpha_{\lambda} \in \mathbb{R}$. Then there exists a positive-definite Hermitian inner product, $\langle \cdot, \cdot \rangle$, on $\mathcal{F}_\lambda$ such that $\lambda(g') = \lambda(g)^*$ for all $g \in \text{GL}_+(d, \mathbb{R})$, where $\lambda(g)^*$ is the adjoint (with respect to the inner product $\langle \cdot, \cdot \rangle$) of $\lambda(g)$.

Under the assumption $\alpha_{\lambda} > d$, Godement proved that (i) $H_{\lambda}(\Sigma)$ is a Hermitian positive-definite operator on $\mathcal{F}_\lambda$; (ii) $H_{\lambda}(I_d)$ commutes with $\lambda(g)$ for all $g \in \text{SO}(d)$, the special orthogonal group; and (iii) for all $\Sigma \in \mathbb{R}^{d \times d}_+$,

$$
H_{\lambda}(\Sigma) = (\det \Sigma)^{(d+1)/2} \lambda(\Sigma^{1/2}) H_{\lambda}(I_d) \lambda(\Sigma^{1/2}).
$$

(3.14)

This latter formula shows that the evaluation of $H_{\lambda}(\Sigma)$ reduces to the case in which $\Sigma = I_d$. 
If \( \lambda \) is any homomorphism of \( \text{GL}(d, \mathbb{R}) \) and \( X \sim W_d(\beta, \Sigma) \) then by applying Godement's argument, we find that there holds an analog of (3.14) for \( E\lambda(X) \). Indeed, by making the standard change of variables in (3.2) then we obtain \( Y \sim W_d(\beta, I_d) \). Assuming that all expectations exist, we apply the homomorphism property of \( \lambda \) to obtain

\[
E\lambda(X) = E\lambda(\Sigma^{1/2}Y\Sigma^{1/2}) = E\lambda(\Sigma^{1/2})\lambda(Y)\lambda(\Sigma^{1/2}) = \lambda(\Sigma^{1/2})E\lambda(Y)\lambda(\Sigma^{1/2}).
\]

(3.15)

Furthermore, a similar argument proves that \( E\lambda(X) \) commutes with \( \lambda(g) \) for all \( g \in O(d) \) which commutes with \( \Sigma \).

As a special case, consider the example for which \( \lambda(x) = x^{\otimes r} \), the Kronecker (or tensor) product of \( r \) copies of \( x \). Then it follows from (3.15) that

\[
E(X^{\otimes r}) = (\Sigma^{1/2})^{\otimes r} \Phi(\Sigma^{1/2})^{\otimes r};
\]

(3.16)

where \( \Phi = E(Y^{\otimes r}) \) with \( Y \sim W_d(\beta, I_d) \). Moreover, all entries of the matrix \( \Phi \) can be computed using the procedure given in Theorem 3.2.

The result (3.16) appears to have gone unobserved in the statistical literature.

4. MOMENTS OF THE COMPLEX WISHART DISTRIBUTIONS

In the case of the complex Wishart distributions, results of a nature similar to those in Section 3 can be derived. Because the arguments are similar, we shall present the essential details and leave it to the reader to provide the final formulas. Denote by \( \mathbb{C}^{d \times d}_+ \) the space of positive-definite, Hermitian, \( d \times d \) matrices. Suppose that \( \Sigma \) denotes a fixed matrix in \( \mathbb{C}^{d \times d}_+ \) and \( \beta \in \mathbb{R} \) with \( \beta > d - 1 \). Following James [10], a random matrix \( V = (V_{i,j}) \in \mathbb{C}^{d \times d}_+ \) is said to have a (complex) Wishart distribution with parameters \( \beta \) and \( \Sigma \) if its probability density function is

\[
\frac{1}{\tilde{\Gamma}_d(\beta)}(\det \Sigma)^{-\beta}(\det v)^{\beta-d} \exp(-\text{tr} \Sigma^{-1}v),
\]

(4.1)

\( v \in \mathbb{C}^{d \times d}_+ \), where the corresponding multivariate gamma function is defined by

\[
\tilde{\Gamma}_d(\beta) := \pi^{d(d-1)/2} \prod_{j=1}^{d} \Gamma(\beta - j + 1).
\]

(4.2)
Whenever (4.1) holds, we write 

$$V \sim W_d(\beta, \tilde{\Sigma})$$

Similar to (3.10), the moment-generating function of $V$ is given by the formula

$$E \exp(\text{tr} \, TV) = \det(I_d - \tilde{\Sigma}T)^{-\beta},$$

where $T$ is any Hermitian $d \times d$ matrix such that $\|T\Sigma\| < 1$.

By expanding (4.3) using the extension (2.3) of MacMahon’s master theorem and proceeding as in Section 3, we can calculate the moments of any polynomial in the entries of $V$. The differential operators which replace those in (3.8) may be described as follows (cf. [4]).

Write the Hermitian matrix $T = (t_{i,j})$ in the form $T = T_1 + \sqrt{-1}T_2$ where $T_1 = (t'_{i,j})$ is real and symmetric and $T_2 = (t''_{i,j})$ is real and antisymmetric. Define the matrix of differential operators $\partial/\partial T$ with entry

$$\frac{\partial}{\partial t_{i,j}} := \begin{cases} \frac{\partial}{\partial t'_{i,j}}, & \text{if } i = j, \\ \frac{1}{2} \left( \frac{\partial}{\partial t'_{i,j}} + \sqrt{-1} \frac{\partial}{\partial t''_{i,j}} \right), & \text{if } i > j, \\ \frac{1}{2} \left( \frac{\partial}{\partial t'_{i,j}} - \sqrt{-1} \frac{\partial}{\partial t''_{i,j}} \right), & \text{if } i < j. \end{cases}$$

(4.4)

It is not difficult to verify that, for arbitrary Hermitian matrices $T$ and $V$, and for any polynomial $P$,

$$P\left(\frac{\partial}{\partial T}\right) \exp(\text{tr} \, TV) = P(V) \exp(\text{tr} \, TV).$$

Suppose $P(V) = \prod_{i,j} V_{i,j}^{k_{i,j}}$, a monomial in the entries of $V$, where $V \sim W_d(\beta, \tilde{\Sigma})$. To calculate $E P(V)$, we may assume, with no loss of generality, that $\tilde{\Sigma} = I_d$. Similar to (3.12), by applying the differential operator $P(\xi)$ to the corresponding moment-generating function, we obtain

$$E \prod_{i,j} V_{i,j}^{k_{i,j}} = P\left(\frac{\partial}{\partial T}\right) \cdot \det(I - T)^{-\beta} \bigg|_{T=0}$$

$$= \frac{1}{k!} P\left(\frac{\partial}{\partial T}\right) \sum_{i_1, \ldots, i_k=1}^{d} \text{per}_d(T(i_1, \ldots, i_k)) \bigg|_{T=0},$$

(4.5)

where $k = \sum_{i,j} k_{i,j}$. Writing out the expansion for $\text{per}_d(T(i_1, \ldots, i_k))$ and applying the operator $P(\partial/\partial T)$, it is straightforward now for the reader to write down a result similar to Theorem 3.2.

Another elegant, and still explicit, solution to the problem of computing the expectation of monomials of a complex Wishart matrix may be obtained.
through the representation theory of the general linear group GL\((d,\mathbb{C})\). Here, it is of interest to calculate \(E_\lambda(V)\) where \(\lambda\) is an irreducible, holomorphic representation of GL\((d,\mathbb{C})\) in a vector space \(\mathcal{F}_\lambda\) of dimension \(d_\lambda\); from this result we may deduce the expectation of arbitrary monomials in \(V\). Again, we denote by \(\alpha_\lambda\) the power to which the determinant appears in the highest weight of \(\lambda\), and we assume \(\alpha_\lambda \in \mathbb{R}\). Further, we denote by \(d_\lambda\) the dimension of \(\lambda\).

The following result is due to Gross and Kunze [6, Theorem 8, p. 104].

**Theorem 4.1.** (Gross and Kunze [6]). Suppose that \(V \sim W_d(\beta, I_d)\) and \(\lambda\) is an irreducible, polynomial representation of GL\((d,\mathbb{C})\) in a finite-dimensional vector space \(\mathcal{F}_\lambda\). If \(\beta + \alpha_\lambda > d - 1\) then \(E_\lambda(V)\) exists and

\[
E_\lambda(V) = c_\lambda I_{d_\lambda},
\]

where \(I_{d_\lambda}\) is the identity operator on \(\mathcal{F}_\lambda\) and \(c_\lambda \in \mathbb{R}\) is a constant.

Gross and Kunze prove (4.6) by showing that \(E_\lambda(V)\) commutes with \(\lambda(u)\) for all \(u \in U(d)\), the group of unitary \(d \times d\) matrices. Then, by Schur’s lemma, it follows that \(E_\lambda(V)\) is a scalar multiple of the identity matrix on \(\mathcal{F}_\lambda\).

An alternative proof of Theorem 4.1 can be obtained by applying to the Wishart probability distribution \(W_d(\beta, I_d)\) the polar decomposition

\[
V = U\Delta U^{-1},
\]

where \(U \in U(d)\) and \(\Delta = \text{diag}(v_1, \ldots, v_d)\) is a diagonal matrix containing the eigenvalues of \(V\). Under this decomposition, it is well-known that the Lebesgue measure on \(\mathbb{C}_+^{d \times d}\) factors according to the formula

\[
dV = c_0 \prod_{1 \leq i < j \leq d} (v_i - v_j)^2 dv_1 \cdots dv_d dU,
\]

where \(c_0\) is a positive constant and \(dU\) denotes the normalized Haar measure on \(U(d)\). By applying (4.7) and (4.8) to the density function (4.1) for the probability distribution \(W_d(\beta, I_d)\), we obtain

\[
E_\lambda(V) = E_\lambda(U\Delta U^{-1}) = E_U[\lambda(U)(E_\Delta \lambda(\Delta))\lambda(U^{-1})],
\]

where \(E_U\) denotes expectation with respect to the Haar measure on \(U(d)\) and \(E_\Delta\) denotes expectation with respect to the probability measure proportional to

\[
\prod_{i=1}^d v_i^{\beta-d} \exp(-v_i) \prod_{1 \leq i < j \leq d} (v_i - v_j)^2 dv_1 \cdots dv_d,
\]
moments of the Wishart distributions 543

\[ v_1, \ldots, v_d > 0. \] The constant of proportionality in (4.10), as well as the condition for existence of \( E_\Delta \lambda(\Delta) \), are consequences of a limiting case of the beta integral formula of Selberg [21] (cf. [2]).

To complete this proof, it needs to be shown that (4.9) is a scalar matrix. To accomplish this we choose an orthonormal basis for the representation space \( \mathcal{F}_\lambda \) and denote by \( \lambda_{i,j}(V) \) the \((i, j)\)th entry of \( \lambda(V) \) relative to this basis. By (4.7) we have

\[ \lambda_{i,j}(V) = \sum_{k,l=1}^{d_i} \lambda_{i,k}(U)\lambda_{k,l}(\Delta)\lambda_{l,j}(U^{-1}). \]

By Weyl’s unitary trick (Knapp [12, p. 115]), the restriction of \( \lambda \) to \( U(d) \) remains irreducible. Then, by the Schur orthogonality relations (Knapp [12, p. 15]), we obtain

\[ \text{E}_U \lambda_{i,k}(U)\lambda_{l,j}(U^{-1}) = \frac{1}{d_\lambda} \delta_{i,j} \delta_{k,l}; \]

hence

\[ \text{E} \lambda_{i,j}(V) = \frac{1}{d_\lambda} \delta_{i,j} \text{E}_\Delta \sum_{k,l=1}^{d_i} \delta_{k,l} \lambda_{k,l}(\Delta) \]

\[ = \frac{1}{d_\lambda} \delta_{i,j} \text{E}_\Delta \sum_{k=1}^{d_i} \lambda_{k,l}(\Delta) = \frac{1}{d_\lambda} (\text{E}_\Delta \text{tr} \lambda(\Delta))\delta_{i,j}. \quad (4.11) \]

This proves that (4.9) is a scalar matrix.

The constant \( c_\lambda \) in (4.6) can be calculated either from (4.11), by applying a limiting case of an expectation formula given by Richards [19, Eq. (31)], or by a method which begins by taking traces of both sides of (4.6). Following either approach, we obtain

\[ c_\lambda = \frac{1}{d_\lambda} \text{E}_\Delta \text{tr} \lambda(\Delta) = \frac{1}{d_\lambda} \text{E} \text{tr} \lambda(V). \quad (4.12) \]

By the Weyl character formula (cf. Knapp [12, p. 105]) the character, \( \text{tr} \lambda(V) \), of the representation \( \lambda \) depends only on the eigenvalues of \( V \) and may be expressed in terms of the ubiquitous Schur functions. Indeed, corresponding to the representation \( \lambda \) there exists a unique partition \( (l_1, \ldots, l_d) \), consisting of integers \( l_1 \geq \cdots \geq l_d \geq 0 \), such that, with \( v_1, \ldots, v_d \) denoting the eigenvalues of \( V \),

\[ \text{tr} \lambda(V) = s_{(l_1, \ldots, l_d)}(V) := \frac{\det(v_j^{l_i+n-j})}{\prod_{1 \leq i < j \leq d} (v_i - v_j)}. \quad (4.13) \]
Moreover, it follows as a limiting case of (4.13) that
\[ d_\lambda = \operatorname{tr} \lambda(I_d) = s_{(l_1, \ldots, l_d)}(I_d) = \frac{\prod_{1 \leq i < j \leq d}(l_i - l_j - i + j)}{\prod_{i=1}^{d}(i - 1)!}. \quad (4.14) \]
Using the notation \( (\beta)_i := \beta(! + 1) \cdots (\beta + l - 1), \) \( l = 0, 1, 2, \ldots, \) for the classical shifted factorial, we define the partitional shifted factorial
\[ [\beta]_{(l_1, \ldots, l_d)} := \prod_{i=1}^{d}(\beta - i + 1)_{l_i}. \]
Then by James [10, Eq. (86), p. 488],
\[ \mathbb{E} s_{(l_1, \ldots, l_d)}(V) = [\beta]_{(l_1, \ldots, l_d)} s_{(l_1, \ldots, l_d)}(I_d) = d_\lambda [\beta]_{(l_1, \ldots, l_d)}. \quad (4.15) \]
Collecting together the results (4.12)–(4.15), we obtain
\[ c_\gamma = \frac{1}{d_\lambda} \mathbb{E} \operatorname{tr} \lambda(V) = \frac{1}{d_\lambda} \mathbb{E} s_{(l_1, \ldots, l_d)}(V) = [\beta]_{(l_1, \ldots, l_d)}. \]
Now that we have calculated the expectation of any irreducible polynomial representation of \( \mathrm{GL}(d, \mathbb{C}) \), we turn to the calculation of \( \mathbb{E} \lambda(V) \) for the case in which \( \lambda \) is a tensor product of \( r \) copies of the identity representation, \( \lambda(g) = \mathfrak{g}^{\otimes r}, \) \( g \in \mathrm{GL}(d, \mathbb{C}) \).

**Theorem 4.2.** Let \( r \) be a positive integer, \( \tilde{\Sigma} \in \mathbb{C}^{d \times d}_+ \), and suppose that \( V \sim W_d(\beta, \tilde{\Sigma}), \) \( \beta > d - 1. \) Then
\[ \mathbb{E}(V^{\otimes r}) = (\tilde{\Sigma}^{1/2})^{\otimes r} \left( \bigoplus \sum \gamma(l_1, \ldots, l_d) [\beta]_{(l_1, \ldots, l_d)} I_{d_\lambda} \right) (\tilde{\Sigma}^{1/2})^{\otimes r}, \quad (4.16) \]
where the direct sum is taken over all partitions \( (l_1, \ldots, l_d) \) satisfying \( l_1 + \cdots + l_d = r; \) \( d_\lambda \), given in (4.14), is the dimension of the irreducible representation \( \lambda \) corresponding to the partition \( (l_1, \ldots, l_d); \) and
\[ \gamma(l_1, \ldots, l_d) = r! \frac{\prod_{1 \leq i < j \leq d}(l_i - l_j - i + j)}{\prod_{i=1}^{d}(l_i + d - i)!}. \quad (4.17) \]

**Proof.** Since the representation \( \lambda(g) = \mathfrak{g}^{\otimes r}, \) \( g \in \mathrm{GL}(d, \mathbb{C}) \), is polynomial, we may apply the argument given by Gross and Kunze [6, Theorem 8, p. 104] to deduce that the expectation \( \mathbb{E}(V^{\otimes r}) \) exists for \( \beta > d - 1. \)

By making the change-of-variables \( V = \tilde{\Sigma}^{1/2} V_0 \tilde{\Sigma}^{1/2} \) where the random matrix \( V_0 \sim W_d(\beta, I_d) \) and applying the homomorphism property of the representation \( \lambda \), we find that
\[ \mathbb{E}(V^{\otimes r}) = \mathbb{E}(\tilde{\Sigma}^{1/2} V_0^{\otimes r} \tilde{\Sigma}^{1/2}) = (\tilde{\Sigma}^{1/2})^{\otimes r} \mathbb{E}(V_0^{\otimes r})(\tilde{\Sigma}^{1/2})^{\otimes r}. \]
Therefore we may assume, with no loss of generality, that \( \tilde{\Sigma} = I_d. \)

To establish (4.16), we decompose \( \lambda \) into its irreducible constituents and then take expectations of each term. It is well known (cf. [11, 26]) that the representation \( \lambda(g) = \mathfrak{g}^{\otimes r}, \) \( g \in \mathrm{GL}(d, \mathbb{C}) \), decomposes explicitly as a
\begin{align}
g^{\otimes r} = \sum \oplus \gamma(l_1, \ldots, l_d)\lambda_{\{l_1, \ldots, l_d\}}(g), \tag{4.18}
\end{align}

where the direct sum is taken over all partitions \((l_1, \ldots, l_d)\) satisfying \(l_1 + \cdots + l_d = r\) and \(\lambda_{\{l_1, \ldots, l_d\}}(V)\) denotes the irreducible representation corresponding to the partition \((l_1, \ldots, l_d)\). Moreover, the coefficients \(\gamma(l_1, \ldots, l_d)\) are defined by the recurrence relations \(\gamma(0, \ldots, 0) = 1\) and otherwise,

\[
\gamma(l_1, \ldots, l_d) = \sum_j \gamma(l_1, \ldots, l_{j-1}, l_j - 1, l_{j+1}, \ldots, l_d), \tag{4.19}
\]

where the sum is taken over all \(j \in \{1, \ldots, d\}\) such that \(l_{j-1} \geq l_j - 1 \geq l_{j+1}\).

The solution to these recurrence relations dates back to Frobenius and MacMahon (cf. [25]) and is well known to be given by (4.17).

Applying (4.18) and Theorem 4.1, we obtain the explicit formula

\[
\mathbb{E}(V^{\otimes r}) = \sum \oplus \gamma(l_1, \ldots, l_d) \mathbb{E}\lambda_{\{l_1, \ldots, l_d\}}(V) = \sum \oplus \gamma(l_1, \ldots, l_d) [\beta]_{\{l_1, \ldots, l_d\}} I_{d_1}.
\]

Now the proof is complete. \(\blacksquare\)

In the literature, much work also has been done on calculating the expectations of polynomial functions of the inverse of complex Wishart random matrices. Here, we obtain a complete solution along the lines of Theorem 4.16.

**Theorem 4.3.** Let \(r\) be a positive integer, \(\tilde{\Sigma} \in \mathbb{C}^{d \times d}_+\), and suppose that \(V \sim W_d(\beta, \tilde{\Sigma})\) where \(\beta > r + d - 1\). Then

\[
\mathbb{E}((V^{-1})^{\otimes r}) = (\tilde{\Sigma}^{-1/2})^{\otimes r} \left( \sum \oplus \gamma(l_1, \ldots, l_d) [\beta]_{\{l_1, \ldots, l_d\}} I_{d_1} \right) (\tilde{\Sigma}^{-1/2})^{\otimes r}, \tag{4.20}
\]

where

\[
[\beta]_{\{l_1, \ldots, l_d\}} := \left( \prod_{i=1}^d \frac{1}{(\beta - l_j - d + j)_{l_j}} \right). \tag{4.21}
\]

**Proof.** As before, we may assume, with no loss of generality, that \(\tilde{\Sigma} = I_d\). Define \(\lambda(g) = (g^{-1})^{\otimes r}, g \in \text{GL}(d, \mathbb{C})\). By following the argument given by Gross and Kunze [6], we deduce that \(\mathbb{E}\lambda(V)\) exists under the stated condition on \(\beta\). Furthermore, it follows from the decomposition (4.18) that

\[
\mathbb{E}((V^{-1})^{\otimes r}) = \mathbb{E} \sum \oplus \gamma(l_1, \ldots, l_d) \lambda_{\{l_1, \ldots, l_d\}}(V^{-1}) = \sum \oplus \gamma(l_1, \ldots, l_d) \mathbb{E}\lambda_{\{l_1, \ldots, l_d\}}(V^{-1}).
\]
Using the polar decomposition (4.7) and applying the argument which led to (4.11), we deduce that $E_{\lambda}(V^{-1})$ is a scalar operator on the underlying representation space $\mathcal{F}_\lambda$. After carrying out the necessary calculations to evaluate the constant, we obtain

$$E_{\lambda}(V^{-1}) = \frac{1}{[\beta]^{d_1}}I_{d_1},$$

where $[\beta]^{d_1}$ is defined in (4.21).

REFERENCES