CHAPTER 3

Derivation and some basic properties of zonal polynomials

In this chapter we define (real) zonal polynomials and derive their basic properties. The results derived in this chapter are sufficient for usual applications of zonal polynomials. Some remarks on notation seem appropriate here. We define zonal polynomials as characteristic vectors of a certain linear transformation τ from V_n to V_n. The normalization is rather arbitrary for a characteristic vector and many properties of zonal polynomials are independent of particular normalization. Corresponding to different normalizations, different symbols such as Z_p, C_p have been used to denote zonal polynomials. We find it advantageous to use still another normalization in addition to those corresponding to Z_p, C_p. Considering these circumstances we use $\mathcal{Y}_p$ for an unnormalized zonal polynomial. ${}_1\mathcal{Y}_p$ is used to denote a zonal polynomial normalized so that the coefficient of $\mathcal{U}_p$ or $\mathcal{M}_p$ is 1.

§ 3.1 DEFINITION OF ZONAL POLYNOMIALS

As mentioned earlier we define zonal polynomials as characteristic vectors of a certain matrix. The matrix in question will be triangular and we begin by a lemma concerning a triangular matrix and its characteristic vectors.

Lemma 1. *Let $\boldsymbol{T} = (t_{ij})$ be an $n \times n$ upper triangular matrix with distinct* *diagonal elements. Let* $\Lambda = \mathrm{diag}(t_{11}, \ldots, t_{nn})$. *Then there exists a nonsingular upper triangular matrix* $\mathcal{B}$ *satisfying*

$$\mathcal{B}T = \Lambda\mathcal{B}. \tag{1}$$

$\mathcal{B}$ *is uniquely determined up to a (possibly different) multiplicative constant for each row.*

Proof is straightforward and omitted. Note that t_{ii}, $i = 1, \ldots, n$ are characteristic roots of T and *i-th* row of $\mathcal{B}$ is the characteristic vector (from the left) associated with t_{ii}.

Remark 1. This lemma seems to be well known to people in numerical analysis although an explicit reference is not easy to find. It is very briefly mentioned on page 365 of Stewart (1973) in connection with the QR algorithm. The QR algorithm is designed to transform a general matrix to a triangular form in order to obtain the characteristic roots and vectors.

For a $k \times k$ matrix $A = (a_{ij})$ we denote its (possibly complex) characteristic roots by

$$\alpha = (\alpha_1, \ldots, \alpha_k) = \lambda(A), \tag{2}$$

and (the determinant of) a principal minor by

$$A(i_1, \ldots, i_\ell) = \begin{vmatrix} a_{i_1 i_1} & \cdots & a_{i_1 i_\ell} \\ \vdots & & \vdots \\ a_{i_\ell i_1} & \cdots & a_{i_\ell i_\ell} \end{vmatrix}. \tag{3}$$

For a matrix argument we define

$$\mathcal{U}_p(A) = \mathcal{U}_p(\alpha) = \mathcal{U}_p(\lambda(A)).$$

As is easily seen by expanding the determinant $|A - \lambda I|$ the *r-th* elementary symmetric function of the roots of a matrix A is equal to the sum of $r \times r$ principal minors, namely

$$\mathcal{U}_{(1^r)}(A) = u_r(\alpha_1, \ldots, \alpha_k) = \sum_{i_1 < \cdots < i_r} A(i_1, \ldots, i_r). \tag{4}$$

(See Theorem 7.1.2 of Mirsky (1955) for example.) Hence

$$(5) \qquad \mathcal{U}_p(\boldsymbol{A}) = \{\sum_{i_1} \boldsymbol{A}(i_1)\}^{p_1-p_2}\{\sum_{i_1<i_2} \boldsymbol{A}(i_1,i_2)\}^{p_2-p_3}\cdots.$$

Note that (4) and (5) holds for general (not necessarily symmetric) matrix $\boldsymbol{A}$.

Now let $\boldsymbol{A}$ be symmetric and consider a (linear) transformation $\tau_\nu : V_n \to V_n$ defined by

$$(6) \qquad (\tau_\nu(\mathcal{U}_p))(\boldsymbol{A}) = (\tau_\nu \mathcal{U}_p)(\boldsymbol{A}) = \mathcal{E}_W\{\mathcal{U}_p(\boldsymbol{AW})\},$$

where $\boldsymbol{W}$ is a random symmetric matrix having a Wishart distribution $\mathcal{W}(\boldsymbol{I}_k,\nu)$, $\nu \geq k$. Here $\mathcal{W}(\boldsymbol{\Sigma},\nu)$ denotes the Wishart distribution with covariance $\boldsymbol{\Sigma}$ and degrees of freedom ν. (τ_ν is defined for the basis $\{\mathcal{U}_p\}$ by (6) and for general elements of $V_n(\boldsymbol{A})$ τ_ν is given by the linearity of expectation.) First we need to verify:

Lemma 2. $\quad \tau_\nu \mathcal{U}_p \in V_n$.

Proof. Since $\boldsymbol{A}$ is symmetric it can be written as $\boldsymbol{A} = \boldsymbol{\Gamma D \Gamma'}$ where $\boldsymbol{\Gamma}$ is orthogonal and $\boldsymbol{D}$ =diag$(\alpha_1,\ldots,\alpha_k)$. Now $\mathcal{U}_p(\boldsymbol{AW})= \mathcal{U}_p(\boldsymbol{\Gamma D\Gamma' W})= \mathcal{U}_p(\boldsymbol{D\Gamma' W\Gamma})$ because the nonzero roots are invariant when the matrices are permuted cyclically. Since the distribution of $\boldsymbol{\Gamma' W\Gamma}$ is the same as the distribution of $\boldsymbol{W}$, we can take $\boldsymbol{A} =$ diag$(\alpha_1,\ldots,\alpha_k)$ without loss of generality. Then

$$(7) \qquad \boldsymbol{AW}(i_1,\ldots,i_r) = (\alpha_{i_1}\cdots\alpha_{i_r})\boldsymbol{W}(i_1,\ldots,i_r).$$

For example

$$(8) \qquad \boldsymbol{AW}(1,2) = \begin{vmatrix} \alpha_1 w_{11} & \alpha_1 w_{12} \\ \alpha_2 w_{21} & \alpha_2 w_{22} \end{vmatrix} = \alpha_1\alpha_2 \begin{vmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{vmatrix}.$$

From (4) and (7) the ***r-th*** elementary symmetric function of the characteristic roots of $\boldsymbol{AW}$ can be written as

$$(9) \qquad u_r(\lambda(\boldsymbol{AW})) = \sum_{i_1<\cdots<i_r} \alpha_{i_1}\cdots\alpha_{i_r}\boldsymbol{W}(i_1,\ldots,i_r).$$

Substituting this into (5) and taking the expectation we obtain

$$(10) \qquad (\tau_\nu \mathcal{U}_p)(\boldsymbol{A}) = \mathcal{E}_W(\sum_{i_1}\alpha_{i_1}\boldsymbol{W}(i_1))^{p_1-p_2}(\sum_{i_1<i_2}\alpha_{i_1}\alpha_{i_2}\boldsymbol{W}(i_1,i_2))^{p_2-p_3}\cdots.$$

Clearly this belongs to V_n. ∎

τ_ν has the following triangular property.

Corollary 1.

(11) $$(\tau_\nu \mathcal{U}_p)(\boldsymbol{A}) = \lambda_{\nu p}\mathcal{U}_p(\boldsymbol{A}) + \sum_{q<p} a_{pq}\mathcal{U}_q(\boldsymbol{A}).$$

Proof. It suffices to show this for a diagonal matrix $\boldsymbol{A} = \mathrm{diag}(\alpha_1, \ldots, \alpha_k)$. As in the proof of Lemma 2.2.2 the highest monomial term in (10) is of the form

(12) $$\alpha_1^{p_1}\alpha_2^{p_2}\cdots\alpha_\ell^{p_\ell}\mathcal{E}_W\left\{\boldsymbol{W}(1)^{p_1-p_2}\boldsymbol{W}(1,2)^{p_2-p_3}\cdots\boldsymbol{W}(1,\ldots,\ell)^{p_\ell}\right\}.$$

Then using (2.2.15) we see that $(\tau_\nu\mathcal{U}_p)(\boldsymbol{A})$ expressed as a linear combination of $\mathcal{U}_q$'s involves only q's such that $q \le p$. In particular the leading coefficient is

(13) $$\lambda_{\nu p} = \mathcal{E}_W\left\{\boldsymbol{W}(1)^{p_1-p_2}\boldsymbol{W}(1,2)^{p_2-p_3}\cdots\boldsymbol{W}(1,\ldots,\ell)^{p_\ell}\right\}.$$

∎

Remark 2. The constants a_{pq} in (11) depend on the degrees of freedom ν.

Remark 3. To be complete we have to verify that (6) does not depend on the number of variables k or more precisely we need to verify

(14) $$(\tau_\nu\mathcal{U}_p)(\alpha_1,\ldots,\alpha_k,0,\ldots,0) = (\tau_\nu\mathcal{U}_p)(\alpha_1,\ldots,\alpha_k),$$

for any number (m) of additional zeros. Note that the left hand side is defined using expectation with respect to $\mathcal{W}(\boldsymbol{I}_{k+m},\nu)$. Now recall that the marginal distribution of the $k \times k$ upper left hand corner of $\mathcal{W}(\boldsymbol{I}_{k+m},\nu)$ is $\mathcal{W}(\boldsymbol{I}_k,\nu)$ and (10) depends only on the $k \times k$ upper left hand corner of the Wishart matrix. Therefore (14) holds.

By Corollary 1 τ_ν expressed in an appropriate matrix form is an upper triangular matrix. In order to apply Lemma 1 we want to evaluate the "diagonal elements" $\lambda_{\nu p}$ in (13). For that purpose we use the following well known result.

Lemma 3. *Let $\boldsymbol{W}$ be distributed according to $\mathcal{W}(\boldsymbol{I}_k, \nu)$. Let $\boldsymbol{T} = (t_{ij})$ be a lower triangular matrix with nonnegative diagonal elements such that $\boldsymbol{W} = \boldsymbol{T}\boldsymbol{T}'$. Then t_{ij}, $i \geq j$, are independently distributed as $t_{ij} \sim \mathcal{N}(0,1)$, $i > j$, $t_{ii} \sim \chi(\nu - i + 1)$ where $\chi(\nu-i+1)$ denotes the chi-distribution with $\nu - i + 1$ degrees of freedom.*

For a proof see Wijsman (1959) or Kshirsagar (1959).

Corollary 2.

$$
\begin{aligned}
\lambda_{\nu p} &= 2^n \prod_{i=1}^{\ell} \Gamma[p_i + \tfrac{1}{2}(\nu + 1 - i)]/\Gamma[\tfrac{1}{2}(\nu + 1 - i)] \\
&= 2^n \prod_{i=1}^{\ell} \left(\frac{\nu + 1 - i}{2}\right)_{p_i} \\
&= \nu(\nu + 2)\cdots(\nu + 2(p_1 - 1)) \\
&\quad \cdot (\nu - 1)(\nu + 1)\cdots(\nu - 1 + 2(p_2 - 1)) \\
&\quad \cdots \\
&\quad \cdot (\nu - \ell + 1)\cdots(\nu - \ell + 1 + 2(p_\ell - 1)),
\end{aligned}
\tag{15}
$$

where $\ell = \ell(p)$ and $(a)_k = a(a+1)\cdots(a+k-1)$.

Proof. Note

$$
\boldsymbol{W}(1, \ldots, r) = (t_{11} \cdots t_{rr})^2. \tag{16}
$$

Substituting this into (13) we obtain

$$
\lambda_{\nu p} = \mathcal{E}\{t_{11}^{2p_1} t_{22}^{2p_2} \cdots t_{\ell\ell}^{2p_\ell}\}. \tag{17}
$$

Now t_{ii}^2 is distributed according to $\chi^2(\nu - i + 1)$ and $\mathcal{E} t_{ii}^{2p_i} = (\nu - i + 1)(\nu - i + 3) \cdots (\nu - i + 1 + 2(p_i - 1))$. From this we obtain (15). ∎

This proof is given in Constantine (1963) in a slightly different form.

Using the vector notation introduced in (2.2.13) let

$$
\tau_\nu(\mathcal{U}) = \begin{pmatrix} \tau_\nu(\mathcal{U}_{(n)}) \\ \tau_\nu(\mathcal{U}_{(n-1,1)}) \\ \cdot \\ \cdot \\ \cdot \\ \tau_\nu(\mathcal{U}_{(1^n)}) \end{pmatrix} \tag{18}
$$

Then Corollary 1 shows that

$$\tau_\nu(\mathcal{U}) = \boldsymbol{T}_\nu \mathcal{U}, \tag{19}$$

where $\boldsymbol{T}_\nu$ is an upper triangular matrix with diagonal elements $t_{pp} = \lambda_{\nu p}$. $\boldsymbol{T}_\nu$ almost fits the condition of Lemma 1. The question now is what ν to take. Actually a particular choice of ν does not matter; we have:

Lemma 4. *There exists a nonsingular upper triangular matrix $\mathcal{B}$ such that*

$$\mathcal{B}\boldsymbol{T}_\nu = \boldsymbol{\Lambda}_\nu \mathcal{B} \qquad \textit{for all } \nu, \tag{20}$$

where $\boldsymbol{\Lambda}_\nu = \operatorname{diag}(\lambda_{\nu p}, p \in \mathcal{P}_n)$. $\mathcal{B}$ is uniquely determined up to a (possibly different) multiplicative constant for each row.

Lemma 4 shows that $\boldsymbol{T}_\nu$ has the same set of characteristic vectors (from the left) for all ν. A proof of this will be given later in this section. Now we define zonal polynomials using this $\mathcal{B}$.

Definition 1. (*zonal polynomials*)

Let $\mathcal{B}$ be as in Lemma 4. Zonal polynomials $\mathcal{Y}_p$, $p \in \mathcal{P}_n$, are defined by

$$\mathcal{Y} = \begin{pmatrix} \mathcal{Y}_{(n)} \\ \mathcal{Y}_{(n-1,1)} \\ \cdot \\ \cdot \\ \cdot \\ \mathcal{Y}_{(1^n)} \end{pmatrix} = \mathcal{B}\mathcal{U}. \tag{21}$$

Remark 4. $\mathcal{B}$ is upper triangular and therefore $\mathcal{Y}_p$ is a linear combination of $\mathcal{U}_q$'s (or $\mathcal{M}_q$'s) with $q \le p$. It follows that $\{ \mathcal{Y}_p, p \in \mathcal{P}_n \}$ forms a basis of V_n.

Remark 5. Since each row of $\mathcal{B}$ is determined uniquely up to a multiplicative constant $\mathcal{Y}_p$ is determined up to normalization. We use $\mathcal{Y}_p$ to denote an unnormalized zonal polynomial.

In order to prove Lemma 4 we first establish that the $\boldsymbol{T}_\nu$'s commute with each other.

Lemma 5.

$$T_\nu T_\mu = T_\mu T_\nu. \tag{22}$$

Proof. For a symmetric positive semi-definite matrix A let $A^{\frac{1}{2}}$ be the symmetric positive semi-definite square root, i.e., $A^{\frac{1}{2}} = \Gamma D^{\frac{1}{2}}\Gamma'$ where Γ is orthogonal and D is diagonal in $A = \Gamma D \Gamma'$. Now let W, V be independently distributed according to $\mathcal{W}(I_k, \nu)$, $\mathcal{W}(I_k, \mu)$ respectively. Consider

$$\mathcal{E}_{W,V}\{\mathcal{U}(A^{\frac{1}{2}}VA^{\frac{1}{2}}W)\}, \tag{23}$$

where $\mathcal{U} = (\mathcal{U}_{(n)}, \mathcal{U}_{(n-1,1)}, \ldots, \mathcal{U}_{(1^n)})'$. Taking expectation with respect to W first we obtain

$$\begin{aligned} &\mathcal{E}_{W,V}\{\mathcal{U}(A^{\frac{1}{2}}VA^{\frac{1}{2}}W)\} \\ &\quad = \mathcal{E}_V\{T_\nu \mathcal{U}(A^{\frac{1}{2}}VA^{\frac{1}{2}})\} \\ &\quad = \mathcal{E}_V\{T_\nu \mathcal{U}(AV)\} \\ &\quad = T_\nu \mathcal{E}_V\{\mathcal{U}(AV)\} \\ &\quad = T_\nu T_\mu \mathcal{U}(A). \end{aligned} \tag{24}$$

We used the cyclic permutation of the matrices since nonzero characteristic roots are invariant. Similarly taking expectation with respect to V first we obtain

$$\mathcal{E}_{W,V}\{\mathcal{U}(A^{\frac{1}{2}}VA^{\frac{1}{2}}W)\} = T_\mu T_\nu \mathcal{U}(A). \tag{25}$$

Hence $T_\nu T_\mu \mathcal{U}(A) = T_\mu T_\nu \mathcal{U}(A)$ for any symmetric positive semidefinite A. Now a polynomial is identically equal to zero if it is zero for all nonnegative arguments. This implies $T_\nu T_\mu = T_\mu T_\nu$. ∎

See Theorem 2.2 of Kushner, Lebow, and Meisner (1981) for an analogous result in a more general framework.

Now we give a proof of Lemma 4.

Proof of Lemma 4: Consider $\lambda_{\nu p}$ given by (15). Let us look at $\lambda_{\nu p}$ as a polynomial in ν. They are different polynomials for different partitions since they have different sets of roots. Now two different polynomials can match only finite number of times. It follows that for a sufficiently large ν_0, $\lambda_{\nu_0 p}$, $p \in \mathcal{P}_n$, are all different. Let ν_0 be fixed such that $\lambda_{\nu_0 p}$, $p \in \mathcal{P}_n$ are all different. Let $\mathcal{B}$ be the matrix in (1) with $\boldsymbol{T}$ replaced by $\boldsymbol{T}_{\nu_0}$. Note that the uniqueness part of Lemma 4 is already established now. Let $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_{\nu_0 p}, p \in \mathcal{P}_n)$. Then for any μ $\boldsymbol{\Lambda}(\mathcal{B}\boldsymbol{T}_\mu) = (\boldsymbol{\Lambda}\mathcal{B})\boldsymbol{T}_\mu = (\mathcal{B}\boldsymbol{T}_{\nu_0})\boldsymbol{T}_\mu = \mathcal{B}(\boldsymbol{T}_{\nu_0}\boldsymbol{T}_\mu) = \mathcal{B}(\boldsymbol{T}_\mu \boldsymbol{T}_{\nu_0}) = (\mathcal{B}\boldsymbol{T}_\mu)\boldsymbol{T}_{\nu_0}$, or $\boldsymbol{\Lambda}\mathcal{B}_1 = \mathcal{B}_1 \boldsymbol{T}_{\nu_0}$ where $\mathcal{B}_1 = \mathcal{B}\boldsymbol{T}_\mu$. Now by the uniqueness part of Lemma 1 we have $\mathcal{B}_1 = \boldsymbol{D}\mathcal{B}$ for some diagonal $\boldsymbol{D}$ or $\mathcal{B}\boldsymbol{T}_\mu = \boldsymbol{D}\mathcal{B}$. Considering the diagonal elements we see that $\boldsymbol{D} = \boldsymbol{\Lambda}_\mu = \mathrm{diag}(\lambda_{\mu p}, p \in \mathcal{P}_n)$. Therefore $\mathcal{B}\boldsymbol{T}_\mu = \boldsymbol{\Lambda}_\mu \mathcal{B}$ for all μ. ∎

We defined zonal polynomials by defining their coefficients. From a little bit more abstract viewpoint they are eigenfunctions of the linear operator τ_ν and the results in this section can be summarized as follows.

Theorem 1. *Let $\mathcal{Y}_p$ be a zonal polynomial then*

$$(\tau_\nu \mathcal{Y}_p)(\boldsymbol{A}) = \mathcal{E}_W \mathcal{Y}_p(\boldsymbol{A}\boldsymbol{W}) = \lambda_{\nu p} \mathcal{Y}_p(\boldsymbol{A}), \tag{26}$$

where $\boldsymbol{W} \sim \mathcal{W}(\boldsymbol{I}_k, \nu)$, $\boldsymbol{A}$ is symmetric and $\lambda_{\nu p}$ is given in (15). Conversely (26) (for all sufficiently large ν and for all symmetric $\boldsymbol{A}$) implies that $\mathcal{Y}_p$ is a zonal polynomial.

Proof. $\mathcal{Y} = (\mathcal{Y}_{(n)}, \mathcal{Y}_{(n-1,1)}, \ldots, \mathcal{Y}_{(1^n)})' = \mathcal{B}\mathcal{U}$. Hence by Lemma 4

$$\begin{aligned} \mathcal{E}_W\{\mathcal{Y}(\boldsymbol{A}\boldsymbol{W})\} &= \mathcal{E}_W\{\mathcal{B}\mathcal{U}(\boldsymbol{A}\boldsymbol{W})\} \\ &= \mathcal{B}\mathcal{E}_W\{\mathcal{U}(\boldsymbol{A}\boldsymbol{W})\} \\ &= \mathcal{B}\boldsymbol{T}_\nu \mathcal{U}(\boldsymbol{A}) \\ &= \boldsymbol{\Lambda}_\nu \mathcal{B}\mathcal{U}(\boldsymbol{A}) \\ &= \boldsymbol{\Lambda}_\nu \mathcal{Y}(\boldsymbol{A}). \end{aligned} \tag{27}$$

Therefore (26) holds. Conversely assume (26). Let $\mathcal{Y}_p = \sum_{q \in \mathcal{P}_n} a_q \mathcal{U}_q$. Then (26) implies

$$\boldsymbol{a}'\boldsymbol{T}_\nu = \lambda_{\nu p} \boldsymbol{a}',$$

where $\boldsymbol{a}' = (a_{(n)}, \ldots, a_{(1^n)})$. Now by the uniqueness part of Lemma 4 $\boldsymbol{a}'$ coincides with the "*p-th*" row of $\mathcal{B}$ up to a multiplicative constant. Therefore $\mathcal{Y}_p$ is a zonal polynomial. ∎

Corollary 3.

$$\mathcal{E}_W \mathcal{Y}_p(\boldsymbol{AW}) = \lambda_{\nu p} \mathcal{Y}_p(\boldsymbol{A\Sigma}), \tag{28}$$

where $\boldsymbol{A}$ is symmetric and $\boldsymbol{W}$ is distributed according to $\mathcal{W}(\boldsymbol{\Sigma}, \nu)$.

Proof. This follows from (26) noting that if $\boldsymbol{W} = \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{W}_1 \boldsymbol{\Sigma}^{\frac{1}{2}}$ then $\boldsymbol{W}_1$ is distributed according to $\mathcal{W}(\boldsymbol{I}_k, \nu)$. ∎

The converse part of Theorem 1 will be strengthened in Theorem 4.1.2 and will be used to show that a particular symmetric polynomial is a zonal polynomial. See Sec. 3.3, Sec. 4.4, and Sec. 4.7.

Remark 6. In (21) zonal polynomials are defined as linear combinations of $\mathcal{U}_q$'s. Therefore by (4) and (5) $\mathcal{Y}_p(\boldsymbol{A})$ makes sense even when $\boldsymbol{A}$ is not symmetric. This has been already used in the form $\mathcal{Y}_p(\boldsymbol{AW})$ in (26). This might be slightly confusing because τ_ν (hence $\boldsymbol{T}_\nu$ and $\boldsymbol{\Xi}$) was defined by considering only symmetric matrices. Indeed in most cases arguments for zonal polynomials are symmetric matrices.

§ 3.2 INTEGRAL IDENTITIES INVOLVING ZONAL POLYNOMIALS

In addition to (3.1.26) the zonal polynomials satisfy other integral identities. The fundamental one (Theorem 1 below) is related to the uniform distribution of orthogonal matrices. The idea of "averaging with respect to the uniform distribution of orthogonal matrices" or "averaging over orthogonal group" was a very important idea of James for the motivation of introducing the zonal polynomials.

A random orthogonal matrix $\boldsymbol{H}$ is said to have the *Haar invariant* distribution or the *uniform distribution* if the distribution of $\boldsymbol{H\Gamma}$ is the same for every orthogonal $\boldsymbol{\Gamma}$. More formally, a probability measure P on the Borel field of orthogonal matrices is *Haar invariant* if

$$P(A) = P(A\boldsymbol{\Gamma}) \tag{1}$$

for every orthogonal $\boldsymbol{\Gamma}$ and every Borel set A. See Anderson (1958), Chapter 13. General theory of Haar measures on topological groups can be found in

Nachbin (1965) or Halmos (1974), for example. For the group of orthogonal matrices, the existence and uniqueness of the Haar invariant distribution can be established easily. For the uniqueness we have

Lemma 1. *Let two probability measures P_1, P_2 satisfy (1). Then $P_1(A) = P_2(A)$ for every Borel set A. Furthermore $P_1(A) = P_1(A')$ where $A' = \{\boldsymbol{H}' \mid \boldsymbol{H} \in A\}$.*

Proof. Let $\boldsymbol{H}_1, \boldsymbol{H}_2$ be independently distributed according to P_1, P_2 respectively. Then

$$(2) \quad Pr(\boldsymbol{H}_1\boldsymbol{H}_2' \in A) = \mathcal{E}_{H_2}\{Pr(\boldsymbol{H}_1\boldsymbol{H}_2' \in A \mid \boldsymbol{H}_2)\} = \mathcal{E}_{H_2}\{P_1(A)\} = P_1(A).$$

Similarly

$$(3) \quad Pr(\boldsymbol{H}_1\boldsymbol{H}_2' \in A) = Pr(\boldsymbol{H}_2\boldsymbol{H}_1' \in A') = \mathcal{E}_{H_1}\{Pr(\boldsymbol{H}_2\boldsymbol{H}_1' \in A' \mid \boldsymbol{H}_1)\} = P_2(A').$$

Hence

$$(4) \qquad P_1(A) = P_2(A').$$

Putting $P_1 = P_2$ we obtain $P_1(A) = P_1(A'), P_2(A) = P_2(A')$. Substituting this into (4) we obtain $P_1(A) = P_2(A)$. ∎

Remark 1. For a more rigorous proof (2) and (3) have to be converted to the form of Fubini's theorem, as is done in standard proofs (see Section 60 of Halmos (1974)). The same remark applies to the proof of Lemma 3 below. Also note that the second assertion of Lemma 1 shows that if $\boldsymbol{H}$ is uniform then $\boldsymbol{H}'$ is uniform.

Existence can be very explicitly established as follows.

Lemma 2. *Let $\boldsymbol{U} = (u_{ij})$ be a $k \times k$ matrix such that u_{ij} are independent standard normal variables. Then with probability 1, $\boldsymbol{U}$ can be uniquely expressed as*

$$(5) \qquad \boldsymbol{U} = \boldsymbol{T}\boldsymbol{H},$$

where $T = (t_{ij})$ is lower triangular with positive diagonal elements and H is orthogonal. Furthermore (i) T and H are independent, (ii) H is uniform, (iii) t_{ij} are all independent and $t_{ii} \sim \chi(k-i+1)$, $t_{ij} \sim \mathcal{N}(0,1)$, $i > j$.

Proof. U is nonsingular with probability 1. Therefore suppose $|U| \neq 0$. Now performing the Gram-Schmidt orthonormalization to the rows of U starting from the first row we obtain $SU = H$ where S is lower triangular with positive diagonal elements and H is orthogonal. Letting $T = S^{-1}$ we obtain (5). Since (5) corresponds to the uniquely defined Gram-Schmidt orthonormalization T, H are unique. Now $W = UU' = TT'$ is distributed according to $\mathcal{W}(I_k, k)$. Hence (iii) follows from Lemma 3.1.3. To show (i) and (ii) we first note that for any orthogonal Γ, $U\Gamma$ has the same distribution as U. Furthermore $U\Gamma = T(H\Gamma)$. Therefore $H\Gamma$ is the resulting orthogonal matrix obtained by performing Gram-Schmidt orthonormalization to the rows of $U\Gamma$ and T is common to U and $U\Gamma$. This implies that given T the conditional distributions of H and $H\Gamma$ are the same. Therefore the conditional distribution of H given T is uniform. Now by unconditioning we see that T and H are independent and H has the uniform distribution. ∎

This lemma has been known for a long time. See Kshirsagar (1959), Example 6 in Chapter 8 of Lehmann (1959), Saw (1970) for example.

Now we prove the following fundamental identity (James (1961a)). The proof is a modification of one in Saw (1977).

Theorem 1. *Let A, B be $k \times k$ symmetric matrices. Then*

$$\mathcal{E}_H \mathcal{Y}_p(AHBH') = \mathcal{Y}_p(A)\mathcal{Y}_p(B)/\mathcal{Y}_p(I_k), \tag{6}$$

where $k \times k$ orthogonal H has the uniform distribution.

Proof. Let $f(A, B)$ denote the left hand side of (6). Let $\lambda(A) = \alpha = (\alpha_1, \ldots, \alpha_k)$ and $\lambda(B) = \beta = (\beta_1, \ldots, \beta_k)$. Let $A = H_1 D_1 H_1'$, $B = H_2 D_2 H_2'$, where H_1, H_2 are orthogonal and $D_1 = \mathrm{diag}(\alpha_1, \ldots, \alpha_k)$, $D_2 = \mathrm{diag}(\beta_1, \ldots, \beta_k)$. Now

$$\begin{aligned} \mathcal{Y}_p(AHBH') &= \mathcal{Y}_p(H_1 D_1 H_1' H H_2 D_2 H_2' H') \\ &= \mathcal{Y}_p(D_1 H_3 D_2 H_3'), \end{aligned} \tag{7}$$

where $H_3 = H_1'HH_2$ which has the uniform distribution. Therefore

$$f(A,B) = \mathcal{E}_H \mathcal{Y}_p(D_1 H D_2 H'). \tag{8}$$

This depends only on $\alpha = (\alpha_1, \ldots, \alpha_k)$, and $\beta = (\beta_1, \ldots, \beta_k)$. Now for any permutation matrix P, $A = (H_1 P)(P' D_1 P)(P' H_1')$. Noting that a permutation matrix is orthogonal we get $\mathcal{E}_H \mathcal{Y}_p\,(D_1 H D_2 H') = \mathcal{E}_H \mathcal{Y}_p(P' D_1 P H D_2 H')$. therefore $f(A,B)$ is symmetric in $\alpha_1, \ldots, \alpha_k$. Similarly it is symmetric in $\beta_1, \ldots, \beta_k$. Now on the left hand side of (6) express $\mathcal{Y}_p$ in terms of $\mathcal{U}_q$'s. Furthermore for each elementary symmetric function $u_r = u_r(AHBH')$ constituting $\mathcal{U}_q$ use the relation (3.1.4). We see that $\mathcal{Y}_p(AHBH')$ and hence $f(A,B)$ are polynomials in $(\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_k)$. Suppose that $f(A,B)$ is completely expanded into monomial terms. Consider the term of the form $c\alpha_1^{q_1} \cdots \alpha_\ell^{q_\ell}$, $(q = (q_1, \ldots, q_\ell) \in \mathcal{P}_n)$ in $f(A,B)$. By symmetry among α_i's $f(A,B)$ has the term $c\alpha_{i_1}^{q_1} \cdots \alpha_{i_\ell}^{q_\ell}$ with the same coefficient c. Collecting these permuted terms in α's we obtain $c\mathcal{M}_q(\alpha)$. However c is a polynomial in β's and by symmetry among β's c can be written as a linear combination of $\mathcal{M}_q(\beta)$'s. Collecting all terms we can write

$$f(A,B) = \sum_{q,q'} a_{qq'} \mathcal{M}_q(\alpha) \mathcal{M}_{q'}(\beta)$$

for some real numbers $a_{qq'}$. Expressing $\mathcal{M}_q$'s in terms of $\mathcal{Y}_q$'s we alternatively have

$$f(A,B) = \sum_{q,q'} c_{qq'} \mathcal{Y}_q(A) \mathcal{Y}_{q'}(B).$$

Note that $c_{qq'} = c_{q'q}$ because $\mathcal{Y}_p(AHBH') = \mathcal{Y}_p(BH'AH)$ and H' has the uniform distribution (see Remark 1). Now let A be distributed independently of H and according to $\mathcal{W}(\Sigma, \nu_0)$ where ν_0 is such that $\lambda_{\nu_0 p}$, $p \in \mathcal{P}$ are all different (see the proof of Lemma 3.1.4). Then by Corollary 3.1.3

$$\mathcal{E}_A \mathcal{E}_H \mathcal{Y}_p(AHBH') = \sum_{q,q'} c_{qq'} \lambda_{\nu_0 q} \mathcal{Y}_q(\Sigma) \mathcal{Y}_{q'}(B). \tag{9}$$

On the other hand taking expectation with respect to A first we obtain

$$\begin{aligned} \mathcal{E}_H \mathcal{E}_A \mathcal{Y}_p(AHBH') &= \lambda_{\nu_0 p} \mathcal{E}_H \mathcal{Y}_p(\Sigma HBH') \\ &= \lambda_{\nu_0 p} \sum_{q,q'} c_{qq'} \mathcal{Y}_q(\Sigma) \mathcal{Y}_{q'}(B). \end{aligned} \tag{10}$$

Therefore

$$0 = \sum_{q,q'} (\lambda_{\nu_0 p} - \lambda_{\nu_0 q}) c_{qq'} \mathcal{Y}_q(\boldsymbol{\Sigma}) \mathcal{Y}_{q'}(\boldsymbol{B}). \tag{11}$$

This holds for any $\boldsymbol{\Sigma}$ and $\boldsymbol{B}$. Hence $(\lambda_{\nu_0 p} - \lambda_{\nu_0 q})c_{qq'} = 0$ for all q, q'. Since $\lambda_{\nu_0 p} \neq \lambda_{\nu_0 q}$ for $p \neq q$ we have $c_{qq'} = 0$ for all q' and all $q \neq p$. But $c_{qq'} = c_{q'q}$. Therefore $c_{qq'} = 0$ unless $q = q' = p$. Therefore

$$\mathcal{E}_H \mathcal{Y}_p(\boldsymbol{AHBH'}) = c_{pp} \mathcal{Y}_p(\boldsymbol{A}) \mathcal{Y}_p(\boldsymbol{B}). \tag{12}$$

Putting $\boldsymbol{B} = \boldsymbol{I}_k$ we obtain

$$\mathcal{Y}_p(\boldsymbol{A}) = c_{pp} \mathcal{Y}_p(\boldsymbol{I}_k) \mathcal{Y}_p(\boldsymbol{A}). \tag{13}$$

Hence $c_{pp} \mathcal{Y}_p(\boldsymbol{I}_k) = 1$ and this proves the theorem. ∎

For more about this proof see Section 4.1.

Theorem 1 implies the following rather strong result.

Theorem 2. *Suppose that a $k \times k$ random symmetric matrix $\boldsymbol{V}$ has a distribution such that for every orthogonal $\boldsymbol{\Gamma}$, $\boldsymbol{\Gamma V \Gamma'}$ has the same distribution as $\boldsymbol{V}$. Then for symmetric $\boldsymbol{A}$*

$$\mathcal{E}_V \mathcal{Y}_p(\boldsymbol{AV}) = c_p \mathcal{Y}_p(\boldsymbol{A}), \tag{14}$$

where

$$c_p = \mathcal{E}_V \{\mathcal{Y}_p(\boldsymbol{V})\} / \mathcal{Y}_p(\boldsymbol{I}_k). \tag{15}$$

Proof. As in the proof of Lemma 3.1.2 $\mathcal{E}_V \mathcal{Y}_p(\boldsymbol{AV}) \in V_n$. Now since the distribution of $\boldsymbol{\Gamma V \Gamma'}$ is the same as $\boldsymbol{V}$ we have

$$\mathcal{E}_V \mathcal{Y}_p(\boldsymbol{A \Gamma V \Gamma'}) = \mathcal{E}_V \mathcal{Y}_p(\boldsymbol{AV}). \tag{16}$$

Letting $\boldsymbol{\Gamma}$ be uniformly distributed independently of $\boldsymbol{V}$

$$\begin{aligned} \mathcal{E}_V \mathcal{Y}_p(\boldsymbol{AV}) &= \mathcal{E}_\Gamma \mathcal{E}_V \mathcal{Y}_p(\boldsymbol{A\Gamma V\Gamma'}) \\ &= \mathcal{E}_V \mathcal{E}_\Gamma \mathcal{Y}_p(\boldsymbol{A\Gamma V\Gamma'}) \\ &= \mathcal{E}_V \{\mathcal{Y}_p(\boldsymbol{A}) \mathcal{Y}_p(\boldsymbol{V}) / \mathcal{Y}_p(\boldsymbol{I}_k)\} \\ &= \mathcal{Y}_p(\boldsymbol{A}) \mathcal{E}_V \{\mathcal{Y}_p(\boldsymbol{V})\} / \mathcal{Y}_p(\boldsymbol{I}_k). \end{aligned} \tag{17}$$

∎

Remark 2. In the sequel we call the distribution of $\boldsymbol{V}$ "orthogonally invariant" if it satisfies the condition of Theorem 2.

Although Theorem 2 has not been explicitly stated, it has been implicitly used for several cases; first with the multivariate beta distribution by Constantine (1963), later with the inverted Wishart distribution by Khatri (1966) etc. These cases will be examined in Section 4.3 together with the evaluation of c_p for each case.

We note that Theorem 2 is a generalization of Theorem 3.1.1. Now suppose that we chose $\boldsymbol{V}$ which has an orthogonally invariant distribution instead of the Wishart matrix $\boldsymbol{W}$ for the construction of zonal polynomials in Section 3.1. Then the construction could have been carried out in exactly the same way provided that c_p, $p \in P_n$ in Theorem 2 are all distinct for $\boldsymbol{V}$. Furthermore if we examine the proof of Theorem 1 closely we find that we could take $\boldsymbol{A} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{V}\boldsymbol{\Sigma}^{\frac{1}{2}}$ in (9) and (10). Once Theorem 1 is proved the identity involving the Wishart distribution can be derived as a special case. Although the Wishart distribution seems to be a natural candidate to take for our construction, we could have used any orthogonally invariant distribution from a purely logical point of view.

Orthogonally invariant distributions are characterized as follows.

Lemma 3. *Let $\boldsymbol{V} = \boldsymbol{H}\boldsymbol{D}\boldsymbol{H}'$ where $\boldsymbol{H}$ is orthogonal and $\boldsymbol{D}$ is diagonal. Let $\boldsymbol{H}$ and $\boldsymbol{D}$ be independently distributed such that $\boldsymbol{H}$ has the uniform distribution. (Diagonal elements of $\boldsymbol{D}$ can have any distribution.) Then $\boldsymbol{V}$ has an orthogonally invariant distribution. Conversely all orthogonally invariant distributions can be obtained in this way.*

Proof. The first part of the lemma is obvious. To prove the converse suppose that $\boldsymbol{V}$ has an orthogonally invariant distribution. Now we form a new random matrix $\tilde{\boldsymbol{V}} = \boldsymbol{H}\boldsymbol{V}\boldsymbol{H}'$ where $\boldsymbol{H}$ has the uniform distribution independently of $\boldsymbol{V}$. Then $\tilde{\boldsymbol{V}}$ has the same distribution as $\boldsymbol{V}$ because for any Borel set A

$$
\begin{aligned}
Pr(\tilde{\boldsymbol{V}} \in A) &= \mathcal{E}_H\{Pr(\boldsymbol{H}\boldsymbol{V}\boldsymbol{H}' \in A \mid \boldsymbol{H})\} \\
&= \mathcal{E}_H\{Pr(\boldsymbol{V} \in A)\} = Pr(\boldsymbol{V} \in A).
\end{aligned}
\tag{18}
$$

Now we evaluate $Pr(\tilde{\boldsymbol{V}} \in A)$ by conditioning on $\boldsymbol{V}$. For fixed $\boldsymbol{V}$ we can write $\boldsymbol{V} = \boldsymbol{\Gamma D \Gamma'}$ where $\boldsymbol{\Gamma}$ is orthogonal and $\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_k)$. We require $d_1 \geq \cdots \geq d_k$ then $\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{V})$ is unique. Then

$$
\begin{aligned}
Pr(\tilde{\boldsymbol{V}} \in A \mid \boldsymbol{V}) &= Pr(\boldsymbol{H\Gamma D}(\boldsymbol{V})\boldsymbol{\Gamma' H'} \in A \mid \boldsymbol{V}) \\
&= Pr(\boldsymbol{HD}(\boldsymbol{V})\boldsymbol{H'} \in A \mid \boldsymbol{V}).
\end{aligned} \tag{19}
$$

Note that we replaced $\boldsymbol{H\Gamma}$ by $\boldsymbol{H}$ since $\boldsymbol{H\Gamma}$ has the uniform distribution. Hence

$$
\begin{aligned}
Pr(\tilde{\boldsymbol{V}} \in A) &= \mathcal{E}_V\{Pr(\tilde{\boldsymbol{V}} \in A \mid \boldsymbol{V})\} \\
&= \mathcal{E}_V\{Pr(\boldsymbol{HD}(\boldsymbol{V})\boldsymbol{H'} \in A \mid \boldsymbol{V})\} \\
&= Pr(\boldsymbol{HD}(\boldsymbol{V})\boldsymbol{H'} \in A).
\end{aligned} \tag{20}
$$

This proves the lemma. ∎

Remark 3. Note that the set of orthogonally invariant distributions is convex with respect to taking mixture of distributions. Lemma 3 implies that the extreme points of this convex set are given by those distributions for which $\boldsymbol{D}$ is degenerate.

We can replace $\boldsymbol{H}$ in Theorem 1 by $\boldsymbol{U}$ whose elements are independent normal variables.

Theorem 3. *Let $\boldsymbol{U} = (u_{ij})$ be a $k \times k$ matrix such that u_{ij} are independent standard normal variables. Then for symmetric $\boldsymbol{A}, \boldsymbol{B}$*

$$
\mathcal{E}_U \mathcal{Y}_p(\boldsymbol{AUBU'}) = \frac{\lambda_{kp}}{\mathcal{Y}_p(\boldsymbol{I}_k)} \mathcal{Y}_p(\boldsymbol{A}) \mathcal{Y}_p(\boldsymbol{B}). \tag{21}
$$

Proof. By Lemma 2 $\boldsymbol{U} = \boldsymbol{TH}$. Then

$$
\begin{aligned}
\mathcal{E}_U \mathcal{Y}_p(\boldsymbol{AUBU'}) &= \mathcal{E}_T \mathcal{E}_H \mathcal{Y}_p(\boldsymbol{ATHBH'T'}) \\
&= \mathcal{E}_T \mathcal{E}_H \mathcal{Y}_p(\boldsymbol{T'ATHBH'}) \\
&= \mathcal{E}_T \mathcal{Y}_p(\boldsymbol{T'AT}) \mathcal{Y}_p(\boldsymbol{B}) / \mathcal{Y}_p(\boldsymbol{I}_k) \\
&= \mathcal{E}_T \mathcal{Y}_p(\boldsymbol{ATT'}) \mathcal{Y}_p(\boldsymbol{B}) / \mathcal{Y}_p(\boldsymbol{I}_k) \\
&= \frac{\lambda_{kp}}{\mathcal{Y}_p(\boldsymbol{I}_k)} \mathcal{Y}_p(\boldsymbol{A}) \mathcal{Y}_p(\boldsymbol{B}).
\end{aligned} \tag{22}
$$

We used the fact that $\boldsymbol{TT'} = \boldsymbol{UU'} \sim \mathcal{W}(\boldsymbol{I}_k, k)$. ∎

Theorem 3 leads to the following important observation.

Theorem 4. *$b_p \equiv \lambda_{kp}/\mathcal{Y}_p(\boldsymbol{I}_k)$ is a constant independent of k.*

Proof. Let $\boldsymbol{A},\boldsymbol{B}$ be augmented by zeros as

$$\tilde{\boldsymbol{A}} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}, \quad k_1 \times k_1, \qquad \tilde{\boldsymbol{B}} = \begin{pmatrix} \boldsymbol{B} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}, \quad k_1 \times k_1.$$

Then $\mathcal{Y}_p(\tilde{\boldsymbol{A}}) = \mathcal{Y}_p(\boldsymbol{A})$, $\mathcal{Y}_p(\tilde{\boldsymbol{B}}) = \mathcal{Y}_p(\boldsymbol{B})$, and $\mathcal{Y}_p(\tilde{\boldsymbol{A}}\tilde{\boldsymbol{U}}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{U}}') = \mathcal{Y}_p(\boldsymbol{AUBU}')$ where $\tilde{\boldsymbol{U}}(k_1 \times k_1)$ is obtained by adding independent standard normal variables to $\boldsymbol{U}$. Now (21) implies the result. ∎

We evaluate the b_p's for a particular normalization of zonal polynomial to be denoted as ${}_1\mathcal{Y}_p$ in Section 4.2. Corresponding to Theorem 2, Theorem 3 can be generalized as follows.

Theorem 5. *Let $\boldsymbol{X}$ be a $k \times k$ random matrix (not necessarily symmetric) such that for every orthogonal $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$, the distribution of $\boldsymbol{\Gamma}_1\boldsymbol{X}\boldsymbol{\Gamma}_2$ is the same as the distribution of $\boldsymbol{X}$. Then for symmetric $\boldsymbol{A},\boldsymbol{B}$*

$$\mathcal{E}_X \mathcal{Y}_p(\boldsymbol{AXBX}') = \gamma_p \mathcal{Y}_p(\boldsymbol{A})\mathcal{Y}_p(\boldsymbol{B}), \tag{23}$$

where

$$\gamma_p = \mathcal{E}_X\{\mathcal{Y}_p(\boldsymbol{XX}')\}/\{\mathcal{Y}_p(\boldsymbol{I}_k)\}^2. \tag{24}$$

Proof. For any orthogonal $\boldsymbol{\Gamma}_1$

$$\mathcal{E}_X \mathcal{Y}_p(\boldsymbol{AXBX}') = \mathcal{E}_X \mathcal{Y}_p(\boldsymbol{AX\Gamma}_1\boldsymbol{B\Gamma}_1{}'\boldsymbol{X}'). \tag{25}$$

Letting $\boldsymbol{\Gamma}_1$ be uniformly distributed we obtain

$$\begin{aligned} \mathcal{E}_X \mathcal{Y}_p(\boldsymbol{AXBX}') &= \frac{\mathcal{Y}_p(\boldsymbol{B})}{\mathcal{Y}_p(\boldsymbol{I}_k)} \mathcal{E}_X \mathcal{Y}_p(\boldsymbol{X}'\boldsymbol{AX}) \\ &= \frac{\mathcal{Y}_p(\boldsymbol{B})}{\mathcal{Y}_p(\boldsymbol{I}_k)} \mathcal{E}_X \mathcal{Y}_p(\boldsymbol{AXX}'). \end{aligned} \tag{26}$$

Now $\boldsymbol{V} = \boldsymbol{XX}'$ has an orthogonally invariant distribution because $\boldsymbol{\Gamma}_2\boldsymbol{V\Gamma}_2{}' = (\boldsymbol{\Gamma}_2\boldsymbol{X})(\boldsymbol{\Gamma}_2\boldsymbol{X})'$. Therefore by Theorem 2

$$\mathcal{E}_X \mathcal{Y}_p(\boldsymbol{AXX}') = \mathcal{Y}_p(\boldsymbol{A})\mathcal{E}_X\{\mathcal{Y}_p(\boldsymbol{XX}')\}/\mathcal{Y}_p(\boldsymbol{I}_k). \tag{27}$$

Substituting (27) into (26) we obtain the theorem. ∎

Remark 4. We call the distribution of $\boldsymbol{X}$ "orthogonally biinvariant" if it satisfies the condition of Theorem 5.

Corresponding to Lemma 3 we have

Lemma 4. *Let $\boldsymbol{X} = \boldsymbol{H}_1\boldsymbol{D}\boldsymbol{H}_2$ where $\boldsymbol{H}_1, \boldsymbol{H}_2$ are orthogonal and $\boldsymbol{D}$ is diagonal. Let $\boldsymbol{H}_1, \boldsymbol{H}_2, \boldsymbol{D}$ be independently distributed such that $\boldsymbol{H}_1, \boldsymbol{H}_2$ have the uniform distribution. ($\boldsymbol{D}$ can have any distribution.) Then $\boldsymbol{X}$ has an orthogonally biinvariant distribution. Conversely all orthogonally biinvariant distributions can be obtained in this way.*

The proof is entirely analogous to the proof of Lemma 3, therefore we omit it.

Remark 5. The notion of orthogonal biinvariance can be applied to rectangular matrices. If $\boldsymbol{X}$ is $k \times m$ in Theorem 5 we obtain

$$\gamma_p = \frac{\mathcal{E}_X \mathcal{Y}_p(\boldsymbol{X}\boldsymbol{X}')}{\mathcal{Y}_p(\boldsymbol{I}_k)\mathcal{Y}_p(\boldsymbol{I}_m)} \tag{28}$$

and in Lemma 4 (for $k \leq m$) we replace $\boldsymbol{X} = \boldsymbol{H}_1\boldsymbol{D}\boldsymbol{H}_2$ by $\boldsymbol{X} = \boldsymbol{H}_1(\boldsymbol{D}, \boldsymbol{0})\boldsymbol{H}_2$.

In the sequel we almost exclusively work with the Wishart and the normal distributions. But in view of Theorem 2 and Theorem 5 there could be other distributions which give information on various aspects of zonal polynomials.

§ 3.3 AN INTEGRAL REPRESENTATION OF ZONAL POLYNOMIALS

We prove an integral representation by Kates (1980) which shows that (i) zonal polynomials are positive for positive definite $\boldsymbol{A}$ and increasing in each root of $\boldsymbol{A}$, (ii) in the normalization Z_p defined below the coefficients a_{pq} in $Z_p = \sum a_{pq}\mathcal{M}_q$ are nonnegative integers. The derivation by Kates is rather abstract but the integral representation can be proved directly in our framework. The representation can be formulated in several ways. James (1973) derived one involving uniform orthogonal matrix. We discuss these variations in Section 4.7.

From Theorem 3.2.4 we see that a constant b_p or equivalently the value of a zonal polynomial at I_k describes a particular normalization. The normalization Z_p is the simplest one in this sense.

Definition 1. A particular normalization of a zonal polynomial denoted by Z_p is defined by

$$(1) \qquad Z_p(I_k) = \lambda_{kp},$$

or $b_p = 1$ in Theorem 3.2.4.

Theorem 1. *(Kates, 1980) Let $p = (p_1, \ldots, p_\ell)$. For $k \times k$ symmetric A*

$$(2) \qquad Z_p(A) = \mathcal{E}_U\{\Delta_1^{p_1-p_2}\Delta_2^{p_2-p_3}\cdots\Delta_\ell^{p_\ell}\},$$

where $\Delta_i = UAU'(1, \ldots, i)$ is the determinant of the $i \times i$ upper left minor of UAU' and U is a $k \times k$ random matrix whose entries are independent standard normal variables.

Proof. For symmetric A let

$$(3) \qquad f(A) = \mathcal{E}_U\{\Delta_1^{p_1-p_2}\cdots\Delta_\ell^{p_\ell}\}.$$

It can be routinely checked that f is a homogeneous symmetric polynomial of degree $n = |p|$ in the roots of A. Furthermore augmenting A to $\tilde{A}$ $(k_1 \times k_1)$ by adding zeros and augmenting U to $\tilde{U}$ by adding independent standard normal variables do not change the upper left part of UAU'. Therefore (3) does not depend on k. Hence $f \in V_n$. Note that we can extend the definition of f to nonsymmetric matrices as well (see Remark 3.1.6). Now we want to show

$$(4) \qquad (\tau_\nu f)(A) = \lambda_{\nu p} f(A)$$

for all sufficiently large ν and for all symmetric A. Let

$$\tilde{A} = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \quad : \nu \times \nu$$

and $\tilde{\boldsymbol{W}} = \boldsymbol{Y}'\boldsymbol{Y}$ where $\boldsymbol{Y}$ is a $\nu \times \nu$ matrix whose entries are standard normal variables. Then

$$
\begin{aligned}
(\tau_\nu f)(\boldsymbol{A}) &= \mathcal{E}_{\tilde{W}}\{f(\tilde{\boldsymbol{A}}\tilde{\boldsymbol{W}})\} = \mathcal{E}_Y\{f(\boldsymbol{Y}\tilde{\boldsymbol{A}}\boldsymbol{Y}')\} \\
&= \mathcal{E}_Y \mathcal{E}_{\tilde{U}}\{\prod_{i=1}^{\ell} [\tilde{\boldsymbol{U}}\boldsymbol{Y}\tilde{\boldsymbol{A}}\boldsymbol{Y}'\tilde{\boldsymbol{U}}'(1,\dots,i)]^{p_i - p_{i+1}}\} \\
&= \mathcal{E}_{\tilde{U}} \mathcal{E}_Y\{\prod_{i=1}^{\ell} [\boldsymbol{Y}\tilde{\boldsymbol{U}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{U}}'\boldsymbol{Y}'(1,\dots,i)]^{p_i - p_{i+1}}\}.
\end{aligned}
\tag{5}
$$

We switched $\tilde{\boldsymbol{U}}$ and $\boldsymbol{Y}$ because they have the same distribution. Now by Lemma 3.2.2 $\boldsymbol{Y} = \boldsymbol{T}\boldsymbol{H}$ and $\boldsymbol{H}$ can be absorbed into $\boldsymbol{U}$. Therefore

$$
\begin{aligned}
&\mathcal{E}_{\tilde{U}} \mathcal{E}_Y\{\prod_{i=1}^{\ell} [\boldsymbol{Y}\tilde{\boldsymbol{U}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{U}}'\boldsymbol{Y}'(1,\dots,i)]^{p_i - p_{i+1}}\} \\
&\quad= \mathcal{E}_{\tilde{U}} \mathcal{E}_T\{\prod_{i=1}^{\ell} [\boldsymbol{T}\tilde{\boldsymbol{U}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{U}}'\boldsymbol{T}'(1,\dots,i)]^{p_i - p_{i+1}}\} \\
&\quad= \mathcal{E}_{\tilde{U}} \mathcal{E}_T\{\prod_{i=1}^{\ell} (t_{11}^2 \cdots t_{ii}^2)^{p_i - p_{i+1}} [\tilde{\boldsymbol{U}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{U}}'(1,\dots,i)]^{p_i - p_{i+1}}\} \\
&\quad= \lambda_{\nu p} \mathcal{E}_{\tilde{U}}\{\prod_{i=1}^{\ell} [\tilde{\boldsymbol{U}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{U}}'(1,\dots,i)]^{p_i - p_{i+1}}\} \\
&\quad= \lambda_{\nu p} f(\boldsymbol{A}).
\end{aligned}
\tag{6}
$$

Hence $f = \mathcal{Y}_p$ by Theorem 3.1.1. Putting $\boldsymbol{A} = \boldsymbol{I}_k$ we obtain

$$
f(\boldsymbol{I}_k) = \mathcal{E}_W\{\boldsymbol{W}(1)^{p_1 - p_2} \cdots \boldsymbol{W}(1,\dots,\ell)^{p_\ell}\},
\tag{7}
$$

where $\boldsymbol{W} \sim \mathcal{W}(\boldsymbol{I}_k, k)$. Again by the triangular decomposition $\boldsymbol{W} = \boldsymbol{T}\boldsymbol{T}'$ (Lemma 3.1.3) we obtain $f(\boldsymbol{I}_k) = \lambda_{kp}$. Therefore $f = Z_p$. ∎

Note that the coefficients of the monomial terms in Z_p are integers, being the expected value of sum of products of independent standard normal variables. Furthermore if $\boldsymbol{A} = \mathrm{diag}(\alpha_1, \dots, \alpha_k)$ then by the Binet-Cauchy theorem (see Gantmacher (1959) for example)

$$
\begin{aligned}
&\boldsymbol{UAU'}(1,\ldots,r)\\
(8)\qquad &= \sum_{i_1<\cdots<i_r}\sum_{j_1<\cdots<j_r} \boldsymbol{U}\begin{pmatrix}1,\ldots,r\\ i_1,\ldots,i_r\end{pmatrix}\boldsymbol{A}\begin{pmatrix}i_1,\ldots,i_r\\ j_1,\ldots,j_r\end{pmatrix}\boldsymbol{U'}\begin{pmatrix}j_1,\ldots,j_r\\ 1,\ldots,r\end{pmatrix}\\
&= \sum_{i_1<\cdots<i_r} \alpha_{i_1}\cdots\alpha_{i_r}\left\{\boldsymbol{U}\begin{pmatrix}1,\ldots,r\\ i_1,\ldots,i_r\end{pmatrix}\right\}^2,
\end{aligned}
$$

where $\boldsymbol{B}\binom{i_1,\ldots,i_r}{j_1,\ldots,j_r}$ denotes the determinant of a minor formed by rows $i_1,\ldots,i_r$ and columns $j_1,\ldots,j_r$ of $\boldsymbol{B}$. (8) is obviously increasing in each α_i when $\boldsymbol{A}$ is positive definite. Furthermore coefficients for monomial terms are nonnegative. These points are discussed in Kates (1980). For more about this see Section 4.1. Generalizations of Theorem 1 will be discussed in Section 4.7.

§ 3.4 A GENERATING FUNCTION OF ZONAL POLYNOMIALS

One of the main contributions of Saw (1977) is his generating function which gives a relatively simple way of computing zonal polynomials. Let

$$
(1)\qquad (\operatorname{tr}\boldsymbol{C})^n = \mathcal{U}_p(\boldsymbol{C}) = \sum_{p\in\mathcal{P}_n} d_p Z_p(\boldsymbol{C}).
$$

Let $\boldsymbol{C} = \boldsymbol{AUBU'}$ where $\boldsymbol{A} = \operatorname{diag}(\alpha_1,\ldots,\alpha_k)$, $\boldsymbol{B} = \operatorname{diag}(\beta_1,\ldots,\beta_k)$ and the elements of $\boldsymbol{U}$ are independent standard normal variables. Then by Theorem 3.2.3

$$
\begin{aligned}
&\mathcal{E}_U(\operatorname{tr}\boldsymbol{AUBU'})^n\\
&= \sum_{p\in\mathcal{P}_n} d_p \mathcal{E}_U Z_p(\boldsymbol{AUBU'})\\
(2)\qquad &= \sum_{p\in\mathcal{P}_n} d_p \frac{\lambda_{kp}}{Z_p(\boldsymbol{I}_k)} Z_p(\boldsymbol{A})Z_p(\boldsymbol{B})\\
&= \sum_{p\in\mathcal{P}_n} d_p Z_p(\boldsymbol{A})Z_p(\boldsymbol{B}).
\end{aligned}
$$

Therefore for sufficiently small θ

$$
\begin{aligned}
(3)\qquad & \mathcal{E}_U\{\exp(\theta \operatorname{tr} \boldsymbol{AUBU'})\} \\
&= \mathcal{E}_U\{\sum_{n=0}^{\infty}(\theta^n/n!)(\operatorname{tr} \boldsymbol{AUBU'})^n\} \\
&= \sum_{n=0}^{\infty}(\theta^n/n!)\sum_{p\in P_n} d_p Z_p(\boldsymbol{A})Z_p(\boldsymbol{B}).
\end{aligned}
$$

On the other hand

$$
(4)\qquad \operatorname{tr} \boldsymbol{AUBU'} = \sum_{i,j}^{k} \alpha_i\beta_j u_{ij}^2.
$$

Hence for sufficiently small θ

$$
\begin{aligned}
(5)\qquad \mathcal{E}_U\{\exp(\theta \operatorname{tr} \boldsymbol{AUBU'})\} &= \mathcal{E}_U\{\exp(\theta \sum_{i,j}^{k} \alpha_i\beta_j u_{ij}^2)\} \\
&= \prod_{i,j}^{k}(1-2\theta\alpha_i\beta_j)^{-\frac{1}{2}}.
\end{aligned}
$$

From (3) and (5) we obtain

Theorem 1.

$$
(6)\qquad \prod_{i,j}^{k}(1-2\theta\alpha_i\beta_j)^{-\frac{1}{2}} = \sum_{n=0}^{\infty}(\theta^n/n!)\sum_{p\in P_n} d_p Z_p(\boldsymbol{A})Z_p(\boldsymbol{B}).
$$

The left hand side of (6) can be expanded as follows.

$$
\begin{aligned}
&\prod_{i,j}^{k}(1-2\theta\alpha_i\beta_j)^{-\frac{1}{2}} \\
&\quad = \exp\{\log\prod_{i,j}^{k}(1-2\theta\alpha_i\beta_j)^{-\frac{1}{2}}\} \\
&\quad = \exp\{\frac{1}{2}\sum_{i,j}^{k}\sum_{r=1}^{\infty}\frac{(2\theta)^r}{r}\alpha_i^r\beta_j^r\} \\
&\quad = \exp\{\frac{1}{2}\sum_{r=1}^{\infty}\frac{(2\theta)^r}{r}t_r(\boldsymbol{A})t_r(\boldsymbol{B})\} \qquad (7)\\
&\quad = \sum_{n=0}^{\infty}\sum_{p\in P_n}(2\theta)^n T_p(\boldsymbol{A})T_p(\boldsymbol{B}) \\
&\qquad\quad \cdot\frac{1}{p_1!}\left(\frac{1}{2}\right)^{p_1}\binom{p_1}{p_1-p_2,p_2-p_3,\dots,p_{\ell(p)}}\left(\prod_{r=1}^{\ell(p)}r^{p_r-p_{r+1}}\right)^{-1} \\
&\quad = \sum_{n=0}^{\infty}(\theta^n/n!)\sum_{p\in P_n}c_p T_p(\boldsymbol{A})T_p(\boldsymbol{B}),
\end{aligned}
$$

where

$$
c_p = |p|!2^{|p|-h(p)}\left\{\prod_{r=1}^{\ell(p)}r^{p_r-p_{r+1}}(p_r-p_{r+1})!\right\}^{-1} \tag{8}
$$

The fourth equality follows from the fact that T_p being a product of p_1 terms comes only from the p_1-*th* power term in the expansion of exp. Comparing the coefficient of θ^n in (6) and (7) we obtain

$$
\sum_{p\in P_n}d_p Z_p(\boldsymbol{A})Z_p(\boldsymbol{B}) = \sum_{p\in P_n}c_p T_p(\boldsymbol{A})T_p(\boldsymbol{B}). \tag{9}
$$

Note that c_p is positive for every $p\in P_n$. Hence the right hand side of (9) is a positive definite quadratic form. Now the left hand side of (9) is the same positive definite quadratic form expressed with the different basis $\{Z_p\}$. The positive definiteness implies $d_p > 0$ for every $p\in P_n$. Now let $\boldsymbol{D} = \mathrm{diag}(d_p, p \in$

$\mathcal{P}_n$), $\boldsymbol{C} = \text{diag}(c_p, p \in \mathcal{P}_n)$. We recall that $\boldsymbol{\mathcal{Z}} = \boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{U}}$ where $\boldsymbol{\mathcal{B}}$ is upper triangular and $\boldsymbol{\mathcal{T}} = \boldsymbol{F}\boldsymbol{\mathcal{U}}$ where $\boldsymbol{F}$ is lower triangular (see(2.2.28)). Therefore in matrix notation (9) is written as

$$\mathcal{U}(\boldsymbol{A})'\mathcal{B}'\boldsymbol{D}\mathcal{B}\mathcal{U}(\boldsymbol{B}) = \mathcal{U}(\boldsymbol{A})'\boldsymbol{F}'\boldsymbol{C}\boldsymbol{F}\mathcal{U}(\boldsymbol{B}), \tag{10}$$

or

$$\mathcal{B}'\boldsymbol{D}\mathcal{B} = \boldsymbol{F}'\boldsymbol{C}\boldsymbol{F}. \tag{11}$$

We note that the left hand side and the right hand side correspond to two different triangular decompositions of the same symmetric positive definite matrix. $\boldsymbol{F}$ can be computed from (2.2.24) or alternatively $\boldsymbol{F}$ can be obtained from tables given in David, Kendall, and Barton (1966) for $n \leq 12$. Therefore we can compute the right hand side of (9) relatively easily, then we decompose the resulting positive definite matrix as $\mathcal{B}'\boldsymbol{D}\mathcal{B}$. Diagonal elements of $\mathcal{B}$ corresponding to Z_p is obtained in (4.2.7). This determines $\boldsymbol{D}$ and $\mathcal{B}$ uniquely.

Remark 1. In terms of $\mathcal{M}_p$'s (11) can be written as $\boldsymbol{A}'\mathcal{B}'\boldsymbol{D}\mathcal{B}\boldsymbol{A}=\boldsymbol{A}'\boldsymbol{F}'\boldsymbol{C}\boldsymbol{F}\boldsymbol{A}$ where $\boldsymbol{A}$ is given in (2.2.14). Saw (1977) defined zonal polynomials or the upper triangular coefficient matrix $\mathcal{B}\boldsymbol{A}$ by this relation and derived the first part of Theorem 3.1.1 from this definition. It seems that (11) should be looked at as providing a convenient algorithm for obtaining $\mathcal{B}$ rather than providing a definition of zonal polynomials because it lacks the conceptual motivation necessary for a definition.

Actually d_p is known to be (James (1964), formula(18))

$$\begin{aligned} d_p &= \chi_{[2p]}(1)2^n n!/(2n)! \\ &= \frac{2^n n! \prod_{i<j}(2p_i - 2p_j - i + j)}{\prod_{i=1}^{\ell(p)}(2p_i + \ell(p) - i)!}, \end{aligned} \tag{12}$$

where $n = |p|$ and $\chi_{[2p]}(1) = (2n)! \prod_{i<j}(2p_i - 2p_j - i + j) \;/ \prod_{i=1}^{\ell(p)}(2p_i + \ell(p) - i)!$ is "the dimension of the representation $(2p) = (2p_1, \ldots, 2p_{\ell(p)})$ of the symmetric group on $2n$ symbols." This is one thing we were unable to

obtain by our elementary approach. It was obtained by James (1961) using group representation theory. We will discuss this point again in Section 4.2 and Section 5.4.

$d_p Z_p$ is usually denoted by C_p so that (1) can be written simply as

$$(\operatorname{tr} \boldsymbol{A})^n = \sum_{p \in \mathcal{P}_n} C_p(\boldsymbol{A}). \tag{13}$$

This notation often makes it simpler to write down various noncentral densities. Our last item in this section is related to this point.

Lemma 1.

$$\begin{aligned} \mathcal{E}_H(\operatorname{tr} \boldsymbol{AH})^{2n} &= \sum_{p \in \mathcal{P}_n} \frac{2^n n! d_p}{(2n)! \lambda_{kp}} Z_p(\boldsymbol{AA}') \\ &= \sum_{p \in \mathcal{P}_n} \frac{2^n n!}{(2n)! \lambda_{kp}} C_p(\boldsymbol{AA}'), \end{aligned} \tag{14}$$

where $k \times k$ orthogonal $\boldsymbol{H}$ is uniformly distributed.

Proof. Let the singular value decomposition be $\boldsymbol{A} = \boldsymbol{\Gamma}_1 \boldsymbol{D} \boldsymbol{\Gamma}_2$ where $\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2$ are orthogonal, $\boldsymbol{D} = \operatorname{diag}(\delta_1, \ldots, \delta_k)$ and $\theta_i = \delta_i^2$, $i = 1, \ldots, k$ are the characteristic roots of $\boldsymbol{AA}'$. Then $(\operatorname{tr} \boldsymbol{AH})^{2n} = (\operatorname{tr} \boldsymbol{\Gamma}_1 \boldsymbol{D} \boldsymbol{\Gamma}_2 \boldsymbol{H})^{2n} = (\operatorname{tr} \boldsymbol{D} \boldsymbol{\Gamma}_2 \boldsymbol{H} \boldsymbol{\Gamma}_1)^{2n}$ and $\boldsymbol{\Gamma}_2 \boldsymbol{H} \boldsymbol{\Gamma}_1$ has the same distribution as $\boldsymbol{H}$. Therefore $\mathcal{E}_H(\operatorname{tr} \boldsymbol{AH})^{2n}$ is a $2n$-*th* degree homogeneous polynomial in $\delta_1, \ldots, \delta_k$. Furthermore the order of $\delta_1, \ldots, \delta_k$ and the sign for each δ_i are arbitrary in the singular value decomposition. It follows that $\mathcal{E}_H(\operatorname{tr} \boldsymbol{AH})^{2n}$ is a homogeneous symmetric polynomial of degree n in $(\theta_1, \ldots, \theta_k)$. Therefore we can write

$$\mathcal{E}_H(\operatorname{tr} \boldsymbol{AH})^{2n} = \sum_{p \in \mathcal{P}_n} a_p Z_p(\boldsymbol{AA}'), \tag{15}$$

for some real numbers a_p. Now let $\boldsymbol{A} = \operatorname{diag}(\alpha_1, \ldots, \alpha_k)$ and $\boldsymbol{U} = (u_{ij})$ be as before. Then $\operatorname{tr} \boldsymbol{AU} = \sum \alpha_i u_{ii}$ is distributed according to $\mathcal{N}(0, \sum \alpha_i^2)$. Hence

$$\begin{aligned} \mathcal{E}_U(\operatorname{tr} \boldsymbol{AU})^{2n} &= \left(\sum \alpha_i^2\right)^n \cdot 1 \cdot 3 \cdots (2n-1) \\ &= \frac{(2n)!}{2^n n!} (\operatorname{tr} \boldsymbol{AA}')^n \\ &= \sum_{p \in \mathcal{P}_n} \frac{(2n)! d_p}{2^n n!} Z_p(\boldsymbol{AA}'). \end{aligned} \tag{16}$$

On the other hand by Lemma 3.2.2 and (15)

$$
\begin{aligned}
\mathcal{E}_U(\operatorname{tr} \boldsymbol{A}\boldsymbol{U})^{2n} &= \mathcal{E}_{T,H}(\operatorname{tr} \boldsymbol{A}\boldsymbol{T}\boldsymbol{H})^{2n} \\
&= \sum_{p \in \mathcal{P}_n} a_p \mathcal{E}_T Z_p(\boldsymbol{A}\boldsymbol{T}\boldsymbol{T}'\boldsymbol{A}') \\
&= \sum_{p \in \mathcal{P}_n} a_p \lambda_{kp} Z_p(\boldsymbol{A}\boldsymbol{A}').
\end{aligned} \tag{17}
$$

Comparing (16) and (17) we obtain (14). ∎