THE 70TH ANNIVERSARY OF THE DISTRIBUTION OF RANDOM MATRICES: A SURVEY

by

Ingram Olkin

February 2000

Department of Statistics
STANFORD UNIVERSITY
Stanford, California
THE 70TH ANNIVERSARY OF THE DISTRIBUTION OF RANDOM MATRICES: A SURVEY

by

INGRAM OLKIN
Department of Statistics
Stanford University

February 2000

Department of Statistics
Sequoia Hall
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu
THE 70TH ANNIVERSARY OF THE
DISTRIBUTION OF RANDOM MATRICES:
A SURVEY

by

Ingram Olkin
Stanford University

Abstract

Although distribution theory dates back over a century, the distributions were essentially univariate. The distribution of the sample covariance matrix was perhaps the beginning of a theory of distributions of random matrices. When the underlying distribution is multivariate normal, the distribution of the sample covariance matrix is the Wishart distribution. We here provide a review of the distribution of a variety of matrices as they arise from matrix factorizations.
February 15, 2000

THE 70TH ANNIVERSARY OF THE
DISTRIBUTION OF RANDOM MATRICES:
A SURVEY

Ingram Olkin
Stanford University

Given the distribution of a scalar random variable $X$ it is generally straightforward to obtain the distribution of a function of $X$. This is far from the case when $X$ is a random matrix. Indeed, methods for obtaining the distributions of functions of random matrices has been a continuing challenge for the last half-century. One might argue that the origin of multivariate distribution theory occurred in 1928 with John Wishart's derivation of the distribution of the covariance matrix of a sample from a standard multivariate normal distribution, thereby generalizing the univariate chi-square distribution. Although other multivariate distributions were known earlier, they did not involve random matrices. There are currently a variety of very different derivations of this distribution, attesting to the tantalizing aspect of this problem. In this paper we provide a review of distributional results for random matrices. The choice is personal, and we do not intend to be comprehensive.

As a starting point, let $X$ be a $p \times p$ random matrix. For simplicity of exposition we assume regularity conditions, which at the very least is an absolutely continuous density $f(X)$. We write $S > 0$ to mean that the symmetric matrix $S$ is positive definite. Eigenvalues of a matrix $X$ are denoted $\lambda(X)$.

Given a random matrix $X$ with density $f(X)$, and a one-to-one transformation $X = g(Y)$, the density of $Y$ is $f(g(Y))J(Y)$, where $J(Y)$ is the Jacobian of the transformation. Consequently, the primary computation is that of the Jacobian. Subsequent marginal distributions require integrating out extraneous variables.

We will have occasion to use the fact that if $G$ is orthogonal, then $G'(dG)$ is skew-symmetric. This is a consequence of $d(G'G) = (dG')G + G'(dG) = 0$.

A simplification in the computation of Jacobians is the fact that the Jacobian $J(X \to Y) = J((dX) \to (dY))$, and that the transformation in differentials is linear.

The fact that random matrices need to be treated with greater respect than for random scalars is exhibited in the following.
A. Distribution of the Inverse: $Y = X^{-1}$

Although the transformation $Y = X^{-1}$ has the same appearance whether $X$ is an arbitrary, symmetric, triangular, or skew-symmetric matrix, each case generates a transformation in different dimensions. The following is a summary of the Jacobians:

(A.1a) $X$ arbitrary, \quad $J = (\det X)^{-2p},$

(A.1b) $X$ symmetric, \quad $J = (\det X)^{-(p+1)},$

(A.1c) $X$ skew-symmetric, \quad $J = (\det X)^{-(p-1)},$

(A.1d) $X$ triangular, \quad $J = (\det X)^{-(p+1)}.$

In particular, if $S$ has a Wishart distribution arising from a sample of size $n$ from a $p$-variate normal distribution with identity covariance matrix, then the density of $V = S^{-1}$ is

$$f(V) = c (\det V)^{-(n+p+1)/2} \exp[-\frac{1}{2} \text{tr} V^{-1}], \quad V > 0,$$

where $c$ is the normalizing constant in the Wishart density.

For a discussion of Jacobians see Deemer and Olkin (1951), Olkin (1953), Mathai (1997).

One might question whether there are other special or patterned matrices of interest in this context.

B. Moore-Penrose Inverse: $X^+ = (XX')^{-1}X$

If $X$ is a rectangular $p \times n$ random matrix, then it does not have an inverse when $n > p$. If $f(X)$ is the density of $X$, then the density of the Moore-Penrose inverse

(B.1) $Y = X^+ = (XX')^{-1}X$

was obtained by Y. Zhang (1985) to be:

(B.2) $(\det YY')^{-n}f(Y^+).$

Alternative proofs were given by Neudecker and Liu (1996) and by Olkin (1998). The essential step in Zhang's proof is to write $X = TG$, where $T$ is a $p \times p$ lower triangular matrix and $G$ is a suborthogonal matrix, from which

(B.3) $Y = (TT')^{-1}TG = T'^{-1}G.$
Using the chain rule, the Jacobian is

(B.4) \[ J(X \rightarrow Y) = J(X \rightarrow T, G) J(T, G \rightarrow T^{-1}, G) J(T^{-1}, G \rightarrow X). \]

Each term in (B.4) is known (e.g., Olkin 1951, 1953) and the result follows from the fact that

(B.5) \[ Y^+ = [(Y^+ Y'^+)^{-1}]Y^+ = [XX'](XX')^{-1}X = X. \]

Of course, when \( p = n \), \( Y^+ = Y^{-1} = X \).

Neudecker and Liu (1996) write out equation (B.1) as a vector, \( \tilde{Y} = (y_{11}, y_{12}, \ldots, y_{pn}) \), from which the Jacobian is obtained. This method is rather cumbersome because the transformation (B.1) is not readily adapted to a vector format.

We describe an alternative method to obtain the Jacobian (Olkin, 1998). By using a sequence of transformations we generate a functional equation that yields the Jacobian. This method was introduced by Olkin and Sampson (1972), and may be illuminating in other contexts.

The density of \( Y \) is \( f(Y^+) J(Y) \), where \( J(Y) \) is a function that needs to be evaluated. To generate a functional equation we let \( U = AX \), for \( A \) nonsingular. The Jacobian is \( (\det A)^{-n} \), so that the density of \( U \) is

(B.6) \[ f(A^{-1}U)(\det A)^{-n}. \]

The connection between \( V = U^+ \) and \( Y \) is \( V = A^{-1}Y \); its density is

(B.7) \[ f(A^{-1}V^+)(\det A)^{-n} J(V) = f(A^{-1}AY^+)(\det A)^{-n} J(V)(\det A')^{-n} = f(Y^+)(\det AA')^{-n} J(A'^{-1}Y), \]

which yields the functional equation

(B.8) \[ J(Y) = (\det AA')^{-n} J(A'^{-1}Y). \]

The choice \( Y = (A, 0) \) then gives

(B.9) \[ J(Y) = \det(YY')^{-n} J(I, 0). \]
Thus, we have evaluated $J(Y)$, except for a constant $J(I,0)$, which can be obtained by letting $f(X)$ have a specific form, as for example, the normal density.

C. The Singular Value Decomposition: $X = GD_2H$

Here $X$ is a $p \times n$ random matrix, $G$ is orthogonal, $H$ is a suborthogonal $p \times n$, $D_s = \text{diag}(s_1, \ldots, s_p)$ and the singular values, $s_i = \lambda_i^{1/2}(XX')$, are the positive square roots of the eigenvalues of $XX'$. The primary concern is to obtain the joint distribution of the singular values.

If the density $f$ of $X$ is a function of the eigenvalues of $XX'$, then the joint density of the ordered singular values $s_1 \geq \cdots \geq s_p \geq 0$ is

\begin{equation}
(c \prod_{i \leq j}(s_i^2 - s_j^2))g(s_1, \ldots, s_p),
\end{equation}

where $g(s_1, \ldots, s_p) = f(D_s^2)$. The constant can be evaluated by letting the elements of $X$ be standard normal random variables.

That the Jacobian is a function of $s_1, \ldots, s_p$ can be seen from the functional equation method of Section B. Let $X = G_1D_sH_1$ and $Y = G_2XH_2 = G_2G_1D_sH_1H_2 \equiv GD_sH$, where $G, H$ are orthogonal. The Jacobian $J(X \to G_1, s, H_1)$ is a function $Q(G_1, s, H_1)$. Using the chain rule,

\[ J(Y \to G, s, H) = J(Y \to X)J(X \to G_1, s, H_1). \]

But $J(Y \to X) = 1$, so we obtain

\[ Q(G_2G_1, s, H_1H_2) = Q(G_1, s, H_1) \quad \forall \text{ orthogonal } G_i, H_i. \]

The choice $G_2 = G'_1, H_2 = H'_1$ shows that $Q(G, s, H)$ is a function of $s$ alone.

To obtain the Jacobian, take differentials in $X = GD_sH$, premultiply by $G'$ and postmultiply by $H'$ to obtain

\[ G'(dX)H' = G'(dG)D_s + D_ds + D_s(dH)H'. \]

Let $Y = G'(dX)H'$, $G'(dG) = U$, $(dH)H' = V$, where $U$ and $V$ are skew-symmetric. Using the chain-rule,

\[ J((dX) \to (dG), (dS), (dH)) \]

\[ = J((dX) \to Y)J(Y \to U, (ds), V)J(U \to (dG)J(V \to (dH)) \]

\[ \equiv J_1J_2J_3J_4. \]
Here $J_1 = 1$ and $J_2$ is obtained from the equation

$$Y = UD_s + D_s V,$$

namely,

$$y_{ii} = ds_i,$$
$$y_{ij} = s_i u_{ij} + s_j v_{ij}, \quad i < j,$$
$$y_{ij} = -s_j u_{ij} - s_i v_{ij}, \quad i > j.$$

For each $(i,j)$ pair the absolute value of the $2 \times 2$ determinant of derivatives:

$$\begin{vmatrix}
  u_{ij} & v_{ij} \\
  (i < j) & y_{ij} \\
  (i > j) & y_{ij}
\end{vmatrix}
\begin{array}{cc}
  s_j & s_i \\
  -s_i & -s_j
\end{array}$$

is $|s_i^2 - s_j^2|$. Consequently, for ordered singular values, we obtain $\Pi_{i \leq j} (s_i^2 - s_j^2)$. The density is a function of $\lambda(XX') = \lambda(GD_s^2G') = \lambda(D_s^2)$, that is $f(\lambda(XX')) = g(s_1, \ldots, s_p)$, a function of the singular values.

This derivation was obtained by Olkin (1951). The density (C.1) provides an alternative derivation of the joint density of the characteristic roots $\ell_i = s_i^2$, $i = 1, \ldots, p$, of a Wishart matrix.

**D. Triangular Factorization: $X = GTG'$**

Here $T$ is lower triangular and $G$ is orthogonal; all matrices are $p \times p$. The Jacobian of this transformation is independent of $G$ as follows. Let $Y = HXH' = HG'TG'H' \equiv KTK'$, where $H$ is orthogonal, in which case $K$ is orthogonal. Then

$$J(Y \to X)J(X \to G, T)J(G, T \to K, T) = J(Y \to K, T).$$

The term $J(Y \to X) = (\det H)^p(\det H')^p = (\det HH')^p = 1$, and $J(G, T \to K, T) = 1$. Let $J(X \to G, T) = F(G, T)$ to obtain the functional equation $F(G, T) = F(HG, T)$ for all orthogonal $G, H$. The choice $H = G'$ yields $F(G, T) = F(I, T)$ independent of $G$.

To obtain the Jacobian using differentials, we have

$$G'(dX)G = G'(dG)T + (dT) + T(dG')G.$$
Let \( Y = G'(dX)G, \ S = G'(dG), \) then by the chain rule

\[
(D.1) \quad J((dX) \rightarrow (dG), (dT)) = J((dX) \rightarrow Y)J(Y \rightarrow S, (dT))J(S, (dT) \rightarrow (dG)(dT)) \\
\equiv J_1 J_2 J_3.
\]

The terms \( J_1 \) and \( J_3 \) need not be evaluated. The equation from which \( J_2 \) is evaluated is

\[
(D.2) \quad Y = ST + (dT) - TS,
\]

where \( S \) is skew-symmetric.

The matrix of derivatives is

\[
\begin{pmatrix}
  (dt_{ii}) & (dt_{ij}), i > j & s_{ij} \\
  y_{ii} & I & A_{12} & A_{13} \\
  y_{ij}, i > j & 0 & I & A_{23} \\
  y_{ij}, i < j & 0 & 0 & A_{33}
\end{pmatrix}
\]

The determinant only depends on \( A_{33} \). An examination of equation (D.2) shows that \( ds_{ij}/dy_{ij} = t_{ii} + t_{jj} \). Consequently, the Jacobian of the transformation \( X = GTG' \) is \( \Pi_{i<j}|t_{ii} + t_{jj}|. \)

E. The LU Decomposition: \( X = LU \) (asymmetric), \( X = LDU \) (symmetric)

As noted by Yeung and Chan (1997): “The numerical instability of Gaussian elimination is proportional to the size of the \( L \) and \( U \) factors that it produces.” In their paper, starting with independent standard normal elements \( x_{ij} \), they obtain the density of a single element \( \ell_{ij} \) and \( u_{ij} \) in the asymmetric version. In particular, they show that each element \( \ell_{pq}, p > q, \) has a Cauchy distribution, namely,

\[
f(\ell_{pq}) = \frac{1}{\pi} \frac{1}{1 + \ell_{pq}^2}, \quad -\infty < \ell_{pq} < \infty.
\]

The factorization \( X = LDU \) is unique when \( X \) is nonsingular. Suppose \( L_1 D_1 U_1 = L_2 D_2 U_2, \) then \( L_2^{-1} L_1 D_1 = D_2 U_2 U_1^{-1}. \) But \( L \equiv L_2^{-1} L_1 \) is lower triangular and \( U \equiv U_2 U_1^{-1} \) is upper triangular with \( \ell_{ii} = u_{ii} = 1, \) so that \( D_1 = D_2 \equiv D. \) Then \( LD = DU \) implies that \( L = U = I, \) from which \( L_1 = L_2, U_1 = U_2. \)
E.1 Symmetric Case: $X = LD_\theta U$

To obtain the Jacobian we first note that a functional equation argument shows that the Jacobian is a function of the $\theta_i$ alone.

To see this, let $X = LD_\theta U$, $Y = L_1 X U_1$, where $L_1$ is lower triangular and $U_1$ is upper triangular. Then $Y = L_1 LD_\theta U U_1 = \overline{L} D_\theta \overline{U}$, and

$$J(Y \to \overline{L}, \theta, \overline{U}) = J(Y \to X) J(X \to L, \theta, U) J(L, \theta, U \to \overline{L}, \theta, \overline{U}).$$

Then, with $J(X \to L, \theta, U) = h(L, \theta, U)$ we obtain the functional equation

$$h(L_1 L, \theta, U U_1) = |L|^p |U|^p h(L, \theta, U) \prod_{i=1}^p \ell_{ii} \prod_{i=1}^p u_{ii}$$

$$= h(L, \theta, U) \quad \text{for all } L_1, L, \theta, U, U_1.$$

The choice $L_1 = L^{-1}, U_1 = U^{-1}$ yields

$$h(I, \theta, I) = h(L, \theta, U),$$

so that the Jacobian is a function of the $\theta_i$ alone.

The Jacobian is obtained from the matrix equation

$$Y = Q D_\theta + D_{d\theta} + D_\theta R,$$

where $Q$ is lower triangular, $q_{ii} = 0$, $R$ is upper triangular, $r_{ii} = 0$, $i = 1, \ldots, p$. Partition $Q$ and $R$ as

$$Q = \begin{pmatrix} 0 & 0 \\ Q_1 & 0 \end{pmatrix}, \quad Q_1 : p-1 \times p-1, \quad R = \begin{pmatrix} 0 & R_1 \\ 0 & 0 \end{pmatrix}, \quad R_1 : p-1 \times p-1,$$

where $Q_1$ is lower triangular and $R_1$ is upper triangular. Let $Q_1 = Q_1 \cdot D_\theta$, $D_\theta = \text{diag}(\theta_1, \ldots, \theta_{p-1})$ in which case the Jacobian is $\prod_{i=1}^{p-1} |\theta_i|^{(p-1)-i+1} = \prod_{i=1}^{p-1} |\theta_i|^{p-1}$. Similarly let $R_1 = R_1 \cdot D_\theta = \overline{R}_1$, in which case the Jacobian is $\prod_{i=1}^{p-1} \theta_1^{p-i}$. Hence the Jacobian is $\prod_{i=1}^{p-1} |\theta_i|^{2(p-i)}$.

When the $x_{ij}$ $(i, j = 1, \ldots, p)$ are independent, each having a standard normal density, we obtain the joint density of $L, \theta, U$ to be

(E.1.1) \[ c \prod_{i=1}^{p-1} |\theta_i|^{2(p-i)} \exp - \frac{1}{2} tr LD_\theta U' D_\theta L', \]

7
\[ \ell_{ii} = u_{ii} = 1, \quad -\infty < \ell_{ij}, u_{ij}, \theta_i < \infty. \]

**E.2 Asymmetric Case: \( X = LU \)**

To obtain the Jacobian of the transformation \( X = LU \), take differentials:

\[ (dX) = (dL)U + L(dU), \]

from which

\[ (E.2.1) \quad L^{-1}(dX)U^{-1} = L^{-1}(dL) + (dU)U^{-1}. \]

Let

\[ Y = L^{-1}(dX)U^{-1}, \quad \tilde{Q} = L^{-1}(dL) = \begin{pmatrix} 1 & 0 \\ Q & I \end{pmatrix}, \]

\[ R = (dU)U^{-1} \]

Then (K.1) becomes

\[ Y = \tilde{Q} + R, \]

so that \( y_{ii} = 1 + r_{ii}, \quad i = 1, \ldots, p, \quad y_{ij} = q_{ij}(i > j), \quad y_{ij} = r_{ij}(i < j). \) By the chain rule,

\[ J((dX) \rightarrow (dL), (dU)) = J((dX) \rightarrow Y) J(Y \rightarrow Q, R) J(Q, R \rightarrow (dL), (dU)) \equiv J_1 J_2 J_3. \]

The evaluation of each Jacobian is known:

\[ J_1 = (\det L)^p (\det U)^p = \Pi_i^p |u_{ii}|^p \]

\[ J_2 = 1 \]

\[ J_3 = \Pi_i^p (\ell_{ii})^{p-i+1} \Pi(u_{ii})^i = \Pi_i^p \ell_{ii}^{(p-i+1)} \Pi |u_{ii}|^{-i} = \Pi |u_{ii}|^{-i}. \]

Consequently, the Jacobian of the transformation \( X = LU \) is \( \Pi_i^p |u_{ii}|^{p-i} \), and hence the joint distribution of \((L, U)\) is

\[ (E.2.2) \quad f(L, U) = c \Pi_i^p |u_{ii}|^{p-i} e^{-\frac{1}{2} \text{tr} L U U^t L}, \]

where \( c = (2\Pi)^{-p^2/2} \).

As noted, Yeung and Chan (1997) show that the distribution of a single element \( \ell_{ij} \) is that of a ratio of normal variates, and thereby has a Cauchy distribution. The distribution
of more than one element will therefore be a multivariate Cauchy distribution with Cauchy marginals.

F. The Wishart Density: \( S = XX' \)

Let \( X \) be a \( p \times n \) random matrix of independent standard normal variates, then the density of \( S = XX' \) is

\[
(F.1) \quad \frac{c}{(\det S)^{\frac{n-p+1}{2}}} \exp[-\frac{1}{2} \text{tr } S].
\]

The density (F.1) was first obtained by Fisher (1915) in the bivariate case \( p = 2 \). Using a geometrical argument, Wishart (1928) obtained the general result (F.1). The Wishart distribution is tantalizing in that it offers opportunities for novel derivations. In part, this is because we start with \( pn \) random variables \( x_{ij} \), and end with \( p(p+1)/2 \) random variables \( s_{ij} \). Thus, there are alternative routes to obtain the distribution on a subspace.

In total there are probably close to twenty derivations of the Wishart distribution or extensions and modifications. There are geometrical arguments, proofs by induction, coordinate-free derivations, alternative derivations using coordinates, derivations using characteristic functions, and so on. Many derivations or extensions appeared prior to 1950. References can be uncovered from Anderson (1984) and Johnson and Kotz (1972). The non-central Wishart distribution and the related distribution of the roots of a determinantal equation have generated many papers and books.

We mention only one extension of (F.1), by Hsu (1940), whereby if \( X \) is a \( p \times n \) random matrix with density \( f(XX') \), then the density of \( S = XX' \) is

\[
(F.2) \quad \frac{c}{(\det S)^{\frac{n-p+1}{2}}} f(S).
\]

We give a proof using functional equations. Let \( S = XX' \), then the joint density of the elements of \( S \) is \( f(S)h(S) \), for some function \( h(S) \). Let \( S = AVA' \), \( A \) nonsingular. The density of \( V \) is

\[
(F.3) \quad f(AVA')h(AVA')|A|^{p+1}.
\]

Alternatively, let \( X = AY \), then the density of \( Y \) is \( f(AYY'A')|A|^n \). Now let \( V = YY' \), from which the density of \( V \) is

\[
(F.4) \quad f(AVA')|A|^n h(V).
\]
Equating (F.3) and (F.4) yields
\[ h(AVA') = h(V)|A|^{n-p-1} = h(V)|AA'|^{(n-p-1)/2}. \]

With \( V = I \) and \( AA' = Z \),
\[ h(Z) = c|Z|^{(n-p-1)/2}, \]
which yields (F.2).

The result (F.1) is the case that \( f \) is the normal density. For further discussion see Anderson (1984).

**G. Distribution of the Square of a Symmetric Matrix: \( V = S^2 \)**

This transformation shows that a seemingly simple problem is not so. We now show that the Jacobian is \( 2^p \prod_{i=1}^p \ell_i \prod_{i,j=1}^p (\ell_i + \ell_j) \), where \( \ell_i = \lambda_i(S) \), \( i = 1, \ldots, p \) are the eigenvalues of \( X \). Here \( X \) is an arbitrary \( p \times p \) matrix.

Although we obtain the Jacobian of the transformation, some care is required in obtaining the density of \( V \) because the transformation is not one-to-one. This problem would not occur if \( S \) were symmetric positive definite because we can define a unique symmetric positive definite square root.

**Proof.** The Jacobian of the transformation is that of the transformation in the differentials:
\[ (dV) = (dS) S + S (dS). \]

Let \( S = GD_\ell G' \), where \( D_\ell = \text{diag} (\ell_1, \ldots, \ell_p) \). Then
\[ G' (dV) G = G' (dS) GD_\ell + D_\ell G' (dS) G. \]

Let \( Z = G' (dV) G, \ W = G' (dS) G \), so that
\[ Z = WD_\ell + D_\ell W. \]

By the chain rule,
\[ J(dV \to dS) = J(dV \to Z)J(Z \to W)J(W \to dS) \equiv J_1 J_2 J_3. \]
It is classical (Olkin, 1951) that \( J_1 = J_3 = (\det G)^{p+1} \); \( J_2 \) is obtained from the equation (G.3): \( z_{ii} = 2\ell_i w_{ii}, \ z_{ij} = \ell_i + \ell_j, \ i, j = 1, \ldots, p. \)

Consequently, the Jacobian is \( \Pi \ell_i \Pi_{i<j}(\ell_i+\ell_j) \). The term \( \Pi \ell_i = \det S \), but \( \Pi_{i<j}(\ell_i+\ell_j) \) is not readily translated to matrix functions. It can be expressed as a function of power sums, which in turn is expressible as the trace of powers.

H. The Eigenvalue Decomposition

Given an orthogonally invariant density \( f(S) \) over positive definite \( S \), that is, \( f(GSG') = f(S) \) for all orthogonal matrices \( G \), the joint density of the ordered eigenvalues \( \ell_1 \geq \ldots \geq \ell_p \geq 0 \) of \( S \) is

\[
(H.1) \quad c \prod_{i<j}(\ell_i - \ell_j)g(\ell_1, \ldots, \ell_p),
\]

where \( g(\ell_1, \ldots, \ell_p) = f(D_{\ell}), \) and \( D_{\ell} = \text{diag}(\ell_1, \ldots, \ell_p). \) The constant may be evaluated by letting \( f(s) \) be a Wishart density.

The singular value decomposition and the eigenvalue decomposition are equivalent in that \( s_i = \ell_i^{1/2}, \ i = 1, \ldots, p. \)

Consequently, we could have started with a random \( p \times n \) matrix \( X \) whose density is orthogonally invariant, that is, \( f(X) = f(GXH) \) for all \( p \times p \) orthogonal \( G \), and all \( n \times n \) orthogonal \( H \). Use of the singular value decomposition \( X = \Gamma(D_s, 0)(\Delta_1^{1/2}) = \Gamma(D_{\ell}^{1/2}, 0)(\Delta_2^{1/2}) \), where \( \Gamma \) and \( \Delta = (\Delta_1^{1/2}) \) are orthogonal, yields the joint density of \( \Gamma \) and \( \Delta_1 \). Integration over \( \Gamma' \) and \( \Delta \) yields (G.1). This derivation was obtained by Olkin (1951).

I. A Duality

It was shown by Parker (1945) and Vinogradove (1950) that if \( A : p \times n, \ B : p \times m, \ m \leq n, \) then \( AA' = BB' \) if and only if \( A = BG \), where \( G : m \times n \) is suborthogonal. Consequently, we obtain the equivalence of the singular value decomposition \( X = GD_s H \) and the eigenvalue decomposition \( S = XX' = GD_{\ell}G' \).

We also obtain an equivalence for rectangular coordinates \( X = TG \), where \( T \) is lower triangular and \( G \) is suborthogonal, and the factorization \( S = XX' = TT' \).

For more details on this duality see Horn and Olkin (1996).
J. Rectangular Coordinates: (i) \( X = TG \), (ii) \( S = TT' \)

The factorization (i) is due to Schmidt (1907). The rectangular coordinates were introduced by Mahalanobis, Bose and Roy (1937) using a geometric construction. If \( X \) is a random matrix with density a function of \( XX' \), then the joint density of the rectangular coordinates is

\[
(J.1) \quad c \prod t_{ii}^{n-i} f(TT'), \quad t_{ii} > 0, \quad -\infty < t_{ij} < \infty.
\]

The constant can be obtained by letting \( f \) denote the normal density.

By letting \( S = TT' \) be a transformation from \( T \) to \( S \), we obtain a derivation of the distribution of \( S = TT' = XX' \). Here the Jacobian is \( 2^{-p} \prod t_{ii}^{-(p-i+1)} \). Consequently, we obtain the density

\[
(J.2) \quad c_1 \prod t_{ii}^{(n-i)-(p-i+1)} f(S) = c_1 |S|^{n-p-1} f(S).
\]

The choice \( f(S) = \exp(-\frac{1}{2} \text{tr} S) \) yields the Wishart density. This method was used by James (1954) and by Olkin and Roy (1954).

K. Simultaneous Decompositions

If \( X \) and \( Y \) are independent \( p \times m \) and \( p \times n \) random matrices, with densities \( f \) and \( g \) that are functions of \( XX' \) and \( YY' \), respectively, then the joint density of the roots of the determinantal equation \( \det(XX' - \ell^2 YY') = 0 \) is obtained as follows. Let

\[
(K.1) \quad X = WD_q G, \\
Y = WH.
\]

where \( D_q = \text{diag}(q_1, \ldots, q_p) \), \( G : p \times n \) and \( H : p \times m \) are suborthogonal, and \( q_i^2 = \ell_i \), \( i = 1, \ldots, p \).

The Jacobian of the transformation \( (K.1) \) is \( \left( \det W \right)^{p+2} \prod_{i<j} (q_i - q_j), \; q_1 \geq \cdots \geq q_p > 0 \), so that the joint density of \( W, q, G, H \) is

\[
(K.2) \quad c(\det W)^{p+2} f(WD_q W') g(WW') \prod_{i<j} (q_i - q_j).
\]

Integration over \( W \) and the orthogonal group yields the joint density of the ordered roots. When \( f \) and \( g \) are the normal densities, then \( (K.2) \) becomes

\[
(K.3) \quad c \prod_{i<j} (q_i - q_j) \prod (1 + q_i^2)^{-(p+1)}.
\]
The derivation was obtained by Olkin (1951).

A more common derivation, especially in the case of underlying normal densities, is to start with symmetric positive definite matrices $U$ and $V$. The roots of $\det(U - \ell V) = 0$ can be obtained from the factorization

(K.4) \[
U = WD\ell W', \\
V = WW',
\]

where $D\ell = \text{diag}(\ell_1, \ldots, \ell_p)$, $\ell_1 \geq \cdots \geq \ell_p > 0$. The Jacobian of this transformation is $(\det W)^{p+2p} \prod_{i<j} (\ell_i - \ell_j)$. (See Deemer and Olkin, 1951; Mathai, 1999.) Consequently, the joint density of $W$ and $\ell$ is

(K.5) \[
cf (WDW')g(WW')(\det W)^{p+2} \prod_{i<j} (\ell_i - \ell_j).
\]

When $f$ and $g$ are Wishart densities, we obtain the joint density

\[
c \prod_{i<j} (\ell_i - \ell_j) \prod_{i} \ell_i^{-\frac{m-p-1}{2}} \prod (1 + \ell_i)^{-\frac{(n+m)}{2}}, \quad \ell_1 \geq \cdots \geq \ell_p > 0.
\]

For more details see Anderson (1984), Muirhead (1982).

REFERENCES


