ELSEVIER

Statistics & Probability Letters 38 (1998) 131–135

STATISTICS & PROBABILITY LETTERS

The density of the inverse and pseudo-inverse of a random matrix

Ingram Olkin

Department of Statistics, Stanford University, CA 94305, USA

Received 1 June 1997; received in revised form 1 July 1997

Abstract

Given an absolutely continuous density of a random matrix X, we study the density of the inverse when X is a $p \times p$ symmetric, triangular and arbitrary matrix, and the pseudo-inverse when X is rectangular. In the latter case we provide alternative proofs to that of Zhang (1985), who first obtained this density. © 1998 Elsevier Science B.V. All rights reserved

Suppose that X is a $p \times n$ $(p \leqslant n)$ random matrix with a density that is a function of XX', say $f(XX')$. Then the density of $S = XX'$ is

$$\frac{\pi^{pn}}{\Gamma_p(n/2)} |S|^{(n-p-1)/2} f(S), \tag{1}$$

where $\Gamma_p(t) = \prod_1^p \Gamma(t-(i-1)/2)/\pi^{p(p-1)/4}$.

The essential idea of this result is due to Hsu (1939) who obtained a similar general result for the joint distribution of the roots of a determinantal equation. This counterpart result is stated in Anderson (1984, p. 533). An alternative proof of (1) is based on the decomposition

$$X = TG, \tag{2}$$

where T is a $p \times p$ lower triangular matrix and G is a $p \times n$ suborthogonal matrix, so that $XX' = TT'$. The Jacobian from $X \to T, G$ is

$$J(X \to T, G) = 2^{p(p-1)/2} \prod_1^p t_{ii}^{n-i} h(G), \tag{3}$$

where $h(G)$ is a function of G alone (Olkin and Roy, 1954). At this point we have the joint distribution of the elements of T to be

$$c \prod_1^p t_{ii}^{n-i} f(TT'), \tag{4}$$

0167-7152/98/$19.00
PII S0167-7152(97)00163-6

where the constant can be determined by the special case in which f is chosen to be the normal density function. The elements of T are called rectangular coordinates by Mahalanobis et al. (1937), who obtained the distribution (4) using a geometrical argument (see also Hsu, 1940).

The transformation $TT' = S$ has the Jacobian

$$J(T \to S) = \left(2^p \prod_1^p t_{ii}^{p-i+1} \right)^{-1}. \tag{5}$$

Consequently, the density of S is $cf(S)(\prod t_{ii}^{n-i})(\prod t_{ii}^{-(p-i+1)})$, which yields (1) except for the constant. The constant can be obtained from (4), or alternatively, letting f denote the Wishart density.

For any random symmetric matrix S with absolutely continuous density $f(S)$, the density of $V = S^{-1}$ is

$$|V|^{-p-1} f(V^{-1}). \tag{6}$$

The factor $|V|^{-p-1}$ arises from the Jacobian which is obtained from the fact that $-V^{-1}(\mathrm{d}V)V^{-1} = (\mathrm{d}S)$; see Olkin (1951, Theorem 3.7).

When T is a random $p \times p$ lower triangular matrix with absolutely continuous density $f(T)$, the density of $U = T^{-1}$ is

$$|U|^{-p-1} f(U^{-1}). \tag{7}$$

The Jacobian of a transformation $Y = AZB$ where Y, A, Z, B are lower triangular is $\prod a_{ii}^i b_{ii}^{p-i+1}$. When $A = B = U^{-1}$ we obtain $|U|^{-p-1}$. (See Deemer and Olkin, 1951 and Olkin, 1953.)

When W is a random $p \times p$ matrix with absolutely continuous density $f(W)$, the density of $Z = W^{-1}$ is

$$|Z|^{-2p} f(Z^{-1}), \tag{8}$$

where the Jacobian is obtained from the fact that $-Z^{-1}(\mathrm{d}Z)Z^{-1} = (\mathrm{d}W)$. (See Deemer and Olkin, 1951, Theorem 3.6.)

When X is a random $p \times n$ $(p \leqslant n)$ rectangular matrix with absolutely continuous density $f(X)$, the density of the Moore–Penrose inverse

$$Y = (XX')^{-1}X \equiv X^+ \tag{9}$$

was obtained by Zhang (1985) to be

$$|YY'|^{-n} f(Y^+). \tag{10}$$

Note that $Y^+ = (YY')^{-1}Y = (XX')Y = X$. To obtain the Jacobian, Zhang uses the decomposition (2), from which

$$Y = (TT')^{-1}TG = T'^{-1}G. \tag{11}$$

Using the chain rule, the Jacobian of the transformation $X \to X^+$ is

$$J(X \to X^+) = J(X \to T, G)\, J(T, G \to T^{-1}, G)\, J(T^{-1}G \to X^+). \tag{12}$$

Each term in (12) is known, from which the result follows.

An alternative proof by Neudecker and Liu (1996) is based on a brute force derivation of the Jacobian $J(X \to X^+)$ from (9). This derivation is exceedingly cumbersome and lengthy, and requires considerable matrix manipulation. The rationale for this proof is the claim that the Zhang proof requires additional statistical reasoning as was used in the derivation of (1) or (4).

An alternative proof is based on a functional equation argument. (See Olkin and Sampson (1972) for other examples of this approach.)

The density of $Y = X^+ = (XX')^{-1}X$ is $f(Y^+)J(Y)$, where $J(Y)$ is a function that needs to be evaluated. Let $U = AX$, for A nonsingular, in which case its density is $f(A^{-1}U)|A|^{-n}$. Then let $V = U^+$ with density $f(A^{-1}V^+)|A|^{-n}J(V)$. Finally, let $V = A'^{-1}Y$, with density

$$f(A^{-1}AY^+)|A|^{-n}J(V)|A'|^{-n} = f(Y^+)|AA'|^{-n}J(A'^{-1}Y).$$

This yields the functional equation

$$J(Y) = |AA'|^{-n}J(A'^{-1}Y)$$

which holds for all nonsingular $p \times p$ matrices A and all $p \times n$ matrices Y. With $Y = (A, 0)$, the functional equation becomes

$$J(Y) = |YY'|^{-n}J(I, 0),$$

where $J(I, 0)$ is a constant, which can be evaluated from the constants in a specific density.

Although the use of transformation (2) is both valid and direct, it is not the most natural transformation. Rather, the singular value decomposition might be a first choice. Let

$$X = GD_\theta H, \tag{13}$$

where G is a $p \times p$ orthogonal matrix, $D_\theta = \text{diag}(\theta_1, \dots, \theta_p)$, $\theta_1 > \theta_2 > \cdots > \theta_p > 0$ are the ordered singular values, and H is a $p \times n$ $(p \leqslant n)$ suborthogonal matrix. It follows that

$$X^+ = (GD_\theta^2 G')^{-1}GD_\theta H = GD_\theta^{-1}H \equiv GD_v H, \tag{14}$$

where $v_i = 1/\theta_i$, $i = 1, \dots, p$. The Jacobian is evaluated using the chain rule

$$\begin{aligned} J(X \to X^+) &= J(X \to G, \theta, H)\, J(G, \theta, H \to G, v, H)\, J(G, v, H \to X^+) \\ &\equiv J_1 J_2 J_3. \end{aligned} \tag{15}$$

Then,

$$J_1 = h_1(G)h_2(H)\prod_1^p \theta_i^{n-p}\prod_{i<j}|\theta_i^2 - \theta_j^2|, \tag{16a}$$

$$J_2 = \prod_1^p v_i^{-2}, \tag{16b}$$

$$J_3 = \left[h_1(G)h_2(H)\prod v_i^{n-p}\prod_{i<j}^p |v_i^2 - v_j^2|\right]^{-1}, \tag{16c}$$

the product of which is $\prod \theta_i^{2n} = |XX'|^n = |YY'|^{-n}$.

The Jacobian J_2 is immediate. J_1 and J_3 arise from the same transformation; it was obtained by Olkin (1951), but is unpublished. We now provide some of the essential details.

To obtain the Jacobian of the transformation (13), rewrite (13) as

$$X = GD_\theta(I, 0)\begin{pmatrix} H \\ K \end{pmatrix} \equiv GD_\theta(I, 0)L, \tag{17}$$

where K is an $(n-p)\times n$ suborthogonal matrix such that $LL'=I_n$. Taking differentials in (17), premultiplying by G', and postmultiplying by L' yields

$$G'(\mathrm{d}X)L' = G'(\mathrm{d}G)(D_\theta,0)+(\mathrm{d}D_\theta,0)+(D_\theta,0)(\mathrm{d}L)L'. \tag{18}$$

Note that for any orthogonal matrix Γ, $\Gamma'(\mathrm{d}\Gamma)+(\mathrm{d}\Gamma)'=0$, so that $S=\Gamma'(\mathrm{d}\Gamma)$ is skew symmetric.

Let $Y=G'(\mathrm{d}X)L'$, $R=G'(\mathrm{d}G)$, $S=(\mathrm{d}L)L'$, so that

$$Y=R(D_\theta,0)+(\mathrm{d}D_\theta,0)+(D_\theta,0)S \tag{19}$$

and employ the chain rule

$$\begin{aligned} J(X &\to G,\theta,H)=J(\mathrm{d}X\to \mathrm{d}G,\,\mathrm{d}\theta,\,\mathrm{d}H)\\ &= J(\mathrm{d}X\to Y)J(Y\to R,\,\mathrm{d}\theta,\,S)J(R,\,\mathrm{d}\theta,\,S\to \mathrm{d}G,\,\mathrm{d}\theta,\,\mathrm{d}H)\\ &\equiv J_1J_2J_3. \end{aligned}$$

The Jacobian $J_1=|G|^n|L|^p=1$ (in absolute value), and J_3 is a function of G and H or of R and S, which we need not evaluate, because these terms drop out as in (16a) and (16c). Consequently, we need only evaluate J_2. Note that where $Y\equiv(Y_1,Y_2)=(RD_\theta,0)+(\mathrm{d}D_\theta,0)+(D_\theta S_1,D_\theta S_2)$, S_1 and S_2 are $p\times p$ and $p\times(n-p)$ submatrices of S. These equations are linear and we obtain as derivatives of the matrix

$$\begin{array}{l c} & \begin{array}{cccc} \mathrm{d}\theta & R & S_1 & S_2 \end{array} \\ \begin{array}{l} y_{ii} \\ y_{ij}\ (i<j\leqslant p) \\ y_{ij}\ (j<i\leqslant p) \\ y_{ij}\ (i\leqslant p,\, j>p) \end{array} & \begin{bmatrix} I_p & 0 & 0 & 0 \\ 0 & D_1 & D_2 & 0 \\ 0 & -D_2 & -D_1 & 0 \\ 0 & 0 & 0 & D_3 \end{bmatrix} \end{array}$$

where

$$D_1=\mathrm{diag}(\underbrace{\theta_2,\theta_3,\ldots,\theta_p}_{p-1},\underbrace{\theta_3,\ldots,\theta_p}_{p-2},\ldots,\underbrace{\theta_{p-1}\theta_p}_{2}\,\underbrace{\theta_p}_{1}),$$

$$D_2=\mathrm{diag}(\underbrace{\theta_1,\ldots,\theta_1}_{p-1},\underbrace{\theta_2,\ldots,\theta_2}_{p-2},\ldots,\underbrace{\theta_{p-1}\theta_{p-1}}_{2}\,\underbrace{\theta_p}_{1}),$$

$$D_3=\mathrm{diag}(\underbrace{\theta_1,\ldots,\theta_1}_{n-p},\ldots,\underbrace{\theta_p,\ldots,\theta_p}_{n-p}).$$

Consequently, the determinant is equal to (in absolute value)

$$|D_3|\begin{vmatrix} D_1 & D_2 \\ -D_2 & -D_1 \end{vmatrix} = \prod \theta_i^{n-p}|D_1^2-D_2^2| = \prod \theta_i^{n-p}\prod_{i<j}|\theta_i^2-\theta_j^2|$$

which completes the derivation.

References

Anderson, T.W., 1984. An Introduction to Multivariate Statistical Analysis, 2nd ed. Wiley, New York.

Deemer, W.L., Olkin, I., 1951. The Jacobians of certain matrix transformations useful in multivariate analysis. Biometrika 38, 345–367.

Hsu, P.L., 1939. On the distribution of roots of certain determinantal equations. Ann. Eugenics 9, 250–258.

Hsu, P.L., 1940. An algebraic derivation of the distribution of rectangular coordinates. Proc. Edinburgh Math. Soc. 6(2), 185–189.

Mahalanobis, P.C., Bose, R.C., Roy, S.N., 1937. Normalisation of statistical variates and the use of rectangular coordinates in the theory of sampling distributions. Sankhyā 3, 1–40.

Neudecker, H., Liu, S., 1996. The density of the Moore–Penrose inverse of a random matrix. Linear Algebra Appl. 237/238, 123–126.

Olkin, I., 1951. On distribution problems in multivariate analysis. Doctoral Dissertation, University of North Carolina, Institute of Statistics Mimeograph Series, No. 43, pp. 1–126.

Olkin, I., 1953. Note on 'The Jacobians of certain matrix transformations useful in multivariate analysis'. Biometrika 40, 43–46.

Olkin, I., Roy, S.N., 1954. On multivariate distribution theory. Ann. Math. Statist. 25, 329–339.

Olkin, I., Sampson, A.R., 1972. Jacobians of matrix transformations and induced functional equations. Linear Algebra Appl. 5, 257–276.

Zhang, Y., 1985. The exact distribution of the Moore–Penrose inverse of X with a density in multivariate – Multivariate Analysis VI, P.R. Krishnaiah, (Ed.), Elsevier, New York, pp. 633–635.