ON MOMENTS OF QUADRATIC FORMS IN NON-SPERICALLY DISTRIBUTED VARIABLES

DOUGLAS P. WIENS†


(Received 8 April, 1991; in final form 11 March, 1992)

Summary. We derive the expectations of the matrices $xx' \otimes xx'$ and $xx' \otimes xx' \otimes xx'$, where $\otimes$ denotes a Kronecker product. From this, second and third moments of quadratic forms are obtained. The moments are derived under the following assumption on the product moments of the elements $X_i$ of $x$:

For any non-negative integers $a_1, \ldots, a_h$ with $\sum a_i \leq 6$, $E[X_{i_1}^{a_1}X_{i_2}^{a_2} \cdots X_{i_h}^{a_h}]$ is

i) finite;
ii) zero, if any $a_i$ is odd; and
iii) invariant under permutations of the indices.


Key words: Random matrices, quadratic forms.

1. INTRODUCTION

In this note, we evaluate the expectations of certain random matrices which arise in powers of quadratic forms. These are the matrices $(xx')^{(2)} = xx' \otimes xx'$, and $(xx')^{(3)} = xx' \otimes xx' \otimes xx'$. Here, $\otimes$ denotes a Kronecker product and $x$ is a $p$-dimensional random vector satisfying a symmetry—type condition, weaker than sphericity. Specifically, we make the following assumption on the product moments of the elements $X_i$ of $x$: For any non-negative integers $a_1, \ldots, a_h$, with $\sum a_i \leq 6$, $E[X_{i_1}^{a_1}X_{i_2}^{a_2} \cdots X_{i_h}^{a_h}]$ is

i) finite;
ii) zero, if any $a_i$ is odd; and
iii) invariant under permutations of the indices. (1.1)

Assumption (1.1) is implied by orthant symmetry, as defined by Efron (1969), in the presence of exchangeability. In particular, the assumption holds if $x$ follows a spherical distribution, or if the $X_i$ are distributed independently, identically and symmetrically.

Magnus and Neudecker (1979) obtained the expectations of $(xx')^{(2)}$ and $(xx')^{(3)}$, under the condition that $x$ have a spherical normal distribution. The results then follow for a general spherical distribution, using the fact that if the distribution of $x$ is spherical, then $x/\|x\|$ is distributed uniformly over the surface of the unit sphere, independently of $\|x\|$. Thus, if $N$ is the spherical normal

† Research supported by the National Sciences and Engineering Research Council of Canada.
distribution and $S$ a general spherical distribution,

$$E_x[(x^T x)^k] = E_x[(x^T x)^k]E_x[||x||^{2k}]/E_x[||x||^{2k}] \quad (1.2)$$

for any natural number $k$.

Pukelsheim (1980) calculates the covariance matrices of $x \otimes x \otimes x$ and $x \otimes x \otimes x$ under the assumption of quasi-normality, i.e. that the moments of the $X_i$, up to order 8, coincide with the corresponding normal moments. Drygas (1984) calculates these same quantities, again assuming quasi-normality, using an approach developed by Kleffe (1978). Similar results are exhibited in McCullagh and Pregibon (1987) using tensor and super-subscript notation.

Further results and references on moments of quadratic forms, under various assumptions, may be found in Chapter 2 of Rao and Kleffe (1988). See also Neudecker and Wansbeek (1987), von Rosen (1988) and Neudecker (1990) for results on second moments of matrix quadratic forms in normally distributed variables.

A derivation of the moment matrices, without assuming normality, sphericity or quasi-normality, requires a development quite separate from that in the aforementioned references. This is carried out in Section 3 below. In Section 2 some notation for moments, and some special matrices needed for the statement and derivation of the results, are presented.

The need for these results arose in an experimental design problem. There, the design points $x_1, \ldots, x_n$ are dictated by a certain design measure—a distribution function on the design space—so that sample moments are viewed as expectations of functions of $x$, with respect to this measure. A first order response is anticipated, but the true response is in fact quadratic. An expansion of the bias of an M-estimate of the regression parameters then turns out to involve the first three central moments of a quadratic form in $x$. It is not difficult to think of other problems in which one requires approximate expectations of functions of quadratic forms in non-normal variables. Our results should find applications in those areas as well.

2. NOTATION

a) Moments and Cumulants

We require the moments

$$\mu_2 = E[X^2], \quad \mu_{22} = E[X_1^2 X_2^2], \quad \mu_4 = E[X_1^4],$$

$$\mu_{222} = E[X_1^2 X_2^2 X_3^2], \quad \mu_{24} = E[X_1^4 X_2^2], \quad \mu_6 = E[X_1^6];$$

and cumulants

$$\kappa_1 = \mu_4 - 3\mu_{22}, \quad \kappa_2 = \mu_{24} - 3\mu_{222}, \quad \kappa_3 = \mu_6 - 15\mu_{24} + 30\mu_{222}.$$

Note that for a spherical or quasi-normal distribution, $\kappa_1 = \kappa_2 = \kappa_3 = 0$. This follows from (1.2), and from the fact that these cumulants of the normal distribution are zero.

b) Matrix Notation

We require the following matrix operations and special matrices.
i) The Kronecker product $A_{pr} \otimes B_{qs} = (a_i b_j)_{pq}$. Note that if $A$ and $B$ are square, then

$$\text{tr}(A \otimes B) = \text{tr} A \cdot \text{tr} B. \quad (2.1)$$

ii) The Hadamard product $A_{pq} \circ B_{pq} = (a_i b_i)_{pq}$.

iii) If $A_{pq}$ has columns $a_1, \ldots, a_q$, then

$$\text{vec} A = (a_1^T, \ldots, a_q^T)^T : pq \times 1.$$

We have the identities

$$\text{vec}(ABC) = (C^T \otimes A) \text{vec} B,$$

$$(\text{vec} A)^T(\text{vec} B) = \text{tr} A B. \quad (2.2)$$

iv) We define
1) $v = \text{vec} I : p^2 \times 1$. (Throughout this paper, $I$ is the $p \times p$ identity matrix.)
2) $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T : p \times 1$, with "1" in the $i$th spot.
3) $E_1 = (e_1 \otimes e_1, \ldots, e_p \otimes e_p) : p^2 \times p$. Note that

$$E_1^T (A \otimes B) E_1 = A \circ B \quad \text{for } A, B : p \times p. \quad (2.3)$$

4) $E_2 = (e_1 \otimes e_1, \ldots, e_p \otimes e_p) : p^3 \times p$. Note that

$$E_2^T (A \otimes B \otimes C) E_2 = A \circ B \circ C \quad \text{for } A, B, C : p \times p. \quad (2.4)$$

v) The commutation matrix $K$ (Magnus and Neudecker, 1979), also called the vec-permutation matrix $I_{(p,p)}$ by Henderson and Searle (1981) and Wiens (1985) is the symmetric, orthogonal, $p^2 \times p^2$ permutation matrix whose $(i, j)$th block is $e_i e_j^T$. It has the following properties.

1) $K \text{vec} A = \text{vec} A^T$ for $A : p \times p. \quad (2.5)$
2) $K(a \otimes b) = b \otimes a$ for $a, b : p \times 1. \quad (2.6)$
3) $K(A \otimes B) K = B \otimes A$ for $A, B : p \times p. \quad (2.7)$
4) $\text{tr} K(A \otimes B) = \text{tr} A B$ for $A, B : p \times p. \quad (2.8)$

vi) We define $Q : p^3 \times p^3$ by its action

$$Q(a_1 \otimes a_2 \otimes a_3) = \sum_\sigma (a_{n(1)} \otimes a_{n(2)} \otimes a_{n(3)}),$$

where the $a_i$ are $p \times 1$ and the sum is over all six permutations $\sigma$ of $(1, 2, 3)$. Explicitly,

$$Q = I \otimes I \otimes I + I \otimes K + K \otimes I + (I \otimes K)(K \otimes I) + (K \otimes I)(I \otimes K)(K \otimes I).$$

vii) We define $P : p^3 \times p^3$ by

$$P = I \otimes I \otimes I + (I \otimes K)(K \otimes I) + K \otimes I.$$
3. RESULTS AND DERIVATIONS

**Theorem.** Assume (1.1). Then

a) \( E[xx^T \otimes xx^T] = \mu_{22}[I \otimes I + K + vv^T] + \kappa_1 E_1 E_1^T. \)

b) For symmetric \( A \) and \( B, \)

\[
E[x^T Ax \cdot x^T Bx] = \mu_{22}[tr A \cdot tr B + 2 tr AB] + \kappa_1 tr A \cdot tr B.
\]

c) \( E[xx^T \otimes xx^T] = \mu_{22}[Q + P(I \otimes vv^T)P^T] + \kappa_2[(E_2 \otimes v^v)^TP + P(E_3^T \otimes v^v) + P(I \otimes E_1 E_1^T)P^T] + \kappa_3 E_2 E_2^T. \)

d) For symmetric \( A, B, C: \)

i) \( E[x^T Ax \cdot x^T Bx \cdot x^T Cx] = \mu_{22}[tr A \cdot tr B \cdot tr C + 2(tr A \cdot tr BC + tr B \cdot tr AC + tr C \cdot tr AB) + 8 tr ABC] + \kappa_2[tr A \cdot tr BC + tr B \cdot tr AC + tr C \cdot tr AB] + 4\{tr A \cdot (BC) + tr B \cdot (AC) + tr C \cdot (AB)\}] + \kappa_3 tr A \cdot tr B \cdot tr C.
\]

ii) \( E[(x^T Ax - E[x^T Ax])(x^T Bx - E[x^T Bx])(x^T Cx - E[x^T Cx])] = 8\mu_{22} tr ABC + (\mu_{22} + 2\mu_2^2 - 3\mu_2 \mu_{22}) tr A \cdot tr B \cdot tr C + 2(\mu_{22} - \mu_2 \mu_{22})[tr A \cdot tr BC + tr B \cdot tr AC + tr C \cdot tr AB] + (\kappa_2 - \mu_2 \kappa_1)[tr A \cdot tr B \cdot tr C + tr B \cdot tr A \cdot tr C + tr C \cdot tr A \cdot tr B] + 4\kappa_2[tr A \cdot (BC) + tr B \cdot (AC) + tr C \cdot (AB)] + \kappa_3 tr A \cdot tr B \cdot tr C.
\]

iii) \( E[(x^T Ax)^3] = \mu_{222}(tr A)^3 + 6 tr A \cdot tr A^2 + 8 tr A^3] + \kappa_2[3 tr A \cdot tr A^2 + 12 tr A \cdot tr A^3] + \kappa_3 tr(A * A \cdot A).
\]

iv) \( E[(x^T Ax - E[x^T Ax]^3] = 8\mu_{22} tr A^3 + (\mu_{22} + 2\mu_2^2 - 3\mu_2 \mu_{22})(tr A)^3 + 6(\mu_{22} - \mu_2 \mu_{22}) tr A \cdot tr A^2 + 3(\kappa_2 - \mu_2 \kappa_1) tr A \cdot tr A \cdot tr A + 12\kappa_2 tr A \cdot tr A^2 + \kappa_3 tr A \cdot tr A \cdot tr A.
\)

**Remarks.** 1. In d) ii) and d iv) above, the second, third and fourth of the six terms vanish if the \( x_i \) are iid; the fifth and sixth vanish if the \( x_i \) are spherically distributed. Thus, for a spherical normal distribution, ii) and iv) become \( 8\mu_2 \) tr \( ABC \) and \( 8\mu_2 \) tr \( A^3 \), respectively.

2. If \( p = 1 \) or \( p = 2 \), then some of the moments in the statement of the Theorem are undefined. They cancel, however, after they are formally inserted in \( \kappa_1, \kappa_2, \) or \( \kappa_3. \) The resulting expressions are then correct.

**Proof of the Theorem**: a) First note that

\[
E[x_i x_j xx^T] = \begin{cases} 
\mu_{22}(e_i e_j^T + e_i e_j^T), & i \neq j; \\
\mu_{22}(I + 2e_i e_i^T) + \kappa_1 e_i e_i^T, & i = j.
\end{cases}
\]

Then

\[
E[xx^T \otimes xx^T] = \sum_{i,j=1}^{p} \{e_i e_j^T \otimes E[x_i x_j xx^T]\}
\]

\[
= \mu_{22} \sum_{i,j=1}^{p} \{e_i e_j^T \otimes (e_i e_i^T + e_j e_j^T)\} + \sum_{i=1}^{p} \{e_i e_i^T \otimes [\mu_{22} I + \kappa_1 e_i e_i^T]\}
\]

\[
= \mu_{22}(vv^T + K) + \mu_{22}(I \otimes I) + \kappa_1 E_1 E_1^T.
\]

b) We have

\[
E[x^T Ax \cdot x^T Bx] = E[(x^T \otimes x^T)(A \otimes B)(x \otimes x)]
\]

\[
= tr(A \otimes B)E[xx^T \otimes xx^T].
\]
This is evaluated using (2.1), (2.2), (2.3) and (2.8).

c) Note that

\[ E[x_i x_j x_k x_l x^{T}] = \mu_{222}(e_i e_j^T + e_k e_l^T); \]
\[ k = i \neq j, l \neq i, l \neq j; \]
\[ = \mu_{222} I + (\mu_{24} - \mu_{222})(e_i e_j^T + e_k e_l^T); \]
\[ k = l = i \neq j; \]
\[ = \mu_{24} I + (\mu_{24} - \mu_{222})e_i e_l^T; \]
\[ k = l = i = j; \]
\[ = 0, \text{ if there are no equalities among } i, j, k, l. \]

This leads to

\[ E[x_i x_j x^{T} \otimes x^{T}] = \sum_{k,l=1}^{p} e_k e_l^T \otimes E[x_i x_j x^{T} \otimes x^{T}] \]
\[ = \mu_{222} ((I + K)(e_i \otimes e_j \otimes \nu^T) + (e_i^T \otimes e_j^T \otimes \nu)(I + K)) \]
\[ + (I + K)((e_i e_j^T + e_k e_l^T) \otimes I)(I + K)) \]
\[ + \kappa_2 (\{(e_i e_j^T + e_k e_l^T) \otimes (e_i e_j^T + e_k e_l^T)\}) \]
\[ + \{(e_i e_j^T + e_k e_l^T) \otimes (e_i e_l^T + e_k e_i^T)\} \]
(3.2)

if \( i \neq j \), and

\[ = \mu_{222} (I \otimes I + K + \nu \nu^T) + (\mu_{24} - \mu_{222})((I + K)(I \otimes e_i e_j^T)(I + K)) \]
\[ + \kappa_2 E_i E_j^T + (\kappa_2 + 8\kappa_2)(e_i e_j^T \otimes e_i e_j^T) \]
(3.3)

if \( i = j \).

Now (3.2) and (3.3) in

\[ E[xx^T \otimes xx^T \otimes xx^T] = \sum_{i,j=1}^{p} (e_i e_j^T \otimes E[x_i x_j x^{T} \otimes x^{T}] \]

give c).

d) Statement d i) follows from c) in the same manner as b) follows from a), using

\[ E[x^{T}Ax \cdot x^{T}Bx \cdot x^{T}C] = \text{tr}(A \otimes B \otimes C)E[xx^T \otimes xx^T \otimes xx^T]. \]
(3.4)

as at (3.1). The reader who is interested in verifying, rather than deriving, d i) may do so quite simply by noting that it suffices that d i) hold whenever A, B, C are each of rank 1, say \( A = aa^T, B = bb^T, C = cc^T \). In this case the right hand side of (3.4) becomes

\[ (a \otimes b \otimes c)^T E[xx^T \otimes xx^T \otimes xx^T](a \otimes b \otimes c), \]

which is easier to evaluate than is the general case of (3.4).

Statements d ii), d iii) and d iv) are immediate consequences of d i).

References

