Biometrika (1984), **71**, 3, *pp.* 461–76
Printed in Great Britain

Tensor notation and cumulants of polynomials

BY PETER McCULLAGH
Department of Mathematics, Imperial College, London, U.K.

SUMMARY

A modified and extended tensor notation is introduced that is sufficient to cover multivariate moments and cumulants as special cases. Using this notation, two basic identities are given. The first of these expresses generalized cumulants in terms of ordinary cumulants. The second gives the joint cumulant generating function of any polynomial transformation in terms of the cumulants of the original variables. Three applications of the basic identities are given. The first application is concerned with sample cumulants or k-statistics, the second to Edgeworth series and the third to exponential family models.

Some key words: Bartlett correction factor; Cumulant; Cumulant tensor; Edgeworth series; Exponential family; Generalized cumulant; Graph; Hermite tensor; k-statistic; Möbius function; Partition lattice; Pattern function; Polynomial transformation; Signed likelihood ratio statistic; Tensor.

1. INTRODUCTION

The purpose of the present paper is to describe and illustrate a simple formula for computing joint cumulants of the q-dimensional statistic $Y = g(X)$ in terms of the known cumulants of the p-dimensional statistic X and the derivatives of $g(.)$. A similar problem was considered and solved by Leonov & Shiryaev (1959) who give a rather unwieldy expression for the cumulants of Y. The simplicity of the result given in the present paper is due primarily to the use of tensor notation as opposed to the conventional 'power' notation for cumulants. Despite the efforts of Kaplan (1952), tensor notation remains unpopular in statistics primarily because the conventional notation is more convenient when the components of X are independent and identically distributed. We hope that the modified tensor notation introduced in § 2 will help reverse the trend and draw attention to the advantages of a unified notation that includes both moments and cumulants as special cases.

It is assumed that the function $g(.)$ may be expanded as a polynomial in X so that if Y has components $Y^1, \ldots, Y^q$ and X has components $X^1, \ldots, X^p$ then the transformation may be written, using the summation convention,

$$Y^r = a^r + a^r_i X^i + a^r_{ij} X^i X^j + a^r_{ijk} X^i X^j X^k + \ldots. \tag{1}$$

Without loss of generality and with considerable gain in simplicity, all arrays are assumed to be symmetric under subscript permutation. In subsequent sections it will be convenient to express the transformation (1) using generalized 'matrix' or operator notation as

$$Y = (A_0 + A_1 + A_2 + \ldots) X. \tag{2}$$

For consistency of notation it would have been better to write A^1_s instead of A_s above to indicate that there is a single superscript in addition to the variable number of

subscripts. However, the superscript can be taken as understood. It will also be convenient to rewrite (2), using an operator P, as

$$Y = PX. \tag{3}$$

The formula $K_Y(\xi) = \exp(\xi P)\kappa$, giving the cumulant generating function of Y in terms of the cumulants of X, is derived in two stages. At the first stage, generalized cumulants of order α and degree β in X are introduced and these are expressed in § 3 in terms of the ordinary cumulants of X for which $\alpha = \beta$. It will be seen that moments are a special case of generalized cumulants for which $\alpha = 1$. In § 4, the formula for $K_Y(\xi)$ is given in terms of the generalized cumulants of X.

Three applications are discussed. The first application is concerned with sample cumulants, otherwise known as k-statistics. It is shown that the associated tensors are mutually orthogonal so the k-statistics may be thought of as symmetric homogeneous orthogonal polynomials. Orthogonality greatly simplifies the calculation of joint cumulants of the k-statistics since many terms vanish. Those terms that do not vanish are scalar products of tensors and are simply related to Fisher's (1929) diagrams.

The second application is to Edgeworth series where it is shown that the simple structure of the univariate series is retained if tensor notation is employed.

The third application is to exponential families where we give the cumulants of $\hat{\theta}$ and of the likelihood ratio statistic correct to order n^{-1}. A simple demonstration of the validity of the Bartlett correction factor is also given.

2. Notational conventions

Standard notation for the joint cumulants of a p-dimensional statistic X (Kendall & Stuart, 1969, Ch. 3) employs p subscripts whose sum gives the order of the cumulant. For example, if $p = 3$, $\kappa_{100} = E(X^1)$, $\kappa_{011} = \operatorname{cov}(X^2, X^3)$, $\kappa_{002} = \operatorname{var}(X^3)$ and so on. A similar convention is typically used for moments so that $\mu'_{100} = E(X^1)$, $\mu'_{011} = E(X^2 X^3)$, $\mu'_{002} = E\{(X^3)^2\}$. The difficulties created by such a cumbersome notation are very forcibly expressed by Ryall (1981).

In contrast, the notation employed here for both moments and cumulants uses a single symbol κ having a variable number of indices as superscripts. The value of the index, rather than its position, refers to the component of X: the number of indices, denoted by β, gives the degree of the moment or cumulant. Commas are used to distinguish cumulants from moments and the order of the indices is immaterial. Thus, for example,

$$\kappa^i = E(X^i), \quad \kappa^{ij} = E(X^i X^j), \quad \kappa^{i,j} = \operatorname{cov}(X^i, X^j), \quad \kappa^{ijk} = E(X^i X^j X^k)$$

and $\kappa^{i,j,k} = \kappa^{ijk} - \kappa^i \kappa^{j,k} - \kappa^j \kappa^{i,k} - \kappa^k \kappa^{i,j} - \kappa^i \kappa^j \kappa^k$ is the third-order mixed cumulant corresponding to the variables X^i, X^j, X^k. There is, of course, no requirement that the indices should be distinct so that it is legitimate to put $i = j = k$ giving $\kappa^{i,i,i} = \kappa^{iii} - 3\kappa^i \kappa^{i,i} - (\kappa^i)^3$ as a special case.

Generalized cumulants are intermediate between moments and cumulants. For example, $\kappa^{i,jk} = \operatorname{cov}(X^i, X^j X^k)$ is said to be of order $\alpha = 2$ and degree 3 or simply of order $(2, 3)$. The order is one more than the number of commas so that a $(1, \beta)$-cumulant is an ordinary moment of degree β and an (α, α)-cumulant is an ordinary cumulant of order α. Generalized cumulants can be expressed in terms of ordinary cumulants as the following

examples show:

$$\kappa^{i,jk} = \kappa^{i,j,k} + \kappa^j \kappa^{i,k} + \kappa^k \kappa^{i,j}, \tag{4}$$

$$\kappa^{ij,kl} = \kappa^{i,j,k,l} + \kappa^i \kappa^{j,k,l}[4] + \kappa^{i,k} \kappa^{j,l}[2] + \kappa^{i,k} \kappa^j \kappa^l[4], \tag{5}$$

where, for example,

$$\kappa^{i,k} \kappa^{j,l}[2] = \kappa^{i,k} \kappa^{j,l} + \kappa^{i,l} \kappa^{j,k}$$

is to be understood from the context: see (6) below.

In the following section we describe an algorithm useful for expressing generalized cumulants in terms of ordinary cumulants. The expressions for moments in terms of cumulants follows as a special case of the general formula.

The tensor notation and terminology used above are justified by observing that under linear transformation $X^i \to a_i^r X^i$ the moments and cumulants obey the transformation law of contravariant tensors, $\kappa^i \to a_i^r \kappa^i$, $\kappa^{ij} \to a_i^r a_j^s \kappa^{ij}$ and so on. See, for example, Speed (1983, proposition 4·2). Thus the cumulants are certainly cartesian tensors (Jeffreys & Jeffreys, 1956, Ch. 4). Furthermore the Hermite tensors introduced in §6 obey the transformation law of covariant tensors but again only under linear transformation.

3. Partitions, cumulants and connected graphs

Without loss of generality we consider (α, β)-cumulants of degree β in the variables $X^1, \ldots, X^\beta$. Corresponding to any such cumulant there is an α-partition $\mathscr{D}^* = \{D_1^*, \ldots, D_\alpha^*\}$ of the integers $1, \ldots, \beta$ into nonempty sets $D_1^*, \ldots, D_\alpha^*$. Conversely, for each α-partition of the integers $1, \ldots, \beta$ there is a corresponding (α, β) cumulant denoted by $\kappa(\mathscr{D}^*)$ or, where convenient, by $\mathscr{D}^*$. Thus, for example, the $(3, 6)$-cumulant $\kappa^{135,26,4}$ corresponds to the partition $\mathscr{D}^* = \{(1, 3, 5), (2, 6), (4)\}$ and is distinct from similar $(3, 2, 1)$-partitions such as $\{(2, 4, 6)\ (1, 5), (3)\}$. The cumulants corresponding to the elements of $\mathscr{D}^*$ are $\kappa(D_1^*) = \kappa^{1,3,5}$, $\kappa(D_2^*) = \kappa^{2,6}$ and $\kappa(D_3^*) = \kappa^4$ so that, in general, $\kappa(D_j^*)$ is an ordinary cumulant whose order and degree are the same.

The expression giving generalized cumulants in terms of ordinary cumulants is most conveniently expressed using the terminology of graph theory. In this development, the partition $\mathscr{D}^*$ and the cumulant $\kappa(\mathscr{D}^*)$ are represented as a graph on β vertices labelled $1, \ldots, \beta$ or by the variables $X^1, \ldots, X^\beta$. The graph comprises α disconnected cliques corresponding to the α sets $D_1^*, \ldots, D_\alpha^*$. In other words, (X^i, X^j) is an edge of the graph if and only if, for some k, i and j, both belong to the set D_k^*. Thus, for $\alpha > 1$ the graph of $\mathscr{D}^*$ is disconnected while for $\alpha = 1$, corresponding to an ordinary moment, the graph of $\mathscr{D}^*$ is a single clique, i.e. a complete graph.

If $\mathscr{D}$ and $\mathscr{D}^*$ are two graphs on the same β vertices with edges E and E^* respectively, we define the edge sum $\mathscr{D} \oplus \mathscr{D}^*$ to be the graph on the same vertices whose edges are $E \cup E^*$ (Behzad, Chartrand & Lesniak-Foster, 1979, p. 13). The edge sum is defined for arbitrary graphs but we will be concerned only with graphs corresponding to partitions. Thus the use of the same notation for cumulants, partitions and graphs causes no confusion. We will be particularly interested in partitions $\mathscr{D}$ such that $\mathscr{D} \oplus \mathscr{D}^*$ is connected. Two particular cases merit special attention:

(i) if $\kappa(\mathscr{D}^*)$ is an ordinary moment then $\mathscr{D}^*$ is the complete graph on β vertices and $\mathscr{D} \oplus \mathscr{D}^* = \mathscr{D}^*$ is connected for every $\mathscr{D}$;

(ii) if $\kappa(\mathscr{D}^*)$ is an ordinary cumulant then E^* is empty so that $\mathscr{D} \oplus \mathscr{D}^* = \mathscr{D}$ is connected only if $\mathscr{D}$ is the complete graph.

The general expression, derived in the Appendix, for $\kappa(\mathscr{D}^*)$ in terms of ordinary cumulants is

$$\kappa(\mathscr{D}^*) = \Sigma_{\mathscr{D}}\, \kappa(D_1)\, \kappa(D_2) \dots \kappa(D_\nu), \tag{6}$$

where $\mathscr{D} = \{D_1, \dots, D_\nu\}$ and the sum is over all partitions $\mathscr{D}$ such that $\mathscr{D} \oplus \mathscr{D}^*$ is a connected graph. This result is given by James (1958) and may be deduced from Leonov & Shiryaev (1959). See also Brillinger (1975, § 2·3). In the particular case corresponding to (i) above the sum is over all partitions, while for (ii) there is only a single partition satisfying the requirement that $\mathscr{D} \oplus \mathscr{D}^*$ be connected so that (6) becomes a trivial identity in this case.

Expressions (4) and (5) are special cases of (6). A third special case is

$$\kappa^{i,jkl} = \kappa^{i,j,k,l} + \kappa^j \kappa^{i,k,l}[3] + \kappa^{i,j} \kappa^{k,l}[3] + \kappa^{i,j} \kappa^k \kappa^l[3],$$

where $\mathscr{D}^* = \{(i), (j,k,l)\}$ and the partitions corresponding to terms on the right are

$$\{(i,j,k,l)\}, \quad \{(j),(i,k,l)\}, \quad \{(k),(i,j,l)\}, \quad \{(l),(i,j,k)\}, \quad \{(i,j),(k,l)\}$$

and so on.

Expression (6) may be viewed as giving the joint cumulant of $Y^1, \dots, Y^\alpha$ corresponding to the particular polynomial transformation

$$Y^1 = \prod_{j \in D_1^*} X^j, \quad Y^2 = \prod_{j \in D_2^*} X^j, \dots,$$

where the components of X, though labelled distinctly for notational convenience, may in fact refer to the same random variable. Thus it comes as no surprise to learn that (6) plays a major role in determining the joint cumulants for arbitrary polynomial transformations.

The converse of (6), giving cumulants or generalized cumulants in terms of the moments may be written

$$\kappa(\mathscr{D}^*) = \sum_{\mathscr{D} \geqslant \mathscr{D}^*} (-1)^{\nu-1} (\nu-1)!\, M(D_1) \dots M(D_\nu),$$

where $M(D_j)$ is the moment corresponding to the indices in D_j and the sum is over all partitions $\mathscr{D}$ such that $\mathscr{D}^*$ is a subpartition of $\mathscr{D}$ (Kendall & Stuart, 1969, p. 319).

A proof of (6) using Möbius inversion on the lattice of set partitions is given by Speed (1983) and a simplified version of that proof avoiding double indices is given in the Appendix. The condition that $\mathscr{D} \oplus \mathscr{D}^*$ be connected is replaced by the equivalent condition $\mathscr{D} \vee \mathscr{D}^* = 1$ meaning that the least upper bound of the two partitions should be the full set. For a discussion of Möbius inversion on arbitrary lattices, see Rota (1964).

4. CUMULANTS OF POLYNOMIALS

4·1. *General points*

We now examine in more detail the operators A and P introduced at (2) and (3) in § 1. Expression (1) defines the action of A and hence of P on X. We now define, in a similar way, the action of A on κ, where κ is a structure representing the cumulants of X. A few simple examples will suffice:

$$1\kappa = 0, \quad A_0 \kappa = a^r, \quad A_1 \kappa = a_i^r \kappa^i, \quad A_2 \kappa = a_{ij}^r \kappa^{ij}, \quad A_3 \kappa = a_{ijk}^r \kappa^{ijk}, \tag{7}$$

and so on. Compound operators introduce commas as follows:

$$A_1 A_1 \kappa = a_i^r a_j^s \kappa^{i,j}, \quad A_2 A_1 \kappa = a_{ij}^r a_k^s \kappa^{ij,k},$$
$$A_2 A_3 A_1 \kappa = a_{ij}^r a_{klm}^s a_n^t \kappa^{ij,klm,n}, \quad A_0 A_1 \kappa = a^r a_i^s \kappa^{,i} = 0, \tag{8}$$

and, more generally, any compound term involving A_0 acting on κ gives zero. The operators A are not commutative.

With the above definitions we may now write the cumulant generating function of Y formally as

$$K_Y(\xi) = \exp(\xi P)\kappa, \tag{9}$$

where the exponent is to be expanded formally as a function of A. This gives

$$\begin{aligned} K_Y(\xi) &= \xi_r(A_0 + A_1 + A_2 + \ldots)\kappa + \tfrac{1}{2}\xi_r \xi_s\{A_1 A_1 + (A_1 A_2 + A_2 A_1) + A_2 A_2 + \ldots\}\kappa \\ &\quad + \tfrac{1}{6}\xi_r \xi_s \xi_t\{A_1 A_1 A_1 + A_1 A_1 A_2[3] + A_1 A_2 A_2[3] + \ldots\}\kappa + \ldots \qquad (10) \\ &= \xi_r\{a^r + a_i^r \kappa^i + a_{ij}^r \kappa^{ij} + \ldots\} + \tfrac{1}{2}\xi_r \xi_s\{a_i^r a_j^s \kappa^{i,j} + 2a_i^r a_{jk}^s \kappa^{i,jk} + a_{ij}^r a_{kl}^s \kappa^{ij,kl} + \ldots\} \\ &\quad + \tfrac{1}{6}\xi_r \xi_s \xi_t\{a_i^r a_j^s a_k^t \kappa^{i,j,k} + 3a_i^r a_j^s a_{kl}^t \kappa^{i,j,kl} + \ldots\} + \ldots . \qquad (11) \end{aligned}$$

There are a number of ways in which (9) may be derived. One route that is not recommended is to reexpress formula IV(b) of Leonov & Shiryaev (1959) using the operator notation described above. A much simpler method is to proceed directly by observing that the moment generating function of Y is given by

$$M_Y(\xi) = \exp(\xi P) * \kappa \tag{12}$$

where the operator $*$ is defined as in (7) and (8), except that the commas are omitted in (8) and $1 * \kappa = 1$. Expansion of (12) gives (10) and (11) with minor modifications such as removal of commas and addition of 1. The effect of taking logarithms is to delete the leading constant, 1, and to insert commas although the latter step requires a small supporting argument. Expression (9) is in fact a concise summary of results derived by James & Mayne (1962) who also provide details sufficient to construct a proof.

4·2. *Quadratic forms*

Suppose that $X^1, \ldots, X^p$ are independent and identically distributed random variables, the rth cumulant of X^i being denoted by κ_r, using power notation. It is required to find the cumulants of the quadratic form $Q = a_{ij} X^i X^j$. From (9), or on further expansion of (11), the first three cumulants are $a_{ij}\kappa^{ij}$, $a_{ij} a_{kl} \kappa^{ij,kl}$ and $a_{ij} a_{kl} a_{mn} \kappa^{ij,kl,mn}$. These may be expressed in terms of ordinary cumulants as

$$a_{ij}(\kappa^{i,j} + \kappa^i \kappa^j), \quad a_{ij} a_{kl}(\kappa^{i,j,k,l} + \kappa^i \kappa^{j,k,l}[4] + \kappa^{i,k} \kappa^{j,l}[2] + \kappa^i \kappa^k \kappa^{j,l}[4]),$$
$$\begin{aligned} a_{ij} a_{kl} a_{mn}(&\kappa^{i,j,k,l,m,n} + \kappa^i \kappa^{j,k,l,m,n}[6] + \kappa^{i,k} \kappa^{j,l,m,n}[12] + \kappa^i \kappa^k \kappa^{j,l,m,n}[12] \\ &+ \kappa^{i,k,m} \kappa^{j,l,n}[4] + \kappa^{i,j,k} \kappa^{l,m,n}[6] + \kappa^i \kappa^{j,k} \kappa^{l,m,n}[24] + \kappa^i \kappa^{k,m} \kappa^{j,l,n}[24] \\ &+ \kappa^i \kappa^k \kappa^m \kappa^{j,l,n}[8] + \kappa^{i,k} \kappa^{j,m} \kappa^{l,n}[8] + \kappa^i \kappa^k \kappa^{j,m} \kappa^{l,m}[24]). \end{aligned}$$

Now if we revert to power notation and use the property of independence, the cumulants of Q may be written

$$a_{ii}\kappa_2+a_{ij}\kappa_1^2,\quad a_{ii}^2\kappa_4+4a_{ij}a_{jj}\kappa_1\kappa_3+2a_{ij}^2\kappa_2^2+4a_{ij}a_{ik}\kappa_1^2\kappa_2,$$
$$a_{ii}^3\kappa_6+6a_{ij}a_{jj}^2\kappa_1\kappa_5+12a_{ij}^2a_{jj}\kappa_2\kappa_4+12a_{ij}a_{kj}a_{jj}\kappa_1^2\kappa_4+(4a_{ij}^3+6a_{ii}a_{jj}a_{ij})\kappa_3^2$$
$$+24(a_{ij}a_{jk}a_{kk}+a_{ij}^2a_{ik})\kappa_1\kappa_2\kappa_3+8a_{ij}a_{ik}a_{il}\kappa_1^3\kappa_3+8a_{ij}a_{ik}a_{jk}\kappa_2^3+24a_{ij}a_{kl}a_{jl}\kappa_1^2\kappa_2^2,$$

with summation over all indices. In particular, if $\kappa_1=0$ and $\{a_{ij}\}=\operatorname{diag}\{\kappa_2^{-1},\ldots,\kappa_2^{-1}\}$, the cumulants become

$$p,\quad p(2+\kappa_4/\kappa_2^2),\quad p(8+10\kappa_3^2/\kappa_2^3+12\kappa_4/\kappa_2^2+\kappa_6/\kappa_2^3).$$

The corresponding calculations for the fourth cumulant are rather tedious because the expression (6) for the fourth cumulant of four products involves 2465 partitions of 30 different types. However, if, as above, Q is a sum of squares of standardized variates, all but 12 of the types vanish and the fourth cumulant becomes

$$p(48+240\kappa_3^2/\kappa_2^3+144\kappa_4/\kappa_2^2+32\kappa_4^2/\kappa_2^4+56\kappa_3\kappa_5/\kappa_2^4+24\kappa_6/\kappa_2^3+\kappa_8/\kappa_2^4).$$

5. SAMPLE CUMULANTS

Suppose that $X_1,\ldots,X_n$ are independent and identically distributed random vectors. Each vector $X_i=X_i^1,\ldots,X_i^p$ has moments $\kappa^r,\kappa^{rs},\kappa^{rst},\ldots$ and cumulants $\kappa^r,\kappa^{r,s},\kappa^{r,s,t},\ldots$. The corresponding sample statistics will be denoted by $k^r,k^{rs},k^{rst},\ldots$ and $k^r,k^{r,s},k^{r,s,t},\ldots$ and have the following properties:

(i) the sample cumulants, including moments, are unbiased estimates of the population cumulants;

(ii) the sample cumulants are symmetric functions, i.e. are invariant under permutations of $X_1,\ldots,X_n$.

It is convenient at this stage to introduce the diagonal arrays $\delta^r,\delta^{rs},\ldots,\delta_r,\delta_{rs},\ldots$ and $\delta_s^r,\delta_t^{rs},\ldots$ which take the value unity when all indices are equal and zero otherwise. Trivially, therefore, $\delta^r=1$ and $\delta_r=1$. The sample moments may now be written $k^r=n^{-1}\delta^iX_i^r$, $k^{rs}=n^{-1}\delta^{ij}X_i^rX_j^s,\ldots$, or, more formally, using matrix notation

$$M^r=n^{-1}\Delta^rX, \tag{13}$$

where M^r is the sample moment tensor of degree r.

From (9) or, alternatively, by direct calculation, it may be seen that the sample moments are unbiased estimates of the population moments. The joint cumulants of the sample moments may also be found from (9). We find, for example, that the covariance of k^{rs} and k^{tu} is

$$n^{-2}\delta^{ij}\delta^{kl}\kappa_{ij,kl}^{rs,tu}, \tag{14}$$

where $\kappa_{ij,kl}^{rs,tu}=\operatorname{cov}(X_i^rX_j^s,X_k^tX_l^u)$. Simplification of (14) gives

$$\begin{aligned}\kappa(rs;tu)&=\operatorname{cov}(k^{rs},k^{tu})=n^{-1}\kappa^{rs,tu}\\&=n^{-1}\{\kappa^{r,s,t,u}+\kappa^r\kappa^{s,t,u}[4]+\kappa^{r,t}\kappa^{s,u}[2]+\kappa^{r,t}\kappa^s\kappa^u[4]\}.\end{aligned} \tag{15}$$

There is an obvious generalization to joint cumulants of several sample moments corresponding to the α-partition $\mathscr{D}^*$ of the indices $r,s,t,\ldots$. Using the notation of § 3, we

may write

$$\kappa_M(\mathscr{D}^*) = n^{1-\alpha}\kappa(\mathscr{D}^*).$$

In particular, if $\alpha = 1$, so that $\mathscr{D}^*$ contains a single set, the above result states simply that the sample moments are unbiased estimates of the population moments.

From a single tensor formula may typically be derived several 'scalar' formulae resulting in considerable economy of tabulation. Formula (15) above is equivalent to seven types of scalar formulae which would appear distinct if the conventional power notation were used. For example, the variances of a sum of squares and a sum of products are

$$\operatorname{var}(m_2) = (\kappa_4 + 4\kappa_3\kappa_1 + 2\kappa_2^2 + 4\kappa_2\kappa_1^2)/n \quad (r = s = t = u),$$

$$\operatorname{var}(m_{11}) = (\kappa_{22} + 2\kappa_{10}\kappa_{12} + 2\kappa_{01}\kappa_{21} + 2\kappa_{11}^2 + 4\kappa_{11}\kappa_{01}\kappa_{10})/n \quad (r = s,\ t = u),$$

where the indices above are of the conventional 'power' type.

In the case of sample cumulants, as opposed to sample moments, we may write, analogous to (13), $K^r = n^{-1}\Phi^r X$, or, in tensor notation,

$$k^s = n^{-1}\phi^i X_i^s, \quad k^{s,t} = n^{-1}\phi^{ij} X_i^s X_j^t, \quad k^{s,t,u} = n^{-1}\phi^{ijk} X_i^s X_j^t X_k^u, \ldots$$

and aim to choose the tensor coefficients ϕ to satisfy the criteria (i) and (ii) above. Since

$$E(X_i^r) = \kappa_i^r = \kappa^r\delta_i, \quad E(X_i^r X_j^s) = \kappa_{ij}^{rs} = \kappa^{r,s}\delta_{ij} + \kappa^r\kappa^s\delta_i\delta_j,$$

$$E(X_i^r X_j^s X_k^t) = \kappa_{ijk}^{rst} = \kappa^{r,s,t}\delta_{ijk} + \kappa^r\kappa^{s,t}\delta_i\delta_{jk}[3] + \kappa^r\kappa^s\kappa^t\delta_i\delta_j\delta_k,$$

it follows that the k's are unbiased if

$$\phi^i\delta_i = n, \quad \phi^{ij}\delta_j = 0, \quad \phi^{ij}\delta_{ij} = n, \quad \phi^{ijk}\delta_k = 0, \quad \phi^{ijk}\delta_{ijk} = n,$$

and, for cumulants of order 4,

$$\phi^{ijkl}\delta_l = 0, \quad \phi^{ijkl}\delta_{kl} = 0, \quad \phi^{ijkl}\delta_{ijkl} = n.$$

It follows by symmetry that $\phi^{ijk}\delta_{jk} = 0$ and more generally that, for $r > s$, the tensor of inner products $\Phi^r\Delta_s$ is zero while $\Phi^r\Delta_r = n$. It may be verified by induction that the components of Φ^r are uniquely determined and are given by

$$\phi^{i\ldots l} = (-1)^{s-1}\Big/\binom{n-1}{s-1}, \tag{16}$$

where s is the number of distinct superscripts. Note that unlike Δ^r, Φ^r is defined only for $r \leqslant n$. There is a close connexion between (16) and the Möbius function for the partition lattice.

As a consequence of the orthogonality criterion, it follows that the tensors Φ^r are themselves mutually orthogonal in the same sense. For example,

$$\phi^i\phi^{jk}\delta_{ij} = 0, \quad \phi^i\phi^{jkl}\delta_{ij} = 0, \quad \phi^{ij}\phi^{klm}\delta_{ik}\delta_{jl} = 0, \quad \phi^{ij}\phi^{klm}\delta_{ijkl} = 0,$$

and so on. Furthermore $\phi^{ii} = \phi^i$, $\phi^{iij} = \phi^{ij}, \ldots$ and, more generally, repeated superscripts may be elided.

In writing down the nonzero scalar products of the Φ's it is convenient to depart temporarily from standard tensor notation by summing over indices repeated, however, frequently, as superscripts. Thus we write $\phi^i\phi^i$ instead of $\phi^i\phi^j\delta_{ij}$, $\phi^i\phi^i\phi^i$ instead of

$\phi^i \phi^j \phi^k \delta_{ijk}$ and so on. The squared Euclidean norms of the Φ's are

$$\begin{gathered}\phi^i \phi^i = n, \quad \phi^{ij} \phi^{ij} = n^2/(n-1), \\ \phi^{ijk} \phi^{ijk} = n^3/\{(n-1)(n-2)\}, \quad \phi^{ijkl} \phi^{ijkl} = n^3(n+1)/\{(n-1)(n-2)(n-3)\}.\end{gathered} \tag{17}$$

There is a variety of nonzero scalar products involving several Φ's. For example,

$$\begin{gathered}\phi^i \ldots \phi^i = n, \quad \phi^{ij} \phi^{ij} \phi^{ij} = n^2(n-2)/(n-1)^2, \\ \phi^{ij} \phi^{ij} \phi^{ij} \phi^{ij} = n^2(n^2-3n+3)/(n-1)^3, \quad \phi^{ij} \phi^{jk} \phi^{ki} = n^3/(n-1)^2, \\ \phi^{ijk} \phi^{jkl} \phi^{il} = n^4/\{(n-1)^2(n-2)\}, \quad \phi^{ij} \phi^{ijk} \phi^{ik} = n^3/(n-1)^2.\end{gathered} \tag{18}$$

We may now compute the joint cumulants of the sample cumulants using (9) and (6). For example, the covariance of $k^{r,s}$ and $k^{t,u}$, denoted by $\kappa(r, s;\, t, u)$ is

$$\begin{aligned}\kappa(r, s;\, t, u) &= n^{-2} \phi^{ij} \phi^{kl} \kappa^{rs,tu}_{ij,kl} \\ &= n^{-2} \phi^{ij} \phi^{kl} \{\kappa^{r,s,t,u} \delta_{ijkl} + \kappa^r \kappa^{s,t,u}[4] \delta_i \delta_{jkl} \\ &\qquad + \kappa^{r,t} \kappa^{s,u}[2] \delta_{ik} \delta_{jl} + \kappa^{r,t} \kappa^s \kappa^u [4] \delta_{ik} \delta_j \delta_l\} \\ &= \kappa^{r,s,t,u}/n + (\kappa^{r,t} \kappa^{s,u} + \kappa^{r,u} \kappa^{s,t})/(n-1),\end{aligned}$$

involving only three terms rather than the eleven terms in the corresponding expression (15) for sample moments. This simplification is purely a consequence of the orthogonality of Φ^r and Δ_s. Similarly we find for the covariance of $k^{r,s}$ and $k^{t,u,v}$

$$\kappa(r, s;\, t, u, v) = n^{-2} \phi^{ij} \phi^{klm} \kappa^{rs,tuv}_{ij,klm}.$$

Application of (6) with $\mathscr{D}^* = \{(i, j), (k, l, m)\}$ gives

$$\kappa(r, s;\, t, u, v) = n^{-2} \phi^{ij} \phi^{klm} \{\kappa^{r,s,t,u,v} \delta_{ijklm} + \kappa^{r,t,u} \kappa^{s,v}[6] \delta_{ikl} \delta_{jm} + \kappa^{r,s,t} \kappa^{u,v}[3] \delta_{ijk} \delta_{lm}\},$$

where partitions involving unit parts have been dropped because $\phi^{ij} \delta_i = 0$ and $\phi^{klm} \delta_m = 0$, and the $(2, 3)$ partition has been split into two parts the second of which contributes nothing because $\phi^{klm} \delta_{lm} = 0$. Thus

$$\kappa(r, s;\, t, u, v) = \kappa^{r,s,t,u,v}/n + \kappa^{r,t,u} \kappa^{s,v}[6]/(n-1).$$

The general expression for the joint cumulant of α sample cumulants corresponding to the partition $\mathscr{D}^*$ is easily seen to be

$$\kappa(\mathscr{D}^*) = n^{-\alpha} \phi(D_1^*) \ldots \phi(D_\alpha^*) \sum_{\mathscr{D}} \kappa(D_1) \ldots \kappa(D_\nu) \delta(D_1) \ldots \delta(D_\nu), \tag{19}$$

where the sum extends over all partitions such that $\mathscr{D} \oplus \mathscr{D}^*$ is connected. In the above formula we have made use of the notation of § 3 and also of the 1–1 correspondence between the sets of indices $i, j, k, \ldots$ and $r, s, t, \ldots$. Furthermore,

$$\phi(D^*) \delta(D) = \begin{cases} 0 & (D \subset D^*) \\ n & (D = D^*), \\ \delta(D - D^*) & (D^* \subset D), \end{cases}$$

no simplification being possible otherwise. Thus any partition $\mathscr{D}$ having a part D such that $D \subset D_k^*$ for some k makes no contribution to (19). In addition, using the above

formula we may write

$$\phi^{ij}\phi^{k}\delta_{jk} = \phi^{ij}\delta_{j} = 0, \quad \phi^{ij}\phi^{klm}\delta_{ijkl} = \phi^{klm}\delta_{kl} = 0,$$

$$\phi^{ij}\phi^{kl}\delta_{ijkl} = \phi^{ij}\delta_{ij} = n, \quad \phi^{ij}\phi^{kl}\delta_{ijklm} = \phi^{kl}\delta_{klm} = \delta_{m},$$

and so on. In the case of the fourth cumulant of a sample variance we may write

$$\mathscr{D}^* = \{(p,q),(r,s),(t,u),(v,w)\}.$$

The partition $\mathscr{D} = \{(p,r),(q,s,t),(u,v,w)\}$ satisfies the requirement that $\mathscr{D} \oplus \mathscr{D}^*$ be connected and is one of the 12 types containing no unit parts that makes a contribution to the fourth cumulant of a sum of squares of standardized variates (§4·2). However, $\phi^{tu}\phi^{vw}\delta_{uvw} = \phi^{tu}\delta_{u} = 0$ so that there is no contribution to (19). Such partitions can therefore be ignored and this corresponds to Fisher's (1929) rules (3) and (4). In particular, all partitions having unit parts are automatically excluded. Finally, Fisher's diagrams, with the columns corresponding to the partition $\mathscr{D}^*$ and the rows to $\mathscr{D}$ provide a simple method for computing the nonzero scalar products $\phi(D_1^*)\ldots\phi(D_\alpha^*)\,\delta(D_1)\ldots\delta(D_\nu)$ as in (17) and (18). For example, the patterns

$$\begin{array}{ccc} \times\ \times\ \times, & \begin{array}{ccc} \times & \times & \times \\ & & \\ . & \times & \times \end{array}, & \begin{array}{ccc} \times & \times & \times \\ & & \\ \times & \times & \times \end{array}, \end{array} \quad \begin{array}{ccc} \times & . & \times \\ \times & \times & . \\ . & \times & \times \end{array} \tag{20}$$

correspond to the scalar products

$$\phi^{i}\phi^{i}\phi^{i} = n, \quad \phi^{i}\phi^{ij}\phi^{ij} = n^2/(n-1) = \phi^{ij}\phi^{ij},$$

$$\phi^{ij}\phi^{ij}\phi^{ij} = n^2(n-2)/(n-1)^2, \quad \phi^{ij}\phi^{jk}\phi^{ik} = n^3/(n-1)^2.$$

These differ from the functions given by Fisher by the factor n^3 or, more generally, by the factor n^α where α is the number of blocks of $\mathscr{D}^*$ or the number of columns in the diagram.

For an application of the general result, we may consider the third cumulant of three variances or covariances which is

$$\begin{aligned}\kappa(r,s;\,t,u;\,v,w) = n^{-3}\{&\kappa^{r,s,t,u,v,w}n + \kappa^{r,s,t,v}\kappa^{u,w}[12]\,n^2/(n-1)\\ &+\kappa^{r,t,v}\kappa^{s,u,w}[4]\,n^2(n-2)/(n-1)^2 + \kappa^{r,t}\kappa^{s,r}\kappa^{u,w}[8]\,n^3/(n-1)^2\},\end{aligned} \tag{21}$$

where the four functions of n correspond to the four diagrams (20). Further formulae of this type are given by Kaplan (1952).

As usual with tensor formulae, scalar formulae may be derived by equating sets of indices. For example, the third cumulant of a sample variance is

$$\kappa_6/n^2 + 12\kappa_2\kappa_4/\{n(n-1)\} + 4\kappa_3^2(n-2)/\{n(n-1)^2\} + 8\kappa_2^3/(n-1)^2,$$

having set $r = s = \ldots = w$ and reverted to 'power' notation. All of Fisher's 32 univariate formulae as well as the bivariate formulae may be derived from (19) provided that the scalar products for the appropriate patterns are available.

Recently, Speed (1985) has given more general, though less transparent, expressions for k-statistics and polykays together with the sampling cumulants of these statistics for both finite and infinite populations.

6. EDGEWORTH SERIES

6·1. *Hermite tensors*

The components of the Hermite tensors $h_r(x,\lambda), h_{rs}(x,\lambda), h_{rst}(x,\lambda), \ldots$ are polynomials in x whose degree is the same as the order of the tensor. Ordinary Hermite tensors, $h_r(x,\delta), h_{rs}(x,\delta), \ldots$ are obtained by successive partial differentiation of the standard p-variate normal density

$$\phi(x,\delta) = (2\pi)^{-\frac{1}{2}p} \exp(-\tfrac{1}{2}x^r x^s \delta_{rs}).$$

Writing the differential operator $d_r = \partial/\partial x^r$ as a covariant tensor and using the 'matrix' notation

$$D_1 = \{d_r\}, \quad D_2 = \{d_r d_s\}, \quad D_3 = \{d_r d_s d_t\}, \ldots,$$

$$H_1(x,\delta) = \{h_r(x,\delta)\}, \quad H_2(x,\delta) = \{h_{rs}(x,\delta)\}, \ldots,$$

we have that $D_j \phi(x,\delta) = (-1)^j H_j(x,\delta)\,\phi(x)$. The first six ordinary Hermite tensors are

$$h_r(x,\delta) = \delta_{rs} x^s = x'_r, \quad h_{rs}(x,\delta) = x'_r x'_s - \delta_{rs},$$

$$h_{rst}(x,\delta) = x'_r x'_s x'_t - x'_r \delta_{st}[3], \quad h_{rstu}(x,\delta) = x'_r \ldots x'_u - x'_r x'_s \delta_{tu}[6] + \delta_{rs}\delta_{tu}[3],$$

$$h_{rstuv}(x,\delta) = x'_r \ldots x'_v - x'_r x'_s x'_t \delta_{uv}[10] + x'_r \delta_{st}\delta_{uv}[15],$$

$$h_{r\ldots w}(x,\delta) = x'_r \ldots x'_w - x'_r \ldots x'_u \delta_{vw}[15] + x'_r x'_s \delta_{tu}\delta_{vw}[45] - \delta_{rs}\delta_{tu}\delta_{vw}[15].$$

The ordinary Hermite tensors can be obtained from the Hermite polynomials. For example, the coefficient of x^4 in $H_{10}(x)$ is 3150. The corresponding term in the Hermite tensor $H_{10}(x,\delta)$ is $x'_r x'_s x'_t x'_u \delta_{vw}\delta_{ij}\delta_{kl}[3150]$, where

$$3150 = \binom{10}{6} 6!/(2^3\, 3!)$$

is the number of partitions of ten indices leading to terms of the required type.

Generalized Hermite tensors, formed by differentiation of $\phi(x,\lambda)$, are obtained by replacing δ_{rs} with λ_{rs} in (29). Thus

$$h_r(x,\lambda) = \lambda_{rs} x^s = x'_r, \quad h_{rs}(x,\lambda) = x'_r x'_s - \lambda_{rs} = \lambda_{rt}\lambda_{su}(x^t x^u - \kappa^{t,u})$$

and so on, where $\kappa^{r,s}$ is the inverse of λ_{rs}. Notice that if the variables are uncorrelated then $h_{rs} = h_r h_s$ $(r \neq s)$, $h_{rst} = h_r h_{st}$ $(r \neq s, t)$ and so on.

To investigate the properties of Hermite tensors with generality, it is necessary to use 'matrix' notation and to distinguish between direct products such as

$$(x_r - d_r)\, h_s(x,\delta) = h_{rs}(x,\delta),$$

which increase the order of the tensor, and inner products such as $d^t h_{rst}(x,\delta) = 3h_{rs}(x,\delta)$, where $d^t = \delta^{ts} d_s$, which decrease the order. In 'matrix' notation we may write

$$(x - D_1)\, H_j(x,\delta) = H_{j+1}(x,\delta), \quad D^1 H_j(x,\delta) = jH_{j-1}(x,\delta).$$

Hermite tensors are also orthogonal in the sense that

$$\int h_r h_s \phi\, dx = \delta_{rs}, \quad \int h_r h_{st} \phi\, dx = 0, \quad \int h_{rs} h_{tu} \phi\, dx = \delta_{rt}\delta_{su}[2!],$$

$$\int h_{rst} h_{uvw} \phi\, dx = \delta_{ru}\delta_{sv}\delta_{tw}[3!], \quad \int h_{rs} h_{tuv} \phi\, dx = 0,$$

and so on, where $\delta_{rt}\delta_{su}[2!] = \delta_{rt}\delta_{su} + \delta_{ru}\delta_{st}$ and integration is over R^p.

We now examine the application of Hermite tensors to Edgeworth series. Suppose therefore that the random variable X has mean zero and cumulants $\kappa^{r,s}$, $\kappa^{r,s,t}/\sqrt{n}$, $\kappa^{r,s,t,u}/n, \ldots$ decreasing in powers of $n^{\frac{1}{2}}$ where n is typically a sample size. The first-order approximation for the density of X, using the central limit theorem, is $\phi(x, \lambda)$, the p-variate normal density. The Edgeworth expansion with two correction terms may be written

$$\phi(x,\lambda)\,[1+\kappa^{r,s,t}\,h_{rst}(x,\lambda)/(6\sqrt{n}) \\ +\{3\kappa^{r,s,t,u}\,h_{rstu}(x,\lambda)+\kappa^{r,s,t}\,\kappa^{u,v,w}\,h_{r\ldots w}(x,\lambda)\}/(72n)]+O(n^{-3/2}).$$

Written in this way, the multivariate Edgeworth series is no more complicated than the univariate series. Furthermore, the fact that the covariance matrix is not the identity does not in any way complicate the formulae. In fact there is no notational advantage in restricting attention to uncorrelated random variables; contrast Barndorff-Nielsen & Cox (1979). Similar formulae, though in different notation, were given by Chambers (1967) who also discusses the validity of Edgeworth approximations. The notation used here is essentially that used by Amari & Kumon (1983).

6·2. *Conditional cumulants*

Suppose now that $X_{(1)}, X_{(2)}$ is a partition of X into components of dimensions $p-q$ and q respectively. For the moment, we let the indices $r, s, t, \ldots$ range from 1 to p while $i, j, k, \ldots$ range only over the components of $X_{(2)}$. Let λ^*_{ij} be the inverse of $\kappa^{i,j}$, the covariance matrix of $X_{(2)}$. The marginal distribution of $X_{(2)}$ is

$$\phi(x,\lambda^*)\,[1+\kappa^{i,j,k}\,h_{ijk}(x,\lambda^*)/(6\sqrt{n}) \\ +\{3\kappa^{i,j,k,l}\,h_{ijkl}(x,\lambda^*)+\kappa^{i,j,k}\,\kappa^{l,m,n}\,h_{i\ldots n}(x,\lambda^*)\}/(72n)]+O(n^{-3/2}).$$

Thus the leading term in the expansion of the conditional distribution of $X_{(1)}$ given $X_{(2)}$ is $\phi(x,\lambda)/\phi(x,\lambda^*)$,

$$(2\pi)^{-\frac{1}{2}(p-q)}\,|\lambda^*_{ij}|^{-\frac{1}{2}}\,|\lambda_{rs}|^{\frac{1}{2}}\exp\{-\tfrac{1}{2}(x^r x^s \lambda_{rs}-x^i x^j \lambda^*_{ij})\},$$

which is a $(p-q)$-variate normal density with mean $\kappa^{r,i}\lambda^*_{ij}x^j = \beta^r_i x^i = \kappa^{r,i}\,h_i(x,\lambda^*)$ and inverse covariance matrix λ_{rs} restricted to the components of $X_{(1)}$. The correction factor in the conditional density, including terms of order $O(n^{-1})$ is

$$1+\{\kappa^{r,s,t}\,h_{rst}(x,\lambda)-\kappa^{i,j,k}\,h_{ijk}(x,\lambda^*)\}/6\sqrt{n}+\{\kappa^{r,s,t,u}\,h_{rstu}(x,\lambda)-\kappa^{i,j,k,l}\,h_{ijkl}(x,\lambda^*)\}/(24n) \\ +\{\kappa^{r,s,t}\,\kappa^{u,v,w}\,h_{r\ldots w}(x,\lambda)-\kappa^{i,j,k}\,\kappa^{l,m,n}\,h_{i\ldots n}(x,\lambda^*)\}/(72n) \\ -\kappa^{i,j,k}\,h_{ijk}(x,\lambda^*)\,\{\kappa^{r,s,t}\,h_{rst}(x,\lambda)-\kappa^{i,j,k}\,h_{ijk}(x,\lambda^*)\}/(36n). \tag{22}$$

Essentially the same expression is given by Barndorff-Nielsen & Cox (1979).

To compute the conditional cumulants of $X_{(1)}$ we first make a linear transformation to uncorrelated variables Y, where $Y^r = X^r-\beta^r_j X^j$, $Y^j = X^j$, and, now, indices $r, s, t, \ldots$

range only over the components of $X_{(1)}$. The cumulants of Y, found using (9), are

$$\kappa_1^r = 0, \quad \kappa_1^{r,s} = \kappa^{r,s} - \beta_i^r \beta_j^s \kappa^{i,j}, \quad \kappa_1^{i,r} = 0,$$

$$\kappa_1^{r,s,t} = \kappa^{r,s,t} - \beta_i^r \kappa^{i,s,t}[3] + \beta_i^r \beta_j^s \kappa^{i,j,t}[3] - \beta_i^r \beta_j^s \beta_k^t \kappa^{i,j,k},$$

$$\kappa_1^{i,r,s} = \kappa^{i,r,s} - \beta_j^r \kappa^{i,j,s}[2] + \beta_h^r \beta_k^s \kappa^{i,j,k}, \quad \kappa_1^{i,j,r} = \kappa^{i,j,r} - \beta_k^r \kappa^{i,j,k},$$

$$\kappa_1^{r,s,t,u} = \kappa^{r,s,t,u} - \beta_i^r \kappa^{i,s,t,u}[4] + \beta_i^r \beta_j^s \kappa^{i,j,t,u}[6] - \beta_i^r \beta_j^s \beta_k^t \kappa^{i,j,k,u}[4] + \beta_i^r \beta_j^s \beta_k^t \beta_l^u \kappa^{i,j,k,l},$$

$$\kappa_1^{i,r,s,t} = \kappa^{i,r,s,t} - \beta_j^r \kappa^{i,j,s,t}[3] + \beta_j^r \beta_k^s \kappa^{i,j,k,t}[3] - \beta_j^r \beta_k^s \beta_l^t \kappa^{i,j,k,l},$$

$$\kappa_1^{i,j,r,s} = \kappa^{i,j,r,s} - \beta_k^r \kappa^{i,j,k,s}[2] + \beta_k^r \beta_l^s \kappa^{i,j,k,l}, \quad \kappa_1^{i,j,k,r} = \kappa^{i,j,k,r} - \beta_l^r \kappa^{i,j,k,l},$$

where the factors $n^{-\frac{1}{2}}$ and n^{-1} have been omitted in the third and fourth cumulants. To first order in n, $\{Y^r\}$ is independent of $\{Y^j\}$ so that the conditional mean and variance of X^r are $\beta_j^r X^j$ and $\kappa_1^{r,s}$, the remaining cumulants being $O(n^{-\frac{1}{2}})$ or smaller.

To the next order of approximation, in which terms of order $O(n^{-\frac{1}{2}})$ are included, we find from (22) that the conditional cumulants of X^r are

$$\kappa^{r,i} h_i(x, \lambda^*) + \kappa_1^{r,i,j} h_{ij}(x, \lambda^*)/(2\sqrt{n}), \quad \kappa_1^{r,s} + \kappa_1^{r,s,i} h_i(x, \lambda^*)/\sqrt{n}$$

and $\kappa_1^{r,s,t}/\sqrt{n}$. In other words, the regression function of $E(X^r)$ on $X_{(2)}$ is quadratic while the conditional covariance is linear in $X_{(2)}$. Fourth and higher order conditional cumulants are $O(n^{-1})$ or smaller.

To the third order of approximation, the conditional cumulants of $X_{(1)}$ given $X_{(2)}$ are

$$\kappa^{r,i} h_i + \kappa_1^{r,i,j} h_{ij}/(2\sqrt{n}) + \{\kappa_1^{r,i,j} \kappa^{k,l,m}(h_{i\ldots m} - h_{ij} h_{klm}) + 2\kappa_1^{r,i,j,k} h_{ijk}\}/(12n),$$

$$\kappa_1^{r,s} + \kappa_1^{r,s,i} h_i/\sqrt{n} + \{6\kappa_1^{r,s,i,j} h_{ij} + 2\kappa_1^{r,s,i} \kappa^{j,k,l}(h_{ijkl} - h_i h_{jkl}) + 3\kappa_1^{r,i,j} \kappa_1^{s,k,l}(h_{ijkl} - h_{ij} h_{kl})\}/(12n),$$

$$\kappa_1^{r,s,t}/\sqrt{n} + \{2\kappa_1^{r,s,t,i} h_i + \kappa_1^{r,s,i} \kappa_1^{t,j,k}(h_{ijk} - h_i h_{jk})\,[3]\}/(2n),$$

$$\kappa_1^{r,s,t,u}/n + \kappa_1^{r,s,i} \kappa_1^{t,u,j}(h_{ij} - h_i h_j)\,[3]/n,$$

where the arguments, (x, λ^*) of $h(.)$ have been suppressed for brevity. Note that $h_{ijklm} - h_{ij} h_{klm}$ is cubic in x^i, the n^{-1} term in the conditional variance is quadratic, the n^{-1} term in the conditional skewness is linear and the conditional kurtosis is constant. If we had first applied a nonlinear transformation to new variables $Y' = (Y'_{(1)}, Y'_{(2)})$, where $Y'_{(2)}$ is a function of $X_{(2)}$ alone, with associated cumulants $\kappa_2^{r,s}, \kappa_2^{r,i} = 0$, $\kappa_2^{i,j}, \kappa_2^{r,s,t}, \kappa_2^{r,s,i} = 0$, $\kappa_2^{r,i,j} = 0$, $\kappa_2^{i,j,k} = 0, \ldots$, then the $O(n^{-1})$ correction terms for the conditional cumulants would be $\kappa_2^{r,i,j,k} h_{ijk}/(6n)$, $\kappa_2^{r,s,i,j} h_{ij}/(2n)$ and $\kappa_2^{r,s,t,i} h_i/n$, a special case of a more general result.

7. Linear exponential family models

7·1. *Cumulants of the likelihood ratio statistic*

Suppose that the log likelihood for $\theta = (\theta_1, \ldots, \theta_p)$ based on X may be written

$$\theta_i X^i - \kappa(\theta). \tag{23}$$

If we use superscipts to denote differentiation with respect to components of θ, the cumulants of X may be written

$$\kappa^r = \kappa^r(\theta), \quad \kappa^{r,s} = \kappa^{r,s}(\theta), \ldots, \tag{24}$$

justifying the notation in (23). Differentiation of (24) with respect to κ^i gives

$$\delta_i^r = \kappa^{r,s}(\theta)\,\theta_{si}(\kappa), \tag{25}$$

where $\theta_{si}(\kappa) = \theta_{is}(\kappa)$ is the derivative of θ_s with respect to κ^i and symmetry justifies the tensor notation. Thus we may write analogous to (24) that

$$\theta_i = \theta_i(\kappa), \tag{26}$$

the notation expressing the fact that the components of θ are derivatives with respect to κ of a scalar field $\theta(\kappa)$. The function $\theta(\kappa) = \theta_i\kappa^i - \kappa(\theta)$ is the solution to the system of partial differential equations (26) and is in fact the Legendre transformation of $\kappa(\theta)$ (Barndorff-Nielsen, 1978, Ch. 5).

In addition to (25) which gives the second derivatives of $\theta(\kappa)$ in terms of the cumulants of X, higher derivatives may also be obtained by further differentiation. For example

$$\theta_{ijk} = -\kappa^{r,s,t}\,\theta_{ri}\,\theta_{sj}\,\theta_{tk},$$

$$\theta_{ijkl} = -\kappa^{r,s,t,u}\,\theta_{ri}\,\theta_{sj}\,\theta_{tk}\,\theta_{ul} + \kappa^{r,s,t}\,\kappa^{u,v,w}\,\theta_{ri}\,\theta_{sj}\,\theta_{tu}\,\theta_{vk}\,\theta_{w}[3],$$

where the arguments θ and κ have been omitted for brevity.

Suppose now that it is required to test the simple null hypothesis H_0: $\theta_i = \theta_i^0$, or equivalently that $\kappa^r = \kappa_0^r = \kappa^r(\theta^0)$, based on a simple random sample of size n for which the log likelihood is $l(\theta; x) = n\{\theta_i\bar{X}^i - \kappa(\theta)\}$. Since $\hat{\kappa}^r = \bar{X}^r$ and $\hat{\theta}_r = \theta_r(\bar{X})$, the log likelihood ratio statistic may be written as $W(\bar{X}; \theta^0)$, where $W(\bar{X}; \theta) = 2n[X^i\theta_i(\bar{X}) - X^i\theta_i - \{\kappa(\theta(\bar{X})) - \kappa(\theta)\}]$. Furthermore,

$$\partial W/\partial\bar{X}^i = 2n\{\theta_i(\bar{X}) - \theta_i\} = 2n(\hat{\theta}_i - \theta_i), \quad \partial W/\partial\theta_j = 2n(\bar{X}^j - \kappa^j)$$

from which we obtain the more convenient expression

$$W(\bar{X}; \theta) = 2n\{\theta(\bar{X}) - \theta - \theta_i(\bar{X}^i - \kappa^i)\}.$$

Thus the log likelihood ratio statistic is simply the Taylor series expansion of $2n\theta(\bar{X})$ about κ_0^j with the first two terms omitted. Expansion in powers of $Z^i = \bar{X}^i - \kappa^i$ gives

$$W = 2n\{\tfrac{1}{2}\theta_{ij}Z^iZ^j + \tfrac{1}{6}\theta_{ijk}Z^iZ^jZ^k + \tfrac{1}{24}\theta_{ijkl}Z^iZ^jZ^kZ^l + \ldots\},$$

a form suitable for application of the identity (9). The Bartlett correction factor (Lawley, 1956; Cordeiro, 1983) is readily obtained as

$$\begin{aligned} E(W/p) &= 1 + \{3\kappa^{r,s,t}\,\kappa^{u,v,w}\,\theta_{rs}\,\theta_{tu}\,\theta_{vw} + 2\kappa^{r,s,t}\,\kappa^{u,v,w}\,\theta_{ru}\,\theta_{sv}\,\theta_{tw} - 3\kappa^{r,s,t,u}\,\theta_{rs}\,\theta_{tu}\}/(12np) \\ &\qquad + O(n^{-2}) \\ &= 1 + b/n + O(n^{-2}) \end{aligned} \tag{27}$$

For scalar parameters the correction may be expressed using power notation as $(5\rho_3^2 - 3\rho_4)/12$, where ρ_3 and ρ_4 are the standardized third and fourth cumulants of $\bar{X}$.

Approximate significance levels, with error $O(n^{-2})$, are obtained by referring the statistic $W/(1+b/n)$ to the χ_p^2 distribution.

The signed or directional likelihood ratio statistic (Barndorff-Nielsen, 1984) is a vector $\bar{Y}^r$ having p components and such that $n\bar{Y}^r\,\bar{Y}^s\theta_{rs} = W$. We may write

$$\begin{aligned} \bar{Y}^r &= Z^r - \kappa^{r,s,t}\,\theta_{si}\,\theta_{tj}\,Z^iZ^j/6 \\ &\quad + \{8\kappa^{r,s,t}\,\kappa^{u,v,w}\,\theta_{si}\,\theta_{tu}\,\theta_{vj}\,\theta_{wk} - 3\kappa^{r,s,t,u}\,\theta_{si}\,\theta_{tj}\,\theta_{uk}\}\,Z^iZ^jZ^k/72 + O_p(n^{-2}) \end{aligned} \tag{28}$$

since $Z^i = O_p(n^{-\frac{1}{2}})$. The mean and variance of $n^{\frac{1}{2}}\,\bar{Y}^r$ may be found using identities (9) and (6) to be

$$-\kappa^{r,s,t}\,\theta_{st}/(6n^{\frac{1}{2}})+O(n^{-3/2}), \tag{29}$$

$$\kappa^{r,s}+(6\kappa^{r,t,u}\,\kappa^{s,v,w}\,\theta_{tv}\,\theta_{uw}+8\kappa^{r,s,t}\,\kappa^{u,v,w}\,\theta_{tv}\,\theta_{uw}-9\kappa^{r,s,t,u}\,\theta_{tu})/(36n)+O(n^{-2}). \tag{30}$$

The third and fourth cumulants are $O(n^{-3/2})$ and $O(n^{-2})$ respectively, higher order cumulants being $O(n^{-3/2})$ or smaller. Thus the likelihood-based transformation (28) induces elliptical symmetry to third order, but the mean and variance differ from the reference values 0 and $\kappa^{r,s}$ by terms of order $O(n^{-\frac{1}{2}})$ and $O(n^{-1})$. It follows therefore, that, with error $O(n^{-2})$, W has a scaled noncentral χ^2_p distribution with noncentrality parameter

$$a/n = \kappa^{r,s,t}\,\kappa^{u,v,w}\,\theta_{ru}\,\theta_{st}\,\theta_{vw}/(36n)+O(n^{-2})$$

and scale factor

$$1+\{6\kappa^{r,t,u}\,\kappa^{s,v,w}\,\theta_{rs}\,\theta_{tv}\,\theta_{uw}+8\kappa^{r,s,t}\,\kappa^{u,v,w}\,\theta_{rs}\,\theta_{tv}\,\theta_{uw}$$
$$-9\kappa^{r,s,t,u}\,\theta_{rs}\,\theta_{tu}\}/(36np)+O(n^{-2}) = 1+c/n,$$

say. The rth cumulant of W (Johnson & Kotz, 1970, p. 134) is

$$\kappa_r = 2^{r-1}(r-1)!\,p(1+c/n)^r\{1+ar/(np)\}+O(n^{-2})$$
$$= \{1+(c+a/p)/n\}^r\,2^{r-1}(r-1)!\,p+O(n^{-2}).$$

Thus, with $b = c+a/p$, all cumulants of $W/(1+b/n)$ differ from those of χ^2_p by terms of order $O(n^{-2})$, justifying the use of the Bartlett correction factor for significance testing.

In the case of scalar parameters, we may write $S = \pm W^{\frac{1}{2}}$, where the sign of S is that of $Z = \bar{X}-\kappa_0$. The asymptotic distribution of S is then given by

$$\{S+\rho_3/6\}\,\{1+(9\rho_4-14\rho_3^2)/72\} \sim N(0,1)+O_p(n^{-3/2}),$$

if we assume that Edgeworth expansion of S is valid.

7·2. *Cumulants of* $\hat{\theta}$

Since $\hat{\theta}_r = \theta_r(\bar{X})$, we may write

$$\hat{\theta}_r = \theta_r+\theta_{ri}\,Z^i+\tfrac{1}{2}\theta_{rij}\,Z^i\,Z^j+\theta_{rijk}\,Z^i\,Z^j\,Z^k/6+\ldots.$$

The cumulants of the standardized statistic $n^{\frac{1}{2}}(\hat{\theta}_r-\theta_r)$ may be found using (9) and (6). The first four are

$$\tfrac{1}{2}n^{-\frac{1}{2}}\,\theta_{rij}\,\kappa^{i,j}+n^{-3/2}(4\theta_{rijk}\,\kappa^{i,j,k}+3\theta_{rijkl}\,\kappa^{i,j}\,\kappa^{k,l})/24+O(n^{-5/2}),$$
$$\theta_{ri}\,\theta_{sj}\,\kappa^{i,j}+\{\theta_{ri}\,\theta_{sjk}\,\kappa^{i,j,k}[2]+\theta_{ri}\,\theta_{sjkl}\,\kappa^{i,j}\,\kappa^{k,l}[2]+\theta_{rij}\,\theta_{skl}\,\kappa^{i,k}\,\kappa^{j,l}\}/(2n)+O(n^{-2}),$$
$$\theta_{ri}\,\theta_{sj}\{\theta_{tk}\,\kappa^{i,j,k}+\theta_{tkl}\,\kappa^{i,k}\,\kappa^{j,l}[3]\}/\sqrt{n}+O(n^{-3/2}),$$
$$\theta_{ri}\,\theta_{sj}\{\theta_{tk}(\theta_{ul}\,\kappa^{i,j,k,l}+3\theta_{ulm}\,\kappa^{i,j,l}\,\kappa^{k,m}[4]+\theta_{ulmn}\,\kappa^{i,l}\,\kappa^{j,m}\,\kappa^{k,m}[4])$$
$$+2\theta_{tkl}\,\theta_{umn}\,\kappa^{i,k}\,\kappa^{j,m}\,\kappa^{l,n}[6]\}/n+O(n^{-2}),$$

where we have included the $O(n^{-3/2})$ term in the first cumulant but not elsewhere, e.g. third and fifth cumulants. The figures in [.] refer to permutations of the indices r, s, t, u.

After elimination of the higher derivatives of θ, these cumulants may be written

$$-\tfrac{1}{2}n^{-\frac{1}{2}}\kappa^{i,j,k}\theta_{ir}\theta_{jk}+O(n^{-3/2}),$$

$$\theta_{rs}+\{3\kappa^{i,j,k}\kappa^{l,m,n}\theta_{ir}\theta_{ls}\theta_{jm}\theta_{kn}+2\kappa^{i,j,k}\kappa^{l,m,n}\theta_{ir}\theta_{js}\theta_{kl}\theta_{mn}-2\kappa^{i,j,k,l}\theta_{ir}\theta_{js}\theta_{kl}\}/(2n)+O(n^{-2}),$$

$$-2\theta_{ri}\theta_{sj}\theta_{tk}\kappa^{i,j,k}/\sqrt{n}+O(n^{-3/2}),\quad \{4\kappa^{i,j,m}\kappa^{k,l,n}\theta_{mn}[3]-3\kappa^{i,j,k,l}\}\theta_{ir}\theta_{js}\theta_{kt}\theta_{lu}/n+O(n^{-2}).$$

Appendix

Derivation of formula (6) *for generalized cumulants*

The following proof of formula (6) uses Möbius inversion on the lattice of set partitions and clarifies a proof given by Speed (1983). Let $\mathscr{L}_n$ be the lattice of partitions of the set of indices $\mathscr{S} = (1, 2, \ldots, n)$. Partitions of $\mathscr{S}$ are denoted by $\mathscr{D}, \mathscr{D}^*$, the blocks of the partition being the sets $D_1, \ldots, D_\nu$ or $D_1^*, \ldots, D_\alpha^*$ as appropriate. The following functions are defined on $\mathscr{L}_n$: $F(\mathscr{D}^*) = \kappa(D_1^*)\ldots\kappa(D_\alpha^*)$, a product of cumulants, $M(\mathscr{D}^*) = M(D_1^*)\ldots M(D_\alpha^*)$, a product of moments and $\kappa(\mathscr{D}^*)$ which is the joint cumulant of the α variables

$$\prod_{i\in D_1^*} X^i, \ldots, \prod_{i\in D_\alpha^*} X^i.$$

Speed (1983, eqn (3·1)) shows that

$$M(\mathscr{D}^*) = \sum_{\mathscr{D}\leqslant\mathscr{D}^*} F(\mathscr{D}) = \sum_{\mathscr{D}\leqslant\mathscr{D}^*} \kappa(D_1)\ldots\kappa(D_\nu). \tag{A1}$$

By Möbius inversion we have that

$$F(\mathscr{D}_1) = \Sigma(-1)^{\nu-1}(\nu-1)!\,M(\mathscr{D}) = \kappa(\mathscr{D}_n), \tag{A2}$$

where $\mathscr{D}_1 = \{(1, 2, \ldots, n)\}$ is the greatest element of $\mathscr{L}_n$ and $\mathscr{D}_n = \{(1), (2), \ldots, (n)\}$ is the least element. Also from (A2) we have more generally that

$$\kappa(\mathscr{D}^*) = \sum_{\mathscr{D}\leqslant\mathscr{D}^*} (-1)^{\nu-1}(\nu-1)!\,M(D_1)\ldots M(D_\nu) \tag{A3}$$

and, using (A1), this becomes

$$\kappa(\mathscr{D}^*) = \sum_{\mathscr{D}\geqslant\mathscr{D}^*} (-1)^{\nu-1}(\nu-1)! \sum_{\mathscr{P}\leqslant\mathscr{D}} \kappa(P_1)\ldots\kappa(P_\sigma). \tag{A4}$$

It is straightforward to see that any partition $\mathscr{P}$ such that the least upper bound $\mathscr{P}\vee\mathscr{D}^* = \mathscr{D}_1$ contributes the term $\kappa(P_1)\ldots\kappa(P_\sigma)$. Other partitions make no contribution. Thus

$$\kappa(\mathscr{D}^*) = \sum_{\mathscr{D}\vee\mathscr{D}^*=\mathscr{D}_1} \kappa(D_1)\ldots\kappa(D_\nu) \tag{A5}$$

and this expression is the same as (6). The least element and the greatest element of the lattice are commonly written as 0 and 1 respectively.

References

Amari, S. I. & Kumon, M. (1983). Differential geometry of Edgeworth expansions in curved exponential family. *Ann. Inst. Statist. Math.* A **35**, 1–24.

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. Chichester: Wiley.

Barndorff-Nielsen, O. E. (1984). Conditionality resolution and the likelihood-ratio ancillary for curved exponential models. *Scand. J. Statist.* **11**. To appear.

Barndorff-Nielsen, O. & Cox, D. R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. R. Statist. Soc.* B **41**, 279–312.

Behzad, M., Chartrand, G. & Lesniak-Foster, L. (1979). *Graphs & Digraphs*. Boston, Mass: Prindle, Weber and Schmidt.

BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart and Winston.

CHAMBERS, J. M. (1967). On methods of asymptotic approximation for multivariate distributions. *Biometrika* **54**, 367–83.

CORDEIRO, G. M. (1983). Improved likelihood-ratio tests for generalized linear models. *J. R. Statist. Soc.* B **45**, 404–13.

FISHER, R. A. (1929). Moments and product moments of sampling distributions. *Proc. Lond. Math. Soc.* Series 2, **30**, 199–238. Reprinted (1972) as paper 74 in *Collected Papers of R. A. Fisher*, **2**, Ed. J. H. Bennett, pp. 314–54. University of Adelaide Press.

JAMES, G. S. (1958). On moments and cumulants of systems of statistics. *Sankhyā* **20**, 1–30.

JAMES, G. S. & MAYNE, A. J. (1962). Cumulants of functions of random variables. *Sankhyā* **24**, 47–54.

JEFFREYS, H. & JEFFREYS, B. S. (1956). *Methods of Mathematical Physics*, 3rd edition. Cambridge University Press.

JOHNSON, N. L. & KOTZ, S. (1970). *Continuous Univariate Distributions*, **2**. Boston, Mass: Houghton-Mifflin.

KAPLAN, E. L. (1952). Tensor notation and the sampling cumulants of k-statistics. *Biometrika* **39**, 319–23.

KENDALL, M. G. & STUART, A. (1969). *The Advanced Theory of Statistics*, **1**, 3rd edition. London: Griffin.

LAWLEY, D. M. (1956). A general method for approximating to the distribution of certain likelihood-ratio criteria. *Biometrika* **43**, 295–303.

LEONOV, V. P. & SHIRYAEV, A. M. (1959). On a method of calculation of semi-invariants. *Theory Prob. Applic.* **4**, 319–29.

ROTA, G.-C. (1964). On the foundations of combinational theory, I. Theory of Mobius functions. *Zeit. f. Warsch.* **2**, 340–68.

RYALL, T. A. (1981). Extensions of the concept of local ancillarity. *Biometrika* **68**, 677–83.

SPEED, T. P. (1983). Cumulants and partition lattices. *Aust. J. Statist.* **25**, 378–88.

SPEED, T. P. (1985). Cumulants and partition lattices, II: Generalized k-statistics. *J. Aust. Math. Soc.* A **39**. To appear.

[*Received December* 1983. *Revised May* 1984]