J. Austral. Math. Soc. (*Series A*) **40** (1986), 34–53

CUMULANTS AND PARTITION LATTICES II: GENERALISED k-STATISTICS

T. P. SPEED

(Received 13 January 1984)

Communicated by T. C. Brown

Abstract

The role played by the Möbius function of the lattice of all partitions of a set in the theory of k-statistics and their generalisations is pointed out and the main results concerning these statistics are derived. The definitions and formulae for the expansion of products of generalised k-statistics are presented from this viewpoint and applied to arrays of random variables whose moments satisfy suitable symmetry constraints. Applications of the theory are given including the calculation of (joint) cumulants of k-statistics, the minimum variance estimation of (generalised) moments and the asymptotic behaviour of generalised k-statistics viewed as (reversed) martingales.

1980 *Mathematics subject classification* (*Amer. Math. Soc.*): 62 A 05.

1. Introduction

In an earlier paper (Speed (1983)) we showed how the definition and basic properties of mth order multivariate cumulants could be obtained elegantly and efficiently by recognising the role that the Möbius function μ of the lattice $\mathscr{P}(\mathbf{m})$ of partitions of $\mathbf{m} = \{1,\ldots,m\}$ plays in the theory. Our aim in this paper is to discuss Fisher's k-statistics and Tukey's polykays from this same viewpoint, once more deriving the main results of the theory by exploiting the role of μ and $\mathscr{P}(\mathbf{m})$.

We begin with a brief historical survey motivating our notion of generalised k-statistics, this being essentially equivalent to a polykay extension of Kaplan's (1952) tensor approach to k-statistics, and coinciding with the approach of Doubilet (1972) to the classical symmetric functions. Most of the useful results

concerning these generalised k-statistics stem from the fact that the (tensor) product of two such expands as a linear combination of higher-order ones, and in Section 3 we review the main results concerning the coefficients in these expansions.

After this discussion of generalised k-statistics we turn to arrays of random variables and, under certain symmetry constraints, define generalised moments and generalised cumulants, these reducing to the moments and cumulants indexed by $\mathscr{P}(\mathbf{m})$ as in our earlier paper when the arrays consist of independent and identically distributed (iid) random vectors. It is then easy to state and prove the main facts concerning the (minimum variance) unbiased estimation of generalised cumulants in terms of generalised k-statistics. In Section 5 we present a very compact proof of the main theorem of Fisher (1929) concerning the joint cumulants of his k-statistics, whilst Section 6 shows how generalised k-statistics are reversed martingales and hence converge almost surely to their expectations.

Most of the results in this paper can be found in some form in the literature, although we feel that the approach we take unifies, generalises and simplifies their presentation, but there is another reason for our collecting and arranging this material in the way we do. It is the fact that this approach generalises immediately to the multi-indexed arrays of random variables found in the analysis of variance. In that context generalised cumulants of order two correspond to components of variance, generalised k-statistics of order two to the standard linear combinations of mean squares used to estimate components of variance, and so on. A rich and fruitful theory arises simply by generalising the index sets and the underlying symmetry constraints, and it contains most of the contents of this paper. However it would appear to be quite unmotivated if we went directly to this theory, bypassing the single index case where all the ideas originally developed.

2. Generalising k-statistics

Problems associated with population moments of sample moments occupied many journal pages in the period 1900–1940 but in retrospect, there seems to be little disagreement that the major work on this topic was Fisher (1929). In that paper Fisher gave the definition of *k-statistics* as the symmetric sample functions which give unbiased estimators of the population cumulants, and he also obtained the joint cumulants of the k-statistics of order at most 10 by an ingenious combinatorial method. In this context *sample* refers to a sequence $X_1, X_2, \ldots$ of mutually independent random variables with a common distribution—the *population*—having cumulants $\kappa_1, \kappa_2, \ldots$, and so the k-statistics $k_1(X_1, \ldots, X_n)$,

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

$k_2(X_1,\ldots,X_n),\ldots$ are characterised by symmetry and the property that for $1 \leqslant m \leqslant n$

$$\mathbb{E}_{iid}\{k_m(X_1,\ldots,X_n)\} = \kappa_m, \tag{2.1}$$

where $\mathbb{E}_{iid}$ denotes expectation under the stated assumptions on the $\{X_i\}$.

Similar problems were studied in the context of simple random sampling from finite populations by a number of authors including Neyman (1923, 1925), who first derived an expression for the variance of the variance under these assumptions. The big breakthrough in this field came when Tukey (1950) introduced *polykays*, also called *l-statistics* by Kendall and Stuart (1969). The polykays of a given order m were labelled by *partitions* $\lambda = (\lambda_1\lambda_2,\ldots)$, $\lambda_1 \geqslant \lambda_2 \geqslant \ldots$, of the number m and reduced to the sample k-statistics when $\lambda = (m)$. These statistics were "inherited on the average" under the simple random sampling hypothesis, i.e. when $\sum_j \lambda_j = m \leqslant n$

$$\mathbb{E}_{srs}\{k_\lambda(X_1,\ldots,X_n)\} = \tilde{k}_\lambda, \tag{2.2}$$

where $\tilde{k}_\lambda$ is the corresponding population statistic and $\mathbb{E}_{srs}$ denotes an average over all $\binom{N}{n}$ samples from the population. It turns out that we also have

$$\mathbb{E}_{iid}\{k_\lambda(X_1,\ldots,X_n)\} = \kappa_{\lambda_1}\kappa_{\lambda_2},\ldots, \tag{2.3}$$

and these properties, together with symmetry, can also serve as a definition of polykays.

Both k-statistics and polykays (which we will call *generalised k-statistics*) may be viewed simply as symmetric functions of a single set (x_i) of indeterminates, and each has an obvious generalisation to a bivariate, trivariate,... form involving indeterminates (x_i), (y_i), (z_i),.... For example,

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i - x.)(y_i - y.) \tag{2.4}$$

is readily recognised as the "cross k-statistic" unbiassedly estimating the "cross-cumulant" $\kappa_{1,1}$ (covariance) of (X_i, Y_i) under i.i.d. sampling.

However, whilst we have no trouble recognising (say) the trivariate k-statistic of order $m = 3$ as

$$\frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}(x_i - x.)(y_i - y.)(z_i - z.) \tag{2.5}$$

where $x. = n^{-1}\sum_1^n x_i$, and similarly for $y.$ and $z.$, a little thought shows that the trivariate analogue of the polykay $k_{(21)}$ takes three forms corresponding to the partitions 1|23, 13|2 and 12|3 of $\{1,2,3\}$, namely

(2.6)

$$\frac{1}{n-1}\sum_{i=1}^{n}x.(y_i - y.)(z_i - z.) - \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}(x_i - x.)(y_i - y.)(z_i - z.)$$

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

and the two similar expressions obtained by pairing x with z and x with y as we have y with z in the first summation. Just how we get these expressions will be explained shortly.

The fact that symmetric functions naturally labelled by partitions of a *number* m might usefully be viewed as being labelled by partitions of the *set* $\mathbf{m} = \{1,\ldots,m\}$ has been noted and exploited very effectively by Doubilet (1972). In our context it is not only *convenient* to do so when studying generalised k-statistics as functions of a single set of indeterminates, it is *necessary* when we wish to discuss (multivariate) generalised k-statistics as functions of a number of sets of indeterminates, the most general case being where there are as many distinct sets of indeterminates as the degree of the statistic. With this in mind we would write (2.4), (2.5) and (2.6) as k_{12}^{xy}, k_{123}^{xyz} and $k_{1|23}^{xyz}$ respectively, whilst the two expressions similar to (2.6) would be written $k_{2|13}^{xyz}$ and $k_{3|12}^{xyz}$. The other generalised k-statistic of order three would, of course, be $k_{1|2|3}^{xyz}$ and the superscript can be suppressed (as is the number $n \geqslant 3$) when no confusion is likely to result. In what follows m is the number of distinct (sets of) indeterminates, later coinciding with the number of distinct sequences of random variables whilst n is the number of indeterminates in each array, later coinciding with the sample size.

The next step in our approach to generalised k-statistics is the removal of the indeterminates, leaving simply their coefficients. This is in essence the tensor approach introduced by Kaplan (1952) although we prefer to use the modern multi-linear form of tensors rather than Kaplan's classical one involving multiple subscripts. As a preliminary we must record some simple facts concerning the action of the symmetric group S_n on the set $\mathbf{n}^{\mathbf{m}}$ of all mappings h: $\mathbf{m} \to \mathbf{n}$. The *kernel* of any such map is the partition $\ker h$ whose blocks are the non-empty inverse images of elements of $\mathbf{n}$ under h, i.e. l_1 and l_2 belong to the same block of $\ker h$ if $h(l_1) = h(l_2)$, $l_1, l_2 \in \mathbf{m}$. The group S_n acts on $\mathbf{n}^{\mathbf{m}}$ as follows: $h^g(l) = h(l^g)$, $l \in \mathbf{n}$, $g \in S_n$, $h \in \mathbf{n}^{\mathbf{m}}$ and for any partition $\pi \in \mathscr{P}(\mathbf{m})$, the set of all partitions of $\mathbf{m}$, we write $\mathscr{O}_\pi = \{h \in \mathbf{n}^{\mathbf{m}}: \ker h = \pi\}$. These subsets constitute the *orbits* of S_n acting on $\mathbf{n}^{\mathbf{m}}$, i.e.

LEMMA 2.1.

(i) $\{\mathscr{O}_\pi: \pi \in \mathscr{P}(\mathbf{m})\}$ *forms a partition of* $\mathbf{n}^{\mathbf{m}}$.

(ii) $\mathscr{O}_\pi$ *is* S_n*-invariant*, $\pi \in \mathscr{P}(\mathbf{m})$.

(iii) S_n *acts transitively on* $\mathscr{O}_\pi$, $\pi \in \mathscr{P}(\mathbf{m})$.

Now let us take the set $\{\delta^1, \delta^2,\ldots,\delta^n\}$ of n-tuples which form the standard basis for the vector space of all $n \times 1$ column vectors, i.e. $(\delta^i)_j = \delta_j^i = 1$ if $i = j$

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

and 0 otherwise, $1 \leqslant i, j \leqslant n$. Then for a map h: $\mathbf{m} \to \mathbf{n}$ we can define the tensors ($mn \times 1$ vectors)

$$\bar{h} = \delta^{h(1)} \otimes \cdots \otimes \delta^{h(m)}.$$

In terms of these we can define three families of tensors indexed by partitions π of the set $\mathbf{m}$, $m \leqslant n$, namely the *association* tensors, which give the *augmented monomial symmetric functions*

$$A_\pi = \sum_{h:\ \ker h = \pi} \bar{h}, \tag{2.7}$$

and the *relationship* tensors, which give the *power sum symmetric functions*

$$R_\pi = \sum_{h:\ \ker h \geqslant \pi} \bar{h}. \tag{2.8}$$

An immediate consequence of these definitions is the following pair of identities:

$$R_\pi = \sum_{\sigma \in \mathscr{P}(\mathbf{m})} \zeta(\pi, \sigma) A_\sigma, \qquad A_\pi = \sum_{\sigma \in \mathscr{P}(\mathbf{m})} \mu(\pi, \sigma) R_\sigma$$

where ζ and μ are the zeta and Möbius function, respectively, of the lattice $\mathscr{P}(\mathbf{m})$, and the familiar identities relating the augmented monomial and power sum symmetric functions follow at once. We can now define the *generalised k-statistic* tensors

$$F_\pi = \sum_{\sigma \in \mathscr{P}(\mathbf{m})} \mu(\sigma, \pi) \frac{1}{(n)_\sigma} A_\sigma, \tag{2.9}$$

where $(n)_\sigma = n(n-1)\cdots(n-b(\sigma)+1)$, with $b(\sigma)$ the number of blocks of $\sigma \in \mathscr{P}(\mathbf{m})$. Identities relating F_π and R_π are discussed in Proposition (2.1) below.

Let us see how (2.7) generalises (apart from the factor $(n)_\pi^{-1}$) Tukey's symmetric means and (2.9) his polykays, whilst (2.8) corresponds to the so-called unrestricted symmetric sums. To do so we write $x = (x_1, \ldots, x_n)'$ as a column vector of indeterminates, and similarly for $y, z, \ldots$, and use the inner product $[u|v] = \sum_\alpha u_\alpha v_\alpha$ where u and v are tensor products of one or more of $x, y, z, \ldots$.

EXAMPLES. (a) $m = 2$.

$$A_{12} = \sum_{h:\ \ker h = 12} \delta^{h(1)} \otimes \delta^{h(2)}$$

and so

$$\left[A_{12} | x \otimes y \right] = \sum_{h:\ \ker h = 12} \left[\delta^{h(1)} \otimes \delta^{h(2)} | x \otimes y \right] = \sum_{h:\ \ker h = 12} x_{h(1)} y_{h(2)} = \sum_{i=1}^{n} x_i y_i.$$

Similarly

$$\left[A_{1|2} | x \otimes y \right] = \sum_{h:\ \ker h = 1|2} x_{h(1)} y_{h(2)} = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} x_i y_j$$

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

whilst $[R_{12}|x \otimes y] = [A_{12}|x \otimes y]$ and

$$\begin{aligned}\left[R_{1|2}|x \otimes y \right] &= \sum_{h:\, \ker h \geqslant 1|2} x_{h(1)} y_{h(2)} \\ &= \sum_{h:\, \ker h = 1|2} x_{h(1)} y_{h(2)} + \sum_{h:\, \ker h = 12} x_{h(1)} y_{h(2)} \\ &= \left[A_{1|2}|x \otimes y \right] + \left[A_{12}|x \otimes y \right].\end{aligned}$$

Finally,

$$\begin{aligned}\left[F_{12}|x \otimes y \right] &= \frac{1}{(n)_{12}} \left[A_{12}|x \otimes y \right] - \frac{1}{(n)_{1|2}} \left[A_{1|2}|x \otimes y \right] \\ &= \frac{1}{n} \sum_{i=1}^{n} x_i y_i - \frac{1}{n(n-1)} \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} x_i y_j\end{aligned}$$

which is well known to simplify to the second (mixed) k-statistic (2.4). In this expression we used the values

$$\mu(1|2, 12) = -1, \qquad \mu(1|2, 1|2) = \mu(12, 12) = +1$$

of the Möbius function in $\mathscr{P}(2)$.

(b) $m = 3$. We refer to Speed (1983) for a diagram of $\mathscr{P}(3)$ from which the values of the Möbius function μ of $\mathscr{P}(3)$ are easily read. Beginning with the finest partition $\pi = 1|2|3$ we see that

$$A_{1|2|3} = \sum_{h:\, \ker h = 1|2|3} \delta^{h(1)} \otimes \delta^{h(2)} \otimes \delta^{h(3)}$$

whence

$$\begin{aligned}\left[A_{1|2|3}|x \otimes y \otimes z \right] &= \sum_{h:\, \ker h = 1|2|3} \left[\delta^{h(1)} \otimes \delta^{h(2)} \otimes \delta^{h(3)}|x \otimes y \otimes z \right] \\ &= \sum_{h:\, \ker h = 1|2|3} x_{h(1)} y_{h(2)} z_{h(3)} \\ &= \sum_{\substack{i=1 \\ i,j,k,\text{ distinct}}}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} x_i y_j z_k\end{aligned}$$

the familiar sum Tukey (1950) writes as $\Sigma^{\neq} x_i y_j z_k$. Similarly we find that

$$\left[A_{1|23}|x \otimes y \otimes z \right] = \sum_{h:\, \ker h = 1|23} x_{h(1)} y_{h(2)} z_{h(3)} = \Sigma^{\neq} x_i y_j z_j$$

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

continuing to use Tukey's notation Σ^*. Since $\mu(1|2|3, 1|23) = -1$ we then have

$$\begin{aligned}\left[F_{1|23}|x \otimes y \otimes z \right] &= \frac{1}{(n)_{1|23}} \left[A_{1|23}|x \otimes y \otimes z \right] - \frac{1}{(n)_{1|2|3}} \left[A_{1|2|3}|x \otimes y \otimes z \right] \\ &= \frac{1}{n(n-1)} \Sigma^* x_i y_j z_j - \frac{1}{n(n-1)(n-2)} \Sigma^* x_i y_j z_k\end{aligned}$$

which may be rearranged into (2.6).

Finally, $[A_{123}|x \otimes y \otimes z] = \Sigma x_i y_i z_i$ and, since $\mu(1|23, 123) = \mu(13|2, 123) = \mu(3|12, 123) = -1$ and $\mu(1|2|3, 123) = 2$,

$$\begin{aligned}\left[F_{123}|x \otimes y \otimes z \right] = \frac{1}{n} \Sigma x_i y_i z_i - \frac{1}{n(n-1)} \Sigma^* x_i y_j z_j - 2 \text{ similar terms} \\ + \frac{2}{n(n-1)(n-2)} \Sigma^* x_i y_j z_k.\end{aligned}$$

This expression simplifies to (2.5), as is well known, but no general results currently exist concerning such simplifications, cf. the remark following proposition (2.1) below.

One of the important aspects of any discussion of symmetric functions such as ours is the ease with which we can pass from one form to the other. Indeed this is the feature of Doubilet's (1972) approach via partition lattices which is most compelling: the entries in the *transition matrices* are easily computed functions of n and the Möbius function of $\mathscr{P}(\mathbf{m})$. Write $M(A, F)$ for the matrix over $\mathscr{P}(\mathbf{m})$ defined by $A_\pi = \Sigma_\rho M(A, F)_{\pi,\rho} F_\rho$, the sum being over $\mathscr{P}(\mathbf{m})$, with similar definitions for $M(A, R)$, etc. In (iii) below we use the notation $n^\sigma := n^{b(\sigma)}$, $\sigma \in \mathscr{P}(\mathbf{m})$, and note the interrelations between these ordinary and the descending powers:

$$n^\sigma = \sum_\tau \zeta(\sigma, \tau)(n)_\tau, \qquad (n)_\sigma = \sum_\tau \mu(\sigma, \tau) n^\tau$$

the sums being over $\tau \in \mathscr{P}(\mathbf{m})$. Furthermore, the notation $\pi \vee \rho$ denotes the least upper bound or sup of the partitions π and ρ taken in $\mathscr{P}(\mathbf{m})$, i.e. the finest partition of $\mathbf{m}$ which is coarser than both π and ρ. The result embodied in the second half of (iii) can be found in Carney (1970, Theorem 1).

PROPOSITION 2.1.

(i) $M(A, R) = (\mu(\pi, \rho)), \qquad M(R, A) = (\zeta(\pi, \rho)).$

(ii) $M(F, A) = \left((n)_\rho^{-1} \mu(\rho, \pi)\right), \qquad M(A, F) = \left((n)_\pi \zeta(\rho, \pi)\right).$

(iii) $M(F, R) = \left(\sum_\sigma (n)_\sigma^{-1} \mu(\sigma, \pi) \mu(\sigma, \rho) \right), \qquad M(R, F) = (n^{\pi \vee \rho}).$

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

PROOF. (i) and (ii) are immediate consequences of the definitions (2.7), (2.8) and (2.9) and Möbius inversion over $\mathscr{P}(\mathbf{m})$. The first matrix in (iii) is just the unsimplified expression resulting from multiplying the earlier expressions $M(F, A)M(A, R)$ whilst the second matrix is derived as follows:

$$R_\pi = \sum_\sigma \zeta(\pi, \sigma) A_\sigma = \sum_\sigma \zeta(\pi, \sigma) \sum_\rho (n)_\sigma \zeta(\rho, \sigma) F_\rho$$

$$= \sum_\rho \left\{ \sum_\sigma \zeta(\pi \vee \rho, \sigma)(n)_\sigma \right\} F_\rho = \sum_\rho n^{\pi \vee \rho} F_\rho .$$

No simplification of $M(F, R)$ seems to be known, although Kaplan (1952, equation 2) gives formulae for $F_{\mathbf{m}}$ in terms of R_σ, $\sigma \in \mathscr{P}(\mathbf{m})$, for $m = 2, 3$ and 4. It is clear from these formulae that the coefficient $M(F, R)_{\mathbf{m},\sigma}$ has the form $\mu(\sigma, \mathbf{m})Q_\sigma(n)$ where $Q_\sigma(n)$ is the ratio of two monic polynomials in n whose degrees differ by $b(\sigma)$, but we have been unable to prove this more generally. Finding a manageable closed form for it seems to be a challenging unsolved problem. Kendall and Sutart (1969, equation (12.8), pp. 280–281) give expansions in a less general form up to $m = 8$, and the above-mentioned pattern persists although in their case the numerical coefficients include a further combinatorial multiplier.

3. Products of generalised k-statistics

Most of the useful results concerning generalised k-statistics stem from the fact that the tensor product of two or more of them can be expanded as linear combinations of higher-order ones. Furthermore, there is a good deal of knowledge available concerning the coefficients in such expansions, much due to Dwyer and Tracy (1968) who systematised and extended earlier results of Fisher, Tukey, Wishart and others. We will restate the more important of their conclusions within our framework, abbreviating F_π by (π). Kinney (1976) gives some results which take a different form from those below and in the works just mentioned, but we will have no occasion to refer to them.

In the result which follows we suppose that $\mathbf{m}_1, \ldots, \mathbf{m}_r$ are disjoint sets containing $m_1, \ldots, m_r$ elements, respectively, and that $\mathbf{m} = \mathbf{m}_1 \cup \cdots \cup \mathbf{m}_r$. In examples it is convenient to let $\mathbf{m}_1 = \{1, 2, \ldots, m_1\}$, $\mathbf{m}_2 = \{m_1 + 1, \ldots, m_1 + m_2\}$, etc.

PROPOSITION 3.1. *For $\sigma_1 \in \mathscr{P}(\mathbf{m}_1), \ldots, \sigma_r \in \mathscr{P}(\mathbf{m}_r)$, we have*

$$(\sigma_1) \otimes \cdots \otimes (\sigma_r) = \sum_\sigma c(\sigma; \sigma_1, \ldots, \sigma_r)(\sigma)$$

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

where the sum is over all $\sigma \in \mathscr{P}(\mathbf{m})$ *and*

$$c(\sigma; \sigma_1, \ldots, \sigma_r) = \sum_{\rho \geqslant \sigma} (n)_\rho \prod_{i=1}^{r} \frac{\mu(\rho \cap \mathbf{m}_i, \sigma_i)}{(n)_{\rho \cap \mathbf{m}_i}}. \tag{3.1}$$

PROOF. We use Proposition 2.1 (iii) and obtain

$$(\sigma_1) \otimes \cdots \otimes (\sigma_r) = \sum_{\pi_1} \sum_{\tau_1} \cdots \sum_{\pi_r} \sum_{\tau_r} \frac{\mu(\pi_1, \sigma_1)\mu(\pi_1, \tau_1)}{(n)_{\pi_1}}$$

$$\cdots \frac{\mu(\pi_r, \sigma_r)\mu(\pi_r, \tau_r)}{(n)_{\pi_r}} R_{\tau_1} \otimes \cdots \otimes R_{\tau_r}.$$

Then we observe that $R_{\tau_1} \otimes \cdots \otimes R_{\tau_r} = R_{\tau_1|\cdots|\tau_r}$, and write this last expression in terms of (σ), $\sigma \in \mathscr{P}(\mathbf{m})$ *without* simplifying it:

$$R_{\tau_1|\cdots|\tau_r} = \sum_{\rho} \sum_{\sigma} \zeta(\tau_1|\cdots|\tau_r, \rho)\zeta(\sigma, \rho)(n)_\rho(\sigma).$$

Now $\zeta(\tau_1| \cdots |\tau_r, \rho) = \zeta(\tau_1, \rho \cap \mathbf{m}_1) \cdots \zeta(\tau_r, \rho \cap \mathbf{m}_r)$ and if we put this into the preceding equation and substitute the result in our expression for $(\sigma_1) \otimes \cdots \otimes (\sigma_r)$, we find that $\tau_1, \ldots, \tau_r$ can be summed out forcing $\pi_1 = \rho \cap \mathbf{m}_1, \ldots, \pi_r = \rho \cap \mathbf{m}_r$.

Since $|\mathscr{P}(\mathbf{m})|$ grows very rapidly with m, it would be hoped that most of the coefficients $c(\sigma; \sigma_1, \ldots, \sigma_r)$ are zero, or at least there is a simplification possible in the formulae for computing c. Our next three propositions are the major results along these lines; a few others can be found in Tracy (1968) to which we also refer for the proofs of Propositions 3.2 and 3.3.

PROPOSITION 3.2. *Suppose that* $\sigma \in \mathscr{P}(\mathbf{m})$ *is a partition having a block* σ^1 *which is a proper subset of a block* σ_1^1 *of* σ_1. *Then* $c(\sigma; \sigma_1, \ldots, \sigma_r) = 0$.

PROPOSITION 3.3. *Suppose that* $\sigma \in \mathscr{P}(\mathbf{m})$ *is a partition having a block* σ^1 *which coincides with a block* σ_1^1 *of* σ_1. *Then* $c(\sigma; \sigma_1, \ldots, \sigma_r) = c(\sigma^*; \sigma_1^*, \ldots, \sigma_r)$ *where* σ^* *and* σ_1^* *denote the partitions* σ *and* σ_1 *without their first block.*

With these propositions it is easy to calculate the coefficients for important special cases. For example, if $\sigma_1 = \mathbf{m}_1$, $\sigma_2 = \mathbf{m}_2$, then $c(\mathbf{m}_1|\mathbf{m}_2; \mathbf{m}_1, \mathbf{m}_2) = 1$. All other $c(\sigma; \mathbf{m}_1, \mathbf{m}_2)$ vanish unless each block of σ intersects both $\mathbf{m}_1$ and $\mathbf{m}_2$, and in this case

$$c(\sigma; \mathbf{m}_1, \mathbf{m}_2) = \sum_{k=1}^{b(\sigma)} \frac{[(k-1)!]^2}{(n)_k} S_{b(\sigma),k} \tag{3.2}$$

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

where $S_{b,k}$ is the Stirling number of the second kind. To see (3.2) we simply note that if each of the blocks of σ meets both $\mathbf{m}_1$ and $\mathbf{m}_2$, then this property holds for all $\rho \geqslant \sigma$, and it is then a consequence of known facts concerning $\mathscr{P}(\mathbf{b})$; see for example Aigner (1979, p. 70). Equation (3.2) is due to Fisher (1929) where it can be found expressed in terms of differences of zero.

The results in Table 1 will be useful later. They are readily checked using the propositions just proved. Note that our usage of (σ) is almost but not quite identical to that of Tukey (1950) and later writers.

Expansions of some products of generalised k-statistics

$$
\begin{aligned}
(1) \otimes (2) &= (1|2) + \frac{1}{n}(12) \\
(1) \otimes (23) &= (1|23) + \frac{1}{n}(123) \\
(1) \otimes (2|3) &= (1|2|3) + \frac{1}{n}[(12|3) + (13|2)] \\
(12) \otimes (34) &= (12|34) + \frac{1}{n}(1234) + \frac{1}{n-1}[(13|24) + (14|23)] \\
(12) \otimes (3|4) &= (12|3|4) + \frac{1}{n}[(123|4) + (124|3)] \\
&\quad - \frac{1}{n(n-1)}[(13|24) + (14|23)] \\
(1|2) \otimes (3|4) &= (1|2|3|4) + \frac{1}{n(n-1)}[(13|24) + (14|23)] \\
&\quad + \frac{1}{n}[(13|2|4) + (14|2|3) + (23|1|4) + (24|1|3)]
\end{aligned}
$$

TABLE 1

The final proposition in this section is not only a useful rule for computing the $c(\sigma; \sigma_1, \ldots, \sigma_r)$, but it plays an important role in our proof of Fisher's theorem given in the next section. It first appeared in James ((1958), Lemma 6.1) in that context, but stated only for the case $\sigma_1 = \mathbf{m}_1, \ldots, \sigma_r = \mathbf{m}_r$ (i.e. for k-statistics and not polykays), whilst the present more general formulation is Rule 6 of Tracy (1968). Our proof is essentially that of Tracy (1968).

DEFINITION 3.1. A partition $\sigma = \sigma^1| \cdots |\sigma^b \in \mathscr{P}(\mathbf{m})$ is said to be *blocked with respect to the partition* $\alpha = \alpha^1| \cdots |\alpha^k \in \mathscr{P}(\mathbf{r})$ *over* $B^h = \cup\{\mathbf{m}_i\colon i \in \alpha^h\}$, $h = 1, \ldots, k$, if there exists a partition $\beta = \beta^1| \cdots |\beta^k \in \mathscr{P}(\mathbf{b}(\sigma))$ having the same number of blocks as α such that for all $h = 1, \ldots, k$,

$$\sigma^j \subseteq B^h \quad \text{whenever } j \in \beta^h. \tag{3.3}$$

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

EXAMPLE. Suppose that $\mathbf{m}_1 = \{1,2\}$, $\mathbf{m}_2 = \{3,4\}$, $\mathbf{m}_3 = \{5,6\}$ and $\mathbf{m}_4 = \{7,8\}$. Then the partition $\sigma = 13|24|57|68$ is blocked with respect to the partition $\alpha = 12|34$ over $B^1 = \{1,2\} \cup \{3,4\}$ and $B^2 = \{5,6\} \cup \{7,8\}$ since $\beta = 12|34$ satisfies (3.3).

PROPOSITION 3.4. *If σ is blocked over $\{B^h\}$, then*

$$c(\sigma; \sigma_1, \ldots, \sigma_r) = \prod_h c\big(\sigma \cap B^h; \sigma_1 \cap B^h, \ldots, \sigma_r \cap B^h\big).$$

PROOF. We will only consider partitions $\alpha = \alpha^1|\alpha^2$ of $\mathbf{r}$ into two blocks; the more general case follows by an obvious induction. The proof begins with the observation that $\sigma = \sigma \cap B^1|\sigma \cap B^2$ and hence $\rho \geqslant \sigma$ if and only if $\rho \cap B^1 \geqslant \sigma \cap B^1$ and $\rho \cap B^2 \geqslant \sigma \cap B^2$. Thus we may break up the sum in Proposition 3.1 over $\rho \geqslant \sigma$ defining $c(\sigma; \sigma_1, \ldots, \sigma_r)$ into three sums: Over $\rho_1 \geqslant \sigma \cap B^1$, over $\rho_2 \geqslant \sigma \cap B^2$, and over ρ with $\rho \cap B^1 = \rho_1$, $\rho \cap B = \rho_2$. The expression inside the summation in (3.1), which we call $f(\rho; \sigma_1, \ldots, \sigma_r)$, almost splits over B^1, B^2. With the above observations we readily find that

$$\sum_{\rho \geqslant \sigma} f(\rho; \sigma_1, \ldots, \sigma_r) = \sum_{\rho_1 \geqslant \sigma \cap B^1} \sum_{\rho_2 \geqslant \sigma \cap B^2} f\big(\rho_1; \sigma_1 \cap B^1, \ldots, \sigma_r \cap B^1\big) \times f\big(\rho_2; \sigma_1 \cap B^2, \ldots, \sigma_r \cap B^2\big) S \tag{3.4}$$

where $S = S(\rho_1, \rho_2)$ is given by

$$S(\rho_1, \rho_2) = \frac{1}{(n)_{\rho_1}(n)_{\rho_2}} \sum_{\rho} (n)_{\rho}, \tag{3.5}$$

the sum being over all ρ with $\rho \cap B^1 = \rho_1$, $\rho \cap B^2 = \rho_2$.

If $b(\rho_1) = b_1$ and $b(\rho_2) = b_2$ then there are $\binom{b_1}{r}\binom{b_2}{r} r!$ partitions ρ with $b_1 + b_2 - r$ blocks satisfying $\rho \cap B^1 = \rho_1$, $\rho \cap B^2 = \rho_2$, and upon substituting these into (3.5) we find that $S(\rho_1, \rho_2) = 1$ follows from the Vandermonde identity in the form

$$\sum_r (n)_{b_1 + b_2 - r} \frac{(b_1)_r (b_2)_r}{r!} = (n)_{b_1}(n)_{b_2}.$$

Putting this back into (3.4) we find that the proposition is proved.

4. Generalised moments and cumulants

In this section we will apply the combinatorial reesults of Section 3 to arrays $X = (X_i: i \in \mathbf{n})$ of random vectors X_i whose joint moments satisfy certain symmetry constraints. As we will be extending some of the results derived below

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

to multi-indexed random vectors it is convenient to begin with a definition somewhat more general than is needed for our present purposes.

Let G be a group acting on an index set $\mathbf{I}$ and suppose that $X = (X_i: i \in \mathbf{I})$ is an array of random vectors $X_i = (X_i(1), \ldots, X_i(m))'$ indexed by $\mathbf{I}$.

DEFINITION 4.1. The array X is said to have *G-invariant moments of order* $q \geqslant m$ if for all $t \leqslant q$, $i_1, \ldots, i_t \in \mathbf{I}$, $l_1, \ldots, l_t \in \mathbf{m}$, and $g \in G$ we have

$$\mathbb{E}\{X_{i_1}(l_1) \cdots X_{i_t}(l_t)\} = \mathbb{E}\{X_{i_1^g}(l_1) \cdots X_{i_t^g}(l_t)\}. \tag{4.1}$$

Here i^g is the image of $i \in \mathbf{I}$ under $g \in G$.

EXAMPLE 4.1. If $m = 2$, $\mathbf{I} = \mathbf{n}$ and $G = S_n$, the full symmetric group on n symbols, and if for convenience we put $X_i(1) = X_i$, $X_i(2) = Y_i$, $i \in \mathbf{N}$, then the array $((X_i, Y_i)': i \in \mathbf{n})$ has S_n-invariant moments of order at most 2 if and only if for all $i, j \in \mathbf{n}$ and $g \in S_n$, $\mathbb{E}\{X_i\} = \mathbb{E}\{X_{i^g}\}$, $\mathbb{E}\{Y_i\} = \mathbb{E}\{Y_{i^g}\}$, $\mathbb{E}\{X_iX_j\} = \mathbb{E}\{X_{i^g}X_{j^g}\}$, $\mathbb{E}\{Y_iY_j\} = \mathbb{E}\{Y_{i^g}Y_{j^g}\}$ and $\mathbb{E}\{X_iY_j\} = \mathbb{E}\{X_{i^g}Y_{j^g}\}$.

Our concern in the present section is solely with $\mathbf{I} = \mathbf{n}$ and $G = S_n$, and such arrays of random vectors whose joint moments satisfy Definition 4.1 will be described as having *symmetric moments*. There are two important ways in which they may arise: (i) when the component vectors X_i are mutually independent and identically distributed with some joint distribution F on the Borel subsets of $\mathbb{R}^m$; and (ii) when there exists an $N \geqslant n$, an array $X = (\tilde{X}_I: I \in \mathbf{N})$ of vectors $\tilde{X}_I = (\tilde{X}_I(1), \ldots, \tilde{X}_I(m))'$ of real numbers called the *population*, such that the n *labels* of $X = (X_i: i \in \mathbf{N})$ are chosen with m-symmetric inclusion probabilities, i.e. for every subset $i_1, \ldots, i_m$ of $\mathbf{n}$ consisting of m *distinct* labels and every $\{I_1, \ldots, I_m\} \subseteq \mathbf{N}$ we have

$$\operatorname{pr}(X_{i_1} = \tilde{X}_{I_1}, \ldots, X_{i_m} = \tilde{X}_{I_m}) = \begin{cases} \dfrac{1}{(N)_m} & \text{if } |\{I_1, \ldots, I_m\}| = m; \\ 0 & \text{otherwise.} \end{cases}$$

We describe this second situation as simple random sampling without replacement (abbreviated *srswor*), although it is in fact somewhat more general.

Now let us suppose that $X = (X_i: i \in \mathbf{n})$ is a random array of m-vectors whose moments of order m are symmetric. Under this assumption the following definition makes sense.

DEFINITION 4.2. The *generalised mth moment* of X corresponding to $\sigma \in \mathscr{P}(\mathbf{m})$ is defined by $\gamma_\sigma = \mathbb{E}\{X_{h(1)}(1) \cdots X_{h(m)}(m)\}$ where $h: \mathbf{m} \to \mathbf{n}$ is any map with $\ker h = \sigma$. The *generalised mth cumulant* of X corresponding to $\sigma \in \mathscr{P}(\mathbf{m})$ is defined by $f_\sigma = \sum_\rho \mu(\rho, \sigma)\gamma_\rho$.

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

COROLLARY. *If the components of X are iid, then* $\gamma_\sigma = \mu'_\sigma$ *and* $f_\sigma = \kappa_\sigma$, *where* μ'_σ *and* κ_σ *are defined as in Speed* (1983).

PROOF. The result for γ_σ is easy for if $\sigma = \sigma^1 | \cdots | \sigma^b$, then the sets $\{\{(X_{h(l)}(l): l \in \sigma^a)\}: a = 1,\ldots,b\}$ of random variables are mutually independent. Then the result for f_σ follows from Speed (1983), equation (3.2).

This corollary justifies our use of the terms generalised moments and cumulants. For the remainder of this section it is convenient to view our array $X = (X_i: i \in \mathbf{n})$ of m-tuples as an array $(X(1),\ldots,X(m))$ of m n-tuples, and we let $X = X(1) \otimes \cdots \otimes X(m)$.

PROPOSITION 4.1. *If X is an array with symmetric moments of order m then* $\mathbb{E}\{X\} = \sum_\sigma \gamma_\sigma A_\sigma = \sum_\rho f_\rho R_\rho$ *where the sums are over* $\mathscr{P}(\mathbf{m})$.

PROOF. This is an immediate consequence of the definition of the generalised moments $\{\gamma_\sigma\}$, the tensors $\{A_\sigma\}$, and of Lemma 2.1. Indeed if we view functions on $\mathbf{n}^\mathbf{m}$ as $mn \times 1$ vectors indexed lexicographically, then it is not hard to see that A_σ is the characteristic or indicator function of the orbit $\mathcal{O}_\sigma$, $\sigma \in \mathscr{P}(\mathbf{m})$. The corresponding relations for generalised cumulants follow from their definitions and those of the tensors $\{R_\rho\}$.

Unbiased estimation of generalised cumulants

We have already introduced the abbreviation (σ) for F_σ; the corresponding one for A_σ is $\langle\sigma\rangle = (n)_\sigma^{-1} A_\sigma$; see Tukey (1950). Let us also use the abbreviation $a_\sigma = [A_\sigma | X] = (n)_\sigma [\langle\sigma\rangle | X]$ and $k_\sigma = [(\sigma)|X]$ for the corresponding symmetric functions of the elements of our random array X.

PROPOSITION 4.2. *If the moments of order m of X are symmetric, then for all* $\sigma \in \mathscr{P}(\mathbf{m})$ *we have*

$$\mathbb{E}\left\{(n)_\sigma^{-1} a_\sigma\right\} = \gamma_\sigma \quad \text{and} \quad \mathbb{E}\{k_\sigma\} = f_\sigma.$$

PROOF. For any $h: \mathbf{m} \to \mathbf{n}$ with $\ker h = \sigma$ we have

$$[\bar{h}|X] = \left[\delta^{h(1)} \otimes \cdots \otimes \delta^{h(m)} | X(1) \otimes \cdots \otimes X(m)\right] = X_{h(1)}(1) \cdots X_{h(m)}(m)$$

whence $\mathbb{E}[\bar{h}|X] = \gamma_\sigma$. This extends by linearity to give the first assertion of the proposition and the second follows from the definitions.

Thus we have proved that the generalised k-statistics give unbiased estimates of generalised cumulants. Now it is well known that, in a sense to be made precise,

"symmetric" functions have minimum variance amongst unbiased estimators of their expectation. We are unable to provide a reference to this result in the form we need it, although other forms are very familiar, see e.g. Pitman (1939), and so we include it as a lemma.

LEMMA 4.1. *If a group G acting on a finite set I has orbits $I^1, I^2, \ldots$ say, and $Y = (Y_i)_{i \in I}$ is a linearly independent array of real random variables indexed by I with G-invariant moments of order 2, then the minimum variance unbiased estimator of $\mathbb{E}\{\Sigma_{i \in I} c_i Y_i\}$ is $\Sigma_p c_\cdot^p \Sigma_{i \in I^p} Y_i$, where $c_\cdot^p = |I^p|^{-1} \Sigma_{i \in I^p} c_i$, $p = 1, 2, \ldots$.*

PROOF. It follows from the G-invariance of $\mathbb{E}\{Y\}$ that $\Sigma_{i \in I} c_i Y_i$ and $\Sigma_p c_\cdot^p \Sigma_{i \in I^p} Y_i$ have the same expectation, and by our linear independence assumption the latter is the unique linear combination of the (Y_i) with this property whose coefficients are constant over the orbits of G. The result will be proved once we show that $\{\Sigma_{i \in I} c_i Y_i\} \geqslant \mathrm{Var}\{\Sigma_p c_\cdot^p \Sigma_{i \in I^p} Y_i\}$, and this will follow from the fact that $\Sigma_p \Sigma_{i \in I^p} (c_i - c_\cdot^p) Y_i$ is uncorrelated with *every* contrast $\Sigma_p d^p \Sigma_{i \in I^p} Y_i$ whose coefficients are constant over the orbits of G. To see this we compute

$$\mathrm{cov}\left\{ \sum_p \sum_{i \in I^p} (c_i - c_\cdot) Y_i, \sum_q d^q \sum_{j \in I^q} Y_i \right\} = \sum_p \sum_q \sum_{i \in I^p} \sum_{j \in I^q} (c_i - c_\cdot^p) d^q \gamma_{a(i,j)}$$

where $\gamma_{a(i,j)}$ is the covariance between $i \in I^p$ and $j \in I^q$. Now the subsets I^q are orbits of G on I and so for every $i \in I^p$ the sum $\Sigma_{j \in I^q} \gamma_{a(i,j)}$ has the same value. But then we can sum over $i \in I^p$ and obtain zero.

The application of this result to our situation gives the following proposition. It is related to but not derivable from results in Halmos (1946) for iid random variables.

PROPOSITION 4.3. *If X is an array whose moments of order $2m$ are symmetric, then $(n)_\sigma^{-1} a_\sigma$ and k_σ have minimum variance amongst all polynomials of degree m in $\{X_i(l): i \in \mathbf{n}, l \in \mathbf{m}\}$ which are unbiased estimators of γ_σ and f_σ, respectively.*

EXAMPLE 4.1 (continued). If $m = 2$ and $\sigma = 12$, then it follows that $k_{12} = (n-1)^{-1} \Sigma_{i=1}^n \Sigma_{i=1}^n (X_i - X_\cdot)(Y_i - Y_\cdot)$ is the best quadratic unbiased estimator of $f_{12} = \mathbb{E} X_i Y_i - \mathbb{E} X_i Y_j$ $(i \neq j)$ whenever $((X_i, Y_i): i \in \mathbf{n})$ has invariant moments of order 4. When the (X_i, Y_i) are iid this proves that the usual unbiased estimates of covariances (and variances) are best amongst quadratic estimates.

Cumulants or k-statistics in sampling

The familiar difficulty of whether to divide the sum of the squared deviations about the mean of a quantity associated with each member of a finite population of size N by N or $N-1$ arises in this context in the following form: the

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

k-statistics (of order greater than unity) and, more generally, the generalised k-statistics computed for the *whole* of a finite population do *not* coincide with the corresponding cumulants or generalised cumulants of the random variables naturally defined by *srswor*. For example, if $X = (X_1, \ldots, X_n)'$ is an n-tuple of values sampled in the manner described from $\tilde{X} = (\tilde{X}_1, \ldots, \tilde{X}_N)'$, and $\mathbb{E}_{srs}$ denotes the average over all $\binom{N}{n}$ samples, taken with equal probability then for any $i \in \mathbf{n}$

$$\kappa_1 = \mathbb{E}_{srs}\{X_i\} = \frac{1}{N}\sum_{I=1}^{N} \tilde{X}_I = \tilde{X}_{\cdot} = \tilde{k}_1,$$

$$\kappa_2 = \mathrm{Var}_{srs}\{X_i\} = \frac{1}{N}\sum_{I=1}^{N} (\tilde{X}_I - \tilde{X}_{\cdot})^2 \neq \frac{1}{N-1}\sum_{I=1}^{N} (\tilde{X}_I - \tilde{X}_{\cdot})^2 = \tilde{k}_2,$$

where $\tilde{k}_1$ and $\tilde{k}_2$ are the first and second k-statistics computed for the whole population. Classical (generalised) k-statistics provide unbiased estimates of (products of) classical (generalised) cumulants under *iid sampling*, but when we turn to *srswor* from finite populations it is the generalised cumulants of the array, which coincide with the generalised k-statistics evaluated for the whole population array, which seem to be the appropriate population summary statistics, *not* the (generalised) cumulants of the individual random variables.

5. Fisher's theorem

This theorem gives an expression for the joint cumulant under iid sampling of a set of r k-statistics in terms products of cumulants of the common underlying distribution. Our proof is a streamlined form of that of James (1958) except that, with no extra effort, we can formulate it for generalised k-statistics (i.e. polykays). Because the main source for this result is Kendall and Stuart (1969) where the statement and proof are spread over many pages in two chapters of the book, we would like to emphasize the simplicity and accessability of the result given Möbius inversion on $\mathscr{P}(\mathbf{m})$. We need only (i) the definition of joint cumulant; (ii) the definition, equation (2.9), of generalised k-statistic; (iii) Proposition 3.1 on products of generalised k-statistics, and (iv) Proposition 3.4 giving a product rule for the coefficients in the expansions of products of generalised k-statistics. (Of course Propositions 3.2 and 3.3 are invaluable aids when one comes to apply the theorem.) The notation and general framework is that of Section 3 above, except that we write $k(\sigma)$ instead of k_σ. The expression $\mathscr{C}(\cdot, \ldots, \cdot)$, which is used in the statement of the theorem and defined in the proof, is explained in greater detail in Speed (1983).

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

THEOREM 5.1 (Fisher (1929)). *If* $\{X_i: i \in \mathbf{n}\}$ *are iid with generalised cumulants* $\{\kappa(\sigma)\}$ *then*

$$\mathscr{C}\{k(\sigma_1),\ldots,k(\sigma_r)\} = \sum_{\sigma}{}^{*}\, c(\sigma;\sigma_1,\ldots,\sigma_r)\kappa(\sigma)$$

where Σ^* *denotes the sum over all partitions* $\sigma \in \mathscr{P}(\mathbf{m})$ *which are not blocked with respect to any partition of* $\mathbf{r}$.

PROOF (cf. James (1958), Theorem 6.1). By definition

$$\mathscr{C}\{k(\sigma_1),\ldots,k(\sigma_r)\} = \sum_{\alpha}\mu(\alpha,\mathbf{r})\prod_{a=1}^{b(\alpha)}\mathbb{E}\Big\{\prod_{i\in\alpha^a}k(\sigma_i)\Big\},$$

the sum being over all $\alpha \in \mathscr{P}(\mathbf{r})$. If we write $U_i = \otimes_{l\in\mathbf{m}_i} X(l)$, then

$$\prod_{i\in\alpha^a}k(\sigma_i) = \prod_{i\in\alpha^a}\big[\,(\sigma_i)|U_i\big] = \Big[\bigotimes_{i\in\alpha^a}(\sigma_i)\Big|\bigotimes_{i\in\alpha^a}U_i\Big]$$

$$= \sum_{\sigma(a)} c(\sigma(a);\{\sigma_i: i\in\alpha^a\})\Big[\,(\sigma(a))|\bigotimes_{i\in\alpha^a}U_i\Big]$$

this sum being over all $\sigma(a) \in \mathscr{P}(\bigcup\{\mathbf{m}_i: i \in \alpha^a\})$. Thus

$$\prod_{a=1}^{b(\alpha)}\mathbb{E}\Big\{\prod_{i\in\alpha^a}k(\sigma_i)\Big\} = \prod_a\sum_{\sigma(a)}c(\sigma(a))\kappa(\sigma(a))$$

$$= \sum_{\{\sigma(a)\}}\prod_a c(\sigma(a))\kappa(\sigma(a)) = \sum_{\sigma}{}^{\alpha}c(\sigma;\sigma_1,\ldots,\sigma_r)\kappa(\sigma),$$

by Proposition 3.4, where Σ^α denotes the sum over all $\sigma \in \mathscr{P}(\mathbf{m})$ which are blocked with respect to $\alpha \in \mathscr{P}(\mathbf{r})$. The final step in the proof is identical to that of Proposition 4.3 of Speed (1983): we break up the sum Σ^α into a sum over $\beta \leqslant \alpha$ of those terms which are blocked with respect to β but no finer partition, and then find by Möbius inversion that

$$\sum_{\alpha}\mu(\alpha,\mathbf{r})\sum_{\sigma}{}^{\alpha}c(\sigma;\sigma_1,\ldots,\sigma_r)\kappa(\sigma) = \sum_{\sigma}{}^{*}\,c(\sigma;\sigma_1,\ldots,\sigma_r)\kappa(\sigma).$$

REMARK. When $\sigma_i = \mathbf{m}_i$, $i = 1,\ldots,r$, the coefficients $c(\sigma;\mathbf{m}_1,\ldots,\mathbf{m}_r)$ depend only on the *pattern* of the intersections of the blocks of σ with the $\mathbf{m}_i$, i.e. on the $b \times r$ pattern matrix $M = (m_{ai})$ where

$$m_{ai} = \begin{cases} 1 & \text{if } \sigma^a \cap \mathbf{m}_i \neq \varnothing; \\ 0 & \text{otherwise.}\end{cases}$$

Thus it is possible to compute the coefficients from the pattern; see Fisher (1929), Fisher and Wishart (1931), and Kendall (1940 a, b, c, 1942). It is also not hard to see that this is not possible in our more general formulation.

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

Even when the components $(X_i: i \in \mathbf{n})$ of X are not iid, it is still possible to derive expressions for the cumulants of generalised k-statistics. Indeed this appears to be Tukey's (1950) motivation for introducing the polykay generalisation. For example, the formulae,

$$(*) \qquad (12) \otimes (34) = \frac{1}{n}(1234) + (12|34) + \frac{1}{n-1}\{(13|24) + (14|23)\}$$

given in Section 3 above quickly leads us to a result apparently first derived by Neyman (1923, 1925). We derive it in stages, starting off with a finite population of size N and four quantities (W_I, X_I, Y_I, Z_I) associated with the unit labelled $I \in \mathbf{N}$. From $(*)$ we can read off

$$\mathbb{E}_{srs}\{k_{12}k_{34}\} = \frac{1}{n}\tilde{k}_{1234} + \tilde{k}_{12|34} + \frac{1}{n-1}\{\tilde{k}_{13|24} + \tilde{k}_{14|23}\}.$$

Similarly

$$\mathbb{E}_{srs}\{k_{12}\}\mathbb{E}_{srs}\{k_{34}\} = \tilde{k}_{12}\tilde{k}_{34} = \frac{1}{N}\tilde{k}_{1234} + \tilde{k}_{12|34} + \frac{1}{N-1}\{\tilde{k}_{13|24} + \tilde{k}_{14|23}\}$$

and so

$$\operatorname{cov}_{srs}(k_{12}, k_{34}) = \left(\frac{1}{n} - \frac{1}{N}\right)\tilde{k}_{1234} + \left(\frac{1}{n-1} - \frac{1}{N-1}\right)\{\tilde{k}_{13|24} + \tilde{k}_{14|23}\}.$$

Putting $w = X$, $Y = Z$ gives us an expression for the covariance of two variances; putting $W = Y$ and $X = Z$ gives us an expression for the variance of a covariance,..., and finally, putting $W = X = Y = Z$ gives us the variance of the sample variance (divisor $n - 1$) in terms of the generalised k-statistics of the finite population. A comparison of this formula with Neyman's will confirm that Tukey's invention of polykays certainly did simplify sampling!

Letting $N \to \infty$ we obtain the "infinite population" result

$$\operatorname{cov}_{srs}(k_{12}, k_{34}) = \frac{1}{n}\tilde{k}_{1234} + \frac{1}{n-1}\{\tilde{k}_{13|24} + \tilde{k}_{14|23}\}.$$

Thus can be compared with the covariance of two covariance under iid sampling, obtainable from Fisher's theorem as

$$\operatorname{cov}_{iid}(k_{12}, k_{34}) = \frac{1}{n}\kappa_{1234} + \frac{1}{n-1}(\kappa_{13|24} + \kappa_{14|23}).$$

6. Generalised k-statistics as reversed martingales

We will now suppose that our array $(X_i: i \in \mathbf{n})$ is a subarray of a larger array $(X_i: i \in \mathbf{N})$, $N \leqslant \infty$, and that $X = (X_i: i \in \mathbf{N})$ has a symmetric *joint distribution*, with finite moments of order m, where symmetry here refers either to the full

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

symmetric group S_N over $\mathbf{N}$ if $N < \infty$, or the restricted infinite symmetric group S_∞ of all finitary permutations of ∞ when $N = \infty$. In the latter case we are dealing with a symmetry which is normally termed (de Finetti) *exchangeability*.

It is well known that under such assumptions the sample average $k_1 = k_n(1) = n^{-1}\sum_1^n X_i$ is a *reversed martingale* with respect to the sequence $\mathscr{F}_n = \sigma(k_p(1): p \geqslant n)$ of σ-fields, and it is also known that $k_n(12) = (n-1)^{-1}\sum_1^n (X_i - X.)^2$ is a reversed martingale with respect to a similar sequence of σ-fields; see Chow and Teicher (1978). We will see that these facts are special cases of a more general one of the same kind, a conclusion which could also be obtained by viewing our functions as U-statistics; see Eagleson and Weber (1978).

For a given $\sigma \in \mathscr{P}(\mathbf{m})$ let us write $a_n(\sigma)$ and $k_n(\sigma)$ for the (augmented) monomial symmetric function and generalised k-statistic, respectively, corresponding to σ and based on $X = \{X_i: i \in \mathbf{n}\}$, and let $\mathscr{F}_n(\sigma) = \sigma(a_p(\sigma): n \leqslant p < N)$, and $\mathscr{G}_n(\sigma) = \vee\{\mathscr{F}_n(\rho): \rho \leqslant \sigma\}$, $1 \leqslant n \leqslant N$.

LEMMA 6.2. *$\{(n)_\rho^{-1} a_n(\rho), \mathscr{G}_n(\sigma): 1 \leqslant n < N\}$ is a reversed martingale if $\rho \leqslant \sigma$.*

PROOF. For h_1, h_2: $\mathbf{m} \to \mathbf{n}$ with $\ker h_1 = \ker h_2 = \rho$, a standard argument using the fact that $h_2 = h_1^g$ for some $g \in S_N$, that $\mathscr{G}_n(\sigma)$ consists of elements invariant under every $g \in S_N$, and the invariance of the joint distribution of X implies

$$\mathbb{E}\{\bar{h}_1(X)|\mathscr{G}_n(\sigma)\} = \mathbb{E}\{\bar{h}_2(X)|\mathscr{G}_n(\sigma)\} \quad \text{a.s.},$$

where we write $\bar{h}(X) = [\bar{h}|X]$ for simplicity. It follows that

$$(n)_\rho^{-1} a_n(\rho) = \mathbb{E}\{(n)_\rho^{-1} a_n(\rho)|\mathscr{G}_n(\sigma)\} = \mathbb{E}\{\bar{h}(X)|\mathscr{G}_n(\sigma)\}$$

where h: $\mathbf{m} \to \mathbf{n}$ is any map with $\ker h = \sigma$.

COROLLARY (to the proof). *$\{k_n(\sigma), \mathscr{G}_n(\sigma): 1 \leqslant n < N\}$ is a reversed martingale.*

PROOF. The argument just given applies when we replace the $\bar{h}$ in the last step by $\sum_\rho \mu(\rho, \sigma)\bar{h}_\rho$ where h_ρ has kernel ρ.

PROPOSITION 6.5. *Suppose that $X = (X_i: i \in \infty)$ has a symmetric joint distribution with finite moments of order m. Then*

(i) *for every $\sigma \in \mathscr{P}(\mathbf{m})$, $k_n(\sigma)$ converges a.s and in L^1 to a limit $k(\sigma)$ as $n \uparrow \infty$. Furthermore, if X has finite moments of order $2m$, then*

(ii) *$k_n(\sigma)$ also converges in L^2 to $k(\sigma)$, and*

(iii) *$k(\sigma)$ is a.s. constant if and only if $f(\sigma|\sigma) = f(\sigma)^2$.*

https://doi.org/10.1017/S1446788700026483 Published online by Cambridge University Press

PROOF. Assertions (i) and (ii) are immediate consequences of known results concerning reversed martingales and submartingales, respectively. The final assertion is a consequence of the easily established fact that $\mathbb{E}\{(k_n(\sigma))^2\} = \mathbb{E}\{k_n(\sigma|\sigma)\} + O(1/n)$ as $n \to \infty$ whence $\lim_n \mathrm{Var}\{k_n(\sigma)\} = \mathrm{Var}\{\lim k_n(\sigma)\} = f(\sigma|\sigma) - f(\sigma)^2$.

COROLLARY. *If the components of* $(X_i: i \in \infty)$ *are iid, then the generalised k-statistics* $k_n(\sigma)$ *converges a.s to the product* κ_σ *of cumulants.*

Acknowledgement

Many thanks are due to the referee for helpful comments and references to relevant material.

References

M. Aigner (1979), *Combinatorial theory* (Springer-Verlag, New York).

E. J. Carney (1968), 'Relationship of generalised polykays to unrestricted sums for balanced complete finite populations', *Ann. Statist.* **39**, 643–656.

Yuan Shih Chow and Henry Teicher (1978), *Probability theory: independence, exchangeability, martingales* (Springer-Verlag: New York, Heidelberg, Berlin).

Peter Doubilet (1972), 'On the foundations of combinatorial theory VII: Symmetric functions through the theory of distribution and occupancy', *Stud. Appl. Math.* **11**, 377–396.

P. S. Dwyer and D. S. Tracy (1964), 'A combinatorial method for products of two polykays with some general formulae', *Ann. Statist.* **35**, 1174–1185.

G. K. Eagleson and N. C. Weber (1978), 'Limit theorems for weakly exchangeable arrays', *Math. Proc. Cambridge Philos. Soc.* **84**, 123–130.

R. A. Fisher (1929), 'Moments and product moments of sampling distributions', *Proc. London Math. Soc.* (2) **30**, 199–238.

R. A. Fisher and J. Wishart (1931), 'The derivation of the pattern formulae of two-way partitions from those of similar patterns', *Proc. London Math. Soc.* (2) **33**, 195–208.

Paul R. Halmos (1946), 'The theory of unbiased estimation', *Ann. Statist.* **17**, 34–43.

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

E. L. Kaplan (1952), 'Tensor notation and the sampling cumulants of *k*-statistics', *Biometrika* **39**, 319–323.

M. G. Kendall (1940a), 'Some properties of *k*-statistics', *Ann. Eugen, London*, **10**, 106–111.

M. G. Kendall (1940b), 'Proof of Fisher's rules for ascertaining the sampling semi-invariants of *k*-statistics', *Ann. Eugen, London*, **10**, 215–222.

M. G. Kendall (1940c), 'The derivation of multivariate sampling formulae from univariate formulae by symbolic operation', *Ann. Eugen, London*, **10**, 392–402.

M. G. Kendall (1942), 'On seminvariant statistics', *Ann. Eugen, London*, **11**, 300–305.

Maurice G. Kendall and Alan Stuart (1969), *The advanced theory of statistics*, Volume 1 (Third Edition, Griffin, London).

John Kinney (1976), 'The k-statistics, polykays and randomised sums', *Sankhyā Ser. A* **38**, 271–286.

J. Splawa Neyman (1925), 'Contributions to the theory of small samples drawn from a finite population', *Biometrika* **17**, 472–479 (reprinted from *La Revue Mensuelle de Statistique* **6**, (1923), 1–29.).

E. J. G. Pitman (1939), 'The estimation of location and scale parameters of a continuous population of any given form', *Biometrika* **30**, 391–421.

T. P. Speed (1983), 'Cumulants and partition lattices', *Austral. J. Statist.* **25**, 378–388.

Derrick S. Tracy (1968), 'Some rules for a combinatorial method for multiple products of generalised k-statistics', *Ann. Statist.* **39**, 983–998.

John W. Tukey (1950), 'Some sampling simplified', *J. Amer. Statist. Assoc.* **45**, 501–519.

Division of Mathematics and Statistics
CSIRO Box 1965 GPO
Canberra, A.C.T. 2601
Australia