Chapter 6
Cumulants and Partition Lattices

Peter McCullagh

This is the first paper to appear in the statistical literature pointing out the importance of the partition lattice in the theory of statistical moments and their close cousins, the cumulants. The paper was first brought to my attention by Susan Wilson, shortly after I had given a talk at Imperial College on the Leonov-Shiryaev result expressed in graph-theoretic terms. Speed's paper was hot off the press, arriving a day or two after I had first become acquainted with the partition lattice from conversations with Oliver Pretzel. Naturally, I read the paper with more than usual attention to detail because I was still unfamiliar with Rota [18], and because it was immediately clear that Möbius inversion on the partition lattice $\mathscr{E}_n$, partially ordered by sub-partition, led to clear proofs and great simplification. It was a short paper packing a big punch, and for me it could not have arrived at a more opportune moment.

The basic notion is a partition σ of the finite set $[n] = \{1,\ldots,n\}$, a collection of disjoint non-empty subsets whose union is $[n]$. Occasionally, the more emphatic term set-partition is used to distinguish a partition of $[n]$ from a partition of the integer n. For example $135|2|4$ and $245|1|3$ are distinct partitions of [5] corresponding to the same partition $3+1+1$ of the integer 5. Altogether, there are two partitions of [2], five partitions of [3], 15 partitions of [4], 52 partitions of [5], and so on. These are the Bell numbers $\#\mathscr{E}_n$, whose exponential generating function is $\exp(e^t - 1)$. The symmetric group acting on $\mathscr{E}_n$ preserves block sizes, and each integer partition is a group orbit. There are two partitions of the integer 2, three partitions of 3, five partitions of 4, seven partitions of 5, and so on.

It turns out that, although set partitions are much larger, the additional structure they provide is essential for at least two purposes that are fundamental in modern probability and statistics. It is the partial order and the lattice property of $\mathscr{E}_n$ that simplifies the description of moments and generalized cumulants in terms of cumulants. This is the subject matter of Speed's paper. At around the same time, from the late 1970s until the mid 1980s, Kingman was developing the theory of partition structures, or partition processes. These were initially described in terms of inte-

P. McCullagh
Department of Statistics, University of Chicago
e-mail: pmcc@galton.uchicago.edu

S. Dudoit (ed.), *Selected Works of Terry Speed*, Selected Works in Probability and Statistics, 277
DOI 10.1007/978-1-4614-1347-9_6,

ger partitions [3, 10], but subsequent workers including Kingman and Aldous have found it simpler and more natural to work with set partitions. In this setting, the simplification comes not from the lattice property, but from the fact that the family $\mathscr{E} = \{\mathscr{E}_1, \mathscr{E}_2, \ldots\}$ of set partitions is a projective system, closed under permutation and deletion of elements. The projective property makes it possible to define a process on $\mathscr{E}$, and the mutual consistency of the Ewens formulae for different n implies an infinitely exchangeable partition process.

In his 1964 paper, Rota pointed out that the inclusion-exclusion principle and much of combinatorics could be unified in the following manner. To any function f defined on a finite partially-ordered set, there corresponds a cumulative function

$$F(\sigma) = \sum_{\tau \le \sigma} f(\tau).$$

The mapping $f \mapsto F$ is linear and invertible with inverse

$$f(\sigma) = \sum_{\tau} m(\tau, \sigma) F(\tau),$$

where the Möbius function is such that $m(\tau, \sigma) = 0$ unless $\tau \le \sigma$. In matrix notation, $F = Lf$, where L is lower-triangular with inverse M. The Möbius function for the Boolean lattice (of sets, subsets and complements) is $(-1)^{\#\sigma - \#\tau}$, giving rise to the familiar inclusion-exclusion rule. For the partition lattice, the Möbius function relative to the single-block partition is $m(\tau, \{[n]\}) = (-1)^{\#\tau - 1}(\#\tau - 1)!$, where $\#\tau$ is the number of blocks. More generally, $m(\tau, \sigma) = \prod_{b \in \sigma} m(\tau[b], b)$ for $\tau \le \sigma$, where $\tau[b]$ is the restriction of τ to the subset b.

Although they have the same etymology, the word 'cumulative' in this context is unrelated semantically to 'cumulant', and in a certain sense, the two meanings are exact opposites: cumulants are to moments as f is to F, not vice-versa.

Speed's paper is concerned with *multiplicative* functions on the partition lattice. To understand what this means, it is helpful to frame the discussion in terms of random variables $X^1, X^2, \ldots, X^n$, indexed by $[n]$. The joint moment function μ associates with each subset $b \subset [n]$ the number $\mu(b)$, which is the product moment of the random variables $X[b] = \{X^i \colon i \in b\}$. Any such function defined on subsets of $[n]$ can be extended multiplicatively to a function on set partitions by $\mu(\sigma) = \prod_{b \in \sigma} \mu(b)$. Likewise, the joint cumulant function κ associates with each non-empty subset $b \subset [n]$ a number $\kappa(b)$, which is the joint cumulant of the random variables $X[b]$. The extension of κ to set partitions is also multiplicative over the blocks. It is a property of the partition lattice that if $f \equiv \kappa$ is multiplicative, so also is the cumulative function $F \equiv \mu$. In particular, the full product moment is the sum of cumulant products

$$\mu([n]) = \sum_{\sigma} \prod_{b \in \sigma} \kappa(b).$$

For zero-mean Gaussian variables, all cumulants are zero except those of order two, and the above expression reduces to Isserlis's theorem [5] for $n = 2k$, which is a the sum over $n!/(2^k k!)$ pairings of covariance products. Wick's theorem, as it is known in the quantum field literature, is closely associated with Feynman diagrams. These are not merely a symbolic device for the computation of Gaussian moments, but also an aid for interpretation in terms of particle collisions [4, Chapter 8]. For an account that is accessible to statisticians, see Janson [8] or the AMS feature article by Phillips [17].

The moments and cumulants arising in this way involve distinct random variables, for example $X^2X^3X^4$, never $X^3X^3X^4$. However, variables that are given distinct labels may be equal, say $X^2 = X^3$ with probability one, so this is not a limitation. As virtually everyone who has worked with cumulants, from Kaplan [9] to Speed and thereafter, has noted, *the general results are most transparent when all random variables are taken as distinct.*

The arguments put forward in the paper for the combinatorial lattice-theoretic approach are based on the simplicity of the proof of various known results. For example, it is shown that the ordinary cumulant $\kappa([n])$ is zero if the variables can be partitioned into two independent blocks. Subsequently, Streitberg [25] used cumulant measures to give an if and only if version of the same result. To my mind, however, the most compelling argument for Speed's combinatoric approach comes in Proposition 4.3, which offers a simple proof of the Leonov-Shiryaev result using lattice-theoretic operations. To each subset $b \subset [n]$ there corresponds a product random variable $X^b = \prod_{i\in b} X^i$. To each partition σ there corresponds a set of product variables, one for each of the blocks $b \in \sigma$, and a joint cumulant $\kappa^\sigma = \mathrm{cum}\{X^b\colon b \in \sigma\}$. One of the obstacles that I had encountered in work on asymptotic approximation of mildly non-linear transformations of joint distributions was the difficulty of expressing such a generalized cumulant in terms of ordinary cumulants. The lattice-theoretic expression is remarkable for its simplicity:

$$\kappa^\sigma = \sum_{\tau:\tau\vee\sigma=\mathbf{1}_n} \prod_{b\in\tau} \kappa(b),$$

where the sum extends over partitions τ such that the least upper bound $\sigma \vee \tau$ is the single-block partition $\mathbf{1}_n = \{[n]\}$. Tables for these connected partitions are provided in McCullagh [14]. For example, if $\sigma = 12|34|5$ the third-order cumulant κ^σ is a sum over 25 connected partitions. If all means are zero, partitions having a singleton block can be dropped, leaving nine terms

$$\kappa^{12,34,5} = \kappa^{1,2,3,4,5} + \kappa^{1,2,3}\kappa^{4,5}[4] + \kappa^{1,3,5}\kappa^{2,4}[4]$$

in the abbreviated notation of McCullagh [13]. Versions of this result can be traced back to James [6], Leonov and Shiryaev [11], James and Mayne [7], and Malyshev [12].

A subject such as statistical moments and cumulants that has been thoroughly raked over by Thiele, Fisher, Tukey, Dressel and others for more than a century, might seem dry and unpromising as a topic for current research. Surprisingly, this is not the case. Although the area has largely been abandoned by research statisticians, it is a topic of vigorous mathematical research connected with Voiculescu's theory of non-commutative random variables, in which there exists a notion of freeness related to, but distinct from, independence. The following is a brief idiosyncratic sketch emphasizing the parallels between Speicher's work and Speed's paper.

First, Speed's combinatorial theory is purely algebraic: it does not impose positive definiteness conditions on the moments or cumulants, nor does it require them to be real-valued, but it does implicitly require commutativity of the variables. In a theory of non-commutative random variables, we may think of $X^1, \ldots, X^n$ as orthogonally invariant matrices of unspecified order. For a subset $b \subset [n]$, the *scalar product* $X^b = \operatorname{tr} \prod_{i \in b} X^i$ is the trace of the matrix product, which depends on the cyclic order. The first novelty is that $\mu(b) = E(X^b)$ is not a function on subsets of $[n]$, but a function on cyclically ordered subsets. Since every permutation $\sigma \colon [n] \to [n]$ is a product of disjoint cycles, every function on cyclically ordered subsets can be extended multiplicatively to a function on permutations $\mu(\sigma) = \prod_{b \in \sigma} \mu(b)$. Given two permutations, we say that τ is a *sub-permutation* of σ if each cycle of τ is a subcycle of some cycle of σ — in the obvious sense of preserving cyclic order [1]. For $\tau \le \sigma$, the crossing number $\chi(\tau, \sigma)$ is the number of 4-cycles (i, j, k, l) below σ such that i, k and j, l are consecutive in τ: $\chi(\tau, \sigma) = \#\{(i, j, k, l) \le \sigma \colon \tau(i) = k,\ \tau(j) = l\}$, and τ is called *non-crossing in* σ if $\chi(\tau, \sigma) = 0$. For a good readable account of the non-crossing property, see Novak and Sniady [16].

Although it is not a lattice, the set Π_n of permutations has a lattice-like structure; each maximal interval $[0_n, \sigma]$, in which 0_n is the identity and σ is cyclic, is a lattice. With sub-permutation as the partial order, $[0_n, \sigma] \cong \mathscr{E}_n$ is isomorphic with the standard partition lattice; with non-crossing sub-permutation as the partial order, each maximal interval is a partition lattice of a different structure. Speicher's combinatorial theory of moments and cumulants of non-commutative variables uses Möbius inversion on this lattice of non-crossing partitions [24]. If $f \equiv \kappa$ is multiplicative, so also is the cumulative function $F \equiv \mu$, and vice-versa. The function $\kappa(b)$ on cyclically ordered subsets is called the free cumulant because it is additive for sums of freely independent variables. Roughly speaking, freeness implies that the matrices are orthogonally or unitarily invariant of infinite order. For further discussion on this topic, see Nica and Speicher [15] or Di Nardo et al. [2].

The partition lattice simplifies the sampling theory of symmetric functions, leading to a complete account of the joint moments of Fisher's k-statistics and Tukey's polykays [19]. It led to the development of an extended theory of symmetric functions for structured and nested arrays associated with a certain subgroup [20, 21, 22, 23]. Elegant though they are, these papers are not for the faint of heart. With some limitations, it is possible to develop a parallel theory of spectral k-statistics and polykays — polynomial functions of eigenvalues having analogous finite-population inheritance and reverse-martingale properties. Simple expressions are easily obtained for low-order statistics, but the general theory is technically rather complicated.

Acknowledgement
I am grateful to Elvira Di Nardo, John Kolassa, Lek Heng Lim and Domenico Senato for helpful comments on an earlier draft. Support for this work was provided by NSF grant DMS-0906592.

References

[1] P. Biane. Some properties of crossings and partitions. *Discrete Math.*, 175: 41–53, 1997.

[2] E. Di Nardo, P. Petrullo, and D. Senato. Cumulants and convolutions via Abel polynomials. *Eur. J. Combin.*, 31(7):1792–1804, 2010.

[3] W. J. Ewens. The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.*, 3:87–112, 1972.

[4] J. Glimm and A. Jaffe. *Quantum Physics*. Springer, 1987.

[5] L. Isserlis. On a formula for the product moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12: 134–139, 1918.

[6] G. S. James. On moments and cumulants of systems of statistics. *Sankhyā*, 20:1–30, 1958.

[7] G. S. James and A. J. Mayne. Cumulants of functions of random variables. *Sankhyā*, 24:47–54, 1962.

[8] S. Janson. *Gaussian Hilbert Spaces*. Cambridge University Press, 1997.

[9] E. L. Kaplan. Tensor notation and the sampling cumulants of k-statistics. *Biometrika*, 39:319–323, 1952.

[10] J. F. C. Kingman. The representation of partition structures. *J. Lond. Math. Soc.*, 18:374–380, 1978.

[11] V. P. Leonov and A. N. Shiryaev. On a method of calculation of semi-invariants. *Theor. Probab. Appl.*, 4:319–329, 1959.

[12] V. A. Malyshev. Cluster expansions in lattice models of statistical physics and the quantum theory of fields. *Russ. Math. Surv.*, 35:1–62, 1980.

[13] P. McCullagh. Tensor notation and cumulants of polynomials. *Biometrika*, 71: 461–476, 1984.

[14] P. McCullagh. *Tensor Methods in Statistics*. Chapman and Hall, London, 1987.

[15] A. Nica and R. Speicher. *Lectures on the Combinatorics of Free Probability*. Cambridge University Press, London, 2006.

[16] J. Novak and P. Sniady. What is a free cumulant? *Notices of the American Mathematical Society*, 58:300–301, 2011.

[17] T. Phillips. Finite-dimensional Feynman diagrams. *American Math. Society feature column*, 2001.

[18] G.-C. Rota. On the foundations of combinatorial theory I. *Z. Wahrscheinlichkeitstheorie und verw. Geb.*, 2:340–386, 1964.

[19] T. P. Speed. Cumulants and partition lattices II: Generalised k-statistics. *J. Aust. Math. Soc. A*, 40:34–53, 1986.

[20] T. P. Speed. Cumulants and partition lattices III: Multiply-indexed arrays. *J. Aust. Math. Soc. A*, 40:161–182, 1986.

[21] T. P. Speed. Cumulants and partition lattices IV: A.S. convergence of generalised k-statistics. *J. Aust. Math. Soc. A*, 40:79–94, 1986.

[22] T. P. Speed and H. L. Silcock. Cumulants and partition lattices V. Calculating generalized k-statistics. *J. Aust. Math. Soc. A*, 44(2):171–196, 1988.

[23] T. P. Speed and H. L. Silcock. Cumulants and partition lattices VI. Variances and covariances of mean squares. *J. Aust. Math. Soc. A*, 44(3):362–388, 1988.
[24] R. Speicher. Multiplicative functions on the lattice of non-crossing partitions and free convolution. *Math. Ann.*, 296:611–628, 1994.
[25] B. Streitberg. Lancaster interactions revisited. *Ann. Stat.*, 18:1878–1885, 1990.

Austral. J. Statist., **25** (**2**), 1983, 378–388

CUMULANTS AND PARTITION LATTICES[1]

T. P. SPEED
CSIRO Division of Mathematics and Statistics, Canberra

Summary

The (joint) cumulant of a set of (possibly coincident) random variables is defined as an alternating sum of moments with appropriate integral coefficients. By exploiting properties of the Möbius function of a partition lattice some basic results concerning cumulants are derived and illustrations of their use given.

1. Introduction

Cumulants were first defined and studied by the Danish scientist T. N. Thiele (1889, 1897, 1899) who called them *half-invariants* (halvinvarianter); see Hald (1981) for a review of this early work. The ready interpretability and descriptive power of the first few cumulants was evident to Thiele, as was their role in studying non-linear functions of random variables, and these aspects of their use have continued to be important to the present day, see Brillinger (1975, Section 2.3). In a sense which it is hard to make precise, all of the important aspects of (joint) distributions seem to be simpler functions of cumulants than of anything else, and they are also the natural tools with which transformations (linear or not) of systems of random variables (independent or not) can be studied when exact distribution theory is out of the question.

The definition of multivariate cumulant most commonly used today involves moment-generating functions. If $X_1, \ldots, X_m$ is a system of m random variables and $\mathbf{r} = (r_1, \ldots, r_m)$ is an m-tuple of non-negative integers, then the cumulants $\{\kappa_\mathbf{r}\}$ of $X_1, \ldots, X_m$ are defined by $\kappa_{0\ldots0} = 0$ and the identity

$$\sum_{\mathbf{r}} \kappa_\mathbf{r} \frac{\boldsymbol{\theta}^\mathbf{r}}{\mathbf{r}!} = \log \sum_{\mathbf{r}} \mathbb{E}\{\mathbf{X}^\mathbf{r}\} \frac{\boldsymbol{\theta}^\mathbf{r}}{\mathbf{r}!}. \tag{1.1}$$

where we have written $\boldsymbol{\theta}^\mathbf{r} = \theta_1^{r_1} \ldots \theta_m^{r_m}$, $\mathbf{X}^\mathbf{r} = X_1^{r_1} \ldots X_m^{r_m}$ and $\mathbf{r}! = r_1! \ldots r_m!$, and summed over $r_1 \geqslant 0, \ldots, r_m \geqslant 0$. Here and below all relevant moments are assumed to exist. An alternative approach which

[1] Manuscript received September 22, 1982; revised February 1, 1983.

is in some respects more convenient defines the joint cumulant $\mathscr{C}(X_1, \ldots, X_m)$ of $X_1, \ldots, X_m$ ($\kappa_{1\ldots1}$ in the notation above) directly:

$$\mathscr{C}(X_1, \ldots, X_m) = \sum_{\sigma} (-1)^{b(\sigma)-1}(b(\sigma)-1)! \prod_{a=1}^{b(\sigma)} \mathbb{E}\left\{\prod_{i \in \sigma_a} X_i\right\} \qquad (1.2)$$

the sum being over all *partitions* σ of $\{1, \ldots, m\}$ into $b = b(\sigma) \geq 1$ *blocks* $\sigma_1, \sigma_2, \ldots, \sigma_b$. For example, if $m = 3$ we have

$$\mathscr{C}(X_1, X_2, X_3) = \mathbb{E}\{X_1X_2X_3\} - \mathbb{E}\{X_1X_2\}\mathbb{E}\{X_3\} - \mathbb{E}\{X_1X_3\}\mathbb{E}\{X_2\} - \mathbb{E}\{X_1\}\mathbb{E}\{X_2X_3\} + 2\mathbb{E}\{X_1\}\mathbb{E}\{X_2\}\mathbb{E}\{X_3\}.$$

Note that we have not required that the random variables $X_1, \ldots, X_m$ are all *distinct*. If $X_1 = X_2 = X_3 = X$ in the last formula, we obtain an expression which in the notation of Kendall & Stuart (1969) we recognise to be the formula $\kappa_3 = \mu'_3 - 3\mu'_1\mu'_2 + 2(\mu'_1)^3$. The general multivariate cumulant $\kappa_{\mathbf{r}}$ can be defined via (1.2) in a similar manner.

The purpose of this expository note is to derive some basic results concerning (joint) cumulants from definition (1.2) and give illustrations of their use. Our approach is based upon the fact that (1.2) is an instance of Möbius inversion over the lattice $\mathscr{P}(\underline{m})$ of all partitions of the set $\underline{m} = \{1, \ldots, m\}$, and further use of this technique leads to some new proofs. None of the results we prove are new; our aim is simply to show how a small investment in modern algebra—in this instance the theory of Möbius functions—helps us to step our way elegantly through some potentially messy classical algebra.

It is a great pleasure to be able to contribute to this number honouring Evan Williams. Amongst many other things he introduced me to cumulants and showed me their usefulness, and I hope that this note can convey some of the enjoyment I have found working with them.

2. Lattice Preliminaries

A *partition* σ of a non-empty set S is simply a family of non-empty subsets $\sigma_1, \ldots, \sigma_b$—called the *blocks* of σ—whose union is S. For example, the family $\sigma = \{\{1, 2\}, \{3\}, \{4\}\}$ is a partition of $S = \{1, 2, 3, 4\}$ and we denote it by $\sigma = 12|3|4$. If σ and τ are two partitions of the same set S and every block of σ is contained in a block of τ, then we say that σ is *finer* than τ (τ is *coarser* then σ) and write $\sigma \leq \tau$ ($\tau \geq \sigma$). In this way we find that the collection $\mathscr{P}(S)$ of all partitions of S becomes a *partially-ordered set* and it is in fact a *lattice*, for every pair $\sigma, \tau \in \mathscr{P}(S)$ has a least upper bound and a greatest lower bound in the partial order. The greatest lower bound $\sigma \wedge \tau$ of σ and τ is easy to describe directly: its blocks are just the non-empty intersections of blocks of σ with blocks of τ. For example, $123|4 \wedge 12|34 = 12|3|4$ and $12|34 \wedge 13|24 = 1|2|3|4$ hold in $\mathscr{P}(\underline{4})$. An excellent general

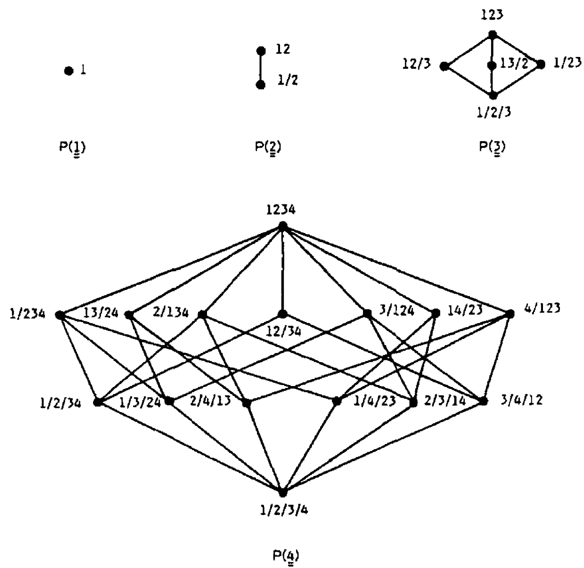

Fig. 1

reference for lattice theory and those results we quote below is Aigner (1979). We illustrate the foregoing with Hasse diagrams of the small partition lattices, see Figure 1.

In these diagrams each element of the partially-ordered set is denoted by a vertex, and an edge is drawn between the vertices corresponding to ρ and τ if $\rho < \tau$ (or $\rho > \tau$) and there is *no* element σ with $\rho < \sigma < \tau$ (or $\rho > \sigma > \tau$).

Associated with *any* finite partially ordered set $(\mathcal{P}, \leq)$ are two important numerical functions defined on $\mathcal{P}$: its *zeta function* $\zeta_{\mathcal{P}}$ given by $\zeta(\sigma, \tau) = 1$ if $\sigma \leq \tau$, and 0 otherwise; and its Möbius function $\mu = \mu_{\mathcal{P}}$ which can be defined in many ways, one simple one being the following:

$$\mu(\rho, \tau) = \begin{cases} 1 & \text{if} \quad \rho = \tau; \\ -\sum_{\rho \leq \sigma < \tau} \mu(\rho, \sigma) & \text{if} \quad \rho < \tau; \\ 0 & \text{otherwise.} \end{cases}$$

It is not hard to prove that $\sum_\sigma \mu(\rho, \sigma)\zeta(\sigma, \tau) = \sum_\sigma \zeta(\rho, \sigma)\mu(\sigma, \tau) = \delta(\rho, \tau)$ where $\delta(\rho, \tau) = 1$ if $\rho = \tau$ and 0 otherwise, i.e. the *matrices* $Z = (\zeta(\sigma, \tau))$ and $M = (\mu(\sigma, \tau))$ over $\mathcal{P}$ are mutually inverse.

Let us suppose that f is a real-valued function on $\mathcal{P}$ and that we define another function F on $\mathcal{P}$ by

$$F(\tau) = \sum_{\sigma \leq \tau} f(\sigma).$$

Thinking of f as a column vector this is saying that $F = Zf$. *Möbius inversion* is just the recovery of f from F: formally, $f = Z^{-1}F = MF$, and more fully

$$f(\tau) = \sum_{\sigma} \mu(\sigma, \tau) F(\sigma).$$

The power of Möbius inversion rests in the fact that for many familiar partially ordered sets, there is a simple formula for μ. Indeed it can be quite a useful technique without even having a formula! These basic ideas apply to any finite partially ordered set and we refer to Aigner (1979) for many illustrations.

It is clear from the definition of μ that if $\rho < \tau$ and there is no σ with $\rho < \sigma < \tau$, then $\mu(\rho, \tau) = -\mu(\rho, \rho) = -1$. Referring to the diagram of $\mathcal{P}(\underline{3})$ we can readily calculate that $\mu(1 \mid 2 \mid 3, 123) = 2$, whilst all other μ-values there are $+1$ or -1. Similarly we find that in $\mathcal{P}(\underline{4})$ the following are true: $\mu(12 \mid 3 \mid 4, 1234) = 2$ whilst $\mu(1 \mid 2 \mid 3 \mid 4, 1234) = -6$. It can be shown that for any $\sigma \in \mathcal{P}(\underline{m})$ we have $\mu(\sigma, \underline{m}) = (-1)^{b-1}(b-1)!$ where $b = b(\sigma)$ is the number of blocks of σ; a product of such expressions gives a formula for $\mu(\sigma, \tau)$ in $\mathcal{P}(\underline{m})$ but we will have no occasion to use it. We refer to Rota (1964), and Aigner (1979) for a proof.

3. Equivalence of the Two Definition

We will begin the proof of the equivalence of the two definitions by seeking an expression for $\mathbb{E}\{X_1 \ldots X_m\}$ in terms of the $\{\kappa_{\mathbf{r}}\}$, and to this end we introduce some notation which plays a fundamental role in what follows. For a partition $\sigma = \sigma_1 \mid \ldots \mid \sigma_b$ of $\{1, \ldots, m\}$ let us write

$$\kappa_\sigma = \prod_{a=1}^{b(\sigma)} \kappa_{\mathbf{r}(\sigma_a)}$$

where $\mathbf{r}(\sigma_a) = (r_1, \ldots, r_m)$ is defined by $r_i = 1$ if $i \in \sigma_a$, $r_i = 0$ otherwise, $a = 1, \ldots, b(\sigma)$. For example, if $\sigma = 1234$, then $\kappa_\sigma = \kappa_{1111}$, whilst if $\sigma = 12 \mid 34$, then $\kappa_\sigma = \kappa_{1100}\kappa_{0011}$.

Now let us exponentiate both sides of (1.1) and calculate the coefficient of $\theta_1 \ldots \theta_m$ on the left-hand side. It is really quite straightforward to see that the answer is $\sum_\sigma \kappa_\sigma$, where the κ_σ have just been defined and the sum is over all partitions σ of $\{1, \ldots, m\}$. For example, $\mathbb{E}\{X_1X_2X_3X_4\}$ is the sum of 15 terms beginning with $\kappa_{1234} = \kappa_{1111}$ and ending with $\kappa_{1|2|3|4} = \kappa_{1000}\kappa_{0100}\kappa_{0010}\kappa_{0001}$. More generally, if τ is an arbitrary partition of $\{1, \ldots, m\}$ with blocks $\tau_1, \ldots, \tau_b$, then we can multiply expressions of the form just derived to obtain the identity

$$\prod_{a=1}^{b(\tau)} \mathbb{E}\left\{\prod_{i \in \tau_a} X_i\right\} = \prod_{a=1}^{b(\tau)} \sum_{\sigma_a \in \mathcal{P}(\tau_a)} \kappa_{\sigma_a} = \sum_{\sigma \leq \tau} \kappa_\sigma. \tag{3.1}$$

For example, $\mathbb{E}\{X_1X_2\}\mathbb{E}\{X_3\}\mathbb{E}\{X_4\} = (\kappa_{12} + \kappa_{1|2})\kappa_3\kappa_4 = \kappa_{12|3|4} + \kappa_{1|2|3|4}$. Now equation (3.1) can be inverted by Möbius inversion and doing so gives us the fundamental relationship:

$$\kappa_\tau = \sum_{\sigma} \mu(\sigma, \tau) \prod_{a=1}^{b(\sigma)} \mathbb{E}\left\{\prod_{i\in\sigma_a} X_i\right\}. \tag{3.2}$$

When $\tau = \underline{m}$ this reduces to (1.2), apart from the identification of $\mu(\sigma, \underline{m})$ as $(-1)^{b(\sigma)-1}(b(\sigma)-1)!$, and we have proved the equivalence of the definitions.

A more abstract and general theory including this equivalence can be found in Doubilet *et al.* (1972).

Example. Putting $m=4$ we see from (3.2) and Figure 1 that $\kappa_{1234} = \kappa_{1111}$ is an alternating sum of 15 terms with coefficients $+1$, -1, $+2$ and -6. If we identify two or more of the random variables $X_1, \ldots, X_4$, additional numerical factors enter because the same expression appears more than once in the 15 terms. At the extreme, when $X_1 = X_2 = X_3 = X_4 = X$, we find cf. Kendall & Stuart (1969, p. 701) the traditional expression

$$\kappa_4 = \mathbb{E}\{X^4\} - 4\mathbb{E}\{X^3\}\mathbb{E}\{X\} - 3(\mathbb{E}\{X^2\})^2 + 12\mathbb{E}\{X^2\}(\mathbb{E}\{X\})^2 - 6(\mathbb{E}\{X\})^4.$$

Here the factors of -4, -3 and 12 are a combination of multiplicities and Möbius function values.

It is a long standing observation of workers with cumulants that the general results are most transparent when all random variables under discussion are taken as distinct. The identification of some or all at a later stage merely introduces extra factors, and at times these multiplicities are not particularly easy to calculate.

4. Properties of Cumulants

Cumulants of order 2 are just variances and covariances and a number of properties which are familiar in this case seem much less well known in general. Our first result provides a good illustration of the way in which Möbius inversion may be used in this context although its proof using (1.1) is also easy. We take as given a set $X_1, \ldots, X_m$ of random variables, and write $\underline{m} = \{1, \ldots, m\}$.

Proposition 4.1. If there is a subset $s \subseteq \underline{m}$ such that the random variables $\{X_i : i \in s\}$ and $\{X_i : i \in t\}$ are independent, $t = \underline{m}\setminus s$, then $\mathscr{C}(X_1, \ldots, X_m) = 0$.

Proof. For each $\pi \in \mathscr{P}(\underline{m})$ we denote the partition induced on s, i.e. that partition having as blocks the non-empty numbers of $\pi_1 \cap s, \ldots, \pi_b \cap s$, by $\pi \cap s$, and similarly for $\pi \cap t$. The proof makes crucial use of the following simple fact: for any $\pi \in \mathscr{P}(\underline{m})$, $\sigma \in \mathscr{P}(s)$

and $\tau \in \mathscr{P}(t)$ we have

$$\pi \geq \sigma \mid \tau \quad \text{iff} \quad \pi \cap s \geq \sigma \quad \text{and} \quad \pi \cap t \geq \tau. \tag{4.1}$$

We can now go to (1.2) and calculate:

$$\begin{aligned}\mathscr{C}(X_1, \ldots, X_m) &= \sum_{\pi} \mu(\pi, \underline{m}) \prod_{a=1}^{b(\pi)} \mathbb{E}\Big\{\prod_{i\in\pi_a} X_i\Big\}\\ &= \sum_{\pi} \mu(\pi, \underline{m}) \prod_{a=1}^{b(\pi\cap s)} \mathbb{E}\Big\{\prod_{i\in\pi_a\cap s} X_i\Big\} \prod_{a=1}^{b(\pi\cap t)} \mathbb{E}\Big\{\prod_{i\in\pi_a\cap t} X_i\Big\}\\ &\qquad \text{by independence,}\\ &= \sum_{\pi} \mu(\pi, \underline{m}) \Big\{\sum_{\sigma} \zeta(\sigma, \pi\cap s)\kappa_\sigma\Big\}\Big\{\sum_{\tau} \zeta(\tau, \pi\cap t)\kappa_\tau\Big\}\\ &\quad \text{by (3.1),}\\ &= \sum_{\pi}\sum_{\sigma}\sum_{\tau} \mu(\pi, \underline{m})\zeta(\sigma, \pi\cap s)\zeta(\tau, \pi\cap t)\kappa_\sigma\kappa_\tau)\\ &= \sum_{\pi}\sum_{\sigma}\sum_{\tau} \mu(\pi, \underline{m})\zeta(\sigma\mid\tau, \pi)\kappa_\sigma\kappa_\tau \quad \text{by (4.1),}\\ &= \sum_{\sigma}\sum_{\tau} \delta(\sigma\mid\tau, \underline{m})\kappa_\sigma\tau_\sigma \quad \text{by Möbius inversion}\end{aligned}$$

and this expression is zero since $\underline{m} \neq \sigma \mid \tau$ for any $\sigma \in \mathscr{P}(s)$, $\tau \in \mathscr{P}(t)$.

For the next two propositions we consider an array $(X_{ij} : j \in \underline{n}_i, i \in \underline{m})$ of real random variables and a similarly indexed array (a_{ij}) of real numbers. The following result also generalises a well known one for variances and covariances: it states that $\mathscr{C}$ is a *multi-linear* operator.

Proposition 4.2.

$$\mathscr{C}\Big(\sum_{j_1} a_{1j_1}X_{1j_1}, \ldots, \sum_{j_m} a_{mj_m}X_{mj_m}\Big) = \sum_{j_1} \cdots \sum_{j_m} a_{1j_1} \cdots a_{mj_m}\mathscr{C}(X_{1j_1}, \ldots, X_{mj_m}).$$

Proof. From (1.2) and the distributive law

$$\begin{aligned}&\sum_{\sigma} \mu(\sigma, \underline{m}) \prod_{a=1}^{b} \mathbb{E}\Big\{\prod_{i\in\sigma_a}\sum_{j_i\in\underline{n}_i} a_{ij_i}X_{ij_i}\Big\}\\ &\quad= \sum_{\sigma} \mu(\sigma, \underline{m}) \prod_{a=1}^{b} \mathbb{E}\Big\{\sum_{j_i\in\underline{n}_i, i\in\sigma_a}\prod_{i\in\sigma_a} a_{ij_i}X_{ij_i}\Big\}\\ &\quad= \sum_{\sigma} \mu(\sigma, \underline{m}) \sum_{j_i\in\underline{n}_i, i\in\underline{m}}\prod_{i} a_{ij_i} \prod_{a=1}^{b} \mathbb{E}\Big\{\prod_{i\in\sigma_a} X_{ij_i}\Big\}\\ &\quad= \sum_{j_1} \cdots \sum_{j_m} a_{1j_1} \cdots a_{mj_m} \sum_{\sigma} \mu(\sigma, \underline{m}) \prod_{a=1}^{b} \mathbb{E}\Big\{\prod_{i\in\sigma_a} X_{ij_i}\Big\}\end{aligned}$$

which is the stated result.

Corollary. If $X_i = Y_i + Z_i$, $i = 1, \ldots, m$, where $\{Y_i\}$ and $\{Z_i\}$ are independent sets of m real random variables, then

$$\mathscr{C}(X_1, \ldots, X_m) = \mathscr{C}(Y_1, \ldots, Y_m) + \mathscr{C}(Z_1, \ldots, Z_m).$$

Proof. This is an immediate consequence of Propositions 4.1 and 4.2.

The following proposition is the core of the main result of Leonov & Shiryaev (1959). Our proof is much more direct than theirs and highlights the power of Möbius inversion. Any partition $\pi = \pi_1 | \ldots | \pi_b$ of the row labels $\underline{m}$ of the (X_{ij}) induces a partition $\tilde{\pi}$ of the full set $S = \{(i, j) : j \in \underline{n}_i, i \in \underline{m}\}$ of labels in a natural way: $\tilde{\pi}$ has blocks $\{(i, j) : j \in \underline{n}_i, i \in \pi_a\}$, $a = 1, \ldots, b(\pi)$. We say that a partition σ of S is *decomposable relative to* a partition π of $\underline{m}$ when $\sigma \leq \tilde{\pi}$, where $\tilde{\pi}$ has just been defined, and we call σ *indecomposable* if no such relation holds other than $\sigma \leq \underline{m}$. Brillinger (1975, p. 20) gives some equivalent formulations, and states without proof the following.

Proposition 4.3.

$$\mathscr{C}\Big(\prod_{j_1 \in \underline{n}_1} X_{1j_1}, \ldots, \prod_{j_m \in \underline{n}_m} X_{mj_m}\Big) = \sum_{\sigma}{}^{*} \prod_{a=1}^{b(\sigma)} \mathscr{C}(X_{ij} : (i, j) \in \sigma_a)$$

where $\sum^*$ denotes the sum over all *indecomposable* partitions σ of S.

Proof. For any $\pi \in \mathscr{P}(\underline{m})$ we have by (3.1)

$$\prod_{a=1}^{b(\pi)} \mathbb{E}\Big\{\prod_{i \in \pi_a} \prod_{j_i \in \underline{n}_i} X_{ij_i}\Big\} = \sum_{\sigma \leq \tilde{\pi}} \prod_{a=1}^{b(\sigma)} \mathscr{C}(X_{ij} : (i, j) \in \sigma_a).$$

The sum on the right, which we denote by $F(\pi)$, is over all $\sigma \in \mathscr{P}(S)$ which are decomposable relative to π. Such σ may also be decomposable relative to some $\rho < \pi$, and so we can use Möbius inversion over $\mathscr{P}(\underline{m})$ to write $f(\pi) = \sum_\rho \mu(\rho, \pi)F(\rho)$ for the corresponding sum over all σ which are decomposable relative to π and no finer partition. With this notation we use (1.2) and Möbius inversion over $\mathscr{P}(\underline{m})$ once more to obtain

$$\mathscr{C}\Big(\prod_{j_1 \in \underline{n}_1} X_{1j_1}, \ldots, \prod_{j_m \in \underline{n}_m} X_{mj_m}\Big) = \sum_{\pi} \mu(\pi, \underline{m})F(\pi) = f(\underline{m})$$

and the proof is complete.

This proposition provides easy access to a number of results due to Isserlis (1918–19a, b) Bergström (1918–19), Wishart (1928–29, 1929) and others.

Example. Let us take $S = \{(1, 1), (1, 2), (2, 1), (2,2)\}$ which we simplify to $\{1, 2, 3, 4\}$. Then we may refer to the lattice $\mathscr{P}(\underline{4})$ and, by omitting the decomposable partitions $12 | 34$, $12 | 3 | 4$, $1 | 2 | 34$ and

$1|2|3|4$ we readily see that

$$\begin{aligned}\mathscr{C}(X_1X_2, X_3X_4) &= \mathscr{C}(X_1, X_2, X_3, X_4)\\ &+ \mathscr{C}(X_1)\mathscr{C}(X_2, X_3, X_4) + \mathscr{C}(X_2)\mathscr{C}(X_1, X_3, X_4)\\ &+ \mathscr{C}(X_3)\mathscr{C}(X_1, X_2, X_4) + \mathscr{C}(X_4)\mathscr{C}(X_1, X_2, X_3) + \mathscr{C}(X_1, X_3)\mathscr{C}(X_2, X_4)\\ &+ \mathscr{C}(X_1, X_4)\mathscr{C}(X_2, X_3) + \mathscr{C}(X_1)\mathscr{C}(X_3)\mathscr{C}(X_2, X_4)\\ &\qquad + \mathscr{C}(X_1)\mathscr{C}(X_4)\mathscr{C}(X_2, X_3)\\ &+ \mathscr{C}(X_2)\mathscr{C}(X_3)\mathscr{C}(X_1; X_4) + \mathscr{C}(X_2)\mathscr{C}(X_4)\mathscr{C}(X_1, X_3).\end{aligned}$$

If X_1, X_2, X_3 and X_4 have a joint normal distribution, then cumulants of order exceeding two all vanish, and in this case if their means are all zero we have

$$\operatorname{cov}(X_1X_2, X_3X_4) = \operatorname{cov}(X_1, X_3)\operatorname{cov}(X_2, X_4) + \operatorname{cov}(X_1, X_4)\operatorname{cov}(X_2, X_4).$$

As a further illustration of this result, let us suppose that $X_1, \ldots, X_n$ are mutually independent and identically distributed random variables with cumulants $\kappa_1 = 0$, $\kappa_2, \kappa_3, \kappa_4, \ldots$ (traditional notation). Then for any matrix (a_{ij}) of coefficients, we have

$$\operatorname{var}\left(\sum_i\sum_j a_{ij}X_iX_j\right) = \kappa_4 \sum_i a_{ii}^2 + \kappa_2^2 \sum_i\sum_j (a_{ij}^2 + a_{ij}a_{ji}).$$

The proof is almost immediate once we observe that we require

$$\mathscr{C}\left(\sum_i\sum_j a_{ij}X_iX_j, \sum_i\sum_j a_{ij}X_iX_j\right) = \sum_{i_1}\sum_{i_2}\sum_{i_3}\sum_{i_4} a_{i_1i_2}a_{i_3i_4}\mathscr{C}(X_{i_1}X_{i_2}, X_{i_3}X_{i_4}).$$

Of the 15 possible combinations of equality and inequality on i_1, i_2, i_3 and i_4, each corresponding to an element of $\mathscr{P}(\underline{4})$ in an obvious way, only three give a non-zero cumulant, namely those corresponding to 1234, $13|24$ and $14|23$. Now

$$s^2 = \frac{1}{n-1}\sum_i (X_i - \bar{X})^2 = \frac{1}{n}\sum_i X_i^2 - \frac{1}{n(n-1)}\sum_{i\neq j}\sum X_iX_j$$

and so if we put

$$a_{ii} = \frac{1}{n} \quad \text{and} \quad a_{ij} = a_{ji} = \frac{-1}{n(n-1)}, \quad i \neq j,$$

in the preceding result we obtain the formula which goes back to Gauss (1823);

$$\operatorname{var}(s^2) = \frac{1}{n}\kappa_4 + \frac{2}{n-1}\kappa_2^2.$$

Our final result, due to Brillinger (1969), and generalizes to higher-order cumulants the familiar identity

$$\mathrm{Var}\,(X) = \mathbb{E}\{\mathrm{Var}\,(X \mid Y)\} + \mathrm{Var}\,(\mathbb{E}\{X \mid Y\})$$

for real random variables X and Y. We will use an obvious notation for conditional cumulants.

Proposition 4.4.

$$\mathscr{C}(X_1, \ldots, X_m) = \textstyle\sum_\pi \mathscr{C}(\mathscr{C}(X_i : i \in \pi_a \mid Y) : a \in \underline{b}(\pi)),$$

the sum being over all partitions π of $\{1, \ldots, m\}$.

Proof. The proof which follows is not as simple as Brillinger's, which uses moment-generating functions. A typical term in the expansion (1.2) for $\mathscr{C}(X_1, \ldots, X_m)$ is a product of terms of the form $\mathbb{E}\{\prod_{i \in \pi_a} X_i\} = \mathbb{E}\{\mathbb{E}\{\prod_{i \in \pi_a} X_i \mid Y\}\}$, and we expand the inner term on the right-hand side of this using (3.1), switch the sum and the outer expectation, and use (3.1) once more. Most terms cancel and the simple result is derived. The notational details are somewhat messy, but we proceed.

$$\begin{aligned}\mathbb{E}\Big\{\mathbb{E}\Big\{\prod_{i \in \pi_a} X_i \mid Y\Big\}\Big\} &= \sum_{\sigma_a \in \mathscr{P}(\pi_a)} \mathbb{E}\Big\{\prod_{k \in \underline{b}(\sigma^a)} \mathscr{C}(X_i : i \in \sigma_a^k \mid Y)\Big\} \\ &= \sum_{\sigma_a \in \mathscr{P}(\pi_a)} \sum_{\tau_a \in \mathscr{P}(\underline{b}(\sigma^a))} \prod_{l \in \underline{b}(\tau_a)} \mathscr{C}((\sigma_a^k \mid Y) : k \in \tau_a^l)\end{aligned}$$

where we have abbreviated $\mathscr{C}(X_i : i \in \sigma_a^k \mid Y)$ by $(\sigma_a^k \mid Y)$. Putting this expression into (1.2) we obtain

$$\mathscr{C}(X_1, \ldots, X_m) = \sum_\pi \mu(\pi, \underline{m}) \sum_{\sigma \leq \pi} \sum_{\substack{\tau_a \in \mathscr{P}(\underline{b}(\sigma_a)) \\ a = 1, \ldots, b(\pi)}} \prod_{a \in \underline{b}(\pi)} \prod_{l \in \underline{b}(\tau_a)} \mathscr{C}((\sigma_a^k \mid Y) : k \in \tau_a^l)$$

where in the third sum we write $\sigma_a = \sigma \cap \pi^a$, $a = 1, \ldots, b(\pi)$. Our result is proved if we can show that only terms involving $\pi = \underline{m}$, i.e. $b(\pi) = 1$, survive.

To this end suppose that $\sigma \in \mathscr{P}(\underline{m})$ and $\tau \in \mathscr{P}(\underline{b}(\sigma))$ and write

$$P(\sigma, \tau) = \prod_{l \in \underline{b}(\tau)} \mathscr{C}((\sigma^k \mid Y) : k \in \tau^l)$$

$$\rho(\sigma, \tau) = \bigcup_{k \in \tau_1} \sigma^k \Big| \bigcup_{k \in \tau_2} \sigma^k \Big| \ldots .$$

Noting that $\rho(\sigma, \tau) \geq \sigma$, we find that the last sum can be written as

$$\sum_\pi \sum_\sigma \sum_\tau \mu(\pi, \underline{m}) \zeta(\pi, \rho(\sigma, \tau)) P(\sigma, \tau) = \sum_\sigma \sum_\tau \delta(\underline{m}, \rho(\sigma, \tau)) P(\sigma, \tau)$$

by Möbius inversion and the result is proved.

Example. For $m = 3$ this result asserts that

$$\begin{aligned}\mathscr{C}(X_1, X_2, X_3) = {} & \mathscr{C}(\mathscr{C}(X_1, X_2, X_3 \mid Y)) + \mathscr{C}(\mathscr{C}(X_1 \mid Y), \mathscr{C}(X_2, X_3 \mid Y)) \\ & + \mathscr{C}(\mathscr{C}(X_2 \mid Y), \mathscr{C}(X_1, X_3 \mid Y)) \\ & + \mathscr{C}(\mathscr{C}(X_3 \mid Y), \mathscr{C}(X_1, X_2 \mid Y)) \\ & + \mathscr{C}(\mathscr{C}(X_1 \mid Y), \mathscr{C}(X_2 \mid Y), \mathscr{C}(X_3 \mid Y)).\end{aligned}$$

If $X_1 = X_2 = X_3 = X$ and we adopt a suggestive notation, the previous expression simplifies to a formula similar to the well-known one for $\mathrm{Var}\,(X)$:

$$\kappa_3(X) = \mathbb{E}\{\kappa_3(X \mid Y)\} + 3\,\mathrm{cov}\,(\mathbb{E}\{X \mid Y\}, \mathrm{Var}\,(X \mid Y)) + \kappa_3(\mathbb{E}\{X \mid Y\}).$$

We note in closing that Proposition 4.4 has been used to obtain the cumulants of random sums of (iid) random variables, see e.g. Lange *et al.* (1981).

5. Closing Remarks

The theory of *k-statistics* developed by Fisher (1928–29) and its generalized form involving the so-called *polykays* due to Tukey (1950) is also simplified greatly by a recognition of the role played by the underlying partition lattices and their Möbius functions. For example, it is possible to give a fairly compact proof of a generalization of Fisher's famous result concerning the joint cumulants of sample k-statistics along the lines of that of Proposition 4.3 above.

In a quite different direction, (joint) cumulants of another kind can be defined for arrays of random variables labelled by multiple indices as in a complex experimental design. Here the second order cumulants turn out to be components of variance, and many interesting generalizations of anova notions appear. We leave this and other work to another time.

References

AIGNER, M. (1979). *Combinatorial Theory.* New York: Springer-Verlag.

BERGSTRÖM, S. (1918–19). Sur les moments de la function de correlation normale de n variables. *Biometrika* **12,** 177–183.

BRILLINGER, D. R. (1969). The calculation of cumulants via conditioning. *Ann. Inst. Statist. Math.* **21,** 375–390.

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

DOUBILET, PETER, ROTA, GIAN-CARLO & STANLEY, RICHARD, P. (1972). On the foundations of combinatorial theory VI: The idea of a generating function. *Proceedings of the 6th Berkeley Symposium on Probability and Mathematical Statistics* 267–318. Edited by L. Le Cam, University of California Press.

FISHER, R. A. (1928–29). Moments and product moments of sampling distributions. *Proc. Lond. Math. Soc.* (2) **30,** 199–238.

GAUSS, C. F. (1823). Theoria Combinationis Observationum Erroribus Minimis Obnoxiae Pars Posterior. *Comm. Soc. Reg. Scient. Gotting.* 5. See also: *Werke* Band IV.

HALD, A. (1981). T. N. Thiele's contributions to statistics. *Int. statist. Rev.* **49,** 1–20.

ISSERLIS, L. (1918–19a). On a formula for the product moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* **12,** 134–139.

ISSERLIS, L. (1918–19b). Formulae for determining the mean value of products of deviations of mixed moment coefficients in two to eight variables in samples from a limited population. *Biometrika* **12,** 183–184.

KENDALL, MAURICE G. & STUART, ALAN (1969). *The Advanced Theory of Statistics.* Volume 1 (Third Edition). London: Griffin.

LANGE, KENNETH, BOEHNKE, MICHAEL & CARSON, RICHARD (1981). Moment computations for subcritical branching processes. *J. Applied Probability* **18,** 52–64.

LEONOV, V. P. & SHIRYAEV, A. N. (1959). On a method of calculation of semi-invariants. *Theor. Prob. Appl.* **4,** 319–329.

ROTA, GIAN-CARLO (1964). On the foundations of combinatorial theory I. Theory of Möbius functions. *Zeit. f. Warsch.* **2,** 340–368.

THIELE, T. N. (1889). Forelaesninger over Almindelig Iagttagelseslaere: Sandsynlighedregning og mindste Kvadraters Methode. København: Reitzel.

THIELE, T. N. (1897). *Elementaer Iagttagelseslaere.* København: Gyldendalske. English translation: *Theory of Observations.* London 1903: Layton. Reprinted in *Ann. Math. Statist.* (1931) **2,** 165–308.

THIELE, T. N. (1899). Om Iagttagelseslaerens Halvinvarienter. *Overs. Vid. Sels. Forh. Nr.* **3,** 135–141.

TUKEY, JOHN W. (1950). Some sampling simplified. *J. Amer. Statist. Assoc.* **45,** 501–519.

WISHART, JOHN (1928–29). A problem in combinatorial analysis giving the distribution of certain moment statistics. *Proc. Lond. Math. Soc.* (2) **30,** 309–321.

WISHART, JOHN (1929). The correlation between product moments of any order in samples from a normal population. *Proc. Roy. Soc. Edin.* **49,** 1.