

Commentary on Speed, T.P. (1983) *Cumulants and partition lattices*

Austral. J. Statist. **25**, 378–388.

Peter McCullagh, April 2011

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’s 1964 paper, and because it was immediately clear that Möbius inversion on the partition lattice \mathcal{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 $[n] = \{1, \dots, n\}$, 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 $\#\mathcal{E}_n$, whose exponential generating function is $\exp(e^t - 1)$. The symmetric group acting on \mathcal{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 \mathcal{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 integer partitions (Ewens 1972; Kingman 1978), 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 $\mathcal{E} = \{\mathcal{E}_1, \mathcal{E}_2, \dots\}$ 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 \mathcal{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

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

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 $\tau \leq \sigma$. In matrix notation, $F = Lf$, where L is lower-triangular with inverse M . The Möbius function for the binary lattice 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 $\tau \leq \sigma$, 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, \dots, 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 $X[b] = \{X^i : i \in b\}$. 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 (1918) theorem 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 (Glimm and Jaffe, 1987, chapter 8). For an account that is accessible to statisticians, see Janson (1997) or the AMS feature article by Phillips (2001).

The moments and cumulants arising in this way involve distinct random variables, for example $X^2 X^3 X^4$, never $X^3 X^3 X^4$. 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 (1952) 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 (1990) 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-Shirayev 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 $\kappa^\sigma = \text{cum}\{X^b : b \in \sigma\}$. 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 (1987). 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 (1984). Versions of this result can be traced back to James (1958), Leonov and Shiryaev (1959), James and Mayne (1962) and Malyshev (1980).

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, \dots, X^n as orthogonally invariant matrices of unspecified order. For a subset $b \subset [n]$, the *scalar product* $X^b = \text{tr} \prod_{i \in b} X^i$ 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 $\sigma: [n] \rightarrow [n]$ 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 sub-cycle of some cycle of σ — in the obvious sense of preserving cyclic order (Biane 1997). For $\tau \leq \sigma$, 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 : $\chi(\tau, \sigma) = \#\{(i, j, k, l) \leq \sigma: \tau(i) = k, \tau(j) = l\}$, 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 (2011).

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, $[0_n, \sigma] \cong \mathcal{E}_n$ 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 (1994) combinatorial theory of moments

and cumulants of non-commutative variables uses Möbius inversion on this lattice of non-crossing partitions. 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 (2006) or Di Nardo, Petrullo and Senato (2010).

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 (Speed, 1986). It led to the development of an extended theory of symmetric functions for structured and nested arrays associated with a certain sub-group (Speed, 1984ab), Speed and Silcock (1988ab). 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 Evira 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] Biane, P. (1997) Some properties of crossings and partitions. *Discrete Mathematics* **175**, 41–53.
- [2] Ewens, W.J. (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**, 87–112.
- [3] Fisher, R.A. (1929) Moments and product moments of sampling distributions. *Proc. Lond. Math. Soc.* **30**, 199–238.
- [4] Glimm, J. and Jaffe, A. (1987) *Quantum Physics*. Springer.
- [5] Isserlis, L. (1918) 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.

- [6] James, G.S. (1958) On moments and cumulants of systems of statistics. *Sankhya* **20**, 1–30.
- [7] James, G.S. and Mayne, A.J. (1962) Cumulants of functions of random variables. *Sankhya* **24**, 47–54.
- [8] Kaplan, E.L. (1952) Tensor notation and the sampling cumulants of k -statistics. *Biometrika* **39**, 319–323.
- [9] Kingman, J.F.C. (1977) The population structure associated with the Ewens sampling formula. *Theoretical Population Biology* **11**, 274–283.
- [10] Kingman, J.F.C. (1978) The representation of partition structures. *J. Lond. Math. Soc.* **18**, 374–380.
- [11] Kingman, J.F.C. (1980) *Mathematics of Genetic Diversity*. CBMS-NSF conference series in applied math, **34** SIAM, Philadelphia.
- [12] Leonov, V.P. and Shiryaev, A.N. (1959) On a method of calculation of semi-invariants. *Theory of Probability and its Applications* **4**, 319–329.
- [13] Malyshev, V.A. (1980) Cluster expansions in lattice models of statistical physics and the quantum theory of fields. *Russian Math. Surveys* **35**, 1–62.
- [14] McCullagh, P. (1984) Tensor notation and cumulants of polynomials. *Bometrika* **71**, 461–476.
- [15] McCullagh, P. (1987) *Tensor Methods in Statistics*. London, Chapman and Hall.
- [16] Nica, A. and Speicher, R. (2006) *Lectures on the Combinatorics of Free Probability*. London Math Society Lecture Note Series, **335** Cambridge University Press.
- [17] Novak, J. and Sniady, P. (2011) What is a free cumulant? *Notices of the American Mathematical Society* **58**, 300–301.
- [18] Phillips, T. (2001) Finite-dimensional Feynman diagrams. American Math. Society feature column, Nov. 2001: <http://www.ams.org/samplings/feature-column/fcarc-feynman1>.
- [19] Rota, G.-C. (1964) On the foundations of combinatorial theory I. Theory of Möbius functions. *Zeit. f. Warsch.* **2**, 340–386.

- [20] Speicher, R. (1994) Multiplicative functions on the lattice of non-crossing partitions and free convolution. *Mathematische Annalen* **296**, 611–628.
- [21] Streitberg, B. (1990) Lancaster interactions revisited. *Ann Statist* **18**, 1878–1885.
- [22] Tukey, J.W. (1950) Some sampling simplified. *J. Amer. Statist. Assoc.* **45**, 501–519.
- [23] Wick, G.C. (1950) The evaluation of the collision matrix. *Physical Review* **80**, 268–272.