View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Applied Mathematics Letters 19 (2006) 968–975
Applied Mathematics Letters
www.elsevier.com/locate/aml

# A symbolic method for $k$-statistics

E. Di Nardo*, D. Senato

*Dipartimento di Matematica, Università degli Studi della Basilicata, C.da Macchia Romana, I-85100 Potenza, Italy*

Received 19 September 2005; accepted 22 September 2005

**Abstract**

Through the classical umbral calculus, we provide new, compact and easy to handle expressions for $k$-statistics, and more generally for $U$-statistics. In addition, this symbolic method can be naturally extended to multivariate cases and to generalized $k$-statistics.
© 2005 Elsevier Ltd. All rights reserved.

*Keywords:* Umbral calculus; Symmetric polynomials; $U$-Statistics; $k$-Statistics; Joint cumulants

## 1. Introduction

In 1929, Fisher [4] introduced $k$-statistics as new symmetric functions of a random sample. The aim of Fisher was to estimate cumulants through free-distribution methods by using only combinatorial techniques. The $k$-statistics are related to the power sum symmetric polynomials in the random variables (r.v.'s) of the sample. The point of view of Fisher is described with a wealth of details by Kendall and Stuart (cf. [8]). The method is straightforward enough; however, his execution leads to intricate computations and some cumbersome expressions, except for very simple cases. That is why many authors tried to simplify the matter later.

One of the most relevant contributions was given by Speed [7] in the 1980's. Speed resumed the Doubilet approach to symmetric functions [3], consisting in labelling symmetric functions through partitions of a set rather than partitions of an integer. Transition matrices are computed via Mœbius function, generalizing and simplifying the presentation of the $k$-statistic theory. Nevertheless, in order to extend such a theory to generalized $k$-statistics (polykays), Speed had to resort to the tensor approach introduced by Kaplan in 1952.

We take a different point of view by using the high computational potential of the classical umbral calculus. This symbolic method was introduced by Rota and Taylor in 1994 [6] and further developed in [1] and [2]. From a combinatorial perspective, we revisit the Fisher theory as exposed by Kendall and Stuart, taking into account the Doubilet approach to symmetric functions. The umbral calculus offers a nimble syntax method that allows both computation without using Mœbius function and a natural extension to generalized $k$-statistics without bringing the tensor device into it. What is more, this language clarifies the role played by the power sum symmetric polynomials in the $k$-statistic expressions.

* Corresponding author.
*E-mail addresses:* dinardo@unibas.it (E. Di Nardo), senato@unibas.it (D. Senato).

0893-9659/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aml.2005.09.012

After recalling in Section 2 the strictly necessary umbral background, in Section 3 we put in an umbral setting the four classical bases of symmetric polynomial algebra. In Section 4 we give a general procedure for writing down $U$-statistics. Such a procedure is based on the umbral relation between moments and augmented symmetric polynomials (Theorem 4.1). In the last section, applications to $k$-statistics, $h$-statistics and multivariate $k$-statistics are given.

## 2. The umbral calculus language

We start by presenting the formal setting of umbral calculus. We shall confine our exposition to what is necessary for the aims of this paper.

The umbral calculus is a syntax consisting of the following data: an alphabet $A = \{\alpha, \beta, \dots\}$ whose elements are named *umbrae*; a commutative integral domain $R$ whose quotient field is of characteristic zero; a linear functional $E$, called *evaluation*, defined on the polynomial ring $R[A]$ and taking values in $R$, and such that $E[1] = 1$, $E[\alpha^i \beta^j \cdots \gamma^k] = E[\alpha^i]E[\beta^j]\cdots E[\gamma^k]$ for any set of distinct umbrae in $A$ and for $i, j, \dots, k$ non-negative integers (*uncorrelation property*); an element $\epsilon \in A$, called *augmentation*, such that $E[\epsilon^n] = 0$ for $n \geq 1$; an element $u \in A$, called *unity* umbra, such that $E[u^n] = 1$, for $n \geq 1$.

A polynomial $p \in R[A]$ is called *umbral polynomial*. The *support* of $p$ is the set of all umbrae occurring in $p$. Two umbral polynomials are said to be *uncorrelated* when their supports are disjoint.

When the evaluation $E$ is considered as expectation operator, an umbra carries the structure of a r.v., making no reference to its probability space. The *moments* of an umbra $\alpha$ are the elements $a_n \in R$ such that $E[\alpha^n] = a_n$ for $n \geq 1$. In that case, we said that the umbra $\alpha$ *represents* the sequence $1, a_1, a_2, \dots$. The *singleton umbra* $\chi$ is the umbra whose moments are all zero but the first, which is equal to 1. The *factorial moments* of an umbra $\alpha$ are the elements $a_{(n)} \in R$ corresponding to the umbral polynomials $(\alpha)_n = \alpha(\alpha-1)\cdots(\alpha-n+1)$, $n \geq 1$ via the evaluation $E$, i.e. $E[(\alpha)_n] = a_{(n)}$. The *Bell umbra* $\beta$ is the umbra such that $E[(\beta)_n] = 1$, $n \geq 1$. The umbra $\beta$ and the umbra $\chi$ provide respectively a symbolic tool for handling composition and inversion of formal power series; for a detailed exposition see [1] and [2].

Two umbrae $\alpha$ and $\gamma$ are said to be *similar* when $E[\alpha^n] = E[\gamma^n]$ for all $n \geq 1$, and we will write $\alpha \equiv \gamma$. So two similar umbrae represent the same moment sequence. A sequence $1, a_1, a_2, \dots$ in $R$ is represented by infinitely many distinct and thus similar umbrae, so one deals with moment products by means of the uncorrelation property. Indeed, we have $a_i\, a_j \neq E[\alpha^i \alpha^j]$ with $a_i = E[\alpha^i]$ and $a_j = E[\alpha^j]$, as well as $a_i\, a_j = E[\alpha^i \alpha'^j]$ with $\alpha \equiv \alpha'$ and $\alpha'$ uncorrelated with $\alpha$. So, given $n \in N$, the sequence

$$\sum_{k=0}^{n} \binom{n}{k} a_{n-k}\, a_k$$

gives the moments of $\alpha + \alpha'$.

Let $p$ and $q$ be two umbral polynomials. We said that $p$ is *umbrally equivalent* to $q$ if and only if $E[p] = E[q]$; in symbols, $p \simeq q$. This last equivalence turns out to be very useful in defining and handling generating functions of umbrae. We point out that all operations among umbrae correspond to analogous operations in the algebra of generating functions. Nevertheless, in the following we make no mention of generating functions (for a detailed exposition see [9]).

The notion of similarity allows us to extend the alphabet $A$ with the so-called *auxiliary umbrae* derived from operations among similar umbrae. That leads to the construction of a *saturated umbral calculus* in which auxiliary umbrae are treated as elements of the alphabet (cf. [6]). Let $\alpha', \alpha'', \dots, \alpha'''$ be $n$ uncorrelated umbrae similar to an umbra $\alpha$. The symbol $n.\alpha$ denotes an auxiliary umbra similar to the sum $\alpha' + \alpha'' + \cdots + \alpha'''$. The symbol $\alpha^{.n}$ denotes an auxiliary umbra similar to the product $\alpha'\alpha'' \cdots \alpha'''$. Properties of such auxiliary umbrae are extensively described in [1] and they will be recalled whenever it is necessary. We will assume both the support of $n.\alpha$, $m.\alpha$ and $\alpha^{.n}, \alpha^{.m}$ to be disjoint whenever $n \neq m$. If $p$ and $q$ are correlated umbral polynomials, then $n.p \simeq p_1 + \cdots + p_n$ is correlated with $n.q \simeq q_1 + \cdots + q_n$, and $p_i$ is correlated with $q_i$ but uncorrelated with $q_j$ with $i \neq j$. In [1], the following identity is stated:

$$E[(n.\alpha)^i] = \sum_{k=1}^{i} (n)_k B_{i,k}(a_1, a_2, \dots, a_{i-k+1}) \quad i \geq 1, \tag{1}$$

where $B_{i,k}$ are the (incomplete) exponential Bell polynomials and $a_i$ is the $i$-th moment of $\alpha$. Moreover, it is easy to verify that $E[(\alpha^{\cdot n})^i] = a_i^n$ for $i \geq 0$.

Two umbrae $\alpha$ and $\gamma$ are said to be *inverse* to each other when $\alpha + \gamma \equiv \varepsilon$. The inverse of the umbra $\alpha$ is denoted by $-1.\alpha$. Note that, in dealing with a saturated umbral calculus, the inverse of an umbra is not unique, but any two inverse umbrae of the same umbra are similar.

Replacing in $n.\alpha$ the integer $n$ with an umbra $\gamma$, we obtain the auxiliary umbra $\gamma.\alpha$ whose moments are

$$E[(\gamma.\alpha)^i] = \sum_{k=1}^{i} g_{(k)} B_{i,k}(a_1, a_2, \ldots, a_{i-k+1}) \quad i \geq 1, \tag{2}$$

where $g_{(k)}$ are the factorial moments of $\gamma$. In particular $\beta.\alpha$ is called $\alpha$*-partition umbra* and its moments are the (complete) exponential Bell polynomials (cf. [1]). Moreover $\chi.\alpha$ is called $\alpha$*-cumulant umbra* and $\alpha.\chi$ is called $\alpha$-*factorial umbra*, with moments equal to the factorial moments of $\alpha$ (cf. [2]). In particular we have

$$\beta.\chi \equiv u \equiv \chi.\beta. \tag{3}$$

Again, replacing in $\gamma.\alpha$ the umbra $\gamma$ with the umbra $\gamma.\beta$, we obtain the *composition* umbra of $\alpha$ and $\gamma$, i.e. $\gamma.\beta.\alpha$. The compositional inverse of an umbra $\alpha$ is the umbra $\alpha^{\langle -1\rangle}$ such that $\alpha^{\langle -1\rangle}.\beta.\alpha \equiv \chi \equiv \alpha.\beta.\alpha^{\langle -1\rangle}$. In particular we have

$$u^{\langle -1\rangle}.\beta \equiv \chi \equiv u^{\langle -1\rangle}.\beta, \tag{4}$$

where $u^{\langle -1\rangle}$ denotes the compositional inverse of $u$. Via the umbral Lagrange inversion formula (cf. [1]), the moments of $u^{\langle -1\rangle}$ are $E[(u^{\langle -1\rangle})^n] = (-1)^{n-1}(n-1)!$. Finally, we have

$$\chi.\chi \equiv u^{\langle -1\rangle} \qquad -\chi.(-\chi) \equiv (-u)^{\langle -1\rangle}. \tag{5}$$

The disjoint sum of $\alpha$ and $\gamma$ is the umbra whose moments are $a_n + g_n$, where $a_n$ and $g_n$ are the $n$-th moments of $\alpha$ and $\gamma$ respectively (cf. [2]); in symbols $(\alpha \dot{+} \gamma)^n \simeq \alpha^n + \gamma^n$. For instance, it turns out that

$$\chi.\alpha \dot{+} \chi.\gamma \equiv \chi.(\alpha + \gamma) \tag{6}$$

which is the well known additive property of cumulants. In the following, we denote the disjoint sum $\underbrace{\alpha \dot{+} \cdots \dot{+} \alpha}_{n}$ by $\dot{+}_n \alpha$.

## 3. Umbral symmetric polynomials

A partition of an integer $m$ is a sequence $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_t)$, where $\lambda_i$ are weakly decreasing and $\sum_{i=1}^t \lambda_i = m$. The integers $\lambda_i$ are said to be *parts* of $\lambda$. A different notation is $\lambda = (1^{r_1}, 2^{r_2}, \ldots)$, where $r_i$ is the number of parts of $\lambda$ equal to $i$. The *monomial symmetric polynomials* in the variables $\alpha_1, \alpha_2, \ldots, \alpha_n$ are $m_\lambda = \sum \alpha_1^{\lambda_1} \cdots \alpha_t^{\lambda_t}$, where the sum is over all distinct monomials having exponents $\lambda_1, \ldots, \lambda_t$. When $\lambda$ ranges over the set of partitions of the integer $m$, $m_\lambda$ is a basis for the algebra of the symmetric polynomials. There are other different bases; we recall just those necessary in the following: the $r$-th power sum symmetric polynomials $s_r = \sum_{i=1}^n \alpha_i^r$; the $k$-th elementary symmetric polynomials $e_k = \sum \alpha_{j_1}\alpha_{j_2} \cdots \alpha_{j_k}$, where the sum is over $1 \leq j_1 < j_2 < \cdots < j_k \leq n$, and the $r$-th complete homogeneous symmetric polynomials $h_i = \sum_{d_1+\cdots+d_n} = i \ \alpha_1^{d_1} \cdots \alpha_n^{d_n}$. It is not hard to see that $h_i = \sum_{|\lambda|=i} m_\lambda$, where $|\lambda| = \sum_j \lambda_j$.

When the umbrae $\alpha_1, \ldots, \alpha_n$ are uncorrelated and similar to each other, these four classical bases can be represented by means of the umbral polynomial $n.(\chi\alpha^i)$ and its moments. Propositions 1, 3 and 4 are stated under such a hypothesis. For power sum symmetric polynomials we have

$$s_r \simeq n.\alpha^r \simeq n.(\chi\alpha^r).$$

Note that $s_r$ are umbrally equivalent to the moments of $\dot{+}_n \alpha$ (cf. [2]):

**Proposition 1** (*Elementary Symmetric Polynomials*).

$$[n.(\chi\alpha)]^k \simeq \begin{cases} k!e_k, & k = 1, 2, \ldots, n, \\ 0, & k = n+1, n+2, \ldots. \end{cases} \tag{7}$$

**Proof.** For $k = 1, \ldots, n$ the result follows by applying the evaluation $E$ to the multinomial expansion of $[n.(\chi\alpha)]^k \simeq (\chi_1\alpha_1 + \cdots + \chi_n\alpha_n)^k$ and observing that the terms having powers of $\chi$ greater than 1 vanish. So just $k!$ monomials of the form $\chi_{j_1}\alpha_{j_1}\chi_{j_2}\alpha_{j_2}\cdots\chi_{j_k}\alpha_{j_k}$ have an evaluation that is not zero. Instead, for $k = n+1, n+2, \ldots$ the result follows by observing that at least one power of $\chi$ greater than 1 occurs in each monomial of the multinomial expansion. □

**Proposition 2.** *We have*

$$\chi.n.(\chi\alpha) \equiv u^{\langle -1 \rangle}(\dot{+}_n\alpha) \quad n.(\chi\alpha) \equiv \beta.[u^{\langle -1 \rangle}(\dot{+}_n\alpha)]. \tag{8}$$

**Proof.** The first equivalence in (8) follows from (6) on observing that

$$\chi.n.(\chi\alpha) \equiv \chi.(\chi_1\alpha_1 + \cdot + \chi_n\alpha_n) \equiv \dot{+}_n\chi.(\chi\alpha) \equiv \dot{+}_n u^{\langle -1 \rangle}\alpha.$$

The second one follows from the first by taking the right dot product with $\beta$ and recalling (3). □

Eq. (8) are the umbral versions of the well known relations between power sum symmetric polynomials $s_r$ and elementary symmetric polynomials $e_k$. Indeed the umbra $(\dot{+}_n\alpha)$ represents the sequence $\{s_r\}$ and the umbra $n.(\chi\alpha)$ represents the sequence $\{e_k\}$. The umbral expression of $m_\lambda$ requires the introduction of the augmented monomial symmetric polynomials $\tilde{m}_\lambda$. Let $\lambda = (1^{r_1}, 2^{r_2}, \ldots)$ be a partition of the integer $m$; such polynomials are defined by

$$\tilde{m}_\lambda = \sum_{1 \leq j_1 \neq \cdots \neq j_{r_1} \neq j_{r_1+1} \neq \cdots \neq j_{r_1+r_2} \cdots \leq n} \alpha_{j_1} \cdots \alpha_{j_{r_1}} \alpha^2_{j_{r_1+1}} \cdots \alpha^2_{j_{r_1+r_2}} \cdots. \tag{9}$$

**Proposition 3.** *It is*

$$\tilde{m}_\lambda \simeq [n.(\chi\alpha)]^{r_1}[n.(\chi\alpha^2)]^{r_2}\cdots.$$

**Proof.** Analogous to the proof of Proposition 1. □

Recalling that $m_\lambda = \tilde{m}_\lambda/[r_1!r_2!\cdots]$ it follows that

$$m_\lambda \simeq \frac{[n.(\chi\alpha)]^{r_1}}{r_1!}\frac{[n.(\chi\alpha^2)]^{r_2}}{r_2!}\cdots. \tag{10}$$

**Proposition 4** (*Complete Symmetric Polynomials*).

$$[-n.(-\chi\alpha)]^m \simeq m!h_m \qquad m = 1, 2, \ldots. \tag{11}$$

**Proof.** We have

$$\begin{aligned}[-n.(-\chi\alpha)]^m &\simeq [-1.(-\chi_1\alpha_1) + \cdots + -1.(-\chi_n\alpha_n)]^m \\ &\simeq m! \sum_{|\lambda|=m} [-1.(-\chi')]^{.r_1}([-1.(-\chi'')]^2)^{.r_2}\cdots \frac{\tilde{m}_\lambda}{(1!)^{r_1}r_1!(2!)^{r_2}r_2!\cdots}\end{aligned}$$

and the result follows from (10), since $([-1.(-\chi)]^i)^{.r_i} \simeq (i!)^{r_i}$. □

Equivalences (7) and (11) are the umbral versions of the well known identities

$$\sum_k e_k t^k = \prod_{i=1}^n (1 + \alpha_i t) \qquad \sum_k h_k t^k = \frac{1}{\prod_{i=1}^n (1 - \alpha_i t)}.$$

**Proposition 5.** *We have*

$$-\chi.n.(-\chi\alpha) \equiv (-u)^{\langle -1 \rangle}(\dot{+}_n\alpha) \qquad -n.(-\chi\alpha) \equiv \beta.[(-u)^{\langle -1 \rangle}(\dot{+}_n\alpha)]. \tag{12}$$

**Proof.** The results follow from (8) on replacing the umbra $\chi$ with $-\chi$ and recalling that $u^{\langle -1\rangle} \equiv \chi.\chi$ must be replaced with $(-u)^{\langle -1\rangle} \equiv -\chi.(-\chi)$. □

Eq. (12) are the umbral versions of the well known relations between power sum symmetric polynomials $s_r$ and complete symmetric polynomials $h_k$. Indeed the umbra $(\dot{+}_n \alpha)$ represents the sequence $\{s_r\}$ and the umbra $-n.(-\chi\alpha)$ represents the sequence $\{h_k\}$.

## 4. $U$-Statistics

Let $X_1, X_2, \ldots, X_n$ be $n$ independent r.v.'s. A statistic of the form

$$U = \frac{1}{(n)_k} \sum \Phi(X_{j_1}, X_{j_2}, \ldots, X_{j_k})$$

where the sum ranges over the set of all permutations $(j_1, j_2, \ldots, j_k)$ of $k$ integers, $1 \leq j_i \leq n$, is called *$U$-statistic* [5]. If $X_1, X_2, \ldots, X_n$ have the same cumulative distribution function $F(x)$, $U$ is an unbiased estimator of the population character $\theta(F) = \int \cdots \int \Phi(x_1, \ldots, x_k)\, \mathrm{d}F(x_1) \cdots \mathrm{d}F(x_k)$. In that case, the function $\Phi$ may be assumed to be a symmetric function of its arguments. Often, in the applications, $\Phi$ is a polynomial in the $X_i$'s, so the $U$-statistic is a symmetric polynomial. By virtue of the fundamental theorem on symmetric polynomials, the $U$-statistic can be considered a polynomial in the elementary symmetric polynomials. The following theorem is an umbral reformulation of the above statement.

**Theorem 4.1** ($U$-*Statistic*). *If $\lambda = (1^{r_1}, 2^{r_2}, \ldots,)$ is a partition of the integer $m \leq n$ then*

$$(\alpha_{j_1})^{.r_1} (\alpha^2_{j_2})^{.r_2} \cdots \simeq \frac{1}{(n)_k} [n.(\chi\alpha)]^{r_1} [n.(\chi\alpha^2)]^{r_2} \cdots \tag{13}$$

*where $j_i \in \{1, 2, \ldots, n\}$ and $\sum_j r_j = k$.*

**Proof.** The result follows on observing that $[n.(\chi\alpha^i)]^{r_i} \simeq (n.\chi)^{r_i} (\alpha^i)^{.r_i}$ are distinct integers and $(n.\chi)^k \simeq (n)_k$. □

Note that equivalence (13) states how to estimate moment products by using only $n$ information drawn out of the population. Then the symmetric polynomial on the right side of (13) is the *$U$-statistic* of the uncorrelated and similar umbrae $\alpha_1, \alpha_2, \ldots, \alpha_n$ associated with $(\alpha_{j_1})^{.r_1} (\alpha^2_{j_2})^{.r_2} \cdots$.

**Example 4.1** (*Powers of Moments*). Set $r_1 = 2$ and $k = 2$; from (13) the $U$-statistic associated with $\alpha^{.2} \simeq a_1^2$ is

$$\alpha^{.2} \simeq \frac{1}{n(n-1)} [n.(\chi\alpha)]^2 \simeq \frac{1}{n(n-1)} \sum_{i \neq j} \alpha_i \alpha_j, \qquad n \geq 2.$$

**Example 4.2** (*h-Statistics*). As shown in [2], we have

$$(\alpha^{a_1})^r \simeq \sum_{k=0}^{r} \binom{r}{k} (-1)^k \alpha^{.k} (\alpha')^{r-k} \simeq \sum_{k=0}^{r-2} \frac{(-1)^k}{r-k} \sum_{k=0}^{r} \alpha^{.k} (\alpha')^{r-k},$$

where $\alpha^{a_1}$ is the central umbra of $\alpha$ about $a_1 = E[\alpha]$ and $\alpha' \equiv \alpha$ is an umbra uncorrelated with $\alpha$. Replacing the product $\alpha^{.k}(\alpha')^{r-k}$ with the corresponding $U$-statistic (13), the result is

$$(\alpha^{a_1})^r \simeq \sum_{k=0}^{r} \binom{r}{k} \frac{(-1)^k}{(n)_{k+1}} [n.(\chi\alpha)]^k n.(\chi\alpha^{r-k}). \tag{14}$$

When the products $[n.(\chi\alpha)]^k n.(\chi\alpha^{r-k})$ are expressed in terms of uncorrelated power sum symmetric polynomials, we have the so-called $h$-statistics. About this key point, we will give more details in the next section.

## 5. *k*-Statistics

The $i$-th $k$-statistic $k_i$ is the unique symmetric unbiased estimator of the $i$-th cumulant $\kappa_i$ of a given statistical distribution, i.e. $E[k_i] = \kappa_i$. In this section we give an umbral syntax that provides a general computational method for generating the expressions of $k$-statistics in terms of sums of the $r$-th powers of the data points. To this end, we digress to introduce the exponential Bell umbral polynomials.

The most widespread expression for incomplete exponential Bell polynomials refers to partitions of an integer. Of course, it is also possible to express such polynomials referring them to partitions of a set. Here we follow this last point of view. We denote by $\Pi_{i,k}$ the set of all partitions of the set $[i] = \{1, 2, \ldots, i\}$ into $k$ blocks. Let $\pi = \{A_1, A_2, \ldots, A_k\}$ be an element of $\Pi_{i,k}$. Then we have $B_{i,k}(a_1, a_2, \ldots) = \sum_{\pi \in \Pi_{i,k}} a_{n_1} a_{n_2} \ldots a_{n_k}$ where $|A_j| = n_j$, $j = 1, 2, \ldots, k$, as we will suppose from now on. Let us consider the following symmetric umbral polynomial:

$$\mathcal{B}_{i,k}(\alpha_1, \alpha_2, \ldots, \alpha_k) = \sum_{\pi \in \Pi_{i,k}} \alpha_1^{n_1} \alpha_2^{n_2} \cdots \alpha_k^{n_k}, \tag{15}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_k$ are uncorrelated umbrae similar to $\alpha$ and $a_{n_j} = E[\alpha^{n_j}]$, $j = 1, 2, \ldots, k$. Obviously we have $E[\mathcal{B}_{i,k}] = B_{i,k}$, so any expression containing the polynomials $B_{i,k}$ could be replaced with an umbrally equivalent expression containing the polynomials $\mathcal{B}_{i,k}$. The polynomials $\mathcal{B}_{i,k}$ will be called (incomplete) umbral exponential Bell polynomials. The combinatorics underlying the polynomial $\mathcal{B}_{i,k}$ is the following: the set $[i]$ is partitioned into $k$ blocks; with each of them one associates the umbra $\alpha^{n_j}$ obtained by firstly replacing the elements in the $j$-th block with the umbra $\alpha$ and then labelling all blocks so that powers belonging to different blocks become uncorrelated. Replacing in (15) the products $\alpha_1^{n_1} \alpha_2^{n_2} \cdots \alpha_k^{n_k}$ with the umbrally corresponding $U$-statistic (13), we get for $i \le n$

$$\mathcal{B}_{i,k}(\alpha_1, \alpha_2, \ldots, \alpha_k) \simeq \sum_{\pi \in \Pi_{i,k}} \frac{1}{(n)_k} n.(\chi\, \alpha^{n_1})\, n.(\chi\, \alpha^{n_2}) \cdots n.(\chi\, \alpha^{n_k}), \tag{16}$$

by which we are able to give the umbral $k$-statistics. Indeed, the $\alpha$-cumulant umbra $\chi.\alpha$ represents the sequence of cumulants $k_i$ (cf. [2]) so, since $(\chi)_k \simeq (u^{\langle -1 \rangle})^k$, from (2) and (16) we have

$$(\chi.\alpha)^i \simeq \sum_{k=1}^{i} (-1)^{k-1} \frac{(k-1)!}{(n)_k} \sum_{\pi \in \Pi_{i,k}} n.(\chi\, \alpha^{n_1})\, n.(\chi\, \alpha^{n_2}) \cdots n.(\chi\, \alpha^{n_k}). \tag{17}$$

Since $n.(\chi\alpha^{n_i})$ is umbrally equivalent to a symmetric power sum polynomial, Eq. (17) gives the moments of the $\alpha$-cumulant umbra in terms of power sum polynomials, i.e. the umbral form of the $k$-statistics. Note that the symmetric power sum polynomials in (17) are correlated. So, in order to make formula (17) effective, we need a device with which to evaluate the product $n.(\chi\, \alpha^{n_1})\, n.(\chi\, \alpha^{n_2}) \cdots n.(\chi\, \alpha^{n_k})$. To this end, by using the umbral exponential Bell polynomials (15), the moments of the umbra $n.(\chi\alpha)$ can be evaluated from (2) recalling $n.\alpha^{n_i} \simeq (\dot{+}_n \alpha)^{n_i}$ and the second equivalence in (8):

$$[n.(\chi\alpha)]^i \simeq \sum_{k=1}^{i} \sum_{\pi \in \Pi_{i,k}} (u'^{\langle -1 \rangle})^{n_1} (u''^{\langle -1 \rangle})^{n_2} \cdots (u'''^{\langle -1 \rangle})^{n_k} [n.\alpha'^{n_1}]\, [n.\alpha''^{n_2}] \cdots [n.\alpha'''^{n_k}]. \tag{18}$$

The previous umbral equivalence is suitable for being generalized to the product of umbral polynomials $[n.(\chi p_1)][n.(\chi p_2)] \cdots [n.(\chi p_i)]$. Indeed, the result is

$$[n.(\chi p_1)] \cdots [n.(\chi p_i)] \simeq \sum_{k=1}^{i} \sum_{\pi \in \Pi_{i,k}} (u'^{\langle -1 \rangle})^{n_1} \cdots (u''^{\langle -1 \rangle})^{n_k} [n.P'_{A_1}] \cdots [n.P''_{A_k}], \tag{19}$$

where $P_{A_j} = \prod_{t=1}^{n_j} p_{j_t}$ and $p_{j_t}$ are polynomials indexed by the elements of the block $A_j$, as we will suppose from now on. Equivalence (19) is the device required to generate the $k$-statistics. Indeed, setting $i = k$ and $p_t = \alpha^{n_t}$ for $t = 1, 2, \ldots, i$, the result is

$$[n.(\chi\alpha^{n_1})] \cdots [n.(\chi\alpha^{n_k})] \simeq \sum_{j=1}^{k} \sum_{\pi \in \Pi_{k,j}} (u'^{\langle -1 \rangle})^{n_1} \cdots (u''^{\langle -1 \rangle})^{n_j} [n.\alpha'^{m_1}] \cdots [n.\alpha''^{m_j}], \tag{20}$$

where $m_j = \sum_{t=1}^{n_j} n_{j_t}$ and $n_{j_t}$ are indexed by the elements of the block $A_j$. Note that the power sum polynomials on the right side of (20) are now uncorrelated. Equivalence (20) gives augmented monomial symmetric polynomials in terms of power sum polynomials and translates Kendall and Stuart tables read downwards [8]. Instead, the following formula translates Kendall and Stuart tables read across (i.e. power sum polynomials in terms of augmented monomial ones):

$$(n.p_1)\cdots(n.p_i) \simeq \sum_{k=1}^{i}(n)_k \sum_{\pi\in\Pi_{i,k}} P'_{A_1}\cdots P''_{A_k} \simeq \sum_{k=1}^{i}\sum_{\pi\in\Pi_{i,k}} n.(\chi P_{A_1})\cdots n.(\chi P_{A_k}). \tag{21}$$

The first equivalence in (21) is obtained from (1) through considerations analogous to those used to state (19); the second equivalence comes from (13) replacing $\alpha_i^{j_i}$ with the umbral polynomial $P_{A_i}$, i.e.

$$P'_{A_1}P''_{A_2}\cdots P'''_{A_k} \simeq \frac{1}{(n)_k} n.(\chi P_{A_1})n.(\chi P_{A_2})\cdots n.(\chi P_{A_k}).$$

Setting in (21) $p_t = \alpha^{n_t}$ for $t = 1, 2, \ldots, i$, we have

$$(n.\alpha^{n_1})\cdots(n.\alpha^{n_i}) \simeq \sum_{k=1}^{i}\sum_{\pi\in\Pi_{i,k}} (n.\chi\alpha^{m_1})(n.\chi\alpha^{m_2})\cdots(n.\chi\alpha^{m_k}), \tag{22}$$

where $m_j = \sum_{t=1}^{n_j} n_{j_t}$ and $n_{j_t}$ are indexed by the elements of the block $A_j$.

**Example 5.1** (*h-Statistics*). In (20) set $n_1 = n - k$ and $n_2 = \cdots = n_{k+1} = 1$; from (14) we get the umbral expression of the $h$-statistics.

**Example 5.2** (*Joint Cumulants*). Let $p_1, p_2, \ldots, p_i$ be umbral polynomials. In the first equivalence of (21), replacing $n$ with $\chi$ we have

$$(\chi.p_1)(\chi.p_2)\cdots(\chi.p_i) \simeq \sum_{k=1}^{i}(\chi)_k \sum_{\pi\in\Pi_{i,k}} P'_{A_1}\cdots P''_{A_k}. \tag{23}$$

When the umbral polynomials $p_i$ are interpreted as r.v.'s, equivalence (23) gives their joint cumulants. So we will call $(\chi.p_1)(\chi.p_2)\cdots(\chi.p_i)$ the joint cumulant of $p_1, \ldots, p_i$. Note that, setting $p_t = \alpha$ for $t = 1, 2, \ldots, i$, one has the $i$-th ordinary cumulant $(\chi.\alpha)^i$. Through this equivalence there results $\chi.(p_1 + \cdots + p_i) \equiv \chi.p_1 + \cdots + \chi.p_i$. Now suppose we split the set $\{p_1, p_2, \ldots, p_i\}$ into two subsets $\{p_{j_1}, \ldots, p_{j_t}\}$ and $\{p_{k_1}, \ldots, p_{k_s}\}$ with $s + t = i$, such that polynomials belonging to different subsets are uncorrelated. Then we have

$$(\chi.p_1)(\chi.p_2)\cdots(\chi.p_i) \simeq 0. \tag{24}$$

Indeed, setting $P = \sum_{l=1}^{t} p_{j_l}$ and $Q = \sum_{l=1}^{s} p_{k_l}$, such polynomials are uncorrelated, so $\chi.(P + Q) \equiv \chi.P \dot{+} \chi.Q$ from (6). Equivalence (24) follows on observing that, due to the disjoint sum, products involving powers of $\chi.P$ and $\chi.Q$ vanish. When the umbral polynomials $p_i$ are interpreted as r.v.'s, equivalence (24) states the following well known result: if some of the r.v.'s are uncorrelated with all others, then their joint cumulant is zero.

**Example 5.3** (*Multivariate k-Statistics*). Equivalence (21) allows compact expression of the multivariate $k$-statistics. In the second equivalence of (21), replacing $n$ with $\chi$ we construct the $U$-statistic of the joint cumulant:

$$(\chi.p_1)(\chi.p_2)\cdots(\chi.p_i) \simeq \sum_{k=1}^{i}(-1)^{k-1}\frac{(k-1)!}{(n)_k}\sum_{\pi\in\Pi_{i,k}} n.(\chi P_{A_1})n.(\chi P_{A_2})\cdots n.(\chi P_{A_k}). \tag{25}$$

Again, in the product on the right side of (25) the umbral polynomials are correlated. In order to make the computation effective, it is necessary to rewrite (25) by using equivalence (19) with $P_{A_t}$ instead of $p_t$. For instance, in order to express $k_{21}$, set in (25) $i = 3$ and $p_1 = p_2 = \alpha_1$, $p_3 = \alpha_2$. The result is

$$\begin{aligned}(\chi.\alpha_1)^2(\chi.\alpha_2) \simeq{}& \frac{\chi}{n} n.(\chi\alpha_1^2\alpha_2) + \frac{(\chi)_2}{(n)_2}\{2\, n.(\chi\alpha_1)\, n.(\chi\alpha_1\alpha_2) + n.(\chi\alpha_1^2)\, n.(\chi\alpha_2)\}\\ &+ \frac{(\chi)_3}{(n)_3}\{[n.(\chi\alpha_1)]^2\, n.(\chi\alpha_2)\}.\end{aligned} \tag{26}$$

Set $s_{p,q} \simeq n.(\alpha_1^p\,\alpha_2^q)$. We have

$$n.(\chi\alpha_1)\,n.(\chi\alpha_1\alpha_2) \simeq (u^{\langle -1\rangle})^2 n.(\alpha_1^2\alpha_2) + (u^{\langle -1\rangle})^{.2} n.\alpha_1'\,n.(\alpha_1\alpha_2) \simeq -s_{2,1} + s_{1,0}\,s_{1,1} \tag{27}$$

$$n.(\chi\alpha_1^2)\,n.(\chi\alpha_2) \simeq (u^{\langle -1\rangle})^2 n.(\alpha_1^2\alpha_2) + (u^{\langle -1\rangle})^{.2} n.\alpha_1'^2\,n.\alpha_2 \simeq -s_{2,1} + s_{2,0}\,s_{0,1} \tag{28}$$

$$\begin{aligned}\{n.(\chi\alpha_1)\}^2 n.(\chi\alpha_2) \;&\simeq\; (u^{\langle -1\rangle})^3 n.(\alpha_1^2\alpha_2) + (u^{\langle -1\rangle})^{.3} n.\alpha_1'\,n.\alpha_1\,n.\alpha_2 \\ &\quad + u'^{\langle -1\rangle}\,(u^{\langle -1\rangle})^2[n.\alpha_1^2\,n.\alpha_2 + 2\,n.\alpha_1'\,n.(\alpha_1\alpha_2)] \\ &\simeq\; 2\,s_{2,1} - s_{2,0}\,s_{0,1} - 2\,s_{1,0}\,s_{1,1} + s_{1,0}^2\,s_{0,1}.\end{aligned} \tag{29}$$

Equivalence (27) comes from (19), setting $i = 2$, $p_1 = \alpha_1$ and $p_2 = \alpha_1\alpha_2$; equivalence (28) comes from (19) setting $i = 2$, $p_1 = \alpha_1^2$ and $p_2 = \alpha_2$; equivalence (29) comes from (19) setting $i = 3$, $p_1 = p_2 = \alpha_1$ and $p_3 = \alpha_2$. Substituting the above equivalences in (26) and rearranging the terms, we have the expression for $k_{21}$:

$$k_{21} \simeq (\chi.\alpha_1)^2(\chi.\alpha_2) \simeq \frac{1}{(n)_3}[n^2 s_{2,1} - 2n\,s_{1,0}\,s_{1,1} - n\,s_{2,0}\,s_{0,1} + 2\,s_{1,0}^2\,s_{0,1}].$$

The expressions for generalized $k$-statistics (as well as the multivariate ones) in terms of power sums come from (23) on replacing some of the umbrae $\chi$ with uncorrelated ones and then constructing the corresponding $U$-statistics.

## References


[1] E. Di Nardo, D. Senato, Umbral nature of the Poisson random variables, in: H. Crapo, D. Senato (Eds.), Algebraic Combinatorics and Computer Science, Springer, Italia, 2001, pp. 245–266.

[2] E. Di Nardo, D. Senato, An umbral setting for cumulants and factorial moments, European J. Combin. (2005) (in press).

[3] P. Doubilet, On the foundations of combinatorial theory VII: Symmetric functions through the theory of distribution and occupancy, Stud. Appl. Math. 11 (1972) 377–396.

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

[5] W. Hoeffding, A class of statistics with asymptotically normal distribution, Ann. Math. Stat. 19 (1948) 293–325.

[6] G.-C. Rota, B.D. Taylor, The classical umbral calculus, SIAM J. Math. Anal. 25 (1994) 694–711.

[7] T.P. Speed, Cumulants and partition lattices II: Generalized $k$-statistics, J. Aust. Math. Soc. Ser. A 40 (1986) 34–53.

[8] A. Stuart, J.K. Ord, Kendall's Advanced Theory of Statistics, vol. 1, Charles Griffin and Company Limited, London, 1987.

[9] B.D. Taylor, Difference equations via the classical umbral calculus, in: Sagan et al. (Eds.), Mathematical Essays in Honor of Gian-Carlo Rota, Birkhauser, Boston, 1998, pp. 397–411.