A NEW METHOD FOR FAST COMPUTING UNBIASED
ESTIMATORS OF CUMULANTS

E. DI NARDO, G. GUARINO, AND D. SENATO

ABSTRACT. We propose new algorithms for generating $k$-statistics, multivariate $k$-statistics, polykays and multivariate polykays. The resulting computational times are very fast compared with procedures existing in the literature. Such speeding up is obtained by means of a symbolic method arising from the classical umbral calculus. The classical umbral calculus is a light syntax that involves only elementary rules to managing sequences of numbers or polynomials. The cornerstone of the procedures here introduced is the connection between cumulants of a random variable and a suitable compound Poisson random variable. Such a connection holds also for multivariate random variables.

1. Introduction

In the last decades, symbolic methods have been successfully used in different mathematical areas (Grossman, 1989; Wang and Zheng, 2005). Symbolic techniques have been recently used in problems arising from computational statistics (see Andrews and Stafford (2000), Zeilberger (2004)). The papers (Di Nardo and Senato, 2006b) and (Di Nardo et al., 2008b) lie within this field. In these papers, the theory of $k$-statistics and polykays was completely rewritten, carrying out a unifying framework for these estimators, both in the univariate and multivariate cases.

This subject goes back to Fisher (1929) and up to today it was treated by means of different languages. Main references are Stuart and Ord (1987), Speed (1983, 1986), McCullagh (1987). A more accurate list of references can be found in Di Nardo et al. (2008b).

The umbral techniques, investigated in Di Nardo et al. (2008b), have allowed us to implement a single algorithm for $k$-statistics, multivariate $k$-statistics, polykays and multivariate polykays (Di Nardo et al. 2008b). Nevertheless, the elegance of a unifying outlook pays a price in computational costs that become comparable with those of MathStatica (Rose and Smith, 2002) for polykays and not competitive for univariate and multivariate $k$-statistics.

By developing the ideas introduced in Di Nardo et al. (2008b) for $k$-statistics, in this paper we introduce radically innovative procedures for generating all these estimators, by realizing a substantial improvement of computational times compared with those in the literature.

A frequently asked question is: why are these calculations relevant? Usually higher order objects require enormous amounts of data to estimate with any accuracy. Nevertheless, there are different areas, such as astronomy, astrophysics and

Key words and phrases. univariate and multivariate $k$-statistics, univariate and multivariate polykays, umbral calculus.

AMS-Primary: 65C60, 05A40, AMS-Secondary: 68W30, 62H99.
biophysics, which need to compute high order \( k \)-statistics in order to recognizing a gaussian population (Ferreira et al., 1997). Undoubtedly, an enjoyable challenge is to have efficient procedures to deal with the involved huge amount of algebraic and symbolic computations.

The algorithms proposed here are based on the umbral language introduced by Rota and Taylor (1994). Applications of the classical umbral calculus are given in Zeilberger (2000) ÷ (2002), where generating functions are computed for many difficult problems dealing with counting combinatorial objects. Applications to bilinear generating functions for polynomial sequences are given in Gessel (2003). Di Nardo and Senato (2001) have developed the classical umbral calculus (1994), with special care to probabilistic aspects. The basic device is the representation of a unital sequence of numbers or polynomials by a symbol \( \alpha \), called an umbra. The umbra \( \alpha \) is related to these unital sequences via an operator \( E \) that resembles the expectation operator of random variables (r.v.’s). This symbolic method provides a light syntax for handling cumulants and factorial moments (Di Nardo and Senato, 2006a), \( k \)-statistics then come in hand since these are the unique symmetric unbiased estimators of cumulants (Di Nardo and Senato, 2006b). These estimators are expressed in terms of power sums in the variables of the random sample.

After recalling basic notions of the umbral language, in Section 3 we show that any cumulant can be evaluated via cumulants of a suitable umbra. In probabilistic terms, cumulants of a r.v. can be obtained via cumulants of a suitable compound Poisson r.v. This link allows us the significant speed up of the algorithms for \( k \)-statistics and polykays. For multivariate cumulants, the basic tool is given by umbrae indexed by multisets. In Section 4, we recall this symbolic device, introduced and largely used in Di Nardo et al. (2008b). In Section 5, we show the connection between multivariate cumulants of an umbra and a suitable multivariate compound Poisson r.v. We then summarize the algorithms for generating multivariate \( k \)-statistics and polykays. Finally, we compare the computational times of the algorithms proposed here with those of MathStatica (Rose and Smith, 2002) and those of Andrews and Stafford (2000). Note that MathStatica does not have a procedure for multivariate polykays. All programs have been executed on a PC Pentium(R)4 Intel(R), CPU 2.08 Ghz, 512MB Ram with Maple version 10.0 and Mathematica version 4.2. We choose the Maple language due to its acknowledged plainness in translating symbolic computations.

2. Background to umbral calculus

This section is aimed to recalling notation and terminology useful to handle umbrae. More details and technicalities can be found in Di Nardo and Senato (2001, 2006a).

Formally, an umbral calculus is a syntax consisting of the following data:

i) a set \( A = \{\alpha, \beta, \ldots\} \), called the alphabet, whose elements are named umbrae;

ii) a commutative integral domain \( R \) whose quotient field is of characteristic zero;

iii) a linear functional \( E \), called the evaluation, defined on the polynomial ring \( R[A] \) and taking values in \( R \) such that

a) \( E[1] = 1 \);

\(^1\)For statistical applications, \( R \) is the field of real numbers.
An umbral polynomial is a polynomial \( p \in R[A] \). The support of \( p \) is the set of all umbrae occurring in \( p \). If \( p \) and \( q \) are two umbral polynomials, \( p \) and \( q \) are uncorrelated if and only if their supports are disjoint. The umbral polynomials \( p \) and \( q \) are umbrally equivalent if and only if
\[
E[p] = E[q], \quad \text{in symbols } p \simeq q.
\]

The moments of an umbra \( \alpha \) are the elements \( a_n \in R \) such that
\[
E[\alpha^n] = a_n, \quad \forall n \geq 0
\]
and we say that the umbra \( \alpha \) represents the sequence of moments \( 1, a_1, a_2, \ldots \).

It is possible that two distinct umbrae represent the same sequence of moments, in this case they are called similar umbrae. More formally, two umbrae \( \alpha \) and \( \gamma \) are said to be similar when
\[
E[\alpha^n] = E[\gamma^n] \quad \forall n \geq 0, \quad \text{in symbols } \alpha \equiv \gamma.
\]

In addition, given a sequence \( 1, a_1, a_2, \ldots \) in \( R \) there are infinitely many distinct and thus similar umbrae representing the sequence.

The factorial moments of an umbra \( \alpha \) are the elements \( a_{(n)} \in R \) corresponding to umbral polynomials \( \langle \alpha \rangle_n = \alpha(\alpha - 1) \cdots (\alpha - n + 1) \), for each \( n \geq 1 \) via the evaluation \( E \), that is \( E[\langle \alpha \rangle_n] = a_{(n)} \).

Two more special umbrae have been defined in the alphabet \( A \): the singleton umbra \( \chi \) and the Bell umbra \( \beta \).

The singleton umbra \( \chi \) is the umbra whose moments are all zero, except the first \( E[\chi] = 1 \). As shown in Di Nardo and Senato (2006a), its factorial moments are
\[
E[\langle \chi \rangle_n] = x_{(n)} = (-1)^{n-1}(n-1)!, \quad \forall n \geq 1.
\]
The Bell umbra \( \beta \) is the umbra whose factorial moments are all equal to 1,
\[
E[\langle \beta \rangle_n] = b_{(n)} = 1, \quad \forall n \geq 1.
\]

Its moments are the Bell numbers. The umbra \( \beta \) is therefore the umbral counterpart of a Poisson r.v. with parameter 1 (Di Nardo and Senato, 2001).

Thanks to the notion of similar umbrae, it is possible to extend the alphabet \( A \) with the so-called auxiliary umbrae, obtained via operations among similar umbrae. As a consequence, a saturated umbral calculus can be constructed where auxiliary umbrae are treated as elements of the alphabet (Rota and Taylor, 1994). Let \( \{\alpha_1, \alpha_2, \ldots, \alpha_n\} \) be a set of \( n \) uncorrelated umbrae, similar to an umbra \( \alpha \). The symbol \( n.\alpha \) denotes an auxiliary umbra similar to the sum \( \alpha_1 + \alpha_2 + \cdots + \alpha_n \) and called the dot product between the integer \( n \) and the umbra \( \alpha \). Powers of \( n.\alpha \) are umbrally equivalent to the following umbral polynomials (Di Nardo et al., 2008a):
\[
(n.\alpha)^i = \sum_{\lambda^i \theta} (n)_{v_\lambda} d_\lambda \alpha_\lambda,
\]

An umbral polynomial is a polynomial \( p \in R[A] \). The support of \( p \) is the set of all umbrae occurring in \( p \). If \( p \) and \( q \) are two umbral polynomials, \( p \) and \( q \) are uncorrelated if and only if their supports are disjoint. The umbral polynomials \( p \) and \( q \) are umbrally equivalent if and only if
\[
E[p] = E[q], \quad \text{in symbols } p \simeq q.
\]

The moments of an umbra \( \alpha \) are the elements \( a_n \in R \) such that
\[
E[\alpha^n] = a_n, \quad \forall n \geq 0
\]
and we say that the umbra \( \alpha \) represents the sequence of moments \( 1, a_1, a_2, \ldots \).

It is possible that two distinct umbrae represent the same sequence of moments, in this case they are called similar umbrae. More formally, two umbrae \( \alpha \) and \( \gamma \) are said to be similar when
\[
E[\alpha^n] = E[\gamma^n] \quad \forall n \geq 0, \quad \text{in symbols } \alpha \equiv \gamma.
\]

In addition, given a sequence \( 1, a_1, a_2, \ldots \) in \( R \) there are infinitely many distinct and thus similar umbrae representing the sequence.

The factorial moments of an umbra \( \alpha \) are the elements \( a_{(n)} \in R \) corresponding to umbral polynomials \( \langle \alpha \rangle_n = \alpha(\alpha - 1) \cdots (\alpha - n + 1) \), for each \( n \geq 1 \) via the evaluation \( E \), that is \( E[\langle \alpha \rangle_n] = a_{(n)} \).

Two more special umbrae have been defined in the alphabet \( A \): the singleton umbra \( \chi \) and the Bell umbra \( \beta \).

The singleton umbra \( \chi \) is the umbra whose moments are all zero, except the first \( E[\chi] = 1 \). As shown in Di Nardo and Senato (2006a), its factorial moments are
\[
E[\langle \chi \rangle_n] = x_{(n)} = (-1)^{n-1}(n-1)!, \quad \forall n \geq 1.
\]
The Bell umbra \( \beta \) is the umbra whose factorial moments are all equal to 1,
\[
E[\langle \beta \rangle_n] = b_{(n)} = 1, \quad \forall n \geq 1.
\]

Its moments are the Bell numbers. The umbra \( \beta \) is therefore the umbral counterpart of a Poisson r.v. with parameter 1 (Di Nardo and Senato, 2001).

Thanks to the notion of similar umbrae, it is possible to extend the alphabet \( A \) with the so-called auxiliary umbrae, obtained via operations among similar umbrae. As a consequence, a saturated umbral calculus can be constructed where auxiliary umbrae are treated as elements of the alphabet (Rota and Taylor, 1994). Let \( \{\alpha_1, \alpha_2, \ldots, \alpha_n\} \) be a set of \( n \) uncorrelated umbrae, similar to an umbra \( \alpha \). The symbol \( n.\alpha \) denotes an auxiliary umbra similar to the sum \( \alpha_1 + \alpha_2 + \cdots + \alpha_n \) and called the dot product between the integer \( n \) and the umbra \( \alpha \). Powers of \( n.\alpha \) are umbrally equivalent to the following umbral polynomials (Di Nardo et al., 2008a):
\[
(n.\alpha)^i = \sum_{\lambda^i \theta} (n)_{v_\lambda} d_\lambda \alpha_\lambda,
\]
where the sum is over all partitions \((\lambda = (1^{s_1}, 2^{s_2}, \ldots))\) of the integer \(i\), \((n)_{\nu_{\lambda}} = 0\) for \(\nu_{\lambda} > n\) and

\[
(2.4) \quad d_{\lambda} = \frac{i!}{r_1!r_2! \cdots} \left(\frac{1}{(1!)^{r_1}(2!)^{r_2}} \cdots \right) \quad \text{and} \quad \alpha_{\lambda} \equiv (a_{j_1})^{r_1}(a_{j_2})^{r_2} \cdots ,
\]

with \(\{j_i\}\) distinct integers chosen in \(\{1, 2, \ldots, n\}\). By evaluating equivalence (2.3) via the linear functional \(E\), we have

\[
(2.5) \quad E[(n.\alpha)^i] = \sum_{\lambda \vdash i} (n)_{\nu_{\lambda}} d_{\lambda} \alpha_{\lambda},
\]

where \(a_{\lambda} = a_1^{r_1} a_2^{r_2} \cdots\). Note that if \(\lambda = (1^{r_1}, 2^{r_2}, \ldots)\) is a partition of the integer \(r\), \(\eta = (1^{s_1}, 2^{s_2}, \ldots)\) is a partition of the integer \(s\) and \(\lambda + \eta = (1^{r_1+s_1}, 2^{r_2+s_2}, \ldots)\), then

\[
\alpha_{\lambda+\eta} = a_{\lambda} a_{\eta} \quad \text{and} \quad \alpha_{\lambda+\eta} \simeq \alpha_{\lambda} \alpha_{\eta}.
\]

Properties of the auxiliary umbra \(n.\alpha\) have been extensively described in Di Nardo and Senato (2001) and these will be recalled whenever it is necessary. It is interesting to remark that \(\alpha.n \equiv n\alpha\), as proved in Di Nardo and Senato (2006a), in agreement with the meaning of the dot product.

A feature of the classical umbral calculus is the construction of new auxiliary umbras by suitable symbolic replacements. For example, if we replace the integer \(n\) in \(n.\alpha\) by an umbra \(\gamma\), equivalence (2.3) gives

\[
(2.6) \quad (\gamma.\alpha)^i \simeq \sum_{\lambda \vdash i} (\gamma)_{\nu_{\lambda}} d_{\lambda} \alpha_{\lambda}.
\]

Equivalence (2.6) has been formally proved by using the notion of generating function of an umbra, for further details see Di Nardo and Senato (2001). Note that, contrary to what happens with \(n.\alpha\), in the dot product \(\alpha.n\) the substitution of \(n\) with an umbra \(\gamma\) does not inherit the symbolic expression of moments. As it is straightforward to show via (2.6), the dot product is therefore not commutative. This circumstance justifies also the falling off of the right distributive law in the dot product, so that

\[
(\alpha + \delta).\gamma \equiv \alpha.\gamma + \delta.\gamma \quad \text{whereas} \quad \gamma.(\alpha + \delta) \not\equiv \gamma.\alpha + \gamma.\delta,
\]

for \(\alpha, \gamma, \delta \in A\). Actually, by considering the parallelism with the r.v.’s theory, the dot product \(\gamma.\alpha\) corresponds to a random sum, and the right distributive law falls off similarly to what it happens for random sums.

In the dot product \(\gamma.\alpha\), by replacing the umbra \(\gamma\) by the umbra \(\gamma.\beta\), we obtain the so-called composition umbra of \(\alpha\) and \(\gamma.\beta\), that is \(\gamma.\beta.\alpha\), whose powers are

\[
(2.7) \quad (\gamma.\beta.\alpha)^i \simeq \sum_{\lambda \vdash i} (\gamma.\beta)_{\nu_{\lambda}} d_{\lambda} \alpha_{\lambda}.
\]

The compositional inverse of an umbra \(\alpha\) is the umbra \(\alpha^{-1} \) satisfying

\[
\alpha^{-1}. \beta.\alpha \equiv \alpha.\beta.\alpha^{-1} \equiv \chi.
\]

\(^2\)Recall that a partition of an integer \(i\) is a sequence \(\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_t)\), where \(\lambda_j\) are weakly decreasing integers and \(\sum_{j=1}^t \lambda_j = i\). The integers \(\lambda_j\) are named parts of \(\lambda\). By the symbol \(\nu_{\lambda}\) we denote the length of \(\lambda\), that is the number of its parts. A different notation is \(\lambda = (1^{r_1}, 2^{r_2}, \ldots)\), where \(r_j\) is the number of parts of \(\lambda\) equal to \(j\) and \(r_1 + r_2 + \cdots = \nu_{\lambda}\). We use the classical notation \(\lambda \vdash i\) to denoting “\(\lambda\) is a partition of \(i\).”
In the following examples, some other fundamental auxiliary umbrae are characterized by means of equivalence (2.6). The properties we are going to recall are proved in Di Nardo and Senato (2006a,b).

**Example 2.1.** The $\alpha$-partition umbra. If $\beta$ is the Bell umbra, the umbra $\beta.\alpha$ is called the $\alpha$-partition umbra. By taking into account (2.6) and (2.2), its powers are

\[(\beta.\alpha)^i \simeq \sum_{\lambda \vdash i} d_{\lambda} \alpha_{\lambda}.\]

By equivalence (2.8), we have

\[(\beta.\alpha)^{\nu_{\alpha}} \equiv \chi, \beta.\chi \equiv u,\]

where $u^{\nu_{\alpha}}$ denotes the compositional inverse of $u$. The umbra $\beta.\alpha$ corresponds to a compound Poisson r.v. of parameter 1.

**Example 2.2.** The $\alpha$-cumulant umbra. If $\chi$ is the singleton umbra, the umbra $\chi.\alpha$ is called the $\alpha$-cumulant umbra. By virtue of equivalence (2.6), its powers are

\[(\chi.\alpha)^i \simeq \sum_{\lambda \vdash i} x_{(\nu_{\alpha})} d_{\lambda} \alpha_{\lambda},\]

where $x_{(\nu_{\alpha})}$ are the factorial moments (2.1) of the umbra $\chi$. By taking into account (2.10), the following equivalences follow

\[\chi.\beta \equiv u, \chi.\chi \equiv u^{\nu_{\chi}}.\]

Equivalence (2.10) recalls the well-known expression of cumulants $\kappa_1, \kappa_2, \ldots$ in terms of moments $a_1, a_2, \ldots$ of a r.v.

\[\kappa_i = \sum_{\lambda \vdash i} (-1)^{\nu_{\lambda}-1}(\nu_{\lambda} - 1)! d_{\lambda} a_{\lambda}.\]

The moments of the $\alpha$-cumulant umbra $\chi.\alpha$ are therefore called *cumulants* of the umbra $\alpha$.

**Example 2.3.** The $\alpha$-factorial umbra. The umbra $\alpha.\chi$ is called the $\alpha$-factorial umbra and its moments are the factorial moments of $\alpha$, since the following equivalence holds

\[(\alpha.\chi)^i \simeq (\alpha)^i.\]

The commutative integral domain $R$ may be replaced by a polynomial ring in any number of indeterminates, having coefficient in a field $K$ of characteristic zero. Suppose therefore to replace $R$ with the polynomial ring $K[y]$, where $y$ is an indeterminate. The uncorrelation property iii) must be rewritten as

\[E[1] = 1; \quad E[y^j \alpha^k \beta^l \cdots] = y^j E[\alpha^k] E[\beta^l] \cdots\]

for any set of distinct umbrae in $A$, and nonnegative integers $j, k, l, \ldots$. In $K[y][A]$ an umbra is said to be a *scalar umbra* when its moments are elements of $K$, while it is said to be a *polynomial umbra* if its moments are polynomials of $K[y]$. A sequence of polynomials $p_0, p_1, \ldots \in K[y]$ is umbrally represented by a polynomial umbra if and only if $p_0 = 1$ and $p_n$ is of degree $n$ for every nonnegative integer $n$. If we replace the integer $n$ by the indeterminate $y$ in (2.5), then

\[E[(y.\alpha)^i] = \sum_{\lambda \vdash i} (y)_{\nu_{\lambda}} d_{\lambda} a_{\lambda},\]
where \((y)_\nu\) denotes the lower factorial polynomials in \(K[y]\).

Some other auxiliary umbrae will be used in the following. The symbol \(\alpha^n\) is an auxiliary umbra denoting the product \(\alpha_1 \alpha_2 \cdots \alpha_n\), where \(\{\alpha_1, \alpha_2, \ldots, \alpha_n\}\) are similar but uncorrelated umbrae. Moments of \(\alpha^n\) can be easily recovered from its definition. Indeed, if the umbra \(\alpha\) represents the sequence 1, \(a_1, a_2, \ldots\), then

\[
E[(\alpha^n)_k] = a_n^k,
\]

for nonnegative integers \(k\) and \(n\). The umbra \(\gamma\) is said to be multiplicative inverse of the umbra \(\alpha\) if and only if \(\alpha \gamma \equiv 1\). Recall that, in dealing with a saturated umbral calculus, the multiplicative inverse of an umbra is not unique, but any two multiplicative inverses of the same umbra are similar. From the definition, it follows:

\[
a_n g_n = 1 \quad \forall n = 0, 1, 2, \ldots \quad \text{that is} \quad g_n = \frac{1}{a_n},
\]

where \(a_n\) and \(g_n\) are moments of \(\alpha\) and \(\gamma\) respectively. The multiplicative inverse of an umbra \(\alpha\) should be denoted by \(\alpha^{-1}\), but in order to simplify the notation and in agreement with our intuition, in the following we will use the symbol \(1/\alpha\).

### 3. \(k\)-statistics via compound Poisson r.v.’s

In this section we resume previous results of the authors (Di Nardo et al., 2008b), useful to simplify the subsequent reading.

The \(i\)-th \(k\)-statistic \(k_i\) is the unique symmetric unbiased estimator of the cumulant \(\kappa_i\) of a given statistical distribution (Stuart and Ord, 1987), that is \(E[k_i] = \kappa_i\). By virtue of (2.11), in umbral terms we shall write

\[
k_i \simeq (\chi; \alpha)^i.
\]

In Di Nardo et al. (2008b), \(k\)-statistics have been related to cumulants of compound Poisson r.v.’s by the following theorem.

**Theorem 3.1.** If \(c_i(y) = E[(n; \chi; y; \beta; \alpha)_i^i], i = 1, 2, \ldots\) then

\[
(\chi; \alpha)^i \simeq c_i \left(\frac{\chi \chi}{n \chi}\right).
\]

The statement of Theorem 3.1 requires some more remarks. As stated in Di Nardo and Senato (2006a), the umbra

\[
(\chi; y; \beta; \alpha) \equiv \chi; (y; \beta; \alpha)
\]

is the cumulant umbra of a polynomial \(\alpha\)-partition umbra, the latter corresponding to a compound Poisson r.v. of parameter \(y\). The polynomial umbra

\[
n.(\chi; y; \beta; \alpha) \equiv n; (\chi; y; \beta; \alpha),
\]

is therefore the sum of \(n\) uncorrelated cumulant umbrae of a polynomial \(\alpha\)-partition umbra. Thus, Theorem 3.1 states that cumulants of \(\alpha\) can be recovered from the moments of \(n.(\chi; y; \beta; \alpha)\), by a suitable replacement of the indeterminate \(y\).

Usually \(k\)-statistics are expressed in terms of the \(r\)-th powers of the data points \(S_r = \sum_{t=1}^{r} X_t^r\). In order to recover this expression for \(k\)-statistics in umbral terms, it is sufficient to express the polynomials \(c_i(y)\) in terms of power sums \(n.\alpha^r = \alpha_1^r + \cdots + \alpha_n^r\). To this aim, the starting point is to express the moments of a generic umbra such as \(n.(\gamma; \alpha)\), with \(\gamma \in A\), in terms of \(r\)-th power sums \(n.\alpha^r\).
Theorem 3.2. If $\alpha, \gamma \in A$, then
\begin{equation}
(n.\gamma\alpha)^i \approx \sum_{\lambda \vdash i} d_{\lambda}(\chi,\gamma)\lambda(n.\alpha)^{r_1}(n.\alpha^2)^{r_2} \cdots
\end{equation}
with $\lambda = (1^{r_1}, 2^{r_2}, \ldots)$.

Equivalence (3.2) is the way for expressing the polynomials $c_i(y)$, umbrally equivalent to moments of $n.\chi.y.\beta.\alpha$, in terms of $r$-th power sums $n.\alpha^r$. Indeed, as $n.\chi.y.\beta.\alpha \equiv n.([\chi.y.\beta]_{\alpha})$ (see (31) in Di Nardo and Senato 2006a), we can use equivalence (3.2), with $\gamma$ replaced by $([\chi.y.\beta]_{\alpha})$. This is the starting point to prove the following result, by which the fast algorithm for $k$-statistics can be easily recovered.

Theorem 3.3. In $K[y]$, let
\begin{equation}
p_n(y) = \sum_{k=1}^{n} (-1)^{k-1}(k-1)!S(n,k)y^k,
\end{equation}
where $S(n,k)$ are the Stirling numbers of second type. For every $\alpha \in A$ we have
\begin{equation}
(\chi.\alpha)^i \approx \sum_{\lambda \vdash i} d_{\lambda}(\chi,\chi)\lambda(n.\alpha)^{r_1}(n.\alpha^2)^{r_2} \cdots
\end{equation}
with $\lambda = (1^{r_1}, 2^{r_2}, \ldots)$ and $p_{\lambda}(y) = [p_1(y)]^{r_1}[p_2(y)]^{r_2} \cdots$.

For the proofs of Theorems 3.2 and 3.3 see (Di Nardo et al., 2008b).

3.1. Polykays via compound Poisson r.v.'s. The symmetric statistic $k_{r,\ldots,t}$ satisfying
\[E[k_{r,\ldots,t}] = \kappa_r \cdots \kappa_t,
\]
where $\kappa_r, \ldots, \kappa_t$ are cumulants, generalizes $k$-statistics and these were originally called generalized $k$-statistics by Dressel (1940). Later they were called polykays by Tukey (1950).

As a product of uncorrelated cumulants, the umbral expression for a polykay is simply
\begin{equation}
k_{r,\ldots,t} \approx (\chi.\alpha)^r \cdots (\chi'.\alpha')^{t},
\end{equation}
with $\chi, \ldots, \chi'$ being uncorrelated singleton umbrae, and $\alpha, \ldots, \alpha'$ uncorrelated umbrae satisfying $\alpha \equiv \cdots \equiv \alpha'$.

Also polykays are usually expressed in terms of power sums. In Di Nardo et al. (2008a), starting from (3.5), we have given a compressed umbral formula in order to express polykays in terms of power sums. Such a formula has been implemented in Maple and the resulting computational times have been presented and discussed in Di Nardo et al. (2008a). Despite the compressed expression for this umbral formula, the computational cost of the resulting algorithm involves the Bell numbers and so increases too rapidly with $r+t$. A different umbral formula may be constructed by generalizing the results of the previous section. Such a formula is not quite expressible in a compressed form, but speeds up the algorithm for building polykays (see the computational times in Table 1 in Section 6).

For plainness, in the following we just deal with two subindexes $k_{r,t}$, the generalization to more than two being straightforward.
Let us consider a polynomial umbra whose moments are all equal to \( y \), up to an integer \( k \), after which the moments are all zero. Let us therefore define the umbra \( \delta_{y,k} \) satisfying
\[
\delta_{y,k} \simeq \begin{cases} 
(\chi y, \beta) & i = 0, 1, 2, \ldots, k, \\
0 & i > k.
\end{cases}
\]

**Lemma 3.4.** Let \( r, t \) be two nonnegative integers. If \( k = \max\{r, t\} \), then
\[
[n.(\delta_{y,k} \alpha)]^{r+t} \simeq \sum_{(\lambda \vdash r, \eta \vdash t)} y^{\nu_{\lambda} + \nu_{\eta}} (n)_{\nu_{\lambda} + \nu_{\eta}} d_{\lambda + \eta} \alpha_{\lambda + \eta}.
\]

**Proof.** By equivalence (2.5), we have
\[
[n.(\delta_{y,k} \alpha)]^{r+t} \simeq \sum_{\xi \vdash (r+t)} (n)_{\nu_{\xi}} d_{\xi} (\delta_{y,k} \xi) \alpha_{\xi},
\]
since \( \delta_{y,k} \) and \( \alpha \) are uncorrelated. The result follows by observing that \( (\delta_{y,k} \xi) \simeq 0 \) for any \( \xi \) satisfying \( \lambda + \eta < \xi \), where \( \lambda \vdash r, \eta \vdash t \) and \( < \) represents the lexicography order on integer partitions. \( \square \)

Let \( p_{r,t}(y) \) be the polynomial obtained by evaluating the right hand side of (3.6), that is
\[
p_{r,t}(y) = \sum_{(\lambda \vdash r, \eta \vdash t)} y^{\nu_{\lambda} + \nu_{\eta}} (n)_{\nu_{\lambda} + \nu_{\eta}} d_{\lambda + \eta} \alpha_{\lambda + \eta}.
\]
The proof of the following theorem is straightforward and allows us to express products of uncorrelated cumulants by using the polynomials \( p_{r,t}(y) \).

**Theorem 3.5.** If \( q_{r,t} \) is the umbral polynomial obtained via \( p_{r,t}(y) \) by replacing \( y^{\nu_{\lambda} + \nu_{\eta}} \) by
\[
\frac{(\chi' \chi)^{\nu_{\lambda}} (\chi' \chi')^{\nu_{\eta}}}{(n \chi)_{\nu_{\lambda} + \nu_{\eta}} d_{\lambda + \eta}},
\]
then
\[
(\chi \alpha)^r (\chi' \alpha')^t \simeq q_{r,t}.
\]
These two last results are sufficient to express the polykay \( k_{r,t} \) in terms of power sums. The steps are summarized in the following:

i) we apply Theorem 3.2 to the polynomial umbra \( n.(\delta_{y,k} \alpha) \), with \( k = \max\{r, t\} \), in order to link the polynomials \( p_{r,t}(y) \) to power sums, that is
\[
[n.(\delta_{y,k} \alpha)]^{r+t} \simeq p_{r,t}(y) \simeq \sum_{\xi \vdash (r+t)} d_{\xi} (\chi, \delta_{y,k} \xi) (n \alpha)^{s_1} (n \alpha^2)^{s_2} \cdots ;
\]

ii) we evaluate the cumulants of the umbra \( \delta_{y,k} \) by means of (2.10), by recalling that moments corresponding to powers greater than \( k \) are zero;

iii) we replace occurrences of \( y^{\nu_{\lambda} + \nu_{\eta}} \) in (3.8) by (3.7), thanks to Theorem 3.5.

The steps i) – iii) are the building blocks of the fast algorithm for generating polykays.
4. The multivariate case: umbrae indexed by multisets

In order to consider multivariate $k$-statistics and polykays, we need of the notion of multivariate moments and multivariate cumulants of an umbral monomial. The umbral tools necessary to deal with multivariate moments are introduced in Di Nardo et al. (2008b). Here we recall basic notation and equivalences in order to generalize Theorems 3.2 and 3.5 to the multivariate case.

A multiset $M$ is a pair $(\bar{M}, f)$, where $\bar{M}$ is a set, called the support of the multiset, and $f$ is a function from $\bar{M}$ to nonnegative integers. For each $\mu \in \bar{M}$, $f(\mu)$ is the multiplicity of $\mu$. The length of the multiset $(\bar{M}, f)$, usually denoted by $|M|$, is the sum of multiplicities of all elements of $\bar{M}$, that is

$$|M| = \sum_{\mu \in \bar{M}} f(\mu).$$

When the support of $M$ is a finite set, say $\bar{M} = \{\mu_1, \mu_2, \ldots, \mu_k\}$, we will write

$$M = \{\mu_1^{f(\mu_1)}, \mu_2^{f(\mu_2)}, \ldots, \mu_k^{f(\mu_k)}\}, \quad \text{or} \quad M = \{\mu_1, \mu_2, \ldots, \mu_k\}^{f(\mu_1), \ldots, f(\mu_k)}.$$

For example the multiset

$$M = \{\alpha, \ldots, \alpha\} = \{\alpha^{(i)}\}$$

has length $i$, support $\bar{M} = \{\alpha\}$ and $f(\alpha) = i$. Recall that a multiset $M_i = (\bar{M}_i; f_i)$ is a submultiset of $M = (\bar{M}, f)$ if $\bar{M}_i \subseteq \bar{M}$ and $f_i(\mu) \leq f(\mu), \forall \mu \in \bar{M}_i$.

**Example 4.1.** If $M = \{\alpha, \alpha, \gamma, \delta, \delta\}$ then $M_1 = \{\alpha, \alpha\}$ is a submultiset with support $\bar{M}_1 = \{\alpha\}$ and $f_1(\alpha) = 2$. Also $M_2 = \{\alpha, \delta, \delta\}$ is a submultiset with support $\bar{M}_2 = \{\alpha, \delta\}$ and $f_2(\alpha) = 1, f_2(\delta) = 2$.

In the following we set

$$\mu_M = \prod_{\mu \in \bar{M}} \mu^{f(\mu)}, \quad (n.\mu)_M = \prod_{\mu \in \bar{M}} (n.\mu)^{f(\mu)}, \quad [n.(\chi \mu)]_M = \prod_{\mu \in \bar{M}} [n.(\chi \mu)]^{f(\mu)}.$$

For instance, if $M = \{\alpha^{(i)}\}$ then $\alpha_M = \alpha^i$, $(n.\alpha)_M = (n.\alpha)^i$, $[n.(\chi \alpha)]_M = [n.(\chi \alpha)]^i$.

A subdivision of a multiset $M$ is a multiset $S = (\bar{S}, g)$ of $k \leq |M|$ non empty submultisets $M_i = (\bar{M}_i; f_i)$ of $M$ satisfying

i) $\bigcup_{i=1}^k \bar{M}_i = \bar{M}$;

ii) $\sum_{i=1}^k f_i(\mu) = f(\mu)$ for any $\mu \in \bar{M}$.

**Example 4.2.** Multisets $S_1 = \{\{\alpha, \gamma\}, \{\alpha\}, \{\delta, \delta\}\}$ and $S_2 = \{\{\alpha, \gamma, \delta\}, \{\alpha, \delta\}\}$ are subdivisions of $M = \{\alpha, \alpha, \gamma, \delta, \delta\}$.

By extending the notation (1.1), we set

$$\mu_S = \prod_{M_i \in \bar{S}} \mu_{M_i}^{g(M_i)}, \quad (n.\mu)_S = \prod_{M_i \in \bar{S}} (n.\mu_{M_i})^{g(M_i)},$$

$$[n.(\chi \mu)]_S = \prod_{M_i \in \bar{S}} [n.(\chi \mu_{M_i})]^{g(M_i)}.$$
Example 4.3. If \( M = \{\mu_1, \mu_1, \mu_2\} \), then \( S_1 = \{\{\mu_1\}, \{\mu_1, \mu_2\}\} \) is a subdivision of \( M \). The support of \( S_1 \) consists of two multisets, \( M_1 = \{\mu_1\} \) and \( M_2 = \{\mu_1, \mu_2\} \), each of one with multiplicity 1, therefore \( [n.(\chi\mu)]_{S_1} = n.(\chi_{M_1})n.(\chi_{M_2}) \). Since \( n.(\chi_{M_1}) = n.(\chi_{M_2}) \) and \( n.(\chi_{M_2}) = n.(\chi_{\mu_1\mu_2}) \), we have \( [n.(\chi\mu)]_{S_1} = n.(\chi_{\mu_1})n.(\chi_{\mu_1\mu_2}) \).

We may construct a subdivision of the multiset \( M \) by a suitable set partition. Recall that a partition \( \pi \) of a set \( C \) is a collection \( \pi = \{B_1, B_2, \ldots, B_k\} \) with \( k \leq n \) disjoint and non-empty subsets of \( C \) whose union is \( C \). We denote by \( \Pi_n \) the set of all partitions of \( C \). Suppose the elements of \( M \) to be all distinct, build a set partition and then replace each element in any block by the original one. By this way, any subdivision corresponds to a set partition \( \pi \) and we will write \( S_\pi \). Note that \( |S_\pi| = |\pi| \) and it could be \( S_{\pi_1} = S_{\pi_2} \) for \( \pi_1 \neq \pi_2 \), as the following example shows.

Example 4.4. If \( M = \{\alpha, \alpha, \gamma, \delta, \delta\} \) suppose to label each element of \( M \) in such a way \( C = \{\alpha_1, \alpha_2, \gamma_1, \delta_1, \delta_2\} \). The subdivision \( S_1 = \{\{\alpha, \gamma\}, \{\alpha\}, \{\delta, \delta\}\} \) corresponds to the partition \( \pi_1 = \{\{\alpha_1, \gamma_1\}, \{\alpha_2\}, \{\delta_1, \delta_2\}\} \) of \( C \). We have \( |S_1| = |\pi_1| \). Note that the subdivision \( S_1 \) also corresponds to the partition \( \pi_2 = \{\{\alpha_2, \gamma_1\}, \{\alpha_1\}, \{\delta_1, \delta_2\}\} \).

In the following, we denote by \( n_\pi \) the number of set partitions in \( \Pi_{|M|} \) corresponding to the same subdivision \( S \) of the multiset \( M \).

If \( M = \{\alpha^{(i)}\} \), then subdivisions are of type

\[
S = \{\underbrace{\{\alpha\}, \ldots, \{\alpha\}}_{r_1}, \underbrace{\{\alpha^{(2)}\}, \ldots, \{\alpha^{(2)}\}}_{r_2}, \ldots\},
\]

with \( r_1 + 2r_2 + \cdots = i \). The support of \( S \) is \( \tilde{S} = \{\{\alpha\}, \{\alpha^{(2)}\}, \ldots\} \), so that (4.2) and (4.3) give

\[
(n.(\alpha))_S \equiv (n.(\alpha))^{r_1} (n.(\alpha^{(2)}))^{r_2} \cdots \quad \text{and} \quad (n.(\chi\alpha))_S \equiv [n.(\chi\alpha)]^{r_1} [n.(\chi\alpha^{(2)})]^{r_2} \cdots.
\]

Before ending this summary, we recall one more notation. Suppose \( S \) is a subdivision of the multiset \( M \) of type \( S = \{M_1^{(g(M_1))}, M_2^{(g(M_2))}, \ldots, M_j^{(g(M_j))}\} \). By the symbol \( \mu^S \) we denote

\[
\mu^S \equiv (\mu_{M_1})^{g(M_1)} \cdots (\mu_{M_j})^{g(M_j)},
\]

where \( \mu_{M_i} \) are uncorrelated umbral monomials. Observe that also \( \mu^S \) is a multiplicative function, that is if \( S_1 \) and \( S_2 \) are subdivisions of \( M \) then

\[
\mu^{(S_1 + S_2)} \equiv \mu^{S_1} \mu^{S_2},
\]

where \( S_1 + S_2 \) denotes the disjoint union of \( S_1 \) and \( S_2 \). If \( M = \{\alpha^{(i)}\} \), then

\[
\alpha^\lambda \equiv \alpha^S
\]

with \( \lambda = (1^{r_1}, 2^{r_2}, \ldots) \). The notation (4.6) comes in handy in order to evaluate umbral polynomials like \( (n.\mu)_M \) in terms of moments of the umbral monomials running in \( M \). Indeed, observe that a different way to write equivalence (2.3) follows by using subdivisions \( S_\pi \) of \( M = \{\alpha^{(i)}\} \), that is

\[
(n.(\alpha))^{(S)} \simeq \sum_{\pi \in \Pi_n} (n.(\chi))^{S_\pi} \alpha^{S_\pi}.
\]
Replace the multiset $M = \{\alpha^{(i)}\}$ by a generic multiset $M$, then it follows

$$\begin{equation}
(n, \mu)_M \simeq \sum_{\pi \in \Pi_i} (n, \chi)^{|\pi_\alpha|} \mu^{|\pi_\beta|} y^{|\pi_\gamma|}.
\end{equation}
$$

We end the section by adding one more remark. Let $N$ be a submultiset of $M$. By the symbol $(\mu_M)_N$ we denote the monomial umbra $\mu_{M\cap N}$. This notation allows us to generalize equivalence (4.10) to umbral polynomial $(n, \mu_M)_N$ with $N \subset M$, that is

$$\begin{equation}
(n, \mu)_N \simeq \sum_{\pi \in \Pi_{N,i}} (n, \chi)^{|\pi_\alpha|}(\mu_M)^{|\pi_\beta|}.
\end{equation}
$$

## 5. Multivariate $k$-statistics via compound Poisson r.v.’s

In Di Nardo et al. (2008), multivariate moments and multivariate cumulants of an umbral monomial are introduced. Let $M = \{\mu_1^{(f(\mu_1))}, \mu_2^{(f(\mu_2))}, \ldots, \mu_r^{(f(\mu_r))}\}$ be a multiset of length $i$. A multivariate moment is the element of $K[y]$ corresponding to the umbral monomial $\mu_M$ via evaluation $E$, that is

$$E[\mu_M] = m_{t_1, \ldots, t_r},$$

where $t_j = f(\mu_j)$ for $j = 1, 2, \ldots, r$. The corresponding multivariate cumulant is the element of $K[y]$ satisfying

$$E[(\chi, \mu)_M] = \kappa_{t_1, \ldots, t_r}.$$

For example, if $M = \{\alpha^{(i)}\}$ then $(\chi, \mu)_M \simeq (\chi, \alpha)^i$. As for $k$-statistics, the umbra $(\chi, y, \beta, \mu)_M$ is the cornerstone for building efficiently multivariate $k$-statistics. Let us observe that the evaluation of $(\chi, y, \beta, \mu)_M$ gives the umbral counterpart of joint cumulants of a multivariate compound Poisson r.v. with parameter $y$. We need therefore to characterize the evaluation of $[n, (\chi, y, \beta, \mu)]_M$.

**Proposition 5.1.** Let $M$ be a multiset of length $i$. The umbra $[n, (\chi, y, \beta, \mu)]_M$ is umbrally equivalent to the umbral polynomial

$$c_M(y) = \sum_{\pi \in \Pi_i} (n, \chi)^{|\pi_\alpha|} y^{|\pi_\beta|} \mu^{|\pi_\gamma|},$$

where $S_\pi$ are the subdivisions of the multiset $M$ corresponding to the partitions $\pi$.

**Proof.** In equivalence (1.9), replace the generic umbral monomial $\mu$ by $\chi, y, \beta, \mu$. We have

$$[n, (\chi, y, \beta, \mu)]_M \simeq \sum_{\pi \in \Pi_i} (n, \chi)^{|\pi_\alpha|}[(\chi, y, \beta, \mu)]^{|\pi_\beta|},$$

where the form on the right-hand side is worked out by means of the equivalence $\chi, y, \beta, \mu \equiv (\chi, y, \beta)_M$. The umbra $(\chi, y, \beta)$ and $\mu$ are uncorrelated, so that

$$[(\chi, y, \beta)]^{|\pi_\beta|} \equiv (\chi, y, \beta)^{|\pi_\beta|},$$

Let $S_\pi = \{M_1^{g(M_1)}, M_2^{g(M_2)}, \ldots, M_j^{g(M_j)}\}$. Equivalence (5.2) follows from (5.3), by observing that

$$(\chi, y, \beta)^{|\pi_\beta|} \simeq y^{|\pi_\beta|},$$

since the umbra $\chi, y, \beta$ has moments all equal to $y$, and $\sum g(M_i) = |S_\pi|$. □
Theorem 5.2. If \( c_M(y) \) are the umbral polynomials given in \([5.2]\), then
\[
(5.4) \quad c_M\left(\frac{\chi \cdot \chi}{n \cdot \chi}\right) \simeq (\chi \cdot \mu)_M.
\]

Proof. The result follows directly from \([5.2]\) by replacing \( y \) by \( \frac{\chi \cdot \chi}{n \cdot \chi} \) and by recalling that
\[
(5.5) \quad (\chi \cdot \mu)_M \simeq \sum_{\pi \in \Pi} (\chi \cdot \chi)^{|S_\pi|} \mu^{S_\pi},
\]
an equivalence which follows from \([4.9]\) by replacing \( n \) by \( \chi \). \( \square \)

Theorem 5.3. Let \( p_{\pi}(x) = [p_1(x)]^{r_1}[p_2(x)]^{r_2} \cdots \) with \( p_n(x) \) given in \([3.3]\) and \( \pi \) a partition of \( \Pi_{|M|} \) with \( r_1 \) blocks of cardinality 1, \( r_2 \) blocks of cardinality 2, and so on, then
\[
(5.6) \quad (\chi \cdot \mu)_M \simeq \sum_{\pi \in \Pi} p_{\pi}(\chi \cdot \chi)^{n \cdot \mu} S_\pi.
\]

Proof. First observe that \([n \cdot (\chi \cdot y \cdot \beta \cdot \mu)]_M \equiv [n \cdot ((\chi \cdot y \cdot \beta) \mu)]_M\), so that
\[
c_M(y) \simeq [n \cdot ((\chi \cdot y \cdot \beta) \mu)]_M.
\]
We need to express \([n \cdot ((\chi \cdot y \cdot \beta) \mu)]_M \) in terms of power sums. To this aim, note that, by using \([4.5]\) and \([4.7]\), equivalence \([3.2]\) can be rewritten as
\[
(5.7) \quad [n \cdot (\gamma \alpha)]_M \simeq \sum_{\pi \in \Pi} (\chi \cdot \gamma)^{S_\pi} (n \cdot \alpha) S_\pi,
\]
where \( S_\pi \) is a subdivision of the multiset \( M = \{\alpha^{(i)}\} \). Replace \( M \) by any multiset. Equivalence \([5.7]\) becomes
\[
(5.8) \quad [n \cdot (\gamma \alpha)]_M \simeq \sum_{\pi \in \Pi} (\chi \cdot \gamma)^{S_\pi} (n \cdot \alpha) S_\pi.
\]
In equivalence \([5.8]\), suppose to replace \( \gamma \) by \( \chi \cdot y \cdot \beta \), then
\[
(5.9) \quad \sum_{\pi \in \Pi} (\chi \cdot \chi \cdot y \cdot \beta)^{S_\pi} (n \cdot \mu) S_\pi.
\]
Let \( S_n = \{M_1^{(g(M_1))}, M_2^{(g(M_2))}, \ldots, M_j^{(g(M_j))}\} \) then
\[
(\chi \cdot \chi \cdot y \cdot \beta)^{S_\pi} \equiv [(\chi \cdot \chi \cdot y \cdot \beta)_{M_1}]^{g(M_1)} \cdots [(\chi \cdot \chi \cdot y \cdot \beta)_{M_j}]^{g(M_j)}.
\]
Observe that
\[
(\chi \cdot \chi \cdot y \cdot \beta)_{M_i} = \prod_{\mu \in \Pi_i} (\chi \cdot \chi \cdot y \cdot \beta)^{f(\mu)} = (\chi \cdot \chi \cdot y \cdot \beta)|_{M_i},
\]
so that
\[
E[(\chi \cdot \chi \cdot y \cdot \beta)^{S_n}] = [p_{M_1}(y)]^{g(M_1)} \cdots [p_{M_j}(y)]^{g(M_j)} = p_\pi(y)
\]
if \( S_\pi \) is the subdivision corresponding to the partition \( \pi \). By replacing \( y \) by \( (\chi \cdot \chi)/(n \cdot \chi) \), equivalence \([5.6]\) is proved. \( \square \)

Recall that the multivariate \( k \)-statistics are the unique symmetric unbiased estimators of joint cumulants. Since these estimators are unbrally equivalent to \((\chi \cdot \mu)_M\), with a suitable choice of the multiset \( M \), the expression for multivariate \( k \)-statistics in terms of power sums is given by the right-hand side of equivalence \([5.6]\).
5.1. Multivariate polykays via compound Poisson r.v.’s. The symmetric statistic \( k_{t_1, \ldots, t_r; l_1, \ldots, l_m} \) satisfying
\[
E[k_{t_1, \ldots, t_r; l_1, \ldots, l_m}] = \kappa_{t_1, \ldots, t_r} \cdots \kappa_{l_1, \ldots, l_m},
\]
where \( \kappa_{t_1, \ldots, t_r}, \ldots, \kappa_{l_1, \ldots, l_m} \) are multivariate cumulants, generalizes polykays. As product of uncorrelated multivariate cumulants, the umbral expression for a multivariate polykay is simply
\[
(5.10) \quad k_{t_1, \ldots, t_r; l_1, \ldots, l_m} \simeq (\chi_m T) \cdots (\chi r L),
\]
with \( \chi, \ldots, \chi' \) being uncorrelated singleton umbrae and \( T, \ldots, L \) multisets of umbral monomials such that
\[
T = \{ \mu_1^{(t_1)}, \ldots, \mu_r^{(t_r)} \}, \ldots, L = \{ \mu_1^{(l_1)}, \ldots, \mu_m^{(l_m)} \}.
\]
Also for multivariate polykays we have given a compressed umbral formula in terms of multivariate power sums \((Di \ Nardo \ et \ al. \ 2008)\). Such a formula has been implemented in \texttt{Maple} and the resulting computational times have been presented and discussed in \cite{Di Nardo et al. 2008}. Here we generalize the procedure given for univariate polykays, speeding up the algorithm.

For plainness, in the following we just deal with two multisets \( T \) and \( L \), the generalization being straightforward.

Let \( N \) be the disjoint union of all submultisets respectively of \( T \) and \( L \) and suppose to denote by + the disjoint union of two multisets.

**Example 5.1.** Let \( T = \{ \mu_1, \mu_2 \} \) and \( L = \{ \mu_1 \} \). The disjoint union of all submultisets respectively of \( T \) and \( L \) is \( N = \{ \{ \mu_1, \mu_2 \}, \{ \mu_1 \} \} \). If \( T = \{ \mu_1, \mu_2 \} \) and \( L = \{ \mu_3 \} \) then \( N = \{ \{ \mu_1, \mu_2 \}, \{ \mu_1 \}, \{ \mu_2 \}, \{ \mu_3 \} \} \).

As before, we need of a polynomial umbra, indexed by a suitable multiset, which behaves as a filter on subdivisions of \( T + L \), by deleting those which are not disjoint unions of subdivisions respectively of \( T \) and \( L \). Suppose therefore \( M_i = (M_i, g) \), a submultiset of \( T + L \). Let us define the umbra \( \delta_{y,N} \) satisfying
\[
(5.11) \quad (\delta_{y,N})_{M_i} \simeq \left\{ \begin{array}{cl} 0 & \text{if } M_i \not\subset N, \\ (\chi, y, \beta)_{M_i} \simeq (\chi y, \beta)^{|M_i|} \simeq y & \text{otherwise.} \end{array} \right.
\]
Let \( S_r \) be a subdivision of \( T + L \), and \( S_\pi \) and \( S_r \) subdivisions respectively of \( T \) and \( L \), without taking into account the distinct labels. Via (5.11), the following equivalence results
\[
(5.12) \quad (\delta_{y,N})_{S_r} \simeq \left\{ \begin{array}{cl} 0 & \text{if } S_r \not\subset S_\pi + S_r, \\ (\chi y, \beta)^{|S_r|} \simeq y^{|S_r|} & \text{otherwise,} \end{array} \right.
\]
where < denotes the natural extension to subdivisions of the refinement relation defined on the lattice of set partitions.

**Lemma 5.4.** If \( \delta_{y,N} \) is the umbra defined in (5.11), then
\[
(5.13) \quad [n, (\delta_{y,N} \mu)]_{T + L} \simeq \sum_{(\pi \in \Pi[T], \tau \in \Pi[L])} (n, \chi)^{|S_\pi| + |S_r| + |S_\pi|} \mu^{(S_\pi + S_r)}.
\]

**Proof.** From (4.10), we have
\[
(5.14) \quad [n, (\delta_{y,N} \mu)]_{T + L} \simeq \sum_{\nu \in \Pi[T + L]} (n, \chi)^{|S_r|} \delta_{y,N}^{|S_r|} \mu^{|S_r|},
\]
since $\delta_{y,N}$ is uncorrelated with any element of $T + L$. Due to (5.12), in the sum on the right hand side of (5.14), the addends which give a non-zero contribution are only those corresponding to subdivisions which can be split in a subdivision of $T$ and a subdivision of $L$, that is

$$
(5.15) \quad [n. \langle \delta_{y,N} \mu \rangle]_{T+L} \simeq \sum_{(\pi \in \Pi|T|, \tau \in \Pi|L|)} (n,\chi)_{S_\pi + |S_\tau|} \delta_{y,N}^{S_\pi} \delta_{y,N}^{S_\tau} \mu^{S_\pi} \mu^{S_\tau}.
$$

Observing that $\delta_{y,N}^{S_\pi} \simeq y^{S_\pi}$ and $\delta_{y,N}^{S_\tau} \simeq y^{S_\tau}$, the result follows immediately since $\mu^{S_\pi} \mu^{S_\tau} \simeq \mu^{(S_\pi + S_\tau)}$, due to (5.12).

By recalling that subdivisions corresponding to different partitions can be equal, equivalence (5.13) may be rewritten as

$$
(5.16) \quad [n. \langle \delta_{y,N} \mu \rangle]_{T+L} \simeq \sum_{(S_\pi, S_\tau)} n_{\pi + \tau} (n,\chi)_{S_\pi + |S_\tau|} y^{S_\pi + |S_\tau|} \mu^{(S_\pi + S_\tau)},
$$

where $n_{\pi + \tau}$ is the number of set partition pairs $(\pi, \tau)$ corresponding to subdivision $S_\pi + S_\tau$. Assume

$$
(5.17) \quad p_{T,L}(y) = \sum_{(S_\pi, S_\tau)} n_{\pi + \tau} (n,\chi)_{S_\pi + |S_\tau|} y^{S_\pi + |S_\tau|} \mu^{(S_\pi + S_\tau)}.
$$

Thanks to (5.16), the next theorem is proved by simple calculations and allows us to express products of uncorrelated multivariate cumulants by using the polynomials $p_{T,L}(y)$.

**Theorem 5.5.** Suppose $n_\pi$ (respectively $n_\tau$) the number of set partitions in $\Pi|T|$ (respectively $\Pi|L|$) corresponding to the subdivision $S_\pi$ (respectively $S_\tau$) the number of set partitions in $\Pi|T+L|$ corresponding to the subdivision $S_\pi + S_\tau$. If $q_{T,L}$ is the umbral polynomial obtained from $p_{T,L}(y)$ by replacing $y^{S_\pi + |S_\tau|}$ with

$$
(5.18) \quad \frac{(\chi - \chi')(\chi - \chi')^{|S_\tau|} n_{\pi + \tau}}{(n,\chi)_{S_\pi + |S_\tau|} n_{\pi + \tau}},
$$

then

$$
(\chi - \chi)(\chi - \chi')(L) \simeq q_{T,L}.
$$

**Proof.** Due to (5.5), product of multivariate cumulants may be written as:

$$
(\chi - \chi)(\chi - \chi')(L) \simeq \sum_{(\pi \in \Pi|T|, \tau \in \Pi|L|)} (\chi - \chi)_{S_\pi} (\chi - \chi')_{S_\tau} \mu^{(S_\pi + S_\tau)}.
$$

The previous equivalence can be rewritten as

$$
(5.19) \quad (\chi - \chi)(\chi - \chi')(L) \simeq \sum_{(S_\pi, S_\tau)} n_{\pi + \tau} (\chi - \chi)_{S_\pi} (\chi - \chi')_{S_\tau} \mu^{(S_\pi + S_\tau)}.
$$

The result follows by comparing the right hand side of (5.19) with $p_{T,L}(y)$ in (5.17) where $y^{S_\pi + |S_\tau|}$ has been replaced by (5.18). $$\square$$

Via equivalence (5.18), the following equivalence holds

$$
(5.20) \quad [n. \langle \delta_{y,N} \mu \rangle]_{T+L} \simeq \sum_{\nu \in \Pi|T+L|} (\chi - \chi)(\chi - \chi')_{S_\nu} (n,\mu)_{S_\nu},
$$

by which it is possible to express multivariate polykays in terms of power sums. The algorithm is summarized in the following:
by equivalence (5.20), we evaluate \( n(\delta_{y,N} \mu) \) in terms of power sums in order to link the polynomials \( p_{L,T}(y) \) to power sums;

ii) we evaluate the cumulants of the umbra \( \delta_{y,N} \) by means of

\[
\langle \chi.\delta_{y,N} \rangle_M \simeq \sum_{\pi \in \Pi_i} \langle \chi.\chi \rangle |S_{\pi}| \delta_{y,N}^{S_{\pi}}
\]

which is an obvious generalization of equivalence (5.3);

iii) we replace occurrences of \( y|S_{\pi} \rangle + |S_{\tau} \rangle \) in \( \langle \chi.\delta_{y,N} \rangle_M \) by (5.18).

### 6. Computational comparisons

Tables 1 and 2 show comparisons of computational times among four different software packages. The first one, which we call AS algorithms, has been implemented in \texttt{Mathematica} and refers to procedures explained in \cite{AndrewsStafford2000}, see http://fisher.utstat.toronto.edu/david/SCSI/chap.3.nb. The second one refers to the package \texttt{MathStatica} \cite{RoseSmith2002}. Note that in this package, there are no procedures devoted to multivariate polykays. The third package, named \texttt{Fast} algorithms, has been implemented in \texttt{Maple 10.x} by using the results of this paper. The last procedure, named \texttt{Polyk}, has been described in \cite{DiNardo2008}. Let us remark that, for all the considered procedures, the results are in the same output form and have been performed by the authors on the same platform. To the best of our knowledge, there is no R implementation for \( k \)-statistics and polykays.

#### Table 1. Comparison of computational times in sec. for \( k \)-statistics and polykays. Missed computational times “means greater than 20 hours”.

| \( k_{1,...,l} \) | AS Algorithms | MathStatica | Fast-algorithms | Polyk-algorithm |
|-----------------|---------------|-------------|----------------|-----------------|
| \( k_5 \)       | 0.06          | 0.01        | 0.01           | 0.08            |
| \( k_7 \)       | 0.31          | 0.02        | 0.01           | 0.03            |
| \( k_9 \)       | 1.44          | 0.04        | 0.01           | 0.16            |
| \( k_{11} \)    | 8.36          | 0.14        | 0.01           | 0.23            |
| \( k_{14} \)    | 396.39        | 0.64        | 0.02           | 1.33            |
| \( k_{16} \)    | 57982.40      | 2.03        | 0.08           | 4.25            |
| \( k_{18} \)    | -             | 6.90        | 0.16           | 13.70           |
| \( k_{20} \)    | -             | 25.15       | 0.33           | 42.26           |
| \( k_{22} \)    | -             | 81.70       | 0.80           | 172.59          |
| \( k_{24} \)    | -             | 359.40      | 1.62           | 647.56          |
| \( k_{26} \)    | -             | 1581.05     | 2.51           | 3906.19         |
| \( k_{28} \)    | -             | 6505.45     | 4.83           | 21314.65        |
| \( k_{32} \)    | 0.06          | 0.02        | 0.01           | 0.02            |
| \( k_{44} \)    | 0.67          | 0.06        | 0.02           | 0.06            |
| \( k_{55} \)    | 0.69          | 0.08        | 0.02           | 0.07            |
| \( k_{77} \)    | 34.23         | 0.79        | 0.11           | 0.70            |
| \( k_{77} \)    | 435.67        | 2.52        | 0.26           | 2.43            |
| \( k_{99} \)    | -             | 27.41       | 2.26           | 23.32           |
| \( k_{10,8} \)  | -             | 30.24       | 2.98           | 25.06           |
| \( k_{4,4,4} \) | 34.17         | 0.64        | 0.08           | 0.77            |
The Polyk algorithm, introduced in Di Nardo et al. (2008a), has the advantage to give \( k \)-statistics, multivariate \( k \)-statistics, polykays and multivariate polykays, depending on input parameters. That is one algorithm for the whole matter. The computational times of Polyk are better than those of AS algorithms in all cases. Polyk works better than MathStatica for polykays but is not competitive for \( k \)-statistics. MathStatica has not a procedure for multivariate polykays.

| \( k_{t_1,...,t_r; l_1,...,l_m} \) | AS Algorithms | MathStatica | Fast-algorithms | Polyk-algorithm |
|----------------|---------------|-------------|-----------------|-----------------|
| \( k_{3,2} \) | 0.25          | 0.03        | 0.01            | 0.03            |
| \( k_{4,4} \) | 28.36         | 0.16        | 0.02            | 0.34            |
| \( k_{5,5} \) | 259.16        | 0.55        | 0.06            | 1.83            |
| \( k_{6,5} \) | 959.67        | 1.01        | 0.16            | 4.61            |
| \( k_{6,6} \) | -             | 2.20        | 0.28            | 12.08           |
| \( k_{7,6} \) | -             | 4.01        | 0.53            | 33.22           |
| \( k_{7,7} \) | -             | 8.49        | 1.04            | 95.19           |
| \( k_{8,6} \) | -             | 7.37        | 1.09            | 91.80           |
| \( k_{8,7} \) | -             | 14.92       | 2.19            | 300.60          |
| \( k_{3,3,3} \) | 1180.03       | 0.88        | 0.47            | 2.90            |
| \( k_{4,3,3} \) | -             | 2.00        | 0.40            | 9.26            |
| \( k_{4,4,3} \) | -             | 4.80        | 0.94            | 34.20           |
| \( k_{4,4,4} \) | -             | 13.53       | 2.30            | 155.03          |
| \( k_{1,1,1} \) | 0.05          | -           | 0.01            | 0.01            |
| \( k_{2,1,1} \) | 0.20          | -           | 0.01            | 0.03            |
| \( k_{2,2,1} \) | 1.22          | -           | 0.03            | 0.05            |
| \( k_{2,2,2} \) | 6.30          | -           | 0.08            | 0.09            |
| \( k_{2,2,2} \) | 33.75         | -           | 0.14            | 0.30            |
| \( k_{2,1,2; 2,1} \) | 78.94        | -           | 0.22            | 0.45            |
| \( k_{2,2,1; 1,1} \) | 30.01        | -           | 0.14            | 0.20            |
| \( k_{2,2,2; 1,1} \) | 126.19       | -           | 0.28            | 0.55            |
| \( k_{2,2,2; 2,1} \) | 398.42       | -           | 0.55            | 1.66            |
| \( k_{2,2,2; 2,1} \) | 464.45       | -           | 0.61            | 1.59            |
| \( k_{2,2,2; 2,2} \) | 1387.00      | -           | 1.25            | 5.52            |
| \( k_{2,2,2; 2,2} \) | 3787.41      | -           | 2.91            | 20.75           |

Finally, from Table 1 and 2, it is evident that there is a significant improvement of computational times realized by the Fast-algorithms, compared to the other three packages. The Fast-algorithms are available at the following web page http://www.unibas.it/utenti/dinardo/fast.pdf.

In Table 3, we quote computational times for the \( k \)-statistics, the polykays and the multivariate ones given in Tables 1 and 2, obtained with forthcoming MathStatica release 2, by using Mathematica 6.0, on Mac OS X, with Mac Pro 2.8GHz (Colin Rose, private communication)
Table 3. Computational times in sec., for the $k$-statistics, the polykays and the multivariate ones given in Tables 1 and 2, obtained with forthcoming MathStatica release 2.

| $k_{t,1}$ | MathStatica 2 | $k_{t_1,t_2,t_3,t_4}$ | MathStatica 2 |
|-----------|---------------|------------------------|---------------|
| $k_5$     | 0.008         | $k_{3,2}$              | 0.012         |
| $k_7$     | 0.017         | $k_{4,4}$              | 0.009         |
| $k_9$     | 0.039         | $k_{5,5}$              | 0.345         |
| $k_{11}$  | 0.084         | $k_{6,5}$              | 0.592         |
| $k_{14}$  | 0.329         | $k_{6,6}$              | 1.230         |
| $k_{16}$  | 0.917         | $k_{7,6}$              | 2.107         |
| $k_{18}$  | 2.804         | $k_{7,7}$              | 4.215         |
| $k_{20}$  | 9.363         | $k_{8,6}$              | 3.595         |
| $k_{22}$  | 32.11         | $k_{8,7}$              | 7.359         |
| $k_{3,2}$ | 0.012         | $k_{3,3,3}$            | 0.529         |
| $k_{4,4}$ | 0.044         | $k_{4,3,3}$            | 2.552         |
| $k_{7,5}$ | 0.434         | $k_{4,4,4}$            | 6.926         |
| $k_{7,7}$ | 1.288         | $k_{1,1,1,1}$          | 0.006         |
| $k_{9,9}$ | 11.89         | $k_{2,1,1,1}$          | 0.014         |
| $k_{10,8}$ | 12.39       | $k_{2,2,1,1}$          | 0.038         |
| $k_{4,4,4}$ | 0.359     | $k_{2,2,2,1}$          | 0.085         |
|           |               | $k_{2,2,2,2}$          | 0.200         |
|           |               | $k_{2,1,2,1,1}$        | 0.227         |
|           |               | $k_{2,2,1,1,1}$        | 0.154         |
|           |               | $k_{2,2,2,1,1}$        | 0.413         |
|           |               | $k_{2,2,2,2,1}$        | 0.928         |
|           |               | $k_{2,2,2,2,1}$        | 1.063         |
|           |               | $k_{2,2,2,2,2}$        | 2.622         |
|           |               | $k_{2,2,2,2,2}$        | 6.402         |

References

Andrews D. F. and Stafford J. E. (2000) Symbolic computation for statistical inference. Oxford Statistical Science Series, 21. Oxford University Press, Oxford.

Di Nardo E. and Senato D., (2001) Umbral nature of the Poisson random variables, In Algebraic combinatorics and computer science, H. Crapo, and D. Senato, eds., Springer Italia, Milano, 245–266.

Di Nardo E. and Senato D. (2006) An umbral setting for cumulants and factorial moments, Europ. Jour. Comb., 27, 394–413.

Di Nardo E. and Senato D. (2006) A symbolic method for $k$-statistics, Applied Math. Letters, 19, 968–975.

Di Nardo E., Guarino G. and Senato D. (2008) Maple algorithms for polykays and multivariate polykays, Adv. Appl. Stat, 8, 19-36.

Di Nardo E., Guarino G. and Senato D. (2008) A unifying framework for $k$-statistics, polykays and their multivariate generalizations, Bernoulli, Vol. 14(2), 440-468.

Dressel P. L. (1940) Statistical seminvariants and their estimates with particular emphasis on their relation to algebraic invariants, Ann. Math. Stat., 11, 33–57.

Fisher R. A. (1929) Moments and product moments of sampling distributions, Proc. London Math. Soc. (2), 30, 199–238.
Ferreira P. G., Maguello J. and Silk J. (1997) Cumulants as non-Gaussian qualifiers, Phys. Rev. D, 56, 4592–4603.

Gessel, I. (2003) Applications of the classical umbral calculus, Algebra Universalis, 49, 397-434.

Grossman R. ed. (1989) Symbolic Computation: Applications to Scientific Computing. Frontiers in Applied Mathematics 5, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

McCullagh P. (1987) Tensor methods in statistics. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Rose C. and Smith M. D. (2002) Mathematical Statistics with Mathematica. Springer Verlag, New York, 2002.

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

Speed T. P. (1983) Cumulants and partition lattices, Aust. J. Stat., 25, 378–388.

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

Stuart A. and Ord J. K. (1987) Kendall’s Advanced Theory of Statistics. Vol. 1, Charles Griffin and Company Limited, London.

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

D. Wang and Z. Zheng eds. (2005) Differential equations with symbolic computations. Trends in Mathematics. Birkhäuser Verlag, Berlin.

Zeilberger, D. (2004) Symbolic Moment Calculus I.: Foundations and Permutation Pattern Statistics. Ann. Comb., 8, 369–378.

Zeilberger, D. (2000) The umbral transfer-matrix method. I: Foundations. J. Comb. Theory, Ser. A, 91, No.1-2, 451–463

Zeilberger, D. (2001) The umbral transfer-matrix method. III: Counting animals. New York J. Math., 7, 223–231

Zeilberger, D. (2001) The umbral transfer-matrix method. IV: Counting self-avoiding polygons and walks. Electron. J. Comb., 8, No.1, Research paper R28, 17 p.

Zeilberger, D. (2002) The umbral transfer-matrix method. V: The Goulden-Jackson cluster method for infinitely many mistakes. Integers, 2, Paper A05, 12 p.

Dipartimento di Matematica e Informatica, Università degli Studi della Basilicata, Viale dell’Ateneo Lucano 10, 85100, Potenza, Italy
E-mail address: elvira.dinardo@unibas.it

Medical School, Università Cattolica del Sacro Cuore (Rome branch), Largo Agostino Gemelli 8, I-00168, Roma, Italy.
E-mail address: giuseppe.guarino@rete.basilicata.it

Dipartimento di Matematica e Informatica, Università degli Studi della Basilicata, Viale dell’Ateneo Lucano 10, 85100, Potenza, Italy.
E-mail address: domenico.senato@unibas.it